

Accuracy of Molecular Simulation-Based Predictions of koff Values: A Metadynamics Study

Original

Accuracy of Molecular Simulation-Based Predictions of koff Values: A Metadynamics Study / Capelli, R.; Lyu, W.; Bolnykh, V.; Meloni, S.; Olsen, J. M. H.; Rothlisberger, U.; Parrinello, M.; Carloni, P.. - In: THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS. - ISSN 1948-7185. - 11:15(2020), pp. 6373-6381. [10.1021/acs.jpclett.0c00999]

Availability:

This version is available at: 11583/2844573 since: 2020-09-08T16:00:23Z

Publisher:

ACS

Published

DOI:10.1021/acs.jpclett.0c00999

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

GENERIC -- per es. Nature : semplice rinvio dal preprint/submitted, o postprint/AAM [ex default]

(Article begins on next page)

Accuracy of Molecular Simulation-based Predictions of k_{off} Values: a Metadynamics Study

Riccardo Capelli,^{*,†,‡,¶,§§} Wenping Lyu,^{†,§,||,§§} Viacheslav Bolnykh,^{†,⊥} Simone Meloni,[#] Jógvan Magnus Haugaard Olsen,^{@,Δ} Ursula Rothlisberger,[⊥] Michele Parrinello,^{∇,††,‡‡} and Paolo Carloni^{†,¶¶}

[†]*Computational Biomedicine Section, IAS-5/INM-9, Forschungszentrum Jülich, Wilhelm-Johnen-straße, D-52425 Jülich, Germany*

[‡]*JARA-HPC, Forschungszentrum Jülich, D-52425 Jülich, Germany*

[¶]*Current Address: Department of Applied Science and Technology (DISAT), Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy*

[§]*Current address: School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, 518172, China*

^{||}*Current address: School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, 230026, China*

[⊥]*Laboratory of Computational Chemistry and Biochemistry, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

[#]*Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Ferrara, Via Luigi Borsari 46, I-44121, Ferrara, Italy*

[@]*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

^Δ*Department of Chemistry, Aarhus University, DK-8000, Aarhus C, Denmark*

[∇]*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*

^{¶¶}*JARA-Institute INM-11: Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich, Wilhelm-Johnen-straße, D-52425 Jülich, Germany*

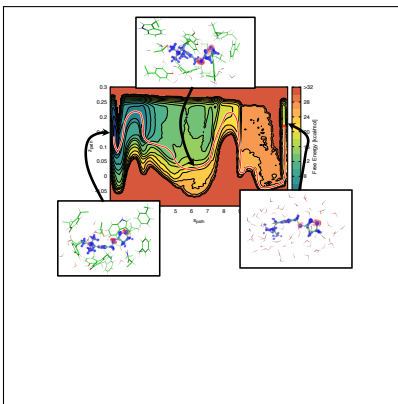
^{§§}*RC and WL share first authorship.*

E-mail: r.capelli@fz-juelich.de;riccardo.capelli@polito.it

Abstract

The k_{off} values of ligands unbinding to proteins are key parameters for drug discovery. Their predictions based on molecular simulation may under- or over-estimate experiment in a system- and/or technique-dependent way. Here we use an established method —infrequent metadynamics, based on the AMBER force field— to compute the k_{off} of the ligand iperoxo (used in the clinics) targeting the muscarinic receptor M_2 . The ligand charges are calculated by either (i) the Amber standard procedure or (ii) B3LYP-DFT. The calculations using (i) turn out not to provide a reasonable estimation of the transition-state free energy. Those using (ii) differ from experiment by two orders of magnitude. Based on B3LYP DFT QM/MM simulations, we suggest that the observed discrepancy in (ii) arises, at least in part, from the lack of electronic polarization and/or charge transfer in biomolecular force fields. These issues might be present in other systems, such as DNA/protein complexes.

Graphical TOC Entry



Keywords

Neuroreceptors, Metadynamics, QM/MM, drugs' residence time

Efficacy and safety of drugs depends critically on their residence time.^{1,2} Indeed, k_{off} values—the drug unbinding rate constant, corresponding to the inverse of the residence time—correlates with clinical efficiency even more than binding affinity.^{3,4} Hence, the k_{off} value is one of the crucial parameters that current drug design strives to improve.^{5,6} While experiments face challenges to identify and characterize rate-limiting transition state(s), simulation approaches are able to predict free energy landscapes and residence times.⁷ Techniques devoted to this aim range from long-time molecular dynamics (MD) with specialized hardware,⁸ to a variety of different enhanced sampling methods such as random acceleration MD (RAMD),^{9,10} hyperdynamics,¹¹ conformational flooding,¹² Markov state models (MSMs),¹³ dissipation-corrected targeted MD¹⁴ and Infrequent¹⁵ or Frequency Adaptive metadynamics.¹⁶ The latter three approaches have also predicted k_{off} values for ligands binding to cytoplasmatic proteins. The values differ by one or two orders of magnitude from experiments.^{13,17–19} The discrepancy, irrespective of the force field used (either Amber99SB²⁰/GAFF²¹ or CHARMM22*²²), could be caused by a variety of factors, including force field accuracy, molecular modeling procedures, and sampling issues. Here, we use a multistep simulation approach to address this important issue. We focus on a ligand, iperoxo (Fig.1 and SI1), routinely used in neuroimaging in the clinics. The ligand targets the human muscarinic acetylcholine receptor 2 (M_2). Overall, the system consists of $\sim 150,000$ atoms (Fig.1).

First, we attempt to calculate the k_{off} value of the ligand by Amber14SB force field²³-based Well-Tempered²⁴ and Frequency-Adaptive MetaD. To minimize errors due to the modeling procedure, we use the same pH and ionic strength as in the experimental conditions.²⁵ We use two approaches to calculate the drug RESP charges. The first one is the Amber standard methodology, based on HF/6-31G calculations (**RESP-HF**).²¹ This has been used to predict the free energy landscape associated with ligand binding to the protein,²⁶ with a calculated binding affinity in excellent agreement with experiment. The second methodology is based on density functional theory (DFT) with the B3LYP^{27–29} exchange-correlation

functional (**RESP-B3LYP**).

Then, we perform quantum mechanics/molecular mechanics (QM/MM) calculations. The QM region consists of the ligand and its interacting residues and it is described by the same B3LYP functional as in the second parametrization. The remaining part is treated as in all the previous calculations performed here, with the Amber14SB force field.²³

The simulations based on **RESP-HF** charges turn out to face difficulties in obtaining a reasonable estimation of the transition-state free energy (and thus k_{off} value). Those with **RESP-B3LYP** charges lead to a calculated k_{off} of $3.7 \pm 0.7 \cdot 10^{-4} \text{ s}^{-1}$. This value is more reasonable but still much smaller than the experimental one ($1.0 \pm 0.2 \cdot 10^{-2} \text{ s}^{-1}$). Comparison of the force-field-based simulations with those based on QM/MM shows a remarkable agreement between the *ab initio* and the force field estimation of the ligand/protein binding energy (Fig.1). However, this is not the case for the transition state. Our analysis indicates that the lack of polarization may be one of the key factors causing this discrepancy, which in turn affects the accuracy of the k_{off} calculation.

