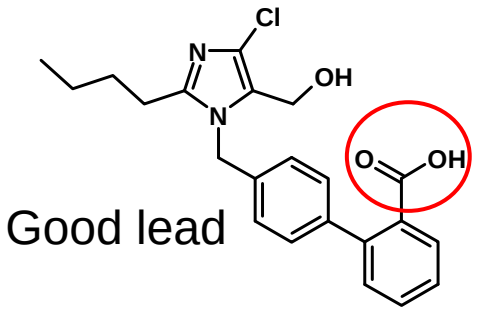
Computational methods for the identification of bioisosteres

James Mills & Carolyn Barker
Pfizer GRD, Sandwich, UK

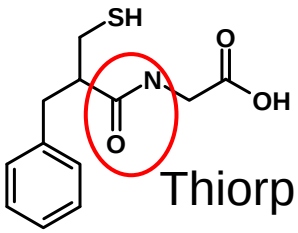
What a chemist needs to know about chemoinformatics and SAR
June 2008

Bioisosteres: a definition

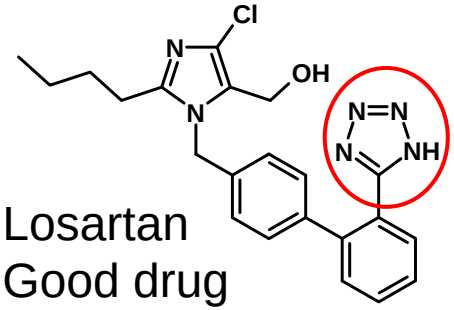
"Substituents imparting similar biological properties on a compound"



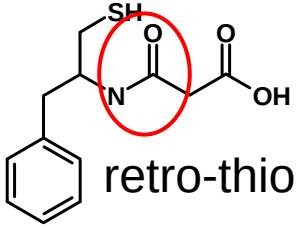
Good lead



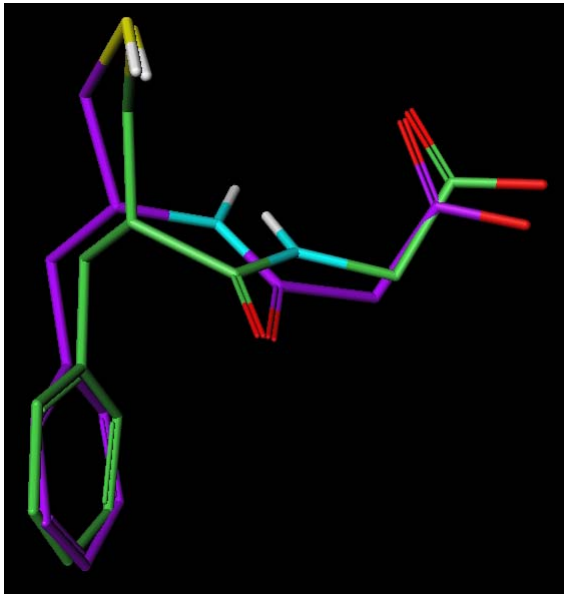
Thiorphan



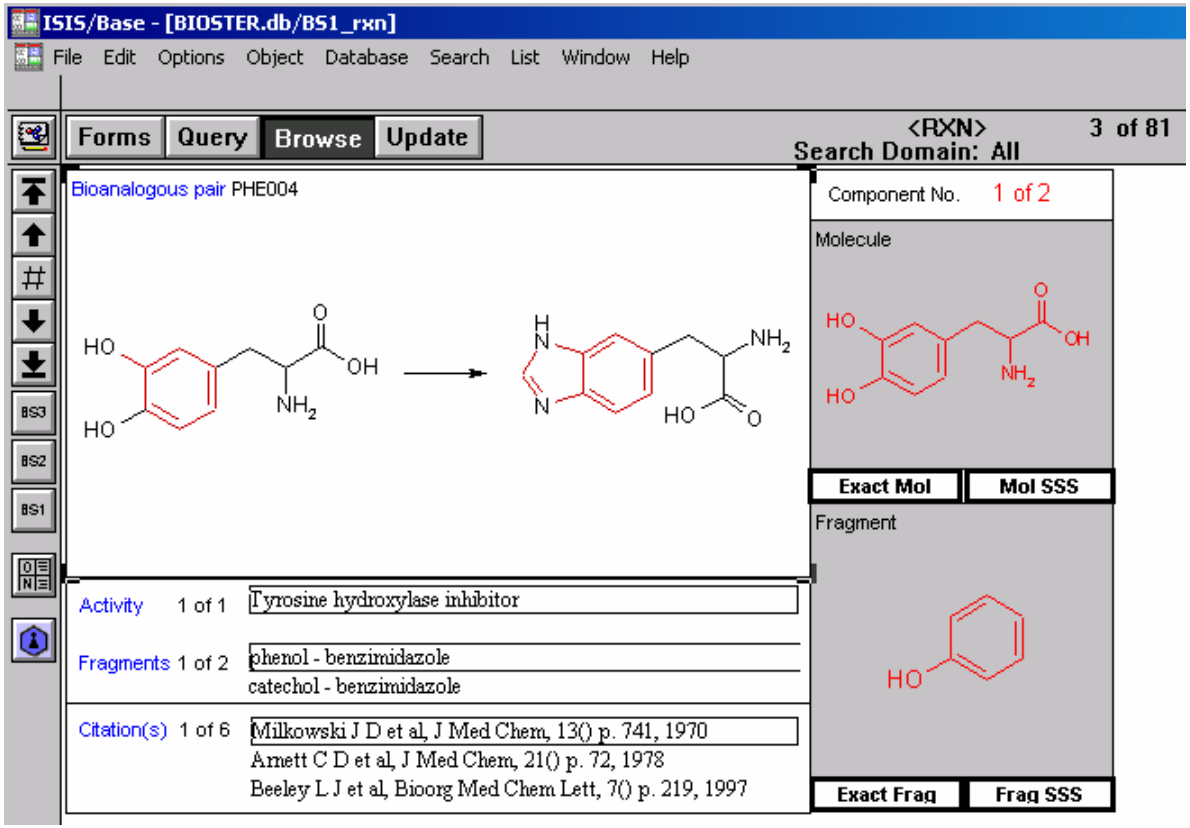
Losartan
Good drug



retro-thiorphan



Commercial tool: BIOSTER.DB



- Lookup database of 14K bioisosteric transformations
- Activity data not directly included
  - Have to look up in lit

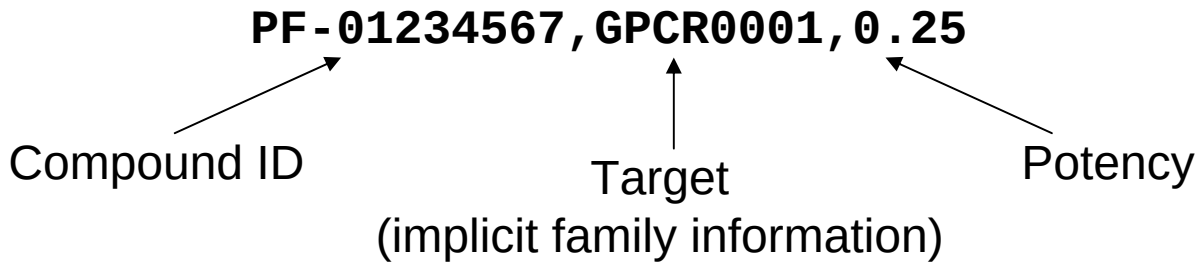
New approaches to bioisosteres

- Considering two ways to identify bioisosteres
  - Analysis of pharmacological screen data
    - Pairs of active ligands with a single difference
      - SWAP
  - Analysis of ligand-protein co-crystals
    - Groups occupying same pocket in crystal structures
      - PDBSEARCH

Bioisosteres from screen data

- Single changes between pairs of molecules
- Can we mine Pfizer screening data for bioisosteres?

