Fragment-Based Screening, What can we learn from published hits?

A work in progress……

Chris Swain

Cambridge MedChem Consulting

Fragment-Based Screening

- Fragment-based screening has become increasingly popular and has proven to be a viable alternative to high-throughput screening.
- Fragment space is smaller
  - A million compounds cover only a small fraction of the suggested 10^{60} Chemical Space, whilst 2000 compounds can probe much of the 10^{6} Fragment Space
- Hit rates for Fragment-based screening appear to be higher, typically 3-10%.
- Binding Efficiency for small molecules is likely to be higher.

Design of the Fragment Library

- Several approaches have been described in the design of fragment libraries. Most comply with the commonly accepted Astex "Rule-of-Three"
  - MW <300, H-bond donors/acceptors <=3, cLogP <3.
- Solubility is key requirement since screening carried out at higher concentrations
  - Often overlooked
- Rather than simply cull available molecules there have been recent attempts to design libraries based on known drugs, PDB ligands, natural products, or enhanced 3D structure.

- Can we use the information from fragment hits reported in the literature to help design fragment libraries?

What can we learn from known fragment hits?

- Compile database of published hits from fragment screens. (Store as SMILES).
- Also include:-
  - Screening technology
  - Target and Uniprot ID, affinity (how measured), PDB code
  - Target type/class, using ChEMBL ontology
- Calculate
  - Physicochemical properties
    - cLogP, cLogD, PSA, HBA, HBD, RotB, pKa, shape descriptors, MR, HAC, fraction aromatic heavy atoms. (ChemAxon, MOE)
  - Functional groups (Checkmol)
  - Cluster analysis

Current Status (1 May 2015)

- 213 Publications
- 1036 Published hits
- 152 Different targets
- 23 Detection technologies

- Finding the data is getting more of a challenge, it seems as fragment screening becomes more mainstream it is often not mentioned in the title or abstract.

Frag_Database

1 — Records — 64 / 706 Found (Unsorted) — Show All — New Record — Delete Record — Find — Sort

Layout: OverView — View As: — Preview — Edit Layout

ChemDraw | smiles_CD | Data | structure — SSS Smiles — Find_List: ID_000006 — SMILES_Query: CCCC

Groups

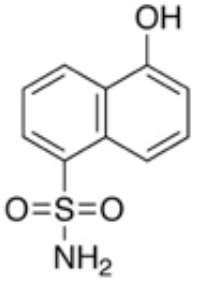
ID_NUM ID_000003 — openbabel_smiles Oc1cccc2c1cccc2S(=O)(=O)N

ChemSpiderID 78505 — ChEMBLID 457047 — inchikey NFVBVKHGDDDCEA- — Structure

Detection_Tech X-Ray
Target CDK2
Target_Type Enzyme
Target_Class kinase
Target Subclass ser-thr
Reference J. Med. Chem. 2008, 51, 4986–4999

UniprotID P24941 — PDB Code 2vth — DOI http://dx.doi.org/10.1021/jm800382h

Affinity 120 uM
AffinityAssay BioAssay



ChemAxon Calc	
ChemAxon_logP	1.27
ChemAxon_logD	1.26
ChemAxon_Mass	223.25
ChemAxon_acceptorcount	3
ChemAxon_donorcount	2
ChemAxon_PSA	80.39
ChemAxon_Rot_bond_count	1
ChemAxon_Most_acid_pka	9.33
ChemAxon_Most_basic_pka	-5.6
ChemAxonFractionArom	0.67
ChemAxon_HAC	15
acid	0
basic	0
neutral	1
a_b_n	neutral

MOE Calc	
SlogP	1.1928
b_ar	11
logS	-3.1274199
a_acc	3
a_don	2
a_hyd	10
b_1rotR	0.0625
mr	5.7425342
ASA	384.48505
ASApositive	189.64247
ASAnegative	184.96759
ASA_H	218.11243
ASA_P	166.37262
logP_ow	1.244
Shape	rodlike
Cluster	3

ChemSpider Data	
ACD_LOGP	1.04
ACD_LOGD	1.01
ACD_PKA	8.58
ACD_PKB	
ExptLOGP	
aLogP	1.2
aLogS	
Elemental_Analysis	C, 53.80; H, 4.06; N,
Exact_Mass	223.030314
Molecular_Formula	C10H9NO3S
Molecular_Weight	223.24836

Checkmol Data	
amine	0
carbonyl	0
carboxylic acid	0
aromatic	1
heterocycle	0
aryl halide	0
phenol	1

100 — Browse

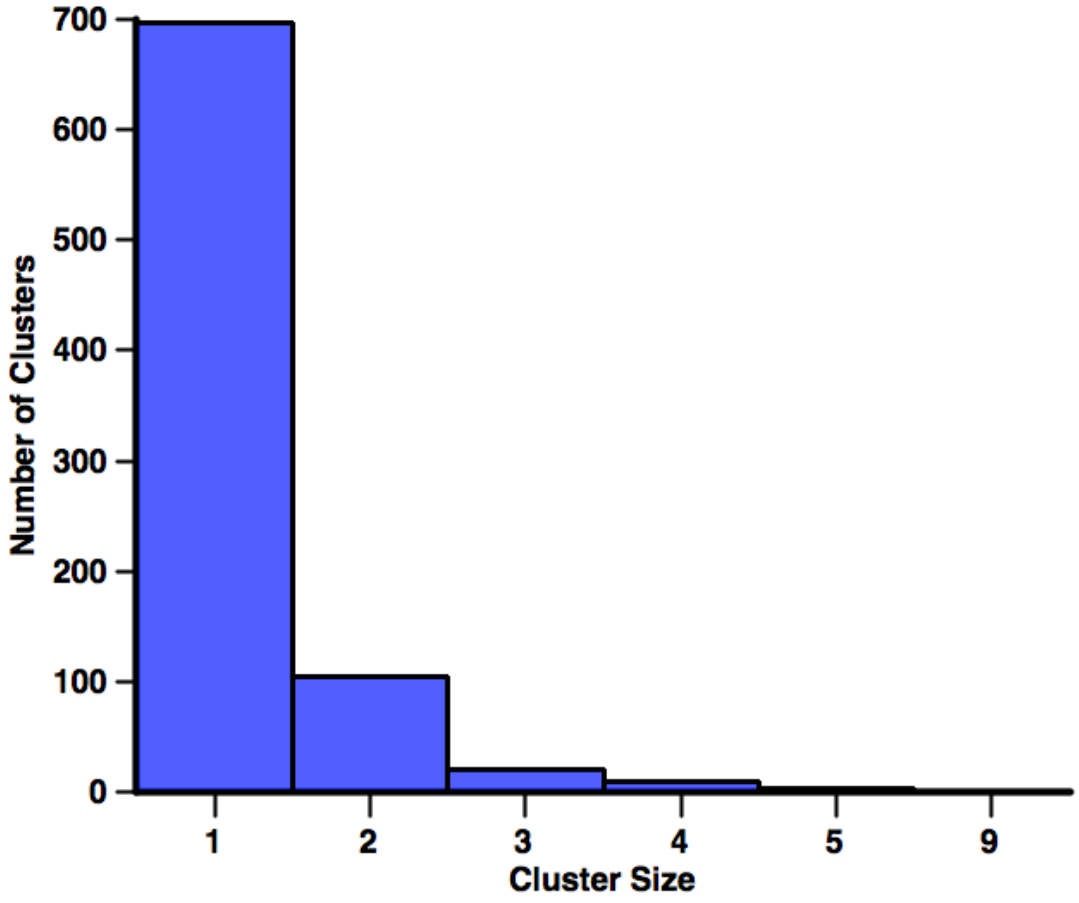
Suppliers of hits

Table 1	PubFragAllData
MaybridgeAll	268
KeyOrganicsBionetPrem	195
Maybridge_2500	164
LifeChemicals_frags	88
Otava	79
Specs	54
KeyOrganicsAll	74
Enamine_frags	48
Prestwick	49
Vitas	39
ChemDiv	28
ChemX	28
TimTec	22
Chembridge	17
Enamine_Golden	5
LCZenobia	5
Asinex	4
3DFragConsortium	2
WuXi	0
Pyxis	0
Infarmatik3D	0
Analyticon	0

Maybridge are the most popular supplier
First major supplier to check solubility of fragments

Diversity

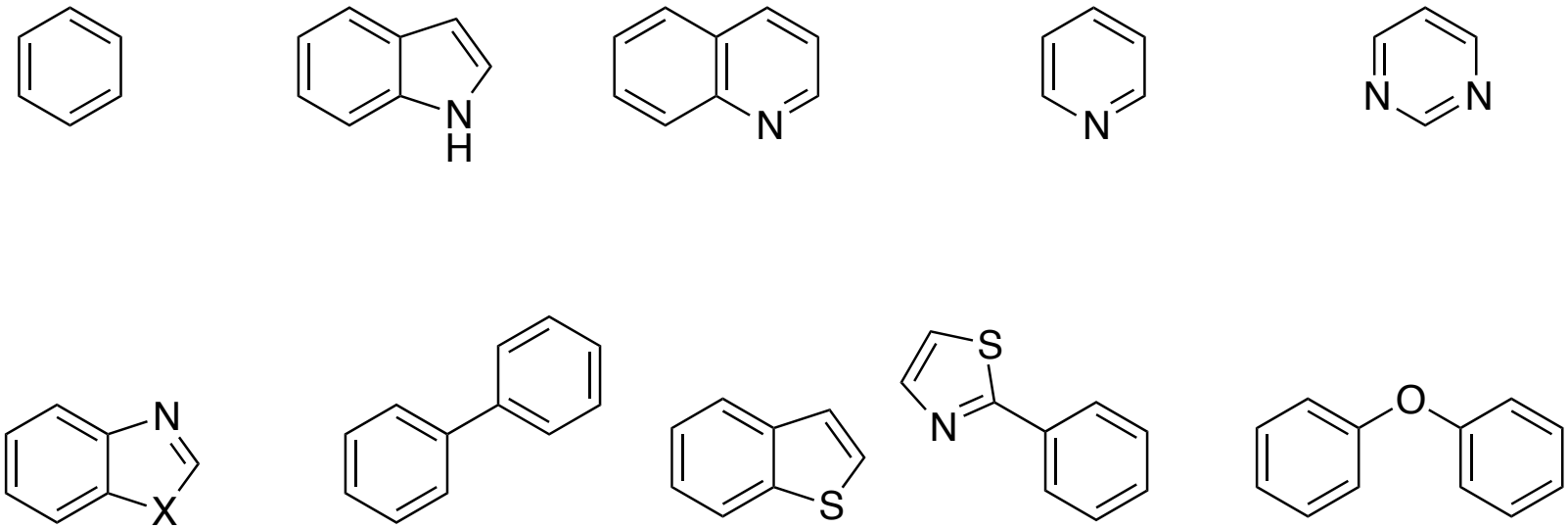
- Clustered using MACCS fingerprints in MOE. Tanimoto 0.85
- Majority are singletons
- Diverse fragments for same target
- Most fragments have sparse fingerprints



Functional Group Analysis

- 990/1036 contain an aromatic ring, 836 of which are heterocyclic
- 214 contain an arylhalide, 112 contain a phenol
- 195 contain an acidic group, 189 a basic group
- 26 contain a nitro group
- 178 contain a hydroxy, 126 an ether
- 416 contain an amine, 192 “anilines” (mainly on heteroaromatic systems)
- 140 amides, 38 esters, 23 ureas