Unbinding process using the RESP-HF parametrization. MetaD-based calculations of k_{off} values require the the free energy landscape of ligand unbinding pathways. Previously, we have predicted this for the ligand iperoxo and its target M₂ muscarinic receptor in its active state by using well-tempered metadynamics.²⁶ The calculations were based on the Amber14SB force field²³ for the protein. The RESP charges for the ligand were calculated at the HF/6-31G* level of theory (**RESP-HF** hereafter). Two different unbinding pathways emerged (here noted with **I** and **II**)²⁶ (see SI, Fig.2). Pathway **I** is the lowest in free energy. Here, the ligand starts from the bound state, rotates around the axis formed by the alkyne bond, passing through the transition state 1 (TS1, Fig.2), to finally reach state **A**. After this step, we observe a rotation of the entire ligand with the trimethylammonium group as a pivot (transition state 2 - TS2), TS2 is the rate-limiting step. A salt bridge between the trimethylammonium group and ASP103, present in the bound state, is broken here and the overall number of intramolecular hydrogen bonds between protein and

the ligand is reduced (Fig. 7 in the SI) and substituted with other H-bonds with solvent. A slightly different rotation around the same pivot can lead to state **B**. This second rotation does not lead to unbinding, and is not further considered here. After reaching state **C**, the trimethylammonium group breaks the salt bridge formed with ASP103 (transition state 3-TS3) and moves toward the extracellular part, reaching the fully solvated state. The rate-limiting step is TS2. Pathway **II** is identical to **I** until the ligand reaches state **C**. Here, the rearrangement of the extracellular loop 2 (ECL2) of the receptor limits the possibility of the ligand to reach the solvated state right after visiting state **C**, forcing it to perform a further rotation, reaching the last metastable state **D**. After that rotation, the ligand breaks the salt bridge formed with ASP103 and it moves towards the solvent, completing its pathway to solvation.

Here we evaluate the k_{off} by a multistep approach as in Casasnovas *et al.*¹⁷ Because kinetic calculations are very expensive, we explore only the lowest free energy pathway, **I**.²⁶ For this, we identify the path collective variables (pathCVs) s_{path} and z_{path} .³¹ These are particularly appropriate to study a single pathway between two reference states, limiting the motion of the system only around this predefined path. s_{path} defines the progression along the pathway, while z_{path} samples deviation from the reference path (in our case **I**). Next, we perform Well-Tempered-MetaD²⁴ to calculate the free energy as a function of s_{path} and z_{path} . Finally, we use Frequency-Adaptive MetaD,¹⁶ for the actual calculation of k_{off} .

The definition of a pathCV is based on a metric that measures the distance of instantaneous configurations from the path. In the first applications of pathCV,³¹ the metric chosen was based on root mean square displacement (RMSD) with respect to the initial bound state. In our case, having seen the presence of multiple intermediate states along pathway **I**²⁶ (Fig. 5), we prefer to define our metrics based on a contact map based on the ligand–protein atom pairs that are crucial for the stabilization of the intermediate states (see SI for detail). To obtain a sequence of conformations along the unbinding trajectory, we employed Ratchet&Pawl MD^{32,33} in a two step-approach. First, we forced our system to perform its