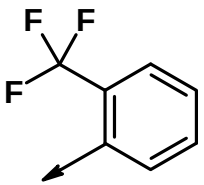
- Data underpinning this: for all Pfizer screening data:

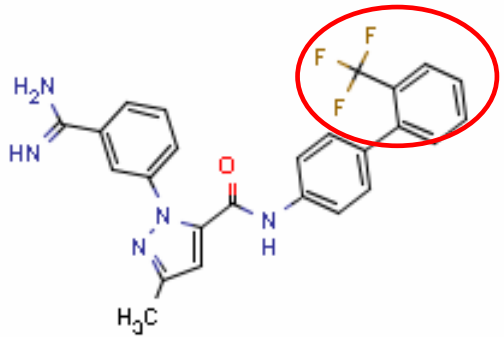


  - 1.4M cpds over 1100 targets
  - Also have assembled set of external data
    - 117K cpds over 737 targets

Mining ligand screening data

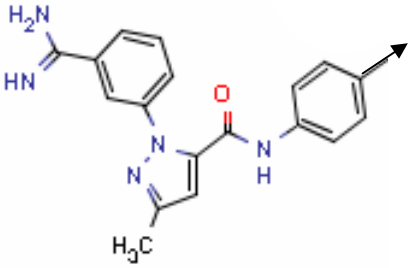
Substructure search of 120K literature actives



