Atomwise

Allosteric hit discovery for phosphatases with AtomNet® virtual screens

Saulo H de Oliveira (Presenter); Pawel Gniewek; Victor Kenyon; Christian Laggner; Teresa Palazzo; Srimukh Prasad; Bradley Worley; Kate Stafford; Brandon Anderson; Michael Mysinger; Henry van den Bedem

Saulo de Oliveira, DPhil
Senior Cheminformatics Scientist
Atomwise Inc.

Why perform virtual screening?

Both practical and scientific reasons why vHTS is a great tool for Pharma

Time

vHTS is faster

Despite automation, screening *in vitro* is still orders of magnitude slower.

Cost

vHTS is cheaper

Testing *in vitro* requires reagents, materials and other overhead costs.

Accessibility of chemical space

***In vitro* HTS**

In traditional screening, synthesis needs to occur **before** screening.

vHTS

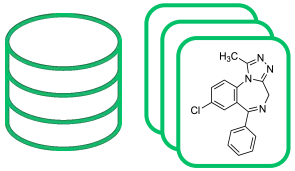
Most of the current accessible chemical space has not yet been synthesized.

Training models for *in silico* hit identification

Step 1: generate protein-compound complexes (poses) via molecular docking

https://blog.atomwise.com/efficient-gpu-implementation-of-autodock-vina

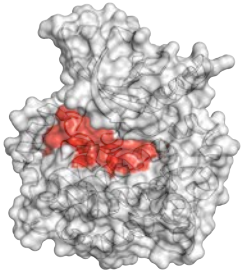
Public + proprietary activity data



ChEMBL

PubChem

GOSTAR

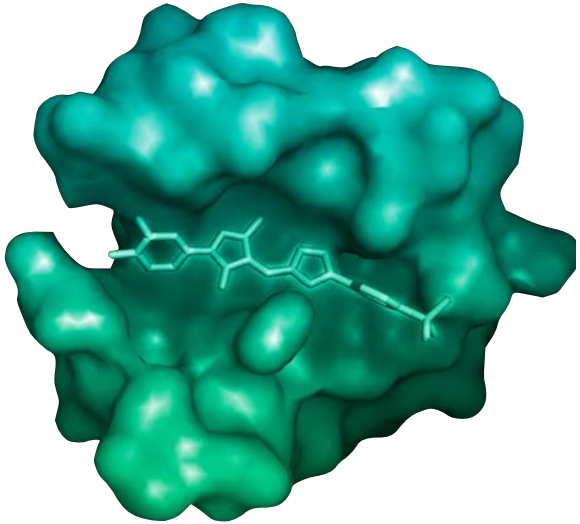


AlphaFold

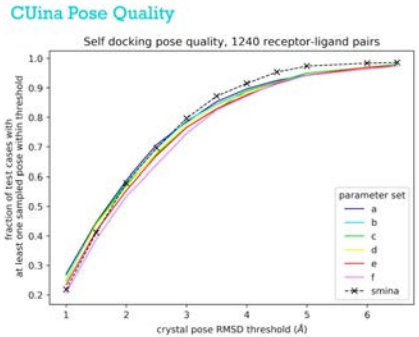
RCSB PDB PROTEIN DATA BANK

Protein structures

CUina



Docked protein-compound complex (pose)



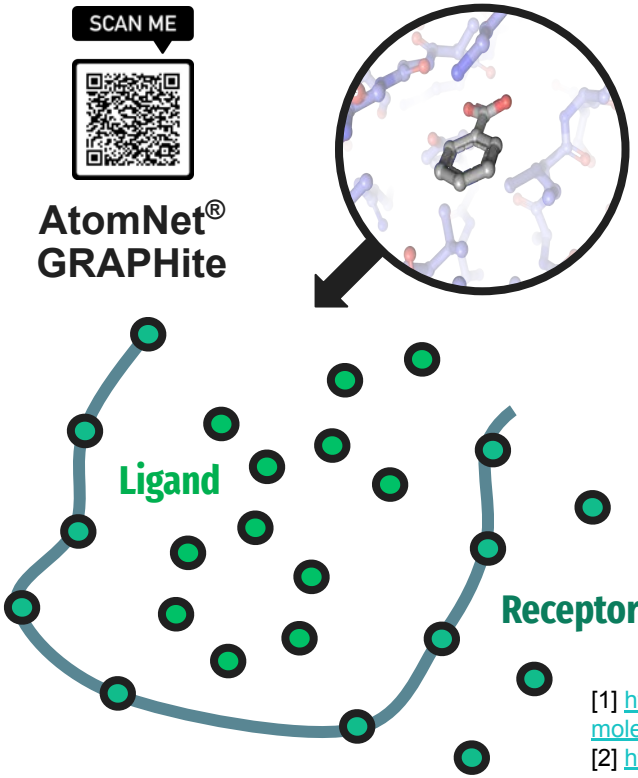
SCAN ME

Training models for *in silico* hit identification

Step 2: feed the docked poses into a neural network and train for different tasks

SCAN ME

AtomNet® GRAPHite



AtomNet® Model

Graph CNN

- Our model architectures are designed to capture and represent physical interactions.
- We have shown that our models are pose-sensitive; bad poses lead to worse scores[2].

