Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Anyone Can Dock: An Online Molecular Docking Tool for Everyone

Kowit Hengphasatporn

kowith@ccs.tsukuba.ac.jp

University of Tsukuba

Thanthapatra Bunchuay

Mahidol University

Lian Duan

University of Tsukuba

Yasuteru Shigeta

University of Tsukuba

software

Keywords: molecular docking, web application, AutoDock Vina, drug design, protein-ligand interaction, virtual screening

Posted Date: May 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9763995/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Molecular docking predicts how small molecules bind to protein targets, but performing a docking calculation has traditionally required software installation, command-line scripting, and manual molecular preparation. Anyone Can Dock is a ready-to-use online platform developed to reduce the technical and procedural burden of conventional docking through three complementary interfaces. First, the Streamlit web application provides a simple four-step workflow: enter a protein, enter target ligands, click dock, and view the result. Second, the Google Colab notebook supports four docking engines, AutoDock Vina 1.2.7, VinaXB, GNINA 1.3, and Uni-Dock, enabling batch virtual screening, teaching, and method development. Third, Anyone Can Dock GPT provides conversational workflow guidance and docking-result interpretation through natural language interaction. These interfaces simplify common practical steps in docking workflows, including receptor preparation for metal-containing structures, chemically consistent pose export after PDBQT conversion, and built-in RMSD-based redocking assessment when a co-crystal reference ligand is available. The web platform is freely available at https://nyelidl.github.io/anyone-docking/.

Scientific Contribution

Anyone Can Dock addresses three practical barriers in accessible molecular docking: automated metal- preserving receptor preparation that preserves metal ion and heme-containing active sites during PDBQT conversion, pH-aware ligand protonation-state assignment via pKaNET Cloud+ with experimental pKa benchmarking, and SMILES-template-based bond-order correction of docked poses for chemically interpretable visualization. The platform unifies these components within a zero-installation browser-based interface requiring only four user steps, validated across standard, zinc-metalloprotein, and heme-containing docking targets. The complete workflow is freely accessible without registration through three complementary interfaces: a Streamlit web application, a Google Colab notebook, and Anyone Can Dock GPT.

1. Introduction

Molecular docking is one of the most widely used techniques in computer-aided drug design, enabling prediction of the non-covalent binding modes of small molecules to target proteins[1]. As a versatile screening tool, molecular docking helps researchers quickly identify the most promising compounds across a wide range of disease areas, from cancer[2, 3] and viral infections[4, 5] to any condition where fast, low-cost computational screening can make a real difference. A wide range of established docking programs is now available, including AutoDock Vina[6], Glide[7], Gold[8], GNINA[9], PyRx[10], iGEMDOCK[11], and Uni-Dock[12]. Together with the continued growth of experimentally determined protein structures in the Protein Data Bank, these developments have increased the demand for accessible structure-based molecular modeling tools.

Although molecular docking has become an indispensable tool in modern research it remains impractical for many potential users. This limitation does not primarily arise from a lack of conceptual understanding, but rather from technical barriers to its practical use. Freely available open-source engines still require command-line literacy, software environment setup, and manual preparation of multiple input files[6]. In addition, ligand preparation is one of the most error-prone steps in molecular docking workflows. Reliable docking requires the assignment of appropriate protonation states, tautomeric forms, formal charges, and three-dimensional geometries that depend on both the intrinsic pK_a values of the ligand and the pH of the binding environment. In conventional workflows, these steps often require multiple software tools and manual pre-processing before docking can be performed, creating opportunities for inconsistent input structures and user-dependent errors. In addition to these usability barriers, several practical workflow limitations remain common in accessible docking platforms. First, metal-containing proteins, such as zinc

metalloproteases, magnesium-dependent kinases, and calcium-binding proteins, may be incompletely supported by simplified docking workflows, and failures during receptor preparation are not always reported clearly to the user[13]. This issue is particularly relevant during automated PDBQT conversion, where metal ion can interfere with charge assignment using the classical pipeline. Second, the PDBQT file format used by AutoDock Vina does not explicitly retain bond-order and formal charge information, so docked poses exported without post-processing may exhibit with simplified bonding patterns that complicate visualization and downstream analysis. Anyone Can Dock addresses this issue by applying SMILES-template-based bond-order reassignment to all docked output poses. Third, few docking platforms provide an integrated RMSD calculation against the co-crystallized ligand, leaving pose accuracy unvalidated relative to experimental structures. Although recent web-based platforms including Webina[14], ezCADD[15], and WeMol[16] have improved usability, few combine these practical workflow components within a single zero-installation interface. Herein, we present Anyone Can Dock, a web-based molecular docking platform that translates the automated, pH-aware molecular preparation concept of pKaNET Cloud+ into an end-to-end docking workflow. Building on the workflow framework introduced in our previous distance-guided fully dynamic docking (DFDD) study[17], Anyone Can Dock does not introduce a new docking engine, instead it integrates established docking tools with automated receptor preparation, ligand protonation-state assignment, pose post-processing, RMSD-based validation, and interactive visualization within a single browser-based interface. Anyone Can Dock has been developed in the form of the web application and Google Colab implementation

In addition, the Streamlit web application and Google Colab implementation, Anyone Can Dock GPT is provided as an optional natural-language companion for workflow guidance and docking-result interpretation. The web application is organized into four steps; (i) enter or upload a receptor, (ii) enter a ligand, (iii) run docking, and (iv) view results through integrated 2D interaction diagrams, RMSD validation, and bond-order-corrected downloadable poses (Fig. 1). Typical single-ligand workflows can be completed within approximately ten minutes using default settings. Beyond the hosted application, the platform supports local deployment for restricted or privacy-sensitive environments, as well as a Google Colab version for browser-based use. No registration is required.

2. Implementation

2.1 Software Architecture and Deployment

Anyone Can Dock is distributed through three complementary interfaces, which is a Streamlit web application and a self-contained Google Colab notebook[18], and Anyone Can Dock GPT, all designed to support docking workflows at different user-expertise levels. The Streamlit and Colab implementations are openly available on GitHub under an open-source license (Fig. 2). Although the Streamlit application is accessible online, it can equally be run locally by cloning the repository and launching it in any standard Python environment (Supporting information). The web interface supports deployment through Streamlit Community Cloud, Google Colab virtual runtime environments, and user-operated local workstations. In the web application, app.py manages the user interface and session state, whereas all scientific operations are delegated to a shared computational backend, core.py. The same backend logic is also exposed through the Colab notebook in an interactive iPython environment. Anyone Can Dock GPT serves as an optional support layer for workflow guidance and result interpretation. The core.py engine implements receptor preparation, ligand preparation, docking execution, bond-order correction, RMSD-based validation, and interaction diagram generation (default: custom built-in engine; optional: PoseView REST API). Runtime setup is automated: the required AutoDock Vina 1.2.7 binary is downloaded on first use, cached locally, and configured without requiring command-line installation. The platform is organized as a four-step workflow comprising receptor preparation, ligand preparation, molecular docking, and results/export. At the end of each session, outputs are bundled into a ZIP archive containing docked poses, prepared receptor files, score tables, interaction diagram images, and summary plots. This

modular separation between interface, runtime setup, computational engine, session workflow, and export layer improves portability, testability, and extensibility. Software dependencies are summarized in Table 1, and detailed installation instructions and environment specifications are provided in the Supporting Information.

Table 1
Software dependencies of Anyone Can Dock. All components are open-source or freely accessible via REST API. Components are grouped by functions and annotated by deployment interface, streamlit and colab versions.

