Predictive ADMET - Promise and Reality

Anne Hersey – ChEMBL Group

EMBL-EBI

Overview

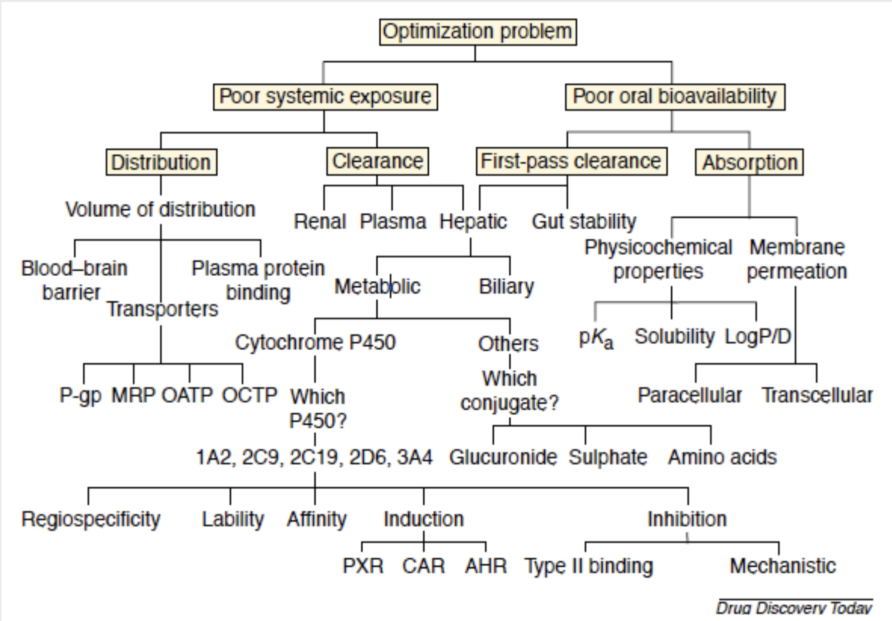
- Why predict ADMET
- Promise of Predictive ADMET
- Reality – what can predictions do
- Conclusions
  "*How modern state of the art methods can accelerate the process of drug design*"

Examples from ChEMBL Database

www.ebi.ac.uk/chembldb

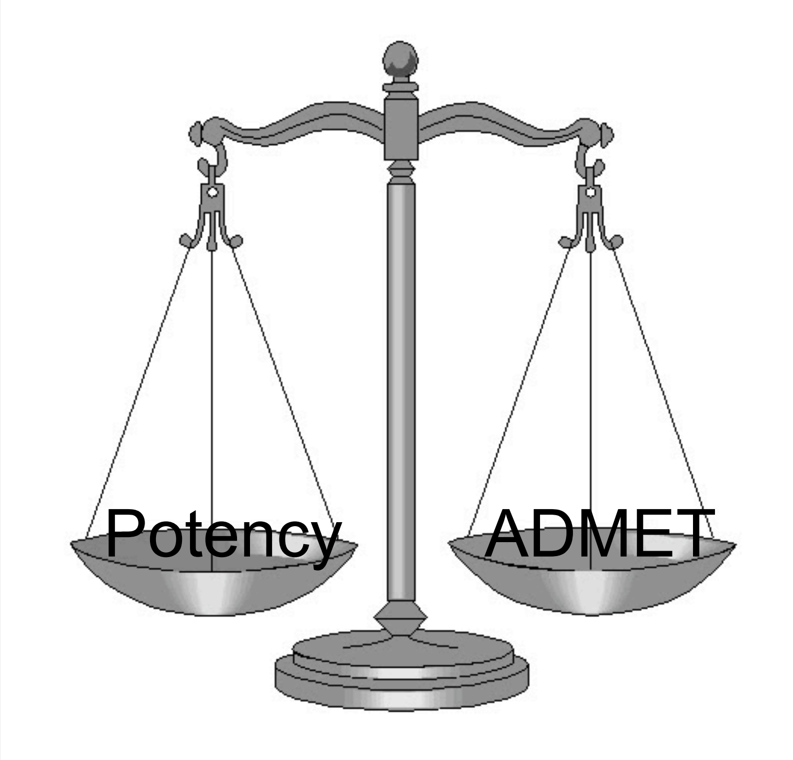
Do we need ADMET Prediction?

- Relatively lower amounts of ADMET data than potency data
- Multiple properties to take account of
  - Absorption, Distribution, Metabolism., Elimination, Toxicity
- in-vivo experiments
  - expensive, ethical issues
- Application in compound design
  - understanding not just screening

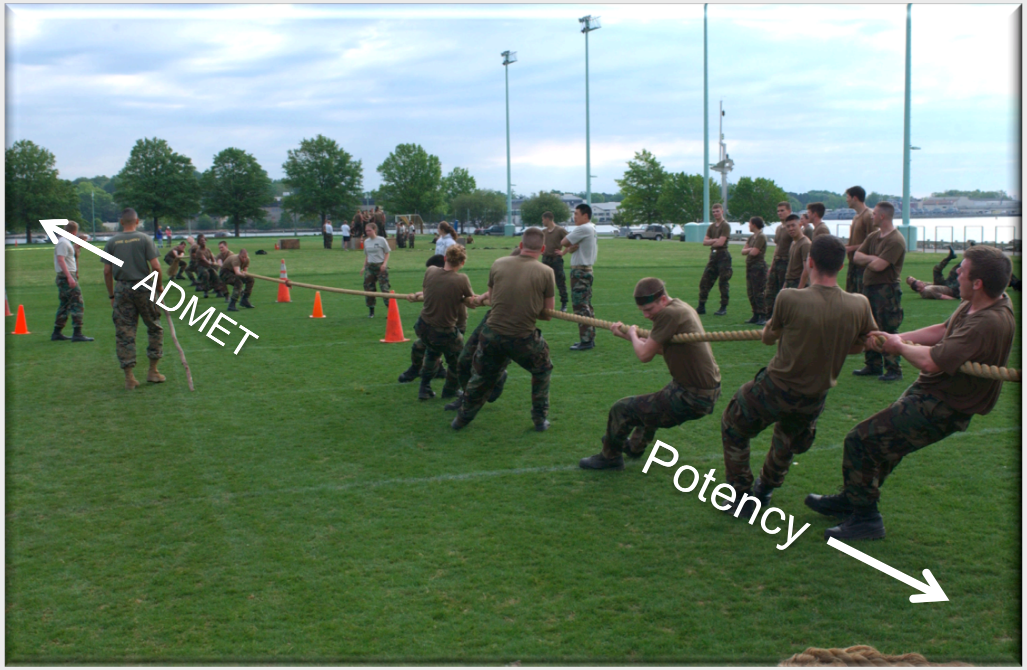


Tarbit 2002

Drug needs to balance potency, exposure and safety

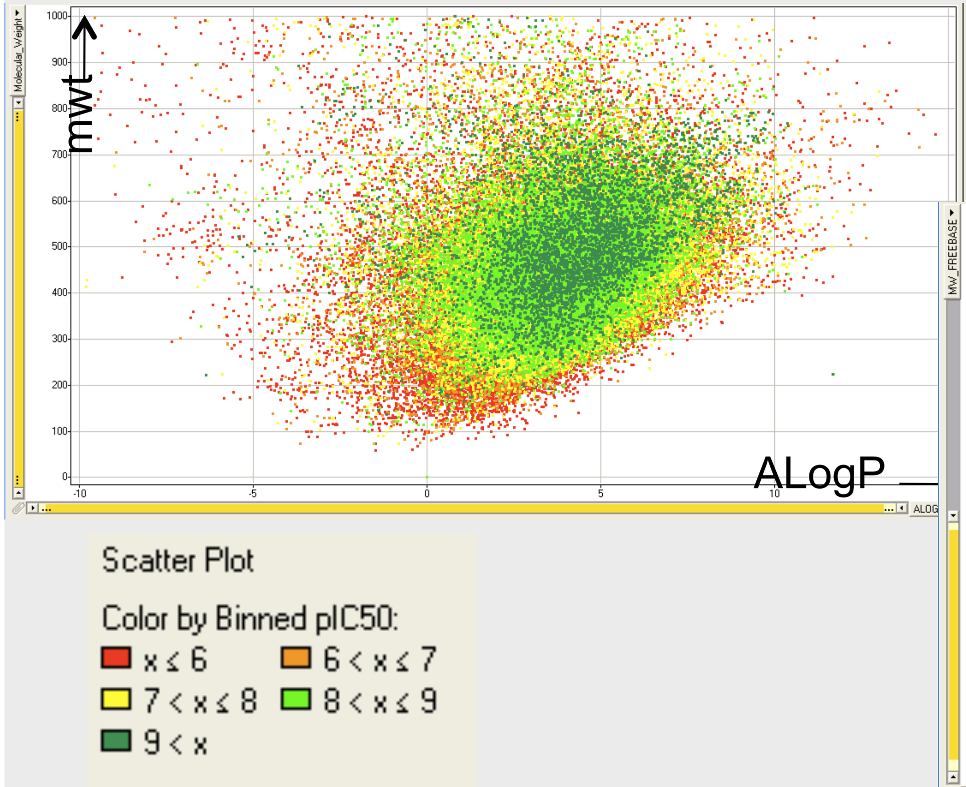


More of a tug of war than a balance!

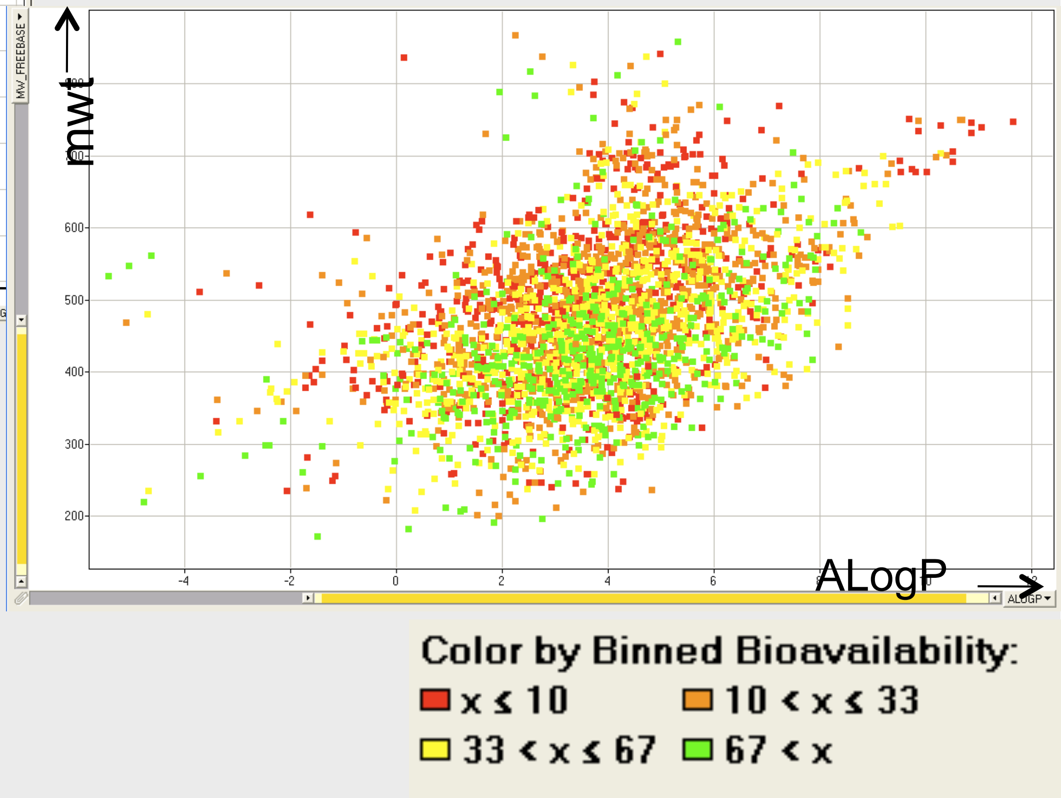


Example - Data from ChEMBL

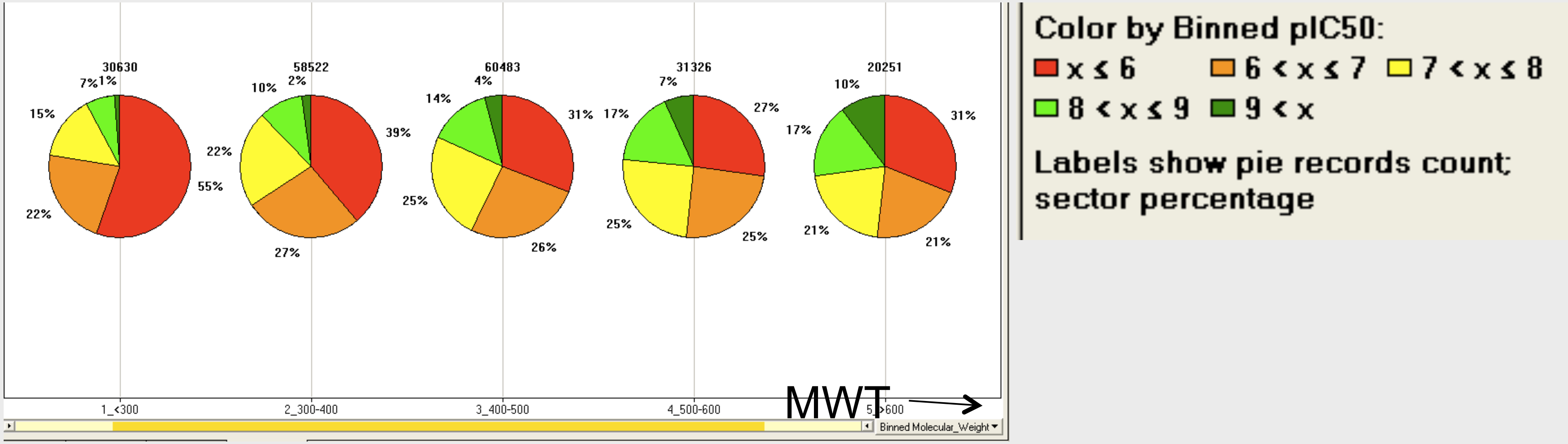
pIC50 binding data ~200K compounds

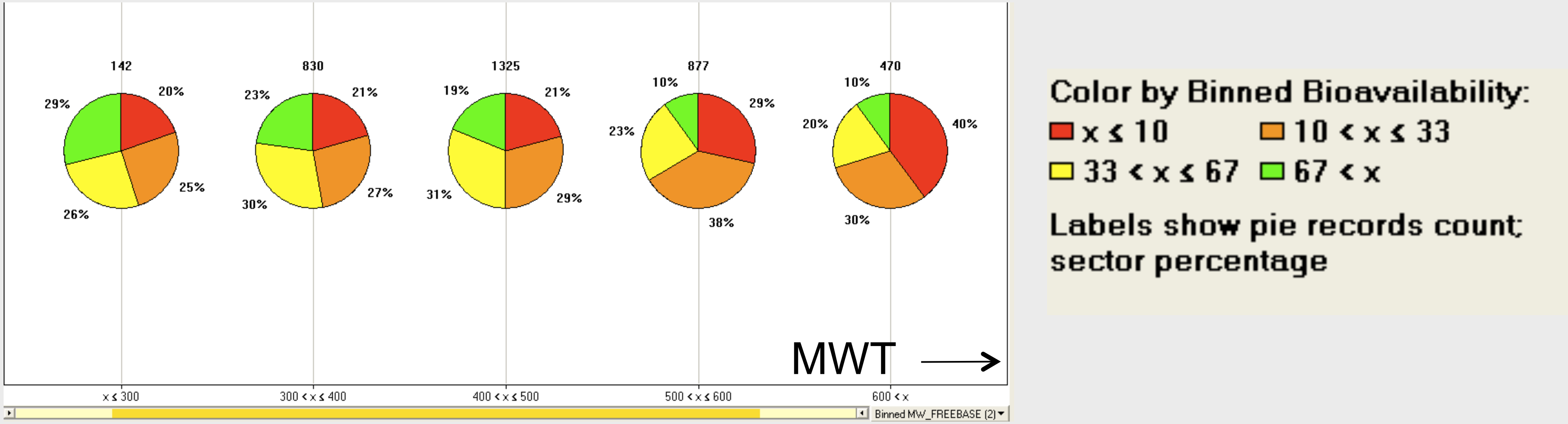


Rat bioavailability ~ 3.6K

