Supporting information for:

Metadynamics simulations of ligands binding to proteins' surfaces: a novel tool for rational drug design

Ke Zuo[a,b,c,d], Agata Kranjc[a], Riccardo Capelli[e], Giulia Rossetti[a,f,g], Rachel Nechushtai[c], Paolo Carloni[a,b,h,*]

[a] Computational Biomedicine, Institute of Advanced Simulation IAS−5 and Institute of Neuroscience and Medicine INM−9, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[b] Department of Physics, RWTH Aachen University, 52074 Aachen, Germany

[c] The Alexander Silberman Institute of Life Science, The Hebrew University of Jerusalem, Edmond J. Safra Campus at Givat Ram, 91904 Jerusalem, Israel

[d] Department of Physics, Università degli Studi di Ferrara, 44121 Ferrara, Italy

[e] Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy

[f] Jülich Supercomputing Center (JSC), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[g] Department of Neurology, Faculty of Medicine, RWTH Aachen University, 52074 Aachen, Germany

[h] JARA Institute: Molecular Neuroscience and Imaging, Institute of Neuroscience and Medicine INM-11, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

1. Computational details

Ligand **1** (2-benzamido-4-(1,2,3,4-tetrahydronaphthalen-2-yl)-thiophene-3-carboxylate) was constructed by Chem3D software. Then it was geometrically optimized based on the MM2 force field[1] by means of Chem3D, with the convergence criteria of the minimal root mean squared gradient less than 0.01 kcal·mol^{-1}·Å^{-1}.

NAF-1 structure was obtained from the PDB database (PDB ID: 4OO7[2]) and energy minimized by Maestro, using OPLS 2005 force field[3, 4]. The ligand was then docked to the NAF-1 structure using the program Glide[5, 6]. For oxidized NAF-1, both iron ions of the [2Fe-2S] cluster were assigned the +3 charges, while for the reduced cases, the histidine-bound iron was assigned charge +2, and the other cysteine-bound iron was kept at charge +3. The Fe-bound histidine was considered deprotonated in all four cases since Glide lacks parameters for the protonated histidine. The protein's docking area was defined with a cubic grid with dimensions of 20 Å and a spacing between the grid points of 1 Å. The center of the cubic grid was determined by three residues at the ligand binding area, Lys95, His114, and Lys81. The remaining docking parameters were left as default. The GlideScore scoring function[2] was used to evaluate the predicted ligand binding modes in the **1**·NAF-1 complex structures, and the top-ranked one was used for MD simulations.

The **1**·NAF-1 complex was embedded in a dodecahedron-shaped water box with a minimum distance of 30 Å from the box edge to the protein surface. Na^+ and Cl^- ions were added to neutralize the system and to simulate the ionic strength used in the experiments (~0.1 mM)[7]. Periodic boundary conditions were applied. The AMBER 99SB- ILDN force field[8, 9], GAFF[10] with AM1-BCC[11] partial charges, TIP3P[12], and Åqvist potentials[13] were used for describing NAF-1 protein, ligand **1**, water molecules, and counterions, respectively. The parameters for the oxidized and reduced [2Fe-2S] clusters were taken from our previous work[14].

In the pre-equilibrium phase, the geometry of the **1**·NAF-1 complexes were optimized by steepest descent minimization run for 50,000 steps followed by 50,000-step conjugate gradient minimization. After the geometrical optimization, the system was heated to 300 K in 1-ns simulated annealing, and then underwent 50-ns NVT (310 K) and 50-ns NPT (310 K and 1 bar) equilibrations, respectively. Finally, the 100-ns production trajectory

was collected. All four states (**I**–**IV**, Tab. 1 in the main text) of NAF-1 were considered. The calculations were performed with GROMACS 2019.2[15, 16].

The particle mesh Ewald (PME) method[17] was used for long-range electrostatic interactions calculation. The cutoff value for short-range electrostatic and van der Waals (vdW) interactions was set to 1.4 nm. During the simulations all chemical bonds were constrained by the LINCS algorithm[18]. The Nosé-Hoover thermostat[19, 20] and Parrinello-Rahman barostat[21] were used to keep the temperature (310 K) and pressure (1.0 bar) constant, respectively. The time constant for the temperature and the pressure coupling was set to 0.4 and 0.8 ps, respectively. The time integration step was 2 fs. The last snapshot of the simulation was used as the initial structure of volume-based metadynamics[22].

