

Online Resource 2 — Reproducibility and configuration summary

Manuscript: *In silico* prioritisation of morroniside as a candidate for Keap1 ETGE-pocket engagement: a molecular docking and molecular-dynamics study of natural iridoid glycosides and coptisine (In Silico Pharmacology) Author: Byungwoong Yoo (ORCID 0009-0002-1797-3100)

This document records the docking and molecular-dynamics configuration used in the study so that the workflow can be reproduced.

1. Software and environment

The molecular-dynamics environment was built from conda-forge into a Python 3.10 environment; AutoDock Vina was used as the official v1.2.5 static Linux binary and Open Babel through the system package. Package versions below were confirmed with `conda list` and reflect the conda-forge environment produced by the command below. The environment was created with:

```
mamba create -n md -c conda-forge python=3.10 openmm openmmforcefields \
    openff-toolkit mdtraj pdbfixer rdkit
mamba install -n md -c conda-forge ambertools
mamba install -n md -c nvidia -c conda-forge cuda-nvcc cuda-version=12.6
```

Component	Version	Source
Python	3.10.20	conda-forge
AutoDock Vina	1.2.5 (static x86_64 binary)	ccsb-scripps/AutoDock-Vina release
Open Babel	3.1.1	Ubuntu system package (apt)
OpenMM	8.5.1	conda-forge
openmmforcefields	0.14.2	conda-forge
openff-toolkit	0.16.10	conda-forge
AmberTools (antechamber/sqm, for AM1-BCC)	24.8	conda-forge
PDBFixer	1.12	conda-forge
MDTraj	1.10.3	conda-forge
RDKit	2025.09.5	conda-forge
CUDA toolkit (cuda-nvcc)	12.6	nvidia/conda-forge

GPU acceleration used the OpenMM CUDA platform.

2. Receptor

- Structure: PDB **1X2R**, **chain A** (mouse Keap1 Kelch/DGR domain, residues 321–609)

- Preparation: chain A extracted; PDBFixer `findMissingResidues` → `findMissingAtoms` → `addMissingAtoms` → `addMissingHydrogens(pH 7.4)` ; missing loops not modelled
- Docking receptor: `obabel ETGE_1X2R.pdb -O ETGE_1X2R.pdbqt -p 7.4 -xr` (2226 atoms)
- Residue numbering: processed-structure positions 92 / 160 / 185 correspond to full-length Keap1 **Arg415** / **Arg483** / **Ser508** (uniform +323 offset); these are the ETGE-Glu79 P1-subpocket anchors monitored in the analysis

3. Ligands (PubChem CIDs)

Catalpol 91520; Geniposide 107848; Morroniside 11228693; Loganin 87691; Coptisine 72322 (monovalent cation, formal charge +1).

- Source coordinates: PubChem 3D SDF (2D + `--gen3d` fallback if 3D unavailable)
- Charges (four iridoids): Gasteiger, assigned with `obabel <in>.sdf -O <name>.pdbqt -p 7.4 --partialcharge gasteiger`
- Charges (coptisine, AM1-BCC sensitivity run): AM1-BCC via openff-toolkit `Molecule.assign_partial_charges("am1bcc")` , computed with AmberTools antechamber/sqm; formal charge +1 preserved

4. Docking (AutoDock Vina 1.2.5)

- Search box: cubic, **22 Å** per side
- Box centre (x, y, z): **13.442, 66.553, -6.668 Å**
- Box centre definition: centroid of the side-chain heavy atoms of Arg415, Ser508 and Tyr334 (17 atoms) in the 1X2R coordinate frame
- Exhaustiveness: 16; num_modes: 9
- Seeds: three independent runs, **101 / 202 / 303**
- Pose selection: the top-ranked pose (mode 1, lowest predicted affinity) from the **seed-101** run was used to seed MD
- Per-seed best scores for the two MD-seeded re-docks (kcal/mol):
- Morroniside: 101 = -8.256 (MD pose), 202 = -8.282, 303 = -8.268
- Loganin: 101 = -7.823 (MD pose), 202 = -8.422, 303 = -8.143
- MD-seeding (top-ranked) pose scores for the original three (kcal/mol; Gasteiger, identical box and protocol): Catalpol -11.05; Geniposide -7.26; Coptisine -10.32
- These five values are the docking scores reported in Table 1 of the manuscript.

5. Molecular dynamics (OpenMM)

- Force fields: Amber **ff14SB** protein parameters (`amber14-all.xml`), **GAFF2** ligand parameters via `openmmforcefields.GAFFTemplateGenerator` with an `openff-toolkit Molecule` , and **TIP3P-FB** water (`amber14/tip3pfb.xml` ; added with `Modeller.addSolvent(model="tip3p")`)
- Solvation: `Modeller.addSolvent(model="tip3p", padding=1.0*nanometer, ionicStrength=0.15*molar, neutralize=True)`
- Energy minimisation: `minimizeEnergy(maxIterations=5000)` for catalpol, geniposide and coptisine (`run_md_openmm.py`); `minimizeEnergy()` at the OpenMM default for morroniside and loganin (`md_compound_fixed.py`)

