Supporting Information for:

COVALENT AND NON-COVALENT BINDING FREE ENERGY CALCULATIONS FOR PEPTIDOMIMETIC INHIBITORS OF SARS-COV-2 MAIN PROTEASE

Ernest Awoonor-Williams* and Abd Al-Aziz A. Abu-Saleh

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

COMPUTATIONAL DETAILS

Protein–Ligand Model System Setup

The initial coordinates of the peptidomimetic inhibitors (i.e., N3, α-ketoamide) were collected from the protein data bank. The PBD identifiers of N3, and α-ketoamide complexed with SARS-CoV-2 M^{pro} used for the study were 6LU7 and 6Y2G, respectively. Molecular dynamics (MD) simulations of the protein–ligand systems were performed in explicit solvent, after which the system was equilibrated. The TIP3P water model[1] was chosen to describe the water molecules. The CHARMM36[2] all-atom protein force field was used to generate parameters for structural models of the protein and the CHARMM General Force Field (CGENFF) was used to obtain parameters for the ligand. **Figure S1** shows the chemical structure of the peptidomimetic inhibitors studied.

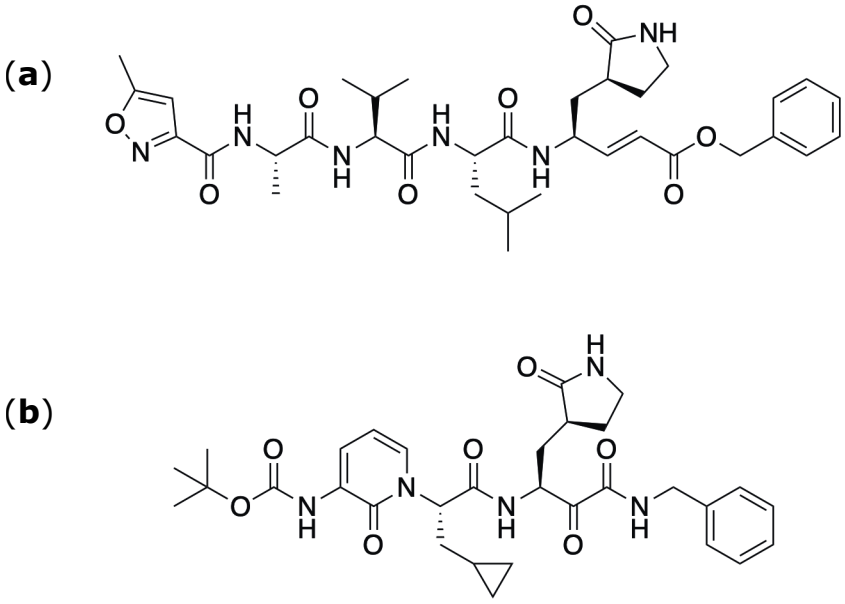

Figure S1. Chemical structure of peptidomimetic inhibitors investigated in this study. (**a**) N3, and (**b**) α-ketoamide 13b.

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

The simulation details are as follows: The initial protein–ligand complex was solvated in a periodic box, with a cutoff distance of 12 Å from the edge of the box. The simulation cell was neutralized with Na^+ and Cl^- ions and the concentration of NaCl was set to 0 M. The system was kept at a constant temperature and pressure of 298.15 K and 1 atm, respectively, by using the Langevin dynamics and Langevin piston method. The temperature and pressure conditions were chosen to match experiment. The SHAKE algorithm was used to constrain covalent bonds involving hydrogen, and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) method.[3,4] A smoothing function was applied from 10 to 12 Å to smoothly truncate van der Waals forces at the cutoff distance. A grid spacing of 1.0 Å was used for all simulation cells. The time step for all the simulations is 2 fs. The model system was initially energy minimized for 1000 steps to eliminate any steric clashes or structural irregularities that may exist within the protein–ligand molecular assembly. The system was then equilibrated for 50 ns at constant pressure and temperature conditions (NpT ensemble) of 1 atm and 300 K, respectively. The equilibration was performed first with restraints on the α-carbon of the protein backbone, and then without restraints on the system. The coordinates of the equilibrated protein–ligand complex was used as a starting structure for the constant-pH molecular dynamics simulations. The parameters used in the configuration file for the equilibration run is provided below and were performed using NAMD 2.13.[5]