In the volume-based metadynamics simulation[22], used here to predict poses and potency of the ligands, the spherical coordinates, radial distance ρ, azimuthal angle θ, and polar angle φ were used as the collective variables (CVs). The ligand was limited to a sphere of radius ρ_s = 30 Å, centered in center of mass of NAF-1. The constant of the repulsive potential at the spherical boundary was set to 10,000 kJ·mol^{-1}·nm^{-1}. The widths of Gaussians for ρ, θ, and φ were 1 Å, π/16 rad, and π/8 rad, respectively. The initial height of the Gaussian was 1.2 kJ·mol^{-1}. The bias factor was set to 20 with a deposition rate of 1 ps^{-1}. All the simulations were carried out employing GROMACS 2019.2[15, 16] patched with PLUMED-2.6.0[23, 24] at the temperature 310 K and the pressure of 1 bar using the same thermostat and barostat as in the MD simulations described above.

The free energy landscape was projected on the distance from the ligand center to the [2Fe-2S] cluster center (CV1) and the number of H-bonds and salt bridges (CV2) defined by Eq. 1:

$$n = \sum_{ij} \frac{1 - \left(r_{ij}/r_0\right)^a}{1 - \left(r_{ij}/r_0\right)^b} \tag{1}$$

where a and b were set to 8 and 12, respectively. r_0 was set to 2.5 Å[25]. Atom i and j represents the donor atom and acceptor atom, respectively. The binding free energy ΔG_b^0 was calculated by Eq. 2 and 3:

$$\Delta G_b^0 = \Delta G_{Metad} - T\Delta S \quad (2)$$

$$-T\Delta S = RT\ln\left(\frac{V^0}{\frac{4}{3}\pi\rho_s^3 - V_{prot}}\right) \quad (3)$$

where R is the gas constant, T is the temperature, and V^0 is the standard volume. ΔG_{Metad} is the binding free energy obtained by metadynamics, and V_{prot} is the volume of the NAF-1 protein (42,400 Å^3).

2. Binding free energy and poses of 1 in different NAF-1 states: NAF-1(II), NAF-1(III), and NAF-1(IV)

The lowest minimum of **II** (−6.1 ± 0.8 kcal/mol) is higher than that of **I** (−7.4 ± 0.8 kcal/mol), indicating that His114 protonation reduces the lowest binding affinity. Contrary to state **I**, in state **II** we observed binding poses of **1** interacting with the [2Fe–2S] cluster binding region, the L2 loops of β-cap, and L3 loop of NAF-1. However, the reduction of [2Fe–2S] clusters (complex **III** and **IV**) decrease the lowest binding affinity of **1** by approximately 1 kcal/mol. The protonation of His114 in state **IV** could be at the origin of decreased binding affinity of **1** (Tab. 1). Unlike in state **II**, in state **IV** there is no local minimum where **1** binds to the L2 loop of β-cap. On the other hand, in one local minimum we observed that **1** binds to the N-terminal. However, it may be biologically irrelevant because the N-terminal is connected to the transmembrane α-helices, which are absent in our simulation. In state **IV**, two local minima are identified where **1** interacts with the β1/L1 loop and L1/L3 loops.

Interestingly, in the case of state **II**, six local minima were found in the free energy landscape. In minimum 1 (and 4, −6.1 ± 0.8 kcal/mol and −5.0 ± 0.8 kcal/mol, respectively), the phenyl ring of **1** forms hydrophobic interaction with Leu91 (at times, with Asp90). In contrast, the thiophene ring forms hydrophobic interactions with Leu93, Ala96, Tyr98, and Ala109 (with Leu91) and a metastable interaction with Thr106 (an H-