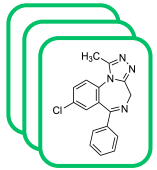
[1] https://blog.atomwise.com/revealing-hidden-determinants-of-molecular-recognition-with-graph-convolutional-attention-mechanisms
[2] https://arxiv.org/pdf/2110.15459.pdf

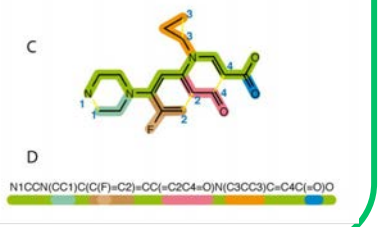
SCAN ME

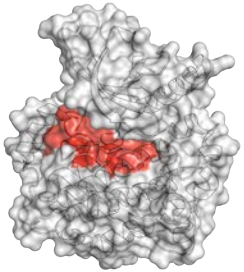
Models can then be used for inference

Step 3: feed docked poses of novel compounds to the model(s), score and rank.

Novel compounds from a library







AlphaFold

RCSB PDB PROTEIN DATA BANK

Protein structures

Pose generation pipeline

Readout

AtomNet® Model

Predictions

-Pose quality[1]
-Active/inactive
-Activity (pKi)
-etc...

SCAN ME

[1] Stafford, K.A. et al. 2022. AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High-Throughput Screens. *Journal of chemical information and modeling*, *62*(5), pp.1178-1189.
https://blog.atomwise.com/identifying-physically-realistic-interactions-between-drug-molecules-and-dynamic-proteins

That looks great, but does it work?

We reported some great results back in 2020 and we have more data now!

Prospective testing results for 100+ targets reported in 2020:

74% project success

6.9% average hit rate

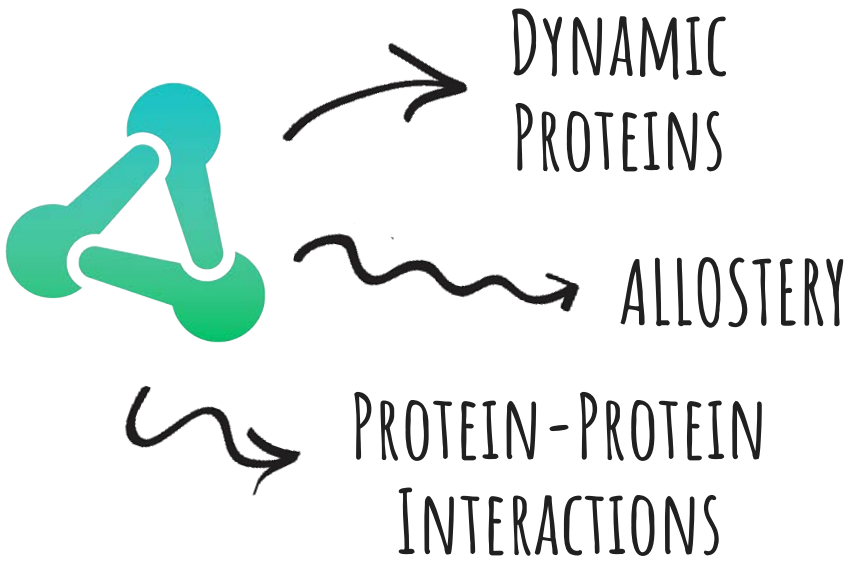
SCAN ME

Stay tuned for our upcoming publication with **updated** results!

https://blog.atomwise.com/results-from-the-worlds-largest-distributed-prospective-application-of-machine-learning-to-small-molecule-hit-discovery

Raising the bar: finding allosteric hits

We have shown our technology works, but can we apply it to challenging cases?



(Personal opinion) Not only should we push the boundaries of the chemical space, but we should also try to innovate on the biology!

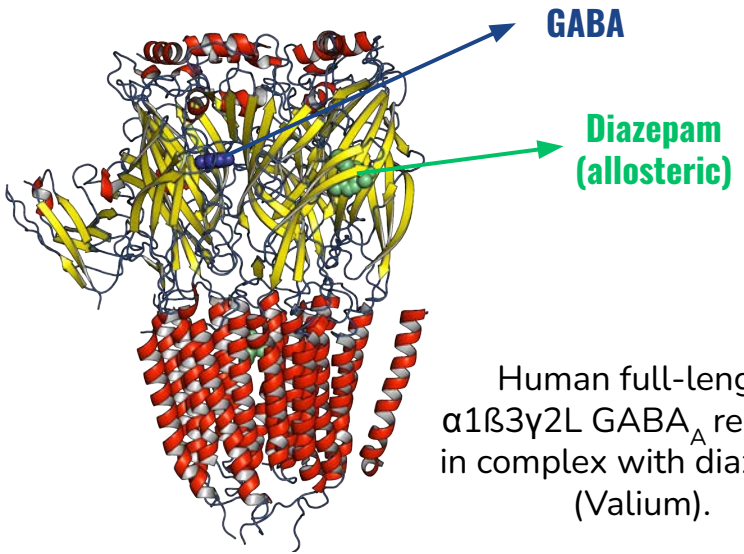
Raising the bar: finding allosteric hits

We have shown our technology works, but can we apply it to challenging cases?

DYNAMIC PROTEINS

ALLOSTERY

PROTEIN-PROTEIN INTERACTIONS



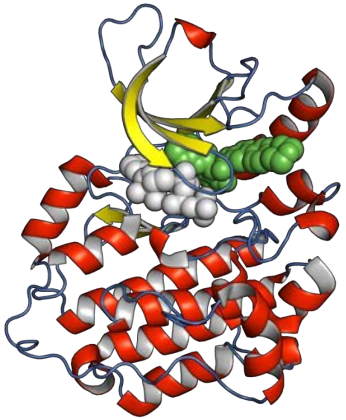
Human full-length α1ß3γ2L $GABA_A$ receptor in complex with diazepam (Valium).

(Personal opinion) Not only should we push the boundaries of the chemical space, but we should also try to innovate on the biology!