Parameters for Protein–Ligand Molecular Dynamics Equilibration Simulation

```
#############################################################
## JOB DESCRIPTION                         ##
#############################################################
structure        complex_ionized.psf
coordinates      complex_ionized.pdb
set temperature   298.15
set outputname    complex_eq
firsttimestep     0

#############################################################
## SIMULATION PARAMETERS                   ##
#############################################################
# Input
paraTypeCharmm      on
parameters toppar/toppar_water_ions.str
parameters toppar/prd.prm
parameters toppar/par_all36_cgenff.prm
parameters toppar/par_all36m_prot.prm
parameters toppar/par_all36_na.prm
parameters toppar/par_all36_carb.prm
temperature       $temperature

exclude          scaled1-4
1-4scaling        1.0
cutoff           12.0
switching         on
switchdist        10.0
pairlistdist      14.0

timestep          2.0 ;# 2fs/step
rigidBonds        all ;# needed for 2fs steps
nonbondedFreq     1
fullElectFrequency  2
stepspercycle     10
```

```
# Constant Temperature Control
langevin          on    ;# do langevin dynamics
langevinDamping    1     ;# damping coefficient (gamma) of 1/ps
langevinTemp       $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens

wrapAll           on
PME               yes
PMEGridSizeX      76
PMEGridSizeY      92
PMEGridSizeZ      120
margin 3

cellbasisvector1 75.335  0.0 0.0
cellbasisvector2 0.0 90.697 0.0
cellbasisvector3 0.0 0.0 116.126
cellOrigin -25.851 12.764 58.031
useGroupPressure     yes ;# needed for rigidBonds
useFlexibleCell      no
useConstantArea      no

langevinPiston       on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100.0
langevinPistonDecay   50.0
langevinPistonTemp    $temperature

outputName         $outputname
restartfreq        10000    ;# 500steps = every 1ps
dcdfreq            10000
xstFreq            10000
outputEnergies     10000
outputPressure     10000

constraints on
conskcol O
conskfile restrained.pdb
consref restrained.pdb

minimize 1000
reinitvels         $temperature
run 25000000
```

Constant-pH Molecular Dynamics Simulation

Nonequilibrium molecular dynamics/Monte Carlo constant-pH simulations[6] were performed to calculate the pK_a of key amino acid residues within the catalytic domain of the SARS-CoV-2 M^{pro}. The main advantage of constant-pH molecular dynamics over standard pK_a calculation tools is its ability to account for pH-induced conformational changes and multiple protonation state changes simultaneously within the model system. The simulations were performed in explicit solvent for the apo form (i.e., no inhibitor bound, PDB ID: 6M03) and holo forms (i.e., inhibitor bound, PDB IDs: 6LU7 and 6Y2G) of the coronavirus M^{pro} model structures. The CHARMM36 protein force field and CGENFF were used to generate model parameters for the protein and ligand structures, respectively. The simulations were implemented using NAMD 2.13 program. Details of the simulation protocol can be found in the Theory and Methods section in the main text of the manuscript and in our recently published paper on druggable cysteine pK_a's in protein kinases.[7] The parameters used for our constant-pH MD simulation can be found below.

Parameters for Constant-pH Molecular Dynamics Simulation

```
# These keywords all follow as usual
#
set temperature 298.15

set topo_dir "../topology"
structure $topo_dir/complex_ionized.psf
coordinates $topo_dir/complex_ionized.pdb
binCoordinates $topo_dir/complex_eq-2.restart.coor
binVelocities  $topo_dir/ complex _eq-2.restart.vel
extendedSystem $topo_dir/ complex_eq-2.restart.xsc

wrapWater on
wrapAll on
wrapNearest on
outputEnergies 5000
DCDFreq 5000

timestep 2.0
fullElectFrequency 2
rigidBonds ALL

langevin on
langevinTemp $temperature
langevinDamping 1.0
langevinHydrogen no

switching on
VDWForceSwitching on
LJCorrection on
switchDist 12.0
cutoff 14.0
pairlistDist 16.0
exclude scaled1-4
```

```
1-4scaling 1.0
PME on
PMEGridSpacing 1.0

# Begin constant-pH MD keywords and modifications
#
# Load the constant-pH Tcl files
source ../namdcph/namdcph.tcl
# Load force field files as usual, but add constant-pH specific parameters
set toppar_dir "../toppar"
paratypecharmm on
parameters $toppar_dir/k36.str
parameters $toppar_dir/par_all36_cgenff.prm
parameters $toppar_dir/par_all36_prot.prm
parameters $toppar_dir/par_all36_na.prm
parameters $toppar_dir/par_all36_carb.prm
parameters $toppar_dir/par_cph36_prot.prm
parameters $toppar_dir/par_all36_solvent.prm
# Load constant-pH specific topology files
cphConfigFile $toppar_dir/conf_cph36_prot.json
topology $toppar_dir/top_all36_prot.rtf
topology $toppar_dir/top_cph36_prot.rtf
topology $toppar_dir/top_solvent.rtf

# We will be running multiple pH values sorted into their own directories, but
# otherwise using the same naming scheme.

source pH.tcl

#set pHList [list 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0 11.5]
#set pH [lindex $pHList [myReplica]]
pH $pH

outputname prot_prod0
stdout prot_prod0.log
cphMDBasename namdcph.md
cphSwitchBasename namdcph.sw

# With the current settings this implies 10 ps between switching attempts,
# which will be 15 ps in length. These settings should be relatively close to
# optimal.
#
cphNumMinSteps      200
cphNumstepsPerSwitch 7500
cphRun 5000 2500
# Don't exit until all simulations have finished.
#replicaBarrier
exit
```

