# SUPPLEMENTARY INFORMATION FOR PUBLICATION:

## Estimating the Binding Energetics of Reversible Covalent Inhibitors of the SARS-Cov-2 Main Protease: An *in-silico* study

Ernest Awoonor-Williams[*,†]

[†] Department of Chemistry, Memorial University of Newfoundland, St. John's, NL, A1B 3X9, Canada

### COMPUTATIONAL DETAILS

### Ligand–Protein Model System Setup

The coordinates of the protein–ligand complexes were downloaded from the protein data bank using the accession codes: 6LZE, 7D1M, 7D3I, and 6XHM which correspond to the inhibitor bound structures of 11a, GC373, MI-30, and PF-00835231 ('PF-231'), respectively. **Figure 1**(b) in the main text shows the chemical structures of the $M^{pro}$ inhibitor compounds studied in this work. For the MI-30 compound, the structure of the inhibitor was built from the complexed ligand, MI-23, which bears a similar scaffold. Given that these are covalent bound complexes of the inhibitors to $M^{pro}$, the covalent bond between the active site Cys145 of the protein and the reactive center of the inhibitors was broken prior to running the equilibration molecular dynamics simulation. A stable and equilibrated protein:ligand complex system was obtained before continuing with the absolute binding free energy simulations. In addition, the preferred protonation state of the active site His41 in the simulated complexes was determined from MD structure analyses to inform the binding free energy simulations. Recent work by Gumbart and coworkers[1] have shown that histidine protonation states can impact ligand binding and protein structure dynamics of the $M^{pro}$. In addition, recent work by Tuñón and coworkers[2] have highlighted the presence of different rotameric states of His41 present in X-ray structures of aldehyde/ketone inhibitor bound $M^{pro}$ complexes. For the complexes studied, the rotameric states of His41 were chosen based on the lowest energy conformation, following structure preparation and energy minimization. For the 11a, GC373, and MI-30 inhibitor complexes, the $\boldsymbol{\varepsilon}$-rotamer of His41 of $M^{pro}$ was used, whereas for the PF-231 inhibitor bound complex, the $\delta$-rotameric state of His41 was used. The simulations were performed using the GROMACS molecular dynamics engine (version 2020.4).[3] Two MD repeats with production run time of 10 ns were performed for all protein–ligand complexes. The average ligand RMSD over the course of the simulations (with respect to the initial starting structures) were analyzed to explore structural changes (**Figure S1**; **Figure S2**).

[*] To whom correspondence should be addressed. Email: esaw83@mun.ca

**Table S1. Summary of Tautomeric states of His41 determined from MD simulation. Cys$^{145}$---ligand warhead distance is the distance between the thiol S atom of C145 of the M$^{pro}$ and the reactive carbonyl center of the ligand. The distances reported were calculated using the final structure of the protein: ligand complex obtained from the unrestrained molecular dynamics simulation.**

| Protein:Ligand Complex | His41 state | Cys$^{145}$---ligand warhead distance (Å) | Nearby H–bond interaction molecule with Cys$^{145}$ (Å) |
|---|---|---|---|
| M$^{pro}$:11a | HID | 3.05 | HID41 $^{N}{}_{\varepsilon}$ (2.40) |
| M$^{pro}$:11a | HIE | 3.25 | --- |
| M$^{pro}$:GC373 | HID | 3.25 | HID41 $^{N}{}_{\varepsilon}$ (2.23) |
| M$^{pro}$:GC373 | HIE | 3.30 | --- |
| M$^{pro}$:MI-30 | HID | 3.50 | $H_2O$ molecule (1.98) |
| M$^{pro}$:MI-30 | HIE | 3.30 | $H_2O$ molecule (2.05) |
| M$^{pro}$:PF-231 | HID | 3.35 | HID41 $^{N}{}_{\varepsilon}$ (2.12) |
| M$^{pro}$:PF-231 | HIE | 3.23 | $H_2O$ molecule (1.86) |

