Supplementary Information

A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitors Design

Carlos A. Ramos-Guzmán, J. Javier Ruiz-Pernía*, Iñaki Tuñón*

Departamento de Química Física, Universidad de Valencia, 46100 Burjassot (Spain)

*To whom correspondence should be addressed:
ignacio.tunon@uv.es
j.javier.ruiz@uv.es

Methodology

Classical Molecular Dynamics simulations

The Michaelis complex was built from the structure with PDB code 7BQY. This contains the crystallographic data for the $3CL^{pro}$ of SARS-CoV-2 forming a covalent complex with the N3 inhibitor (with a resolution of 1.7 angstrom).[1] Crystallographic water molecules were preserved because they can play an important role in modulating the properties of the catalytic dyad, in particular the water molecule bridging between His41-His164 and Asp187.[2,3]

The N3 inhibitor was parameterized following the non-standard residue parameterization procedure implemented in Amber using the Antechamber program[4] from the AmberTools18[5] package. For this purpose, terminal residues of the inhibitor (named as 02J PJE and 010 in the PDB file) were capped using NME and ACE groups respectively. Atomic charges were then obtained using the Restrained Electrostatic Potential (RESP) method[6] at the HF/6-31G* level.

The protonation states of the titratable groups were determined using PROPKA3.0[7] at pH 7.4. For neutral histidine residues, the δ/ε protonation state was determined after visual inspection of the x-ray structure. In particular, the pair of histidine residues closer to the active site (His41 and His164) were modelled in their neutral states, protonated at δ and ε positions, respectively (HD41-HE164). This combination has been recently shown to be the most stable N3-bound state .[2] Then, the system was built using the tleap tool from AmberTools,[5] with regular amino acids being described using the ff14SB forcefield.[8] The charge of the enzyme-inhibitor complex was neutralized adding Na^+ atoms. The system was solvated into a box of TIP3P water molecules,[9] with a buffer region of at least 12 Å from any protein/substrate atom to the limits of the simulation box. The resulting system was then minimized using 500 steps of steepest descent method followed by the conjugate gradient method, until the root mean square of the gradient was below 10^{-3} kcal·mol^{-1}Å^{-1}. A linear heating ramp was used to rise the temperature from 0 to 300 K along 120 ps followed by 20 ps simulation at 300 K. During the heating process a harmonic potential was used to restrain the backbone heavy atoms with a force constant of 20 kcal·mol^{-1}·Å^{-2}. Afterwards, a set of equilibration simulations in the NPT ensemble (300K and 1 bar) were carried out where the force constants for the positional restraint were gently reduced from 15 to 0 kcal·mol^{-1}·Å^{-2}, decreasing 3 units every 1.25 ns. After 6.25 ns the positional restraint was completely removed and the system run restraint-free for another 1.25 ns. Finally, two replicas of NVT simulations at 300 K were performed with a 2 fs time step using SHAKE[10] during 3.0 and 1.0

μs, respectively. The long-range electrostatic interactions were described using the Particle Mesh Ewald Method,[11,12] while a 10 Å cut-off radius was used to evaluate the short-range interactions. Berendsen barostat and Langevin thermostat were used to control pressure and temperature, respectively, where required. For all classical molecular dynamic simulations AMBER19 GPU version of PMEMD[13,14] was employed.

QM/MM calculations

The free energy surfaces associated to the reaction mechanism were explored using QM/MM simulations. The QM region includes the side chains of the catalytic dyad (Cys145 and His41) and a water molecule (in the mechanism presented in Figure 4) . The QM region selected for the N3 inhibitor includes the backbone of P1 and P1' residues corresponding plus the carbonyl carbon atom of P2. The P1 side chain and the benzene ring in P1' were not included in the QM subsystem. The rest of the system was described at the MM level as previously explained. To describe the QM subsystem we used the B3LYP functional[15,16] and D3 dispersion corrections.[17] Mechanisms were explored using the 6-31G* and 6-31+G* basis sets (see below). As previously reported, this is a good choice to describe the acylation of a peptide substrate by the SARS-CoV-2 protease,[3] providing results in excellent agreement with experiments. A systematic study on cysteine-histidine proton transfer found that the B3LYP functional was the most adequate to obtain an electronic description in agreement with higher level methods.[18] For the reaction step we also used the M06-2X functional[19] with D3 corrections to verify the robustness of our conclusions. As usual, Lennard-Jones parameters of QM atoms were imported from the ff14SB forcefield without additional optimization. All QM/MM calculations were performed using a modified version of Amber18[5,20] coupled to Gausssian16[21] for Density Functional Theory calculations. A cutoff-radius of 15 Å was used for all QM-MM interactions and the temperature was 300 K.