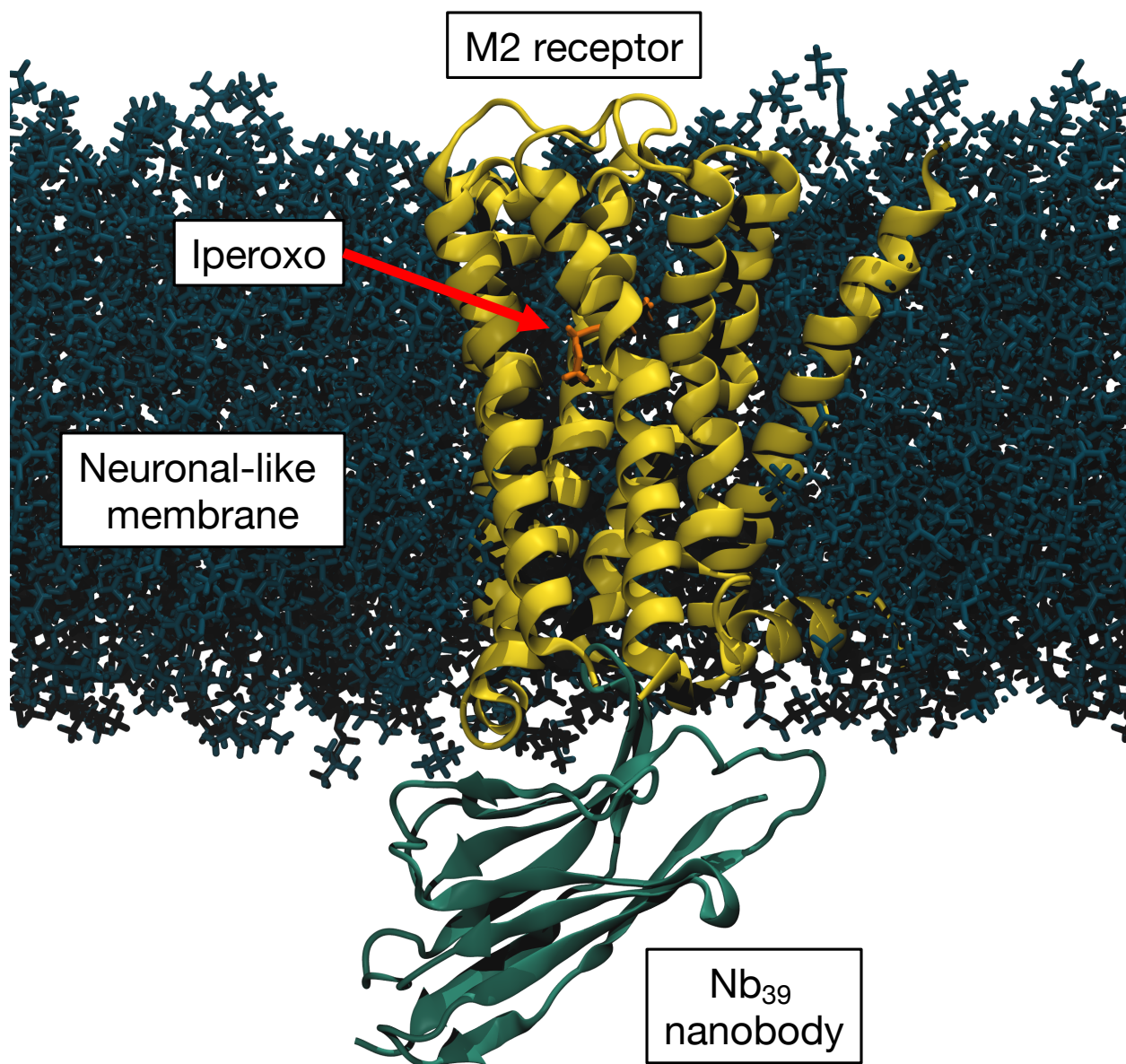


Figure 1: Representation of the simulated system. The M₂ receptor (yellow), with its agonist iperoxo (orange), is embedded in a neuronal-like membrane³⁰ (dark blue) and it is bound to the Nb₃₉ nanobody (green) which keeps it in its active state. Water and Na⁺/Cl⁻ solvated ions are removed for sake of clarity.

unbinding transition using as CV the distance between the binding pocket and the center of mass of the ligand (in the same spirit as a previous work³⁴); from these structures we built a first pathCV. As a second step, we apply Ratched&Pawl to this first pathCV. From the set of conformations obtained during this run, we built the final pathCVs that we employed in the MetaD simulations.

Unbinding process using the RESP-B3LYP parametrization . Multiple-walkers³⁵ Well-tempered²⁴ MetaD along the pathCV previously obtained lead to a large number of recrossing events in all the 10 replicas. The calculated free energy surface along pathway **I** (Fig. 2) shows that the unbinding process was the same as that obtained by **RESP-HF**²⁶ calculations (Fig. 2 in the SI). In particular, all the intermediate states identified by the two setups are the same. Therefore, we use the same pathCVs for both **RESP-HF** and **RESP-B3LYP** Frequency-Adaptive MetaD¹⁶ calculations of k_{off} .

Calculations of k_{off} . We performed 10 different Frequency-Adaptive MetaD¹⁶ runs, biasing both s_{path} and z_{path} . For **RESP-HF**, the bias needed to perform the first ring rotation from bound state to state **A**, in the first run exceeded 140 kJ/mol. This bias corresponds roughly to a residence time in the order of years (because the acceleration factor is exponentially proportional to the bias deposited – see SI). Therefore, this parametrization could not be used for the calculations of the k_{off} . For **RESP-B3LYP**, 5 production runs could be collected. They covered 0.9 to 1.7 μs , for a total simulation time of $\sim 8\mu\text{s}$.¹

The resulting distribution of calculated residence times (Fig. 3) was fitted with a Poisson distribution. From this, we obtained a residence time $2.7 \pm 0.5 \cdot 10^3$ s and a $k_{\text{off}} = 3.7 \pm 0.7 \cdot 10^{-4}$ s⁻¹. To validate the correctness of the calculation performed, we performed a Kolmogorov–Smirnov test between the obtained distribution and the theoretical one.³⁶ The p-value turned out to be 0.87. This shows that the obtained distribution is statistically indistinguishable from a theoretical rare event distribution.

In conclusion, for **RESP-B3LYP** calculations we observe a rare event (as confirmed

¹Other 5 runs had to be removed because they deposited bias on the transition state, invalidating the sampling performed.¹⁵

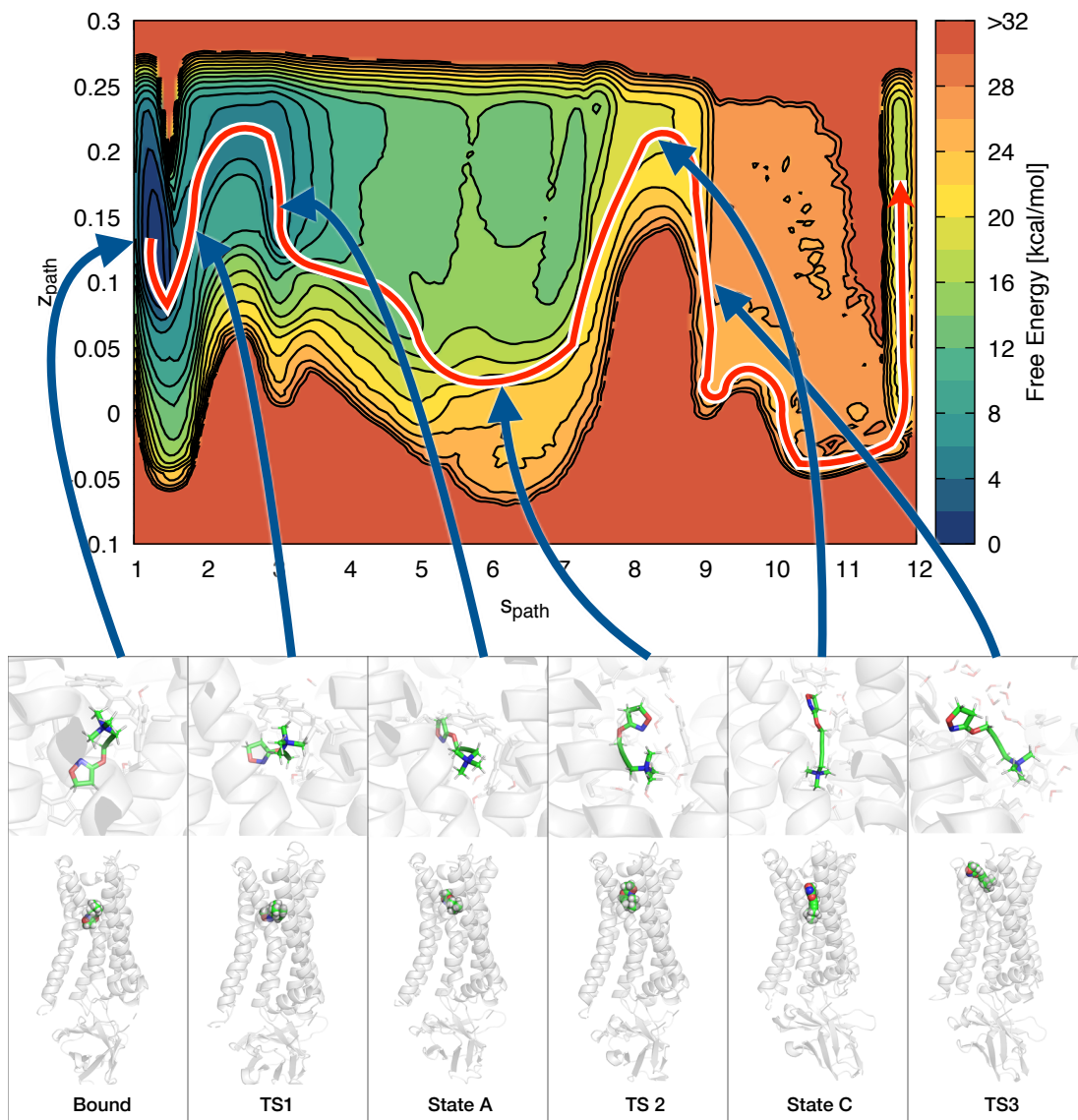


Figure 2: Free energy surface of binding with the observed unbinding pathway and representative structures. The top part shows the M_2 /iperoxo free energy surface as a function of s_{path} and z_{path} , with the unbinding pathway followed by the ligand. The bottom part represents the Bound, TS1, A, TS2, C, and TS3 states: in the upper panel, both the ligand and its surrounding atoms within 4.5 Å are rendered; in the lower panel, the ligand is rendered in sphere mode with all the receptor and the nanobody. Water and ions are not represented for sake of clarity.