- Equilibration (NPT, the Monte Carlo barostat is added at system construction and is therefore engaged throughout; **no position restraints**): velocities initialised at 300 K, then **200 ps** (two 100 ps `step(50000)` stages) for catalpol/geniposide/coptisine (`run_md_openmm.py`), or **100 ps** (one `step(50000)` stage) for morroniside/loganin (`md_compound_fixed.py`)
- Production ensemble: NPT, 300 K
- Integrator: `LangevinMiddleIntegrator(300 K, 1.0/ps, 0.002 ps)` (2 fs step)
- Barostat: `MonteCarloBarostat(1.0 atm, 300 K, 25)`
- Electrostatics: particle-mesh Ewald; real-space nonbonded cutoff 1.0 nm; Ewald error tolerance left at the OpenMM default (5e-4)
- van der Waals: Lennard-Jones truncated at the same 1.0 nm cutoff with the long-range analytical dispersion correction enabled and no switching function (OpenMM `NonbondedForce` defaults for a PME system)
- Constraints: bonds to hydrogen
- Production lengths: morroniside 100 ns; loganin 100 ns; catalpol 80.6 ns; geniposide 77.4 ns; coptisine (Gasteiger) 72.2 ns; coptisine (AM1-BCC) 50 ns
- Coordinate save interval: 10 ps (DCDReporter stride 5000 at 2 fs)

6. Trajectory analysis (MDTraj)

- Periodic boundaries reconstructed with `find_molecules()` followed by `image_molecules()` (protein as anchor) before any distance/RMSD calculation
- Protein and ligand selected as `protein` and `not resname UNK` / `resname UNK`
- Backbone RMSD after backbone superposition to frame 0
- Minimum ligand-anchor heavy-atom distance; pocket occupancy = fraction of frames with this distance < 5 Å
- Per-anchor contact occupancy (Arg415/Arg483/Ser508): closest-heavy distance < 4 Å
- Ligand centre-of-mass to anchor-centroid distance
- Binding-site residues: closest-heavy < 6 Å in the first frame
- Native contacts: ligand-residue closest-heavy < 4.5 Å in the first frame; retention = fraction still < 4.5 Å
- Hydrogen bonds: Baker-Hubbard, present in > 10% of frames
- Buried SASA: Shrake-Rupley, probe radius 1.4 Å
- All statistics computed over the full available trajectory of each system
- Reproducing the Table 1 contact / COM / native-contact / H-bond / SASA metrics: set the data root via an environment variable and run the analysis script, e.g. `export MD_ROOT=/path/to/KEAP1_NRF2_ABC_AUTOSAVE && python analyze_contacts.py`. The script iterates over all six systems (Morroniside, Catalpol, Geniposide, Loganin, Coptisine, Coptisine_am1bcc), prints the per-system summary and writes a per-frame `*_contacts.csv` (frame, time_ns, the three per-anchor distances and the ligand-COM distance; the contact count and native-contact fraction are also written when the native-contact computation succeeds) for each. A consolidated `Table1_summary_metrics.csv` with the exact Table 1 values is provided in Online Resource 1.
- Transparency note: `analyze_contacts.py` reproduces the contact / COM / native-contact / H-bond / SASA metrics when the local trajectory (`traj.dcd`) and topology

(`system_topology.pdb`) files are available. The raw MD trajectories are not deposited in Online Resource 1 because of their size, but the per-frame analysis CSVs and the consolidated summary metrics are provided, and the raw trajectories are available from the corresponding author on reasonable request.

7. Associated files

- **Online Resource 1:** per-frame minimum-distance / RMSD CSVs for all six systems (Catalpol, Geniposide, Morroniside, Loganin, Coptisine, Coptisine_am1bcc); a consolidated `Table1_summary_metrics.csv` reproducing every Table 1 value; and representative per-frame contact CSVs for catalpol and geniposide (`Catalpol_contacts.csv` , `Geniposide_contacts.csv` ; per-frame Arg415/Arg483/Ser508 distances and ligand-COM distance). Contact-level outputs for the remaining systems can be regenerated from the raw trajectories with `analyze_contacts.py` .
- Docking driver and Vina configuration (`dock_iridoids.py`)
- OpenMM production scripts (`run_md_openmm.py` ; `md_compound_fixed.py`)
- Trajectory-analysis script (`analyze_contacts.py`)

8. MD driver scripts

The same force field, water model, solvation, integrator, barostat, nonbonded settings and coordinate save interval (Section 5) were used for all five ligands; the two driver scripts differ only in ligand preparation and in the minimisation/equilibration length noted above.

- `run_md_openmm.py` (catalpol, coptisine, geniposide): the ligand graph is taken from the PubChem record (by CID) and the docked seed-101 pose is applied when heavy-atom counts match.
- `md_compound_fixed.py` (morroniside, loganin): the ligand is built directly from the docked pose (PDBQT → PDB) with bond orders assigned from the PubChem SMILES, preserving the docked coordinates; the CID is supplied at run time (`MD_CID`), and the run aborts unless the ligand is placed within 8 Å of the ETGE anchors.

The PubChem CID for morroniside is **11228693** (C17H26O11) throughout this study.