On the 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,@, $^{\triangle}$ Ursula Rothlisberger, $^{\perp}$ Michele Parrinello, $^{\nabla}$,††,‡‡ and Paolo Carloni†,¶¶

†Computational Biomedicine Section, IAS-5/INM-9, Forschungzentrum Jülich, Wilhelm-Johnen-straße, D-52425 Jülich, Germany

‡JARA-HPC, Forschungszentrum Jülich, D-54245 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

 $\triangle Department\ of\ Chemistry,\ Aarhus\ University,\ DK-8000,\ Aarhus\ C,\ Denmark$

∇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, Forschungzentrum Jülich, Wilhelm-Johnen-straße,

D-52425 Jülich, Germany

 $\S\S{RC}\ and\ WL\ share\ first\ authorship.$

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

Supplementary Information

Supplementary Table 1: Contact map atom pairs

List of the pairs that form the contact map used in the pathCV definition. If a list of atoms is inserted, we consider the center of mass of those. The atom names for the protein residues are the standard ones from the Amber14SB force field. The names of ligand atoms are given in Supplementary Figure 2.

ID	atoms	protein	iperoxo
		residue	atoms
$\overline{d_1}$	CG,CD1,CE1,CZ,CE2,CD2	TYR262	N1
d_2	CG,CD1,CE1,CZ,CE2,CD2	TYR239	N1
d_3	CG,CD1,CE1,CZ,CE2,CD2	TYR104	N1
d_4	OD1,OD2	ASP103	N1
d_5	CG,CD1,NE1,CE2,CZ2,CH2,CZ3,CE3,CD2	TRP236	C8,N2,O2,C10,C9
d_6	CB,CG1,CG2	VAL111	C8,N2,O2,C10,C9
d_7	CG2,CG1,CD1	ILE178	C8,N2,O2,C10,C9
d_8	$_{\mathrm{CB,SG}}$	CYS176	N1
d_9	CB,CG2,OG1	THR187	C8,N2,O2,C10,C9

Supplementary Table 2: Ligand/protein interaction energies at the QM/MM and force field (FF) levels

The intermolecular interaction energies ΔE (equation 1) of the ligand in different states along the unbinding pathway are calculated at the QM/MM (B3LYP/BLYP) and CFF levels. The differences in interaction energy $\Delta \Delta E$ are defined with respect to the unbound state (i.e. $\Delta \Delta E = \Delta E - \Delta E$ (unbound). All the reported values are in kcal/mol. N is the number of conformations considered in statistics.

		Bound	Transition	Unbound
	N	State	State (TS2)	State
$\Delta E(\text{B3LYP})$	10	-127.6 ± 2.4	-122.3 ± 2.1	-109.5 ± 2.0
$\Delta E(\mathrm{BLYP})$	10	-124.7 ± 2.5	-119.8 ± 2.1	-106.6 ± 1.9
$\Delta E({\rm CFF})$	10	-52.5 ± 0.5	-52.9 ± 0.8	-35.6 ± 1.0
$\Delta\Delta E(\mathrm{B3LYP})$	10	-18.1 ± 3.1	-12.8 ± 2.9	0 ± 2.8
$\Delta \Delta E(\text{BLYP})$	10	-18.1 ± 3.1	-13.2 ± 2.8	0 ± 2.7
$\Delta\Delta E(\mathrm{CFF})$	10	-17.3 ± 1.5	-17.8 ± 1.3	0 ± 1.5
$\Delta E(\text{BLYP})$	45	-123.2 ± 2.9	-118.5 ± 2.6	-106.4 ± 2.4
$\Delta E({\rm CFF})$	45	-52.7 ± 0.3	-52.0 ± 0.3	-35.0 ± 0.4
$\Delta\Delta E(\text{BLYP})$	45	-16.8 ± 1.8	-12.1 ± 1.7	0 ± 1.6
$\Delta\Delta E(\mathrm{CFF})$	45	-17.7 ± 0.5	-17.0 ± 0.6	0 ± 0.6
$\Delta E(\text{BLYP})$	90	-127.5 ± 1.2	-123.2 ± 1.0	-110.2 ± 0.6
$\Delta E({\rm CFF})$	90	-53.1 ± 0.2	-50.6 ± 0.3	-33.0 ± 0.3
$\Delta\Delta E(\text{BLYP})$	90	-18.9 ± 1.5	-12.4 ± 1.4	0 ± 1.4
$\Delta \Delta E(\text{CFF})$	90	-20.1 ± 0.3	-17.5 ± 0.4	0 ± 0.4

Supplementary Table 3: Details of the QM/MM systems.

