

Supporting Information

Anyone Can Dock: An Online Molecular Docking Tool for Everyone

Kowit Hengphasatporn^{1}, Thanthapatra Bunchuay^{2,3*}, Lian Duan¹, Yasuteru Shigeta^{1,4}*

¹Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan

²Department of Chemistry and Centre of Excellence for Innovation in Chemistry (PERCH-CIC), Faculty of Science, Mahidol University, Bangkok 10400, Thailand

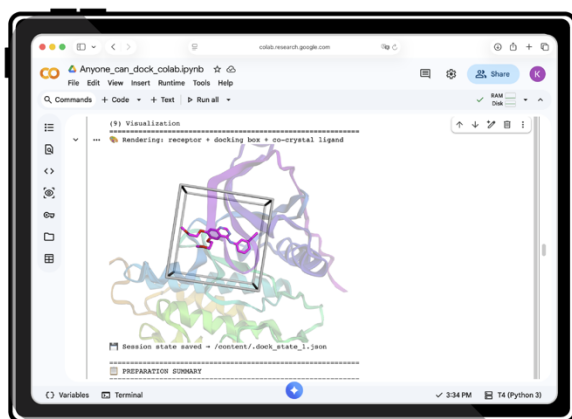
³Center of Sustainable Energy and Green Materials, Mahidol University, Salaya, Nakhon Pathom, Thailand. ⁴Center of Excellence in Structural and Computational Biology, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

*Correspondence: kowith@ccs.tsukuba.ac.jp, thanthapatra.bun@mahidol.ac.th

KEYWORDS: molecular docking, web application, AutoDock Vina, drug design, protein–ligand interaction, virtual screening

Table S1 Supported protein and target types in the platform. The table summarizes target classes currently supported by the docking workflow and notes practical limitations relevant to receptor preparation, grid definition, visualization, and interaction analysis.

Protein or target type	Support	Notes
Standard single-chain proteins	Full	Primary use case.
Multi-chain / homo-oligomers	Full	Duplicate chains can be deduplicated; when the same ligand appears in equivalent chains, the chain A ligand is auto-selected without unnecessary dropdowns.
Heme proteins (CYP450, peroxidases, Hb, Mb)	Full	Fe-porphyrin is handled separately; the grid can be auto-centered on Fe and the heme group is shown in the viewers.
Metal-binding proteins (zinc fingers, carbonic anhydrase)	Full	Common metal ions such as ZN, MG, CA, MN, FE, and CU are re-injected with assigned formal charges after receptor preparation.
MD simulation outputs (GROMACS, AMBER)	Full	Blank chain IDs are automatically assigned to chain A.
Non-standard ligand names (MOL, LIG, UNL, INH)	Full	HETATM keyword handling bypasses potential ProDy misclassification.
Modified amino acids (CYP, MSE, TPO, SEP)	Full	Backbone atom checks retain modified residues as part of the receptor when appropriate.
Multiple co-crystal ligands	Full	Ligand-like HETATM records are auto-scanned. A dropdown is shown only when multiple reference-ligand candidates exist; the selected ligand defines the grid center and is removed from the receptor.
Cofactor-binding proteins (FAD, NAD, ATP, CoA)	Full	Cofactors are classified separately from ligand candidates and can be kept or stripped independently.
Glycoproteins	Partial	Glycans are kept in the receptor, but they are not currently shown in the 2D interaction diagram.
Antibodies / very large proteins	Partial	Supported, but 3D visualization and interaction detection may be slow for very large systems.
Covalent docking	Not supported	AutoDock Vina is designed for non-covalent docking.
Dimer interface binding sites	Not supported	Automatic chain deduplication can remove the second chain, making interface-site docking unsuitable in the current workflow.



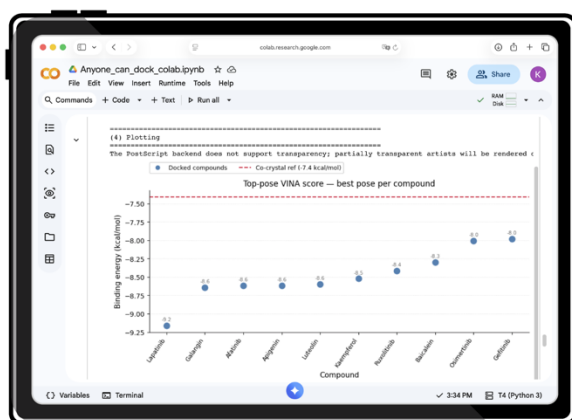
(a) Protein preparation

Type	Name	pKa	Source	Charge @ pH 7.4	Zwitterion?	Starred	Input source	Files
Redock	Erlotinib	7.06	pKaPredict (ML)	0	No	—	SMILES input	molock_Erlotinib_min.pdb, molock_Erlotinib_min.sdf, molock_Erlotinib_min.pdb
Dock	Apigenin	8.16	pKaPredict (ML)	0	No	—	SMILES input	Apigenin_min.pdb, Apigenin_min.sdf, Apigenin_min.pdb
Dock	Bicalutol	7.82	pKaPredict (ML)	0	No	—	SMILES input	Bicalutol_min.pdb, Bicalutol_min.sdf, Bicalutol_min.pdb
Dock	Luteolin	8.10	pKaPredict (ML)	0	No	—	SMILES input	Luteolin_min.pdb, Luteolin_min.sdf, Luteolin_min.pdb
Dock	Kamplifant	7.86	pKaPredict (ML)	0	No	—	SMILES input	Kamplifant_min.pdb, Kamplifant_min.sdf, Kamplifant_min.pdb
Dock	Quercetin	7.65	pKaPredict (ML)	+1	No	—	SMILES input	Quercetin_min.pdb, Quercetin_min.sdf, Quercetin_min.pdb
Dock	Gallatin	6.66	pKaPredict (ML)	+1	No	—	SMILES input	Gallatin_min.pdb, Gallatin_min.sdf, Gallatin_min.pdb
Dock	Lapatinib	6.82	pKaPredict (ML)	+1	No	—	SMILES input	Lapatinib_min.pdb, Lapatinib_min.sdf, Lapatinib_min.pdb
Dock	Aflatinib	7.18	pKaPredict (ML)	+1	No	—	SMILES input	Aflatinib_min.pdb, Aflatinib_min.sdf, Aflatinib_min.pdb
Dock	Galangin	7.82	pKaPredict (ML)	0	No	—	SMILES input	Galangin_min.pdb, Galangin_min.sdf, Galangin_min.pdb