Most common scaffolds

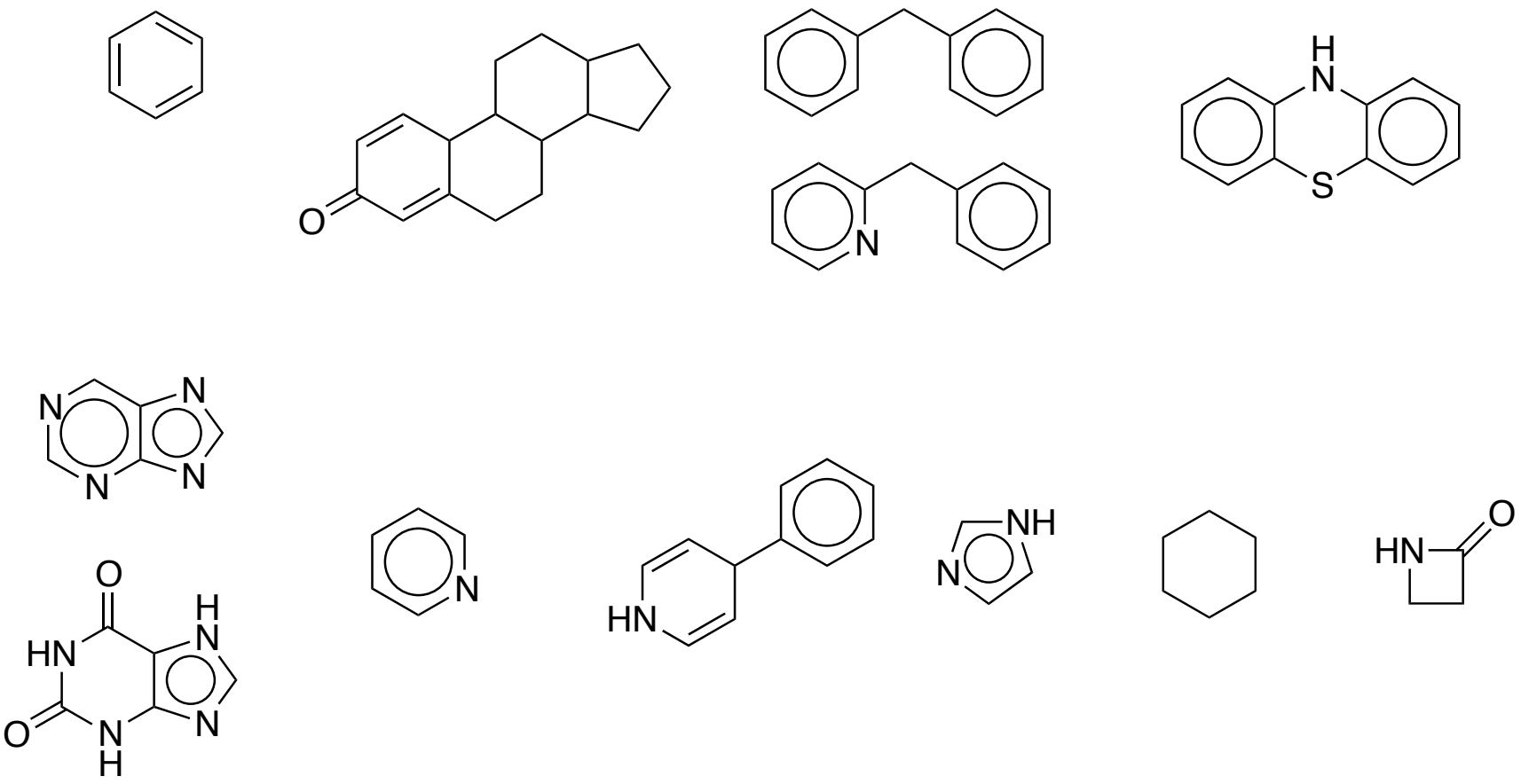


X=S, N

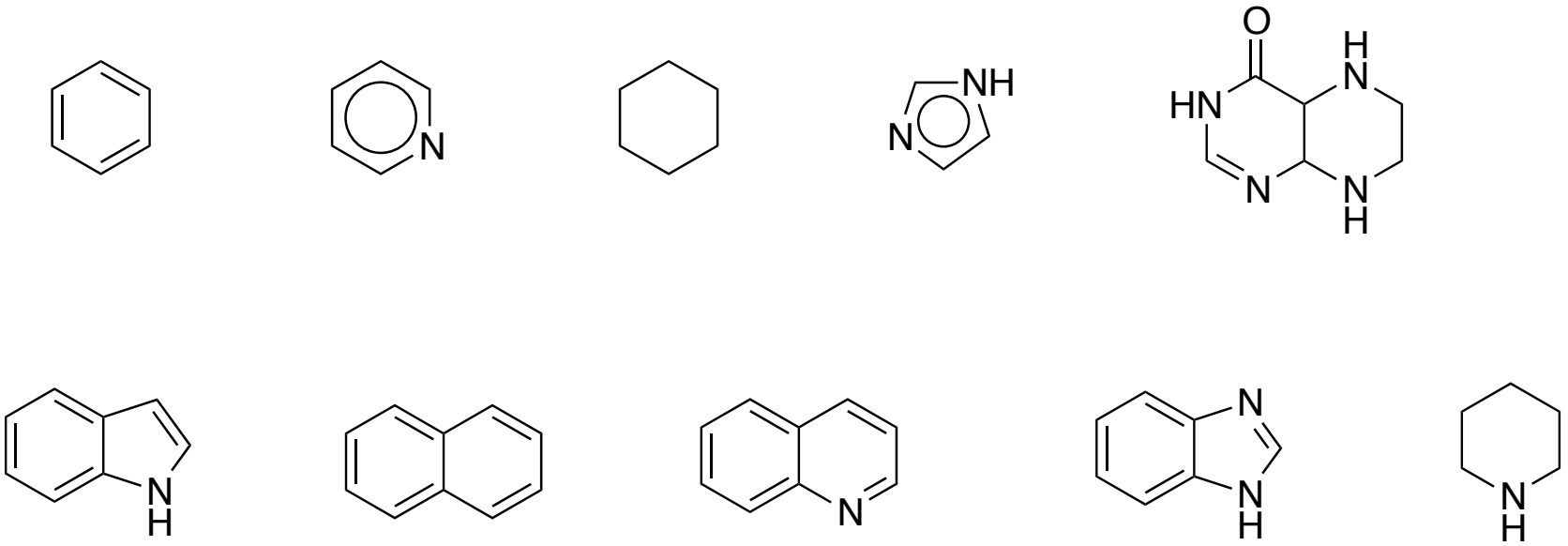
How does this compare with known ligands?

- Compare with
  - DrugBank
  - PDB
  - BindingDB

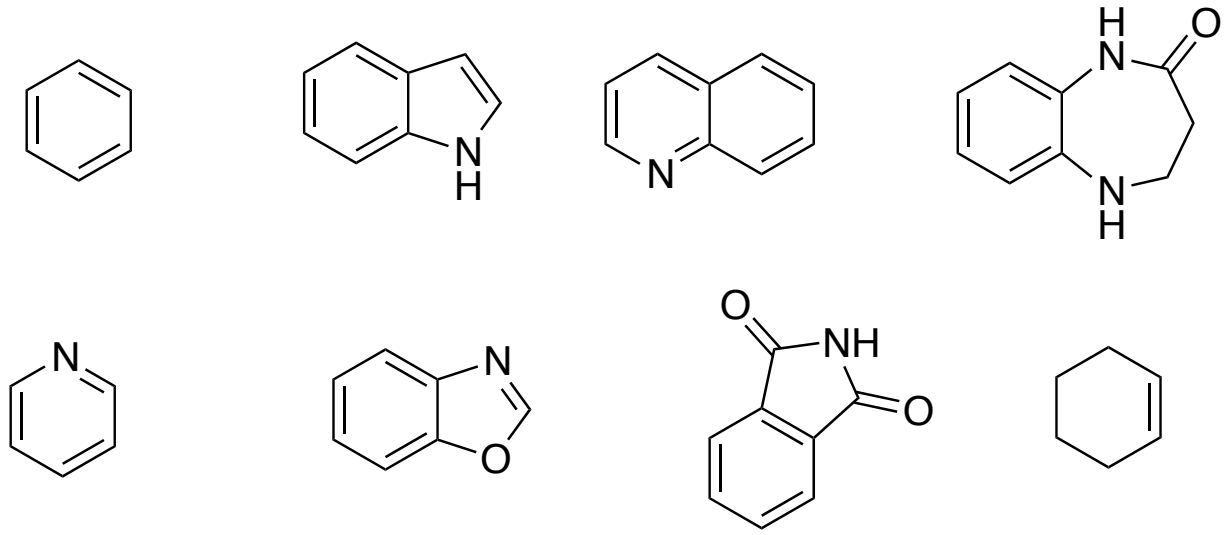
Most common Scaffolds DrugBank



Most common fragments in PDB



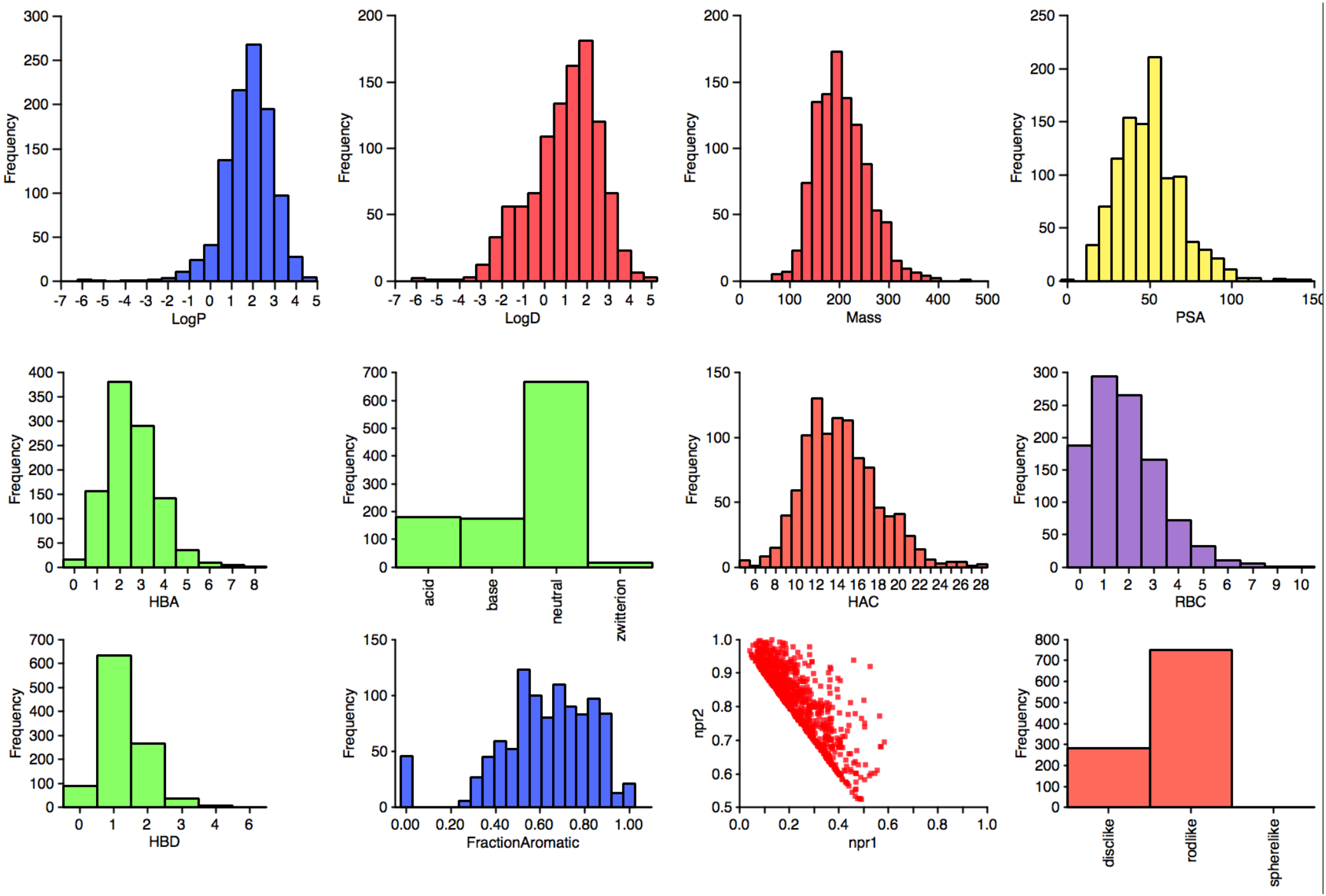
Most common scaffolds in BindingDB



Conclusions

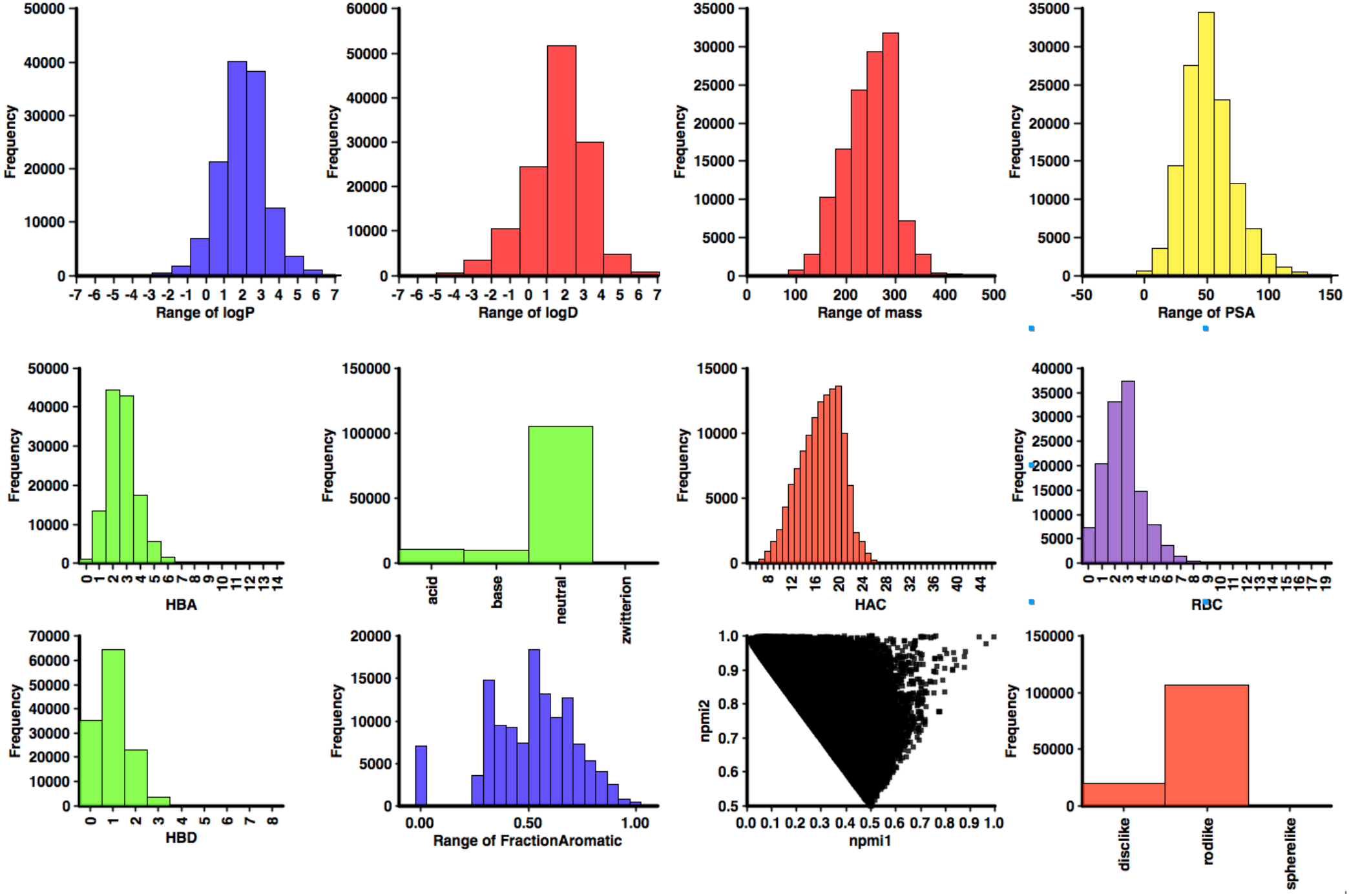
- Analysis of reported fragment hits highlights the preponderance of aromatic systems.
- Exploration of three public data sources of ligands indicates a similar observation.