Atoms treated at QM level in the bound, unbound and the TS2 states, respectively. This number changes depending on the selected MD snapshot. The average value, standard deviation, minimum value, and maximum value of the total number of QM atoms (Natoms), the number of QM atoms belong to protein (Nres), the number of QM atoms belong to water molecules (Nsol) are reported, respectively.

	Natoms	Nres	Nsol
Bound	248 ± 15	196 ± 13	21 ± 5
TS2	262 ± 16	205 ± 13	26 ± 7
Unbound	219 ± 9	0 ± 0	188 ± 9
Bound(min)	210	170	9
TS2(min)	235	186	18
Unbound(min)	196	0	165
Bound(max)	304	246	27
TS2(max)	308	228	39
Unbound(max)	238	0	207

Supplementary Note 1: Ratchet&Pawl MD

Ratchet&Pawl Molecular Dynamics $(rMD)^{2,3}$ is an out-of-equilibrium sampling technique. Starting from one conformation of the system, we define a scalar ratcheting coordinate r(t) that can drive the system toward a second state of interest (reaching the ratcheting coordinate r_{final} that corresponds to the end point). We then apply a bias that dumps thermal fluctuations in the direction opposite to the end point. This is achieved by imposing the following "ratchet" potential:

$$V_{\text{rMD}}(\xi(t)) = \begin{cases} \frac{k}{2} (\xi(t) - \xi_m)^2 & \xi(t) > \xi_m(t) \\ 0 & \xi(t) \le \xi_m(t) \end{cases}$$
(1)

where

$$\xi(t) = (r(t) - r_{\text{final}})^2$$

and

$$\xi_m(t) = \max_{t \in [0,T]} \xi(t)$$

Supplementary Note 2: Infrequent and Frequency Adaptive Metadynamics

In the case of zero deposition of bias in the transition state, metadynamics allows to predict the timescale of a transition between two different states of the system. ⁴ To achieve this, one deposits bias with a slower rate than standard metadynamics (i.e. infrequent metadynamics), giving the possibility to the system to have a transition between those deposition. In these conditions, the simulated time t_{sim} in a metadynamics run turns out to be related to the real time t with the equation

$$t = \alpha t_{\rm sim},\tag{2}$$

where α is the acceleration factor, and can be computed from the bias deposited by metadynamics $V(s, t_{\text{sim}})$ as

$$\alpha = \langle \exp(\beta U(s, t_{\rm sim})) \rangle \tag{3}$$

where $\beta = (k_B T)^{-1}$, s is the position of the system in the CV space at time t_{sim} , and $\langle \cdot \rangle$ is the average over all the simulation.

After a certain number of different metadynamics runs, we obtain a set of residence times of the ligand inside the receptor. To verify that we are simulating a rare event and no bias is deposited on the transition state. (i) we fit on our empirical distribution the theoretical cumulative distribution function (CDF) of a homogeneous Poisson process of characteristic time

$$CDF_{Poisson} = 1 - \exp(-t/\tau) \tag{4}$$

to the empirical CDF of our residence times. (ii) We can perform a Kolmogorov-Smirnov test between the fitted rare event CDF and the residence times empirical CDF: if the p-value is >0.15, we can safely assume that the obtained times are extracted from a rare event distribution.⁵