bond is found between Ser92 and the carbonyl oxygen). In addition, the tetralin group forms hydrophobic interactions with Ala97, Cys110, and Lys95 (with Leu93, Ala96, Ala97, Tyr98, Ala107, Cys110, also, at times, with Asp111). In minimum 2 (−5.7 ± 0.8 kcal/mol), the phenyl ring of **1** forms hydrophobic interactions with Ala97, Pro127, and Ile129. The carbonyl oxygen forms an H-bond with His114 and interacts with Lys81 via two water molecules. On the other hand, the thiophene ring forms vdW interactions with Gly112. The tetralin ring interacts, at times, with Asn115 and Lys116. As for minima 3 and 5 (−5.7 ± 0.8 kcal/mol, −4.7 ± 0.8 kcal/mol, respectively), **1** is bound to the L2 loop. Its phenyl group interacts, at times, with Asp90, Leu91, and Ser92 (at times, with Glu89 and Ser92). The thiophene ring has polar vdW interactions with Glu89 (Thr94). The carboxylate oxygen forms a salt bridge with Lys132 (and interacts to the carbonyl oxygen of the main chain of Leu93 via water-mediated H-bond). In minimum 6 (−4.7 ± 0.8 kcal/mol), the phenyl ring of **1** forms hydrophobic interactions with Val83, Pro127, Ile128, and His114. The thiophene ring and the carboxylate group form vdW interactions with Lys95, Ala97, and Cys110. Besides, the tetralin ring forms hydrophobic interactions with Leu93, Ala96, and Ala109 (see Fig. S1).

In case **III**, four minima are identified in the free energy landscape. In minimum 1 (−6.7 ± 0.8 kcal/mol), the phenyl ring forms hydrophobic interactions with Pro80 and Val83. The carboxylate oxygens form H-bonds with two water molecules; the latter forms H-bonds to Lys81 and Gln76, respectively; and the tetralin group forms metastable interactions with Glu78 and Asn79. In minimum 3 (−4.3 ± 0.8 kcal/mol), the phenyl ring forms hydrophobic interactions with Ala96, Ala97, and Cys110, whereas the carboxylate oxygen forms salt bridges with Lys95. A metastable interaction is also observed between the tetralin group and Glu85. In minimum 4 (−4.1 ± 0.8 kcal/mol), the tetralin ring forms hydrophobic interactions with Ile75, and Phe107, and it also interacts, at times, with Lys74, Gln76, and Lys77. As for minimum 2, **1** binds close to the N-terminal where the transmembrane domain connects, which may be biologically irrelevant, thus is not discussed here. (Fig. S2).