  - Is there something special about aromatic scaffolds?

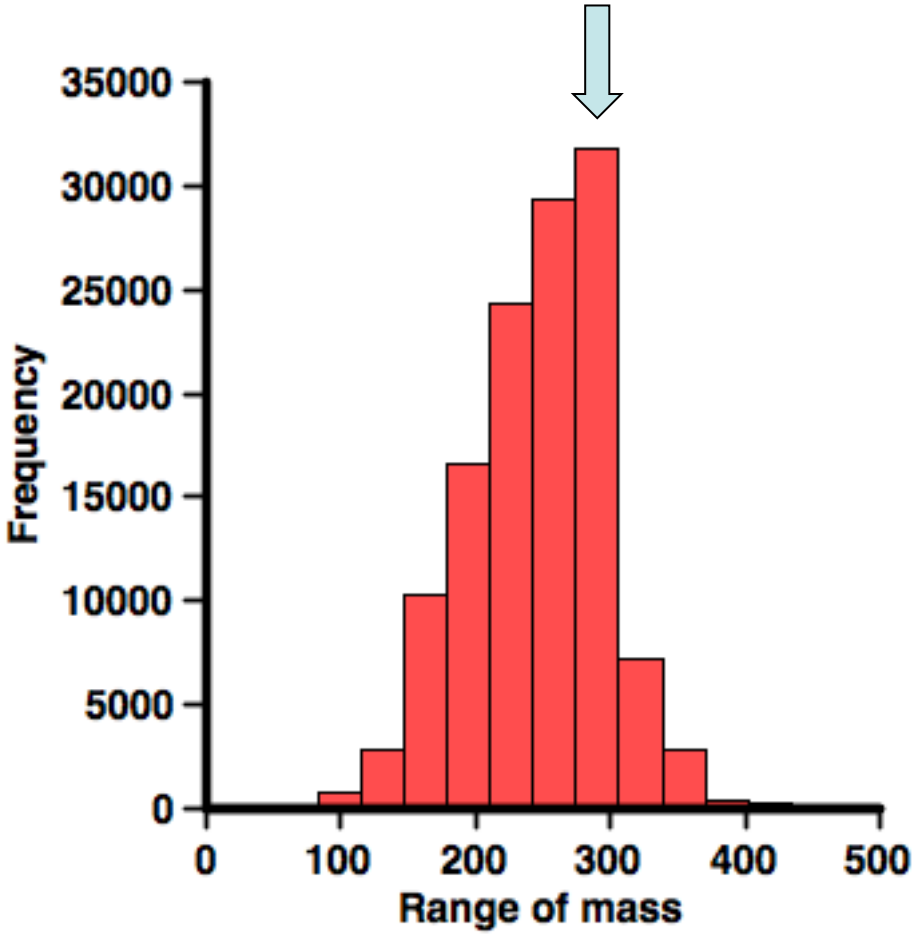

Frequency
LogP
LogD
Mass
PSA
HBA
acid
base
neutral
zwitterion
HAC
RBC
HBD
FractionAromatic
npr2
npr1
disclike
rodlike
spherelike

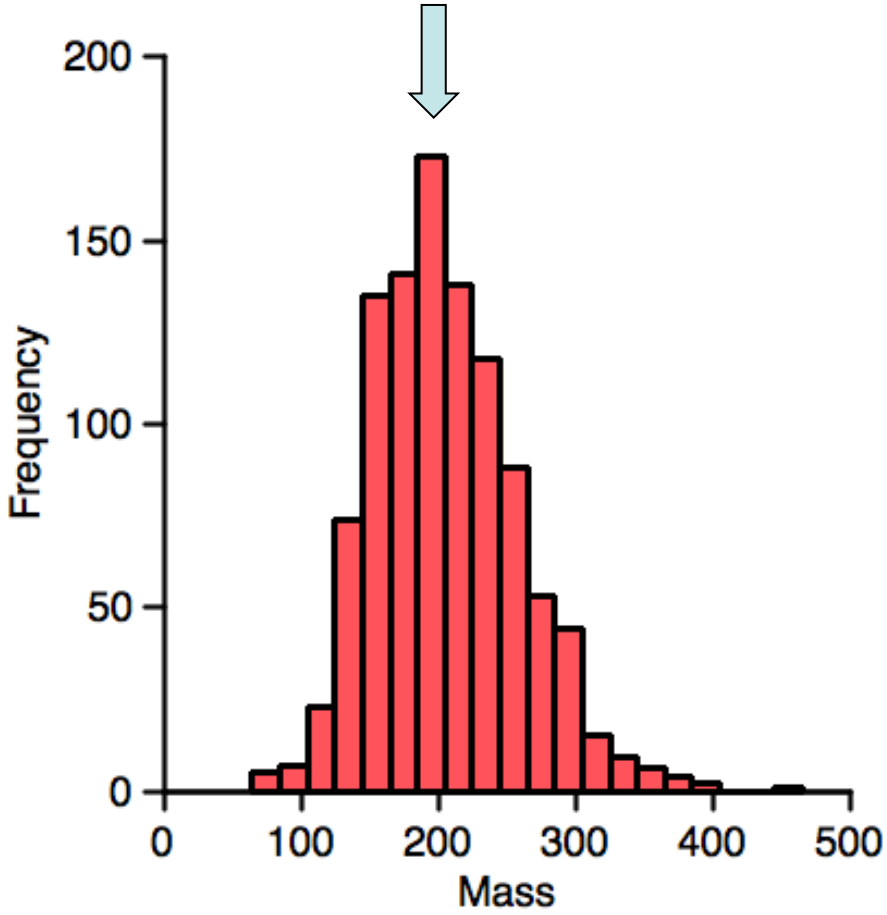
You can only test what is available

- Some papers describe the source of the screening compounds, many do not.
- Looking at the hits we can make a guess at the likely source of the screening collection used.
- Use same tools to calculate profile of putative screening compounds.


Frequency
Range of logP
Range of logD
Range of mass
Range of PSA
HBA
acid
base
neutral
zwitterion
HAC
RBC
HBD
Range of FractionAromatic
npmi2
npmi1
disclike
rodlike
spherelike

Comparison of Molecular Weight

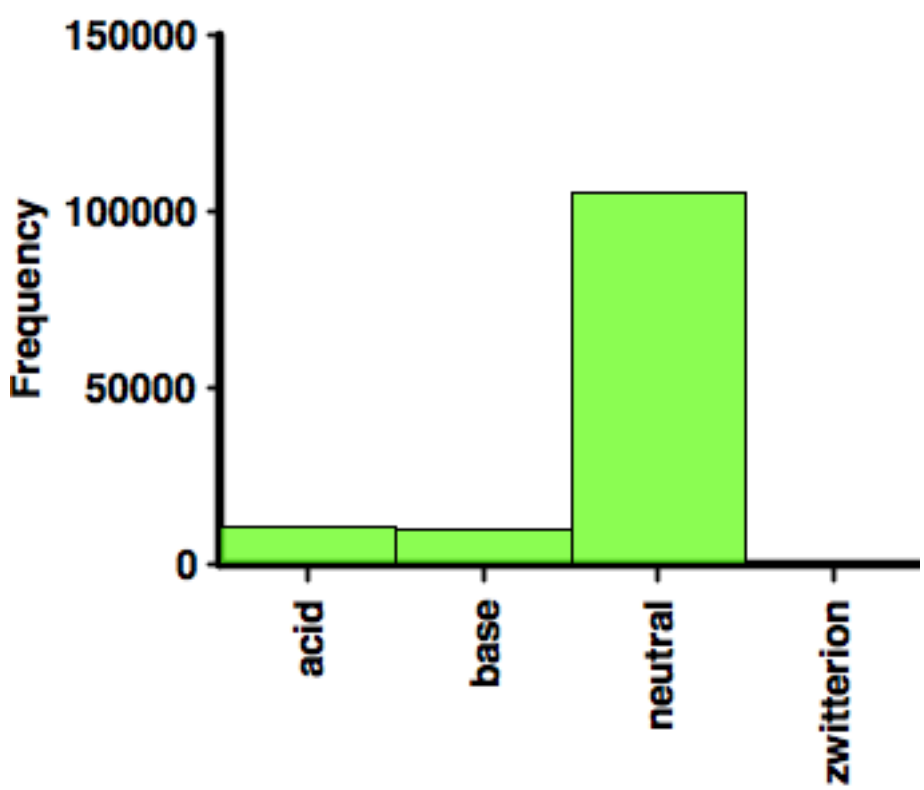


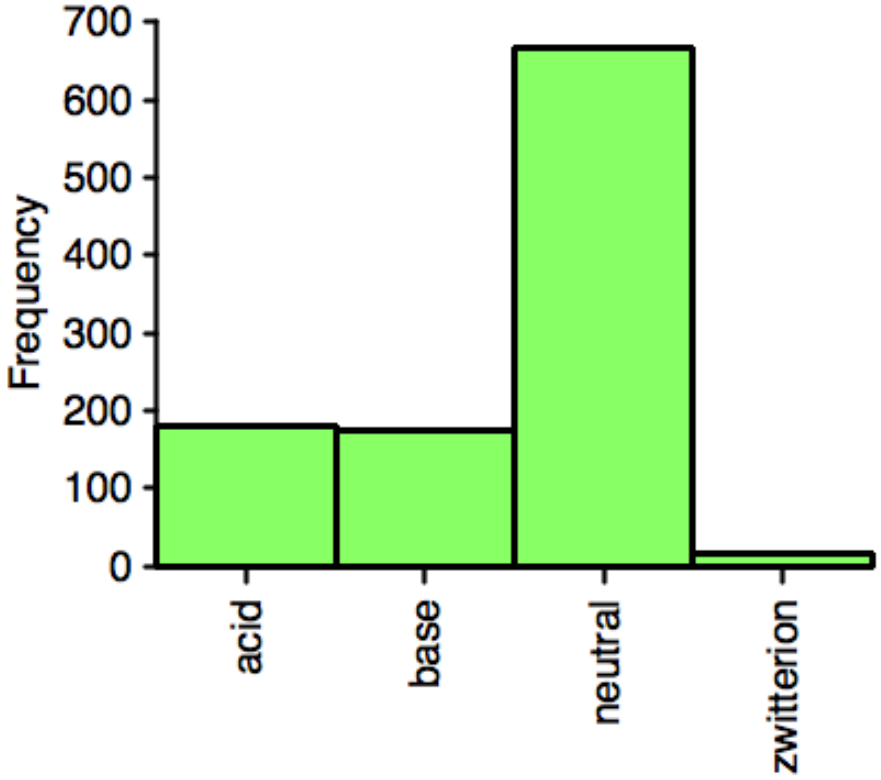


"Screening Collection"

Hits

Comparison of ionisation

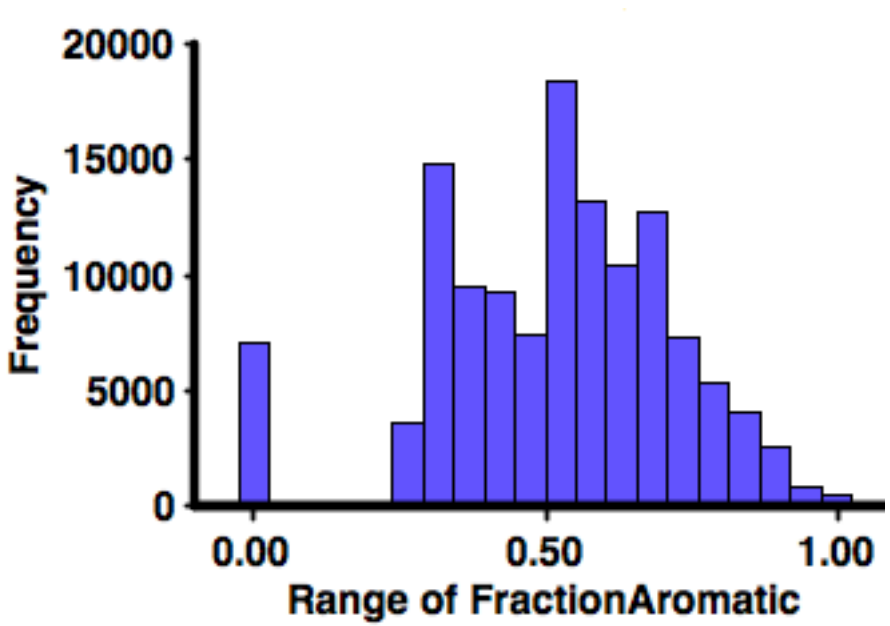


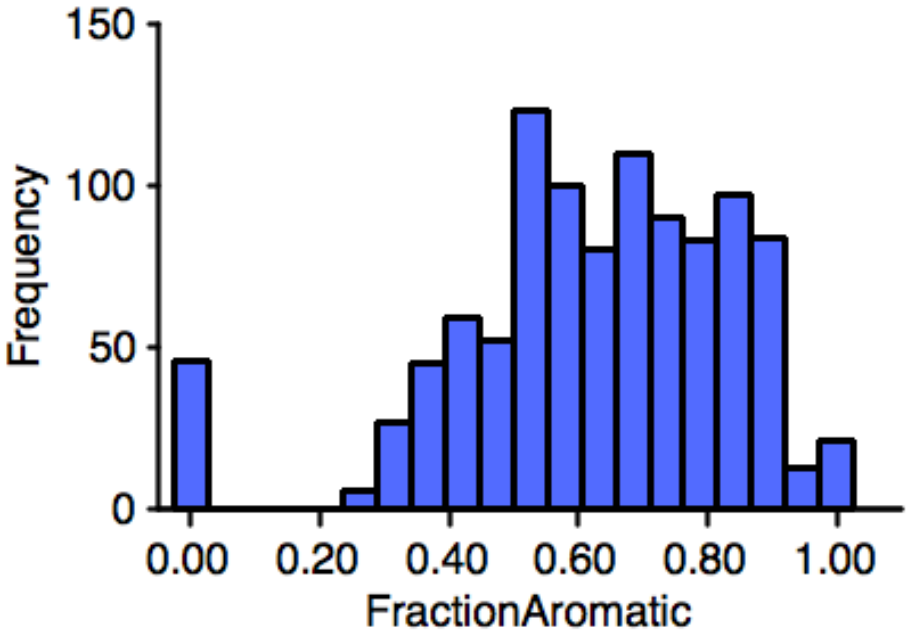


"Screening Collection"

