

Corresponding author(s): Wanxiang Shen, Marinka Zitnik

Last updated by author(s): Jun 6, 2026

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

N/A

Data analysis

Codes used in this study are available from the corresponding author upon request. In addition, the source code and Jupyter notebooks are available from Zenodo (<https://doi.org/10.5281/zenodo.20566775>), including PyTorch and TensorFlow 2.x implementations of the ACA loss. A Molecular Property Cliff Task has been introduced to the Therapeutics Data Commons (<https://tdcommons.ai/>), allowing programmatic access to the data used in the paper. A tutorial has been included in the TDC-2 Tutorials (<https://github.com/mims-harvard/TDC/blob/main/README.md>).

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 data that support the findings of this study, including the 9 small-sample size narrow scaffold (LSSNS) datasets, the 30 large-sample size mixed scaffold (HSSMS)

datasets, the 3 MMP datasets and 10 ADEMT datasets are released at <https://github.com/bidd-group/MPCD>, are available from Zenodo (<https://doi.org/10.5281/zenodo.20566671>) and from the corresponding authors upon request. In addition, datasets are being made available via the Molecular Property Cliff Task (https://tdcommons.ai/single_pred_tasks/MPC/) in Therapeutics Data Commons (<https://tdcommons.ai>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	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

Life sciences study design

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

Sample size	Molecular property prediction datasets. We consider a total of 52 datasets, including 9 LSSNS activity regression datasets, 30 HSSMS activity regression datasets, 3 MMP activity cliff classification datasets, and 10 ADMET delta regression datasets (Figure 2). The 30 HSSMS benchmark datasets (Supplementary Table S2) were curated by van Tilborget al. (2022) using data from ChEMBL. It serves as a benchmark for exposing the limitations of molecular machine learning with activity cliffs (ACs). The HSSMS consists of 30 mixed-scaffold datasets totaling 48,707 molecules. The HSSMS datasets cover a wider range of sample sizes (from 615 to 3,657 compounds), scaffold diversity, bioassay types, and target families. The 9 LSSNS datasets (Supplementary Table S1, Supplementary Figure S1a) contain small molecules with relatively conserved scaffold (similar structures) but varying activity labels, with sample sizes ranging from less than 50 to over 100 compounds (3 sets are less than 50 compounds, three sets are in 50-100, 3 sets are greater than 100 compounds). The LSSNS datasets were collected from fragment-to-lead medicinal chemistry publications in 2019 and the ChEMBL database. The 3 MMP benchmark datasets (Supplementary Table S3) were curated by Park et al. (2022) for AC classification, focusing on the targets thrombin, Mu opioid receptor, and melanocortin receptor 4. The binary AC labels of MMPs are defined as MMP-cliff (AC) if $\Delta pKi \geq 2$, and MMP-nonCliff (non-AC) if $\Delta pKi \leq 1$. The 10 ADMET benchmark datasets (Supplementary Table S4) were collected and preprocessed by Fralish et al. (2023) for delta prediction of molecular ADMET properties.
Data exclusions	Online mining of the activity cliff triplets. In this study, we employed online mining of activity cliff triplets (ACTs) for model contrastive learning in the latent space. During the mining process, we identify a positive compound and a negative compound for each anchor compound to form an ACT. The positive compound has an activity value close to the anchor, while the negative compound exhibits significantly different activity. To facilitate real-time conditional ACT sampling in each batch, we introduce "cliff lower" (cl) and "cliff upper" (cu) as cliff cut-off hyperparameters (Figure 1b). By comparing the activity difference between compounds and the anchor, we categorize compounds as positive examples if the difference is within the "cliff lower" threshold (cl) and as negative examples if the difference exceeds the "cliff upper" threshold (cu). When cl is equal to cu, the triplets are obtained using a single "cliff" hyperparameter, maximizing the number of ACTs acquired. It is important to note that cl should be smaller than cu, and leaving a suitable margin ensures the obtained triplets are reasonable. To obtain highly valuable ACTs (HV-ACTs) in each batch, only an ACT with its triplet loss greater than zero is counted (Figure 1b). In this setting, anchor–Negative (A–N) pairs are typically close enough in the latent space, that is, structurally similar, but at the same time differ greatly in activity, representing ACs that need to be pushed apart, whereas anchor–Positive (A–P) pairs are functionally consistent and should be pulled closer. This ensures that the selected triplets are informative and contribute effectively to the model's learning. The number of mined HV-ACTs (denoted as M') reflects the degree of ACs in the current stage of molecular representations. If the model is already ACA during training, the number of mined HV triplets should gradually decrease as training progresses. This reduction occurs because the ACA of the model optimizes the latent feature space to reduce prediction errors. Gradually decreasing the number of mined HV-ACTs helps the model focus on refining the latent feature space for compounds that are more challenging to predict accurately.
Replication	N/A

Randomization

Molecules within the HSSMS are categorized as either “cliff molecules” or not. Specifically, the structural similarity of ACs in HSSMS is determined by three string-based methods: substructure similarity, scaffold similarity, and SMILES similarity. A pair of molecules is considered structurally similar if any of these similarities are greater than 0.9. The activity difference criterion is defined as a 10-fold difference in potency. A molecule is classified as a cliff molecule if it can find at least one molecule in the dataset that satisfies both the structural similarity criterion and has a potency difference greater than 10-fold. To ensure both similar structural and label distributions between the training and test sets, molecules are first spectrally clustered based on their ECFP fingerprints. Within each cluster, an 80/20 split is performed using stratified sampling based on the cliff tag for training and testing, respectively.

Blinding

N/A

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	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A