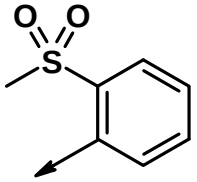


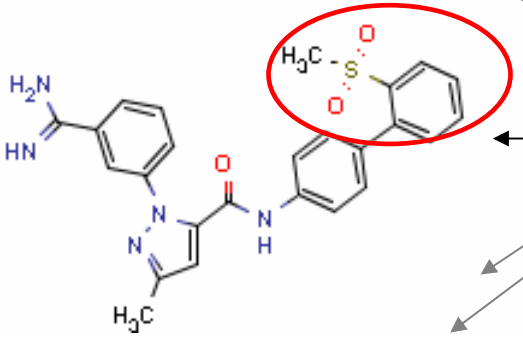
FacXa IC50 = 40pM

Crop input substructure



Substructure search of 2625 FXa actives

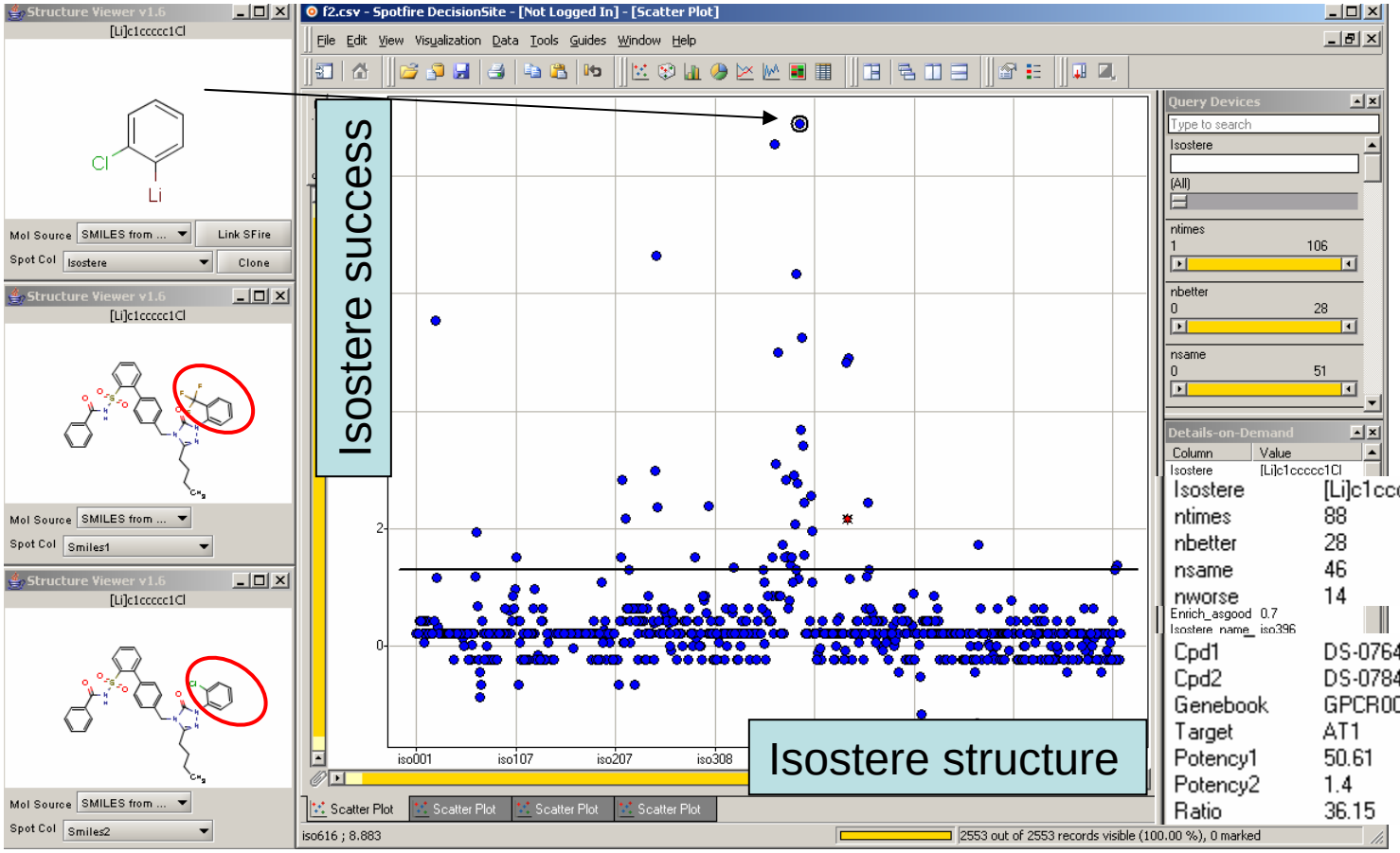




FacXa IC50 = 8pM

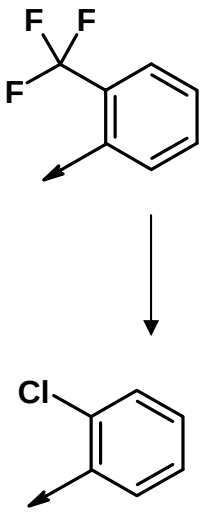
Isosteres for *o*-CF3-Ph

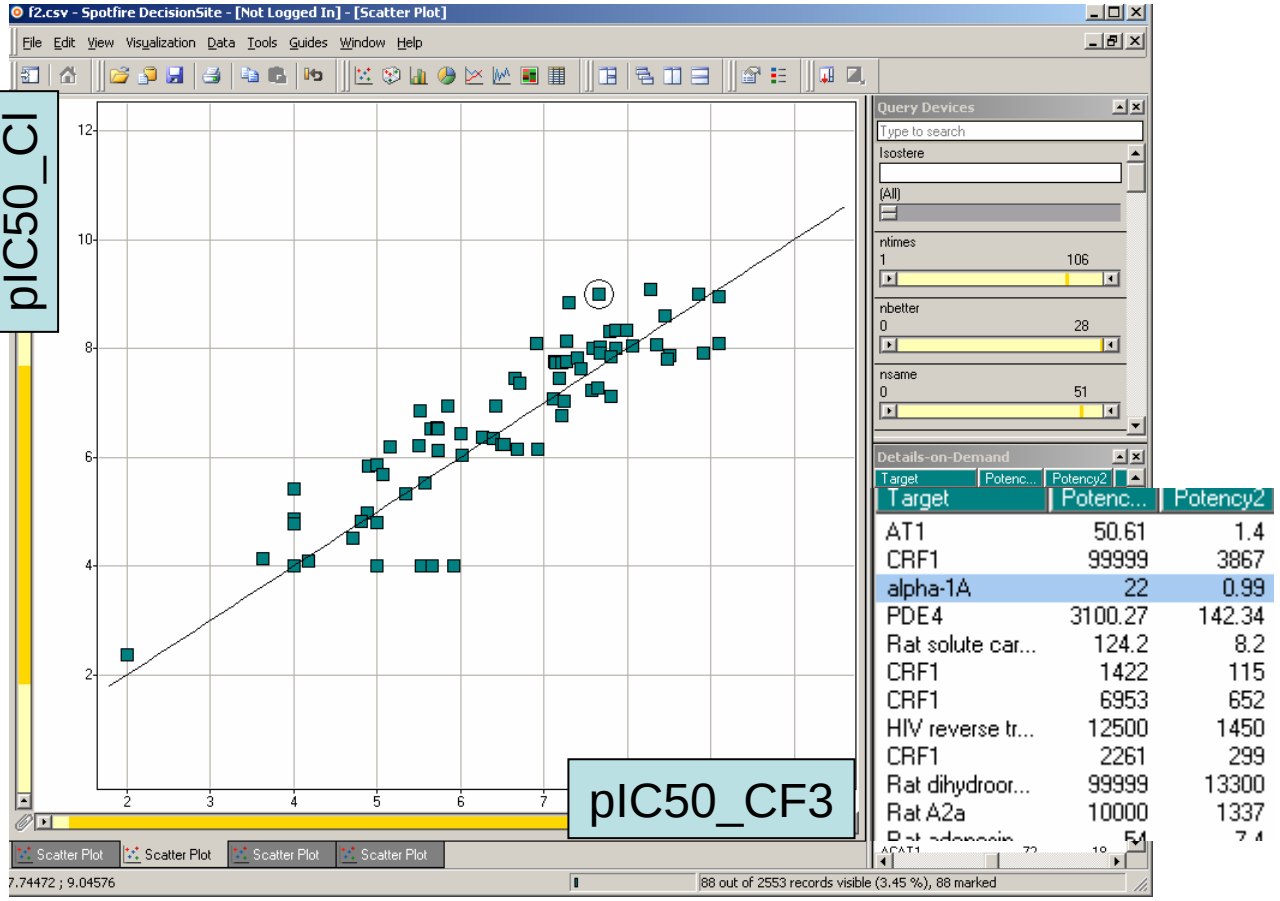
*Li = linker atom



- *o*-Cl is statistically most likely to at least retain potency

Drilling down: *o*-CF3-Ph to *o*-Cl-Ph

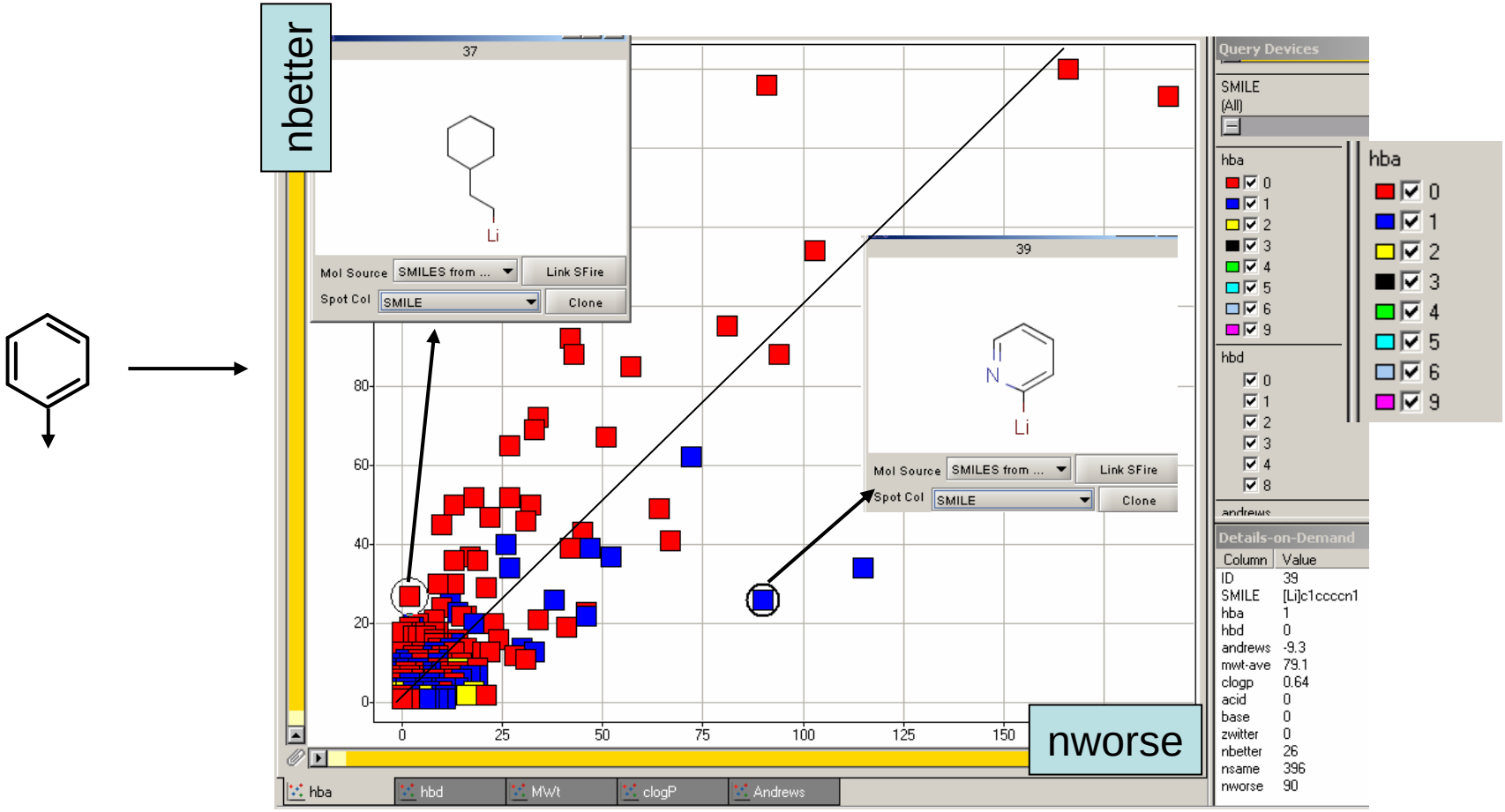




Can drill further down to target or family level

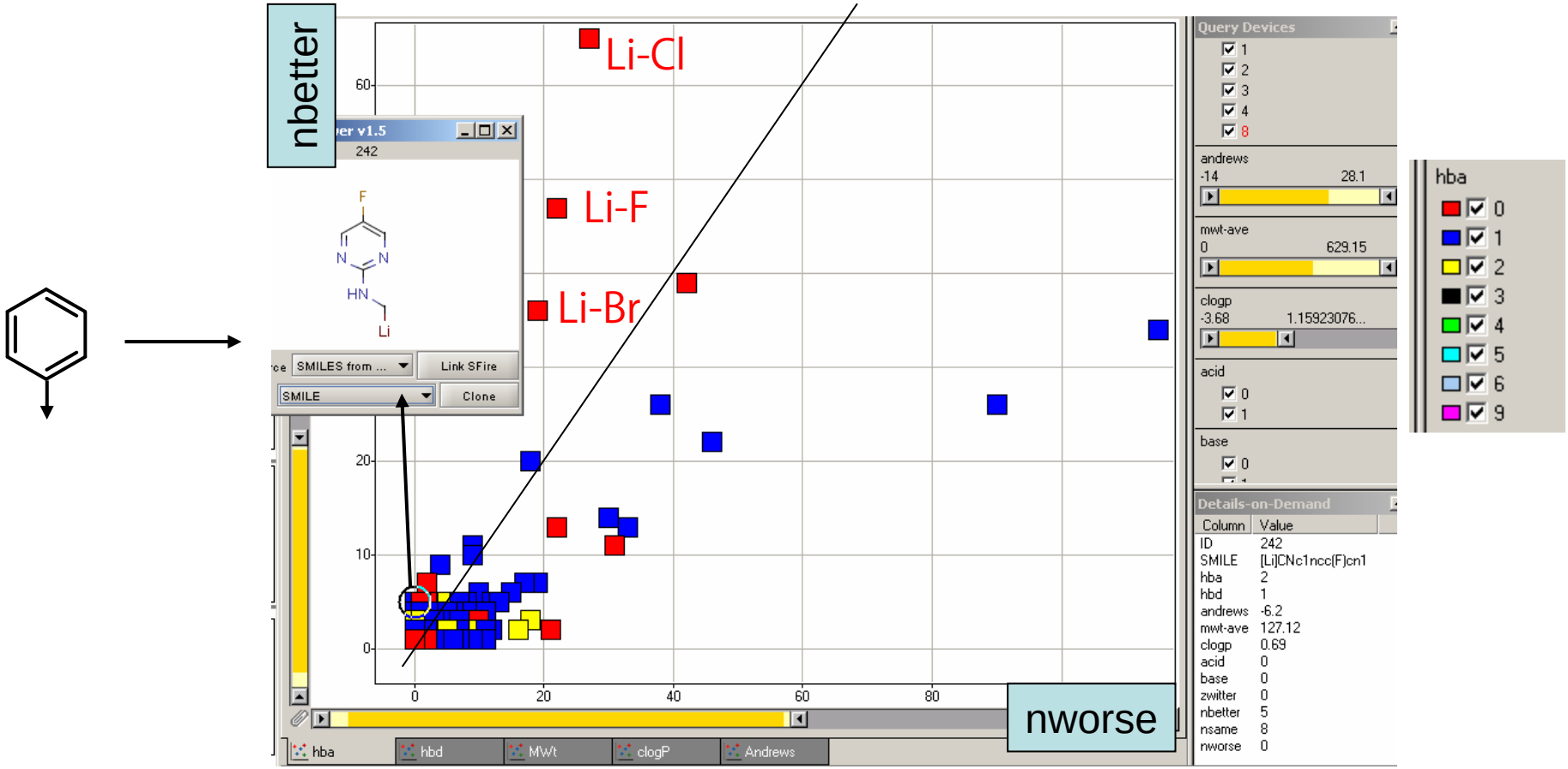
- Appears that *o*-Cl generally is “better” than *o*-CF3