Hits

Comparison of Aromaticity

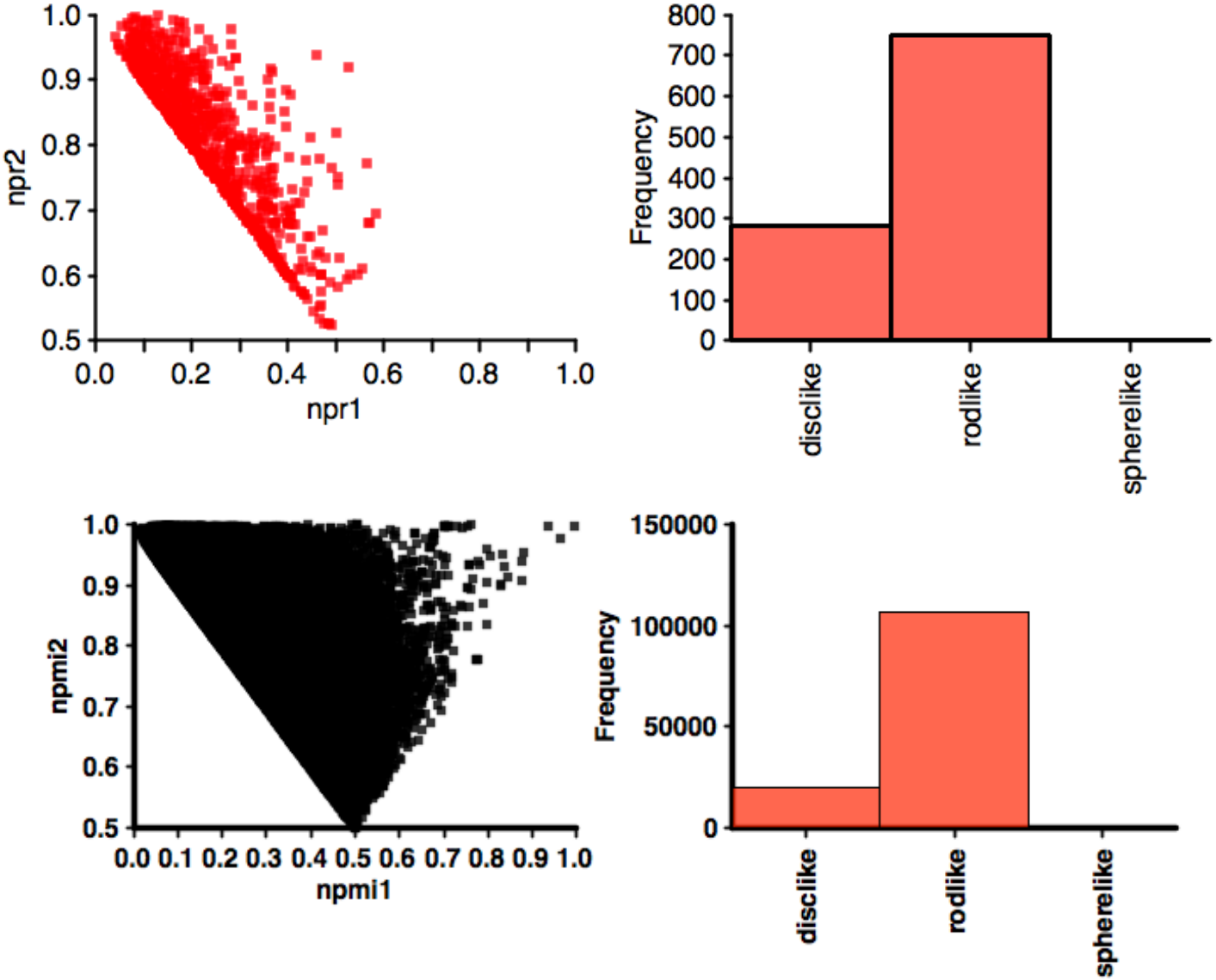




"Screening Collection"

Hits

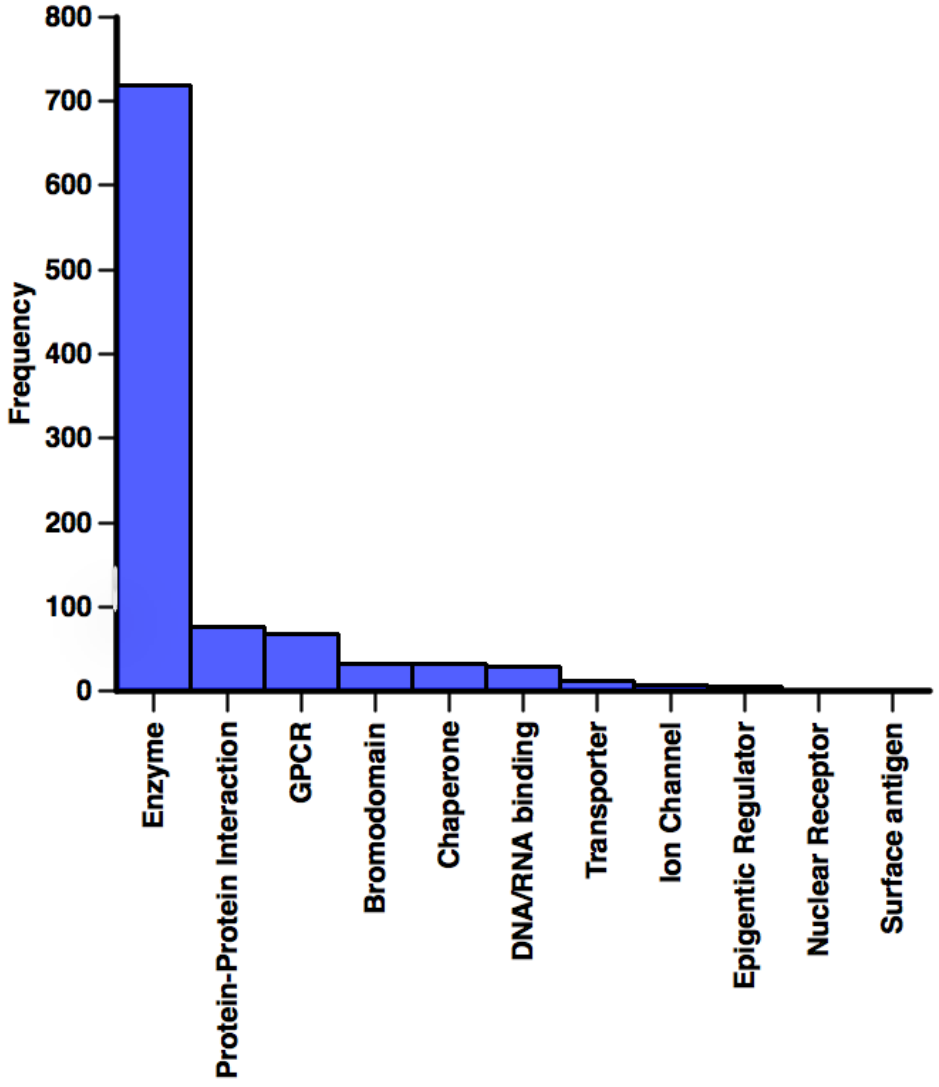
Comparison of Shape

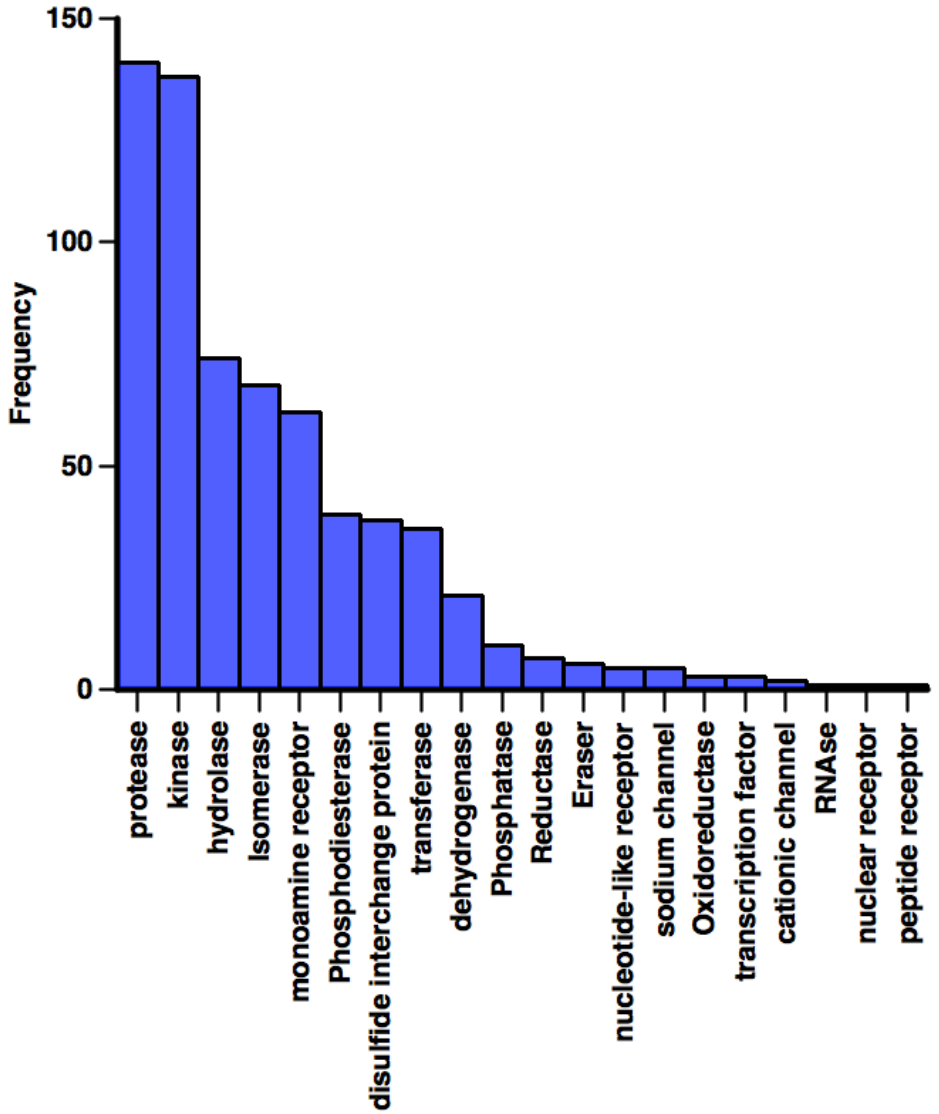


Conclusions

- Published fragments are lower molecular weight
- They contain a greater proportion of ionisable groups
- They contain a greater proportion of aromatics rings
- They contain a greater proportion of “disc-like” shaped molecules
- The role of increased 3D shape is unproven.

Targets

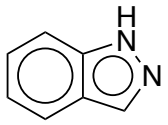

Frequency
800
700
600
500
400
300
200
100
0
Enzyme
Protein-Protein Interaction
GPCR
Bromodomain
Chaperone
DNA/RNA binding
Transporter
Ion Channel
Epigentic Regulator
Nuclear Receptor
Surface antigen


Frequency
150
100
50
0
protease
kinase
hydrolase
Isomerase
monoamine receptor
Phosphodiesterase
disulfide interchange protein
transferase
dehydrogenase
Phosphatase
Reductase
Eraser
nucleotide-like receptor
sodium channel
Oxidoreductase
transcription factor
cationic channel
RNAse
nuclear receptor
peptide receptor

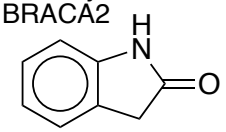
Multiple targets

- Over 80 fragment hits have been shown to be active against multiple targets.
- Whilst a few are active against similar targets (e.g. kinases), many are active against seemingly unrelated proteins.