The allure of allostery to Pharma research

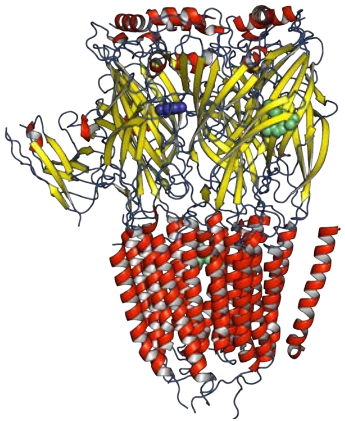
Non-orthosteric molecules offer a path to tackle common problems in the pipeline

SELECTIVITY



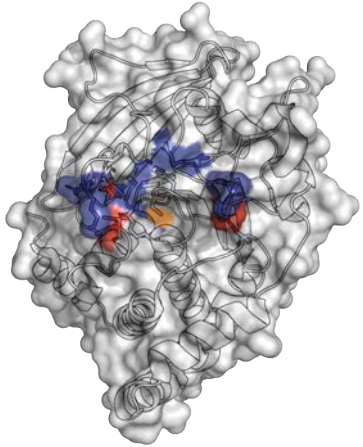
Allosteric sites are less conserved and can be exploited to attain selectivity (EGFR - PDB: 6DUK).

MODULATION



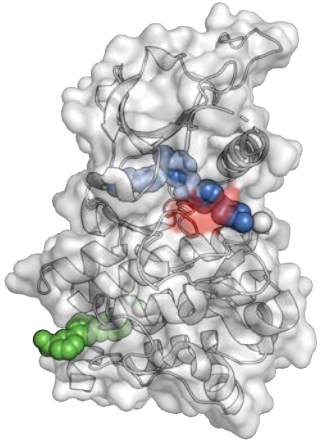
Non-orthosteric binders can act as activators instead of inhibitors ($GABA_A$ receptor - PDB: 6HUP).

CHARGED SITES



Crystal structure of PTP1B, which contains a highly charged orthosteric site (PDB: 2QBP).

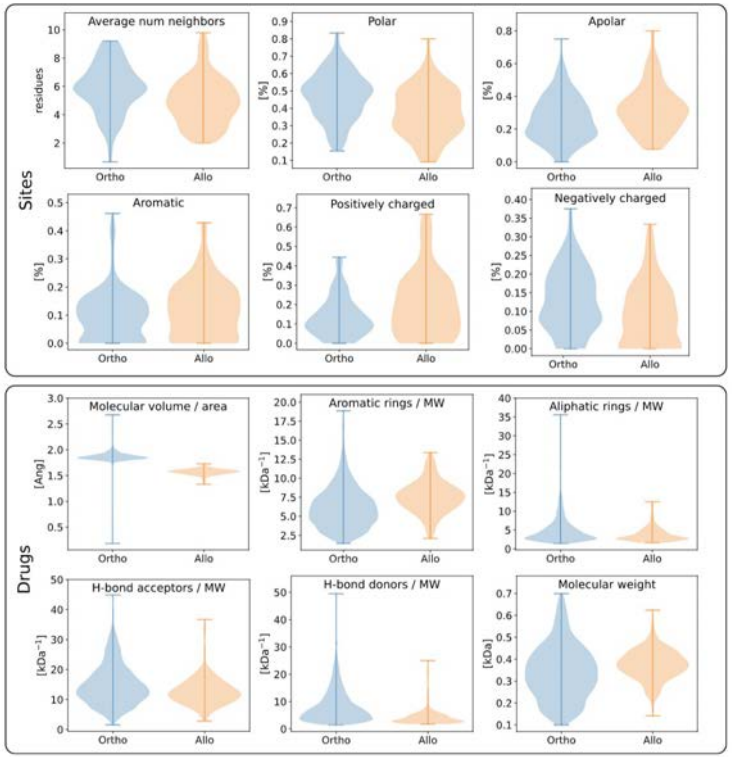
DRUG RESISTANCE



Allosteric binders can provide additive inhibitory activity against T315I mutant human Bcr-Abl (PDB: 3K5V).

Finding allosteric hits is a challenge for AI

Mislabeling, fewer data, different characteristics of pockets and compounds…



Site-labeling is inconsistent at best

A majority of compounds with measured activity cannot easily be assigned to a binding site.

Allosteric hits (and drugs) are much less common, so we also deal with fewer data points.