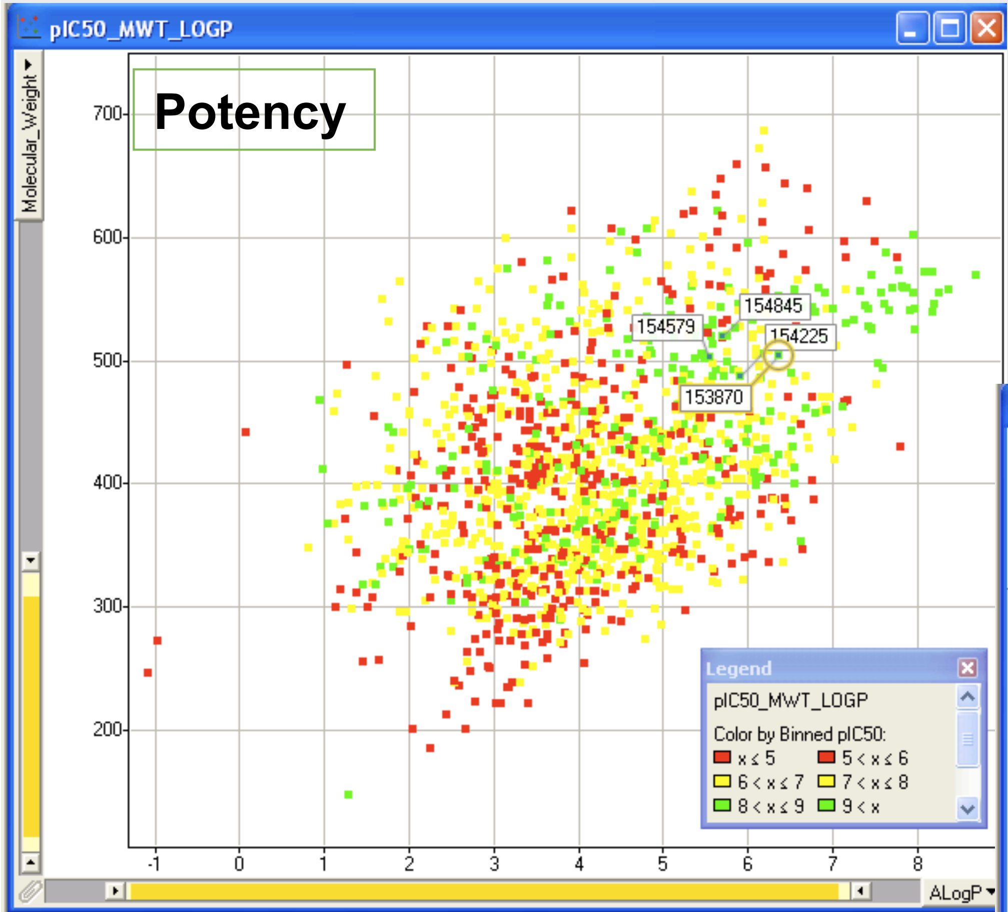


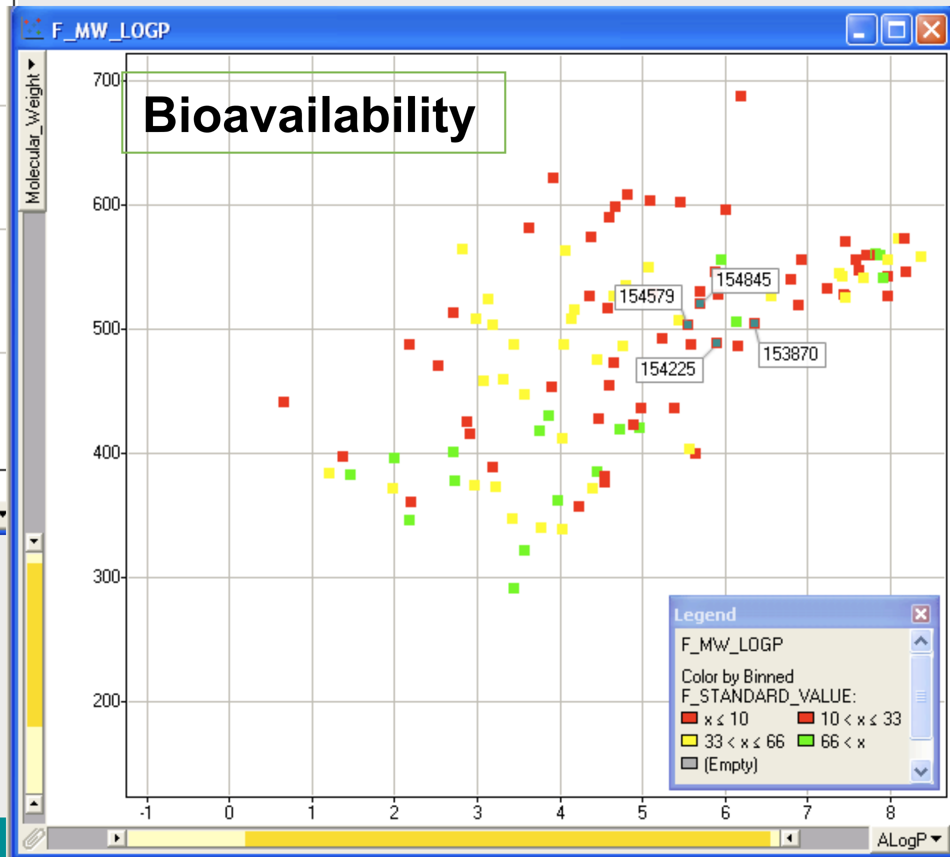
Opposing Properties





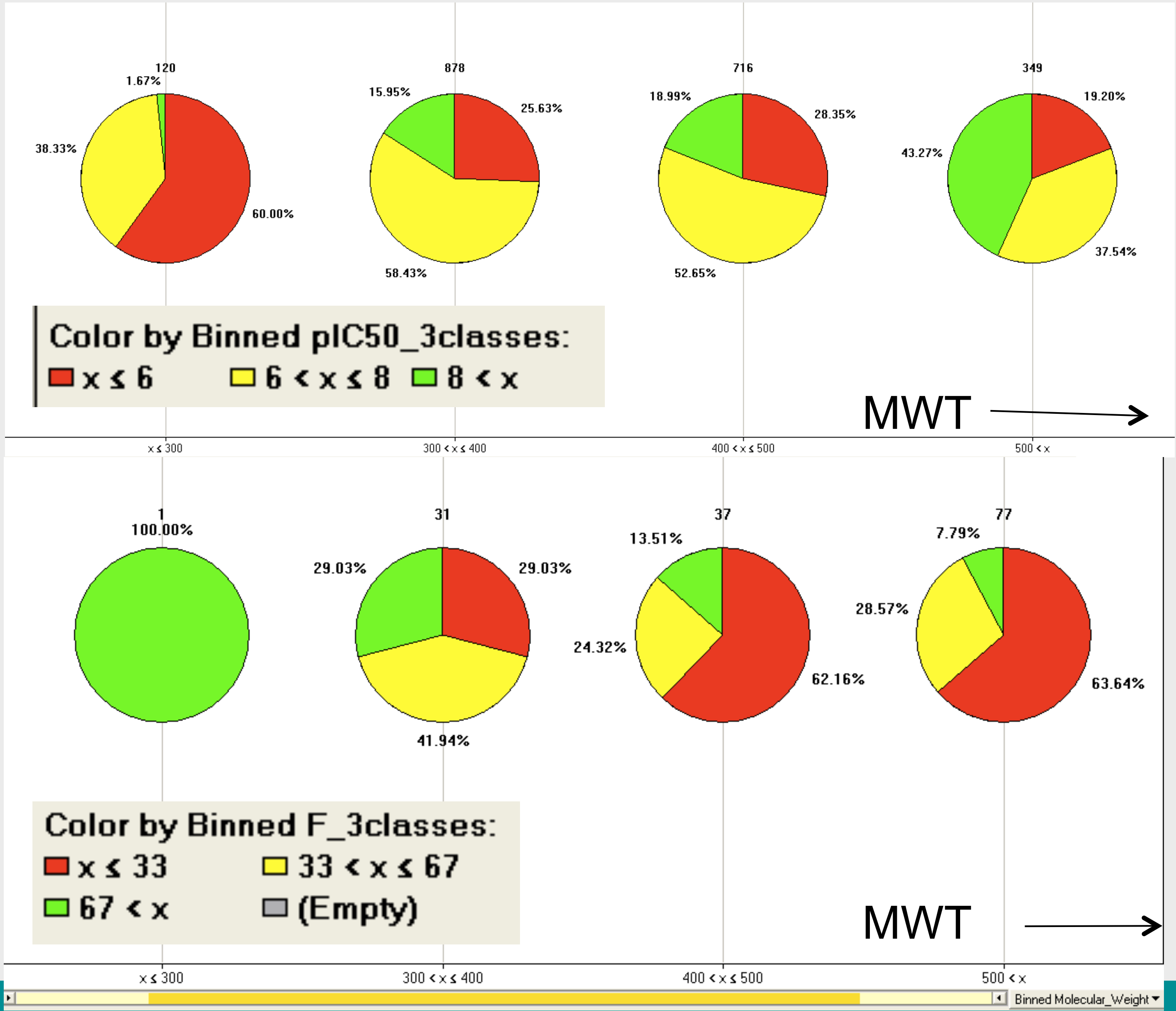
Same Trend on Data on Specific Target (P38alpha)