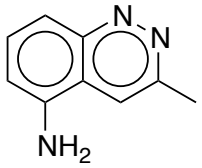
Fragments active against multiple targets



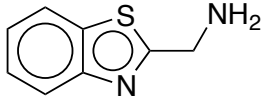
CDK2
DNA Gyrase
BRACA2



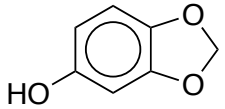
Inositol-3-phosphate synthase
DNA Gyrase
CD54



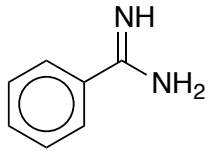
Inositol-3-phosphate synthase
thymidylate synthase



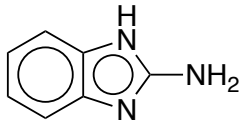
Inositol-3-phosphate synthase
ASIC3
Mycobacterium tuberculosis transaminase



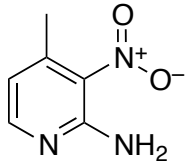
Inositol-3-phosphate synthase
HIV Integrase



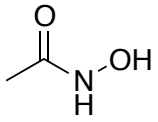
Factor Xa
Urokinase
Tryptase
Thrombin



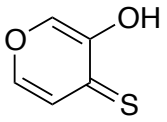
phenylethanolamine N-methyltransferase
Urokinase
Tryptase
DNA Gyrase



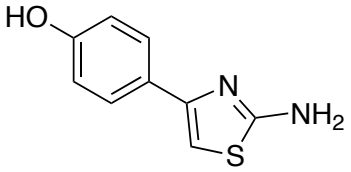
PDE10A
PDE4a
thiM



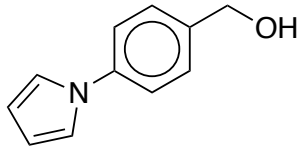
MMP-2
anthrax lethal factor
tyrosinase
Stromelysin



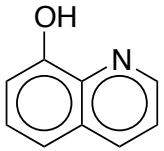
MMP-2
anthrax lethal factor
tyrosinase



LTA4H
Trypanosoma brucei Choline Kinase
Apical Membrane Antigen 1

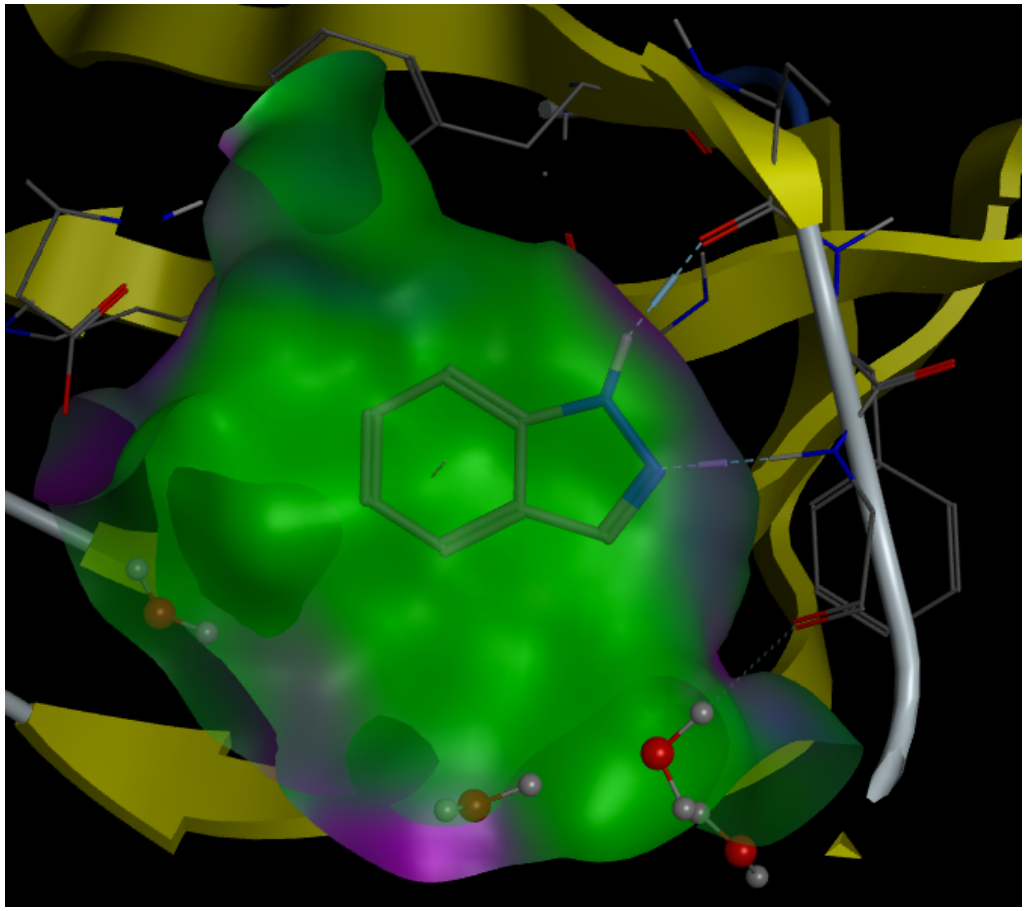


Inositol-3-phosphate synthase
Trypanosoma brucei Choline Kinase

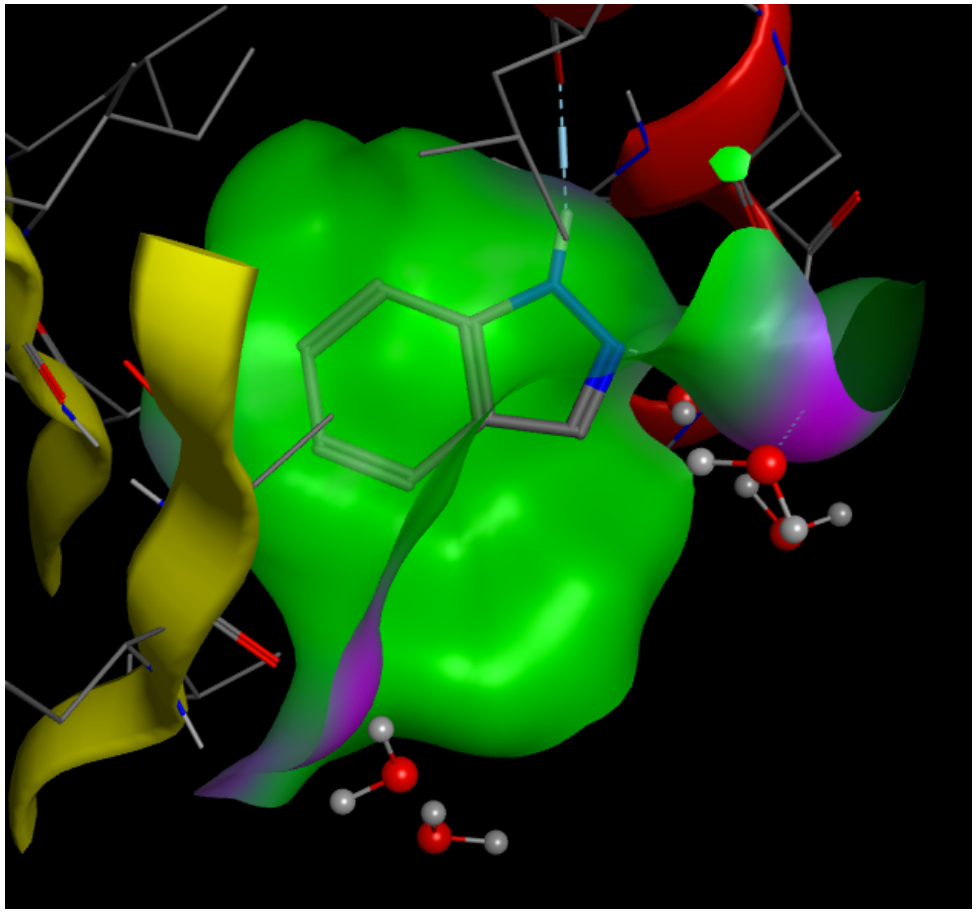


HIV-1 Integrase-Lens Epithelium-Derived Growth Factor/p75 (IN-LEDGF/p75) Interaction
MMP-2

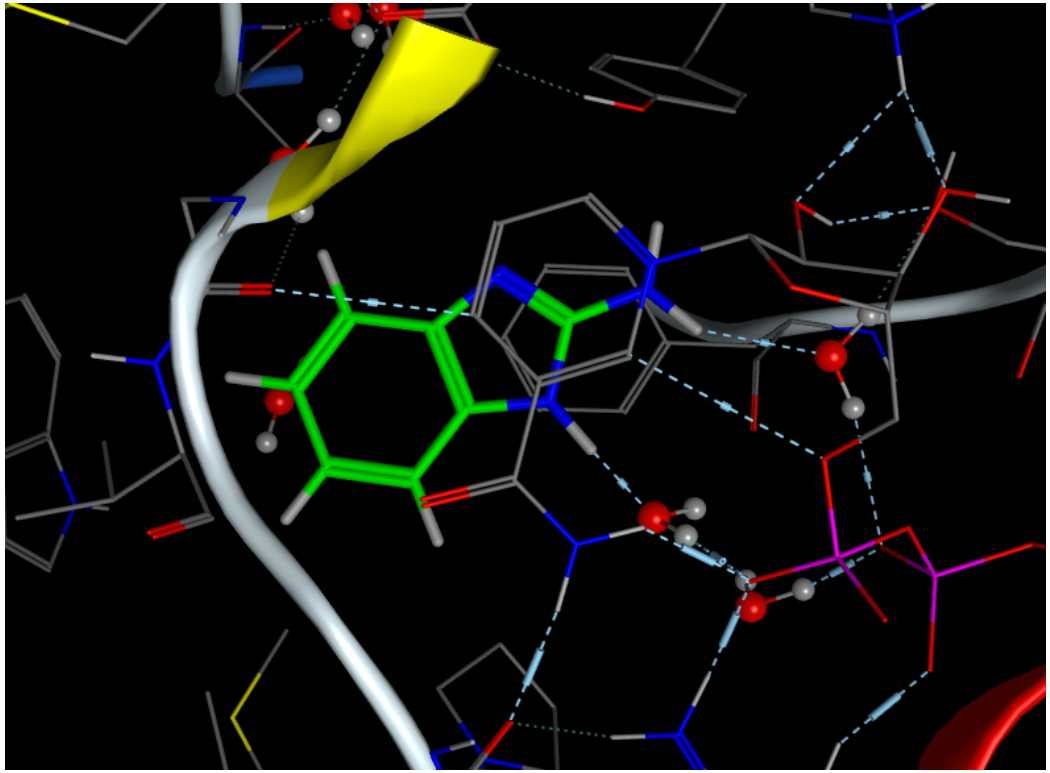
Do identical fragments bind in a similar manner to different targets?