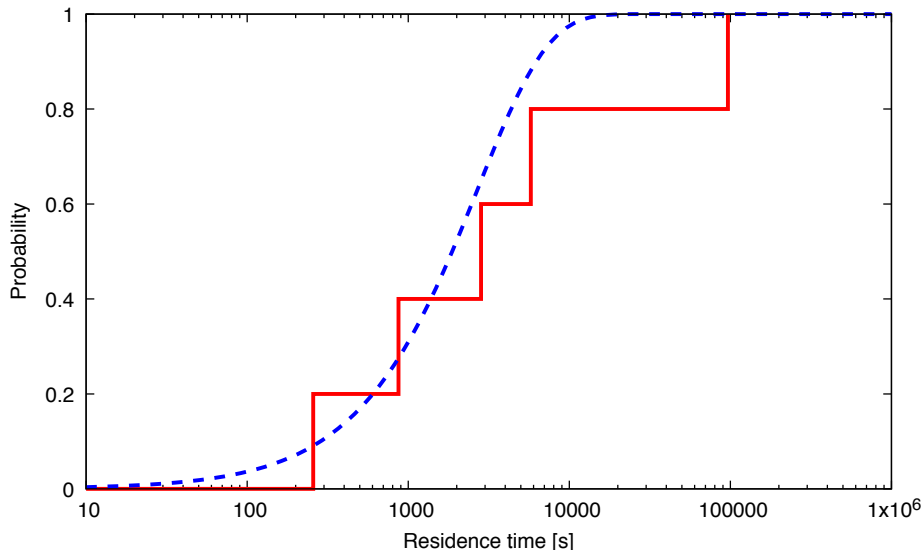


Figure 3: Comparison between the calculated (red line) and the theoretical Poisson distributions (blue dashed curve).

by that displays a free energy barrier that is higher with respect to the one experimentally observed). For **RESP-HF**, the free energy barrier between the bound and the first intermediate state **A** is so high that the residence time related to this transition is at least in the order of years. The discrepancy of the **RESP-B3LYP** and, more, of the **RESP-HF** results does not seem to origin from the enhanced sampling approach adopted but it is likely to be ascribed mostly to the underlying potential energy function, associated either with the ligand distortions and/or with ligand/protein interactions along the pathway. As for the first issue, we find that small and similar distortions of the ligand’s internal degrees of freedom (such as bond lengths (Fig. 3 in SI), bond angles (Fig. 4 in SI), and dihedrals (Fig. 5 in SI)) are present in both TS2 and Bound states ². Thus, the distortion of the molecules are not likely to contribute largely to the increase of potential energy on passing from the bound state to TS2. We address the second issue by recalculating the ligand/protein energetics at the quantum mechanics level for selected configurations in the two states, considering the unbound state —the ligand in solution— as our reference state (i.e., computing the $\Delta\Delta E$ with

²Although the ligand undergo to larger structural fluctuation at the TS2 than in the bound state (Fig. 6 in SI).

respect to the solvated state). We use a QM/MM approach as implemented in the MiMiC multiscale interface.^{37,38} The QM part of the system consists of the ligand and its interacting groups (Fig. 5a-c, See Methods for details). It is treated at the DFT level, using either the B3LYP²⁷⁻²⁹ or the BLYP^{27,28} exchange-correlation functionals. The rest is described by the Amber14SB force field,²³ as in all the other calculations presented in this paper. The energy of the total system is calculated by adding the QM/MM interaction energy to the QM and MM energies, while that of the ligand is given by the QM energy. The $\Delta\Delta E$ values are also calculated at a purely classical force field (CFF) level using **REST-B3LYP**.

In the bound state $\Delta\Delta E$, calculated for the same 10 representative conformations, turn out to be very similar using the force field and DFT (-17.3 ± 1.5 kcal/mol (CFF) and -18.1 ± 3.1 (B3LYP)/ -18.1 ± 3.1 (BLYP) kcal/mol), i.e. the values are not significantly affected by the exchange-correlation functional (Fig. 4, Table 2 in SI). Increasing the number of representative conformations to 45 and 90, calculated for CFF and BLYP does not modify significantly this picture (-17.7 ± 0.5 (CFF) and -16.8 ± 1.8 (BLYP) and -20.1 ± 0.3 (CFF) and -18.9 ± 1.5 (BLYP) kcal/mol, respectively Fig.4 or Table 2 in SI).

Taken together, these results are consistent with previous Well-Tempered-MetaD-based free energy calculations using **RESP-HF**, which showed excellent agreement between calculated and experimental affinities for this^{26,34} and other systems. This is expected and it is indeed confirmed by countless examples, both in protein-ligand³⁹ and protein-protein⁴⁰ interactions. This further supports the conclusion that modern force-field-based calculations accurately predict ligand binding affinities.

A dramatically different scenario takes place at the transition state of the unbinding process, TS2. Here the $\Delta\Delta E$ values obtained with the force field (-17.8 ± 1.3 kcal/mol) differs significantly from the DFT ones, while the latter are still similar to each other (-12.8 ± 2.9 (B3LYP)/ -13.2 ± 2.8 (BLYP) kcal/mol). The trend is preserved when the number of conformations is increased (-17.0 ± 0.6 (CFF) and -12.1 ± 1.7 (BLYP) and -17.5 ± 0.4 (CFF) and -12.4 ± 1.4 (BLYP), for 45 and 90 conformations, respectively, see Fig. 4 or Tab. 2 in SI).

These results suggest that, within the limitation of our statistics, CFF overestimates the potential energy increase (associated with disruption of several intermolecular interactions) at TS2. This suggestion is fully consistent with our calculations of residence times, which are longer with respect to the experimental data.