Example: phenyl isosteres



- Lipohilicity is successful at increasing activity
- Pyridines are least successful at increasing potency

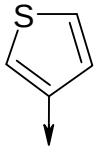
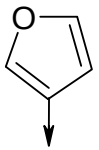
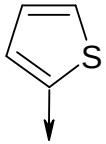
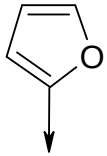
Lower-clogP phenyl isosteres



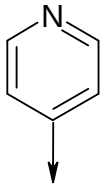
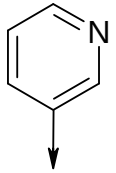
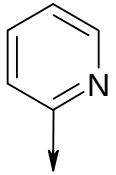
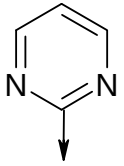
- Halogens are the obvious replacements
  - but there are others that could be interesting

Phenyl to heterocycle

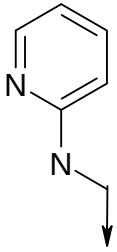
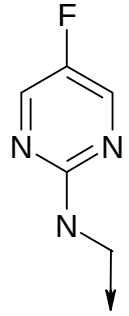
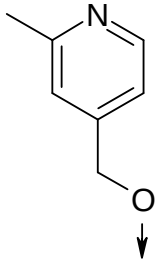
Apolar hets

	n	As good	Better
	82	93%	11%
	45	96%	18%
	196	95%	7%
	129	95%	8%

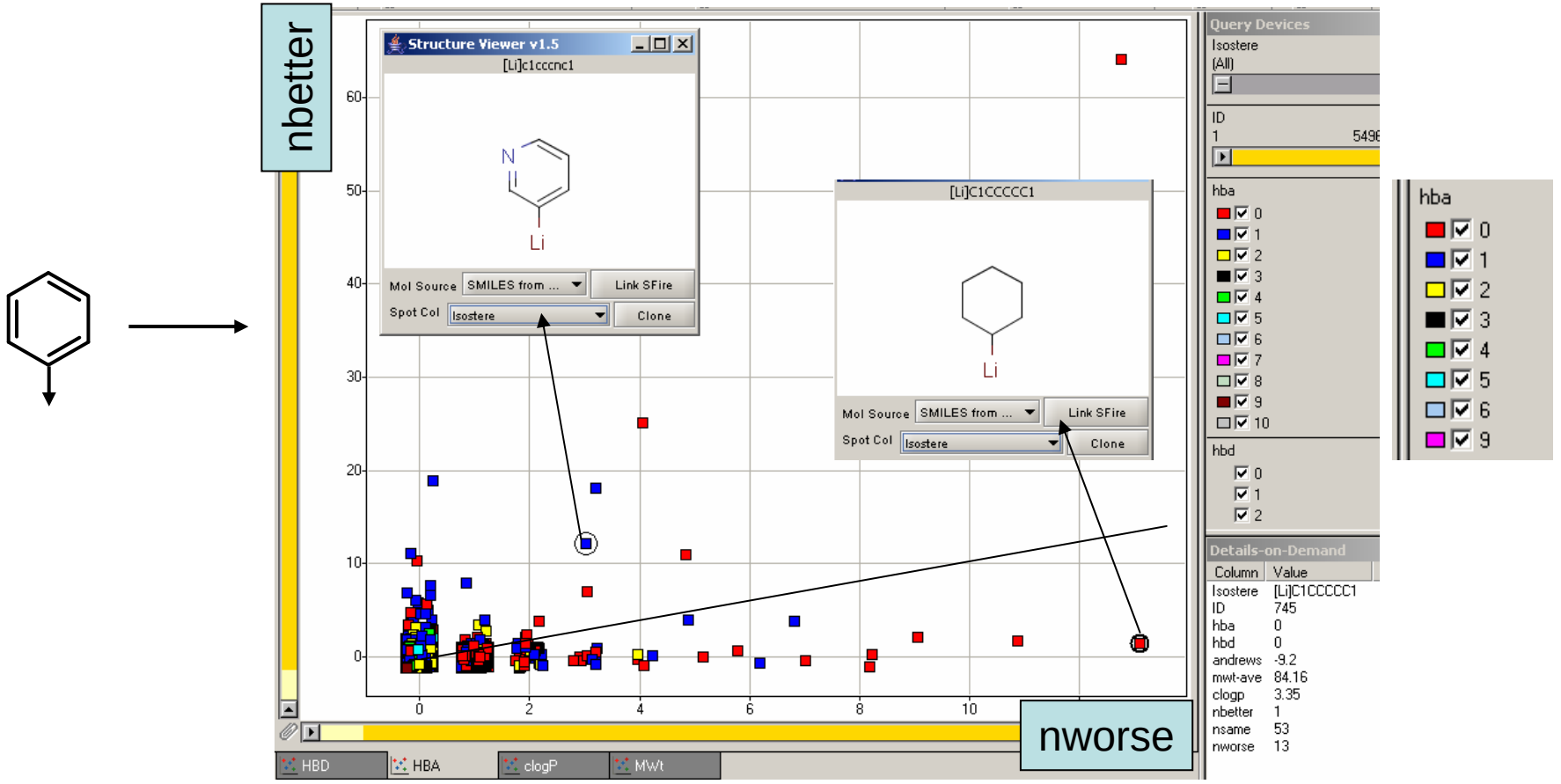
Polar hets

	251	85%	10%
	499	77%	7%
	422	79%	6%
	47	62%	6%

"Best" polar hets

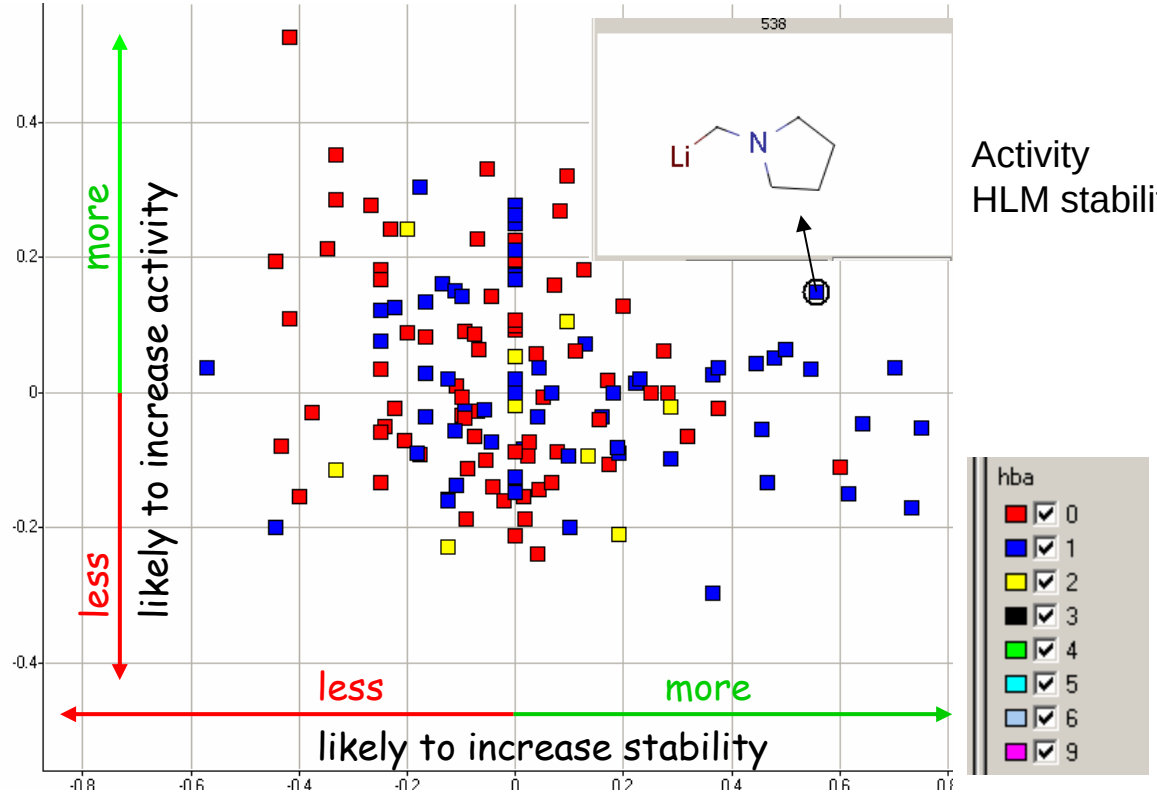
	12	100%	42%
	13	100%	38%
	8	100%	50%

Phenyl isosteres and metabolic stability



- Polarity protects against metabolism

Isosteres improving potency and stability?