(b) Ligand preparation



(c) Pose-by-pose navigation



(d) Score aggregation and download

Figure S1. Google Colab notebook interface running on iPad. (a) Receptor preparation with metal-safe processing. (b) Batch ligand preparation. (c) Interactive pose browser with RMSD validation. (d) Score aggregation and download. Fixed-seed reruns used canonical RCSB CCD SMILES, RDKit ETKDG seed 71 and Vina seed 71.

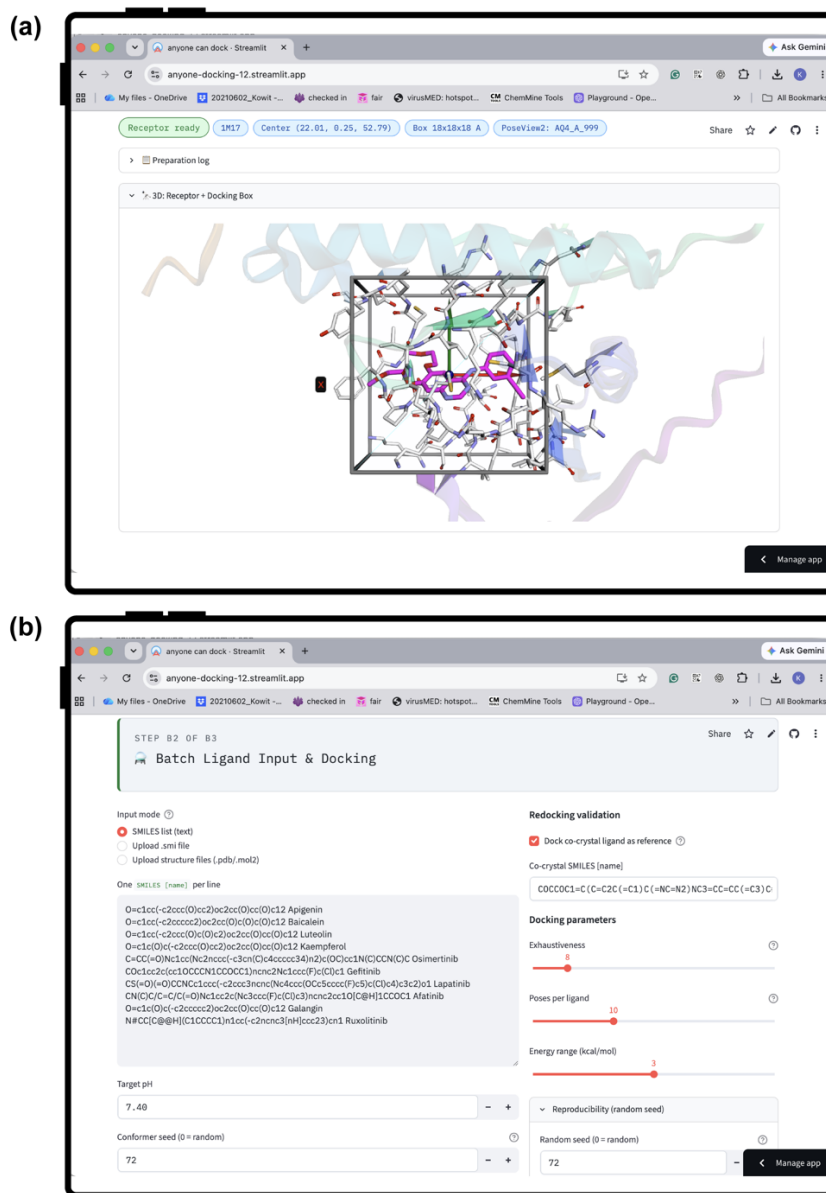


Figure S2. Worked example of the complete batch docking workflow using EGFR kinase (PDB: 1M17) with lapatinib as the test ligand and erlotinib as the co-crystal reference. (a) Receptor preparation: the EGFR kinase domain loaded by PDB accession code with automatic grid centring on the co-crystallised erlotinib ligand (AQ4); the $18 \times 18 \times 18$ Å docking box is displayed in the integrated py3Dmol viewer with spectrum cartoon colouring. (b) Batch ligand input and docking parameters: ten ligands entered as SMILES strings with names, target pH set to 7.4, erlotinib entered as the co-crystal redocking reference, and docking executed at exhaustiveness 8 with 10 poses per ligand. Fixed-seed reruns used canonical RCSB CCD SMILES, RDKit ETKDG seed 71 and Vina seed 71.