Frequency Adaptive metadynamics⁶ is a variant of Infrequent Metadynamics. There, the deposition frequency is not fixed, but it is adaptively slowed down when the acceleration factor α grows. This change permits to speed up the bias filling of the initial free energy minima, allowing to reduce the simulation time, and maintaining the "infrequent" behavior near the transition states. Practically, one choose an initial deposition τ_0 , an upper deposition time limit τ_c (that have to be in the same order of a usual infrequent metadynamics deposition time), and a parameter θ , that drives the dynamics of the deposition time. The deposition time τ_{dep} at time t reads

$$\tau_{\text{dep}} = \min\left(\tau_0 \cdot \max\left(\frac{\alpha(t)}{\theta}, 1\right), \tau_c\right)$$
(5)

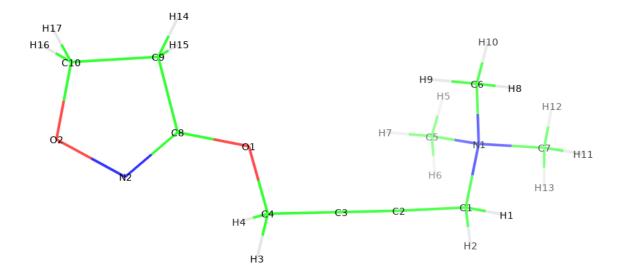
where $\alpha(t)$ is the instantaneous acceleration factor, defined as

$$\alpha(t) = \exp(\beta U(s(t), t)) \tag{6}$$

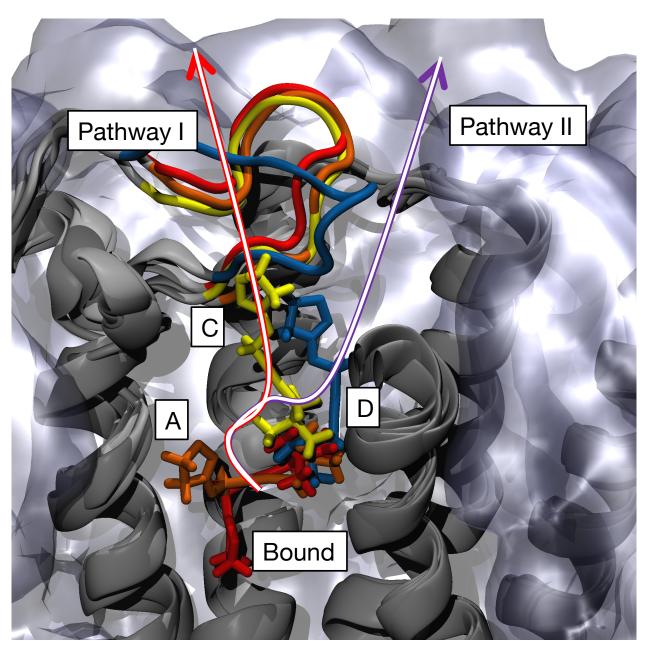
Supplementary Note 3: Details on System Preparation

Our preparation protocol is the following. The X-ray structure of the M₂/iperoxo complex in presence of a nanobody was obtained from the Protein Data Bank (code: 4MQS⁷). The lipids composition was chosen to represent that of a model neuronal membrane 8 (i.e., 48%cholesterol, 16% phosphatidylcholine (DPPC and POPC), 16% phosphatidylethanolamine (DOPE), 14% sphyngomielin (SM 18:0), 4% phosphatidylserine (DOPS), and 2% phosphatidylinositol (SOPC)). The receptor was inserted into membrane bilayers following the orientation of the OPM database. 9 The system was then solvated with explicit water molecules. We neutralized the charge of the system and added a salt concentration of 150 nM, following the condition used in experiments.^{7,10} The final simulation box was 112 x 113 x 149 Å³, containing around 150.000 atoms in total. The parametrization of the model is the same as in our previous work, 11 except for the calculation of the atomic charges for iperoxo: before the calculation of the charges with the restrained electric potential fitting method (RESP)¹² at HF/6-31G* level of theory, we performed a further geometric optimization at B3LYP/6-31+G(d,p) level of theory to have a better estimation of the single-point charges employing Gaussian 09.¹³ The simulations were performed using GROMACS 2018.4.¹⁴ patched with PLUMED 2.5. 15 The protein, membrane, counterions, and water were described by the AM-BER ff14SB force field, ¹ the CHARMM-GUI¹⁶ and Slipids, ^{17,18} Joung and Cheatham force fields, 19 and TIP3P, 20 respectively. We used the generalized AMBER force field (GAFF) 21 for the ligand. A cutoff distance of 12 Å was used for the van der Waals and short-range electrostatic interactions, and the long-range electrostatic interactions were computed with the particle-mesh Ewald (PME) summation method²² using a grid point spacing of 1 Å. Long-range dispersion corrections to the pressure and potential energy were considered. ²³