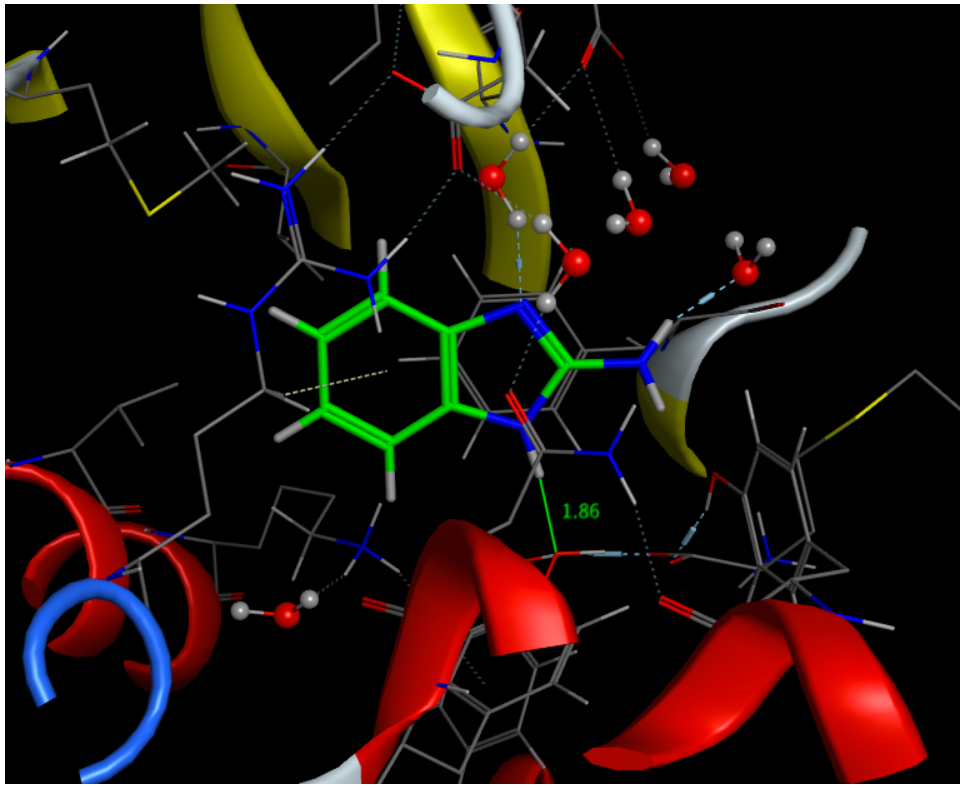
RadA-BRAC2



CDK2

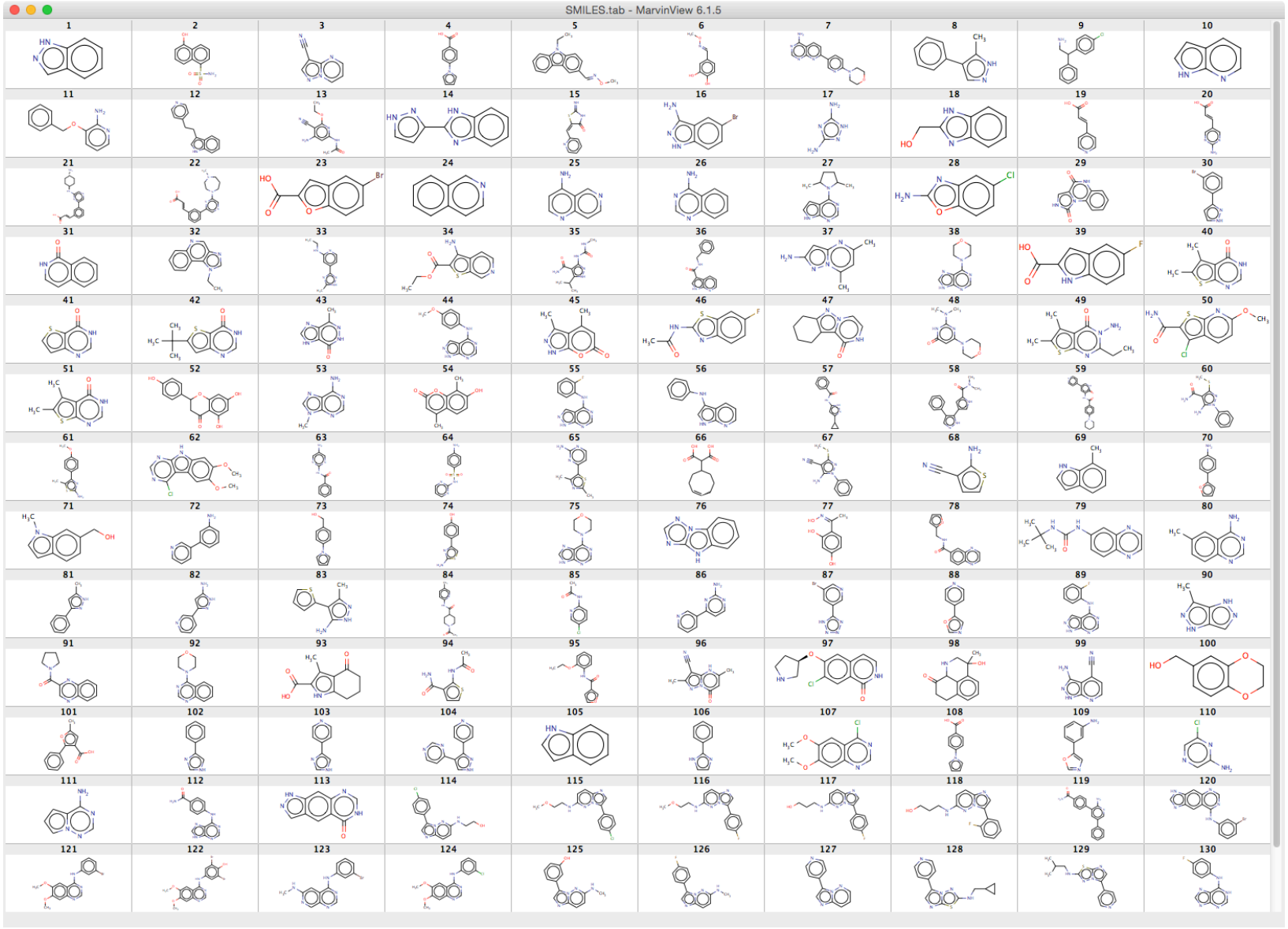


PTR1

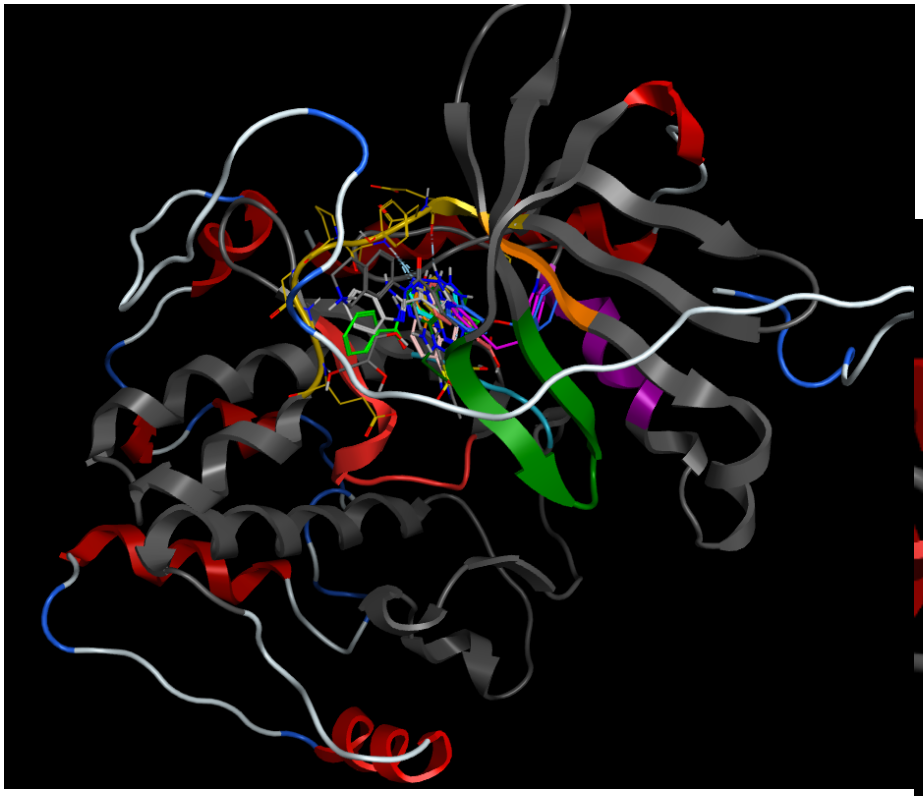


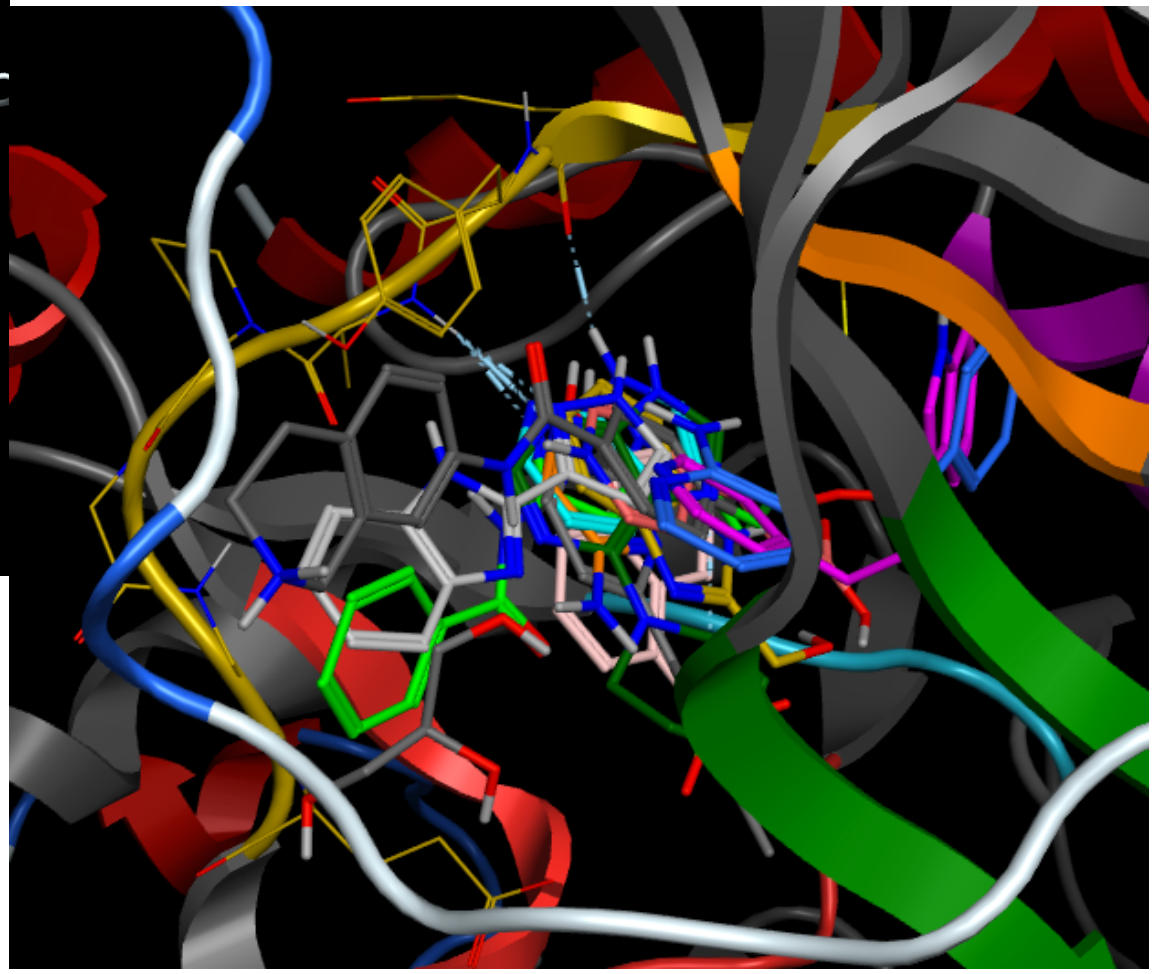
hPNMT

Kinase Fragments



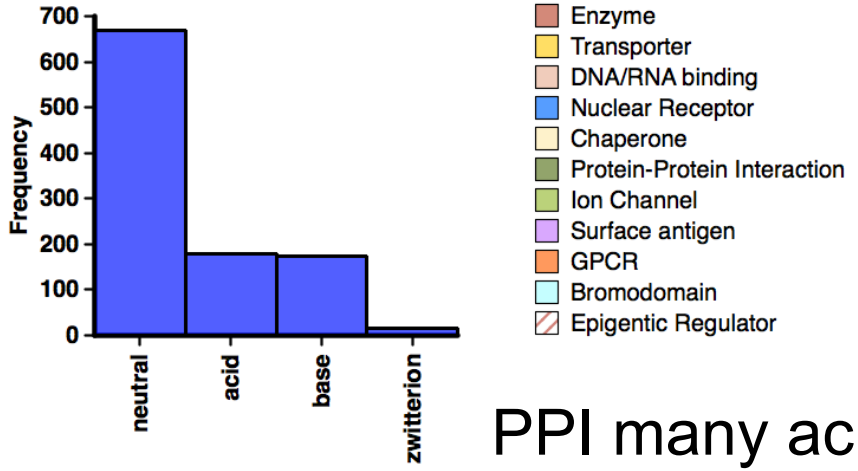
All Fragments in PDB bind to hinge region





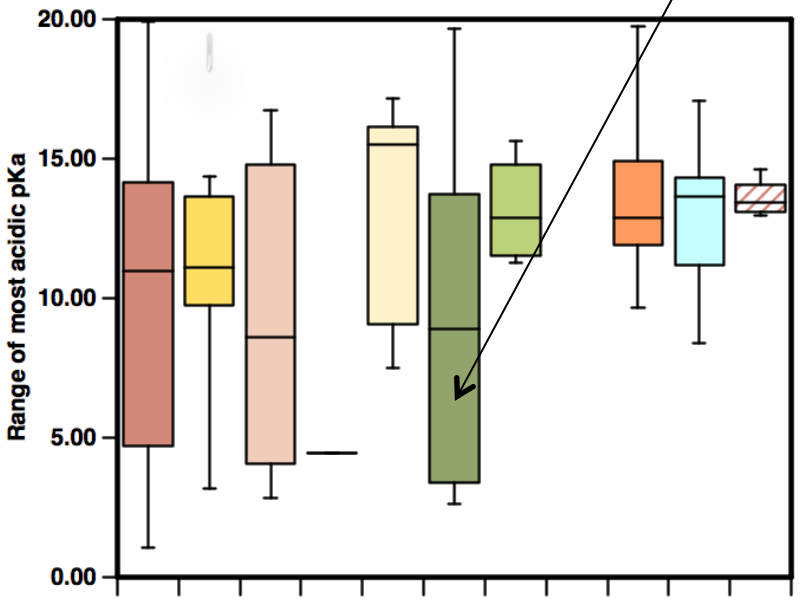
Of 137 fragment hits identified against kinase targets, 12 are in PDB.