Absolute Binding Free Energy Protocol

The absolute binding free energy of the N3 and α-ketoamide peptidomimetic inhibitors were calculated using the rigorous alchemical free energy protocol reported by Aldeghi *et al.*[8] using the GROMACS molecular dynamics software.[9] The AMBER99SB-ILDNP[10] and GAFF[11] force fields were used to model the protein and ligand parameters, respectively. Two sets of simulations were performed in order to determine the absolute binding free energy of the inhibitors. In these simulations, first the ligand was decoupled from bulk solution (i.e., unbound state) and then from the complex (i.e., bound state). The relative position and orientation of the bound ligand in the complex was described using six internal coordinates: one distance (r), two angles (θ_a, θ_A), and three dihedrals (ϕ_{ba}, ϕ_{aA}, ϕ_{AB}), **Figure S2**. Details of the simulation protocol can be found in the Theory and Methods section in the main text of the manuscript. The input files pertaining to the calculations are provided as a compressed archived file.

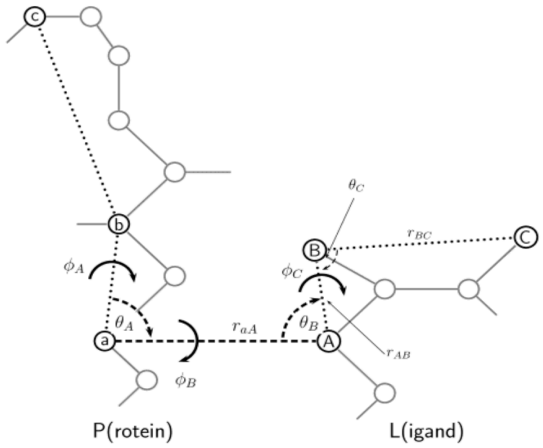

Figure S2. Set of restraints proposed by Boresch et al.[12] for use in binding free energy calculations. The atoms and terms involved in this set of restraints are shown. Atoms "a," "b," and "c" belong to the protein (on the left), while atoms "A," "B," and "C" belong to the ligand (on the right). There is one distance restraint (r_{aA}), two bond angle restraints (θ_A, θ_B), and three dihedral restraints (ϕ_A, ϕ_B, ϕ_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.

Table S1: Tautomeric State Assignment for Key Histidine Residues in the N3 and α-ketoamide M^{pro} Complexes.

Residue	M^{pro}: N3 complex	M^{pro}: α-ketoamide Complex
Histidine-41 [a]	HID	HIE
Histidine-163	HIE	HID
Histidine-164 [a]	HIE	HID
Histidine-172	HIE	HIE

[a] Histidine tautomeric states were chosen based on molecular dynamics simulation results reported by Gumbart and coworkers[13] on the structural stability of the M^{pro} as a function of protonation state assignments. The histidine states listed follow the Amber force field naming system.