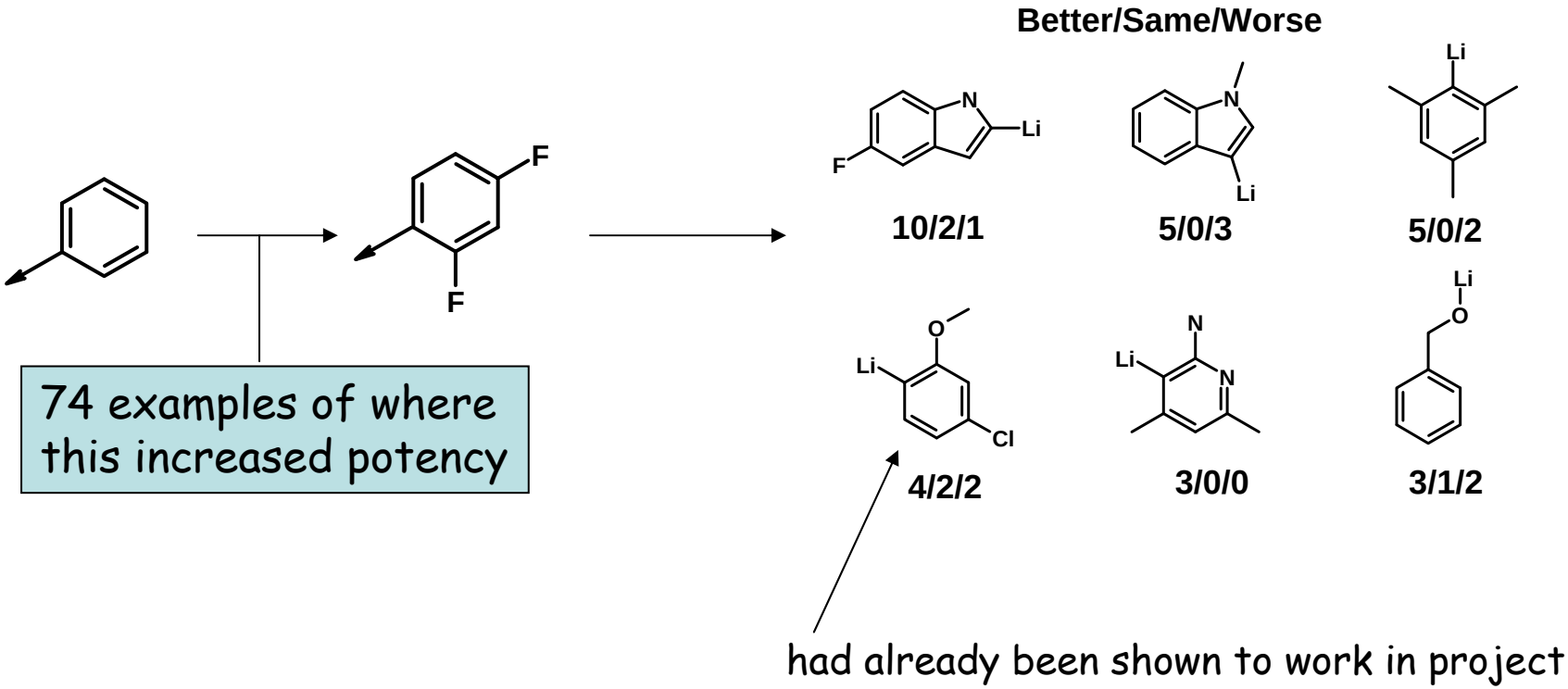


	#pairs	As good	Better
Activity	27	59%	26%
HLM stability	9	100%	56%

- Rare, but there are some suggestions here

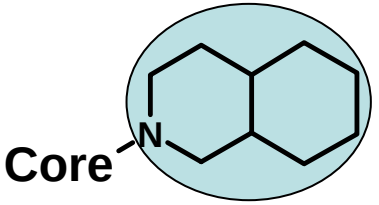
Application of SWAP to design

- In our project, Ph to 2-4-diF-Ph improves potency
- Which changes further improved potency in other projects?
  - *i.e.* build your own Topliss trees



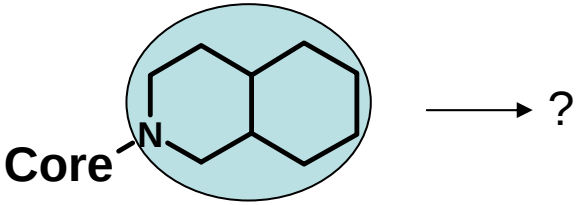
SWAP applied to HTS data

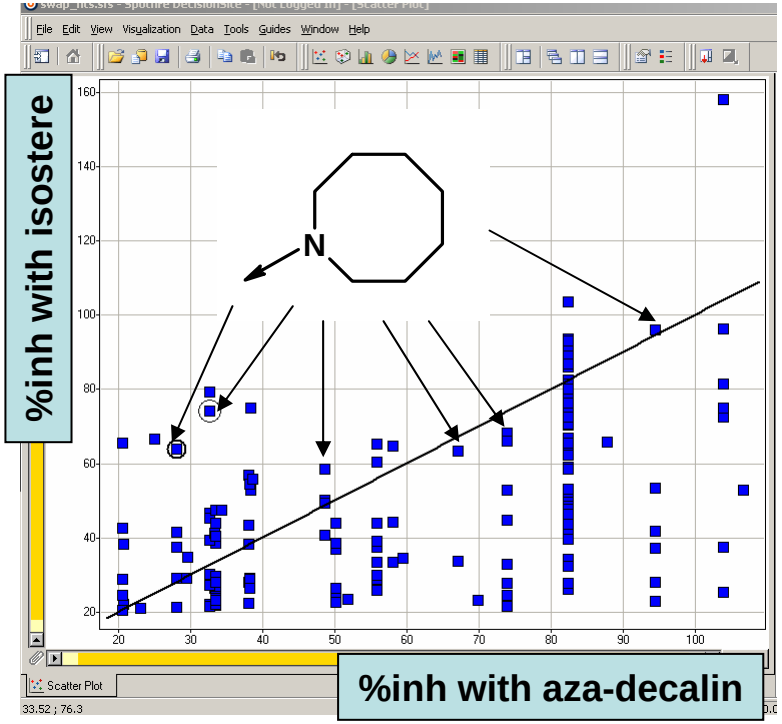
- Can run SWAP on any dataset



- HTS identified cpd as hit for target X
- Library follow-up: looking to replace aza-decalin
- Run SWAP on whole target X HTS?
  - More data than IC50 dataset
    - albeit lower quality
  - All pertinent to target X
    - assuming aza-decalin binds in same subpocket