Component	Version	Role	Interface
Receptor Preparation			
ProDy	≥ 2.4	PDB parsing, metal-safe processing, co-crystal ligand detection, grid centre calculation	*Both*
OpenBabel	≥ 3.1	Receptor hydrogen addition and PDBQT conversion	*Both*
Ligand Preparation			
pKaNET Cloud+	latest	Tautomer-aware Henderson–Hasselbalch microstate ranking for pH-aware protonation state assignment (Dimorphite-DL used internally for ionization-state enumeration)	*Both*
RDKit	≥ 2023	SMILES parsing, ETKDGv3 conformer generation, MMFF94 minimisation, bond-order correction, RMSD (MCS), 2D interaction diagram	*Both*
Meeko	latest	Ligand PDBQT preparation	*Both*
Docking Engines			
AutoDock Vina	1.2.7	Primary docking engine — pose sampling and scoring	*Both*
VinaXB	latest	Halogen-bond-aware scoring	*Colab*
GNINA	1.3	CNN-based rescoring (CPU/GPU)	*Colab*
Uni-Dock	1.1.0	GPU-accelerated batch docking	*Colab*
Visualisation & Interaction Diagrams			
py3Dmol	latest	Interactive 3D browser-based molecular viewer	*Both*
RDKit (2D)	≥ 2023	Default 2D protein–ligand interaction diagram	*Both*
PoseView v1	REST API	2D interaction diagram — docked pose (optional)	*Both*
PoseView2	REST API	2D interaction diagram — co-crystal reference (optional)	*Both*
Interface & Input			
Streamlit	latest	Web UI framework and session-state management	*Streamlit*
Ketcher	latest	In-browser 2D molecular structure drawing	*Streamlit*
condacolab / Miniforge	latest	Conda environment setup in Google Colab	*Colab*

2.2 A Four-Step Workflow

2.2.1 Receptor Preparation

The target protein is loaded by PDB accession code or by uploading a local PDB file (.pdb or .cif). The Streamlit interface also provides optional keyword-based protein search through the RCSB Protein Data Bank API, allowing users to identify a target structure and automatically populate the PDB ID field directly from the search results. Structural parsing and grid-center detection are performed using ProDy[19]. The platform supports docking workflows for diverse protein systems, including single- and multi-chain proteins, metalloproteins, heme proteins, cofactor-binding proteins, modified residues, and MD-derived structures, with current limitations for glycans, very large complexes, membrane/nucleic-acid targets, covalent docking, and dimer-interface binding sites, as detailed in the Supporting information (Table S1). The grid box center is determined in one of three modes: (i) *Auto-detect*, the geometric center of the co-crystallized ligand, identified automatically from HETATM records after filtering out water, ions, glycans, and common crystallographic additives; (ii) *Manual XYZ*, direct coordinate entry; or (iii) *ProDy selection*, any valid ProDy atom selection string (e.g., *resname ATP*). The interface also allows users to manually select whether detected HETATM records should be retained, removed, or used as the reference ligand for grid generation. The default grid box size is 16 × 16 × 16 Å. Receptor PDBQT preparation follows a metal-preserving three-stage protocol designed to avoid the Gasteiger charge calculator crash that affects OpenBabel when metal HETATM records are present. In stage (a), all metal ion HETATM records (Mg, Zn, Ca, Mn, Fe, Cu, Co, Ni, Cd, Hg, Na, K, and selected lanthanides) and heme group atoms (HEM, HEC, HEA, and related variants) are temporarily stripped before OpenBabel processing. In stage (b), hydrogen atoms are added and the file is converted to PDBQT format with Gasteiger partial charges applied cleanly. In stage (c), the stripped metal ions are re-injected into the final PDBQT file with correct formal charges (e.g., Zn^{2+}, Mg^{2+}, Fe^{3+}), restoring the complete receptor. This protocol is intended to preserve metal-containing structures during receptor preparation and visualization. It does not model explicit coordination geometry, polarization, or covalent metal–ligand bonding, and metal-coordinating poses should therefore be interpreted cautiously.

2.2.2 Ligand Preparation via pKaNET Cloud+

The ligand preparation pipeline, adapted from pKaNET Cloud+, executes a four-stage automated protocol: (i) protonation state assignment via pKaNET Cloud+, a tautomer-aware Henderson–Hasselbalch microstate ranking engine with a corrected heuristic pK_a table, which uses Dimorphite-DL[20] for ionization-state enumeration, at a user-specified pH (default 7.4, range 0–14); the current implementation standardizes the input SMILES with RDKit, optionally queries PubChem for experimental dissociation constants, ranks candidate protonation/tautomer microstates, and selects the recommended state for docking. (ii) 3D conformer generation using RDKit ETKDGv3[21]; (iii) energy minimization with the MMFF94 force field[22] (UFF fallback); and (iv) PDBQT conversion via Meeko[23] for AutoDock Vina input. Formal charge and 2D/3D structure previews are presented for user verification before docking proceeds. Ligands can also be searched by compound name through the PubChem API, which retrieves the compound CID, molecular formula, molecular weight, and SMILES representation for direct ligand preparation within the interface. Users may also choose a neutral “add H only” mode for pre-prepared ligands or uploaded structures when they do not want automated pH-based protonation. The pipeline is validated for standard drug-like small molecules (MW < 700, Lipinski-like chemical space); macrocycles, highly flexible molecules (> 15 rotatable bonds), and covalent inhibitors require additional preparation. The protonation-state assignment module in pKaNET Cloud + was benchmarked against 27,283 molecules, a curated docking-relevant subset (65 compounds) and a drug-like subset drawn from the experimental pK_a database pKaHub (Sipos-Szabó et al. [24]; 27,218 compounds), achieving an overall net-charge agreement of 69.3% at pH 7.4; full benchmark data and extended ligand preparation methodology are provided in the Supporting Information.

2.2.3 Molecular Docking

In the web application, molecular docking is performed by AutoDock Vina 1.2.7[6]. The Vina binary is downloaded and cached automatically on first run, requiring no manual installation. The user adjusts exhaustiveness (default 16, range 4–64), number of binding modes (default 10, range 5–20), and energy range (default 3 kcal/mol, range 1–5). The AutoDock Vina binary is automatically downloaded and executed on the host machine, whether Streamlit Cloud, Google Colab, or a local workstation. On the Streamlit Community Cloud free tier, using the default exhaustiveness value of 16, the end-to-end single-ligand workflow typically completes in approximately 6–10 min, including receptor preparation (~ 1 min), ligand preparation (~ 1 min), docking (~ 2–3 min), and results visualization (~ 2–5 min). For batch docking of ten ligands with exhaustiveness 8, the docking step requires approximately 5–8 min on the same platform. In the Google Colab implementation, additional docking engines are available, including VinaXB[25], GNINA 1.3[9], and Uni-Dock 1.1.0[12]. Their timing and accuracy depend on the protein–ligand system size, hardware configuration, and available computational resources, particularly GPU availability; readers are referred to the respective original publications for performance benchmarks. Anyone Can Dock GPT additionally provides a conversational interface for docking setup assistance, workflow guidance, and interpretation of docking results through natural language interaction. The web application also supports optional batch docking through multi-SMILES input, enabling simultaneous preparation and docking of multiple ligands within a single workflow. The PDBQT format used in AutoDock Vina does not preserve bond-order information during docking output generation. As a result, exported SDF structures may contain simplified bonding patterns that are less suitable for downstream visualization and cheminformatics analysis. Anyone Can Dock therefore applies an automatic SMILES-template-based bond-order reassignment step using RDKit's[26] AssignBondOrdersFromTemplate function after docking. The corrected structures are used for 2D interaction visualization, RMSD calculation, and downloadable SDF/PDB outputs. If reassignment fails, the original structure is retained and reported with a warning message.

2.2.4 Results Visualization and Analysis

Docking results are presented through an integrated visualization and analysis interface. Docked poses are visualized interactively in three dimensions using py3Dmol[27], including binding-pocket inspection and ligand pose comparison. Metal ions and retained HETATM records are also displayed in the visualization environment when present in the receptor structure. Two-dimensional protein–ligand interaction diagrams are generated by default using the platform's built-in custom interaction diagram engine, Anyonc Can Dock diagram (implemented in RDKit); an optional alternative diagram may be requested through the PoseView REST API provided by ProteinsPlus[28], with downloadable PNG and SVG outputs available through the interface. If PoseView generation fails or is unavailable, the workflow continues without interrupting the docking calculation.