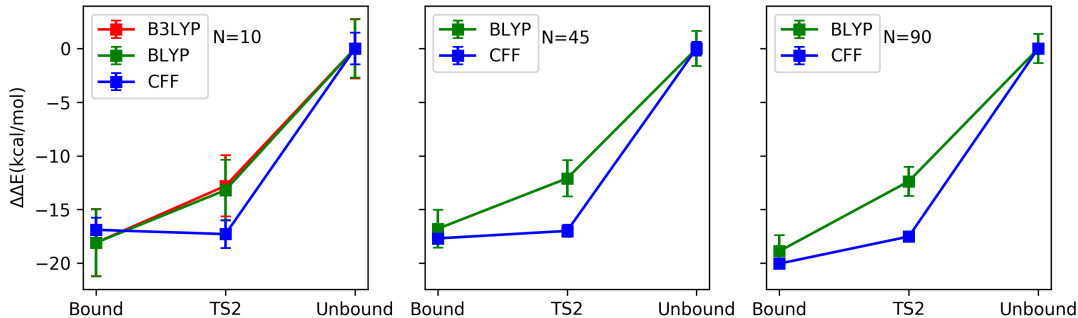


Figure 4: Relative interaction energies ($\Delta\Delta E$) of the ligand to the M_2 receptor at QM/MM level (with BLYP and B3LYP exchange–correlation functionals) and CFF level. N is the number of conformations considered in statistics for each state. $\Delta\Delta E$ is not directly related to the free energy of the binding/unbinding process.

Finally, we investigate if changes in electronic polarization⁴¹ and charge transfer (CT) effects (which are lacking in routinely used biomolecular force fields) occur in the unbinding process. We calculate, using QM/MM, the rearrangement of electronic density of the ligand while passing from *in vacuo* to the bound state, to the transition state TS2, and to the unbound state (Fig. 5 a-c). The calculations are carried out using both the Voronoi⁴² and Bader⁴³ approaches to the partition of the atomic charges (Fig. 8 in SI). Polarization effects turn out to involve mostly the positively charged trimethylammonium group (forming a salt bridge with ASP103 or interacting with water) and the two oxygen atoms (Fig. 5 a-c). The overall CT (ΔQ_{CT}) as well as the polarization (ΔQ_{Pol}), albeit small in magnitude (a fraction of an elementary charge), differ slightly on passing from the bound state to the TS (Fig. 5d and Fig. 8 in SI). Obviously, the fixed charge schemes of commonly used force fields cannot take into account these changes in charge redistribution during the unbinding processes.

Correcting this overestimation would lead to a better representation of the energetics, and

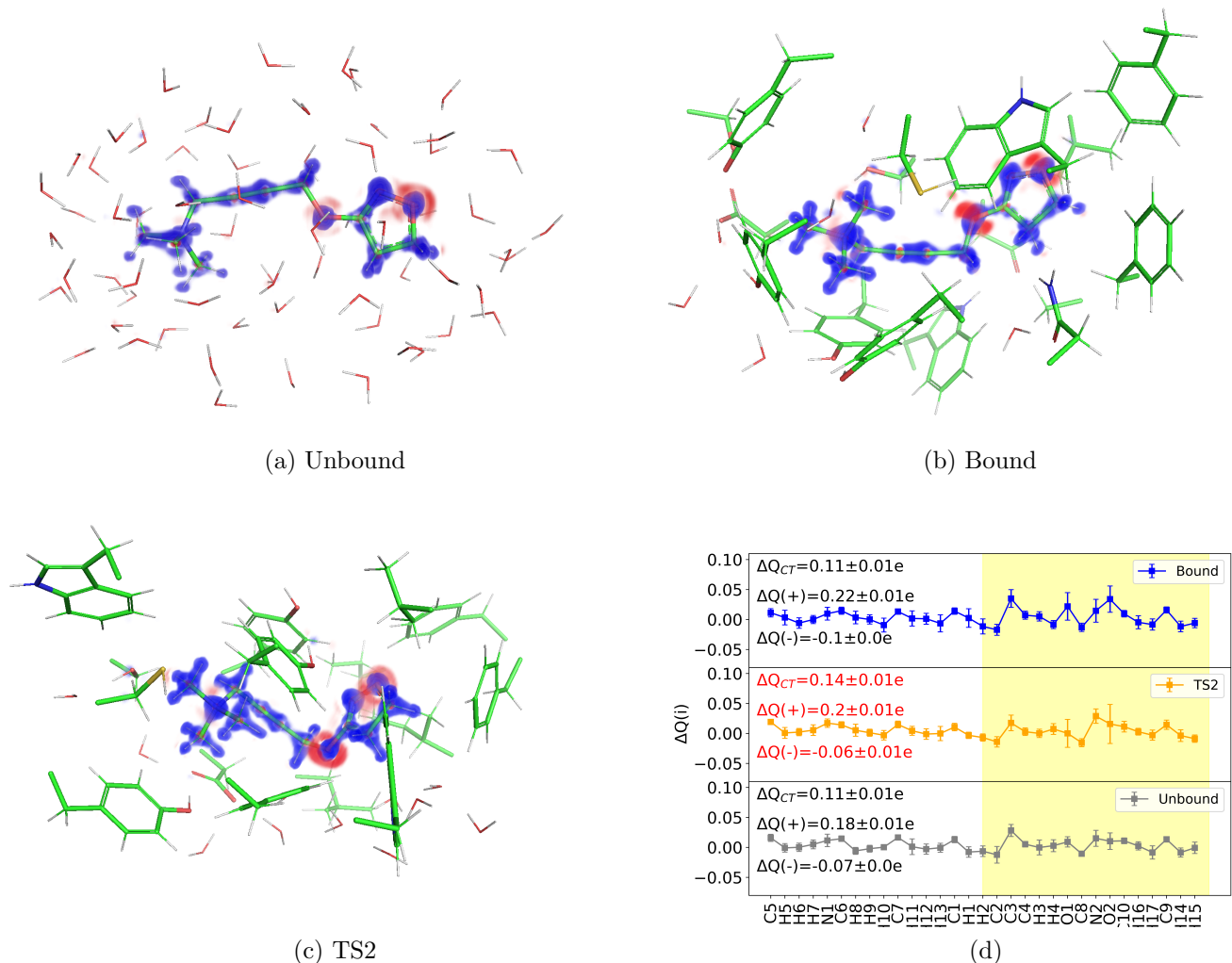


Figure 5: (a-c) Change of electronic density of the ligand on passing from *in vacuo* to in water ($\Delta\rho = \rho^{cplx} - \rho^{lig} - \rho^{rest}$) for the unbound (a), bound (b), and TS2 (c) states (blue: $\Delta\rho > 0$, red: $\Delta\rho < 0$). The atoms are displayed in sticks mode. (d) Correspondent change in atomic charge for each atom ($\Delta Q(i)$, $i=0-30$). ΔQ_{CT} is the overall charge transferred from the protein to the ligand, ranging from 0.11 to 0.14 electrons. $\Delta Q(+)$ and $\Delta Q(-)$ are the polarizations of the ligand contributed from atoms with positive and negative $\Delta Q(i)$, respectively, ranging from -0.1 to 0.22 electrons. The yellow color band highlights the most significant differences among the three states. The calculations have been carried out using the Voronoi partition scheme.⁴²

lower the residence time. However, simply adding a correction might not be enough. Indeed, using more accurate potential energy functions (coming, for instance, from apt polarizable force fields⁴⁴) could give a different pathway and thus a different transition state, as already suggested.⁴⁵ In addition, the entropic contribution to the free energy of the transition state may be affected by the accuracy of the potential energy surface of force fields.