Isosteres from HTS data

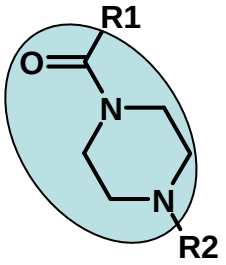


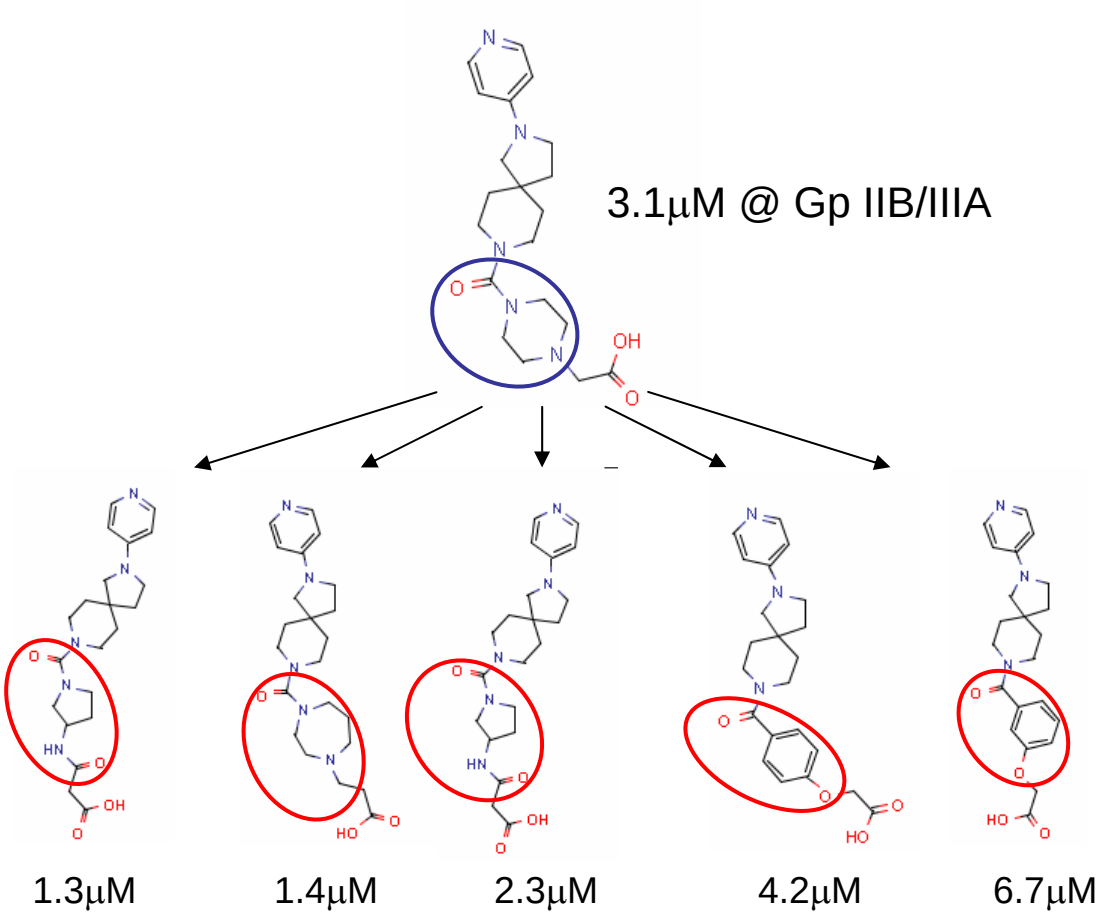


- Suggested a number of replacements
  - Incorporated into library design

Linker isosteres

- SWAP with multiple attachment points
  - Identify alternative cores/templates/linkers





Isosteres from screen data: a summary

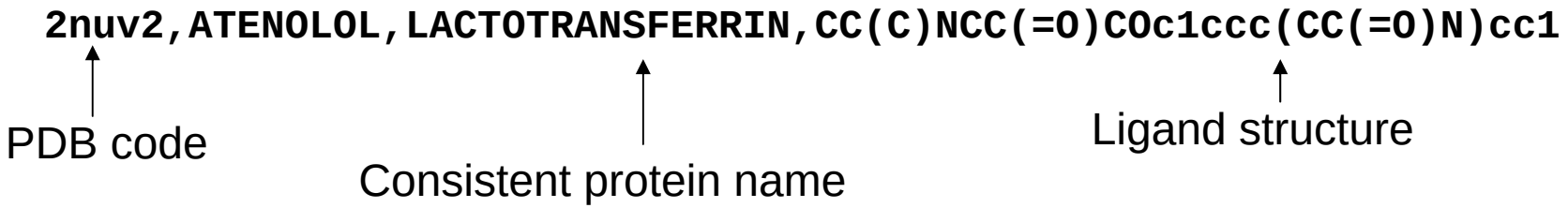
- Can generate ideas from isosteres by pairwise analysis
- Novel “right answers” rarely stand out
  - *e.g.* pick the most successful statistically?
    - Likely to give non-specific activity
- Often requires further mining
  - Generate more appropriate slices of data
    - *e.g.* bespoke Topliss tree design
  - Use more relevant datasets
    - *e.g.* HTS, ADMET data

* Please be aware that past performance is not necessarily indicative of future results

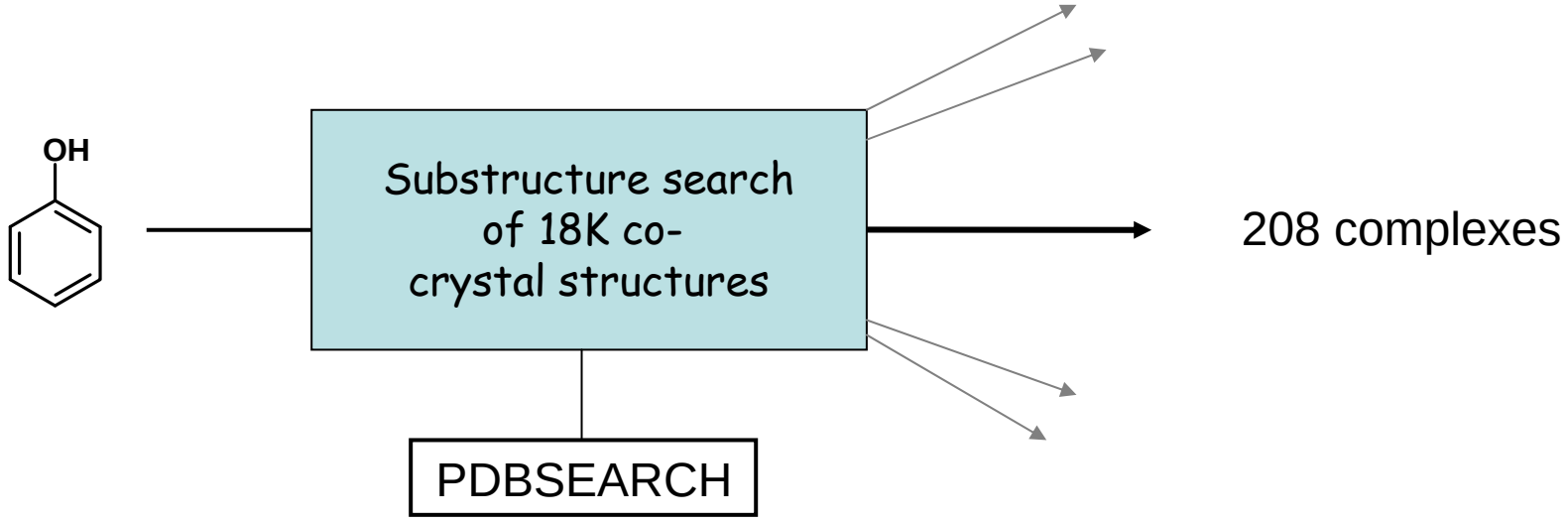
Bioisosteres from crystal-structure analysis

- Isosteres = groups that occupy the same pocket
- Need not come from ligands in the same series
- Require crystal structures of ligands bound in their pockets

- Data underpinning this:
  - 18K ligand/protein complexes (Pfizer+Protein DataBank)

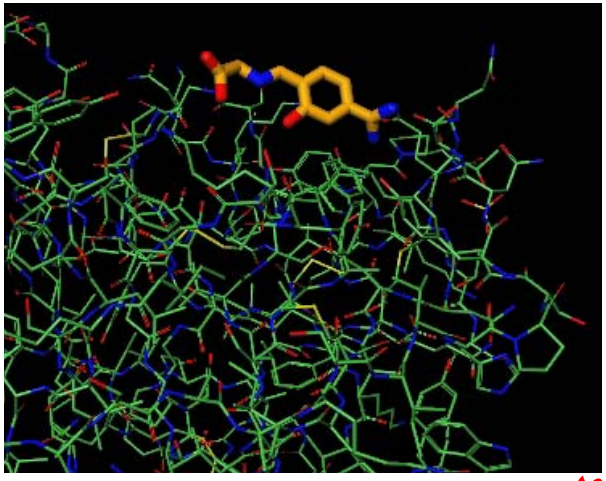


Bioisosteres from crystal-structure analysis



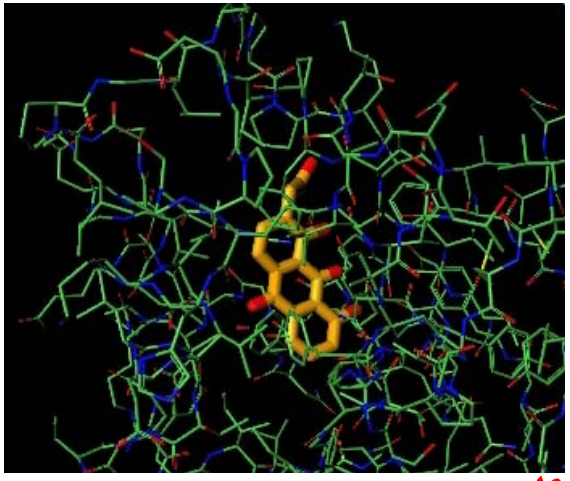
PDBSEARCH on phenols

- PDBSEARCH finds 208 phenols bound in PDB
  - 79 different proteins
  - Triage on basis of ligand burial, interaction patterns…