For redocking workflows, pose accuracy can be assessed through RMSD comparison against the co-crystallized reference ligand when chemically comparable structures are available. RMSD calculations are performed after automatic bond-order correction and maximum common substructure (MCS) alignment using RDKit[26]. RMSD values should be interpreted only for structurally related ligands and are not intended for comparison between chemically unrelated compounds.

The platform additionally provides rapid calculation of common physicochemical descriptors and selected ADME-related properties, including molecular weight, LogP, hydrogen bond donor and acceptor counts, topological polar surface area (TPSA), rotatable bonds, Lipinski rule-of-five assessment, and bioavailability-related descriptors using RDKit[26] and, when the optional ADMET-AI package is installed, machine-learning-based ADMET predictions[29]. Results can be exported as downloadable structure and analysis files, including SDF, PDB, CSV, PNG, and SVG formats, depending on the selected workflow.

2.3 Understand Your Results with AI

An AI prompt section is provided below the 2D interaction diagram and automatically adapts to the current docking session. For docked ligands, the prompt generates a plain-language interpretation of the observed interactions together with a ready-to-use summary paragraph. When a co-crystal reference is available, the prompt is expanded to support comparison with the reference ligand shown in the 2D diagram, and, if redocking has been performed, it also includes the reference binding energy. For heme-containing targets, the prompt describes the gold dashed line used to indicate metal or heme coordination in the diagram legend. Users can copy the prompt together with a screenshot of the 2D diagram into Claude, GPT-4o, or Gemini to obtain an interaction-based interpretation; the prompt ends with a "Ready-to-use summary" section containing a concise 3–4 sentence paragraph suitable for reports or presentation slides.

2.4 Built-in RMSD Validation

When a co-crystallized reference ligand is available, Anyone Can Dock provides an optional built-in redocking validation workflow. The docked and reference ligands are processed through automatic bond-order correction and aligned using RDKit[26] maximum common substructure (MCS)-based atom mapping before RMSD calculation. RMSD analysis is intended for redocking or structurally related ligands and should not be interpreted for chemically unrelated compounds. The calculated RMSD value is reported directly in the interface together with docking scores and interaction visualizations. Docking poses are additionally classified into common qualitative ranges (< 2 Å, 2–3 Å, and > 3 Å) to assist rapid interpretation.

2.5 Batch Mode for Virtual Screening

Anyone Can Dock supports optional batch docking through multi-SMILES input in both the Streamlit web application and Google Colab workflow. Multiple ligands can be submitted in a single run, where each structure undergoes automated ligand preparation, protonation-state assignment, docking calculation, and result generation independently. Batch results are summarized in a table, along with docking scores, physicochemical descriptors, and downloadable structure files. In the Google Colab implementation, batch workflows additionally support multiple docking engines, enabling rapid virtual screening and educational benchmarking workflows without local software installation.

3. Results and Discussion

3.1 User Interfaces and Workflow

The Streamlit web application is designed around a simplified four-step workflow: (i) receptor preparation, (ii) ligand preparation, (iii) molecular docking, and (iv) results visualization and export (Fig. 3, Table S2). The interface supports direct upload of local structure files as well as optional database-assisted input through the RCSB Protein Data Bank and PubChem APIs. Protein structures can be identified through keyword-based RCSB searches, whereas ligand structures can be retrieved directly from PubChem using either compound names or identifiers. The workflow is designed to minimize manual preprocessing and command-line operations for non-expert users. Receptor preparation, ligand protonation-state assignment, 3D conformer generation, docking execution, and post-processing analysis are performed automatically within the interface. Anyone Can Dock GPT additionally provides conversational workflow assistance and docking-result interpretation through natural language interaction.

The Google Colab implementation provides an alternative notebook-based interface for batch docking, teaching, and reproducible workflows. In contrast to the Streamlit interface, the Colab workflow additionally exposes multiple docking engines, including AutoDock Vina 1.2.7, VinaXB, GNINA, and Uni-Dock, together with editable runtime parameters for advanced users.

3.2 Illustrative Case Studies

To illustrate the workflow (Figures S2–S4), three protein–ligand systems were examined using the Streamlit batch mode of Anyone Can Dock: a virtual-screening application against the EGFR kinase domain (PDB 1M17), and two redocking benchmarks against the metalloenzymes human carbonic anhydrase II (PDB 4M2U) and wild-type cytochrome P450cam (PDB 2H7Q). In every case, ligands were entered as SMILES strings, protonated at pH 7.4, and docked at an exhaustiveness of 8 using automatic grid centering based-on the co-crystallized ligand. For redocking benchmarks, pose accuracy was evaluated by heavy-atom RMSD relative to the crystallographic ligand pose, using the conventional 2.0 Å threshold as an indicator of successful pose reproduction. For the EGFR virtual-screening example, RMSD values relative to the erlotinib reference were used only to assess whether candidate ligands adopted erlotinib-like or distinct binding orientations, rather than as a direct measure of pose accuracy for chemically different ligands.

Case 1 EGFR Kinase Domain – Virtual Screening Against the Erlotinib Reference

A set of ten ligands was docked into the EGFR kinase domain, with erlotinib included as the co-crystallized reference ligand for redocking and internal comparison. Redocking of erlotinib produced a top-ranked pose with a Vina score of − 7.26 kcal/mol and a crystallographic-pose RMSD of 1.59 Å, providing a reference for assessing the docking setup. Among the remaining nine candidates, lapatinib gave the most favorable predicted Vina score of − 9.26 kcal/mol (Fig. 4a). Overlay with the co-crystallized erlotinib pose (Fig. 4b) showed that lapatinib occupied the ATP-binding cleft but adopted a substantially different binding orientation, with an RMSD of 6.74 Å relative to the erlotinib reference. The Anyone Can Dock 2D interaction diagram (Fig. 4d) identified a hydrogen bond with Leu694A (3.1 Å), a halogen contact with Phe699A (3.5 Å), and a cation–π interaction with Lys721A (4.7 Å), together with hydrophobic contacts to Leu768A, Met742A, Met769A, Val702A, Pro770A, Gly695A, Leu764A, Gly772A, Ala719A, and Leu820A.

Although the Vina score suggests favorable predicted binding relative to redocked erlotinib, the large RMSD indicates that lapatinib does not reproduce the erlotinib binding mode, and the result should not be overinterpreted given the empirical nature of the Vina scoring function. The mechanistic interpretation presented here was generated with the assistance of an AI language model, using receptor identity, ligand identity, predicted affinity, key contact residues, and RMSD as input; this example is fully reproducible at https://nyelidl.github.io/anyone-docking/ using PDB code 1M17.

Case 2 and 3: Metalloprotein Redocking Case Studies: Zinc- and Heme-Containing Systems

Additionally, two redocking case studies were carried out to assess the practical handling of metal-containing receptors in Anyone Can Dock. These systems were selected because metal-associated binding modes are commonly challenging for automated receptor preparation and empirical docking scores. The examples include a drug-sized sulfonamide inhibitor anchored to an active-site Zn^{2+} in human carbonic anhydrase II (dorzolamide in PDB 4M2U), and a small N-donor fragment coordinating the heme iron in wild-type cytochrome P450cam (imidazole in PDB 2H7Q).

For carbonic anhydrase (CA) II, the top-ranked dorzolamide pose reproduced the crystallographic binding mode with an RMSD of 1.39 Å and a Vina score of -7.07 kcal/mol (Fig. 5a,b). Key interaction characteristics of sulfonamide-class CA inhibitors were reproduced including close contact of the sulfonamide group near the catalytic Zn^{2+} center (Zn301A coordinated by His94, His96, and His119), together with hydrogen-bonding contacts involving His94, Thr199, and Gln92. Additional contacts with Thr200, Pro201, His119, Val121, Phe131, Val135, Leu198, and Pro202 helped define the binding pocket around the thienothiopyran scaffold and ethylamino substituent.