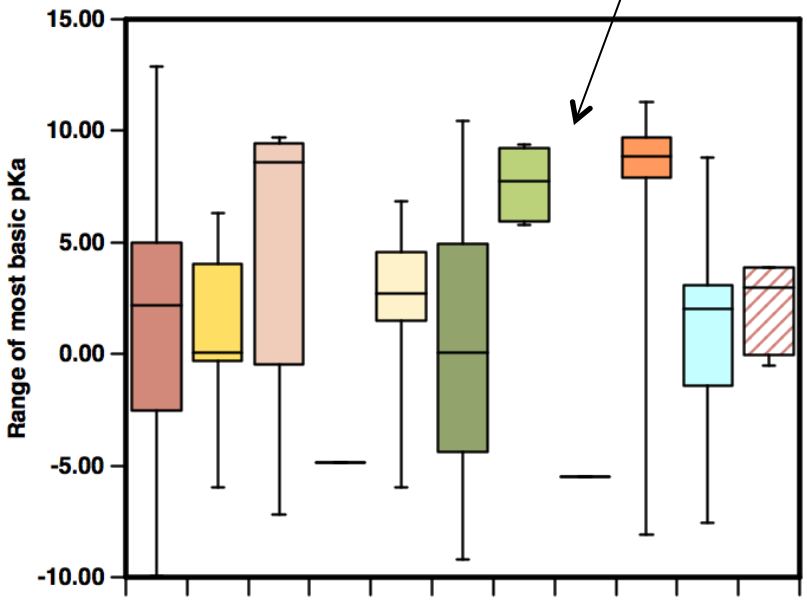
Effect of pKa and Target Type



Ion Channel and GPCR no acids but number of basic

PPI many acids

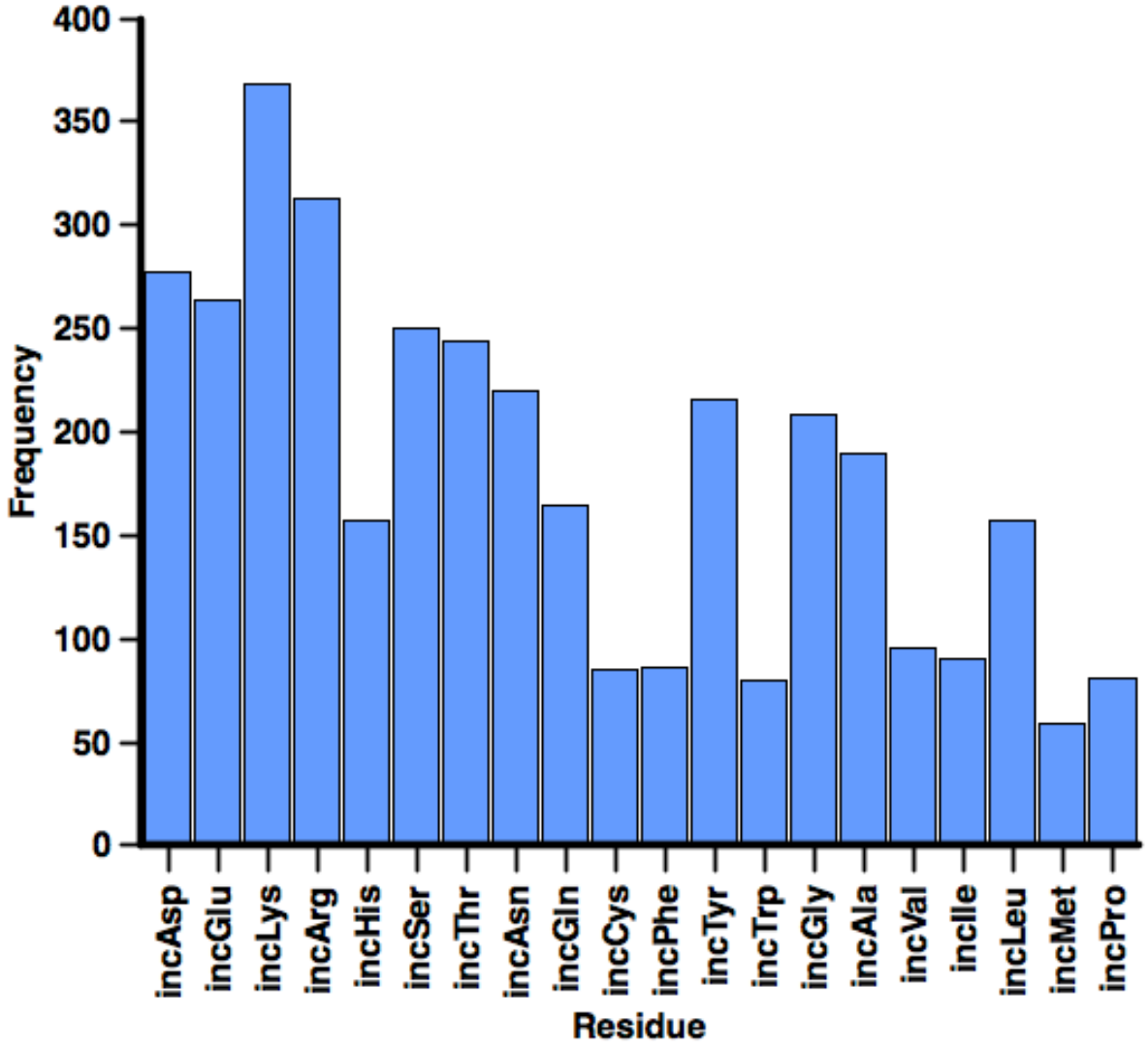




TIMBAL Database

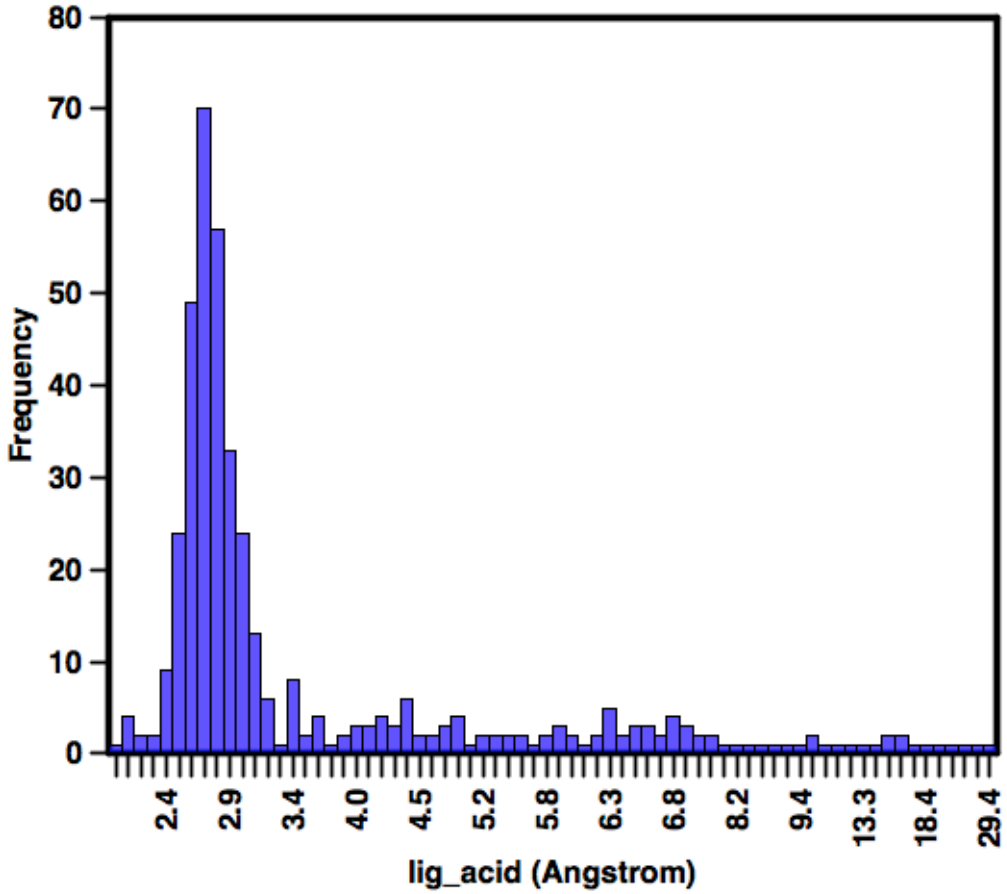
- Curated database containing small molecules that modulate protein-protein interactions. Integrins form a significant proportion (50%, but only 139 with PDB).
- Also contains PDB codes if available.
- If we use those 689 PDB records for which there is a ligand present we can calculate which residues of the protein are with 3A of the ligand using a script within MOE.

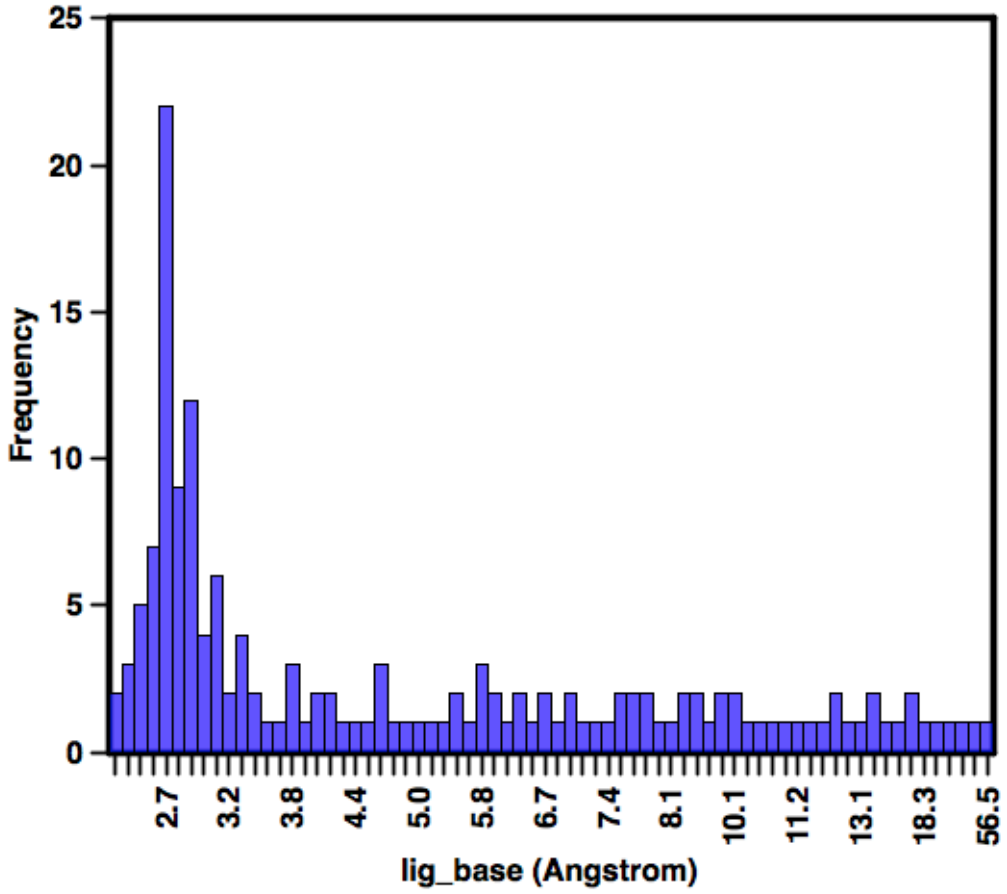
Amino Acids in the binding site.



Do acidic ligands bind to basic residues?

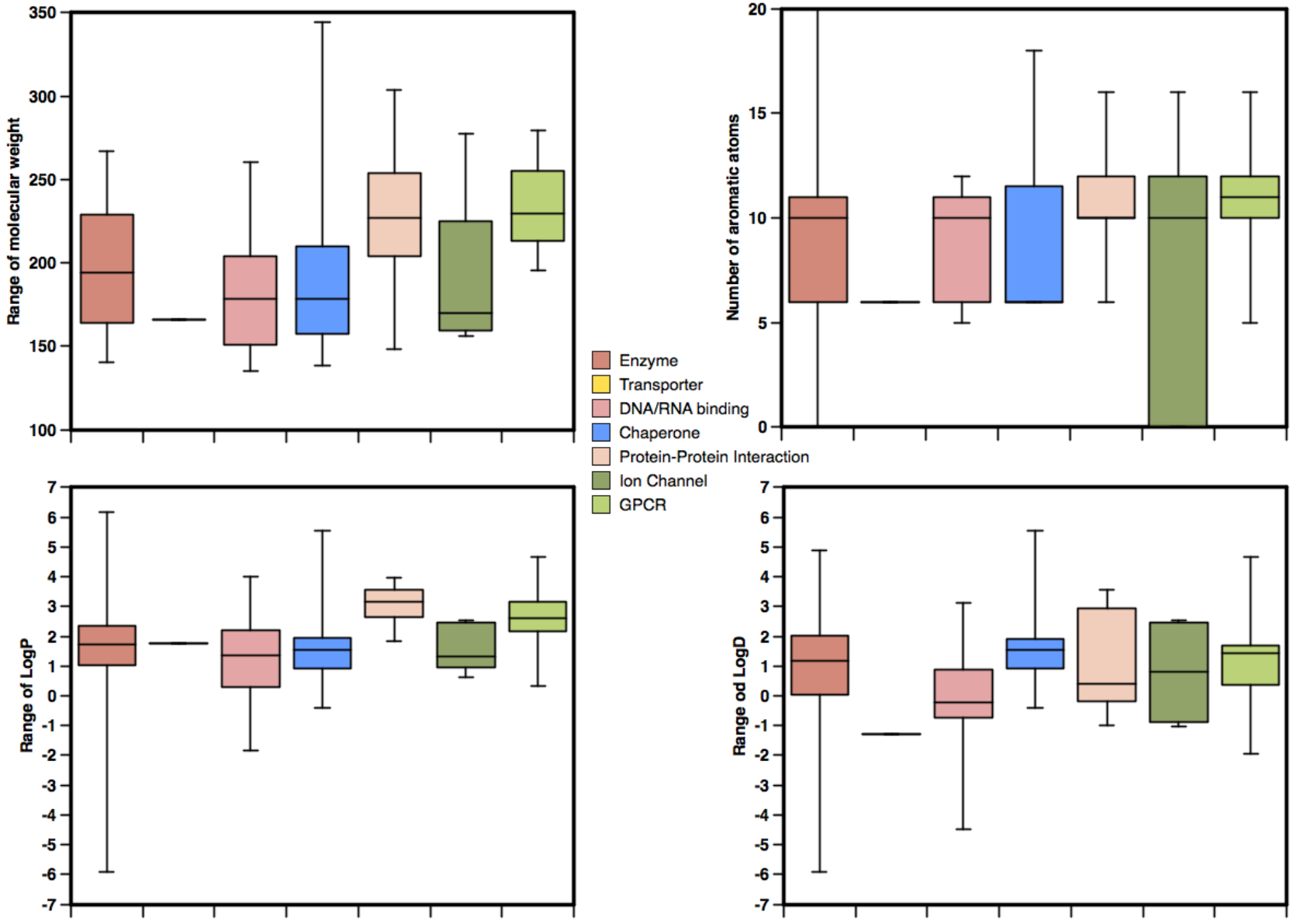
Measure distance between ionisable groups in ligand and protein





Ionizable groups in the ligand should be able to bind to the appropriate amino acids.