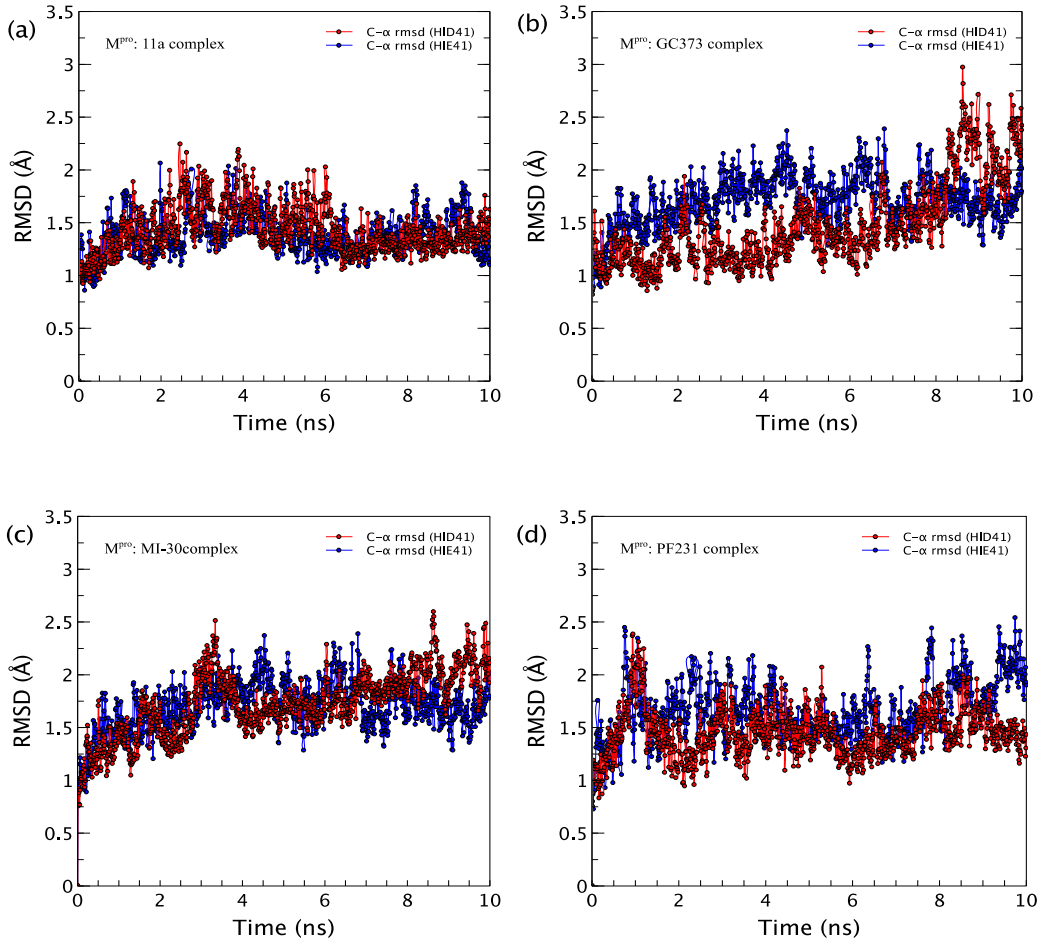


**Figure S1. RMSD of the C$^{\alpha}$ of the protein for the ligand bound Mpro complexes studies: (a) M$^{pro}$:11a complex; (b) M$^{pro}$:GC373 complex; (c) M$^{pro}$:MI-30 complex; and (d) M$^{pro}$:PF231 complex.**

The Amber MD program was also used to perform simulations of the protein–ligand complex with His41 in different states (i.e., HID–neutral, $\delta$-N protonated; or HIE neutral, $\varepsilon$-N protonated) —to provide an independent test on the chosen protonation states for histidine-41 used in this study. MD simulations of 10 ns and 100 ns production runs were performed following minimization (steep descent and conjugate gradient), annealing, and equilibration. For the annealing step, restraints were applied to the heavy atoms of the protein via a harmonic potential of force constant 5 kcal $mol^{-1}$ $Å^{-2}$ over 500 ps until a temperature of 300 K was achieved. Following the annealing step, the system was equilibrated for 500 ps under isothermal-isobaric ensemble conditions (i.e., NPT) at a temperature and pressure of 300 K and 1 bar, respectively. The protein and ligand parameters were derived from the Amberff19SB and GAFF2 force fields, respectively. The average ligand and complex RMSD over the course of the simulations (with respect to the initial structure) for the different states were analyzed to explore changes in the simulated structure (**Figures S2** & **S3**). The results from the simulations are summarized **Table S2**.