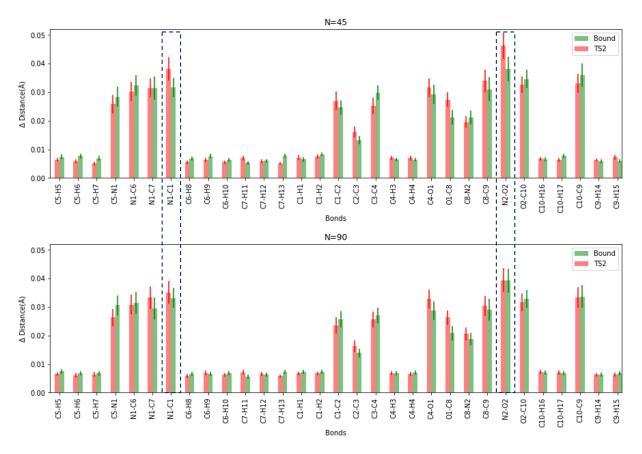
Our equilibration protocol started with the lipid tails: restraining all the other atoms, the lipid tails were energy minimized for 1000 steps using the steepest descent algorithm and subsequently melted with an NVT run for 0.5 ns at 310 K. After this initial run, we activated pressure coupling for a 10 ns run, equilibrating the entire system to a stable volume, fixing only the positions of the protein and the ligand. In this equilibration phase, we noticed that after the first 5 ns the volume of the box reached a plateau, fluctuating around its equilibrium value without any canonical drift. We then released the protein restraints for a further 0.5 ns keeping the pressure of 1 bar and physiological temperature (i.e., 310 K). After these minimization and equilibration procedures, we performed a long unrestrained MD simulation for 0.7 μ s. The temperature was kept constant using velocity rescaling thermostat²⁴ with solvent, solute, and membrane coupled to separate heat baths with coupling constants of 0.5 ps. The pressure was maintained constant with a semi-isotropic scheme, so that the pressure in the membrane plane was controlled separately from the pressure in the membrane normal direction, and the Parrinello-Rahman barostat ^{25,26} was applied with a reference pressure of 1 bar, a coupling constant of 2 ps, and a compressibility of $4.5 \cdot 10^{-5}$ bar⁻¹ for both the directions.



Supplementary Figure 1: Iperoxo molecule in lines representation. The atom names used in our parameterization are labeled on each atom.



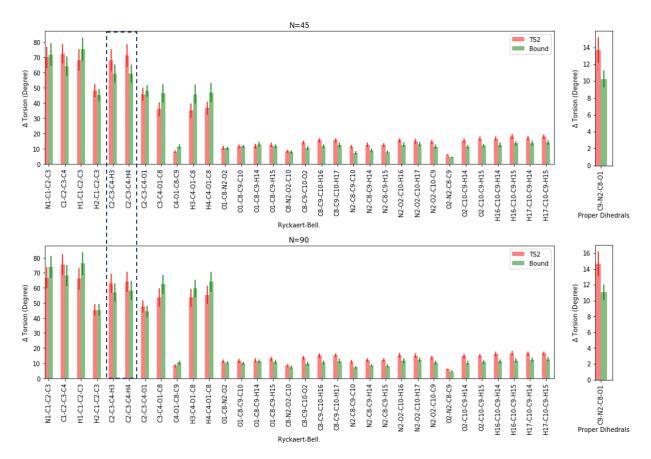
Supplementary Figure 2: Visual representation of the two unbinding pathways of iperoxo ligand from the M_2 receptor as identified in ref. ¹¹ Iperoxo (stick representation) starts from the Bound state (red) and for both the pathways I (red arrow) and II (violet arrow) passes through state A (orange) and C (yellow). If the extracellular loop 2 (ECL2) is the crystallographic position (yellow, orange and red), the ligand can make a translation toward the solvent, reaching the fully solvated state (pathway I); if ECL2 is rotated, it hampers the motion of the ligand, that is forced to reach the intermediate state D (blue), moving then toward the solvent.