The k_{off} constant is exponentially related to the height of the free energy barrier for dissociation. Small errors in the force field may therefore introduce large errors in the k_{off} of drugs, a key parameter in pharmacology. Here, we take iperoxo, a superagonist, and we find that the bound state to its target M₂ receptor is excellently described by both the **RESP-HF** and the **RESP-B3LYP** parametrization of the ligand effective point charges. This confirms the adequacy of the classical force-field representation in the minimum-energy states of the system. Thus, although the electrostatic parametrization in AMBER currently lacks polarization of the system, a high level of accuracy has been achieved by tuning all the parameters of the force field for more than 4 decades.^{46,473} In contrast, the accuracy of the force field at the transition state appears to be limited.⁴⁵ This may be caused, at least in part, by the fact that standard non-polarizable force fields cannot capture the slight changes of electronic polarization and charge transfer effects (Fig. 5) on passing from the bound to the transition state. Furthermore, a more detailed representation of the interaction could in principle find different unbinding pathways, that can change the energetics of the intermediate states. Polarizable force fields,^{44,48,49} reactive MD,⁵⁰ and/or corrections of the free energy landscape derived from quantum mechanical calculations⁴⁵ might alleviate this problem. Similar and possibly even more severe issues may be expected in other unbinding processes, such as those involving protein/protein and protein/DNA complexes.

³Although the total energy function is highly accurate, its single contribution to electrostatics may not provide the electrostatic energy.

Computational Methods

System preparation. We followed a similar protocol as that of ref.²⁶ We obtained the structure from the protein data bank (PDB code: 4MQS⁵¹), parametrized the ligand using GAFF,²¹ and embedded it in a neuronal-like³⁰ membrane. The drug charges were obtained, after an initial geometry optimization, by restrained electric potential fitting method (RESP)⁵² with HF/6-31G* (**RESP-HF**) and B3LYP/6-31G* (**RESP-B3LYP**) levels of theory. Then the system was solvated, ions were added to reach the experimental ionic force, and finally minimizing and equilibrating the system (details in the SI). All the simulations were performed using GROMACS 2018.4.⁵³ patched with PLUMED 2.5.⁵⁴

rMD Simulation and pathCV identification. To identify a collection of conformations to set up our pathCV variable,³¹ we used Ratchet&Pawl MD.^{32,33} As an initial ratcheting coordinate, we considered the distance between the center of mass of our ligand and the center of mass of the pocket (defined by TYR104, SER107, VAL111, PHE195, and TYR239), projected along the axis normal to the membrane. We fixed the bias factor to $k = 500$ kJ/mol/nm and the final ratchet coordinate to $r_{\text{final}} = 2.5$ nm. After 10 different 20-ns long rMD runs with these parameters, we selected 11 frames that describe well the progression of the ligand toward the solvent, and we used them to define a pathCV based on the contact map⁵⁵ between a list of atom pairs of the ligand and the receptor (list in the SI). With this variable, we performed 10 new 20-ns long rMD runs to verify and eventually refine the new variable, choosing again by visual inspection 11 different frames from the unbinding trajectories and redefining a final pathCV that was then used in our MetaD simulation.

Well-Tempered MetaD along the pathCV. Using the pathCV as identified above, we performed multiple-walkers³⁵ Well-Tempered MetaD²⁴ along the z_{path} and s_{path} using 10 different walkers. We set the bias factor to 24, an initial Gaussian height of 1.2 kJ/mol, and a frequency deposition of 1 ps⁻¹. To limit the phase space exploration to path **I** only, we prevented our system from reaching pathway **II**, by putting a restraint to avoid its motion along that pathway (i.e. we put a harmonic wall at $z_{\text{path}} = 0.25$, where paths **I** and **II**

diverge). The total simulation time was 1.8 μ s. The free energy surface was reweighted *a posteriori* with the Tiwary and Parrinello algorithm.⁵⁶

Frequency-Adaptive MetaD. We carried out 10 different Frequency-Adaptive MetaD¹⁶ runs. The approach is a variant of I-MetaD¹⁵ that speeds up the calculations (details in SI). We performed 10 different Frequency-Adaptive MetaD runs, with a bias factor of 24, an initial Gaussian height of 1.2 kJ/mol, an initial frequency deposition of 1 ps⁻¹, an acceleration parameter $\theta = 100$ (details in the SI), and a minimum frequency deposition of 10⁻² ps⁻¹. Out of all simulation performed, 5 have deposited bias on the transition state, and thus we discarded them, obtaining the 5 residence times from the remaining simulations.

QM/MM single point calculations. A selection of N=45 or 90 structures associated with the Bound, TS2, and Unbound states (Fig. 2) underwent 1000 steps of energy minimization using the steepest descent algorithm at the CFF level. Then, for each structure, we considered the *total system*, the *rest* (i.e. the system without the ligand) and the *ligand* without the systems (i.e. in vacuum). Overall, 270 structures were considered.

The QM regions in the *total system* consisted of the ligand, and the side-chains (up to the -C $_{\beta}$) directly interacting with it as well as water molecules within 4.5 Å from it. They ranged from 196 to 308 atoms (see Table 3 in SI). The QM regions of the *rest* were the same except that the ligand was not included. They ranged from 165 to 277 atoms. Those of the *ligand* included only the latter (31 atoms).

The QM part was described at the DFT level (QM part, Fig. 2), using either the B3LYP or BLYP exchange–correlation functional.⁵⁷ A plane-wave basis set with a cutoff of 90 Ry was used. The core electrons were described through norm-conserving Troullier–Martins pseudopotentials.⁵⁸ Isolated system conditions were achieved by using the Martyna–Tuckerman scheme.⁵⁹ For the Bound and TS2 states, (i) covalent C $_{\alpha}$ -C $_{\beta}$ bonds across the QM/MM interface were described by an adapted monovalent carbon pseudopotential;⁶⁰ (ii) the net charge of the residues considered at QM level were reweighted to their sidechain atoms by neutralizing the sum of the partial charges of the remaining backbone atoms in MM region.