The ligand was prepared with a net charge of + 1, corresponding to a protonated secondary amine and neutral sulfonamide under the automated ligand-preparation workflow via pKaNET Cloud+, in line with standard Vina input conventions; although the experimentally bound state involves a deprotonated N–H sulfonamide coordinating zinc Reproduction of both the zinc-anchoring geometry and the Thr199 hydrogen-bond network, the two hallmarks of

clinical sulfonamide CA inhibitors, indicates that the protocol captures the dominant determinants of binding for this class.

For cytochrome P450cam, the top-ranked imidazole pose reproduced the crystallographic binding mode with an RMSD of 0.46 Å and a Vina score of − 3.43 kcal/mol (Fig. 5c,d). The defining geometric feature was placement of an imidazole nitrogen near the heme iron at 1.9 Å, consistent with the type-II binding mode of nitrogen-donor ligands in P450 active sites. Additional contacts involved Leu244, Val247, and Gly248, which line the small distal cavity above the heme. Although the docking pose closely matched the crystal structure (RMSD of 0.46 Å vs co-crystal imidazole), the predicted Vina score was relatively weak. This is expected for a small fragment such as imidazole because it has only a limited contact surface with the protein. In addition, the key Fe–N coordination at 1.9 Å is not explicitly treated as a coordinate covalent bond in the Vina scoring function. Therefore, the low binding score does not mean that the pose is incorrect. In this case, the docking result should be evaluated mainly by the binding geometry, which showed excellent agreement with the crystallographic pose. Therefore, the two metalloprotein examples show that the platform can reproduce experimentally observed binding poses with good accuracy, as indicated by RMSD values below 2.0 Å. However, the docking score should be interpreted more carefully for small metal-coordinating fragments than for larger drug-like ligands, because metal coordination can be underrepresented by empirical scoring functions such as Vina. This interpretation is consistent with the known challenge of docking metal-containing or metal-coordinating systems, where standard empirical docking scores often require specialized treatment of metal parameters and coordination geometry[13]. Apart from the three representative case studies discussed in the main text, additional docking tests were performed to further evaluate the applicability of the workflow across diverse protein target classes, as summarized in Table S3.

3.3 Comparison with Existing Platforms

We compared Anyone Can Dock against nine representative freely accessible docking platforms across nineteen functional dimensions (Table 2). Platforms were selected to span the full range of accessible modalities: command-line (Vina CLI[6]), desktop GUI (PyRx[10]), web server (SwissDock[30]), blind docking (CB-Dock2[31]), WebAssembly (MolModa[32]), local GUI (DockingApp[33]), browser-based (Webina[14]), deep-learning web server (GNINA web[9]), and cloud AI platform (WeMol[16]). Minimum user actions from a clean start to viewing a docked pose with 2D interaction diagram were counted against each platform’s documented interface as verified in February 2026. Anyone Can Dock requires just four steps, fewer steps than the other available tools (Table 2).

Table 2

Feature comparison of accessible molecular docking platforms. “Steps to first result” counts discrete user actions from a clean start to viewing a docked pose with 2D interaction diagram. “Yes (Colab)” denotes a feature available via the *Anyone Can Dock* Google Colab notebook rather than the Streamlit web interface. Bold entries in the *Anyone Can Dock* column indicate features not available in combination across other freely accessible platforms, to the best of our knowledge, among accessible docking tools.

Feature	**Vina CLI**	**PyRx**	**Swiss Dock**	**CB-Dock2**	**Mol Moda**	**Docking App**	**Webina**	**GNINA web**	**Anyone Can Dock**	**WeMol**
Steps to first result	18	14	8	5	8	12	6	7	**4**	8
Installation required	✓	✓	–	–	–	✓	–	–	–	–
Account / login required	–	–	–	–	–	–	–	✓	–	✓
Local deployment option	✓	✓	–	–	–	✓	–	–	✓	–
Mobile browser support	–	–	~	~	~	–	~	~	✓	✓
Metal-safe receptor prep	–	–	–	–	–	–	–	–	✓	–
pH-aware protonation	*Man.*	–	~	–	~	*Man.*	–	–	✓	~
Bond-order correction	–	–	–	–	–	–	–	–	✓	–
Blind docking	–	~	~	✓	–	–	–	–	~	~
Multiple engines	–	–	–	–	–	–	–	–	✓*	✓
CNN / DL scoring	–	–	–	–	–	–	–	✓	✓*	✓
GPU-accelerated	–	–	–	–	–	–	–	–	✓*	*N/A*
Built-in RMSD validation	–	–	–	–	–	–	–	–	✓	–
Interactive 3D viewer	–	✓	✓	✓	✓	✓	✓	–	✓	✓
2D interaction diagram	–	–	–	–	–	–	–	–	✓	*N/A*
Batch virtual screening	*Script.*	✓	~	–	✓	–	✓	~	✓	✓

Symbols

✓ supported – not supported ~ partial, indirect, or unconfirmed support ✓* backend/module-dependent Man. manual Script. Scripted. Data accessed on 17 May 2026.

Under the counting methodology used here, the Streamlit interface of Anyone Can Dock required four discrete user actions from a clean start to a first docked pose with 2D interaction diagram. We emphasize that “steps to first result” is a workflow-design metric and depends on the counting rules applied; the value should be interpreted in that light rather than as an absolute measure of usability.

3.4 Optional natural-language companion (Anyone Can Dock GPT)

To support workflow guidance and result interpretation for non-specialist users, we additionally provide an optional custom large-language-model assistant, *Anyone Can Dock GPT*, hosted as a public custom GPT on a third-party platform. This companion is not a docking engine and is not required to use the open-source platform. It is intended solely as an interactive help layer, providing natural-language guidance for setup and human-readable summaries of docking output. Because Anyone Can Dock GPT is hosted on an external conversational-AI platform, its availability and access requirements are independent of the open-source Streamlit and Google Colab implementations. All workflows and case studies reported here are fully reproducible through the Streamlit application or Google Colab notebook without using Anyone Can Dock GPT.

3.5 Use of large language models

A large language model was used to assist in drafting the EGFR mechanistic narrative reported above, with the receptor identity, ligand identity, predicted affinity, key contact residues, and RMSD provided as input. The model used was Anthropic Claude, accessed on 25 April 2026 via https://claude.ai. All AI-assisted text was reviewed and approved by the authors, who take full responsibility for the content. Large language models do not satisfy the authorship criteria of *Journal of Cheminformatics* and are not listed as authors of this manuscript.

Limitations

Anyone Can Dock employs rigid-receptor docking consistent with the default behavior of AutoDock Vina. Protein flexibility, induced-fit effects, and conformational selection are not explicitly modeled; users working with targets known to undergo substantial conformational change should consider ensemble docking or flexible-residue approaches. The ligand preparation workflow assigns a single recommended protonation state for docking, whereas extended tautomer and stereoisomer enumeration workflows are available through the companion tool pKaNET Cloud+. AutoDock Vina binding affinities are empirical scoring-function estimates and should not be interpreted as experimental K_i or K_d values or compared quantitatively across structurally unrelated scaffolds.

The current implementation is primarily intended for standard non-covalent docking of drug-like small molecules. The platform does not support covalent docking, explicit-solvent methods, or quantum mechanical treatment of metal coordination. Although the metal-preserving protocol re-injects metal ions with appropriate formal charges after receptor preparation, it does not model explicit coordination geometry or polarization effects.

The automated preparation pipeline does not resolve alternate conformations, insertion codes, or missing residues in experimental structures. In addition, the Streamlit deployment operates in a shared cloud computing environment and is therefore intended primarily for small-scale or educational workflows. For practical usage, batch docking is recommended at no more than approximately 20 ligands per run on the public Streamlit deployment, whereas larger screening workflows are better suited to the Google Colab or local deployment versions (Supporting Information).

