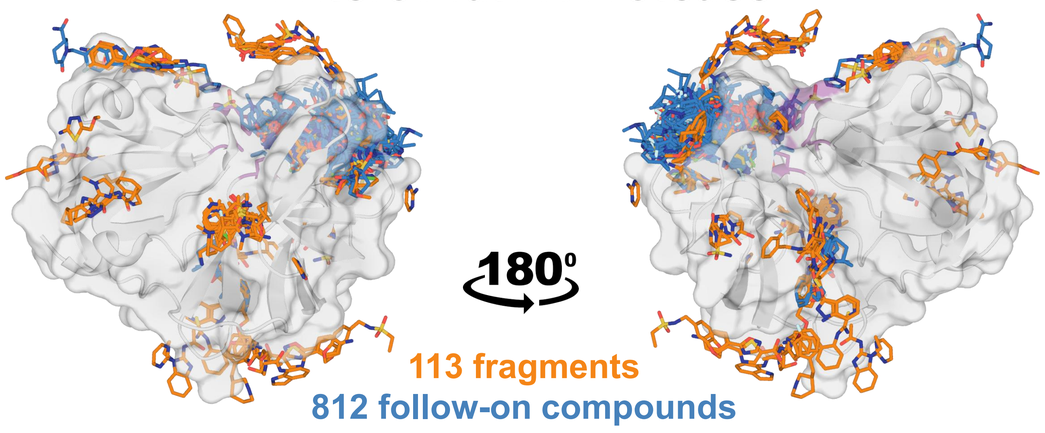
The first OpenBind release: An open experimental structure–affinity dataset and benchmark for structure-based AI

Jochem Nelen, O. Khan, E. Adams, J. C. Aschenbrenner, W. Thompson, A. Ebrahim, E. Çapkin, C. Vallée, OpenBind, E. J. Shotton, E. J. Griffen, J. D. Chodera, C. M. Deane, F. von Delft, M. AlQuraishi, and F. Imrie

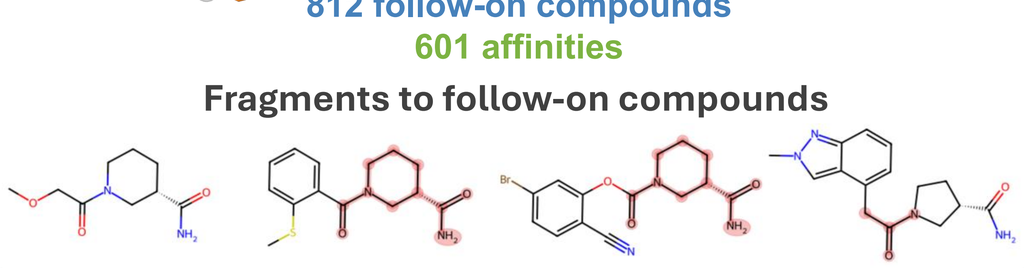
1. A coherent experimental structure–affinity dataset

Enteroviral 2A Protease



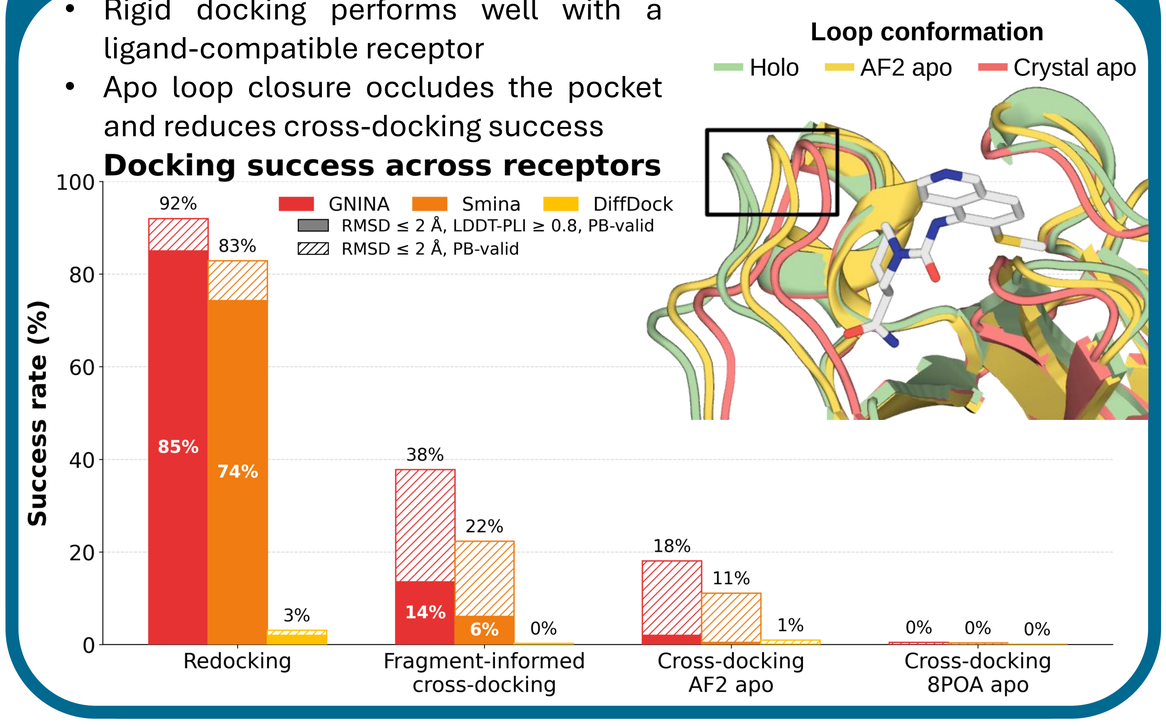
113 fragments
812 follow-on compounds
601 affinities

