

# 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



## Training models for *in silico* hit identification

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



## Training models for in silico hit identification

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



- 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<sup>2</sup>.



#### Models can then be used for inference

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



#### 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:**







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?



(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



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



MODULATION



**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 T3151 mutant human Bcr-Abl (PDB: 3K5V).

### Finding allosteric hits is a challenge for Al

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 >>>)



>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.

\land Atomwise

### 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<sup>1</sup>. 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

# Kinases

PHOSPHORYLATION

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: 60PG) - 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: 3071) - not phosphorylated.

## **Phosphatases**

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

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



\land Atomwise

#### 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.

PTP domains (excluding domains D2)														
	٥	a1' ()	0	a2'	0	( a1	T.		<b>B1</b>			62		63
Amino anid	In hPTP1	#10		#20	#30		#40		50	850		#70	47	0
hPTPaloha	DKLEEF	INREMADON-	K1 .	REFENAL	RACPICA		SKEENKEK	RYVNEL	PYDHSRMH	TPVEGV	PDSDY	NASFINGY	OF KNK	LAADGP
hPTPepsilor	EHLEEE	IRIRSADDC -	KQ	REEFNSUS	SGHIQG	TFELA	NKEENREK	RYPHIL	PNDHSRVI	LSQLDGI	PCSDY	NASYIDGY		AAOOP
hPTPkappa	ADLLOH	INLMKTSDS -	YG!	KEEYESP	EGQSA	SWOVA	KKDQNRAK	RYONII	AYDHSRVI	LOPVEDD	PSSDY	NANYIDGY	QRPSH	ATOOP
hPTPmu	ADLLQH	I TOMKCAEG -	YG	KEEYESF	FEGQSA · · ·	PWDSA	KKDENRMK	RYGNII	AYDHSRVA	LQTIEGD	TNSDY	NGNYIDGY	······HRPNH	TATOOP
hPTPrho	ADLLQH	I TOMKROOG -	YGI	KEEYEALS	EGQTA	····SWDTA	KEDENRNK	RYCNII	SYDHSRVR	LLVLDGD	PHEDY	NANYIDGY	HRPRH	TATOOP
hPTPlamda	ADLLQH	INGMETAEG -	YG!	KQEYESF	EGWDA	· · · · T · · · ·	KKKDKVKG	ROEPMP	AYDRHRVK	LHPMLGD	PNADY	NANYIDGY	HRSNHI	ATOOP
hPTPdelta	LELADH	IERLKANDN -	LKI	SQEYESI	PGQQF	···· TWEHS	NLEVNKPK	RMANVE	AYDHSRVL	LSAIEGI	PGSDYN	NANYIDGY	RKQNAT	TATOOS
hPTPsigma	ADMAEH	TERLKANDS -	LKI	SQEYESI	PGQQF	····TWENS	NLEVNKPK	RMANVI	AYDHFRUI	QPIEGI	MGSDY	MANYVDGY	RRQNA	ATQGP
heregamm	AKGFVKH	IGELTENNO-	TEFEE	BEDFEEVE	CIADMN-	TADES	NHPENKHK		ATONGRUM	LAPLPOK	DOKNEDT	NANTYDGT	NORY	ATOOP