4. Conclusion

Anyone Can Dock transforms protein–ligand docking into a browser-based workflow designed to reduce the technical barriers associated with conventional docking pipelines. The platform integrates receptor preparation, ligand preparation, molecular docking, bond-order-corrected pose export, RMSD-based redocking assessment, and interactive result visualization within a single accessible interface. The workflow additionally supports the preservation of metal-containing receptor structures during preparation, the automated post-processing of docked poses for chemically interpretable visualization, and an optional built-in RMSD-based redocking analysis. The ligand preparation pipeline performs automated pH-aware protonation-state assignment, 3D conformer generation, and force-field minimization with minimal user intervention.

Anyone Can Dock serves as a workflow integration platform built around established docking methodologies. The platform supports deployment through Streamlit Community Cloud, Google Colab runtime environments, and user-operated local workstations, enabling usage across educational, exploratory, and small-scale virtual-screening workflows. Anyone Can Dock GPT further extends accessibility by providing conversational workflow guidance and docking-result interpretation through natural language interaction. Going forward, future development will focus on expanded benchmarking, additional docking engines and scoring functions, and broader validation across diverse protein families and ligand classes.

Declarations

Availability and requirements

- **Project name**: Anyone Can Dock
- **Project home page**: https://nyelidl.github.io/anyone-docking/
- **Source code**: https://github.com/nyelidl/anyone-docking
- **Archived version**: DOI:10.5281/zenodo.20300481
- **Operating system(s)**: Platform independent (Linux, macOS, Windows; Streamlit Community Cloud, Google Colab)
- **Programming language**: Python 3.10+
- **Other requirements**: OpenBabel ≥ 3.1, RDKit ≥ 2023, AutoDock Vina 1.2.7 (auto-downloaded on first run)
- **License**: MIT
- **Any restrictions to use by non-academics**: None

Supporting Information

The Supporting Information is available free of charge.

Funding

K.H. acknowledges JSPS KAKENHI (Grant Numbers 21H05269 and 24K20888) and Toyota Riken Scholar Phase 1. T.B. acknowledges Mahidol University (Fundamental Fund: fiscal year 2025 by the National Science Research and Innovation Fund (NSRF), grant number FF-077/2568). Y.S. acknowledges JSPS KAKENHI (Grant Numbers 23H04879 and 23H02427), JST CREST (Grant No. JPMJCR20B3), and the Center for Quantum Information and Life Science (QiLS) at the University of Tsukuba. We gratefully acknowledge the Ministry of Higher Education, Science, Research

and Innovation of Thailand (MHESI) and the American Chemical Society for supporting the Open Access publication fee under their institutional agreement. The funding support was approved for Order No. 10001106675.

Author Contribution

K.H.: Conceptualization, Investigation, Methodology, Data curation, Validation, Visualization, Writing—original draft, Funding acquisition, Project administration. L.D.: Software, Program assembly. T.B.: Conceptualization, Writing—review and editing. Y.S.: Funding acquisition, Supervision, Resources, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Acknowledgement

K.H. acknowledges JSPS KAKENHI (Grant Numbers 21H05269 and 24K20888) and Toyota Riken Scholar Phase 1. T.B. acknowledges Mahidol University (Fundamental Fund: fiscal year 2025 by the National Science Research and Innovation Fund (NSRF), grant number FF-077/2568). Y.S. acknowledges JSPS KAKENHI (Grant Numbers 23H04879 and 23H02427), JST CREST (Grant No. JPMJCR20B3), and the Center for Quantum Information and Life Science (QiLS) at the University of Tsukuba. We gratefully acknowledge the Ministry of Higher Education, Science, Research and Innovation of Thailand (MHESI) and the American Chemical Society for supporting the Open Access publication fee under their institutional agreement. The funding support was approved for Order No. 10001106675.

Data Availability

All data supporting the findings of this study are available within the manuscript and its Supplementary Information. The docking examples are reproducible using the PDB identifiers and ligand SMILES strings provided in the manuscript and Supplementary Information. The source code for Anyone Can Dock is publicly available in the GitHub repository nyelidl/anyone-docking under the MIT license, and the archived version is available on Zenodo under DOI: 10.5281/zenodo.20300481. The hosted web application is available at https://nyelidl.github.io/anyone-docking/, and the companion pKaNET Cloud+ tool is available at https://github.com/nyelidl/pKaNET_Cloud.

References

1. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161(2):269–288
2. Todsaporn D, Sanachai K, Aonbangkhen C, Geronikaki A, Kartsev V, Lichitsky B, Komogortsev A, Maitarad P, Rungrotmongkol T (2025) Exploring novel furochochicine derivatives as promising JAK2 inhibitors in HeLa cells: Integrating docking, QSAR-ML, MD simulations, and experiments. Comput Struct Biotechnol J 27:3625–3639
3. Todsaporn D, Sanachai K, Aonbangkhen C, Poo-Arporn RP, Kartsev V, Pukhov S, Afanasyeva S, Geronikaki A, Rungrotmongkol T (2025) Identification of Novel JAK2 Inhibitors from Amino Derivatives of Epoxyalantolactone: In Silico and In Vitro Studies. Int J Mol Sci 27(1)
4. Nutho B, Wilasluck P, Deetanya P, Wangkanont K, Arsakhant P, Saeeng R, Rungrotmongkol T (2022) Discovery of C-12 dithiocarbamate andrographolide analogues as inhibitors of SARS-CoV-2 main protease: In vitro and in silico studies. Comput Struct Biotechnol J 20:2784–2797
5. Kerdsiri J, Hengphasatporn K, Pitaksuteepong T, Nuengchamnong N, Suroengrit A, Chupinidsakulwong P, Shigeta Y, Bhattarakosol P, Boonyasuppayakorn S, Waranuch N (2025) In silico and bioassay-guided identification of

potential anti-SARS-CoV-2 tentative candidate compounds from Andrographis paniculata extract. Comput Struct Biotechnol J 27:5479–5492

6. Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model 61(8):3891–3898
7. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47(7):1739–1749
8. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. J Mol Biol 267(3):727–748
9. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, Sunseri J, Koes DR (2021) GNINA 1.0: molecular docking with deep learning. J Cheminform 13(1):43
10. Dallakyan S, Olson AJ (2015) Small-molecule library screening by docking with PyRx. Methods Mol Biol 1263:243–250
11. Hsu K-C, Chen Y-F, Lin S-R, Yang J-M (2011) iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. BMC Bioinformatics 12(Suppl 1):S33–S33
12. Yu Y, Cai C, Wang J, Bo Z, Zhu Z, Zheng H (2023) Uni-Dock: GPU-Accelerated Docking Enables Ultralarge Virtual Screening. J Chem Theory Comput 19(11):3336–3345
13. Hakkennes MLA, Buda F, Bonnet S (2023) MetalDock: An Open Access Docking Tool for Easy and Reproducible Docking of Metal Complexes. J Chem Inf Model 63(24):7816–7825
14. Kochnev Y, Hellemann E, Cassidy KC, Durrant JD (2020) Webina: an open-source library and web app that runs AutoDock Vina entirely in the web browser. Bioinformatics 36(16):4513–4515
15. Tao A, Huang Y, Shinohara Y, Caylor ML, Pashikanti S, Xu D (2019) ezCADD: A Rapid 2D/3D Visualization-Enabled Web Modeling Environment for Democratizing Computer-Aided Drug Design. J Chem Inf Model 59(1):18–24
16. Liu H, Yan X, Fang H, Ge H, Hou X (2026) WeMol: A Cloud-Based and Zero-Code Platform for AI-Driven Molecular Design and Simulation. J Chem Inf Model 66(4):2369–2381
17. Hengphasatporn K, Duan L, Harada R, Shigeta Y (2026) DFDD: A Cloud-Ready Tool for Distance-Guided Fully Dynamic Docking in Host–Guest Complexation. J Chem Inf Model 66(4):1955–1963
18. Bisong E (2019) Google Colaboratory. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners.* Edited by Bisong E. Berkeley, CA: Apress; : 59–64
19. Bakan A, Meireles LM, Bahar I (2011) ProDy: protein dynamics inferred from theory and experiments. Bioinformatics 27(11):1575–1577
20. Ropp PJ, Kaminsky JC, Yablonski S, Durrant JD (2019) Dimorphite-DL: an open-source program for enumerating the ionization states of drug-like small molecules. J Cheminform 11(1):14
21. Riniker S, Landrum GA (2015) Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. J Chem Inf Model 55(12):2562–2574
22. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17(5–6):490–519
23. Santos-Martins D, He Y, Eberhardt J, Sharma P, Bruciaferri N, Holcomb M, Llanos MA, Hansel-Harris A, Barkdull AP, Tillack AF et al (2025) Meeko: Molecule Parametrization and Software Interoperability for Docking and Beyond. J Chem Inf Model 65(24):13045–13050
24. Sipos-Szabó L, Bajusz D, Balogh GT, Keserű GM (2026) Benchmarking pKa Prediction Algorithms against an Extensive, Public Data Set. J Chem Inf Model 66(8):4607–4619