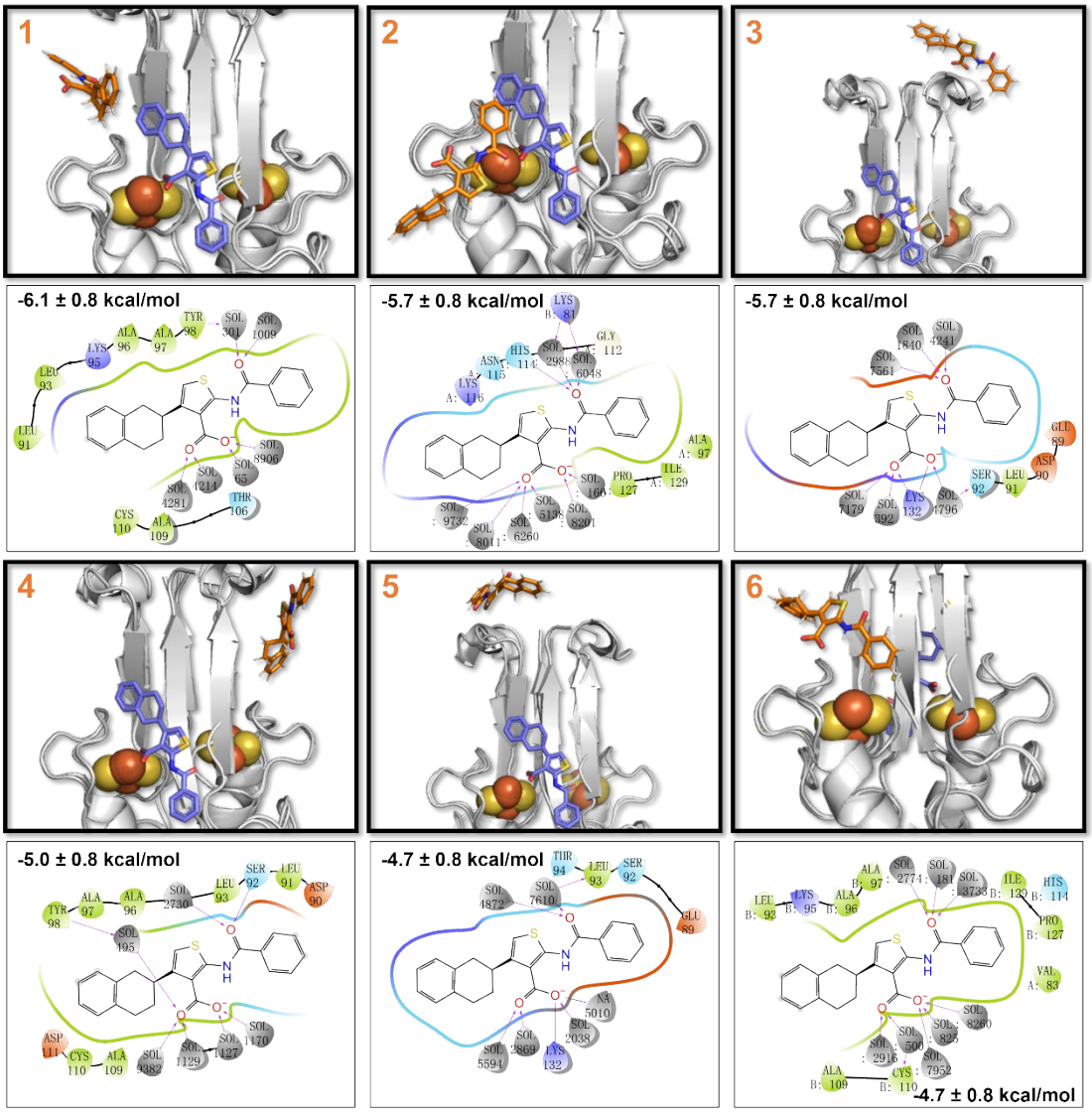

Figure S1. The ligand/protein interactions of 1·NAF-1(II) as observed in the six local minima of the free energy landscape at 310 K. The crystal binding pose is shown in the violet sticks as a reference, while the local minima representative binding poses found in metadynamics simulations are depicted in orange.