N AR	TOLADN	LEPIKANDO.	TEEFE	BOEVESU	REGOOF		NIEVNERK		AYDHSBY	TSIDOV	POSOY	NYIDAY	RECNA	ATOSE
hCD45	DILLET	YKRKIADEG-	RP	LAFFOSIS	PRVFSKF	PIKEA	RKPFNONK	RMYDEL	PYDYNRVE	LSEINGD	AGENY	MASYIDGE	KEPRK	AAGGP
bGLEPP1	DDFDAY	IKDMAKDSD .	YK	SLOFFEL	LIGLDI	PHEAA	DUPUNRCK	RYTNIL	PYDESRUS	VEMNEE	FGADY	NANYIPOY	NSPOR	ATOOP
hPTPS31	KSFLQH	VEELCTNNN -	LK	OFFFSEL	KFLODL	55 T DA	DLPWNRAK	REPAIK	PYNNNRVA	LIADABY	PGEDY	NASYISGY	LCPNE	LATOOP
hDEP1	ENFEAY	FKKQQADSN -		AREYEDL	LVGISQ	PKYAA	ELAENRGK	RYNNVL	PYDISRVA	LSVQTHS	· TDDY	NANYMPGY	HSKKDI	ATOOP
hPTPbeta	NOFEGH	FMKLQADSN -	YL1	SKEYEEL	OVGRNQ	SCD   A	LLPENRGK	RYNNEL	PYDATRY	LSNVDDD	PCSDY	NASYIPGN	NFRRE	TUTOCP
hSAP1	EDFADH	VRKNERDSN -	· · · · · · CGI	ADEYOOL	LVGHSQ	····SQMVA	SASENNAK	RYRNVL	PYDWSRVP	LKPIHEE	PGSDY	NASFMPGL	········WSPQEI	ATOOP
hSTEP	SRVLQA	EELHEKALDF	FL1	QAEFFEI	PMNFVV	PKEY	DIPGRCRK	RYKTIL	PNPHSRVC	LTSPDPD	D - PLSSY	MANYIRGY	GGEEKV	ATOOP
NPCPTP1	SRILTR	SQLRDVVASS	9 HLI	QSEFME I	MNFVD · · ·	PKEI	DIPRHGTK	RAKTEL	PNPLSRVC	LRPKNVT	D-SLSTY	MANYIRGY	S GKEKA	ATOGP
PHOP TP	GHPLTR	WALQROPPSF	KQI	EEEFLKI	SNEVS	PEDL	DIFGHASK	DE ME TIL	PNPQSRVC	LGRAQSO	E · · DGDY	MANYIRGY	DGKEKV	ATOSP
NSHP1	ADIENK	VEELNKKGES	TONYKOO	NEEPESCH	COLORALI	QRLEG	OBOENKEK	30315	PFDHSKYI	LUGROSN	- PUSDI	ANTIAN-	-QLLGPD-ENAKI	ASGEC
NDECT	E.L.B.C.F	LOBYOANKER	DUNCEDH	ABDENBL	DEL ETVVET	EXIVETATO	EVEENUVV			TIXTOR	0000		PEINUMMONPANO	TOOR
NUPTP	FILOFF	I DEADSKKIT	×	ANDELKI	POSTRYKA	OKTYPTTVA	ENAKNIKK	A PARTICULA	PYDYSPME	GLITED	EDSSY	NELEGY	YOPKA	ATOOP
NBDP1	DBARSF	LERLEARGON	EGAVI	AGEFEDI	ACBAAWKA	DOVCSTVAG	GRPENVEK	MAMKOVL	PYDOTRY	LALLQEE	GHEDY	MONFIRGY	DOBLA	ATOOP
NPTPD1	ATNDER	CKILEGRLEG	GM	FTEYERI	KKRLVDG-	ECSTA	RLPENAER	RFODVL	PYDDARVE	LVPTKEN	NTGY	NASHIKVS	V SGIEWD	ATOOP
hPTPD2	VPMDER	FRTLKKKLEE	G M 1	FTEYEQU	KKKANG	IFSTA	ALPENAER	SRIREVV	PYEENRVE	LIPTKEN	NTGY	NASHIKVV	VG GAEWH	ATOOP
hMEG1	HSLRES	MIQLAEGLIT	rg	LTOFDOL	RKKPGM	TMBCA	KLPQNISK	RYRDIS	PYDATRY	LKG	· NEDY	NANYINME	IP 55511NQ	LACOOP
hPTPH1	DTLEGS	MAGLKKGLES	3G TI	LIGFEQU	WRKKPGL	AITFA	KLPQNLDK	RYKDVL	PYDTTRY	LQG	NEDY	NASYVNME	IP AANLVNK	ATOP
<b>hPTPBAS</b>	KSVIRV	LRGLLDQGI -		SKELENL	GELKPLD	QCLIG	QTKENRRK	NRYKNIL	PYDATRY	LGD - · · ·	· EGGY	NASFIKIP	VGKEEFV	ACOOP
NPTP1B	MEMEKE	FEQIDKSGS.		VAAIYQDI	RHEASDF	PCRVA	KLPKNKNR	RMRDVS	PFDHSRIN	LHQE	· DNDY	NASLIKME	EAGRS	LTOOP
hitpip	TTIERE	FEELDTORR-		VOPLYLEIN	NESHDY	PHRVA	KFPENRNR	RUNS	PTDHSHUP	LQNA	ENDY	NASLVDIE	EAGRS	LTOCH
MMEG2	MTIGEL	VDYVNARQKO	2	YEEYEDI	RENPVG -	····TFHCS	MSPGNLEK	H MG DVP	CLDQTAN	TKRSGH	T. QTDY	NASFMOGY	KQKNA	GTOGP
hDTDIA2ba	PIMILAT	MEDHLENKOP		EVENUAL	ATUAEPN.	PREVA	OREENVEY	COL AVE	TYDUCER	LAUREN	S- REDT	COLNOR	DPRMPA	TOCH
NOTOTION	TAIKOC	I KILEEKTAA	YO	MOLENAU	ELKNIRG	EEYSO	NOPSNEEK	a vie ni i	PYDETEM	OF	SKOY	A AY I B I Y	N COREYE	ATOOR