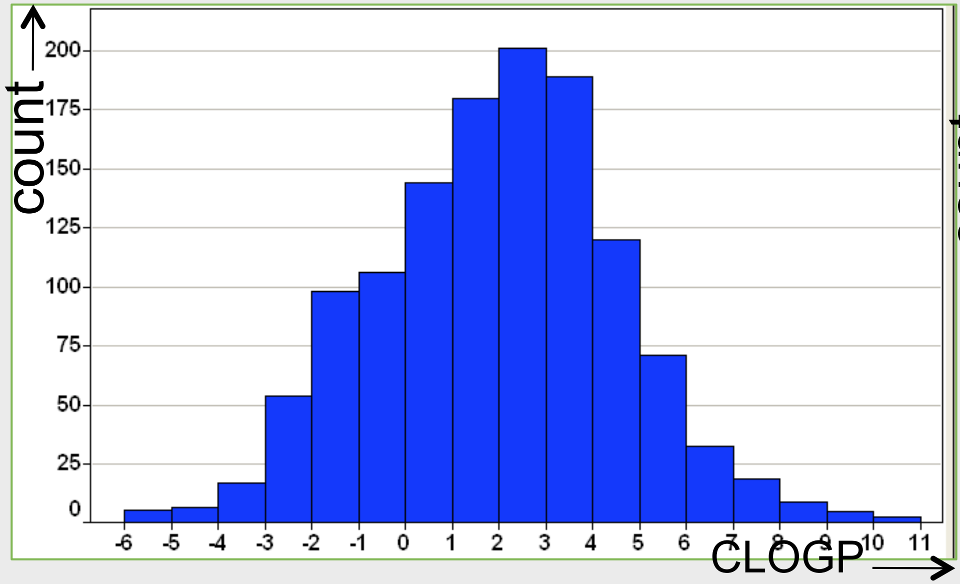
IC50 - 2000 compounds
%F - 150 compounds