For the *total system* and *rest*, the atoms other than those in the QM regions ('MM region') were described using exactly the same setup and force field as for the metadynamics. The interactions between the QM and MM parts were described as in ref. 37: the electrostatic interactions were calculated explicitly for MM atoms that are within 30 Å of the centroid of the QM part, whereas the interactions with the rest of the system was evaluated using a 5th-order multipole expansion of the electrostatic potential.³⁷ In all the calculations, the Grimme's correction⁶¹ was used to describe dispersion interactions.

Single point electronic structure calculations were performed, with a convergence criteria of 10^{-7} au, using the highly scalable MiMiC-based QM/MM interface,^{37,38} which combines CPMD 4.1⁶² with GROMACS 2019.4.⁶³

The ligand binding energy ΔE was calculated either at QM/MM ($\Delta E(\text{B3LYP}/\text{BLYP})$) or at the classical force field (CFF) (that is, at the the **RESP-B3LYP**) level ($\Delta E(\text{CFF})$). It reads

$$\Delta E = E_{\text{total}} - E_{\text{ligand}} - E_{\text{rest}} \quad (1)$$

where E_{total} and E_{rest} are the potential energies of the *total* system and of the *rest* (given by summing the QM energy with the MM energy and the QM/MM interaction energies), and E_{ligand} the potential energy of the *ligand*.

We computed ΔE at BLYP and CFF levels for N=45 and N=90 conformations for each state, in order to verify the consistency of our evaluation. To verify the same effect at a higher level of theory, we chose 10 structures covering the same spreading range of the calculated energies at the BLYP level for the more expensive and accurate B3LYP calculations⁴.

The change in electron density upon ligand binding was calculated at the B3LYP level for N=10.

$$\Delta \rho = \rho^{\text{total}} - \rho^{\text{ligand}} - \rho^{\text{rest}} \quad (2)$$

Here, ρ^{total} is the electron density of the QM part embedded in the MM part of the *total*

⁴For comparison, the statistical estimate at BLYP and CFF were re-evaluated for the 10 structures.

system, ρ^{ligand} is that of the *ligand* and ρ^{rest} is that of the *rest*.

The electron charge transfer (CT) associated with atom i of the ligand reads:⁴²

$$\Delta Q(i) = \int_{VP_i} \Delta \rho(\mathbf{r}) d\mathbf{r} \quad (3)$$

The integral is solved numerically over the grid points within the Voronoi partition,⁴² or Bader’s atom in molecules partition⁶⁴ of atom i (VP_i). An in house code (cpmd-cube-tools: <https://pypi.org/project/cpmd-cube-tools/>) and the Bader code (Bader charge analysis: <http://theory.cm.utexas.edu/henkelman/code/bader/>) were used, respectively.

The CT effect of the whole ligand molecule reads

$$\Delta Q_{CT} = \sum_i \Delta Q(i) \quad (4)$$

An estimation of the change in charge distribution is given by electric polarization as:

$$\Delta Q_{Pol} = |\Delta Q(+)| + |\Delta Q(-)| \quad (5)$$

where, $\Delta Q(+)= \sum_i \Delta Q(i)$, $i \in \{\Delta Q(i) > 0\}$ and $\Delta Q(-)= \sum_i \Delta Q(i)$, $i \in \{\Delta Q(i) < 0\}$.

Acknowledgement

The authors thank Anna Bochicchio and Rodrigo Casasnovas for the initial preparation of the system, Emiliano Ippoliti, Luca Maggi, and GiovanniMaria Piccini for useful discussion. The authors gratefully acknowledge the computing time granted through VSR on the supercomputer JURECA⁶⁵ at Forschungszentrum Jülich (Project ID: jias5d) and acknowledge the JSC for the computing time on the supercomputer JURECA Booster module. This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HBP SGA2) and the Marie Skłodowska-Curie

Grant Agreement No. 642069. W.L. appreciates the National Natural Science Foundation of China No. 21505134. J.M.H.O. acknowledges financial support from the Research Council of Norway through its Centres of Excellence scheme (Project ID: 262695). U.R. acknowledges funding from the Swiss National Science Foundation via the NCCR MUST and individual grants.

Supporting Information Available

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,⁶⁶ as plumID:20.005.

Details on Ratchet&Pawl MD, FA MetaD, system preparation and choice on the CV and on the QM/MM calculations are available in the Supporting Information.

References

- (1) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug–target residence time and its implications for lead optimization. *Nature reviews Drug discovery* **2006**, *5*, 730–739.
- (2) Swinney, D. C. The role of binding kinetics in therapeutically useful drug action. *Current opinion in drug discovery & development* **2009**, *12*, 31–39.
- (3) Lu, H.; Tonge, P. J. Drug–target residence time: critical information for lead optimization. *Current opinion in chemical biology* **2010**, *14*, 467–474.
- (4) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug–receptor binding kinetics. *Drug discovery today* **2013**, *18*, 667–673.
- (5) Schmidtke, P.; Luque, F. J.; Murray, J. B.; Barril, X. Shielded hydrogen bonds as

- structural determinants of binding kinetics: application in drug design. *Journal of the American Chemical Society* **2011**, *133*, 18903–18910.
- (6) Dickson, A.; Tiwary, P.; Vashisth, H. Kinetics of ligand binding through advanced computational approaches: a review. *Current topics in medicinal chemistry* **2017**, *17*, 2626–2641.
 - (7) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Recent progress in molecular simulation methods for drug binding kinetics. *arXiv preprint arXiv:2002.08983* **2020**,
 - (8) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proceedings of the National Academy of Sciences* **2011**, *108*, 13118–13123.
 - (9) Kokh, D. B.; Kaufmann, T.; Kister, B.; Wade, R. C. Machine learning analysis of tauRAMD trajectories to decipher molecular determinants of drug-target residence times. *Frontiers in molecular biosciences* **2019**, *6*, 36.
 - (10) Kokh, D. B.; Doser, B.; Richter, S.; Ormersbach, F.; Cheng, X.; Wade, R. C. A Workflow for Exploring Ligand Dissociation from a Macromolecule: Efficient Random Acceleration Molecular Dynamics Simulation and Interaction Fingerprints Analysis of Ligand Trajectories. 2020.
 - (11) Voter, A. F. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters* **1997**, *78*, 3908.
 - (12) Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Physical Review E* **1995**, *52*, 2893.
 - (13) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature communications* **2015**, *6*, 7653.

- (14) Wolf, S.; Lickert, B.; Bray, S.; Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nature Communications* **2020**, *11*, 2918.
- (15) Tiwary, P.; Parrinello, M. From metadynamics to dynamics. *Physical review letters* **2013**, *111*, 230602.
- (16) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K. Frequency adaptive metadynamics for the calculation of rare-event kinetics. *The Journal of chemical physics* **2018**, *149*, 072309.
- (17) 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.
- (18) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chemical science* **2017**, *8*, 6466–6473.
- (19) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nature communications* **2019**, *10*, 1–8.
- (20) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.
- (21) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.
- (22) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophysical journal* **2011**, *100*, L47–L49.