25. Koebel MR, Schmadeke G, Posner RG, Sirimulla S (2016) AutoDock VinaXB: implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina. J Cheminform 8(1):27
26. Bento AP, Hersey A, Félix E, Landrum G, Gaulton A, Atkinson F, Bellis LJ, De Veij M, Leach AR (2020) An open source chemical structure curation pipeline using RDKit. J Cheminform 12(1):51
27. Rego N, Koes D (2015) 3Dmol.js: molecular visualization with WebGL
28. Stierand K, Rarey M (2010) PoseView -- molecular interaction patterns at a glance. J Cheminform 2(1):P50
29. Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, Zou J (2024) ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. Bioinformatics 40(7):btae416
30. Bugnon M, Röhrig UF, Goullieux M, Perez Marta AS, Daina A, Michielin O, Zoete V (2024) SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. Nucleic Acids Res 52(W1):W324–W332
31. Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y (2022) CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. Nucleic Acids Res 50(W1):W159–w164
32. Kochnev Y, Ahmed M, Maldonado AM, Durrant JD (2024) MolModa: accessible and secure molecular docking in a web browser. Nucleic Acids Res 52(W1):W498–w506
33. Macari G, Toti D, Pasquadibisceglie A, Polticelli F (2020) DockingApp RF: A State-of-the-Art Novel Scoring Function for Molecular Docking in a User-Friendly Interface to AutoDock Vina. *Int J Mol Sci* 21(24)

Figures

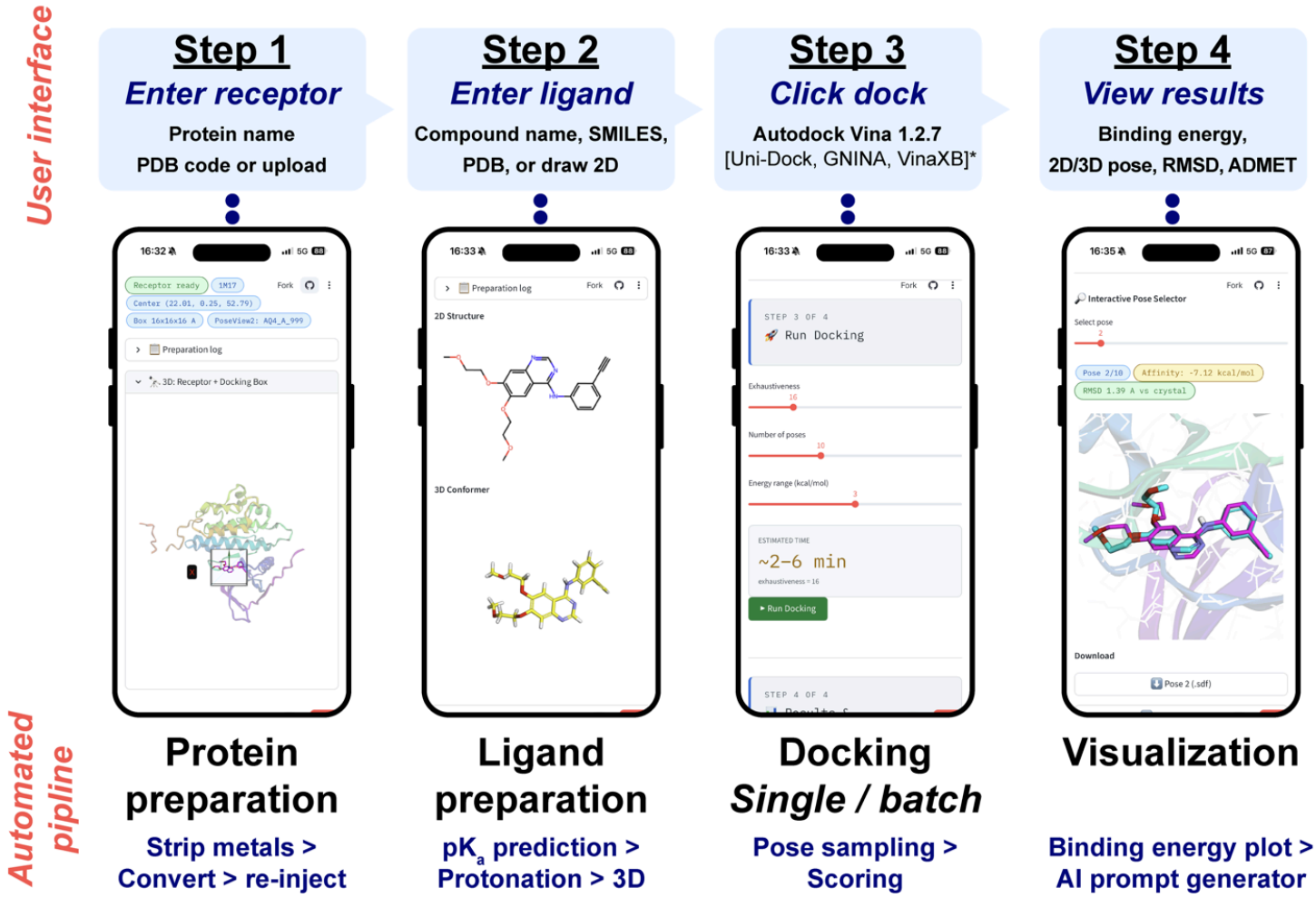


Figure 1

Overall workflow of Anyone Can Dock in Streamlit interface. The four-step user interface, receptor input, ligand input, docking, and results, is presented alongside the computational pipeline executed automatically at each stage, including metal-safe PDBQT preparation, pH-dependent protonation recommendation with pKaNET Cloud+, docking engine execution, bond-order correction, and integrated RMSD validation.