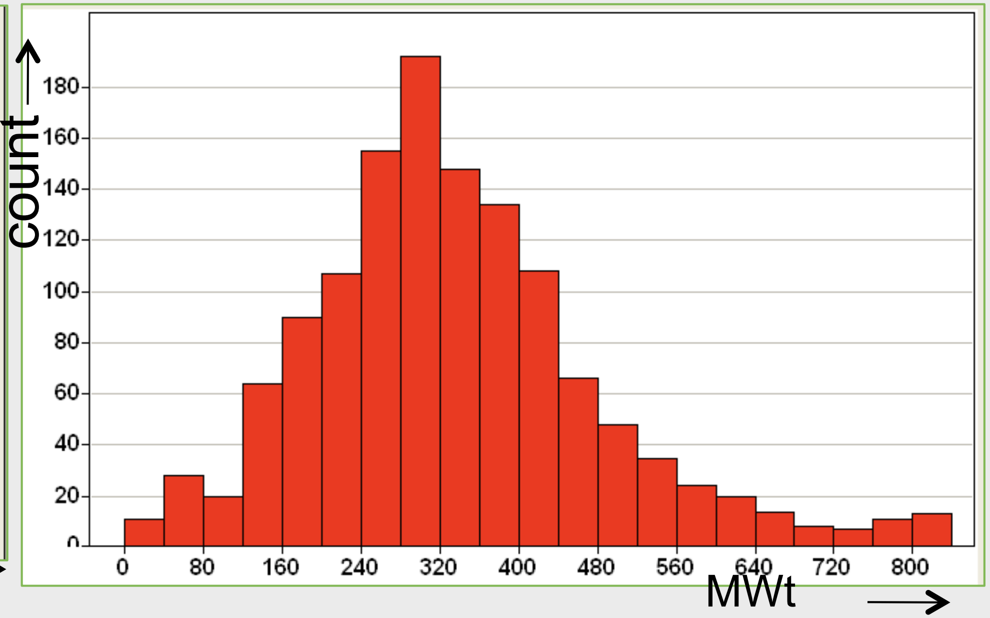
P38alpha – Opposing Properties

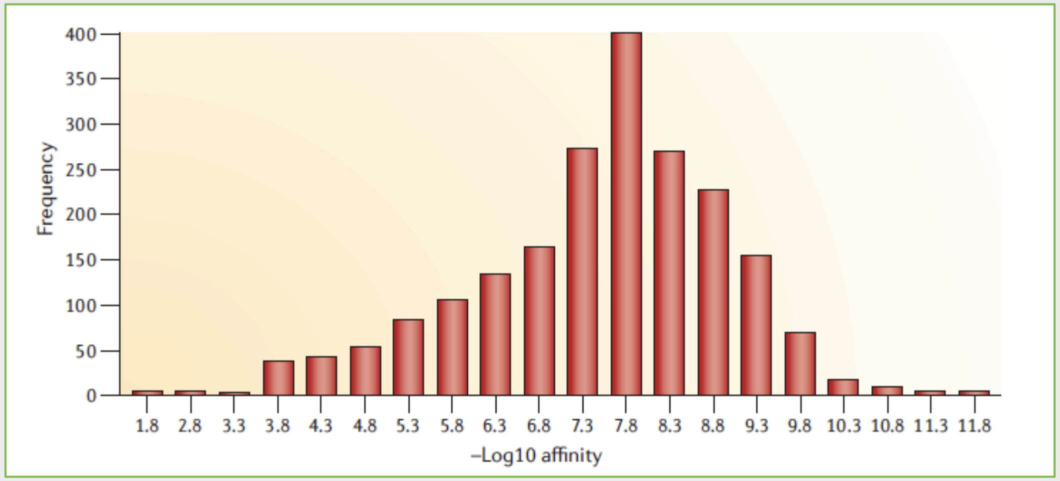


EMBL-EBI

Properties of Drug





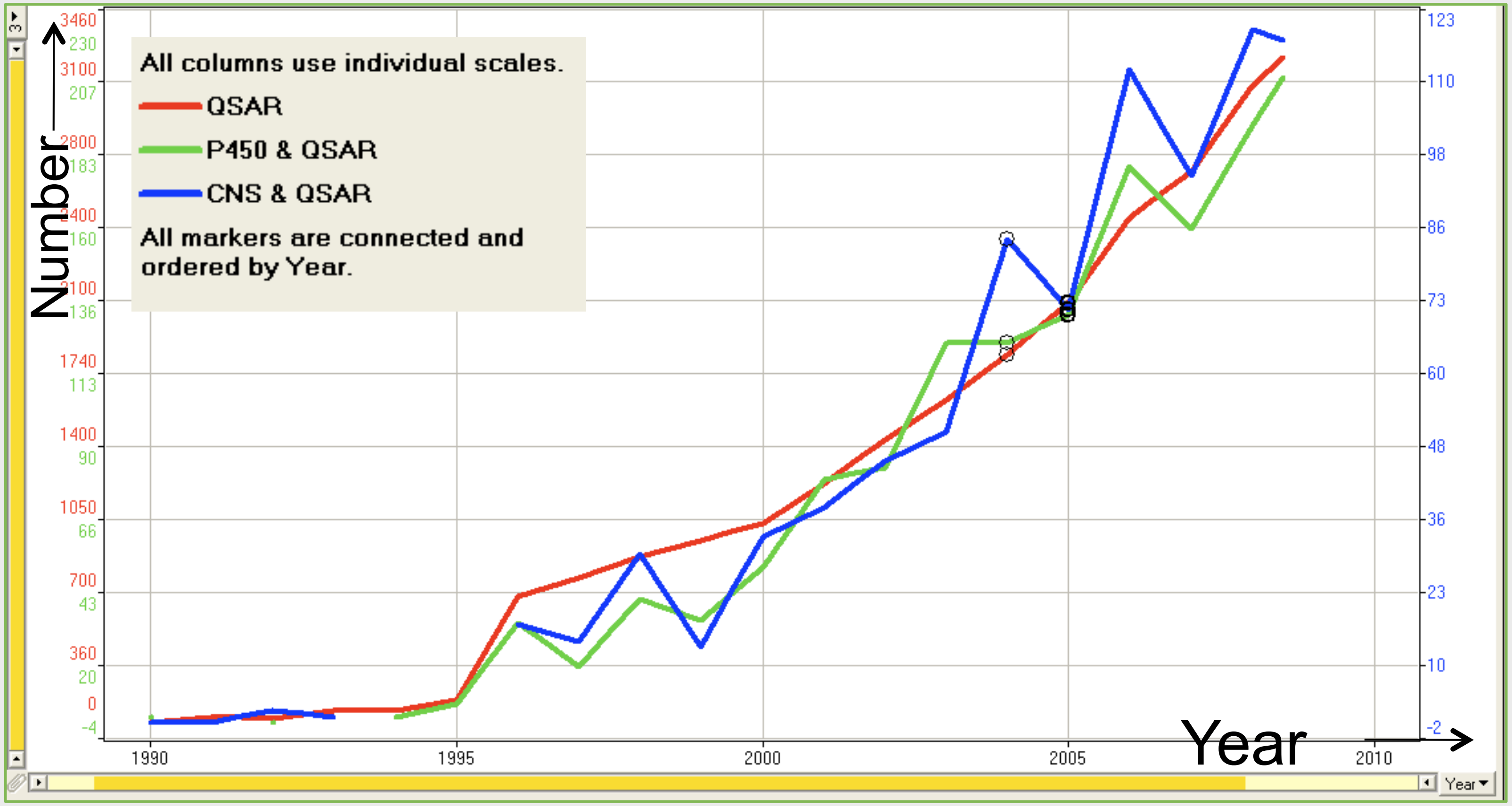


Drug Potencies median = 20nM
Overington et al
Nat Rev Drug Disc 2006

EMBL-EBI

Promise of ADMET Prediction

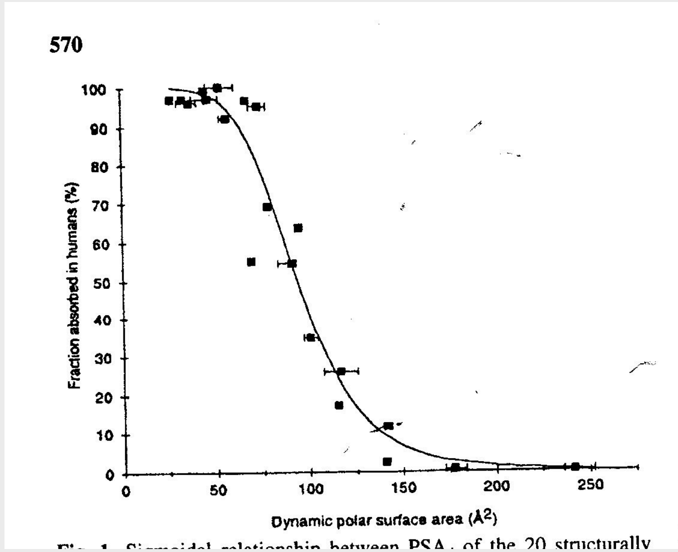
No. of Published ADMET Modelling papers



(Data from SCOPUS Search – any fields)

1990's

- ADMET models showed promise that ADMET could be predicted by simple descriptors and methodology
- Intestinal Absorption
    - Lipinski Rule of 5, Adv Drug Disc Rev 1997
    - PSA – K Palm, Pharm Res 1997

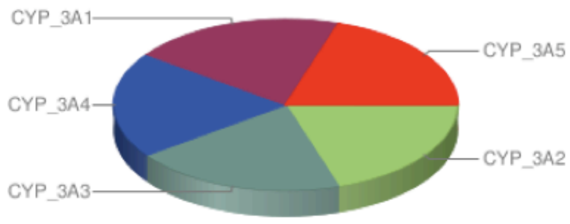


- Brain/Blood Ratio
    - MH Abraham, J Pharm Sci 1994
        - logBB=-0.04+0.20E-0.69S-0.72A-0.70B+1.00V (*n=57, r2= 0.91)*
    - D Clark, J Pharm Sci 1999
        - logBB=-0.015PSA+0.15ClogP + 0.14 (*n=55, r2=0.79)*

ChEMBL ADMET Data

Protein Target Classification Hierarchy

- Enzyme (2462)
  - Kinase (484)
  - Protease (334)
  - Phosphatase (70)
  - Phosphodiesterase (57)
  - Cytochrome P450 (56)
    - CYP_2 (17)
    - CYP_11 (8)
    - CYP_4 (5)
    - CYP_3 (5)
      - CYP_3A (5)
        - CYP_3A2 (1)
        - CYP_3A3 (1)
        - CYP_3A4 (1)
        - CYP_3A1 (1)
        - CYP_3A5 (1)
    - CYP_1 (4)
    - CYP_51 (3)
    - CYP_7 (3)
    - CYP_17 (2)
    - CYP_19 (2)
    - CYP_5 (2)
    - CYP_24 (1)
    - CYP_8 (1)
    - CYP_26 (1)
    - CYP_CAM (1)
    - CYP_21 (1)



ChEMBL Target Search Results: 1 Hits

	Target ID	Preferred Name	UniProt Accession	Gene Names	Description	Organism	Compounds	Endpoints
☑	11673	Cytochrome P450 3A4	P08684	CYP3A4; CYP3A3	Cytochrome P450 3A4	Homo sapiens	1305	1607

Compound ID	Target ID	Bioactivity	Activity Comment	Operator	Value	Units	Assay Type	Description	Target Name	Organism	Target Confidence	Reference	Name in Reference
529308	11673	IC50		=	1	nM	A	Inhibition of CYP3A4	Cytochrome P450 3A4	Homo sapiens	0	Bioorg. Med. Chem. Lett., (2008) 18:8:2725	7c
529265	11673	IC50		=	1	nM	A	Inhibition of CYP3A4	Cytochrome P450 3A4	Homo sapiens	0	Bioorg. Med. Chem. Lett., (2008) 18:8:2725	7f
402023	11673	IC50		=	4	nM	B	Selectivity towards cytochrome P450 3A4 enzyme activity	Cytochrome P450 3A4	Homo sapiens	8	Bioorg. Med. Chem. Lett., (2005) 15:6:1669	15d
402130	11673	IC50		=	4	nM	B	Selectivity towards cytochrome P450 3A4 enzyme activity	Cytochrome P450 3A4	Homo sapiens	8	Bioorg. Med. Chem. Lett., (2005) 15:6:1669	16b
419852	11673	IC50		=	6.3	nM	B	In vitro inhibitory concentration against Cytochrome P450 3A4	Cytochrome P450 3A4	Homo sapiens	9	J. Med. Chem., (2005) 48:7:2270	1

ChEMBL ADMET Data

- Currently ~200K datapoints
- Many in-vivo and in-vitro endpoints
- Manual curation in progress to "standardise" activity_types
- Extract from database where assay_type='A'

Can ADMET Models be built using ChEMBL data?