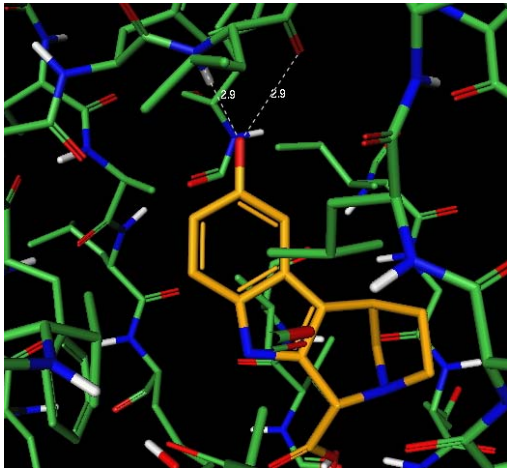
✘

Trypsin
177 other structures in PDB
Not buried



✘

ACTVA-ORF6 monooxygenase
Only 4 other structures in PDB
Atypical phenol (quinone tautomer)



✓

CDK2
388 other structures in PDB
Buried and interacting

  - In this case, 26 proteins selected

Triaging phenols in PDB

- Some degree of automation possible

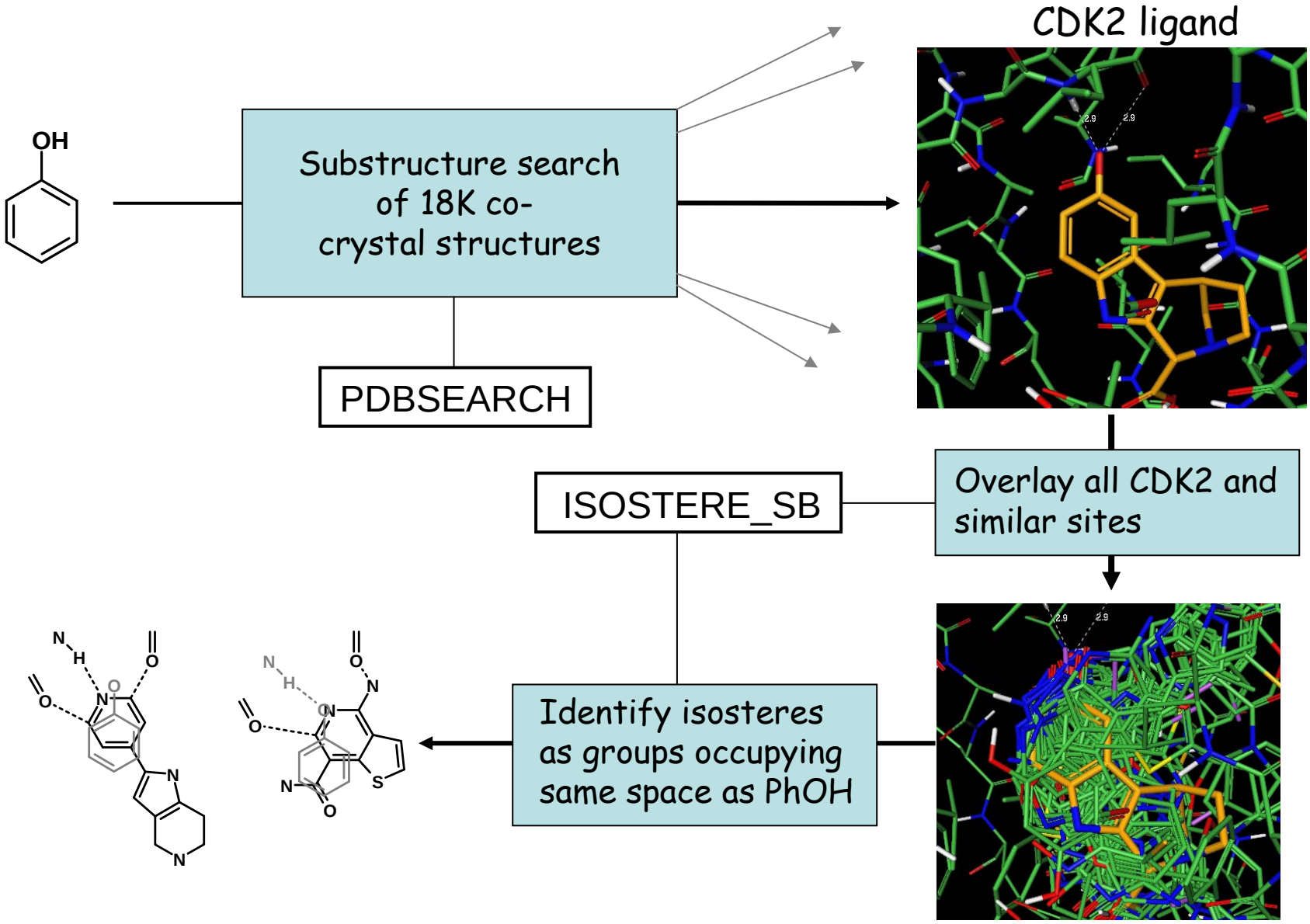
Ligand not buried: discard

Only one other non-phenol in this protein: discard

	A	B	C	D	E	F	G	
1	PDB	Protein_Pocket	EC	Burial	H-bonds	Ngrps_poc	Ncplx_prote	Ligand
2	1n5v1	ACTVA-ORF6 MONOOXYGENASE_1	NOEC	0.99	5H;6H	4	5	7-HYDR
3	1n5q1	ACTVA-ORF6 MONOOXYGENASE_1	NOEC	0.97	3H;5H	4	5	4-DIMET
4	1n5s1	ACTVA-ORF6 MONOOXYGENASE_1	NOEC	0.96	5H;6H	4	5	(1'8-DIH
5	1n5t1	ACTVA-ORF6 MONOOXYGENASE_1	NOEC	0.99	5H;6H	4	5	(1'8-DIH
6	1g3c2	BETA-TRYPSIN_1	3.4.21.4	0.6	5H;5H;5H;5H	1	178	2-(4-CAR
7	1g3e1	BETA-TRYPSIN_2	3.4.21.4	0.94	5H	1	178	2-(4-CAR
8	Pdww	CDK2_1	2.7.1.37	0.9	2A	6	394	PF-0237
9	Pcro	CDK2_1	2.7.1.37	0.91	4A	6	394	PF-0266
10	Paek	CDK2_1	2.7.1.37	0.92	2A;5H;5H	6	394	AG-0127
11	Paep	CDK2_1	2.7.1.37	1	2A;5H	6	394	AG-0127
12	Pdzj	CDK2_1	2.7.1.37	0.91	3A	6	394	PF-0319