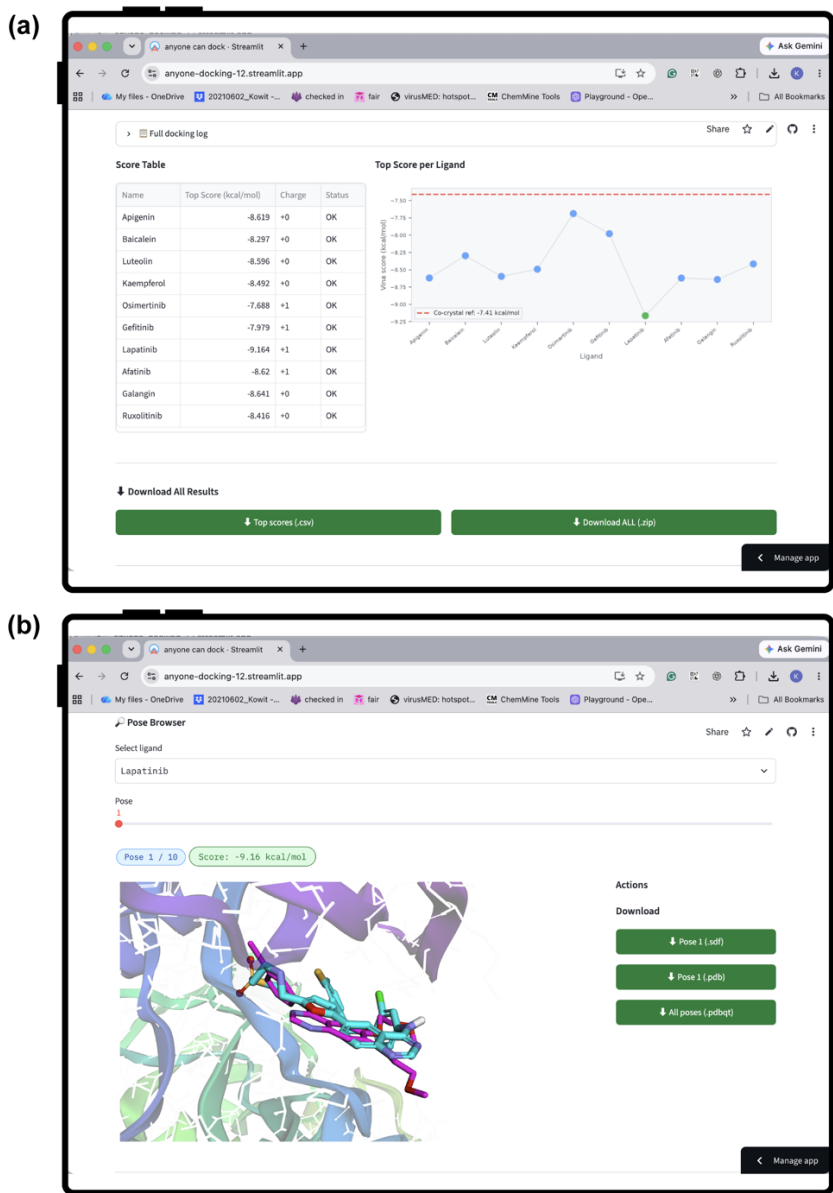


Figure S3. Batch docking results for ten ligands against EGFR kinase (PDB: 1M17). (a) Score table (left) ranking all ten ligands by top predicted binding affinity, and affinity scatter plot (right) with the confirmed erlotinib reference top pose (-7.41 kcal/mol, RMSD 1.39 Å) shown as a dashed red line. (b) Pose browser showing erlotinib redocking top pose (cyan) overlaid with the co-crystal reference (magenta) in the py3Dmol viewer. RMSD 1.39 Å for this single fixed-seed run (seed 71); across ten independent-seed replicates the top-ranked-pose RMSD was 6.42 ± 2.19 Å while a near-native pose (minimum RMSD 1.52 ± 0.16 Å) was sampled in every replicate (see "Validation using 1M17").

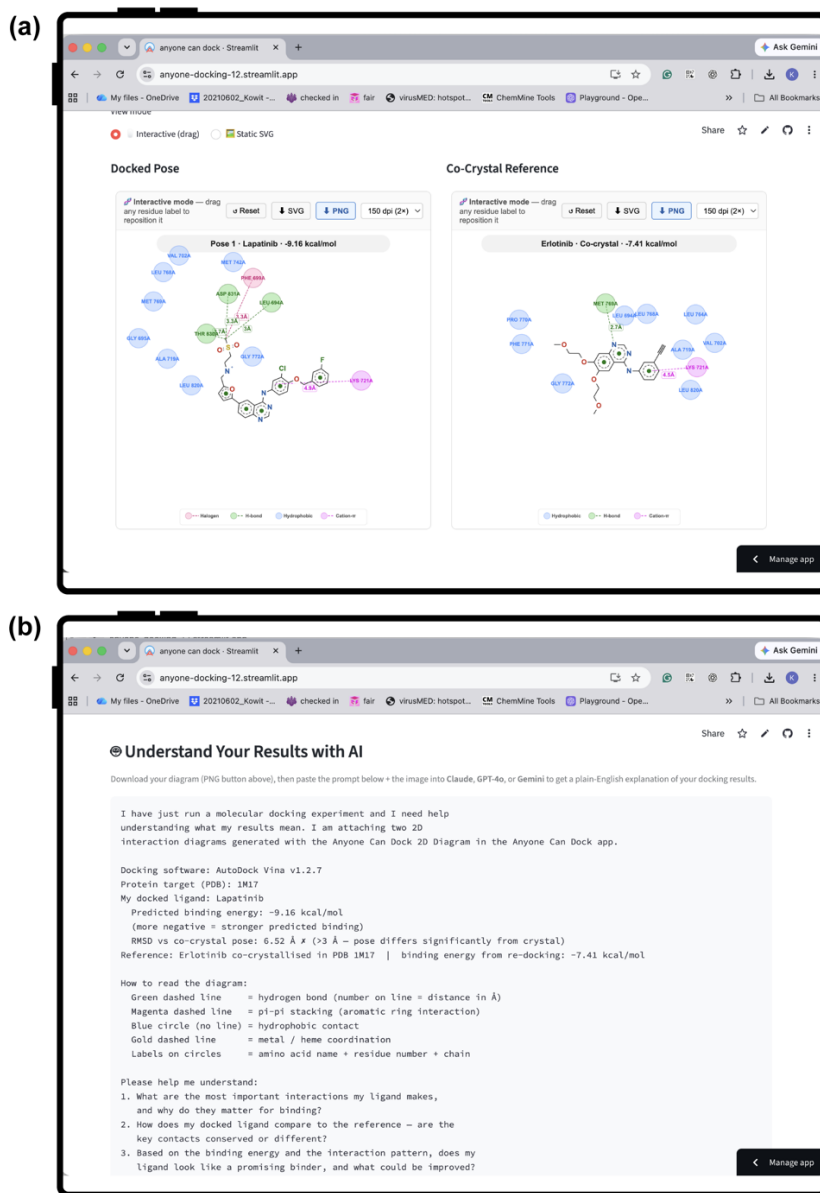


Figure S4. Two-dimensional interaction diagrams and AI-assisted analysis prompt. (a) 2D interaction diagrams for the top-ranked lapatinib docked pose (left, -9.16 kcal/mol) and the erlotinib co-crystal reference (right, -7.41 kcal/mol), with interactions colour-coded by type according to the 2D diagram model. (b) Automatically generated AI analysis prompt, incorporating ligand name, SMILES, binding energy, PDB code, and interaction legend, formatted for direct submission to any large language model to support binding mode interpretation. The AI-generated interpretation is optional and may be confidently incorrect; the prompt embeds the underlying interaction data (contacts, measured distances, binding energy, and legend) so that any model output can be checked against the diagram and coordinates, and it must be verified against the numerical docking results and relevant literature before being treated as a scientific conclusion.

Interface Overview: Streamlit Web App, Google Colab Notebook, and Anyone Can Dock GPT

Anyone Can Dock is delivered through two reproducible computational interfaces that implement the same docking pipeline but serve different user profiles (Figure 3 in main text). The Streamlit web application prioritises zero-friction access: four guided steps, no account, no file conversion, and results within minutes from any browser on any device. The Google Colab notebook is a fully self-contained, standalone iPython notebook in which every stage of the pipeline is written as an editable code cell that users can read, modify, copy, and extend directly without any dependency on an external library module. Anyone Can Dock GPT provides a conversational interface for docking setup assistance, workflow guidance, and interpretation of docking results through natural language interaction; it does not execute docking calculations independently, depends on an external conversational-AI platform, and is not required to reproduce any reported docking result. Both computational interfaces are compared in Table S2.

The web app is appropriate when a result is needed quickly from any device, when the user has no Python background, or when docking 1–20 ligands. The Colab notebook is appropriate when screening more than 20 ligands, GPU acceleration is needed (Vina-GPU 2.1), access to additional docking engines (VinaXB, GNINA, Vina-GPU 2.1) is required, or when the user wishes to modify, inspect, or reuse pipeline code. Anyone Can Dock GPT is appropriate when the user requires guidance on workflow decisions, interpretation of docking scores, or plain-language explanation of results.

Because GPT access depends on an external conversational-AI platform, it is not required for reproducing any docking result reported in the manuscript.

For users who require offline access, institutional data privacy, or repeated high-throughput runs without network dependency, both computational interfaces can alternatively be run on a local machine. The Streamlit app can be launched locally by cloning the repository and running streamlit run app.py in a Python environment with the listed dependencies installed; the Colab notebook can likewise be executed in any local Jupyter environment after installing the same dependency set via the provided requirements.txt. In either case, all docking logic, receptor preparation, and scoring remain identical to their hosted counterparts.

Switching between interfaces, whether hosted or local, does not affect docking accuracy: both use the same receptor preparation logic, the same pK_aNET Cloud+-based protonation-state selection workflow, and the same AutoDock Vina 1.2.7 scoring function.

Table S2 Comparison of the two Anyone Can Dock interfaces. The bottom row lists pipeline components that are identical in both interfaces.

	Web interface	Google Colab notebook
Primary user	Any scientist — no coding needed	Computational researchers, developers
Interface type	Point-and-click GUI, 4-step workflow	Standalone iPython notebook, fully transparent
Installation	None — browser only	Cell-based install per session
Account required	No	Google account
Mobile / tablet	Full — tested iOS and Android	Limited
Docking engines	AutoDock Vina 1.2.7	Vina 1.2.7, VinaXB, GNINA 1.3, Vina-GPU 2.1 (GPU)
Batch size (recommended)	≤ 20 ligands per session	Unlimited
Code visibility	Hidden (black-box interface)	100% visible — every step editable
Copy code for own project	Via GitHub repository only	Directly from any cell
Reproducible notebook format	No	Yes — shareable .ipynb file
Structure drawing	Yes (Ketcher)	No
PoseView 2D diagrams	Yes — auto-generated, side by side	Via API call
AI analysis prompt	Auto-generated	Manual
One-click ZIP download	Yes	Manual from files panel
Shared pipeline (both interfaces)	Metal-preserve receptor prep · pK _a NET Cloud+ protonation selection (Dimorphite-DL initial enumeration → candidate supplementation → tautomer filtering/scoring → multi-criteria ranking/final selection) → ETKDGv3 → MMFF94 → Meeko · AutoDock Vina 1.2.7 · bond-order correction · RMSD validation	

Extended Ligand Preparation Methodology

This section provides full methodological details for the ligand preparation pipeline referenced in the main text. The pipeline executes four sequential automated stages: pK_aNET Cloud+-based protonation-state selection, 3D conformer generation, energy minimisation, and PDBQT conversion.

