Snowdrop Science

Predicting the activity of Cu compounds from 3D structure

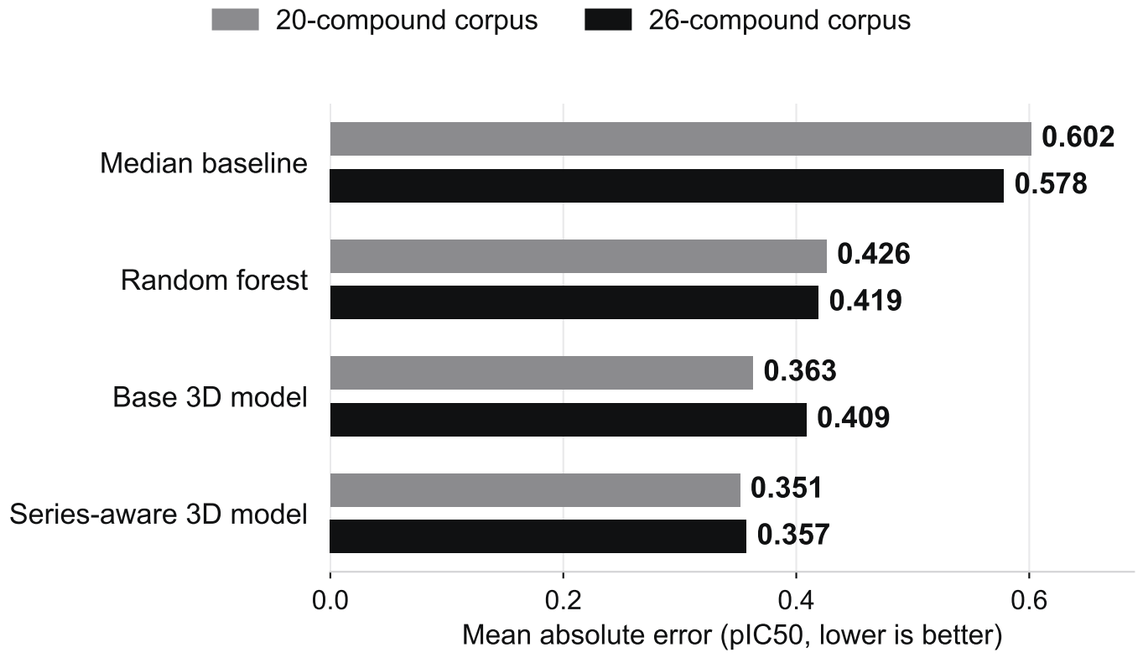
MCF-7 activity of the original 20 compounds, predicted using 15 or 21 training compounds.
The larger corpus adds six literature compounds to training in every matched test.

Ordering held-out compounds

Model	20-compound corpus	26-compound corpus
Median baseline	**All tied**	**All tied**
Random forest	**57 / 80**	**60 / 80**
Base 3D model	**63 / 80**	**59 / 80**
Series-aware 3D model	**64 / 80**	**58 / 80**

A pair is correct when its more active compound is ranked first.
Tied predictions count ½. The median baseline cannot rank.

Error on held-out IC50 predictions



Retrospective evaluation: two four-fold partitions test 20 original compounds, five at a time: 8 test sets and 40 predictions per model.
Each five-compound test set contains 10 pairs, giving 80 ordering comparisons. Predictions average three model seeds.

Sources: [1] 10.1039/D4SC06628F [2] 10.1016/j.jphotobiol.2016.11.024 [3] 10.1039/D4MD00543K [4] 10.1007/s10534-015-9820-5 [5] 10.1039/D4DT01075B
[6] 10.1002/cmdc.202100520 [7] 10.1039/C9NJ04975D [8] 10.1039/D1QI00233C [9] 10.1016/j.ica.2018.11.039 [10] 10.1016/j.poly.2021.115621

Original model predictions and figures: Snowdrop Science.

Snowdrop Science

Cu predictions beyond the observed training activity range

A selected seven-query test from the 26-compound corpus.
The 19 training compounds had measured IC50 values from 0.57 to 39.2 µM.