Table S2: Breakdown of Free Energy Binding Results for Peptidomimetic Inhibitors of SARS-CoV-2 M^{pro}

Ligand	Replicates	$\Delta G^{o}_{binding}$ [a]	$\Delta G^{solv}_{elec+vdw}$ [a]	ΔG^{solv}_{restr} [a]	$\Delta G^{prot}_{elec+vdw+restr}$ [a]
N3	I	-6.33 ± 0.23	36.34 ± 0.09	6.98	-49.65 ± 0.21
	II	-6.46 ± 0.27	36.13 ± 0.08	6.98	-49.57 ± 0.26
	III	-7.60 ± 0.20	37.00 ± 0.07	6.98	-51.58 ± 0.19
α-ketoamide	I	-2.70 ± 0.25	34.76 ± 0.06	7.16	-44.62 ± 0.24
	II	-2.61 ± 0.14	34.88 ± 0.06	7.16	-44.65 ± 0.13
	III	-3.08 ± 0.23	34.92 ± 0.06	7.16	-45.16 ± 0.22

[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, ΔG^{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.

The production run of the binding free energy calculation of the inhibitors to M^{pro} was calculated using Hamiltonian replica-exchange (HREX) molecular dynamics method. There were 31 replica windows for decoupling the ligand in bulk solution and 42 λ windows for the complex simulation. Each replica window was run for 12 ns, with the first 2ns discarded as equilibration during analysis of the free energy results. All simulations were performed in triplicate. The final binding free energy values are the averages and standard deviation of the three independent simulations. **Table S2** summarizes the results from the binding free energy calculations of N3 and α-ketoamide to the M^{pro} target.

The binding free energy calculations were also performed using the CHARMM36 protein force field[2] and the CHARMM General Force Field (CGENFF),[14] for the protein and ligand parameters, respectively. The simulation time and conditions were the same as described above. The binding free energy results from this approach ($\Delta G^{o}_{binding}$, N3 = –4.9 ± 0.3 kcal mol^{-1}; $\Delta G^{o}_{binding}$, α-ketoamide = –1.2 ± 0.2 kcal mol^{-1}) moderately agrees with the results from our Amber simulations above—both different force fields and parameters predict the same trend for the ligand binding energies to the M^{pro}. This provides an independent check on our binding free energy results and makes us confident in the results.

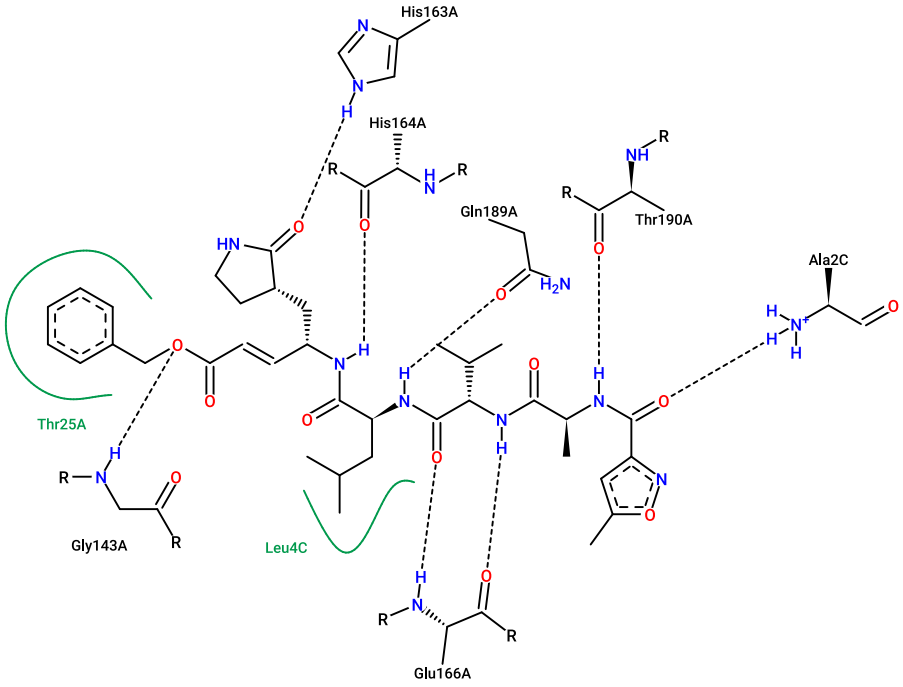

Figure S3. 2D ligand interaction diagram of N3 in complex with SARS-CoV-2 M^{pro} (PDB code: 6LU7). Hydrogen bonds are depicted as black dashed lines and hydrophobic contacts are represented as green spline segments around the participating ligand functional group. Figure was generated using the PoseView software.[15]