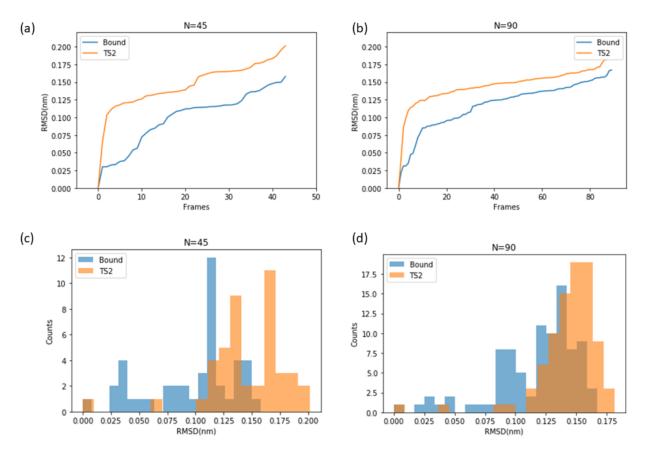
Supplementary Figure 3: Comparison of the bond lengths of the ligand at the Bound and TS2 states. Here, N is the number of conformations considered in statistics. The bond lengths of N2-O2 and N1-C1 are deformed less significantly at TS2 state when more conformations are considered in statistics. Atom names are labeled on the ligand in Figure 1.



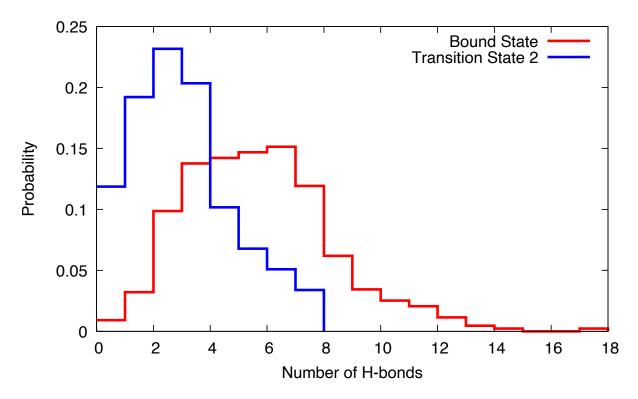
Supplementary Figure 4: Comparison of the bond angles of the ligand at the Bound and TS2 states. Here, N is the number of conformations considered in statistics. Atom names are labeled on the ligand in Figure 1.