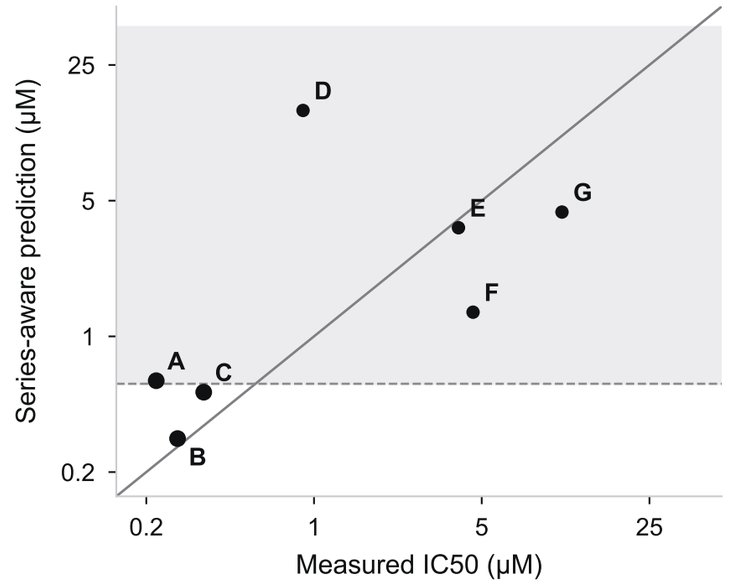
Three most potent held-out compounds: IC50, µM

Query / activity rank	Measured	Median	Random forest	Base 3D	Series-aware 3D
A / 1	0.220	6.010	1.497	0.842	0.591
B / 2	0.270	6.010	2.137	0.485	0.298
C / 3	0.347	6.010	0.821	0.851	0.516

Ranks 1, 2 and 3 are the most potent compounds in all 26.
The series-aware model predicts B and C below 0.57 µM;
A remains above this training boundary.

Full seven-query series-aware result: MAE 0.404 pIC50;
15 of 21 pairs ordered correctly. Related ligand chemistry remains in training.

All seven held-out queries



Shading: observed training IC50 range.
Diagonal: exact prediction.

This test was identified retrospectively. The table zooms in on three of its seven queries; the graph includes all seven.
It supports a limited activity-range extrapolation example, with familiar chemistry represented in training.

Sources: [1] 10.1039/D4SC06628F [2] 10.1016/j.jphotobiol.2016.11.024 [3] 10.1039/D4MD00543K [4] 10.1007/s10534-015-9820-5 [5] 10.1039/D4DT01075B
[6] 10.1002/cmdc.202100520 [7] 10.1039/C9NJ04975D [8] 10.1039/D1QI00233C [9] 10.1016/j.ica.2018.11.039 [10] 10.1016/j.poly.2021.115621

Original model predictions and figures: Snowdrop Science.

Snowdrop Science

Cu activity extremes remain difficult to predict

Two deliberate tests, each withholding both compounds at one end of the activity range.
Each test uses 24 training compounds from the 26-compound corpus.

Both strongest compounds held out

Query	Measured IC50 (µM)	Series-aware prediction (µM)
A	0.220	0.622
B	0.270	0.486

Training IC50 range: 0.347–39.2 µM

The two predictions reverse the measured order.
Both remain inside the training activity range.
All four main 3D arms reverse this strongest pair.

Series-aware MAE: 0.353 pIC50

Both weakest compounds held out

Query	Measured IC50 (µM)	Series-aware prediction (µM)
A	31.900	3.377
B	39.200	5.197

Training IC50 range: 0.22–19.35 µM

The model orders the pair correctly, but predicts substantially lower IC50 values than measured.
Both remain inside the training activity range.

Series-aware MAE: 0.926 pIC50

In the additional five- and six-query joint tests, no main 3D arm predicts the extreme queries beyond the training boundaries.
Correct ordering and occasional boundary crossings do not establish reliable prediction of unseen activity extremes.

Sources: [1] 10.1039/D4SC06628F [2] 10.1016/j.jphotobiol.2016.11.024 [3] 10.1039/D4MD00543K [4] 10.1007/s10534-015-9820-5 [5] 10.1039/D4DT01075B
[6] 10.1002/cmdc.202100520 [7] 10.1039/C9NJ04975D [8] 10.1039/D1QI00233C [9] 10.1016/j.ica.2018.11.039 [10] 10.1016/j.poly.2021.115621

Original model predictions and figures: Snowdrop Science.