Fragments to follow-on compounds

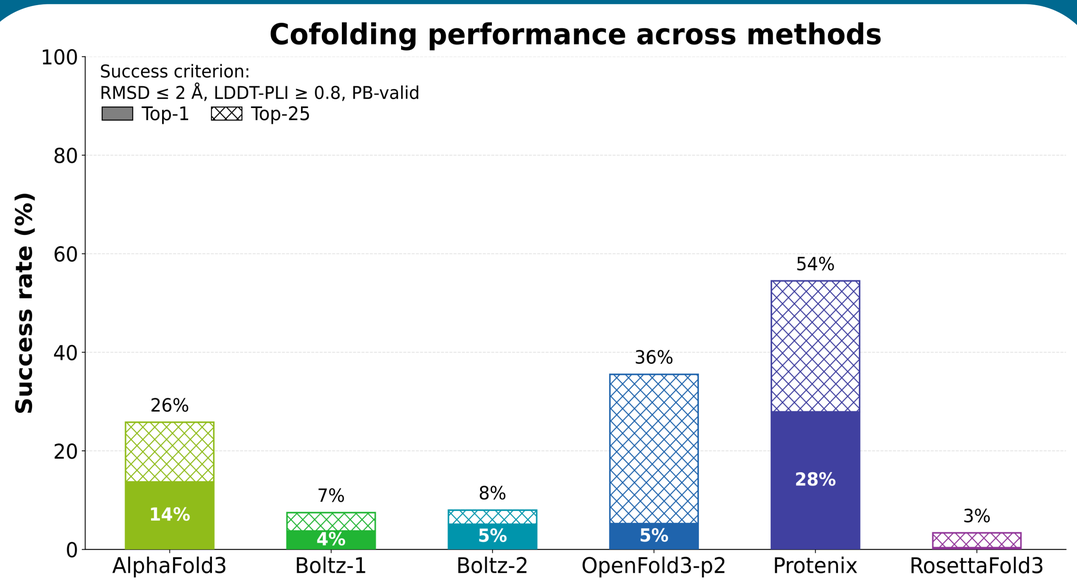


2. Receptor state influences docking performance

- Rigid docking performs well with a ligand-compatible receptor
- Apo loop closure occludes the pocket and reduces cross-docking success

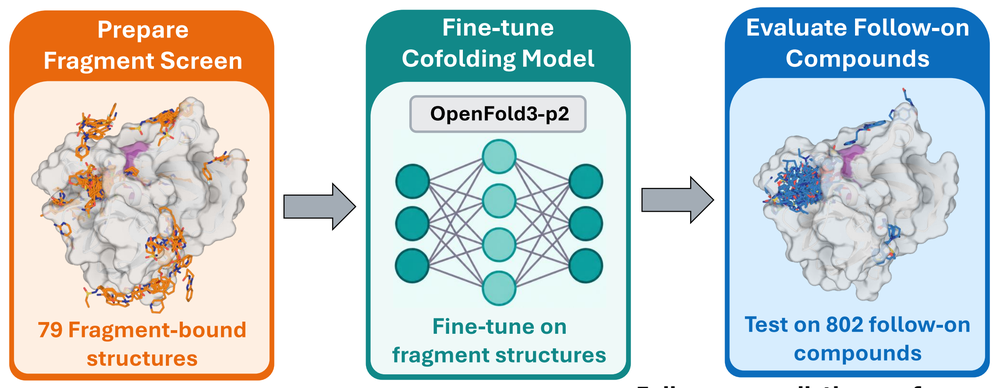


3. Pose recovery and ranking limit cofolding performance

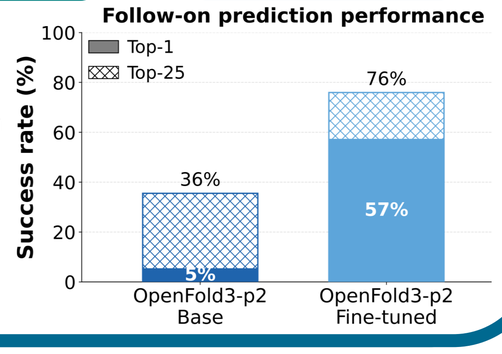


- Performance varies substantially across current cofolding methods
- Top-25 success rate remains limited, with pose ranking further constraining performance