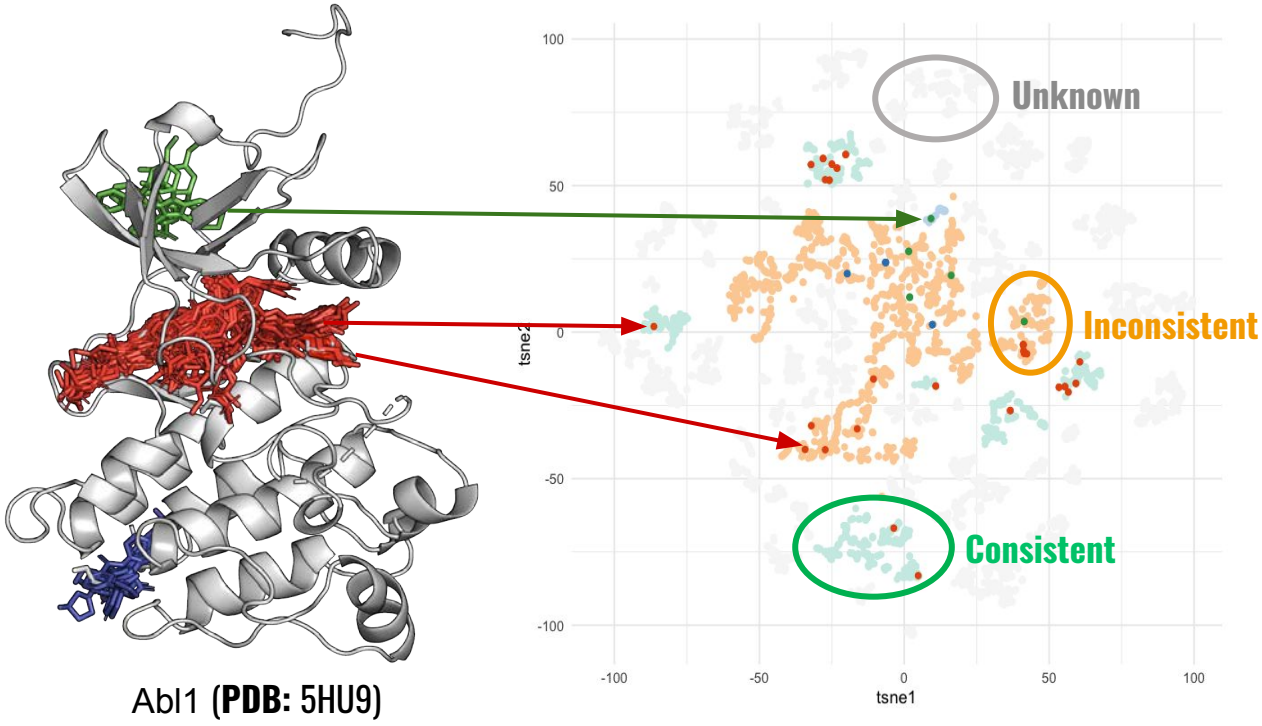
Allosteric and orthosteric sites/compounds have different characteristics (see figure on the right).

[*] Tan, Z.W., Tee, W.V. and Berezovsky, I.N., 2022. Learning about Allosteric Drugs and Ways to Design Them. *Journal of Molecular Biology*, p.167692.

We built a pipeline to address site-labeling

If you are interested, I presented this at ACS Fall 2021 (scan QR >>>)

SCAN ME



Abl1 (**PDB:** 5HU9)

>0.5 M

COMPOUNDS

>500

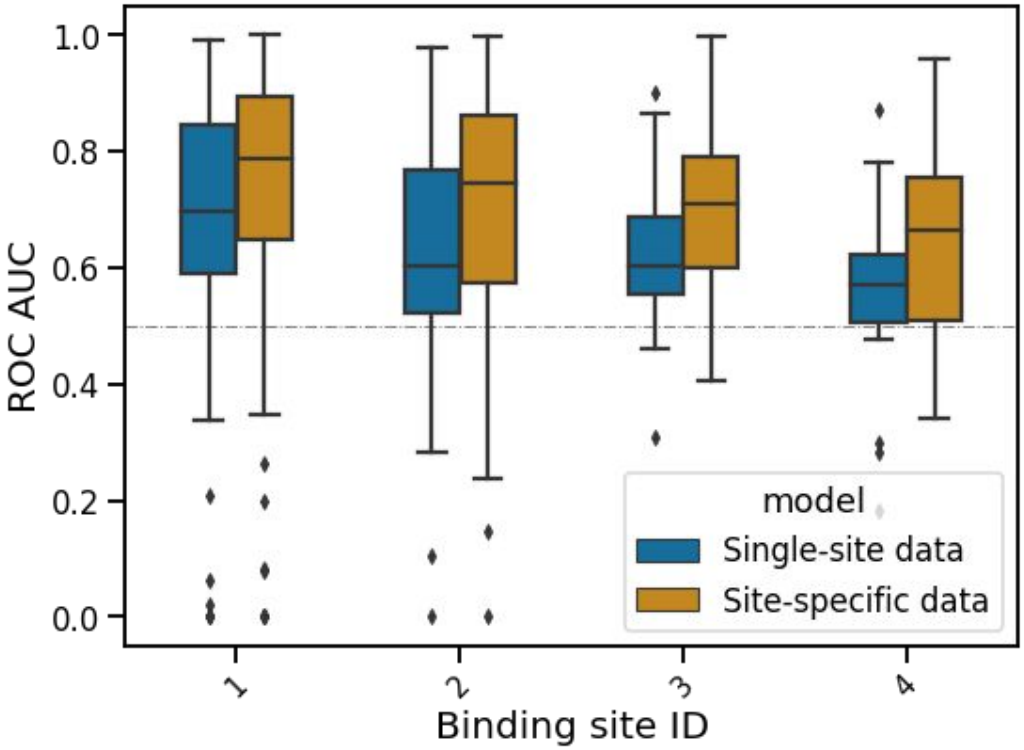
MULTI-SITE PROTEINS

18% of compounds

OF KNOWN ALLOSTERIC TARGETS MAPPED TO NON-PRIMARY SITES

Multi-site data improves model performance

Performance improvement is also observed for primary site



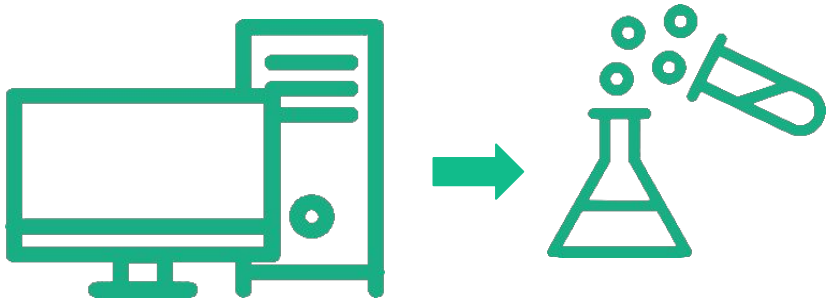
- These exploratory models were trained specifically for ACS using data from public databases such as ChEMBLdb.
- No binding sites for the proteins in our test set were used during training (70% sequence similarity split).
- Improvement in performance for primary site highlights the impact of incorrect data labeling.

