Protein Structure-Informed Molecular Fragment Replacement with InteractionDB

Giulio Mattedi, Mahad Gatti Iou, Andrew Potterton and Yi Mok

Benevolent AI

INTRODUCTION

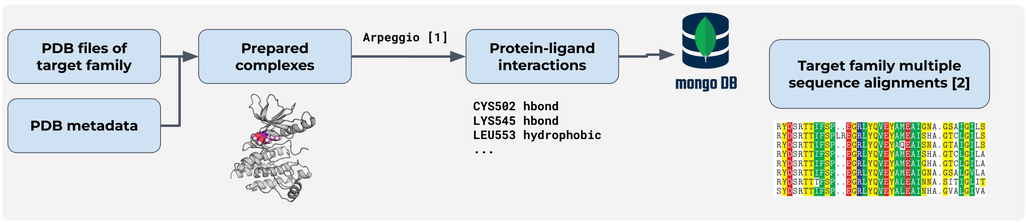
Matched molecular pair analysis is a well-established approach for generating compound ideas by studying the effects of medicinal chemistry modifications on specific endpoints. However, existing methods often produce transformations that are incompatible with the binding mode of molecules in the target protein structure.

Here, we present InteractionDB (IntDB), a protein structure-informed molecular replacement tool. InteractionDB replaces molecular fragments with groups that engage structurally equivalent contacts in experimental protein-ligand complexes in the Protein Data Bank (PDB).

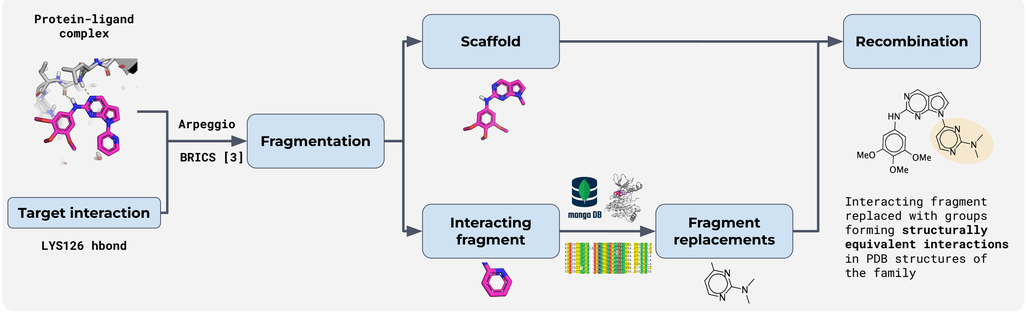
In this study, we demonstrate InteractionDB designs on a set of kinase targets. We replace molecular fragments of crystallised small-molecule ligands and compare the designs to literature data reported in ChEMBL. Our results demonstrate that InteractionDB successfully replicates molecular replacements that have been proven effective in medicinal chemistry optimisation. By considering protein structural information, InteractionDB offers a valuable strategy for generating optimised molecular designs that are compatible with the binding site of the target protein.