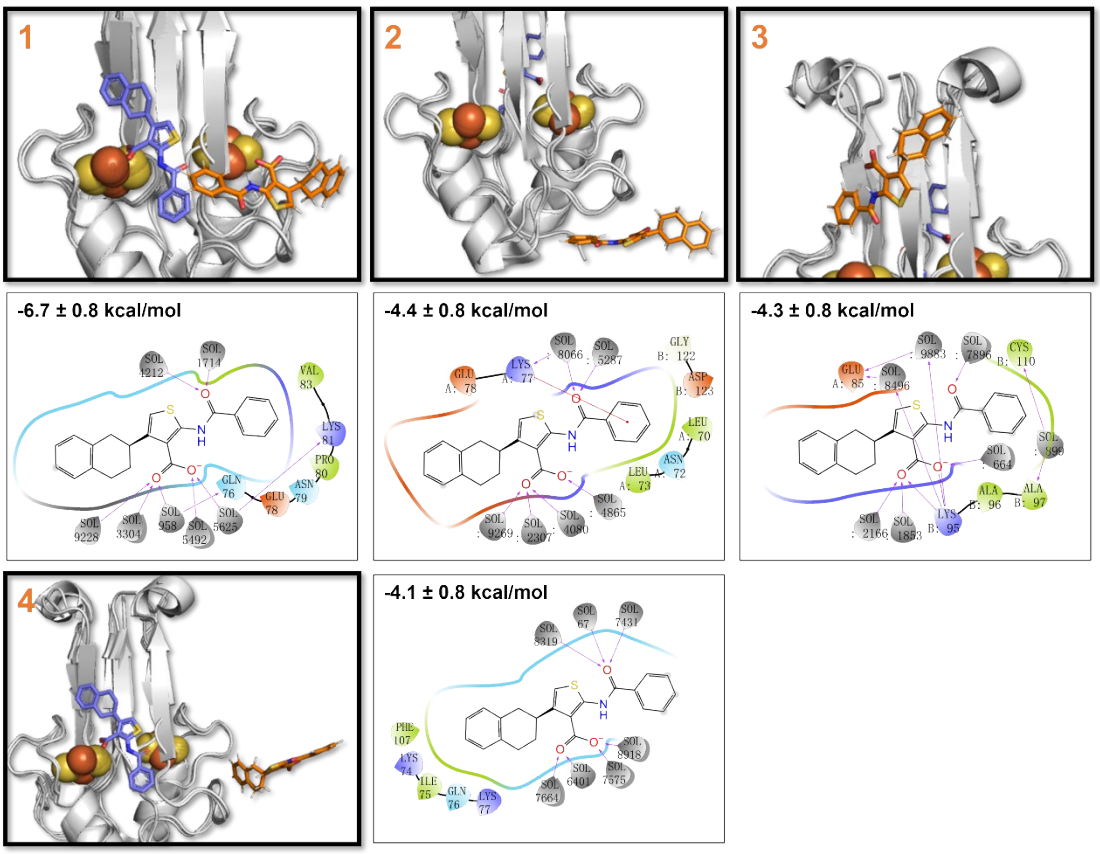

Figure S2. The ligand/protein interactions of 1·NAF-1(III) as observed in the four local minima of the free energy landscape at 310 K. The crystal binding pose is shown in the violet sticks as a reference, while the representative binding poses of **1** are depicted in orange.

In the case of the state **IV**, four local free energy minima were found. Specifically, in minimum 1 (−5.8 ± 0.8 kcal/mol), the phenyl ring of the ligand **1** forms hydrophobic interactions with Ile86, Pro108, and sometimes, with Asn84. The carbonyl oxygen forms an H-bond with Tyr98. The thiophene ring forms hydrophobic interactions with Leu91, Leu93, Ala96, and Ala109, whereas the tetralin ring forms hydrophobic interactions with Ala97 and Cys110 and metastable interactions with Lys95. In minimum 3 (−3.9 ± 0.8 kcal/mol), the phenyl ring of **1** forms hydrophobic interactions with Pro80 and Val83. The carbonyl oxygen forms water-bridged H-bonds with Asn79, whereas the carboxylate oxygen forms a direct H-bond with Gln76. In addition, the tetralin ring interacts, at times, with Asn84 and Glu85. In minimum 4 (−3.4 ± 0.8 kcal/mol), the phenyl ring forms hydrophobic interactions with Pro80 and Val83. The carboxylate oxygens forms a water-bridged H-bond with Lys81 and Gln76, respectively. Additionally, the tetralin group forms

metastable interactions with Glu78 and Asn79 (Fig. S3). As for minimum 2, **1** binds to NAF-1 where the transmembrane helices connected, therefore it may be biological irrelevant.