Target type physicochemical properties



Conclusions

- Fragment screening hits tend to be lower molecular weight, contain aromatic rings and ionizable groups.
- Some targets (GPCR, Ion channels, PPI) select for specific physicochemical properties
- Detection technology does not appear to influence properties of hits identified.
- Measured affinities of fragment hits are in uM to mM range

Future work

- Collaboration with Chris Hunter (Cambridge)
  - Is there something special about aromatic fragments?
  - Can we use predicted/observed binding affinities of fragments to score docking results

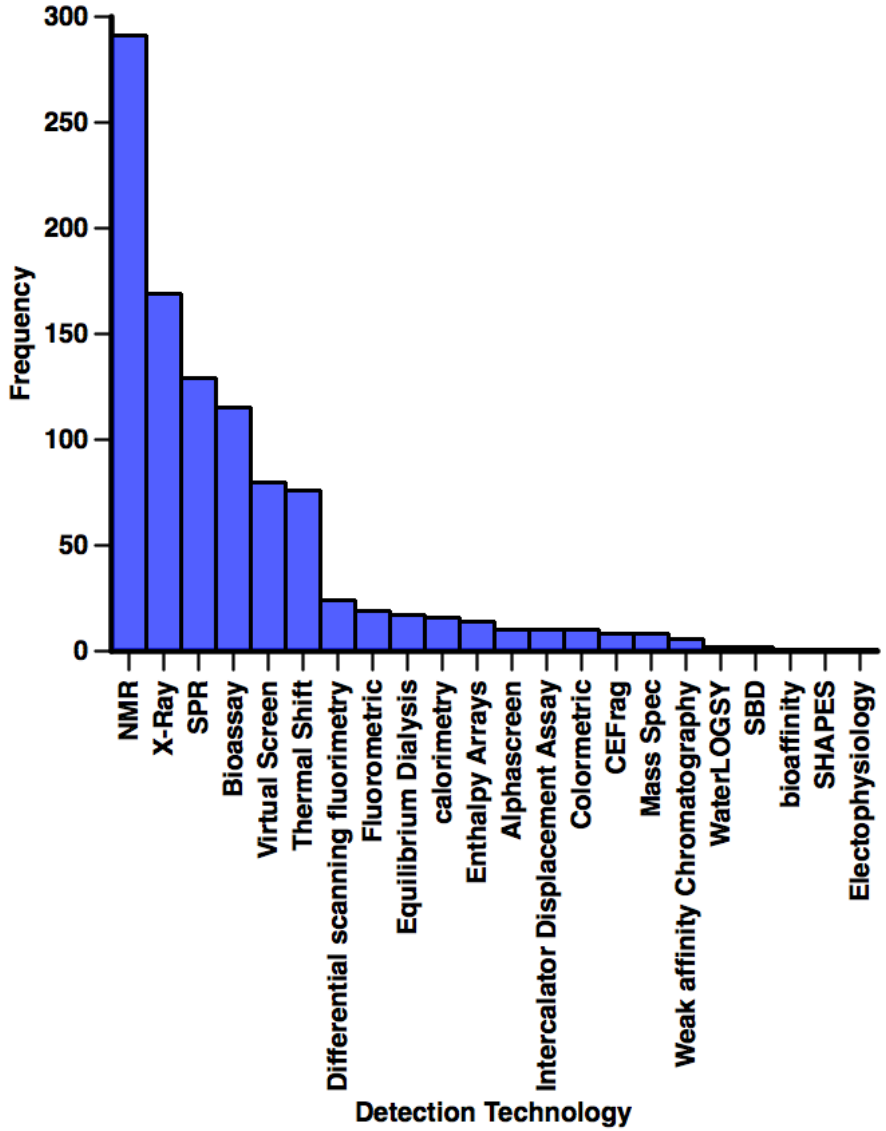
Acknowledgements

- All those who published results
- Chemical Computing Group
- ChemAxon

- And you for your attention!

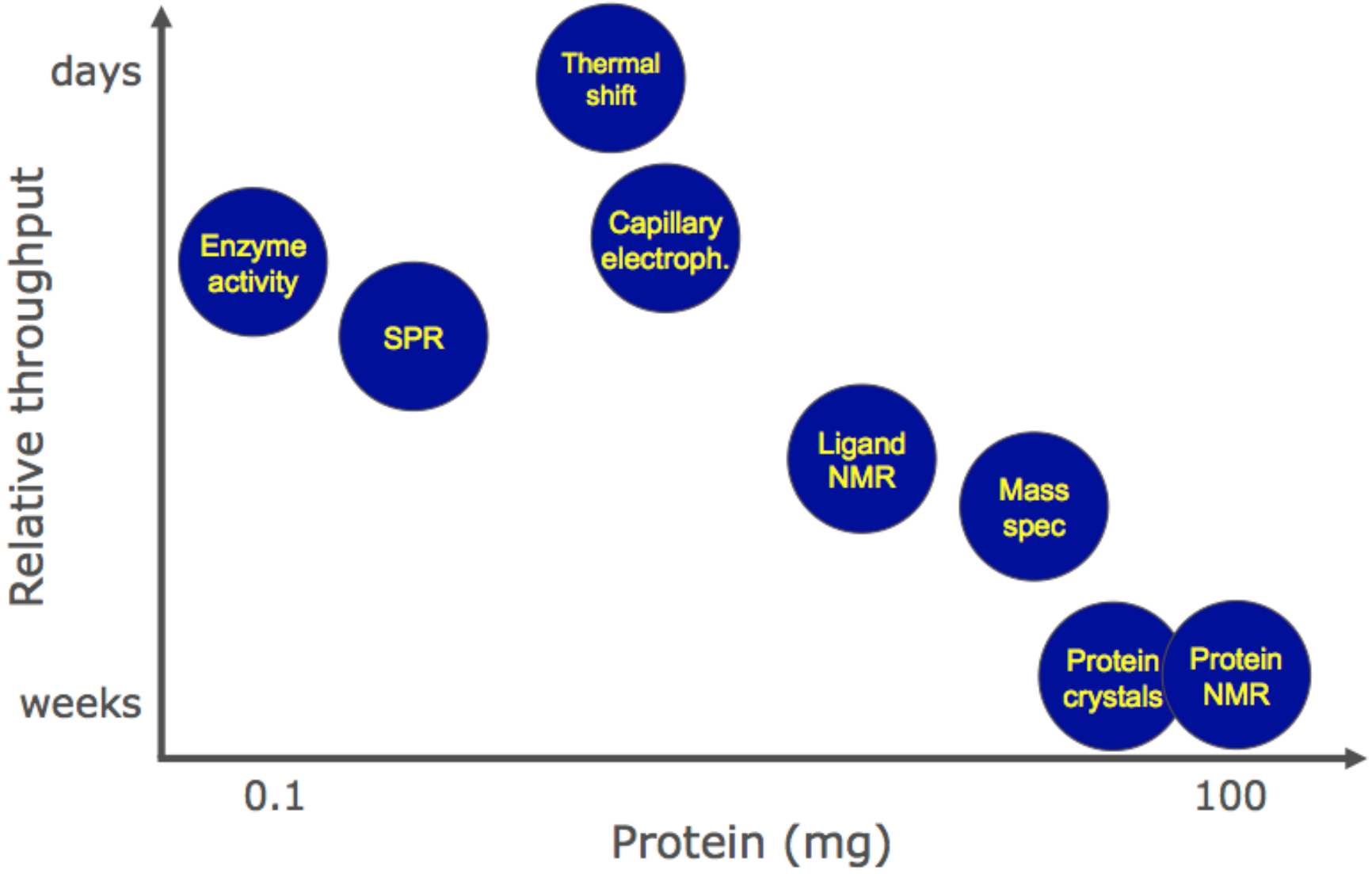
Spare Slides

Detection technology

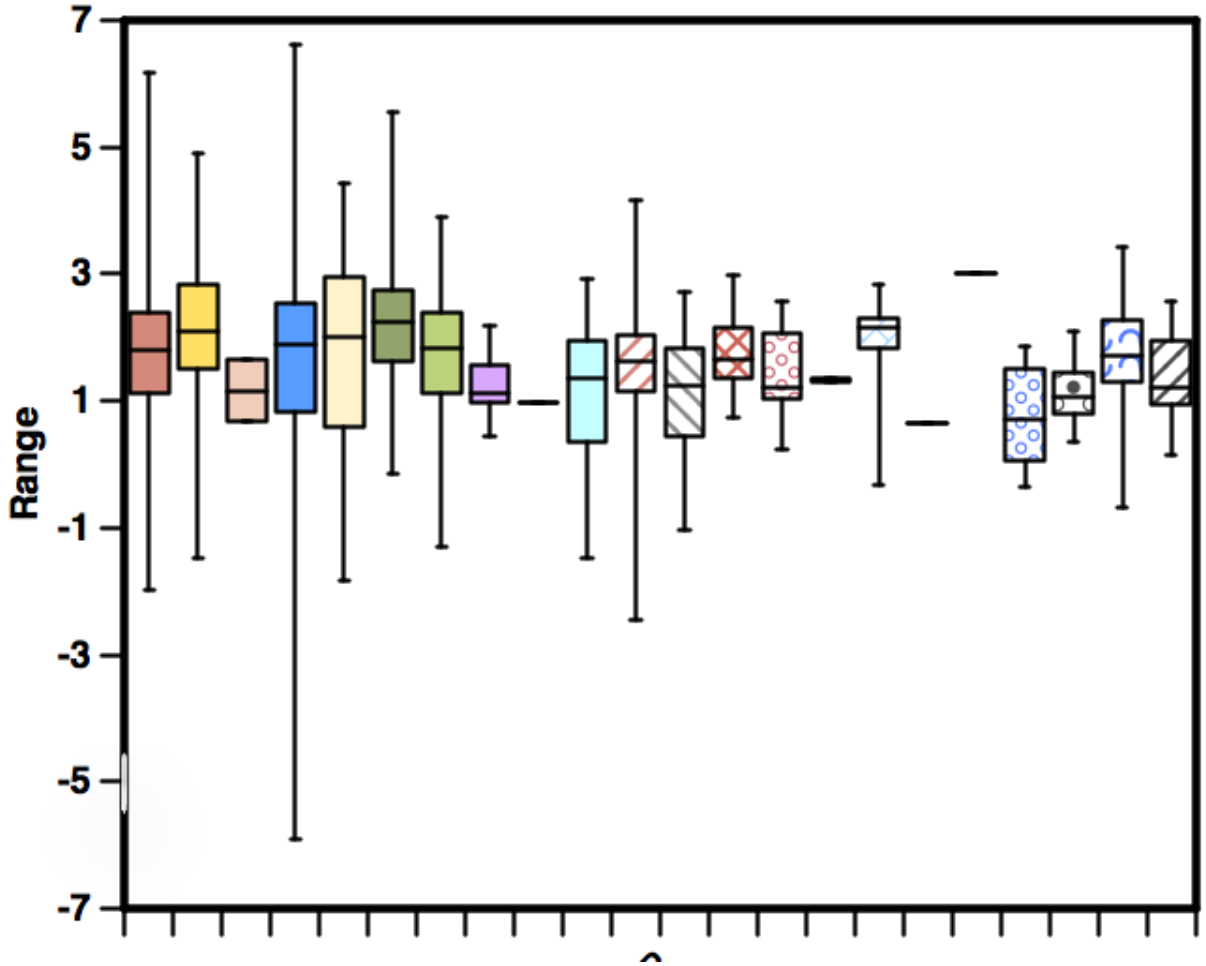


NMR and X-ray dominate
Thermal shift increasing recently

Choice of technology



Detection Technology and LogP of hits

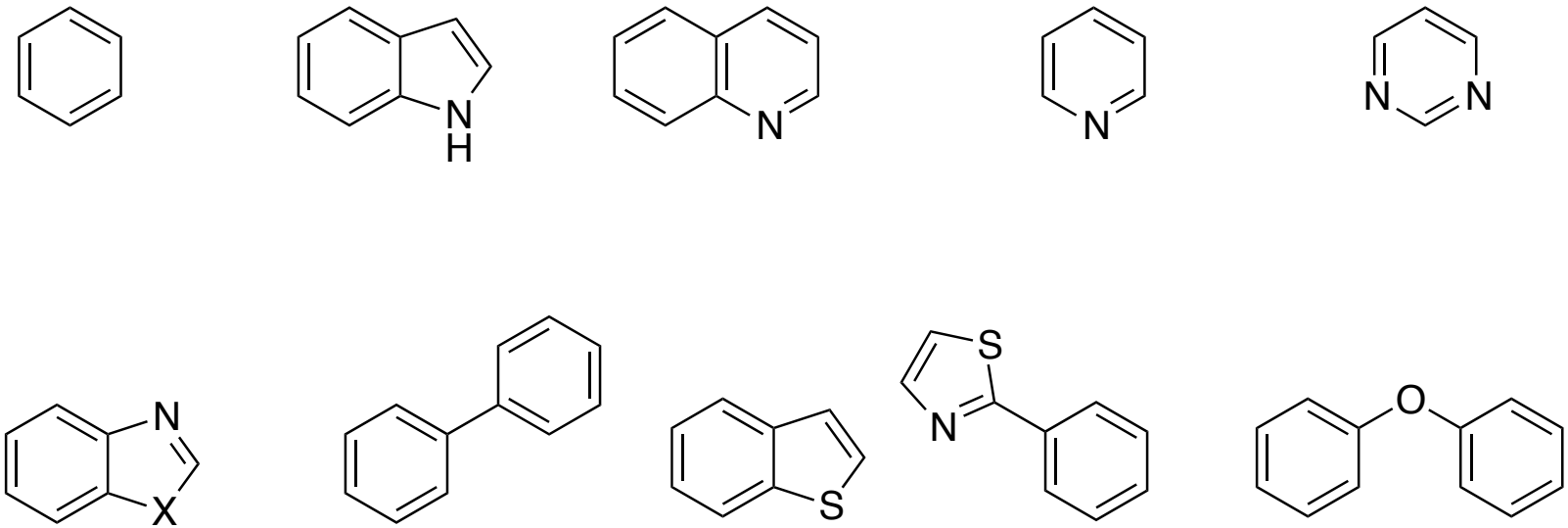


- X-Ray
- NMR
- SBD
- Bioassay
- Mass Spec
- Virtual Screen
- SPR
- Colormetric
- bioaffinity
- Equilibrium Dialysis
- Thermal Shift
- Fluorometric
- CEFrag
- Enthalpy Arrays
- WaterLOGSY
- Weak affinity Chromatography
- Electophysiology
- SHAPES
- Alphascreen
- Intercalator Displacement Assay
- Differential scanning fluorimetry
- calorimetry

Detection Technology

- Evidence from literature that different technologies can identify hits for a single target.
- No evidence that detection technology influences the physiochemical properties of the hits identified.
  - Some technologies (e.g. SPR) are thought to have a higher false positive rate.

Most common scaffolds



X=S, N

How does this compare with drugs

- Search DrugBank (www.drugbank.ca)
  - Approved, small molecule drugs.
- 1474 molecules exported
- Import into MOE database
- Use sca.svl to identify scaffolds
  - The script finds all scaffold in a database, writes them to a separate database
  - A New Approach to Finding Natural Chemical Structure Classes; J. Med. Chem. 2002, 45, 5311-5320
    - http://dx.doi.org/10.1021/jm010520k

How does this compare with ligands in PDB?

- Download all ligands 149,282 structures
- Import into MOE database
- Remove solvent/buffers
- Remove co-factors (porphyrins)
- Remove DNA/RNA
- Remove metal complexes
- Identify fragments

How does this compare with BindingDB

- BindingDB is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules
- Select all molecules for which a binding affinity was measured
- Identify most common fragments

Measured affinities