Informative benchmarking is not trivial

Prospective experiment is the only reliable way to evaluate ML algorithms in our domain

Let us test our predictions in the lab

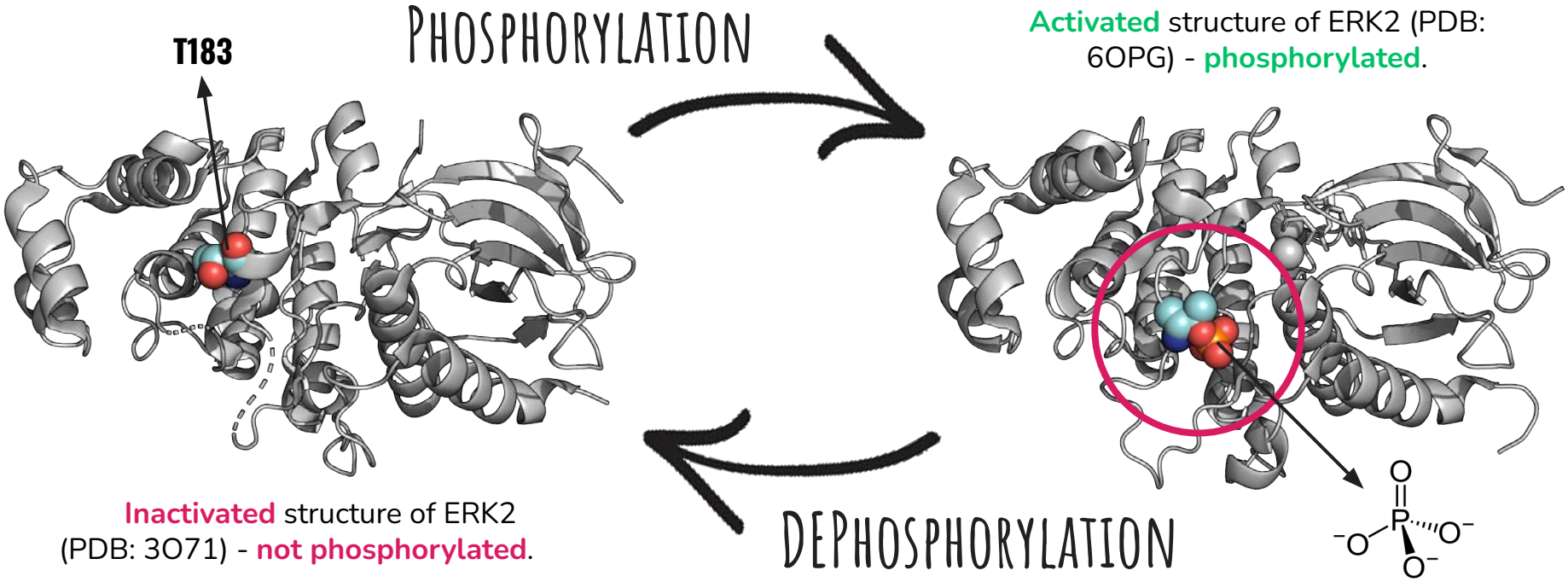
Existing benchmarks tend to reward overfitting, not accuracy[1]. Therefore, the only way to know for sure our models are working is to perform prospective experimentation.



[1] “Most Ligand-Based Classification Benchmarks Reward Memorization Rather than Generalization”, I. Wallach and A. Heifets, JCIM 2018 58 (5), 916-932.

We chose phosphatases to perform this test

Why? Vastly important class of targets that has been notoriously difficult to drug



We chose phosphatases to perform this test

Why? Vastly important class of targets that has been notoriously difficult to drug

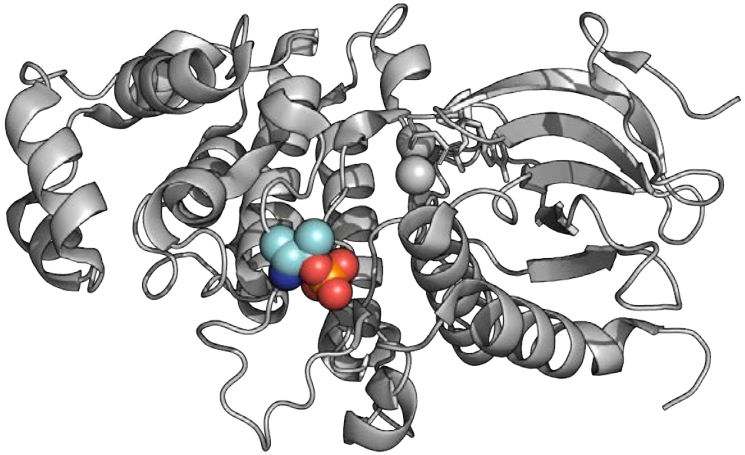
PHOSPHORYLATION

Kinases

One of the, if not the most important drug target class, with more than 70 drugs approved by the FDA targeting a kinase.