**Table S2. Summary of Tautomeric states of His41 determined from Amber MD simulation.**

| Protein:Ligand Complex | His41 state | Average Ligand RMSD (Å) |
|---|---|---|
| $M^{pro}$:11a | HID | 2.99 ± 0.13 |
| $M^{pro}$:11a | HIE | 39.76 ± 9.86 |
| $M^{pro}$:GC373 | HID | 3.17 ± 0.24 |
| $M^{pro}$:GC373 | HIE | 3.35 ± 0.80 |
| $M^{pro}$:MI-30 | HID | 2.98 ± 0.55 |
| $M^{pro}$:MI-30 | HIE | 4.71 ± 0.76 |
| $M^{pro}$:PF-231 | HID | 3.28 ± 0.36 |
| $M^{pro}$:PF-231 | HIE | 2.48 ± 0.20 |

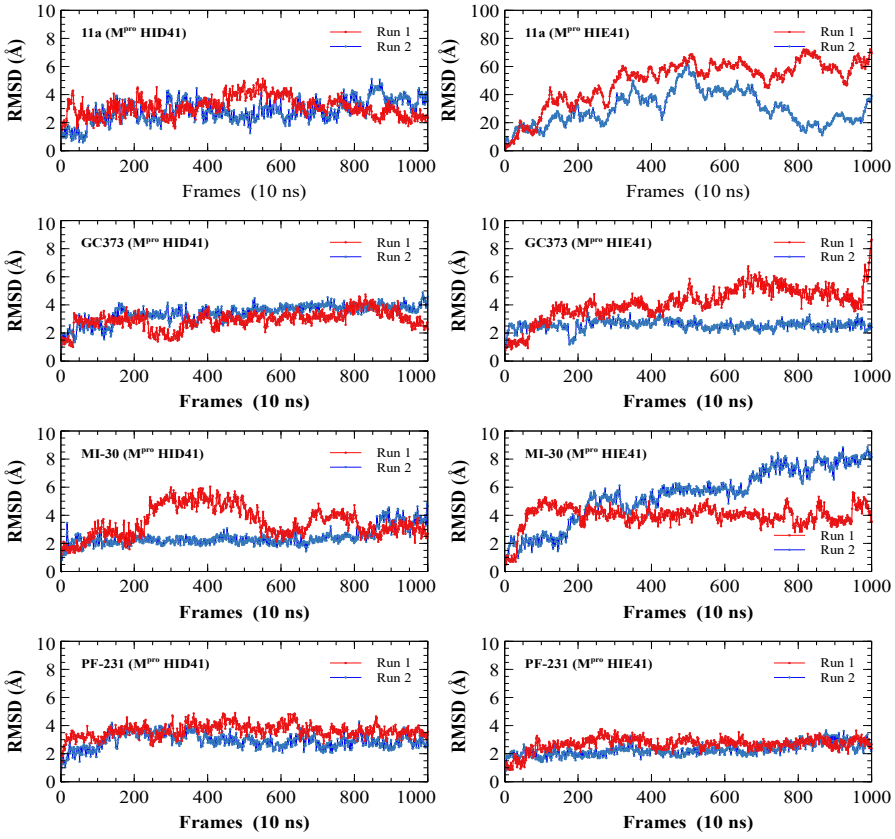


**Figure S2. Average ligand RMSD of the $M^{pro}$ inhibitor bound complexes with different His41 protonation states.**

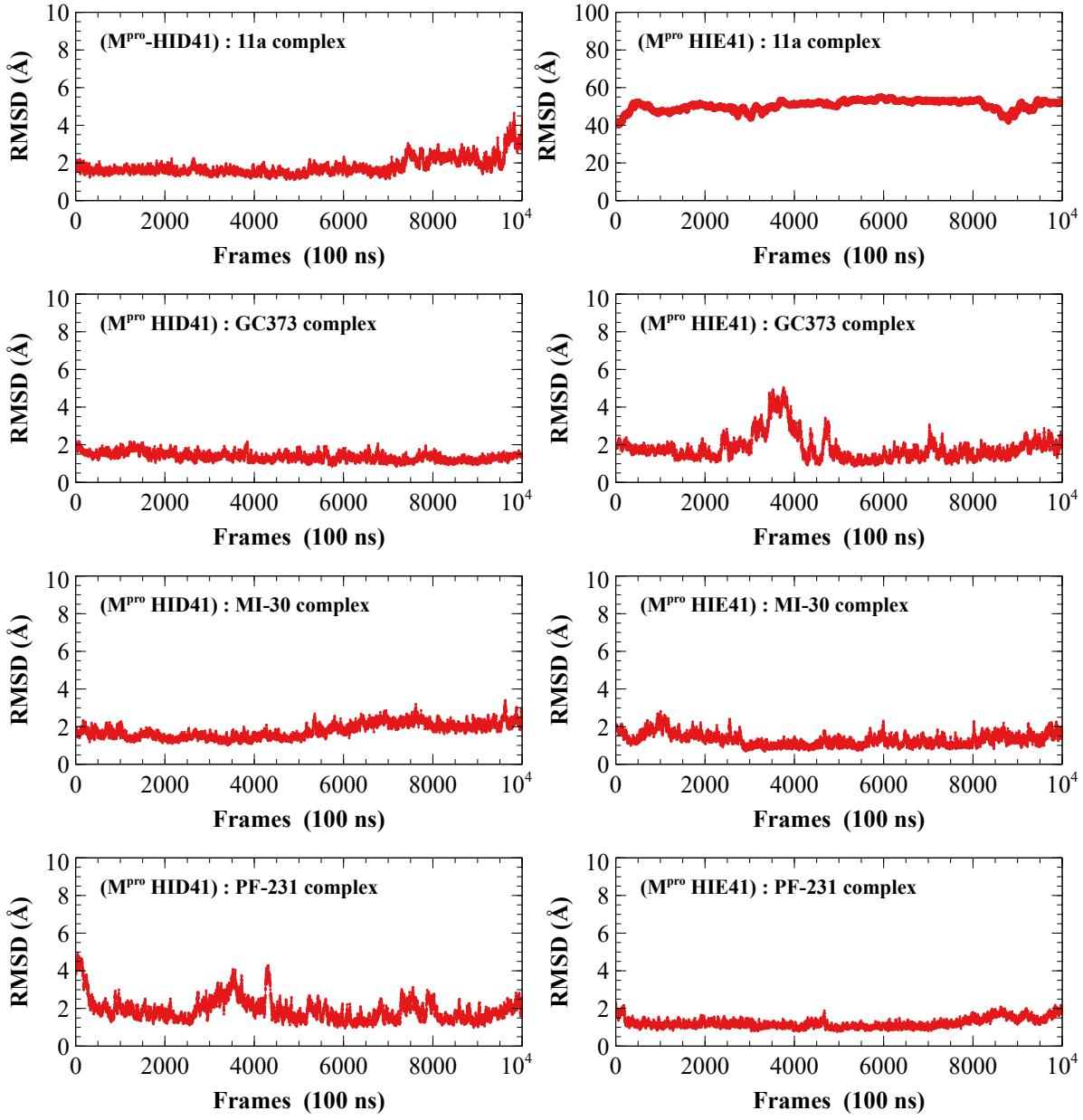


**Figure S3. Average RMSD of the $M^{pro}$ inhibitor bound complexes with different His41 protonation states over the course of 100 ns MD simulation.**