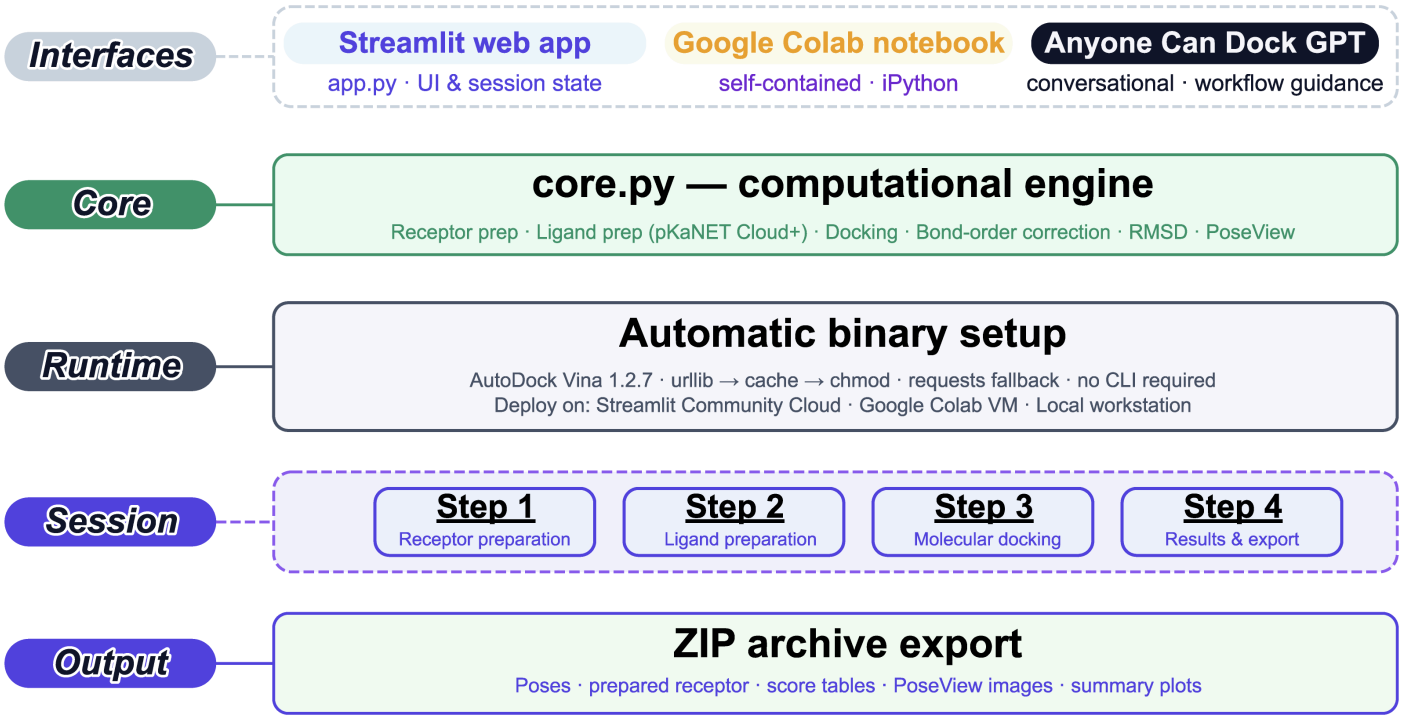


Figure 2

Software architecture of Anyone Can Dock. Three independent interfaces, a Streamlit web application, a Google Colab notebook, and Anyone Can Dock GPT, share a common computational engine (core.py) that handles receptor preparation, ligand protonation-state assignment via pKaNET, docking execution, bond-order correction, and RMSD-based validation. An automatic binary setup layer downloads and configures AutoDock Vina 1.2.7 without command-line installation across Streamlit Community Cloud, Google Colab virtual machines, and local workstations. Session outputs are bundled into a ZIP archive for download.

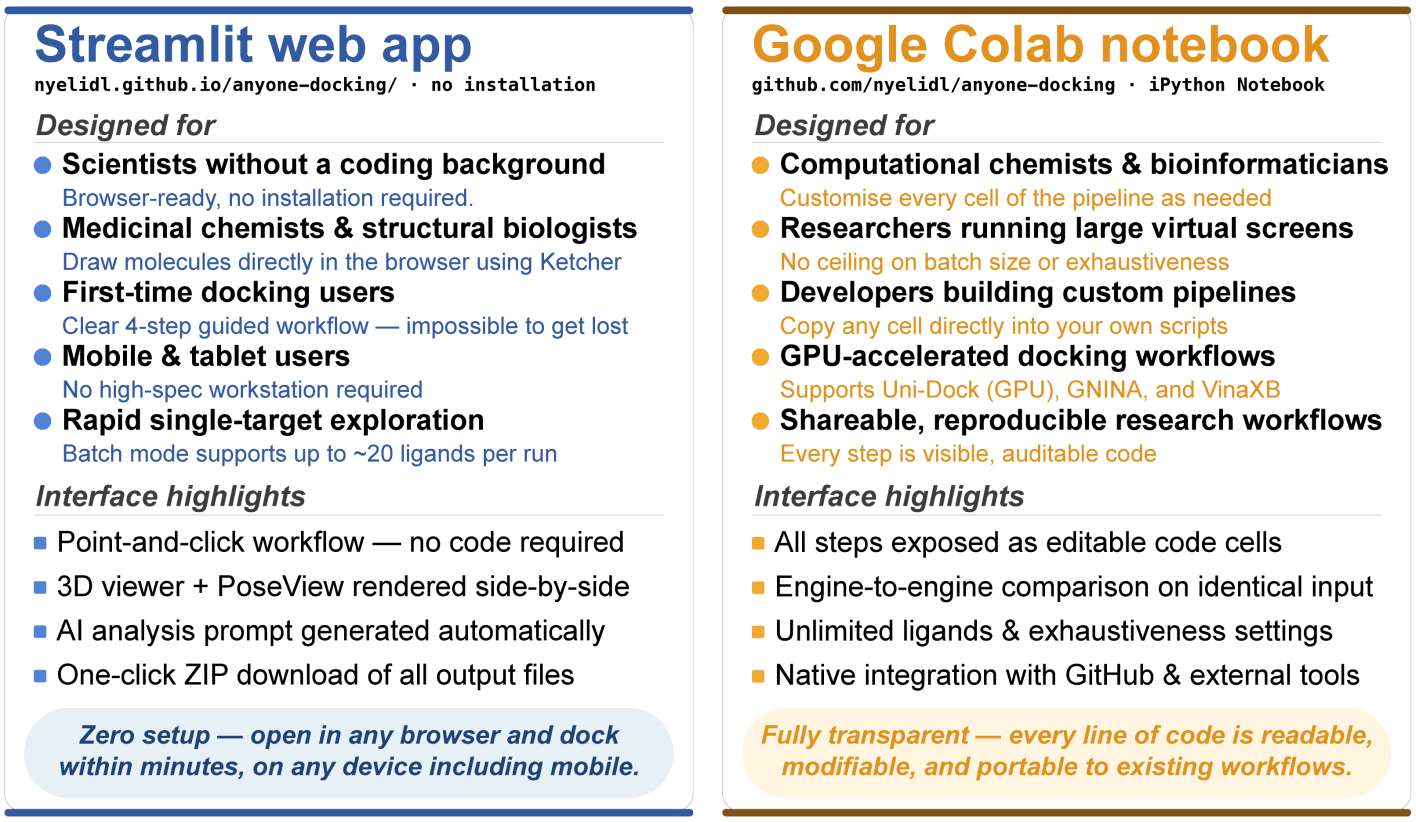


Figure 3

Complementary user interfaces of Anyone Can Dock. The platform provides two interfaces tailored to different user groups. The Streamlit web application supports non-specialist users through a guided four-step workflow with browser-based ligand drawing, integrated 3D and 2D interaction diagram visualization, AI-generated analysis prompts, ZIP export, and small-scale batch docking. The Google Colab notebook targets advanced users by exposing the full workflow as editable Python code, enabling unrestricted parameter control, custom pipeline development, access to multiple docking engines (Uni-Dock, GNINA, and VinaXB), and direct comparison of docking engines on identical inputs.