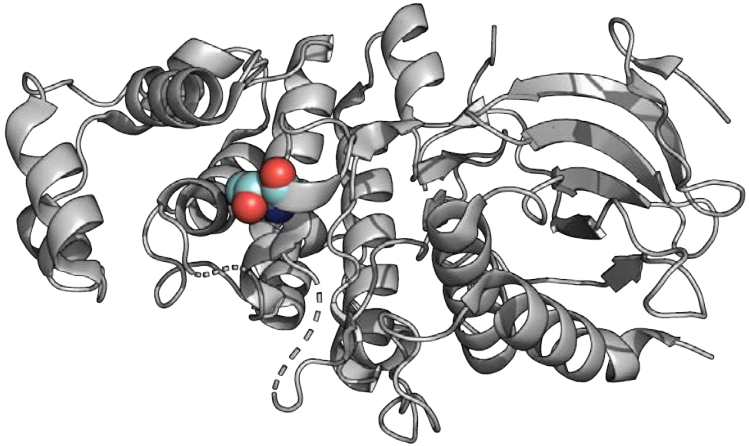
Perform phosphorylation of other proteins, critical for cell signalling.

Activated structure of ERK2 (PDB: 6OPG) - **phosphorylated**.



We chose phosphatases to perform this test

Why? Vastly important class of targets that has been notoriously difficult to drug



Inactivated structure of ERK2 (PDB: 3O71) - **not phosphorylated**.

Phosphatases

Perform dephosphorylation of other proteins, also critical for cell signalling.

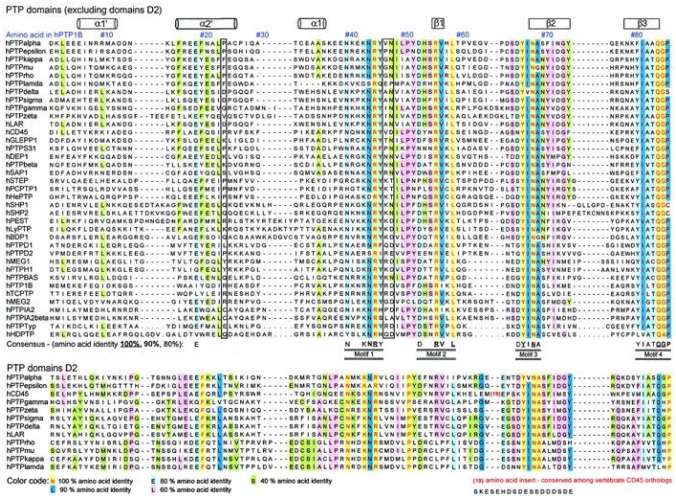
Yet, not a single drug approved by the FDA targets phosphatases.

DEPHOSPHORYLATION

Why allosteric hits in phosphatases?

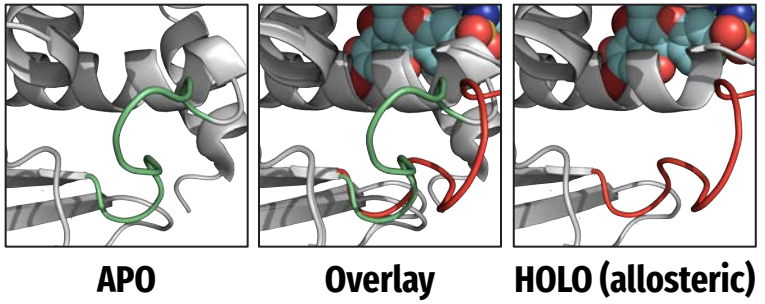
Poor selectivity and highly charged orthosteric site lead to ADMET liabilities

Orthosteric site is very conserved so selectivity is challenging. Charged residues are also an issue.

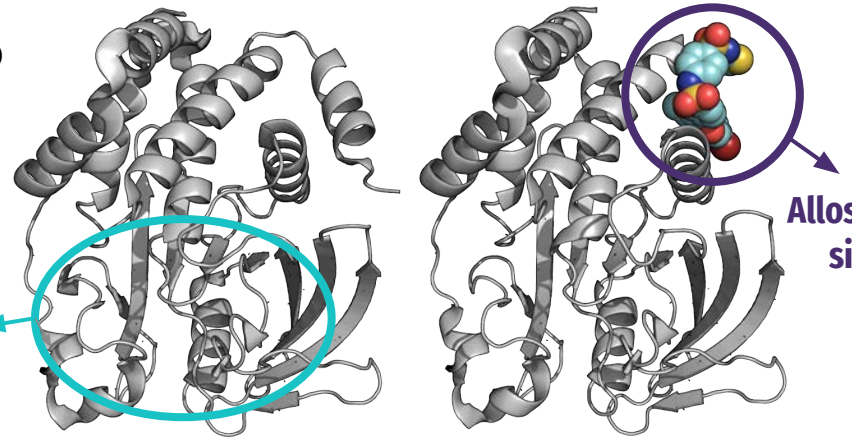


[*] Andersen, J.N., et al, 2001. Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Molecular and cellular biology*, *21*(21), pp.7117-7136.