## Absolute Binding Free Energy Simulations

To estimate the noncovalent binding energy contributions of the inhibitors to the M$^{pro}$ target, absolute binding free energy calculations were performed. The protocol follows from the work of Aldeghi *et al*.[4] and we have recently employed this approach to estimate the binding free energy of peptidomimetic inhibitors (N3 and $\alpha$-ketoamide) of the SARS-CoV-2 M$^{pro}$.[5] The GROMACS MD engine[6] was used for the simulations and the AMBER99SB-ILDNP[7] and GAFF[8] force fields were used to model the protein and ligand parameters, respectively. The simulations were performed using both enhanced and non-enhanced sampling techniques via the Hamiltonian replica-exchange algorithm (H-REMD).[9] As is typical for alchemical binding free energy simulations, a set of restraints were imposed on the bound ligands to keep it in the binding site following the decoupling steps in the simulation. These restraints were defined by one distance (r), two angles ($\theta_a$, $\theta_A$), and three dihedral harmonic potentials ($\phi_{ba}$, $\phi_{aA}$, $\phi_{AB}$) following the work of Boresch et al.[10], **Figure S4**.

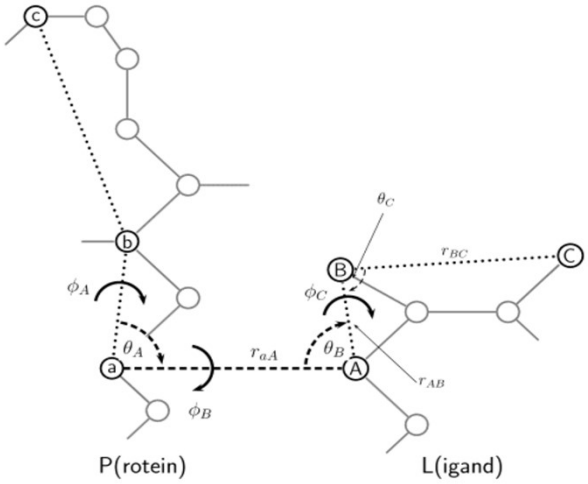


**Figure S4**. Set of restraints proposed by Boresch et al.[10] for use in binding free energy calculations. The atoms and terms involved in this set of restraints are shown. Atoms "a," "b," and "c" belong to the protein (on the left), while atoms "A," "B," and "C" belong to the ligand (on the right). There is one distance restraint ($r_{aA}$), two bond angle restraints ($\theta_A$, $\theta_B$), and three dihedral restraints ($\phi_A$, $\phi_B$, $\phi_C$). Image adapted with permission from Boresch *et al*., Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B,* **2003**, 107 (35), 9535–9551. Copyright © 2003, American Chemical Society.

The binding free energy of ligand to a target protein can be estimated by the summing up the contributions resulting from the interaction energy accompanying the binding event, **Eqn**. (**1**).

$$\Delta G^{o}_{binding} = \Delta G^{prot}_{elec + vdw + restr} + \Delta G^{solv}_{elec + vdw} + \Delta G^{solv}_{restr\_on} \qquad (1)$$

$\Delta G^{prot}_{elec + vdw + restr}$ represents the sum of the electrostatic, van der Waals, and restraint energies following the interaction of the ligand with the protein; $\Delta G^{solv}_{elec + vdw}$ represents the electrostatic and van der Waals interaction energy of the ligand in bulk solution; and $\Delta G^{solv}_{restr\_on}$ represents the restraint energy cost imposed on the bound ligand following decoupling steps from the free energy calculation. **Table S3** and **S4** provide a breakdown of the binding free energy contributions for of the inhibitors studied in this work.