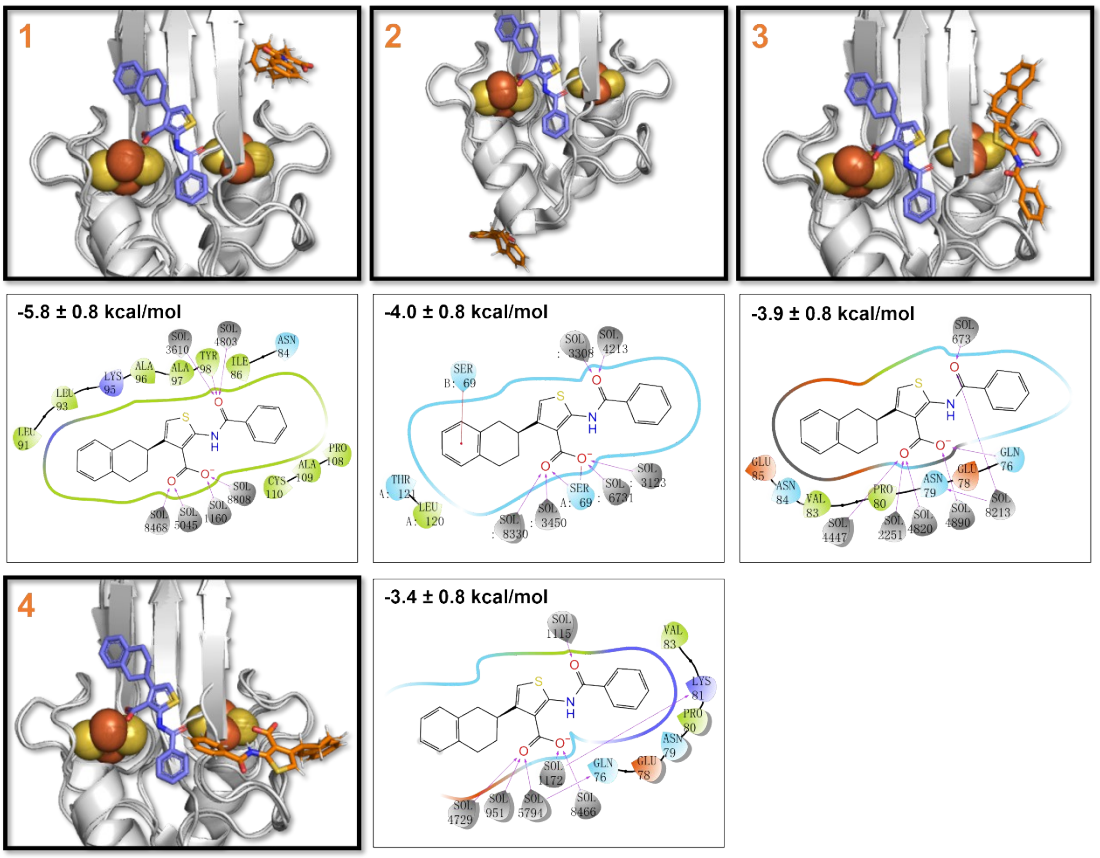

Figure S3. The ligand/protein interactions of 1·NAF-1(IV) as observed in the local minima of the free energy landscape at 310 K. The crystal binding pose is shown in the violet sticks as a reference, while the representative binding poses are presented in orange.

References

1. N. L. Allinger, *J Am Chem Soc*, 1977, **99**, 8127-8134.
2. A. R. Conlan, H. L. Axelrod, A. E. Cohen, E. C. Abresch, J. Zuris, D. Yee, R. Nechushtai, P. A. Jennings and M. L. Paddock, *J Mol Biol*, 2009, **392**, 143-153.
3. W. L. Jorgensen, D. S. Maxwell and J. TiradoRives, *J Am Chem Soc*, 1996, **118**, 11225-11236.
4. E. Harder, W. Damm, J. Maple, C. J. Wu, M. Reboul, J. Y. Xiang, L. L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J Chem Theory Comput*, 2016, **12**, 281-296.
5. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J Med Chem*, 2004, **47**, 1739-1749.

6. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J Med Chem*, 2004, **47**, 1750-1759.
7. H. B. Marjault, O. Karmi, K. Zuo, D. Michaeli, Y. Eisenberg-Domovich, G. Rossetti, B. de Chassey, J. Vonderscher, I. Cabantchik, P. Carloni, R. Mittler, O. Livnah, E. Meldrum and R. Nechushtai, *Commun Biol*, 2022, **5**.
8. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins-Structure Function and Bioinformatics*, 2010, **78**, 1950-1958.
9. E. J. Sorin and V. S. Pande, *Biophys J*, 2005, **88**, 2472-2493.
10. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J Comput Chem*, 2004, **25**, 1157-1174.
11. A. Jakalian, D. B. Jack and C. I. Bayly, *J Comput Chem*, 2002, **23**, 1623-1641.
12. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys*, 1983, **79**, 926-935.
13. J. Aqvist, *J Phys Chem-Us*, 1990, **94**, 8021-8024.
14. K. Zuo, H. B. Marjault, K. L. Bren, G. Rossetti, R. Nechushtai and P. Carloni, *J Biol Inorg Chem*, 2021, **26**, 763-774.
15. H. J. C. Berendsen, D. Vanderspoel and R. Vandrunen, *Comput Phys Commun*, 1995, **91**, 43-56.
16. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19-25.
17. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J Chem Phys*, 1995, **103**, 8577-8593.
18. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J Comput Chem*, 1997, **18**, 1463-1472.
19. S. Nose, *Mol Phys*, 1984, **52**, 255-268.
20. W. G. Hoover, *Phys Rev A*, 1985, **31**, 1695-1697.
21. M. Parrinello and A. Rahman, *J Appl Phys*, 1981, **52**, 7182-7190.
22. R. Capelli, P. Carloni and M. Parrinello, *J Phys Chem Lett*, 2019, **10**, 3495-3499.
23. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput Phys Commun*, 2014, **185**, 604-613.
24. M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banas, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. C. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Lohr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Sponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth and A. White, *Nat Methods*, 2019, **16**, 670-673.
25. A. Bochicchio, G. Rossetti, O. Tabarrini, S. Krauss and P. Carloni, *J Chem Theory Comput*, 2015, **11**, 4911-4922.