Change in conformation upon allosteric binding:



APO and HOLO (allosteric) structures of phosphatase PTP1B. (PDBs: 7RIN and 1T4J)

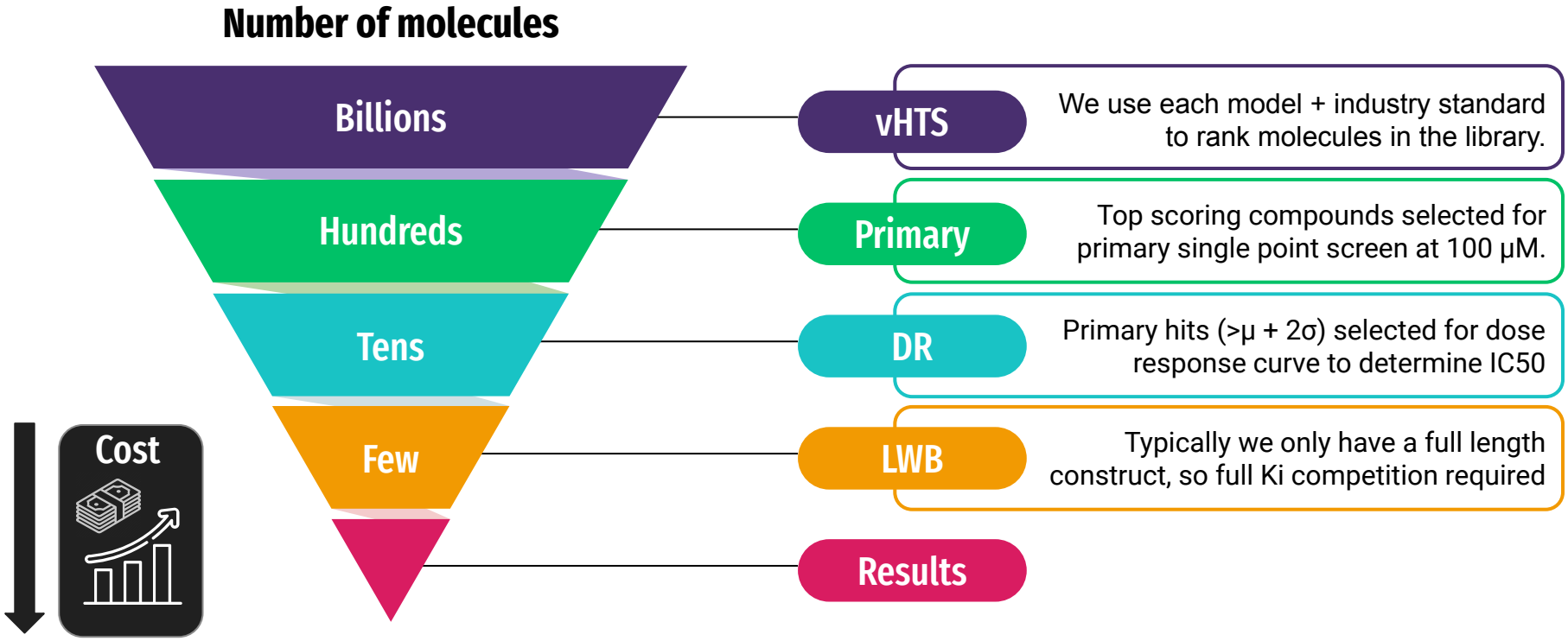


Fairness + measuring incremental progress

We tested several models and docking score in our prospective experiments

Detailing our hit funnel/screening cascade

Goal is to determine if we have an allosteric binder **cheaply**



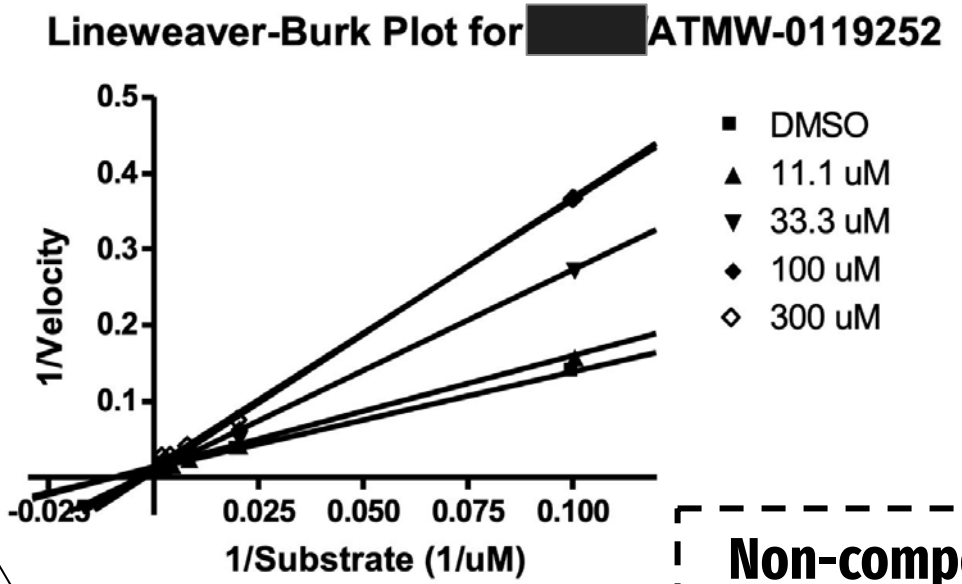
How do we verify if hits are allosteric?