- (23) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.
- (24) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters* **2008**, *100*, 020603.
- (25) Schrage, R.; Holze, J.; Klöckner, J.; Balkow, A.; Klaue, A. S.; Schmitz, A.-L.; De Amici, M.; Kostenis, E.; Tränkle, C.; Holzgrabe, U. et al. New insight into active muscarinic receptors with the novel radioagonist [3H] iperoxo. *Biochemical pharmacology* **2014**, *90*, 307–319.
- (26) Capelli, R.; Boichichio, A.; Piccini, G.; Casasnovas, R.; Carloni, P.; Parrinello, M. Chasing the full free energy landscape of neuroreceptor/ligand unbinding by metadynamics simulations. *Journal of chemical theory and computation* **2019**, *15*, 3354–3361.
- (27) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (28) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (29) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- (30) Chan, R. B.; Oliveira, T. G.; Cortes, E. P.; Honig, L. S.; Duff, K. E.; Small, S. A.; Wenk, M. R.; Shui, G.; Di Paolo, G. Comparative lipidomic analysis of mouse and human brain with Alzheimer disease. *Journal of Biological Chemistry* **2012**, *287*, 2678–2688.
- (31) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *The Journal of chemical physics* **2007**, *126*, 054103.

- (32) Marchi, M.; Ballone, P. Adiabatic bias molecular dynamics: a method to navigate the conformational space of complex molecular systems. *The Journal of chemical physics* **1999**, *110*, 3697–3702.
- (33) Tiana, G.; Camilloni, C. Ratcheted molecular-dynamics simulations identify efficiently the transition state of protein folding. *The Journal of chemical physics* **2012**, *137*, 235101.
- (34) 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.
- (35) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *The journal of physical chemistry B* **2006**, *110*, 3533–3539.
- (36) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the reliability of the dynamics reconstructed from metadynamics. *Journal of chemical theory and computation* **2014**, *10*, 1420–1425.
- (37) Olsen, J. M. H.; Bolnykh, V.; Meloni, S.; Ippoliti, E.; Bircher, M. P.; Carloni, P.; Rothlisberger, U. MiMiC: A Novel Framework for Multiscale Modeling in Computational Chemistry. *Journal of Chemical Theory and Computation* **2019**, *15*, 3810–3823.
- (38) Bolnykh, V.; Olsen, J. M. H.; Meloni, S.; Bircher, M. P.; Ippoliti, E.; Carloni, P.; Rothlisberger, U. Extreme Scalability of DFT-Based QM/MM MD Simulations Using MiMiC. *Journal of Chemical Theory and Computation* **2019**, *15*, 5601–5613.
- (39) Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A. Advances in free-energy-based simulations of protein folding and ligand binding. *Current opinion in structural biology* **2016**, *36*, 25–31.

- (40) Siebenmorgen, T.; Zacharias, M. Computational prediction of protein–protein binding affinities. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2019**, e1448.
- (41) Gao, J.; Xia, X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258*, 631–635.
- (42) Carloni, P.; Andreoni, W.; Hutter, J.; Curioni, A.; Giannozzi, P.; Parrinello, M. Structure and Bonding in Cisplatin and Other Pt(II) Complexes. *Chemical Physics Letters* **1995**, *234*, 50–56.
- (43) Bader, R. F. Atoms in molecules. *Accounts of Chemical Research* **1985**, *18*, 9–15.
- (44) Li, H.; Chowdhary, J.; Huang, L.; He, X.; MacKerell Jr, A. D.; Roux, B. Drude polarizable force field for molecular dynamics simulations of saturated and unsaturated zwitterionic lipids. *Journal of chemical theory and computation* **2017**, *13*, 4535–4552.
- (45) Haldar, S.; Comitani, F.; Saladino, G.; Woods, C.; van der Kamp, M. W.; Mulholland, A. J.; Gervasio, F. L. A multiscale simulation approach to modeling drug–protein binding kinetics. *Journal of chemical theory and computation* **2018**, *14*, 6093–6101.
- (46) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. Calculation of the relative change in binding free energy of a protein-inhibitor complex. *Science* **1987**, *235*, 574–576.
- (47) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature* **1987**, *328*, 551–554.
- (48) Villa, F.; MacKerell Jr, A. D.; Roux, B.; Simonson, T. Classical drude polarizable force field model for methyl phosphate and its interactions with mg^{2+} . *The Journal of Physical Chemistry A* **2018**, *122*, 6147–6155.

- (49) Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of ion–protein interactions using the drude polarizable force-field. *The Journal of Physical Chemistry B* **2015**, *119*, 9401–9416.
- (50) Meuwly, M. Reactive molecular dynamics: From small molecules to proteins. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2019**, *9*, e1386.
- (51) Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J.; Hu, K.; Eitel, K.; Hübner, H.; Pardon, E.; Valant, C.; Sexton, P. M. et al. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **2013**, *504*, 101–106.
- (52) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.
- (53) 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.
- (54) 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.
- (55) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. The unfolded ensemble and folding mechanism of the C-terminal GB1 β -hairpin. *Journal of the American Chemical Society* **2008**, *130*, 13938–13944.
- (56) Tiwary, P.; Parrinello, M. A time-independent free energy estimator for metadynamics. *The Journal of Physical Chemistry B* **2015**, *119*, 736–742.
- (57) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Physical Review B* **1988**, *37*, 785–789.

- (58) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. *Physical Review B* **1991**, *43*, 1993–2006.
- (59) Martyna, G. J.; Tuckerman, M. E. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. *Journal of Chemical Physics* **1999**, *110*, 2810–2821.
- (60) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Variational optimization of effective atom centered potentials for molecular properties. *Journal of Chemical Physics* **2005**, *122*.
- (61) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010**, *132*.
- (62) Hutter, A., J.A.; Deutsch, T.; Bernasconi, M.; Goedecker, S.; Marx, D.; Tuckerman, M.; Parrinello, M. CPMD Copyright IBM Corp 1990–2019, MPI für Festkörperforschung, Stuttgart, 1997–2012.
- (63) 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.
- (64) Tang, W.; Sanville, E.; Henkelman, G. A grid-based Bader analysis algorithm without lattice bias. *Journal of Physics: Condensed Matter* **2009**, *21*, 084204.
- (65) Krause, D.; Thörnig, P. JURECA: general-purpose supercomputer at Jülich supercomputing centre. *Journal of large-scale research facilities JLSRF* **2016**, *2*, 62.
- (66) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G.; Bonas, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D. et al. Promoting transparency

and reproducibility in enhanced molecular simulations. *Nature methods* **2019**, *16*, 670–673.