Our Adaptative String Method (ASM)[22] implementation was used to explore the free energy landscape associated to the chemical reaction. *N* replicas of the system (the string nodes) were used to connect the reactants and product regions. These nodes are evolved according to the averaged forces and kept equidistant, converging into the minimum free energy path (MFEP) in a space of arbitrary dimensionality defined by the collective variables (CVs). The bond lengths selected as CVs coordinates to describe the two reaction mechanisms explored in this work can be seen in the panels a) of Figures 4 and S2. After convergence, a path Collective Variable (path-

CV) is defined to measure the advance of the system along the MFEP. This path-CV is the reaction coordinate employed to trace the reaction free energy.

Two mechanistic proposals were explored at the B3LYPD3/MM level. Each of the strings was composed of 96 nodes, which were propagated with a time step of 1 fs until the RMSD of the string felt below 0.1 amu$^{1/2}$·Å for at least 2 ps. Umbrella sampling[23] technique was employed to obtain the free energy profiles along the path-CVs. At least 10 ps simulations were run for every node along the converged path and the Weighted Hystogram Analysis Method (WHAM)[24] was selected as the integration method. To ensure a probability density distribution of the reaction coordinate as homogeneous as possible, the values of the force constants employed to bias the ASM simulations were determined on-the-fly.[22] In these calculations we firstly used the 6-31G* basis set and then the best path was recalculated using the 6-31+G* basis set. Inclusion of diffuse functions in the basis set resulted in a modest reduction of the activation free energy for the mechanism presented in Figure 4 of about 2 kcal·mol^{-1}.

For the study of the proton transfer from Cys145 to His41, a simple coordinate, defined as the antisymmetric combination of the distances of the proton to the donor and acceptor atoms of the catalytic dyad (d(Sγ-H)-d(Nε-H)), was employed. Then, the free energy profile associated to the formation of the catalytic dyad ion pair (IP) was obtained using Umbrella Sampling.[23] The integration was carried out using WHAM method.[22] For this profile only the side chains of the two involved residues were included in the QM region (using the B3LYPD3/6-31+G* level of theory). 40 windows were used separated every 0.06 Å along the reaction coordinate, and a force constant of 600 kcal·mol^{-1}·Å^{-2} was used to drive the reaction coordinate change. For each simulation window the system was first minimized, heated and then production was run for 20 ps. The free energy change was calculated has the average between the forwards and backwards profiles. All the rest of details of the simulations were as described previously.

The calculated activation free energies are compared to experimentally derived values, using standard Transition State Theory expression to convert between rate constants (k_r) and activation free energies ($\Delta G^{\ddagger}$):

$$k_r = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \qquad \text{(eq. S1)}$$

where k_B, h and R are the Boltzmann, Planck and gases constants and T the temperature.

Examples of MM and QM/MM (the TS) trajectories can be accessed through covid.molssi.org/simulations:

https://covid.molssi.org//simulations/#695-us-simulation-of-sars-cov2-3clpro-in-complex-with-the-n3-inhibitor-all-atom-simulation

https://covid.molssi.org//simulations/#11-ps-qmmm-simulation-of-transition-state-of-sars-cov2-3clpro