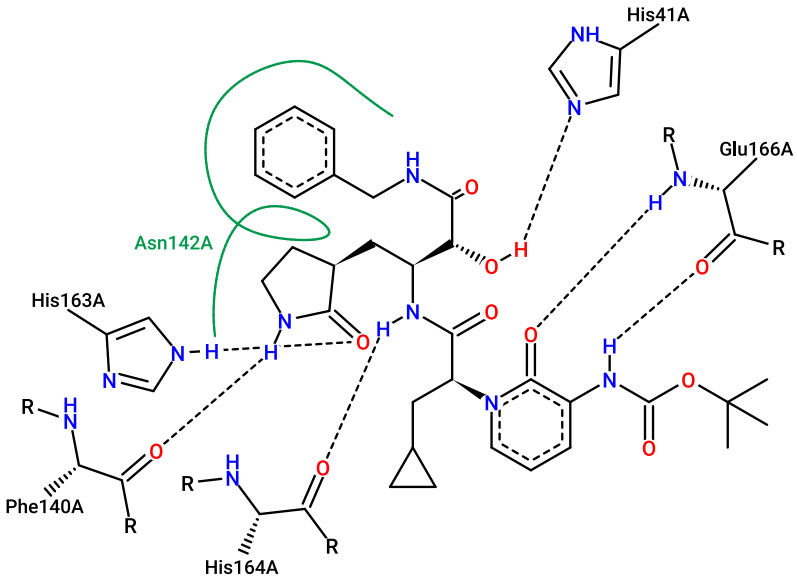

Figure S4. 2D ligand interaction diagram of α-ketoamide in complex with SARS-CoV-2 M^{pro} (PDB code: 6Y2G). Hydrogen bonds are depicted as black dashed lines and hydrophobic contacts are represented as green spline segments around the participating ligand functional group. Figure was generated using the PoseView software.[15]

Sample NAMD Configuration file section for Umbrella Sampling/Replica-Exchange MD Simulation

```
Structure complex_ionized.psf
Coordinates          complex_ionized.pdb

paraTypeCharmm     on
parameters       toppar/toppar_water_ions.str
parameters toppar/n3.prm
parameters toppar/par_all36_cgenff.prm
parameters toppar/par_all36m_prot.prm
parameters toppar/par_all36_na.prm
parameters toppar/par_all36_carb.prm
margin                10.0

# Force-Field Parameters
exclude          scaled1-4
1-4scaling       1.0
cutoff          12.0
switching        on
switchdist       10.0
pairlistdist     14.0

# Integrator Parameters
timestep         2  ;# 2fs/step
rigidBonds       all  ;# needed for 2fs steps
nonbondedFreq     1
fullElectFrequency  2
```

```
stepspercycle      10

# Periodic Boundary Conditions
wrapAll          on
PME              yes
PMEGridSpacing     1.0

colvars       on
colvarsConfig   colvars.tcl

langevin on
langevinDamping 10.0
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens
langevinTemp $temperature

langevinPiston       on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100.0
langevinPistonDecay   50.0
langevinPistonTemp    $temperature
```

