Supporting Information:
Simulating Protein–Ligand Binding with Neural Network Potentials

Shae-Lynn J. Lahey and Christopher N. Rowley

Department of Chemistry, Memorial University of Newfoundland, St. John’s, NL, Canada
E-mail: crowley@mun.ca

Table S1: Table of crystallographic data. Ligands that did not have a common drug name are listed by their Drugbank ID or their compound ID number (CID).

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Protein</th>
<th>EC</th>
<th>Ligand</th>
<th>Resolution (Å)</th>
<th>Temp (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1XOZ</td>
<td>phosphodi-esterase 5A</td>
<td>3.1.4.17</td>
<td>tadalafil</td>
<td>1.37</td>
<td>93</td>
</tr>
<tr>
<td>3EYG</td>
<td>JAK1</td>
<td>2.7.10.2</td>
<td>tofacitinib</td>
<td>1.9</td>
<td>100</td>
</tr>
<tr>
<td>2W6N</td>
<td>biotin carboxylase</td>
<td>6.3.4.14</td>
<td>DB08315</td>
<td>1.87</td>
<td>100</td>
</tr>
<tr>
<td>4HJO</td>
<td>EGFR</td>
<td>2.7.10.1</td>
<td>erlotinib</td>
<td>2.21</td>
<td>110</td>
</tr>
<tr>
<td>4NCT</td>
<td>Human DYRK1A</td>
<td>2.7.12.1</td>
<td>midostaurin</td>
<td>2.6</td>
<td>100</td>
</tr>
<tr>
<td>2HYY</td>
<td>Abl kinase</td>
<td>2.7.10.2</td>
<td>imatinib</td>
<td>2.4</td>
<td>100</td>
</tr>
<tr>
<td>3EIG</td>
<td>dihydrofolate reductase</td>
<td>1.5.1.3</td>
<td>methotrexate</td>
<td>1.7</td>
<td>93</td>
</tr>
<tr>
<td>3ETA</td>
<td>IGF-1R</td>
<td>2.7.10.1</td>
<td>CID 45272927</td>
<td>2.6</td>
<td>93</td>
</tr>
</tbody>
</table>
Table S2: Table of RMSD of the calculated structures of the ligands relative to the PDB structure.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>CGenFF</th>
<th>NNP/MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1XOZ</td>
<td>0.19</td>
<td>0.13</td>
</tr>
<tr>
<td>3EYG</td>
<td>0.24</td>
<td>0.50</td>
</tr>
<tr>
<td>2W6N</td>
<td>0.54</td>
<td>0.57</td>
</tr>
<tr>
<td>4HJO</td>
<td>0.79</td>
<td>0.59</td>
</tr>
<tr>
<td>4NCT</td>
<td>1.1</td>
<td>0.60</td>
</tr>
<tr>
<td>2HYY</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>3EIG</td>
<td>0.37</td>
<td>0.28</td>
</tr>
<tr>
<td>3ETA</td>
<td>0.26</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Figure S1: Trajectory of the RMSD of the calculated structure of 4HJO vs the PDB Structure vs time.
Figure S2: Calculated poses of ligands. CGenFF is in green and ANI is in red. The crystallographic electron density of the ligands are shown in blue. The PDB ID, protein name, and ligand name are included beneath the image.
Figure S3: The potential of mean force for the deviation of the structure of a ligand from its bound conformation when it is bound to its protein target.

Figure S4: The representative structures of the ionic ligands from the lowest energy of the PMF for the NNP/MM simulations of the ligands in explicit water (the solvent is not shown for clarity). The charged functional groups of the ligands form spurious intramolecular contacts.
NNP Technical Details

The ANI-1ccX model was used in all instances where NNP calculations. The TorchANI implementation was used.\textsuperscript{59} No additional intramolecular force terms for the ligand were included. In this NNP, there are 16 radial elements and the radial cutoff is 5.2 Å. The complete details are described in the Supplementary Materials of Smith et al.\textsuperscript{7}

The parameters of the NNP used here are defined in the TorchANI parameter file.\textsuperscript{510}

Figure S5: Parameters used by the ANI-1ccX NNP (rHCNO-5.2R_16-3.5A_a4-8.params)

\[\begin{align*}
TM &= 1 \\
Rcr &= 5.20000e+00 \\
Rca &= 3.50000e+00 \\
EtaR &= [1.6000000e+01] \\
ShfR &= [9.0000000e-01, 1.1687500e+00, 1.4375000e+00, 1.7062500e+00, \\
&1.9750000e+00, 2.2437500e+00, 2.5125000e+00, 2.7812500e+00, \\
&3.0500000e+00, 3.3187500e+00, 3.5875000e+00, 3.8562500e+00, \\
&4.1250000e+00, 4.3937500e+00, 4.6625000e+00, 4.9312500e+00] \\
Zeta &= [3.2000000e+01] \\
ShfZ &= [1.9634954e-01, 5.8904862e-01, 9.8174770e-01, 1.3744468e+00, \\
&1.7671459e+00, 2.1598449e+00, 2.5525440e+00, 2.9452431e+00] \\
EtaA &= [8.0000000e+00] \\
ShfA &= [9.0000000e-01, 1.5500000e+00, 2.2000000e+00, 2.8500000e+00] \\
Atyp &= [H,C,N,O]
\end{align*}\]

NAMD Input File Section for NNP/MM Simulation

\begin{verbatim}
qmforces on
qmParamPDB qmmm.pdb
qmSoftware custom
qmexecpath client.py
qmBaseDir /dev/shm/
QMColumn occ
qmChargeMode none
qmElecEmbed off
\end{verbatim}
Sample ORCA RI-MP2 Input File

! RI-MP2 RIJCOSX def2-TZVP def2-TZVP/C def2/J TIGHTSCF Opt PAL8
% maxcore 15000

* xyzfile 0 1 pes.xyz
%geom
Constraints
{D 2 3 4 5 0.0 C }
{D 3 4 5 6 0.0 C }
end
end

Sample ORCA DLPNO-CCSD(T) Input File

! DLPNO-CCSD(T) def2-TZVP def2-TZVP/C TIGHTSCF
% pal nprocs 8 end
% maxcore 15000

* xyzfile 0 1 pes.xyz
end
end

References


(S3) Mochalkin, I.; Miller, J. R.; Narasimhan, L.; Thanabal, V.; Erdman, P.; Cox, P. B.; Prasad, J. V. N. V.; Lightle, S.; Huband, M. D.; Stover, C. K. Discovery of An-


(S10) Roitberg Group, ANI-1ccX Parameters. 2019; https://github.com/isayev/ASE_ANI/blob/master/ani_models/ani-1ccx_8x/rHCNO-5.2R_16-3.5A_a4-8.params.