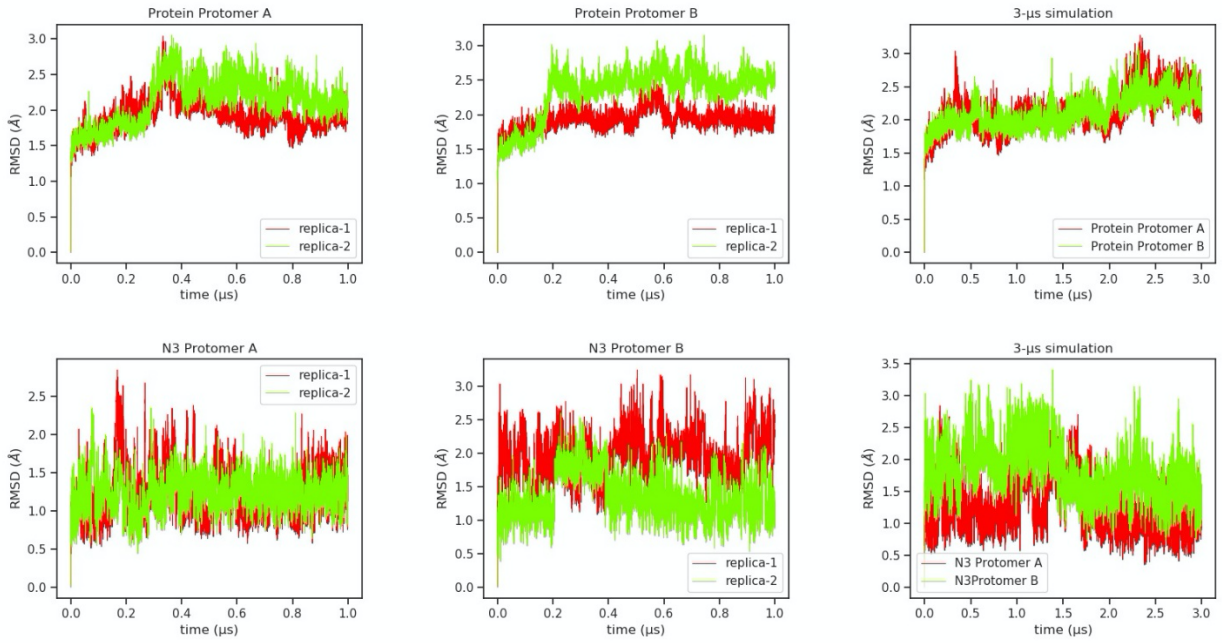

Figure S1. Plot of the RMSD values obtained for the protein (up) and the substrate (down) during the 2 replicas (simulation time of 3.0 μs and 1 μs, respectively) of the noncovalent enzyme-inhibitor complex of the SARS-CoV-2 protease with N3. From left to right, RMSD of protomer A, protomer B. The RMSD has been computed using all non-hydrogen atoms taking as reference the initial structure prepared for the simulations.

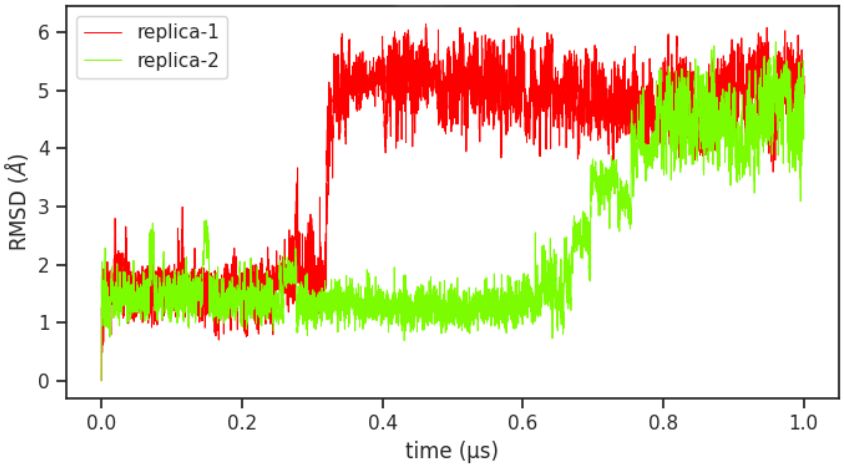

Figure S2. Plot of the RMSD values obtained for the substrate during the 2 replicas (simulation time of 1.0 μs each) of the noncovalent enzyme-inhibitor complex of the SARS-CoV-2 protease with N3 with the catalytic dyad in the IP form.