NHOPTP	ERLEGI	QOFLEAFRO	LODY GAL	DTYMBEL	DAGENDAR	GREIA	LARCYSLK	RHODYM	PYDENES	LRSG	KDDY	MASCYEGL	S. PYCPP	VATOAR
Concentration	Inmine e	aid identity 1000	( ontic entry	t r	Jourgenon			NPV	D BW		0.00			VIII TOOR
Compension	- farmino a	ord rouning and	a							-				
							Moon	1	MOOT 2	-	MO	08.3		MORE 4
PTP dom	ains D2													
hPTPalpha	TSLETH	LOKIYNKIPO	TENNG	LEEFKKU	TSIKIGN	DKMRTG	NUPANMER	RVLQII	PYEFNRVI	IPVKRGE	ENTOY	NASFIDO	RQKDS	TASOOP
hPTPepsilo	SSLEKH	LOTMHOTTTH	FDKIG	LEEEFRKU	TNVRIMK	ENMRTO	NLPANMER	ARVIQII	PYDENRVI	LSMKROO	EYTDY	MASFIDGY	RQKDY	TATOOP
hCD45	SELHPY	LHNMKKROPA	P BEPSPI	LEAFFOR	PSYRSMR -	TQHIG	NGEENKSK	RNSNVI	PYDYNRVE	LKHELEM	ITHE PEKY	NASFIMSY	WKPE VI	ADDAAIM
hPTPgamm	ANGLHST	VNSILIPGVO	GKTR	LEKOFKLV	TOCNAKY	VECFSA	OKECNKEK	RNSSVV	PSERARVO	LAPLPON	KGTDY	NASYINGY	·YRSNE	ELITONP
hPTPzeta	SHIHAT	VNALLIPGP/	GKTK	LEKOFOL	SQSNIQQ -	SDY SAA	LKOCNREX	RTSSII	PVERSRVG	ISSLSG-	EGTDY	NASYIMGY	· · · · · · · YQSNE	EL L TOHP
hPTPsigma	RELYAT	IGKLAQVEPO	EHVTO	MELEFKRU	ANSKAHT -	SRFISA	NLPONKEK	RLVNIM	PYESTRVC	LOPIRGY	EGSDY	NASFIDGY	ROOKA	TATOOP
hPTPdelta	RNLYAY	IQKLTQIETO	SENVTO	MELEFKRU	ASSKANT	SRFISA	NLPONKEK	RLVNIM	PYESTRVO	LQPIRGV	EGSDY	NASFIDGY	ROOKA	TIATOGP
hLAR	RNLYAP	ICKLOQVPPC	ESVTA	MELEFKLU	ASSKAHT	SRFISA	NLPONKER	NRLVNEM	PYELTRVO	LOPIROV	EGSDY	NASFLOGY	ROOKA	TIATOGP
hPTP/ho	CEFREL	YYNISRLOPC	TNSSQ	KDEFGT	NIVTPRVRI	EDCBIG	LEPRNHOK	REMOVE	LDRCLPH	LISVDG-	ESSNY	NAALMOSH	KQPAA	VVTOHP
hPTPkasea	CEEKA	YEDNINKLOP	THEEN	KEEFRICT	NEVTRELO	EDCELA	CLERNNEY	REMONE.	PORCLE		EDONT	SAALNDS .	ROBAA	LUX TOWN
hPTPlanda	REFEAT	YKEMIRIDEC	D SNSSO	REFLOT	NEVTPELDI	EECALA	LEPENEDK	REMOVI	PROBELPS	LISTOR	DENNY	NAAL TOST	TRRSA	MUTINE
Marrida		RIDPO			A PECO		C. S. C. ARDA	ano abye	- BROLFI			1001	TRROA	Care of the second
Color code	9: N 100 % a	mino acid identity	E 80 % a	emino acid ident	tity 5 40	% amino acid ider	ntky			(19) 8	mino acid ins	ert - conserved	among vertebrate CD-	45 orthologs
	E 90 % a	mino acid identity	L 60 % a	amino acid ident	tity					SKE	SEHDSDE			

[\*] 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

Overlay HOLO (allosteric)



#### Fairness + measuring incremental progress

We tested several models and docking score in our prospective experiments





**AW1** 

#### Model 1

No site-labeling, no sitespecific data augmentation

#### Model 2

Some site-labeling data, no data augmentation



All site-labeling data + data

**AW2** 

## Detailing our hit funnel/screening cascade

Goal is to determine if we have an allosteric binder cheaply

#### Number of molecules



### How do we verify if hits are allosteric?

We use a Lineweaver-Burk (LWB) analysis



### 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)



[\*] 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/