Stage 1 — pK_aNET Cloud+ Protonation-State Selection with Initial Dimorphite-DL Microstate Enumeration

The implemented protonation-state selection is performed by the pK_aNET Cloud+ workflow. Dimorphite-DL8 is used only as the initial microstate-enumeration component: it identifies ionisable functional groups via SMARTS-based substructure matching against a curated internal database of 1,938 experimentally characterised small molecules and generates candidate protonation states. Each functional group class carries an empirically derived mean pK_a (μ) and standard deviation (σ), together defining a characteristic pK_a range:

$$\text{range_pK}_a = [\mu - n\sigma, \mu + n\sigma]$$

where the precision factor n (default 1.0) governs the breadth of chemical uncertainty captured. For each ionisable moiety, this range is compared against the user-specified target pH window $[\text{pH}_{\min}, \text{pH}_{\max}]$, and one of three outcomes is applied:

- **Deprotonated:** $\mu + n\sigma < \text{pH}_{\min} \Rightarrow$ moiety is deprotonated
- **Protonated:** $\text{pH}_{\max} < \mu - n\sigma \Rightarrow$ moiety is protonated
- **Both forms enumerated:** $\text{range_pK}_a \cap \text{range_pH} \neq \emptyset \Rightarrow$ both forms generated

RDKit tautomer enumeration and filtering precede pH-dependent microstate enumeration. For each accepted tautomer, Dimorphite-DL generates an initial candidate set. These candidates do not constitute the final assignment: pK_aNET Cloud+ supplements the set with additional rule-based microstates, evaluates ionizable sites, applies Henderson–Hasselbalch-informed scoring, and ranks the candidates using multiple criteria. pK_aNET Cloud+ selects the final state used for docking. A standalone Dimorphite-only fallback branch is distinct from successful Dimorphite-DL enumeration inside the integrated workflow.

Two limitations of the initial enumeration step merit explicit acknowledgement. First, Dimorphite-DL evaluates each ionisable site in isolation; long-range electronic coupling between adjacent ionisable groups, such as mutual pK_a perturbation in polyprotic acids or molecules bearing nearby electron-withdrawing substituents, lies outside the scope of that enumerator. Second, Dimorphite-DL addresses ionisation states only and does not by itself enumerate prototropic tautomers, which arise from intramolecular proton transfer rather than net proton gain or loss. In the implemented pipeline, these limitations are mitigated at the workflow level because pK_aNET Cloud+ supplements the candidate set, applies tautomer filtering, and selects the final state through scoring and ranking.

Context-dependent protonation examples. The isolated-site approximation can miss substituent and intramolecular effects that shift pK_a values by several units; representative examples (substituted phenols, coupled hydroxyls, carboxylates, and sulfonates) are given in Section 2.2.2. These shifts show why Dimorphite-DL output should be treated as an initial candidate set rather than a definitive bound-state assignment. pK_a NET Cloud+ validation summary. The updated offline validation used `test_pkanet.py` with the heuristic path, Dimorphite-DL active, full tautomer processing, and Henderson–Hasselbalch scoring; no ML pK_a backend and no PubChem network access were used. The curated regression suite contains 67 tests across 13 functional groups, with 67 passed, 0 failed, and a 100% pass rate. The previous 65 tests remained unchanged, and two tests were added: dorzolamide and NAH bis-acylhydrazone negative control.

New rule. The added ionizable-site rule is ("amine_gamma_ring_sulfonyl", "[NX3;H1,H2;!\$(NC=O);!\$([nH])][CX4;R][CX4;R][CX4;R][SX4;R](=O)(=O)", 6.5, "base"). It identifies primary or secondary amines positioned gamma to a ring sulfone through three ring-constrained tetrahedral carbons; the ring constraints are intended to reduce false positives involving open-chain amines. Dorzolamide had expected net charge 0, predicted net charge 0, and PASS; this corrects the previous +1 prediction produced by the generic aliphatic_amine rule with pK_a 9.5. NAH bis-acylhydrazone had expected charge 0, predicted charge 0, and PASS; because it contains no sulfur matching the new pattern and no relevant amine rule is triggered, it serves as a negative control against obvious unintended activation of the new SMARTS rule.

Benchmark accuracy. For the pK_a Hub drug-like benchmark (pkahub.ttk.hu), the source benchmark comprised 27,218 records, of which 27,183 were included in the reported accuracy calculation. The previous heuristic was correct for 18,128 of 27,183 evaluated molecules (66.69%). The updated heuristic_net_charge method was correct for 18,850 of 27,183 evaluated molecules (69.34%), representing 722 additional correct assignments and a 2.65-percentage-point absolute increase. The full automatic `predict_charge(mode="auto")` pipeline is reported as at least 18,850 of 27,183, or at least 69.34%. Per-charge accuracy improved from 67.8% to 77.1% for expected +1 species (+9.3 percentage points), from 54.8% to 57.7% for expected -1 species (+2.9 percentage points), and from 76.0% to 76.1% for expected neutral species (+0.1 percentage point, essentially stable). Comparative net-charge benchmark. The refined comparison evaluated 27,096 benchmark rows with experimental integer charge labels at pH 7.4. Exact total-charge agreement was 21,951/27,096 (81.01%) for Uni- pK_a with Open Babel recovery for 102 unsupported rows, 20,139/27,096 (74.32%) for MolGpKa, 19,725/27,096 (72.80%) for pK_a NET Cloud+ in automatic mode, 18,975/27,096 (70.03%) for Open Babel, and 16,693/27,096 (61.61%) for pK_a Solver. Every workflow returned a final prediction for all evaluated rows. This comparison evaluates total-charge assignment rather than docking outcomes.