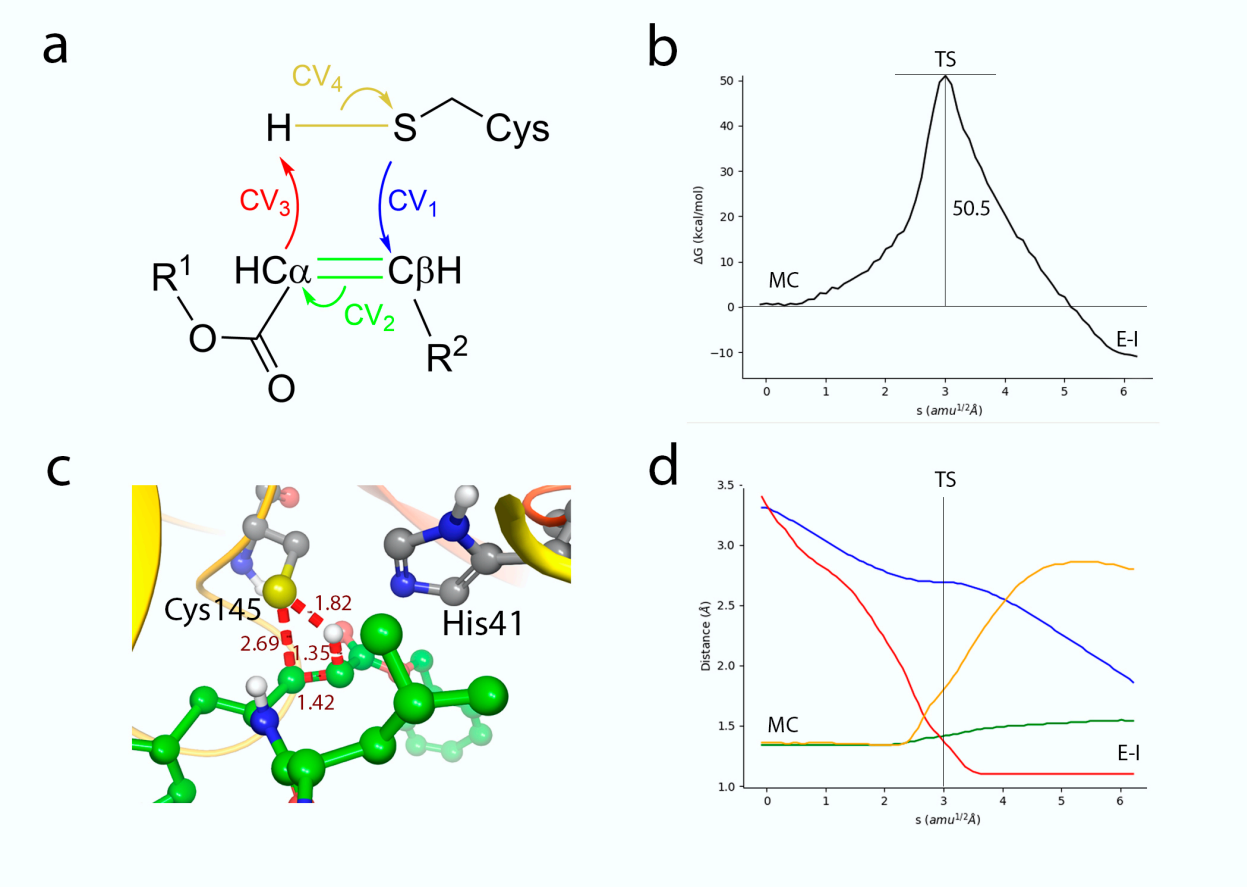

Figure S3. Simulation of the reaction from noncovalent (EI) to covalent complex (E-I) without a proton transfer from Cys145 to His41. **(a)** Collective Variables (CVs) employed to explore the MFEP. **(b)** B3LYPD3/6-31G*/MM free energy profile along the path-CVs for the formation of the covalent E-I complex from EI. **(c)** Representation of TS. The values of the distances (in Å) correspond to the coordinates of the MFEP where the TS is located. Evolution of the distances selected as CVs along the MFEP. The color code corresponds to those shown in Figure S2a. (**d**) Evolution of the distances selected as CVs along the MFEP. The color code corresponds to those shown in Figure S2a.