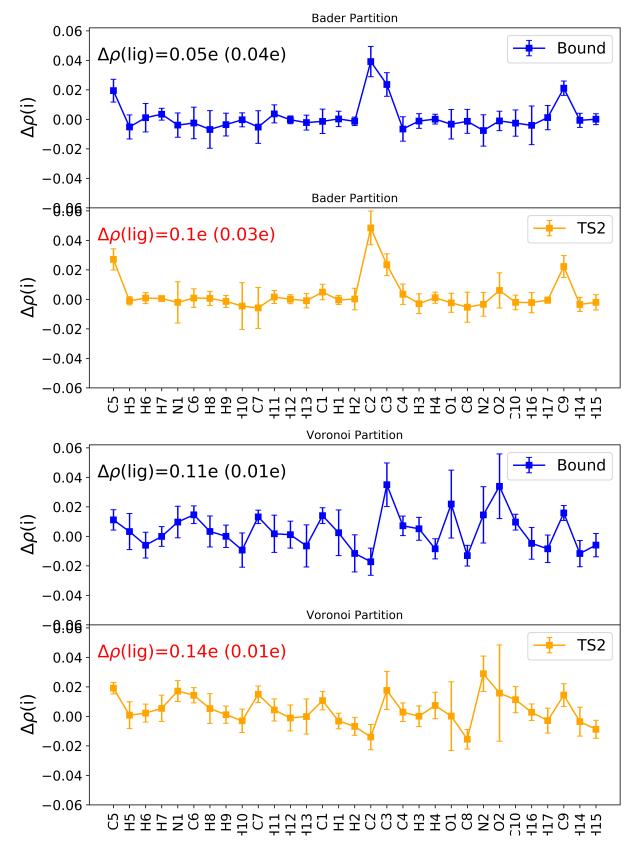
Supplementary Figure 5: Comparison of the dihedrals of the ligand at the Bound and TS2 states. Here, N is the number of conformations considered in statistics. The torsion of C2-C3-C4-H4 are deformed less significantly at TS2 state when more conformations are considered in statistics. Atom names are labeled on the ligand in Figure 1.



Supplementary Figure 6: Comparison of the RMSD fluctuation range of the ligand at the Bound and TS2 states. Here, N is the number of conformations considered in the statistics. A random conformation within N was used as a reference in RMSD calculation, then the calculated RMSD values were sorted ascendingly in (a) and (b).



Supplementary Figure 7: Number of ligand-protein H-bonds for 400 conformations of the bound state (red) and 130 conformations of TS2 (blue), as obtained with **RESP-B3LYP** calculations. We computed the number of H-bonds within a cutoff, using a threshold of 3.5Å and 30° for the distance between the donor and the acceptor and the angle defined by donor, hydrogen and acceptor, respectively. This can lead to a higher number of H-bonds with respect to the real ones.



Supplementary Figure 8: Change in electronic density for single atoms, partitioned using the Voronoi²⁷ and Bader²⁸ schemes, upon formation of the bound state and TS2. In both schemes, the change in density increases on passing from bound state to TS2.

References

- (1) 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.
- (2) 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.
- (3) Tiana, G.; Camilloni, C. Ratcheted molecular-dynamics simulations identify efficiently the transition state of protein folding. *The Journal of chemical physics* **2012**, *137*, 235101.
- (4) Tiwary, P.; Parrinello, M. From metadynamics to dynamics. *Physical review letters* **2013**, *111*, 230602.
- (5) 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.
- (6) 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.
- (7) 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.
- (8) 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.
- (9) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research* **2012**, *40*, D370–D376.
- (10) Schrage, R.; Holze, J.; Klöckner, J.; Balkow, A.; Klause, 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.
- (11) 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. *Journal of chemical theory and computation* **2019**, *15*, 3354–3361.
- (12) 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.
- (13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian 09. 2009.
- (14) 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.
- (15) 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.

- (16) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 1859–1865.
- (17) Ermilova, I.; Lyubartsev, A. P. Extension of the slipids force field to polyunsaturated lipids. *The Journal of Physical Chemistry B* **2016**, *120*, 12826–12842.
- (18) Jämbeck, J. P.; Lyubartsev, A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *The journal of physical chemistry B* **2012**, *116*, 3164–3179.
- (19) Joung, I. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The journal of physical chemistry B* **2008**, *112*, 9020–9041.
- (20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics 1983, 79, 926–935.
- (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) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. The Journal of chemical physics 1995, 103, 8577–8593.
- (23) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *The journal of physical chemistry B* 2007, 111, 13052–13063.
- (24) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling.

 The Journal of chemical physics 2007, 126, 014101.

- (25) Parrinello, M.; Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters* **1980**, *45*, 1196.
- (26) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.
- (27) 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.
- (28) Bader, R. F. Atoms in molecules. Accounts of Chemical Research 1985, 18, 9–15.