

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	The JKlustor module of software ChemAxon was employed to conduct Bemis-Murcko clustering for active molecules identification;
Data analysis	Six AI-based tools and four physics-based tools were employed for protein-ligand docking and scoring, including CarsiDock, KarmaDock, FlexPose, DiffDock, RTMScore, EquiScore, Glide (schrodinger v2022-1), LeDock (v1.0), Surflex (v2.706), rDock (v2021.1). Additionally, in re-docking experiments, the data analysis was conducted by Posebuster package (v 0.3.1); In virtual screening experiments, the data analysis was conducted by oddt package (v 0.7); The post-processing method was based on rdkit package; The code can be obtained at https://github.com/shukai1997/VSDS-VD/ ;

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw data set is collected from <https://www.bindingdb.org/rwd/bind/index.jsp>, <https://www.ebi.ac.uk/chembl/>, and <https://pubchem.ncbi.nlm.nih.gov/>. The prepared benchmarking set is available at <https://zenodo.org/records/13684010>. All structures are available in the Protein Data Bank (<https://www.rcsb.org/>). All target information can be found at <https://www.uniprot.org/>. Source data are available with this paper.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The TrueDecoy set contains 147 targets, with no sample-size calculation performed. Its size was determined by the quantity of active and inactive compounds available in the database that fulfill the requirements. Its validity stems from being a benchmark for virtual screening, as it covers multiple protein families and is currently one of the datasets with the highest number of targets. The RandomDecoy set contains 68 targets, partially derived from TrueDecoy set as a VS benchmark. Its justification lies in the fact that it also encompasses multiple protein families; the number of targets is within a reasonable range compared to previous virtual screening datasets. Furthermore, randomly selecting 68 targets from TrueDecoy set multiple times yielded consistent conclusions, indicating that the sample size in the RandomDecoy set is sufficient. The MassiveDecoy set contains eight targets, intended to simulate a realistic large-scale virtual screening setting. No sample-size calculation was performed, and it is primarily presented in a case format, which also supports its validity.

Data exclusions

Compounds in the database with vague label was abandoned; Targets with inactive compounds fewer than 800 were eliminated;

Replication

During the construction of RandomDecoy set, the decoy molecules were randomly selected from the ChemDiv database. This selection process was performed three times, and the results were consistent across all iterations. All attempts at replication were successful.

Randomization

The covariates are different virtual screening methods. Different virtual screening methods works on the same virtual screening datasets.

Blinding

The study focuses on virtual screening, so it is not related to blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging