Assessment Report

Yuchen Yuan

July 11 2025

Why this target

The theophylline-binding RNA aptamer (PDB: 1O15) was chosen because it has high ligand specificity, high-quality structure, and small size for simulation. With just 33 nucleotides, this aptamer can be simulated quickly with classical MD, yet still captures all of the key hydrogen-bonding and stacking interactions around the theophylline (TEP) binding pocket. The NMR structure comes at high resolution, reducing uncertainty in initial setup and facilitating a more realistic simulation. In addition, because 1O15 has been the subject of many computational studies, a large amount of published data are available to compare against, both for structural dynamics and free energy benchmarks.

Functionally, the aptamer has high selectivity. TEP has 10,000x larger binding affinity than caffeine, even though they differ by only one methyl group. The dissociation constant of TEP is around 0.4 μM under physiological conditions. This characteristic makes it a good example for testing binding free energy methods, such as MM/PBSA or alchemical approaches (FEP/TI).

The aptamer has real-world applications in engineered riboswitches, RNA-based drug delivery systems, and molecular biosensors. Therefore, learning to model its binding energetics can directly inform design strategies for RNA-targeted therapeutics.

Key setup choices

System setup: 1) Look locally before fetching from RCSB to avoid repeated downloads; 2) A ligand residue name is required as an input to avoid ambiguity; 3) Use PDBFixer to add missing atoms and hydrogens at a given pH.

Force field parameterization: Amber ff14SB + OL3 for protein/RNA; GAFF2 via AmberTools for small-molecule ligands

Solvation & Neutralization: TIP3P water with user-defined padding (nm) and ionic strength (M), plus automatic neutralization.

Simulation protocol: 1) Two-stage NVT/NPT equilibration runs at 1 fs: heavy-atom restraints on RNA + ligand (harmonic, kJ/mol·nm²) and then unrestrained; 2) Production run at 2 fs in NPT ensemble; 3) Save dcd file, simulation log file, and OpenMM checkpoint file

MMPBSA setup: 1) Use ParmEd or cpptraj to strip solvent/ions and split prmtop/inpcrd into complex, receptor, and ligand. 2) Convert DCD to NetCDF trajectory for better I/O with MMPBSA.py.

Caveats

System setup: 1) This pipeline assumes that a ligand SDF file exists in the RCSB database, while in real cases, ligand files may be stored in internal databases. Further development should consider these cases. 2) PDBFixer doesn't build missing loops. In the current pipeline, any missing residues will remain as a "gap" in both sequence and 3D structure. A more robust way is using other tools (such as MODELLER, MOE, or Schrodinger software suite) to model missing residues, especially when they are in critical regions (e.g. binding or functional site)

Simulation protocol: A more careful choice is setting up a multi-step equilibration, where the restraints are gradually released

Failure modes

Force field parameterization: mismatch or errors in force field name, such as the Amber RNA foce field (OL3) and the version of GAFF.

MD runs: Make sure the system is fully minimized and equilibrated.

MMPBSA calculations: When the mask selection is not correctly specified, MMPBSA.py cannot correctly identify different components of the complex system and thus the calculation fails.

ANI-2x hybrid scheme: Production run with 2 fs timestep is unstable. Cause of failure was not identified due to the time constraint of the assessment.

Interpretation of binding number vs. PDB co-crystal pose

The binding free energy from MMPBSA calculation is -13.7005 kcal/mol. The standard error of mean is 0.3464 kcal/mol. This translates to a dissociation constant of 0.103 nM, which is significantly stronger than the literature result (~ 0.4 uM). From the PDB co-crystal pose, the ligand is in a stable configuration. It's likely that the very short simulation can only sample the highly stable conformations around the binding pose and therefore the binding free energy is remarkably overestimated. The other possible reason is the intrinsic inaccuracy of MMPBSA method. Alchemical approaches such as absolute free energy calculations using FEP should be able to provide a more reliable estimate.