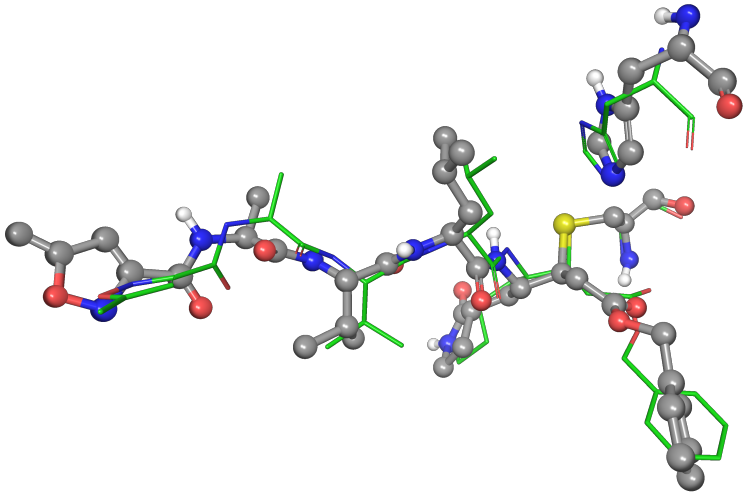

Figure S4. Overlap between the QM/MM (ball & sticks representation) and x-ray (licorice style) structures of the inhibitor in the acylenzyme complex. The x-ray structure corresponds to the PDB file with code 7BQY.

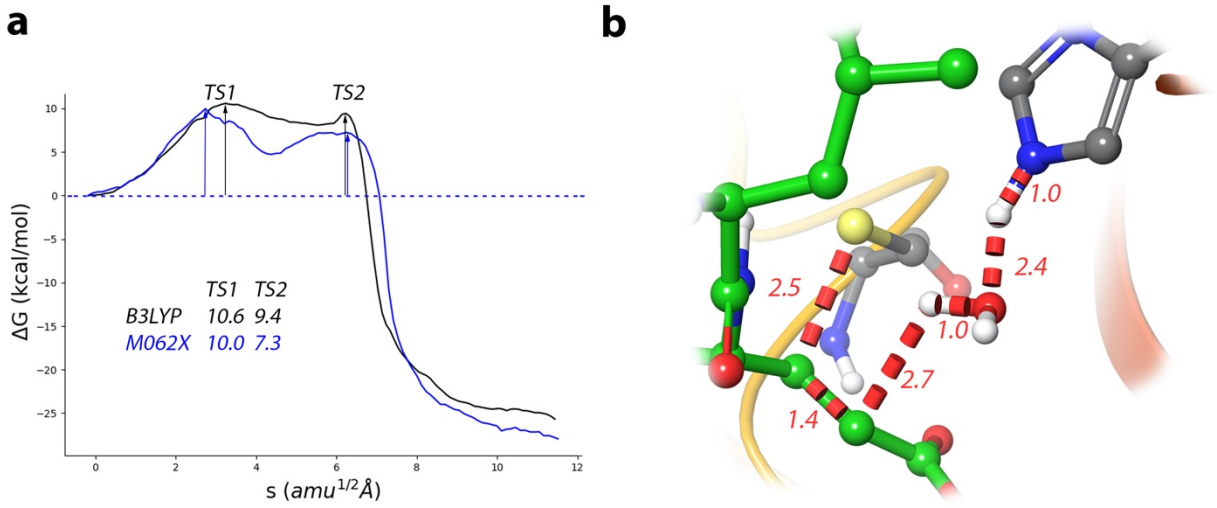

Figure S5. Simulation of the reaction from the IP to the covalent E-I complex. **(a)** B3LYPD3/6-31G*/MM and M062xD3/6-31+G*/MM free energy profiles along the path-CV for the formation of the covalent E-I complex. **(b)** Representation of the M062X structure for TS1, to be compared with the B3LYP one presented in Figure 4d. The values of the distances (in Å) correspond to the coordinates of the MFEP where TS1 is located.