Table S3 Comparison of protonation methods based on exact net-charge agreement and coverage across the benchmark set.

Method	Exact net-charge agreement	Coverage
Uni-pK _a + fallback	81.01%	100%
MolGpK _a	74.32%	100%
pK _a NET Cloud+	72.80%	100%
Open Babel	70.03%	100%
pK _a Solver	61.61%	100%

Comparison of total net-charge prediction accuracy and coverage under a common evaluation protocol at pH 7.4. Accuracy represents the percentage of evaluated compounds for which the predicted integer total formal charge matched the reference charge. Coverage represents the percentage of compounds for which the method returned a prediction. Dataset and evaluation notes. The refined comparison contains 27,096 evaluated rows with experimental integer charge labels at pH 7.4. Coverage denotes the proportion of evaluated rows for which a workflow returned a final integer charge. For Uni-pK_a, 102 unsupported rows were recovered using Open Babel at pH 7.4; this combined workflow is therefore labelled Uni-pK_a and Open Babel recovery rather than standalone Uni-pK_a. The evaluation protocol defines the handling of unsupported, ambiguous, salt, and zwitterionic records; these methodological limitations do not imply experimental microstate validation.

Benchmark limitations. Exact total-net-charge agreement does not guarantee correct assignment of every individual ionizable site, tautomeric state, or crystallographic metal-coordinating microstate. The 67-case suite is a curated regression test suite and is not equivalent to a prospective external clinical or experimental validation. The validation report does not state why 35 of the 27,218 source records were not included in the 27,183-molecule accuracy calculation. Brinzolamide is not corrected by the current SMARTS because its gamma path includes an unsaturated carbon; a separate CX3 variant has not yet been implemented.

This benchmark evaluates exact agreement in total integer net charge. It does not establish atom-specific protonation accuracy, tautomer correctness, or reproduction of an experimentally bound microstate. The benchmark does not directly evaluate the consequences of protonation assignments for Vina score, docking pose RMSD, or virtual-screening enrichment.

Stage 2 — 3D Conformer Generation (RDKit ETKDGv3)

A three-dimensional starting conformation is generated using the ETKDGv3 distance geometry method⁹ implemented in RDKit. ETKDGv3 combines experimentally derived torsional angle distributions from the Cambridge Structural Database with distance geometry embedding and a knowledge-based torsion correction potential, providing an appropriate starting point for force-field refinement. A fixed random seed (randomSeed=71) is used for reproducibility. If

ETKDGv3 is unavailable in the installed RDKit build, ETKDG is used as a fallback. If embedding fails with default parameters, a second attempt is made with useRandomCoords=True to handle molecules where distance geometry fails due to complex ring systems or unusual connectivity. Molecules for which 3D embedding fails after both attempts are flagged with an error and excluded from docking.

Stage 3 — Energy Minimisation (MMFF94 / UFF)

The embedded conformer is energy-minimised using the Merck Molecular Force Field 94 (MMFF94)¹⁰ as implemented in RDKit, with a maximum of 500 iterations. MMFF94 is applied when all force-field parameters are available for the molecule (AllChem.MMFFHasAllMoleculeParams() returns True), relaxing steric clashes and refining bond lengths and angles to yield a docking-ready geometry. For molecules outside the MMFF94 parameterisation scope, primarily those containing atoms or bonding patterns not covered by the original training set, the Universal Force Field (UFF) is applied automatically as a fallback to ensure pipeline continuity without user intervention. Both force fields are applied to the fully hydrogenated molecule.

Stage 4 — PDBQT Conversion (Meeko)

The minimised RDKit molecule is converted to PDBQT format using Meeko, the ligand preparation library developed by the AutoDock Vina team. Meeko assigns AutoDock atom types, sets rotatable bonds, handles ring systems, and writes a PDBQT file compatible with AutoDock Vina 1.2.7. The PDBQTWriterLegacy.write_string() API is used, with automatic fallback to prep.write_pdbqt_string() for older Meeko builds. The formal charge of the prepared ligand, derived from the final protonation state selected by pK_aNET Cloud+, is calculated and presented to the user alongside 2D and 3D structure previews, enabling chemical verification of the protonation outcome before docking proceeds. The SDF file of the minimised conformer is retained as the SMILES template for bond-order correction after docking.

Accepted Input Formats and Ligand Scope

The pipeline accepts three input modalities:

- (i) **SMILES string** — typed or pasted directly. Any valid canonical or isomeric SMILES parseable by RDKit is accepted.
- (ii) **Structure file** — SDF, MOL2, or PDB format. The first molecule is extracted, converted to canonical SMILES via RDKit (SDF) or Open Babel (MOL2, PDB), then passed through the four-stage pipeline. Only the largest fragment is retained if multiple disconnected fragments are present.
- (iii) **Ketcher structure drawing** — the in-browser Ketcher editor (Streamlit app only) returns a SMILES string passed directly to the pipeline.