4. Learning from fragment-bound structures



- 79 curated fragment-bound structures were used for target-specific fine-tuning
- Fine-tuning increased Top-1 success from 5% to 57% on 802 curated follow-on complexes
- For this dataset, experimental fragment data provided a strong target-specific signal for follow-on pose prediction



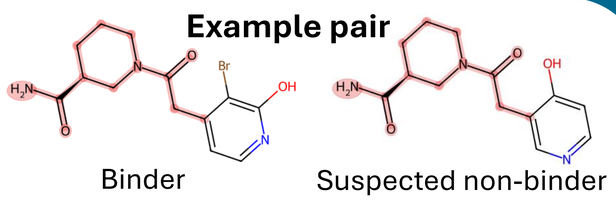
5. Virtual screening among related follow-on compounds

Benchmark composition

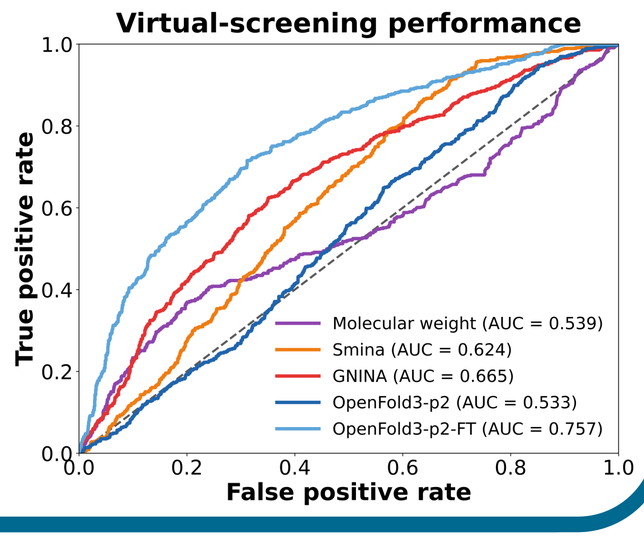
2,084 compounds

566 binders | **1,518 suspected non-binders**

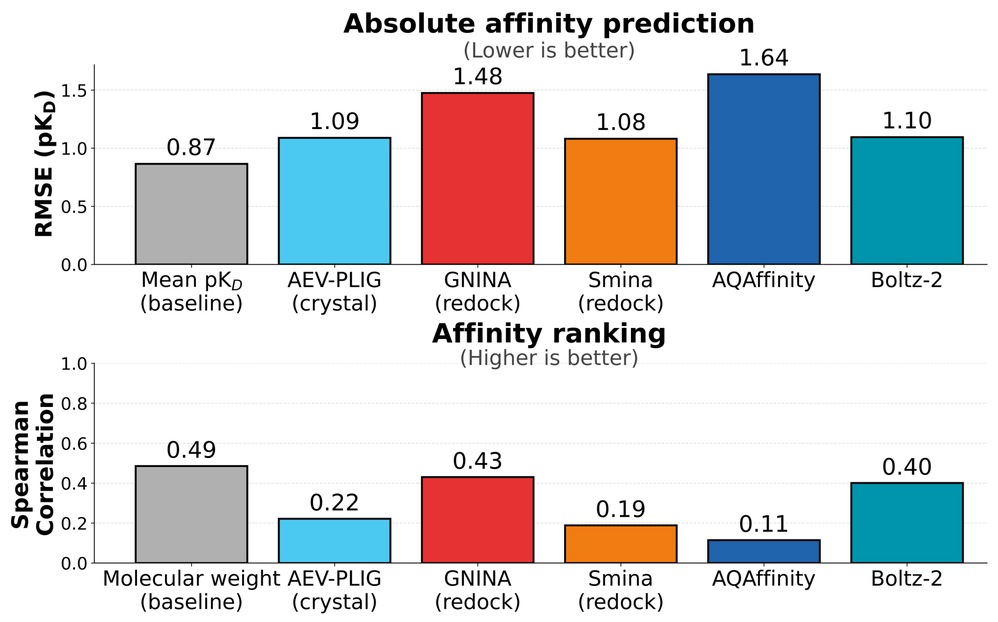
Example pair



- Suspected non-binders showed no crystallographic binding despite a successful soak
- Many binders have close analogues among suspected non-binders
- Fine-tuning increased OpenFold3-p2 ROC-AUC from 0.533 to 0.757, moving it from worst to best performer



6. Structure-based affinity prediction is challenging



- Absolute affinity prediction and ranking are weak across tested methods
- Simple baselines outperform the evaluated structure-based methods

7. Summary and conclusions

- OpenBind releases a coherent experimental structure–affinity dataset
- Docking performance depends strongly on receptor state, while cofolding remains limited by sampling and ranking
- Fragment-informed fine-tuning can substantially improve both follow-on pose prediction and virtual-screening performance
- Current methods struggle to translate structure into quantitative affinity, with simple baselines still competitive

Affiliations and funding

This work was funded by UK DSIT, grant G2-SCH-2025-06-16537

OpenBind
www.openbind.uk

Nelen J, Khan O, Adams E, *et al.*, bioRxiv (2026)
doi: 10.64898/2026.08.27.747600

Preprint