**Table S3: Breakdown of Binding Free Energy Binding Results without Enhanced Sampling (i.e, no H-REMD)**

| Protien:Ligand Complex | Replicates | $\Delta G^{o}_{binding}$ [a] | $\Delta G^{solv}_{elec+vdw}$ [a] | $\Delta G^{solv}_{restr}$ [a] | $\Delta G^{prot}_{elec+vdw+restr}$ [a] |
|---|---|---|---|---|---|
| **M$^{pro}$:11a** | I | –5.33 ± 0.20 | 28.08 ± 0.12 | 7.20 | –40.61 ± 0.16 |
| | II | –5.06 ± 0.20 | 28.86 ± 0.06 | 7.20 | –41.12 ± 0.19 |
| | III | –3.85 ± 0.19 | 28.83 ± 0.05 | 7.20 | –39.88 ± 0.18 |
| **M$^{pro}$:GC373** | I | –5.17 ± 0.22 | 24.49 ± 0.12 | 7.19 | –36.85 ± 0.18 |
| | II | –3.55 ± 0.23 | 25.18 ± 0.14 | 7.19 | –35.92 ± 0.18 |
| | III | –5.69 ± 0.16 | 24.59 ± 0.05 | 7.19 | –37.47 ± 0.15 |
| **M$^{pro}$:MI-30** | I | –7.66 ± 0.18 | 24.84 ± 0.06 | 7.24 | –39.79 ± 0.17 |
| | II | –8.08 ± 0.21 | 24.78 ± 0.13 | 7.24 | –40.10 ± 0.17 |
| | III | –7.09 ± 0.13 | 24.46 ± 0.07 | 7.24 | –38.79 ± 0.11 |
| **M$^{pro}$:PF231** | I | –9.29 ± 0.17 | 30.28 ± 0.05 | 7.21 | –46.78 ± 0.16 |
| | II | –7.48 ± 0.20 | 30.92 ± 0.12 | 7.21 | –45.61 ± 0.16 |
| | III | –7.74 ± 0.20 | 30.52 ± 0.14 | 7.21 | –45.47 ± 0.14 |

**for the Inhibitor-bound M$^{pro}$ Complexes Studied in this Work.**

**Table S4: Breakdown of Binding Free Energy Binding Results with Enhanced Sampling (i.e, H-REMD) for the**

| Protien:Ligand Complex | Replicates | $\Delta G^{o}_{binding}$ [a] | $\Delta G^{solv}_{elec+vdw}$ [a] | $\Delta G^{solv}_{restr}$ [a] | $\Delta G^{prot}_{elec+vdw+restr}$ [a] |
|---|---|---|---|---|---|
| **M$^{pro}$:11a** | I | –5.90 ± 0.26 | 28.04 ± 0.22 | 7.20 | –41.14 ± 0.14 |
| | II | –4.73 ± 0.19 | 28.27 ± 0.12 | 7.20 | –40.20 ± 0.15 |
| | III | –5.59 ± 0.15 | 28.05 ± 0.11 | 7.20 | –40.48 ± 0.10 |
| **M$^{pro}$:GC373** | I | –5.01 ± 0.19 | 24.76 ± 0.15 | 7.19 | –36.98 ± 0.11 |
| | II | –5.37 ± 0.12 | 24.81 ± 0.06 | 7.19 | –37.38 ± 0.10 |
| | III | –5.04 ± 0.12 | 24.59 ± 0.07 | 7.19 | –36.83 ± 0.10 |
| **M$^{pro}$:MI-30** | I | –7.46 ± 0.21 | 24.65 ± 0.17 | 7.24 | –39.35 ± 0.12 |
| | II | –7.75 ± 0.13 | 24.48 ± 0.07 | 7.24 | –39.47 ± 0.11 |
| | III | –7.29 ± 0.14 | 24.65 ± 0.07 | 7.24 | –38.18 ± 0.12 |
| **M$^{pro}$:PF231** | I | –8.38 ± 0.15 | 30.93 ± 0.11 | 7.21 | –46.52 ± 0.10 |
| | II | –8.11 ± 0.15 | 30.88 ± 0.09 | 7.21 | –46.20 ± 0.12 |
| | III | –7.92 ± 0.36 | 31.06 ± 0.14 | 7.21 | –46.19 ± 0.33 |

**Inhibitor-bound M$^{pro}$ Complexes Studied in this Work.**

[a] All energies are in kcal mol$^{-1}$. $\Delta G^{o}_{binding}$ represents the free energy of binding of the inhibitor to the protein, $\Delta G^{solv}_{elec+vdw}$ is the interaction energy of the inhibitor in bulk solution, $\Delta G^{solv}_{restr}$ is the restraint energy term of the inhibitor in bulk solution, and $\Delta G^{prot}_{elec+vdw+restr}$ is the sum of the interaction energy and restraint energy of the inhibitor in the binding pocket of the protein.