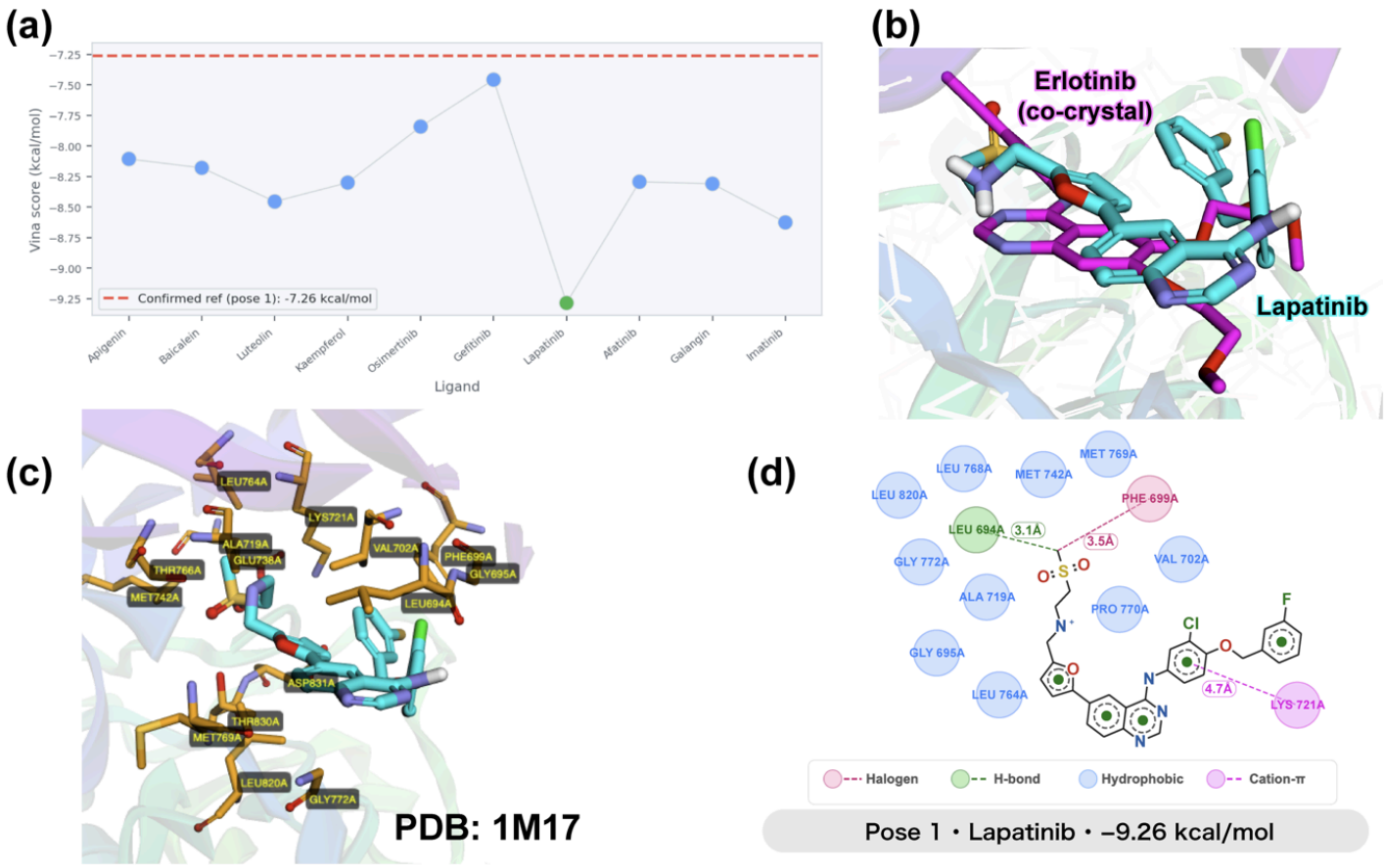


Figure 4

Batch docking of ten ligands into the EGFR kinase domain (PDB: 1M17) using the Streamlit interface, with erlotinib redocked as co-crystal reference. (a) Dot-plot of top-ranked AutoDock Vina binding affinities across all screened ligands, with a dashed reference line at the confirmed erlotinib redocking affinity. (b) Three-dimensional overlay of the top-ranked lapatinib pose against the co-crystallised erlotinib in the ATP-binding cleft. (c) Binding pocket view showing the docked ligand surrounded by key contact residues obtained from single-ligand docking. (d) Anyone Can Dock 2D protein–ligand interaction diagram showing hinge-region hydrogen bonds and hydrophobic contacts.

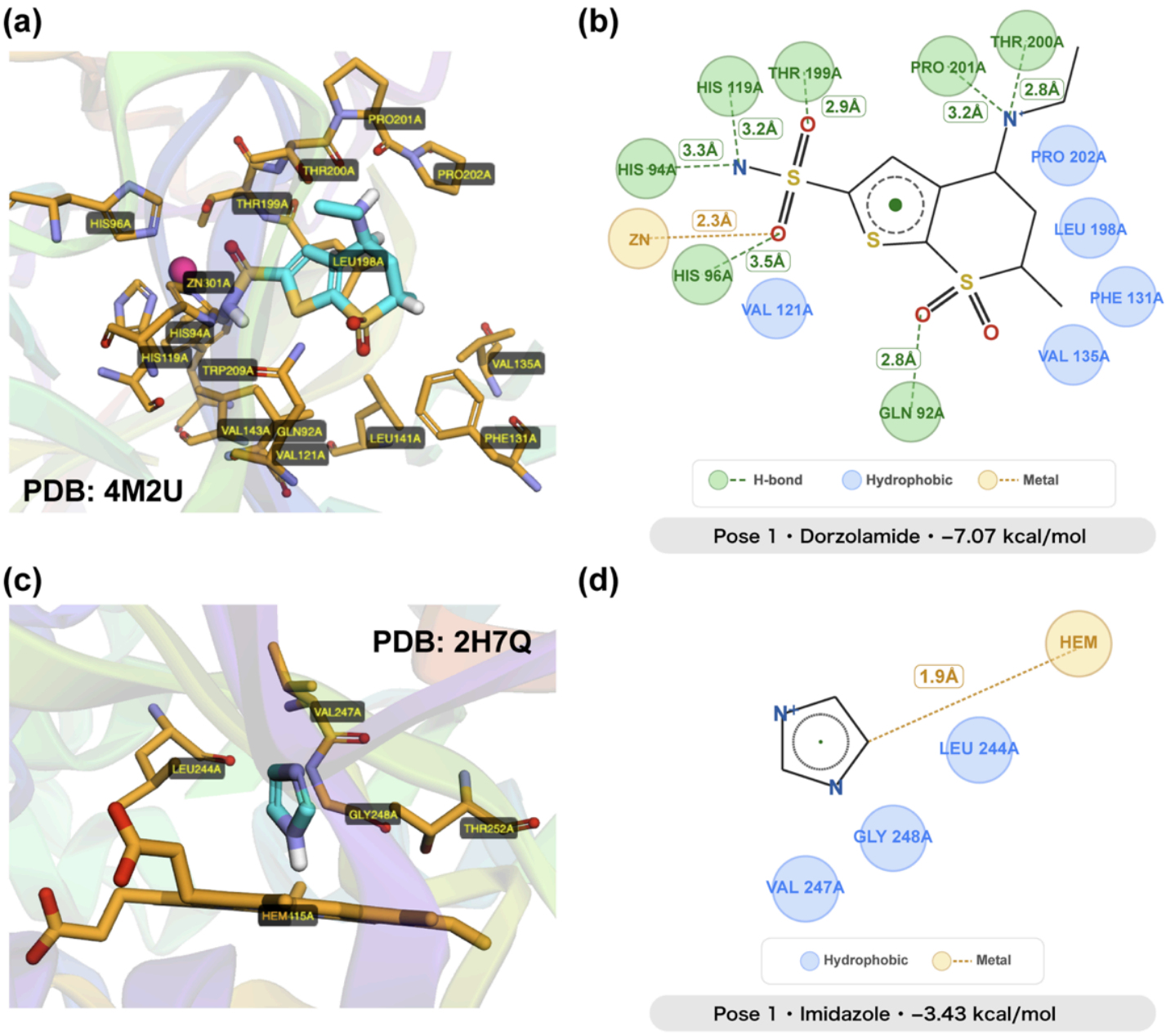


Figure 5

Docked poses and RDKit-generated 2D interaction diagrams of dorzolamide in carbonic anhydrase II (PDB: 4M2U) and imidazole in cytochrome P450 (PDB: 2H7Q), highlighting zinc/heme coordination and key active-site interactions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- floatimage1.png
- AnyoneCanDockSIJchem.pdf