InteractionDB Workflow

Building

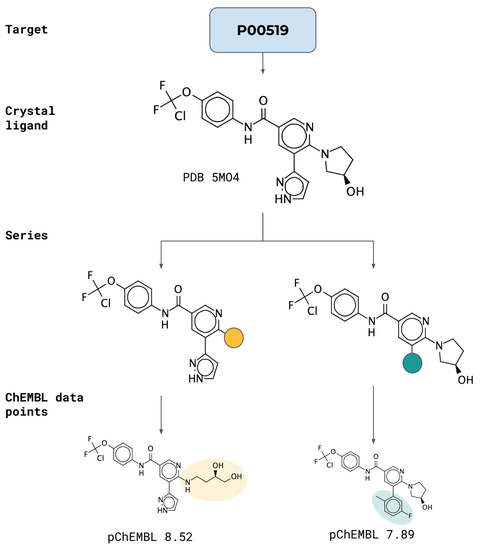


Querying



METHODS

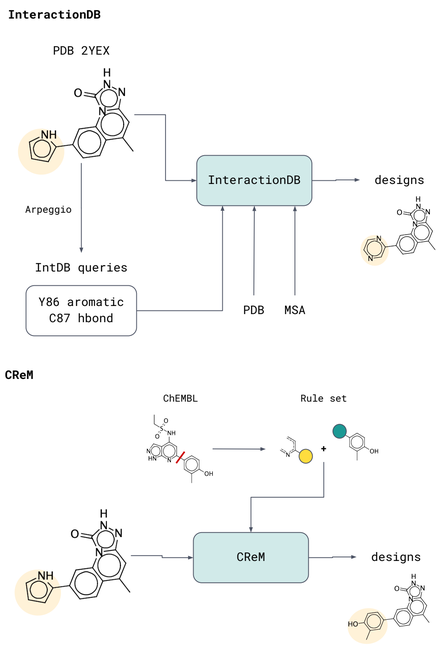
ChEMBL Data Mining



Organising experimental data points

- ChEMBL 32 was mined for pChEMBL data points of **human kinase targets**. The data was organised in a hierarchical structure, where root nodes are ligands crystallised in PDB structures
- For each crystal ligand, a number of **series** were defined as **single R-group replacements** of the parent molecule with BRICS fragmentations
- Data points were assigned to their respective series, with only series containing at least 20 data points considered for further analysis
- The dataset comprised 5140 data points, covering 42 targets, 92 crystal ligands, and 125 series

R-group Replacement



InteractionDB

- For each series, **interactions** formed by the R-group of the parent crystal ligand were used as **queries to IntDB**. Arpeggio was used to detect the interactions between the ligands and the proteins
- IntDB identifies R-group replacements from molecular fragments that form **structurally equivalent interactions** in PDB structures of other kinases according to a multiple sequence alignment (**MSA**) of the human kinome

Baseline

- **Chemically Reasonable Mutations** (CReM)[4] was used as a baseline method. CReM learns connection rules from a set of compounds and enumerates them on query molecules

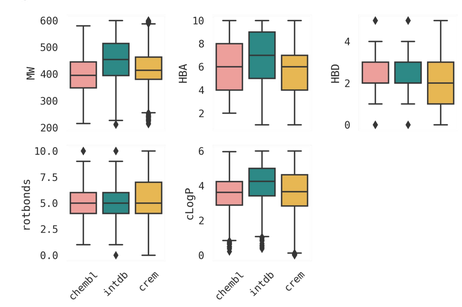
Filtering

- ChEMBL data points and designs generated using IntDB and CReM were filtered according to modified Ro5 (MW ≤ 600 Da and cLogP ≤ 6)

RESULTS

Overview

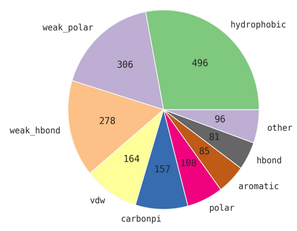
Figure 1



Compound pool generation

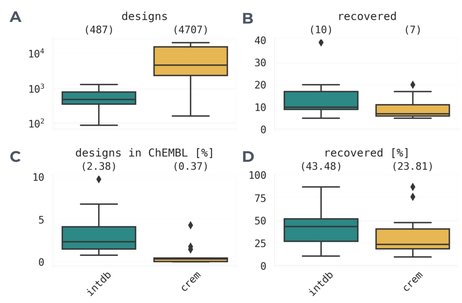
- Series definitions were used as inputs for IntDB and CReM **R-group replacements**. For IntDB, **ligand-protein interactions** that are formed by R-groups in the parent crystal ligands were targeted
- Across all targets, crystal ligands and series, **IntDB** generated 44k molecules and **CReM** 474k molecules. **Figure 1** shows the distributions of physicochemical properties across the two sets, together with those in the retrieved ChEMBL data points
- Hydrophobic, weak polar and weak hydrogen bond **interaction types** constitute the majority of IntDB queries due to their less strict directionality and pattern definitions (**Figure 2**)

Figure 2



Metrics

Figure 3



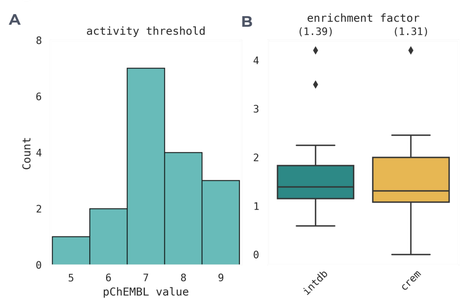
Recovery of experimental compounds

- IntDB and CReM were evaluated for their capacity to **recover compounds** reported in ChEMBL
- **Figure 3** shows statistics of the **intersection** between IntDB or CReM designs and ChEMBL data points **aggregated by target and series,** where both methods were able to recover at least 5 compounds, covering 10 targets, 17 crystal ligands and 17 series
- On average, CReM generated a much larger number of compounds compared to IntDB (**Figure 3A**). Despite this, IntDB was able to recover a **larger percentage of ChEMBL data points** (**Figure 3D**)
- **Enrichment factors** (EF) of IntDB and CReM with respect to baseline activity rates across ChEMBL data points for each series were compared. **Activity thresholds** were set to the pChEMBL value of the parent crystal ligand of each series (**Figure 4A**). The EF for a given series is calculated as:

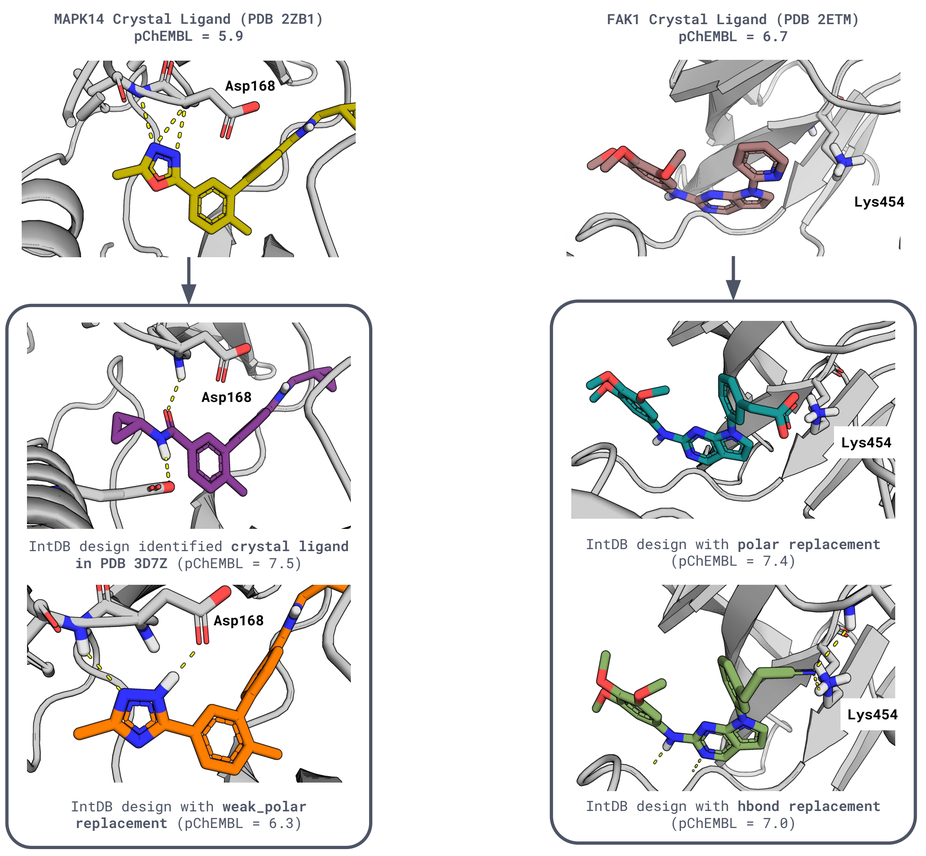
$$EF = \frac{N^{active}_{recovered} / N_{recovered}}{N^{active}_{ChEMBL} / N_{ChEMBL}}$$

- The enrichment factors of the two methods are comparable (**Figure 4B**)

Figure 4



Examples



CONCLUSIONS

- InteractionDB is a molecular replacement tool that allows **molecular designs targeting specific protein-ligand interactions**
- InteractionDB successfully identifies **molecular fragment replacements forming structurally equivalent interactions** in PDB entries of the target protein family
- Comparison with CReM indicates that InteractionDB can **efficiently recover molecules reported in ChEMBL for the respective series** with a smaller number of designs, highlighting its **capacity to imitate design strategies reported in the literature**
- InteractionDB could identify **designs with improved potency** that are **compatible with the binding site** of the target protein

Acknowledgements

We thank Nathan Brown, Beth Caine, Michael Carter, Susan Leung, Josh Meyers and Márton Vass for their contribution to InteractionDB

References

[1] Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montaño, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L. **Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures**. *J. Mol. Biol.* **2017**, *429* (3), 365–371
[2] Modi, V.; Dunbrack, R. L. **A Structurally-Validated Multiple Sequence Alignment of 497 Human Protein Kinase Domains**. *Sci. Reports* **2019**, *9* (1), 1–16
[3] Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. **On the Art of Compiling and Using "Drug-Like" Chemical Fragment Spaces**. *ChemMedChem* **2008**, *3* (10), 1503–1507
[4] Polishchuk, P. **CReM: Chemically Reasonable Mutations Framework for Structure Generation**. *J. Cheminformatics* **2020**, *12* (1), 28