We use a Lineweaver-Burk (LWB) analysis

AW2

Model 2

Some site-labeling data, no data augmentation

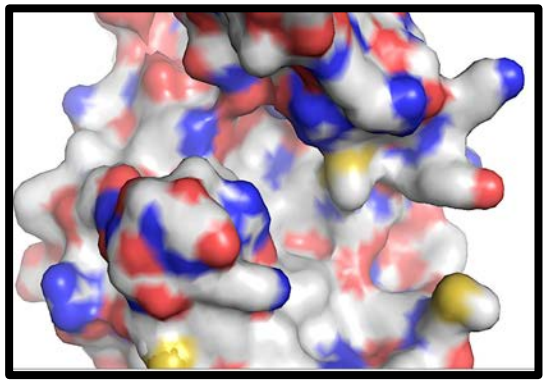


Non-competitive due to convergence at x=0

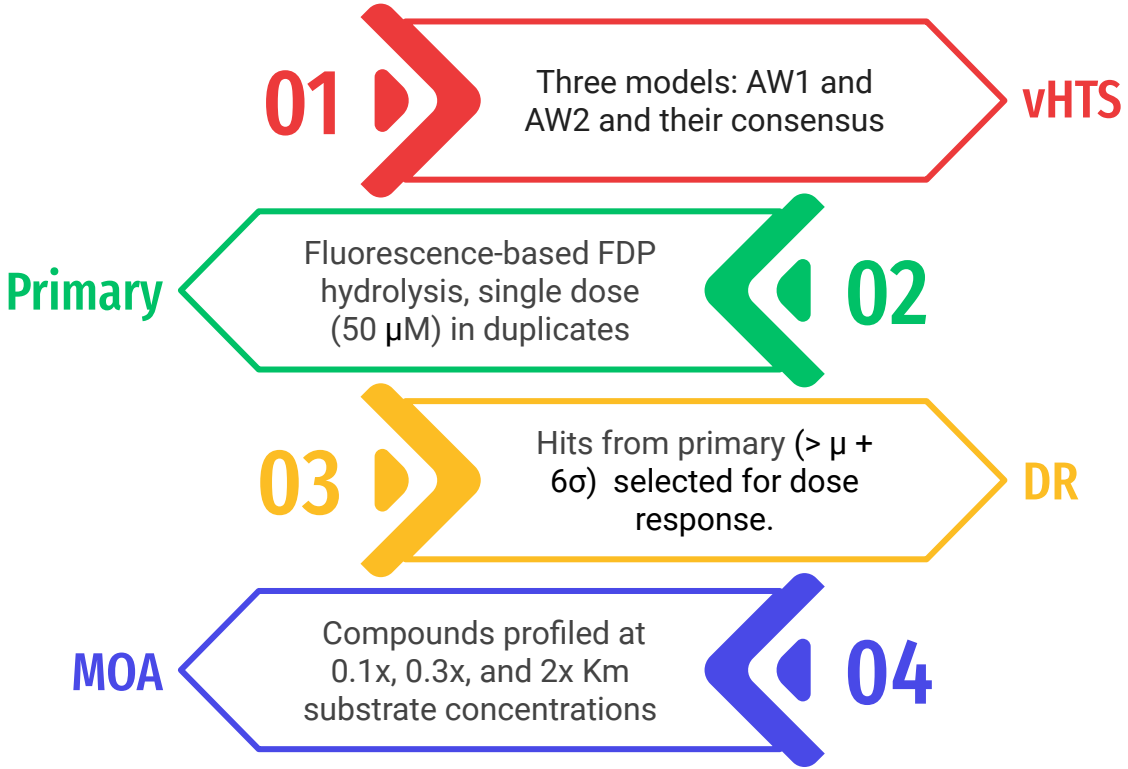
What if no target structure is available?

For this case study, we used a slightly different setup and an AlphaFold2 model

No available structure for this phosphatase



Screen was conducted with an AlphaFold2 model



Summary table of results so far

Complete results are available for two targets: Phosphatases T01 and T05

Protein	Primary screen	Dose response	LWB/MOA	Most potent
Phosphatase T01*	13/310 (4.2%)	2/13	1/2	51μM (AW2)
Phosphatase T02	21/406 (5.1%)	5/21	2/4 tested** LWB pending	14μM (AW3)
Phosphatase T03	13/379 (3.4%)	6/13	TBD	15μM (AW3)
Phosphatase T04	36/487 (7.4%)	TBD	TBD	est. <10μM (AW3)
Phosphatase T05 (no structure)	63/530 (11.8%)	37/63	23/37	11μM (AW2)

IS: 4 hits, 1 allo hit*

AW1: 4 (22) hits, 1* (23) allo hits

AW2: 5 (26) hits, 1 (23) allo hits

AW3: 5 hits, TBD

[*] The allosteric site was identified from a peptide

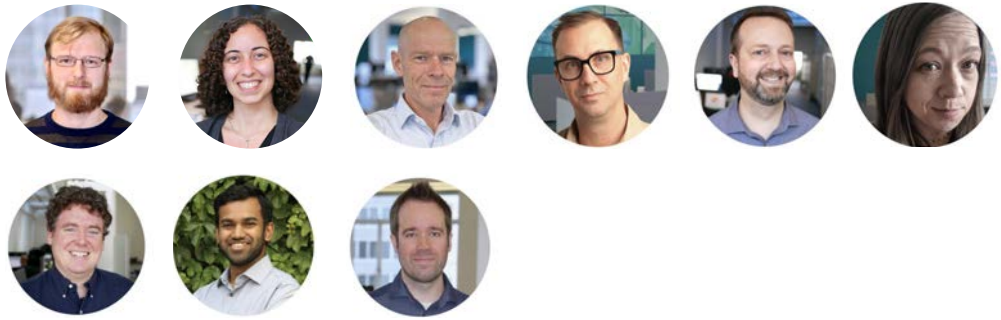
[**] Results are shown for an assay with an alternative construct, allosteric MOA needs to be further validated via LWB analysis

Conclusions

- Single to double digit micromolar inhibition: similar in magnitude to other allosteric phosphatase inhibitors.
- Experimental *n* is still small, but suggestive of success. Working towards n = 5.
- Large scale, systematic vHTS campaign to unlock allosteric regulation of phosphatases.

Acknowledgments

Atomwise contributors:



Thanks to the entire Atomwise team!

Our Other Talks & Posters

https://info.atomwise.com/acs_fall2022

Join Us!

https://www.atomwise.com/careers/