The automated ligand-preparation workflow is intended primarily for standard drug-like small molecules (MW < 700 Da, Lipinski-like chemical space, and ≤ 15 rotatable bonds). Reliability for macrocycles, organometallic compounds, prodrugs requiring metabolic activation, and molecules with unusual valence has not been established. Covalent inhibitors and metal-binding pharmacophores requiring explicit coordination geometry need additional preparation outside this workflow.

Deployment Instructions

Streamlit Cloud (No Installation Required)

The recommended route for most users is the hosted Streamlit deployment, accessible at the URL below. No account, local installation, or system configuration is required. AutoDock Vina 1.2.7 (ca. 4 MB) is retrieved automatically on first use and cached for the duration of the session.

Live application: <https://nyelidl.github.io/anyone-docking/>

Status indicator: A green “Vina 1.2.7 ready” pill in the application header confirms a successful environment. A red error banner and application halt indicate a missing Open Babel installation.

Local Deployment

Linux (Ubuntu / Debian)

```
sudo apt update
sudo apt install -y python3.11 python3.11-venv openbabel libcairo2-dev libpango1.0-dev
libpangocairo-1.0-0

git clone https://github.com/nyelidl/anyone-docking-local.git
cd anyone-docking-local
python3.11 -m venv venv
source venv/bin/activate

python -m pip install --upgrade pip
python -m pip install -r requirements.txt

streamlit run app.py
```

macOS

```
brew update
brew install python@3.11 open-babel cairo pango

git clone https://github.com/nyelidl/anyone-docking-local.git
cd anyone-docking-local

python3.11 -m venv venv
source venv/bin/activate

python -m pip install --upgrade pip
python -m pip install -r requirements.txt

streamlit run app.py
```

Google Colab

A Colab-compatible notebook is provided in the repository. Open the notebook in Google Colab and execute the installation cell at the top of the notebook prior to running any other cells.

Notebook repository: <https://github.com/nyelidl/anyone-docking>

Verification of Dependencies

Once all checks pass, the Streamlit application header displays a green “Vina 1.2.7 ready” status pill confirming that AutoDock Vina 1.2.7 has been located and is executable.

Reproducibility and validation environment

All reported fixed-seed examples used canonical RCSB Chemical Component Dictionary SMILES, RDKit ETKDG randomSeed=71, AutoDock Vina seed=71, and an $18 \times 18 \times 18$ Å box. The worked 1M17 batch example used exhaustiveness 8 and 10 poses; the separate representative docking protocol in Table S4 used exhaustiveness 16, 10 poses, and an energy range of 3 kcal/mol. Validation runs were performed through the hosted web server, so no local shell command or workstation-specific hardware is part of the reported procedure; runtime values are therefore approximate and environment-dependent. The box_18a_support analysis compared tested box-size groups and supported 18 Å as the default compromise (51.39% top-1 success for 18 Å versus 32.76% for 30 Å). The 4M2U and 2H7Q metal-containing systems are illustrative main-text case studies only; they are not represented as rows in Table S4.

Table S4 Representative docking test cases used to evaluate the applicability of the workflow across different target classes. The table summarizes selected PDB structures, corresponding ligands, and reported AutoDock Vina 1.2.x or Vina-compatible docking scores from the literature as contextual reference values rather than harmonized validation benchmarks. Because these literature values originate from heterogeneous studies using different AutoDock Vina versions, receptor and ligand preparation, grid definitions, and search settings, absolute affinities are not directly comparable across sources; the meaningful cross-study comparison is native-ligand pose reproduction ($\text{RMSD} < 2 \text{ \AA}$), for which all retained cases reproduce the crystallographic pose. A harmonized, recognized-dataset benchmark is provided separately in Table S5 (CASF-2016). Native-ligand redocking cases were used for direct pose-reproduction assessment using “acd redock” command; broader target-class entries indicate technical applicability of the workflow inputs and should not be interpreted as full validation of every listed target class. Fixed-seed reruns used canonical RCSB CCD SMILES, RDKit ETKDG seed 71, Vina seed 71, an $18 \times 18 \times 18 \text{ \AA}$ box, exhaustiveness 16, 10 poses, and an energy range of 3 kcal/mol. These validation runs were performed through the hosted web server; no local shell command or workstation-specific hardware was used. Ligand charge/protonation handling used the deterministic heuristic pK_aNET backend with ML pK_a backends disabled; runtimes are approximate.

PDB ID	Ligand	Other platform (Vina 1.2)	Anyone Can Dock (Vina 1.2.7)	RMSD vs crystal ($\text{\AA}$)	Runtime (minute)
5LX6	Veliparib	−8.40[1]	−8.85	1.01	~1
4O2B	Colchicine	−7.80[2]	−7.14	0.63	~1
6M2N	Baicalein	−6.70[3]	−6.91	0.74	~1
6O0K	Venetoclax	−10.60[4]	−10.70	2.06	~6

Box-size rationale. Box-size analysis supported $18 \text{ \AA}$ as a compromise default: the $18 \text{ \AA}$ group achieved 51.39% top-1 success, whereas the larger $30 \text{ \AA}$ group achieved 32.76%; very large boxes therefore reduced pose recovery.