Data:
- P450 3A4 & 2C9
- PPB
- hERG
- Volume of distribution
- BBratio

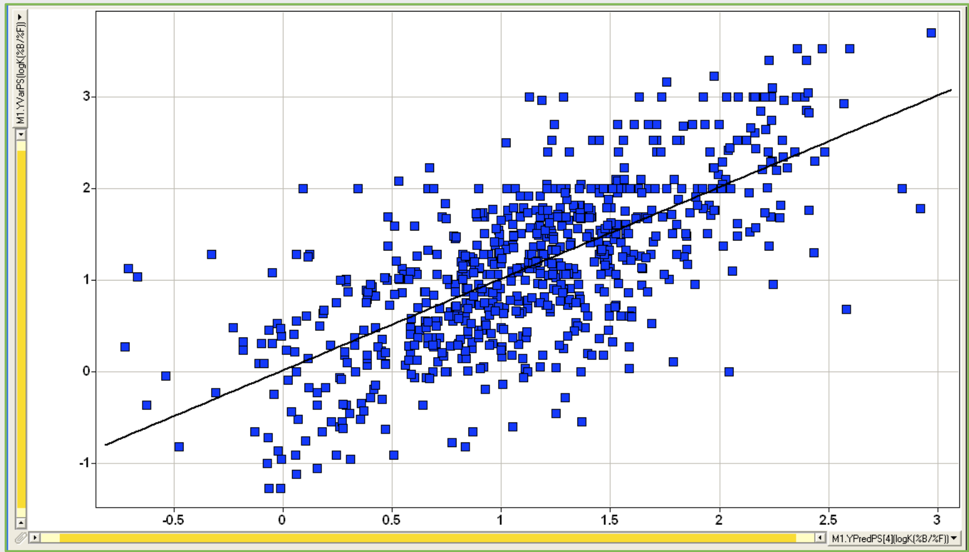
MWt>1000 removed
multiple values averaged

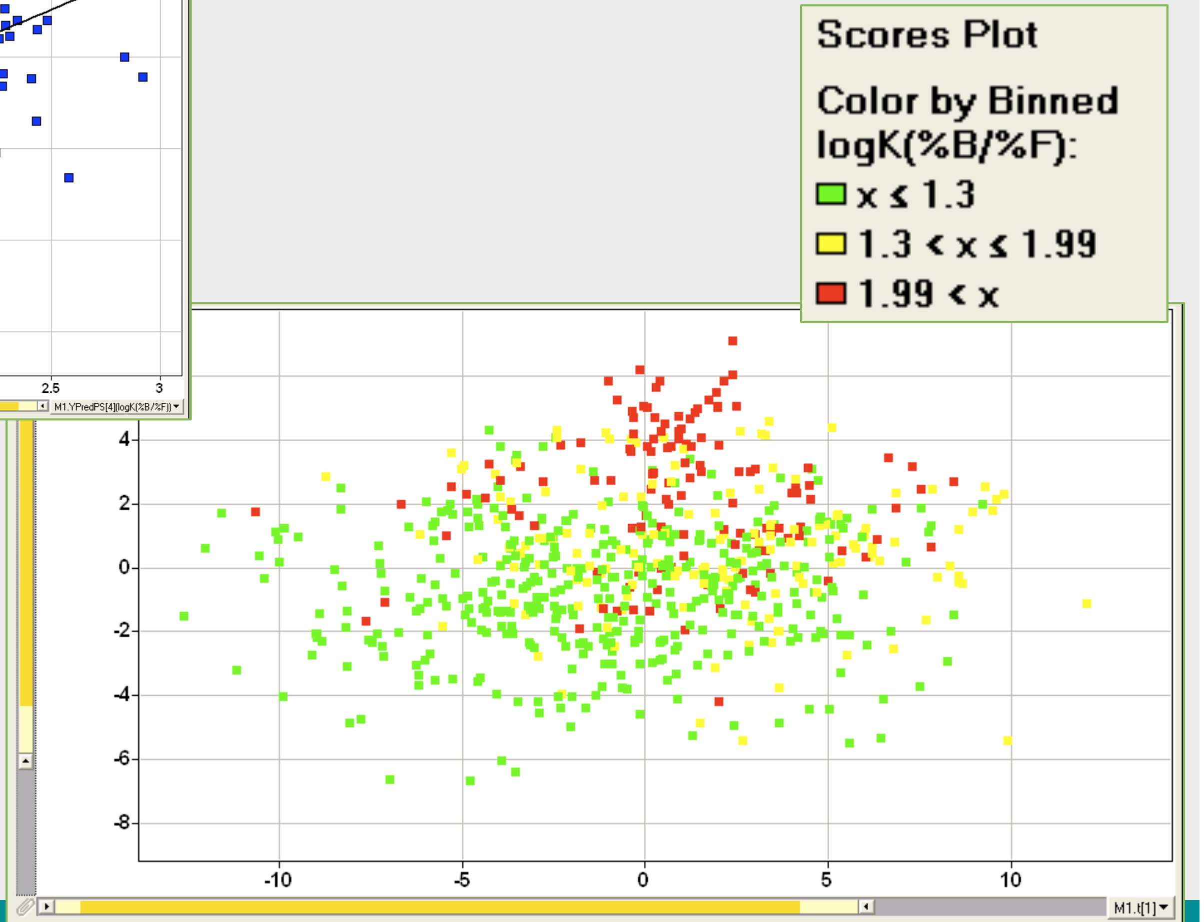
Descriptors:
- simple physchem & topological descriptors from pipeline pilot
- logP/logD/pKa from ACDlabs