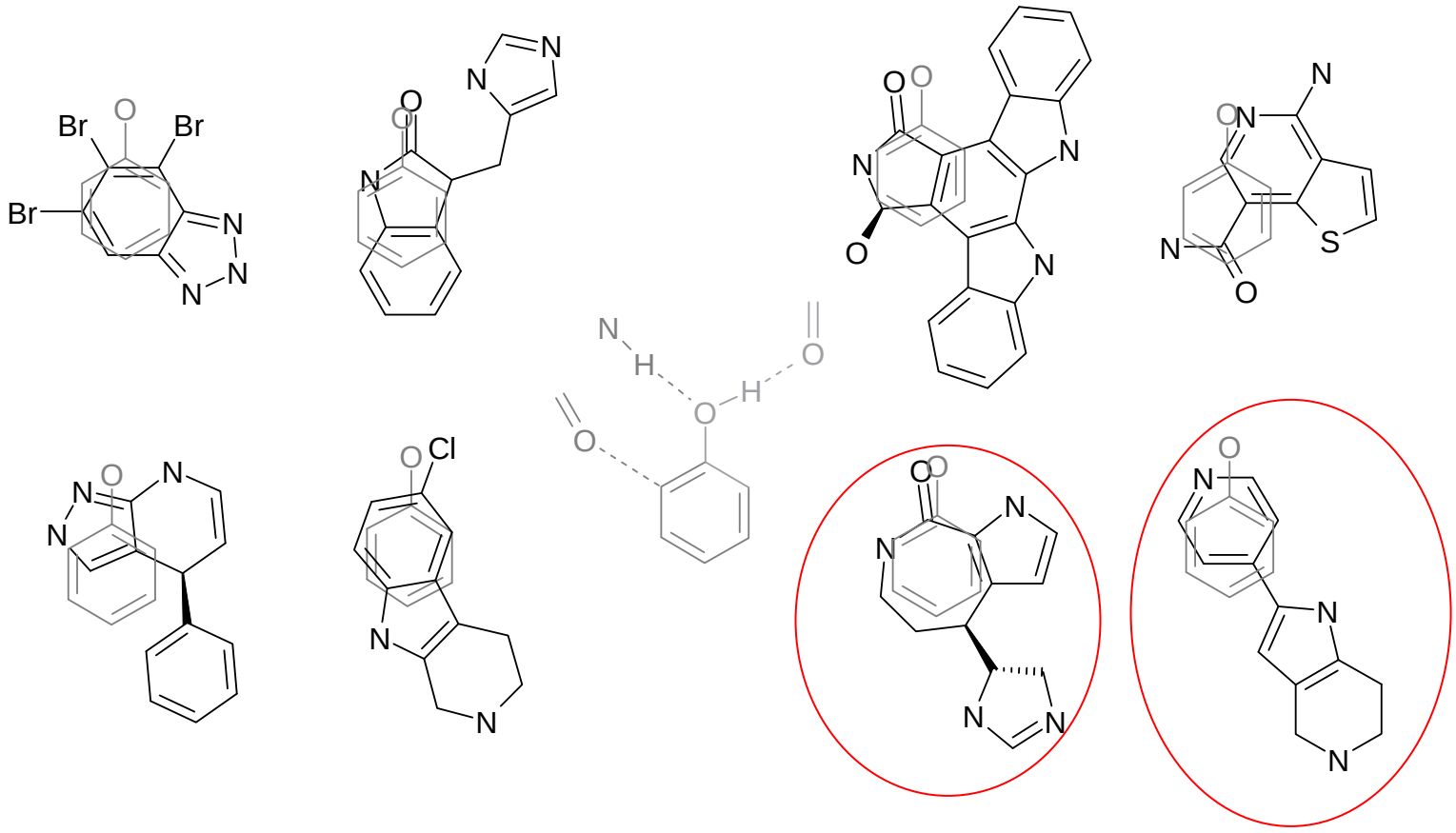
Well buried, makes 2 H bonds and an Aromatic stack, many more complexes involving this protein: keep

Structure-based bioisosteres

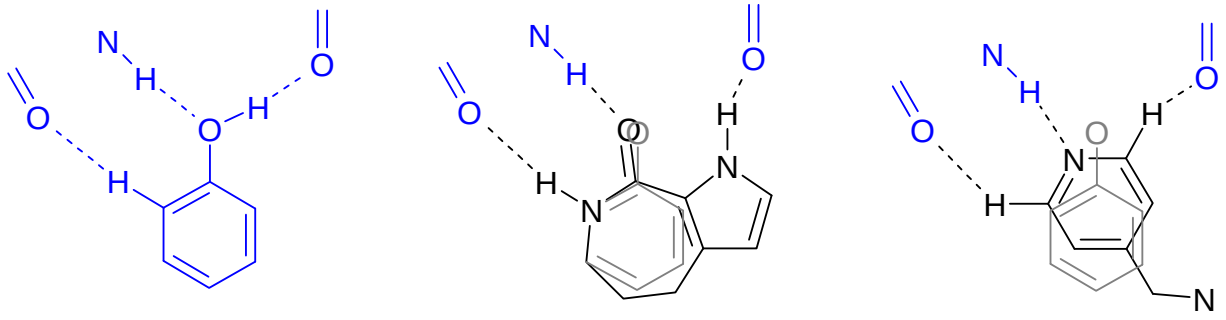


Phenol isosteres from PDB

- Large variety of isosteres



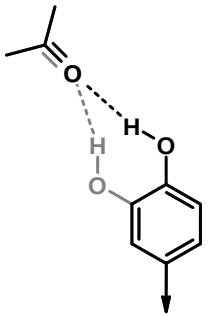
Interactions

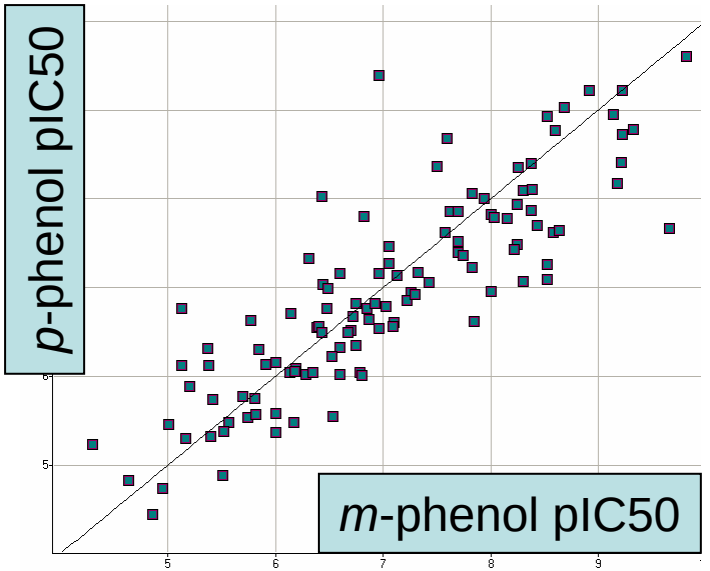


- Power of this method: *how* the group is isosteric
  - How they overlay and which interactions are mimicked
- Spatial orientation: not an atom-based overlap
  - Not always what you expect
  - Could explain inconsistency in isostere performance

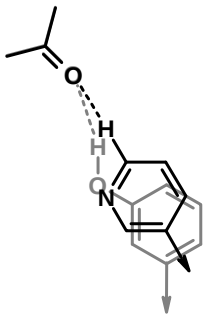
Effect of overlap on isostere consistency

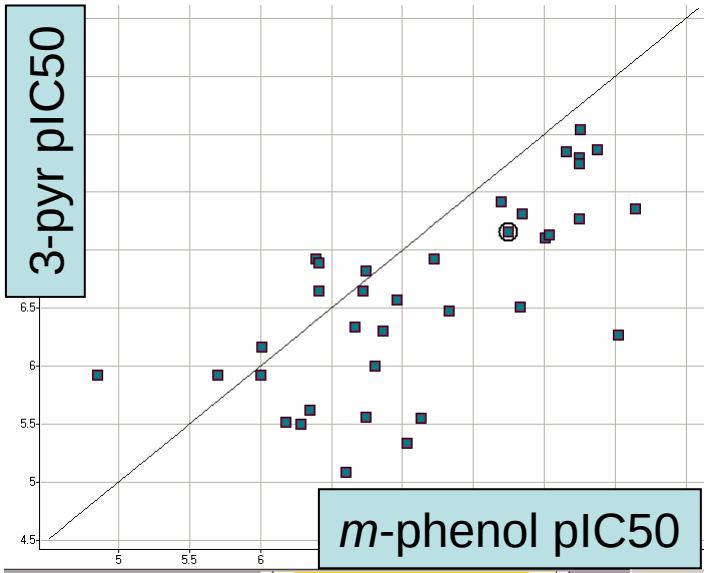
- *m*- and *p*-phenol
- Can overlap linker vector





- *m*-phenol and 3-pyridyl
- Can't overlap linker vector





Summary

	Ligand-based	**Structure-based**
# actives in database	1400K	18K
Activity values known?	Yes	No
Overlap of isosteres known?	Think so	Yes
Do we know they're buried?	No	Yes
Direct isostere incorporation?	Yes	Sometimes
Draws from data across series?	No	Yes
User input required?	No	Yes
Wildcards allowed in search?	No	Yes

- Both methods serve as idea generators
  - Often get complementary results
- Allows us to make use of all our historical data

Acknowledgements

- Lee Harland
- Anna Gaulton
- Chris Kibbey
- Julian Blagg
- Dave Hepworth
- Paul Fish
- Lourdes Cucurull-Sanchez
- Jens Loesel
- Sarah Skerratt
- Alan Stobie
- Mark Gardner
- David Millan
- Samantha Hughes
- Michelle Styles
- Paul Brennan
- Charlotte Lane
- Chris Barber
- Andy Bell
- Wolfgang Klute
- Alan Brown