## ONIOM(QM:MM) Calculations

The covalent binding free energy contribution of the inhibitors to the $M^{pro}$ target was estimated by performing QM/MM simulations using the ONIOM method implemented in Gaussian16. Using the crystallographic x-ray structure as a starting point for the calculations, the relative Gibbs energy difference between the final covalent complex (from the crystallographic x-ray structure) and the non-covalent complex (from the binding free energy simulations) was obtained. The M06-2X/def2-TZVP:AMBER method within an electronic embedding scheme was used for the calculations. The QM region consisted of the full inhibitor structure, Cys145, and His41 side chains. The remaining atoms were included in the low-level molecular mechanical region. The initial structures were fully geometry optimized, followed by frequency calculations to verify that the optimized structures were minima and to calculate the Gibbs energy. Several initial structures were considered to obtain an average value and standard deviation of the estimated reaction free energy for the complexes, **Table S5**. These structures were obtained from different snapshots from the absolute binding free energy simulations.

**Table S5: Summary of the Estimated Covalent Binding Free Energy Contribution of the inhibitor-bound M$^{pro}$**

| Inhibitor | No. QM atoms | Replicates | $G^{o}_{non-covalent}$ **(Hartree)** | $G^{o}_{covalent}$ (Hartree) | $\Delta G^{\circ}_{rxn}$ (Hartree) | $\Delta G^{\circ}_{rxn}$ (kcal mol$^{-1}$) |
|---|---|---|---|---|---|---|
| | 81 | I | –2176.130203 | –2176.139753 | –0.009550 | –5.99 |
| **11a** | 81 | II | –2176.130071 | –2176.139675 | –0.009604 | –6.03 |
| | 81 | III | –2176.129917 | –2176.139681 | –0.009764 | –6.13 |
| | | | | | $\Delta G^{\circ}_{rxn}$ **(avg.)** | –6.05 ± 0.07 |
| | 74 | I | –2042.459253 | –2042.472458 | –0.01321 | –8.29 |
| **GC373** | 74 | II | –2042.459244 | –2042.472386 | –0.01314 | –8.25 |
| | 74 | III | –2042.459479 | –2042.472324 | –0.01285 | –8.06 |
| | | | | | $\Delta G^{\circ}_{rxn}$ **(avg.)** | –8.20 ± 0.12 |
| | 76 | I | –3037.570135 | –3037.585835 | –0.015700 | –9.85 |
| **MI-30** | 76 | II | –3037.570830 | –3037.586681 | –0.015851 | –9.95 |
| | 76 | III | –3037.570205 | –3037.585346 | –0.015141 | –9.50 |
| | | | | | $\Delta G^{\circ}_{rxn}$ **(avg.)** | –9.77 ± 0.23 |
| | 82 | I | –2288.513251 | –2288.524078 | –0.010827 | –6.79 |
| **PF-231** | 82 | II | –2288.513598 | –2288.524007 | –0.010409 | –6.53 |
| | 82 | III | –2288.513550 | –2288.52465 | –0.011100 | –6.97 |
| | | | | | $\Delta G^{\circ}_{rxn}$ **(avg.)** | –6.76 ± 0.22 |

**complexes.**

# References


(1) Pavlova, A.; Lynch, D. L.; Daidone, I.; Zanetti-Polzi, L.; Smith, M. D.; Chipot, C.; Kneller, D. W.; Kovalevsky, A.; Coates, L.; Golosov, A. A.; et al. Inhibitor Binding Influences the Protonation States of Histidines in SARS-CoV-2 Main Protease. *Chem. Sci.* **2021**, *12* (4), 1513–1527.

(2) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes in the Reaction Mechanism. *ACS Catal.* **2021**, *11* (7), 4157–4168.

(3) 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**, *1–2*, 19–25.

(4) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, *7* (1), 207–218.

(5) Awoonor-Williams, E.; Abu-Saleh, A. A.-A. A. Covalent and Non-Covalent Binding Free Energy Calculations for Peptidomimetic Inhibitors of SARS-CoV-2 Main Protease. *Phys. Chem. Chem. Phys.* **2021**, *23* (11), 6746–6757.

(6) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.

(7) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional Timescale Predictions by Molecular Dynamics Simulations: Case Study Using Proline and Hydroxyproline Sidechain Dynamics. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (2), 195–215.

(8) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

(9) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058.

(10) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107* (35), 9535–9551.