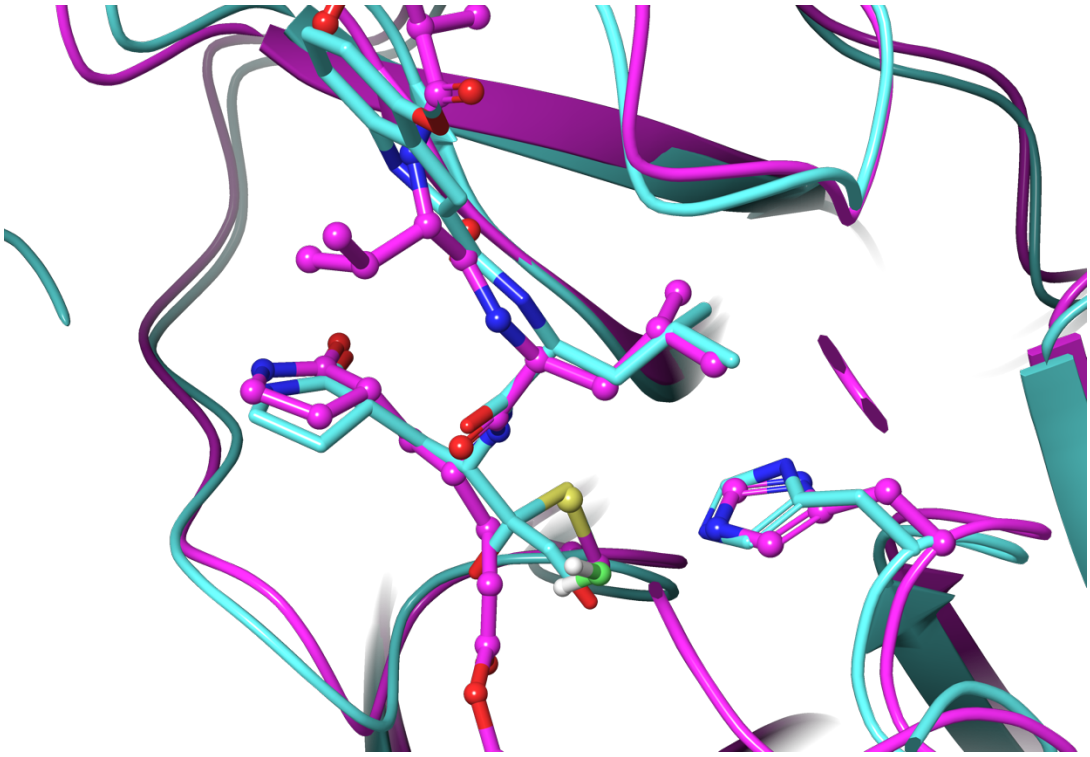

Figure S6. Overlap between the QM/MM TS1 structure (ball & sticks representation with carbon atoms in purple, the oxygen of the water molecule is highlighted in light green) and x-ray structure of the acylenzyme complex formed with PF-00835231 (licorice style with carbon atoms in light blue). The x-ray structure corresponds to the PDB file with code 6XHM.

References

1 Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. X. X. Yang, T. You, X. X. X. X. Liu, X. X. X. Yang, F. Bai, H. Liu, X. X. X. X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, *Nature*, 2020, **582**, 289–293.

2 A. Pavlova, D. L. Lynch, I. Daidone, L. Zanetti-Polzi, M. D. Smith, C. Chipot, D. W. Kneller, A. Kovalevsky, L. Coates, A. A. Golosov, C. J. Dickson, C. Velez-Vega, J. S. Duca, J. V. Vermaas, Y. T. Pang, A. Acharya, J. M. Parks, J. C. Smith and J. C. Gumbart, *Chem. Sci.*, , DOI:10.1039/D0SC04942E.

3 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2020, **10**, 12544–12554.

4 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.

5 D. A. Case, D. S. Cerutti, T. E. I. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *Univ. California, San Fr.*

6 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

7 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.

8 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

9 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

10 J.-P. Ryckaert, G. Ciccotti and H. J. . Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

11 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

12 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

13 S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.

14 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.

15 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

16 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

17 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

18 A. Paasche, T. Schirmeister and B. Engels, *J. Chem. Theory Comput.*, 2013, **9**, 1765–1777.

19 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

20 K. Zinovjev, String-Amber, https://github.com/kzinovjev/string-amber, (accessed 24 June 2020).

21 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, 2016.

22 K. Zinovjev and I. Tuñón, *J. Phys. Chem. A*, 2017, **121**, 9764–9772.

23 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

24 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.