Colvars Module Configuration for Umbrella Sampling/Replica-Exchange MD

```
colvarsTrajFrequency 1000

colvar {
  name p1_l1_dist
  distance {
            group1 {
       atomsFile  protein_sel.pdb
       atomsCol O
       atomsColValue 1.0
            }
            group2 { #lig1
              atomsFile ligand.pdb
              atomsCol O
              atomsColValue 1.0
            }
  }
}

colvar {
  name p2_p1_l1_angle
  angle {
            group1 {
              psfSegID PRO
              atomNameResidueRange CA 166-166
            }
            group2 {
```

```
        atomsFile  protein_sel.pdb
        atomsCol O
        atomsColValue 1.0
            }

            group3 { #lig1
              atomsFile ligand.pdb
              atomsCol O
              atomsColValue 1.0
            }
  }
}

colvar  {
  name p1_l1_l2_angle
  angle {
            group1 {
        atomsFile  protein_sel.pdb
        atomsCol O
        atomsColValue 1.0
            }

            group2 { #lig1
              atomsFile ligand.pdb
              atomsCol O
              atomsColValue 1.0
            }

            group3 {
              psfSegID LIG
              atomNameResidueRange C15 1-1
            }
  }
}

colvar {
  name p3_p2_p1_l1_dihedral
  dihedral {
            group1 {
              psfSegID PRO
              atomNameResidueRange CA 192-192
            }
            group2 {
              psfSegID PRO
              atomNameResidueRange CA 166-166
            }

            group3 {
        atomsFile  protein_sel.pdb
```

```
        atomsCol O
        atomsColValue 1.0
            }

            group4 { #lig1
              atomsFile ligand.pdb
              atomsCol O
              atomsColValue 1.0
            }
  }
}

colvar {
  name p2_p1_l1_l2_dihedral
  dihedral {
            group1 {
              psfSegID PRO
              atomNameResidueRange CA 166-166
            }
            group2 {
        atomsFile  protein_sel.pdb
        atomsCol O
        atomsColValue 1.0
            }
            group3 { #lig1
              atomsFile ligand.pdb
              atomsCol O
              atomsColValue 1.0
            }
            group4 {
              psfSegID LIG
              atomNameResidueRange C15 1-1
            }
  }
}

colvar {
  name p1_l1_l2_l3_dihedral
  dihedral {
            group1 {
        atomsFile  protein_sel.pdb
        atomsCol O
        atomsColValue 1.0
            }

            group2 { #lig1
              atomsFile ligand.pdb
              atomsCol O
              atomsColValue 1.0
```

```
        }

        group3 {
          psfSegID LIG
          atomNameResidueRange C15 1-1
        }

        group4 {
          psfSegID LIG
          atomNameResidueRange N2 1-1
        }
  }
}

colvar {
  name RMSD
  rmsd {
        atoms {
          atomsfile avg_ligand_noh_all.pdb
          atomsCol O
          atomsColValue 1.00
        }
        refPositionsFile avg_ligand_noh_all.pdb
        refPositionsCol O
        refPositionsColValue 1.00
  }
}

harmonic {
 name rmsdpot
 colvars RMSD
 centers 0.0
 forceConstant 50.0
}
```

Table S3: The Gibbs Energy of the Reaction for the Formation of the M^{pro}-N3 Covalent Adduct.

Complex	Energy Component [b]	Energy (Hartree)
Reactant	Electronic + Gibbs Correction	–1754.036004
Product	Electronic + Gibbs Correction	–1754.053258
ΔG°_{rxn} (Product − Reactant)	Electronic + Gibbs Correction	–0.017254

[b] The ONIOM(M06-2X/def2-TZVP:AMBER) method within an electrostatic embedding formalism was used for the calculations.

Table S4: The Gibbs Energy of the Reaction for the Formation of the M^{pro}– α-ketoamide Covalent Adduct.

Complex	Energy Component [b]	Energy (Hartree)
Reactant	Electronic + Gibbs Correction	–1770.338849
Product	Electronic + Gibbs Correction	–1770.349546
ΔG°_{rxn} (Product − Reactant)	Electronic + Gibbs Correction	–0.010697

[b] The ONIOM(M06-2X/def2-TZVP:AMBER) method within an electrostatic embedding formalism was used for the calculations.

Table S5: Effect of QM Region Size on the Covalent Binding Free Energy

Complex	QM Region	$\Delta G_{covalent}$ (kcal mol^{-1})
M^{pro} –N3	Chosen QM Region (see **Fig. 3** in main text)	–10.83
	Full ligand + C145 side chain	–10.14
M^{pro} – α-ketoamide	Chosen QM Region (see **Fig. 3** in main text)	–6.71
	Full ligand + C145 side chain	–6.86

REFERENCES

(1) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.

(2) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257–3273.

(3) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089.

(4) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

(5) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

(6) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Constant-PH Molecular Dynamics Simulations for Large Biomolecular Systems. *J. Chem. Theory Comput.* **2017**, *13* (12), 5933–5944.

(7) Awoonor-Williams, E.; Rowley, C. N. How Reactive Are Druggable Cysteines in Protein Kinases? *J. Chem. Inf. Model.* **2018**, *58* (9), 1935–1946.

(8) 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.

(9) 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.

(10) 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.

(11) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. **2004**, *56531*.

(12) 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.

(13) 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**, 1–38.

(14) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671–690.

(15) Stierand, K.; Rarey, M. Drawing the PDB: Protein-Ligand Complexes in Two Dimensions. *ACS Med. Chem. Lett.* **2010**, *1* (9), 540–545.