Validation using 1M17

To assess the reproducibility of the docking workflow, ten independent redocking runs were performed for the EGFR–erlotinib complex (PDB 1M17). Each replicate used a different random seed, with the same seed applied consistently to both RDKit conformer generation and AutoDock Vina docking within that replicate. Across the ten replicates, the mean top-ranked docking affinity was -7.228 ± 0.126 kcal/mol, while the mean RMSD of the top-ranked pose relative to the crystallographic ligand was $6.416 \pm 2.190 \text{ \AA}$. In contrast, the minimum RMSD obtained among all generated poses in each replicate was $1.518 \pm 0.155 \text{ \AA}$, and all ten replicates produced at least one pose within $2.0 \text{ \AA}$ of the experimental binding mode.

These results indicate that the docking protocol reproducibly sampled a near-native binding pose in every replicate. However, due to the stochastic nature of the search algorithm and the limitations of the empirical scoring function, the near-native pose was not consistently assigned the top rank, resulting in variability in the RMSD of the highest-scoring pose.

Table S5 CASF-2016 ligand-preparation comparison between pK_aNET Cloud+ and Open Babel across 147 paired complexes. All docking used fixed seeds (RDKit ETKDG seed 71, AutoDock Vina seed 71), an $18 \times 18 \times 18$ Å box, exhaustiveness 16, 10 poses, and a 3 kcal/mol energy range. Python code for validation is available in github (<https://github.com/nyelidl/anyone-docking>).

Metric	pK_aNET Cloud+	Open Babel
Charge matches	123/147 (83.67%)	93/147 (63.27%)
Median top affinity	−8.664 kcal/mol	−8.519 kcal/mol
Top-affinity wins	74	73
Median top-pose RMSD	2.370 Å	2.160 Å
Top-pose RMSD ≤ 2 Å	68/147 (46.3%)	71/147 (48.3%)
Median minimum RMSD	1.020 Å	1.210 Å
Minimum RMSD ≤ 2 Å	112/147 (76.2%)	105/147 (71.4%)
Minimum-RMSD wins	85	57

Table S6 Troubleshooting

Error / symptom	Solution
"Zzzz" This app has gone to sleep due to inactivity. Would you like to wake it back up?" in streamlit web application	To wake the application, simply click " <i>Yes, get this app back up!</i> "
"OpenBabel not found" red banner on startup	Install Open Babel: <code>conda install -c conda-forge openbabel</code> or <code>sudo apt-get install openbabel</code> . Restart the Streamlit server after installation.
"Could not download Vina binary"	Check internet connection. If behind a firewall, download <code>vina_1.2.7_linux_x86_64</code> manually from the AutoDock-Vina GitHub releases page, place at <code>/tmp/vina_1.2.7</code> , and run <code>chmod +x /tmp/vina_1.2.7</code> .
Receptor preparation fails for a metal-containing PDB	Verify the PDB file is valid and not truncated. The metal-safe protocol should handle all standard HETATM metal records. Report the PDB code as a GitHub issue if the problem persists.
PoseView returns error or times out	PoseView is an external REST API (Proteins.plus). Errors indicate temporary service downtime — retry after a few minutes. 3D viewing and all downloads remain available regardless of PoseView status.
"RDKit could not parse SMILES"	Verify the SMILES using an external validator. Common issues: mismatched parentheses, invalid valence, unsupported element symbols. Use the Ketcher drawing tool as an alternative input mode.
Bond-order correction warning flag	Occurs when RDKit cannot map the docked pose to the input-SMILES template (<code>AssignBondOrdersFromTemplate</code> fails); common causes are symmetric scaffolds, tautomers, or incompletely perceived connectivity. The uncorrected pose is returned and flagged (status written to the SDF); inspect the raw SDF before downstream use. (The 60% MCS-coverage cutoff is separate and applies to RMSD validity, not bond-order correction.)
Batch job terminates partway through	On Streamlit Cloud, jobs exceeding session CPU limits are terminated. Reduce batch size to ≤ 20 ligands or lower exhaustiveness to 8. For large batches, use the Google Colab notebook or local deployment.
py3Dmol viewer shows blank or black screen	Browser extensions (ad-blockers, script-blockers) may block WebGL. Try in an incognito window or a different browser. All viewers require WebGL support.
RMSD shows "no reference" indicator	The co-crystal ligand was not detected at the receptor preparation step. Use Manual XYZ or ProDy selection

mode for grid centering. RMSD display requires a co-crystal reference identified during receptor preparation.

Reference

1. Boualia B, Sandeli Ae-K, Evren E, Özdemir İ, Karci H, Dündar M, Özdemir İ, Gürbüz N, Koç A, Boulebd H: Synthesis, characterization, and anticancer potential of novel NHC ligands and their selenium complexes: a combined in vitro and in silico investigation. *RSC Advances* 2026, 16(1):409-424.
2. Al-Wahaibi LH, Elshamsy AM, Ali TFS, Youssif BGM, Bräse S, Abdel-Aziz M, El-Koussi NA: Design, synthesis, and antiproliferative activity of novel thiazole-based derivatives as tubulin polymerization inhibitors targeting the colchicine binding site. *RSC Adv* 2026, 16(27):25103-25115.
3. Sandalli C, Bostancioglu SM, Sandalli A, Turumtay EA, Almohazey D, Alsaeed M, Alhamid G, Rabaan AA, Turumtay H, Ami E-i *et al*: Targeting SARS-CoV-2 main protease (3CLpro) with Paeonia-derived phytochemicals. *In Silico Pharmacology* 2026, 14(2):122.
4. Almansour NM, Allemailem KS, Abd El Aty AA, Ismail EIF, Ibrahim MAA: In Silico Mining of Natural Products Atlas (NPAtlas) Database for Identifying Effective Bcl-2 Inhibitors: Molecular Docking, Molecular Dynamics, and Pharmacokinetics Characteristics. *Molecules* 2023, 28(2).