PLS models

ChEMBL Models - Results

Plasma Protein Binding (logB/F) n=731

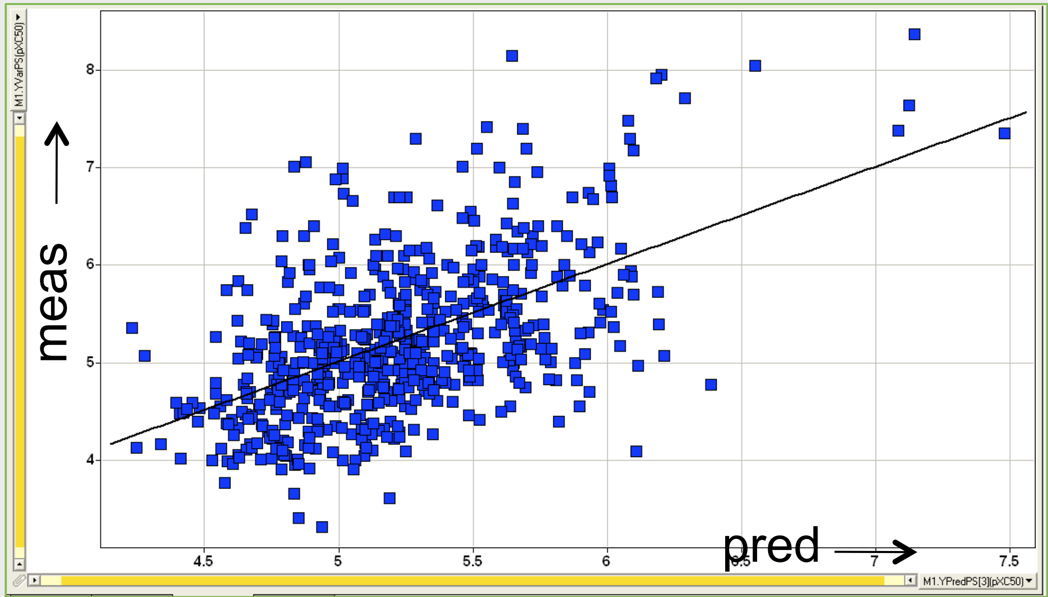


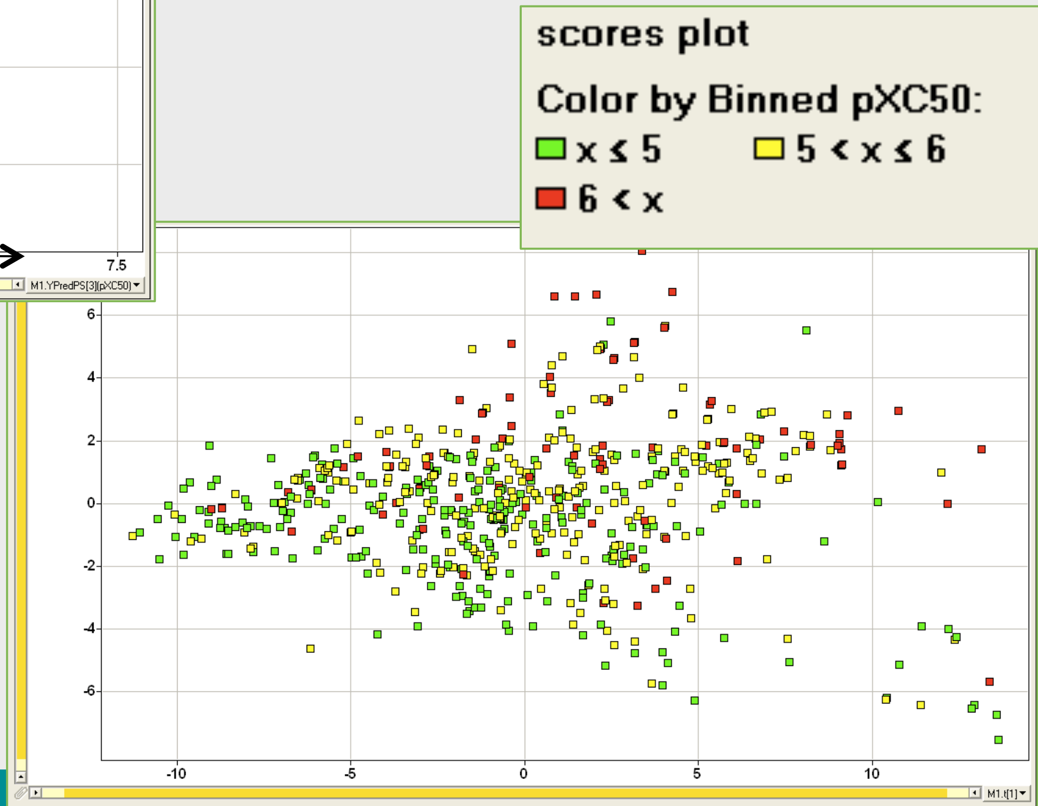


EMBL-EBI

ChEMBL Models - Results

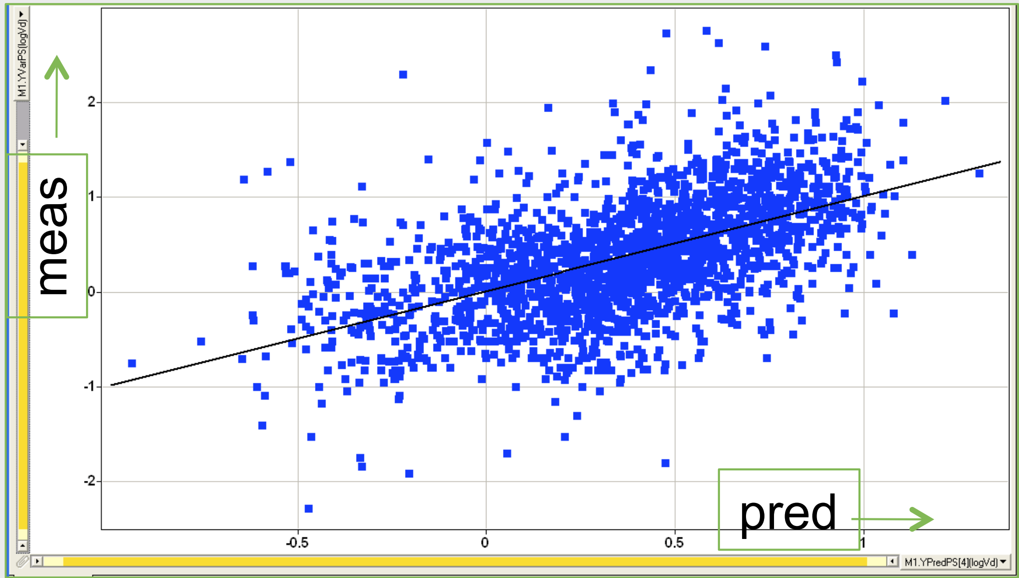
P450 2C9 Inhibition n=616

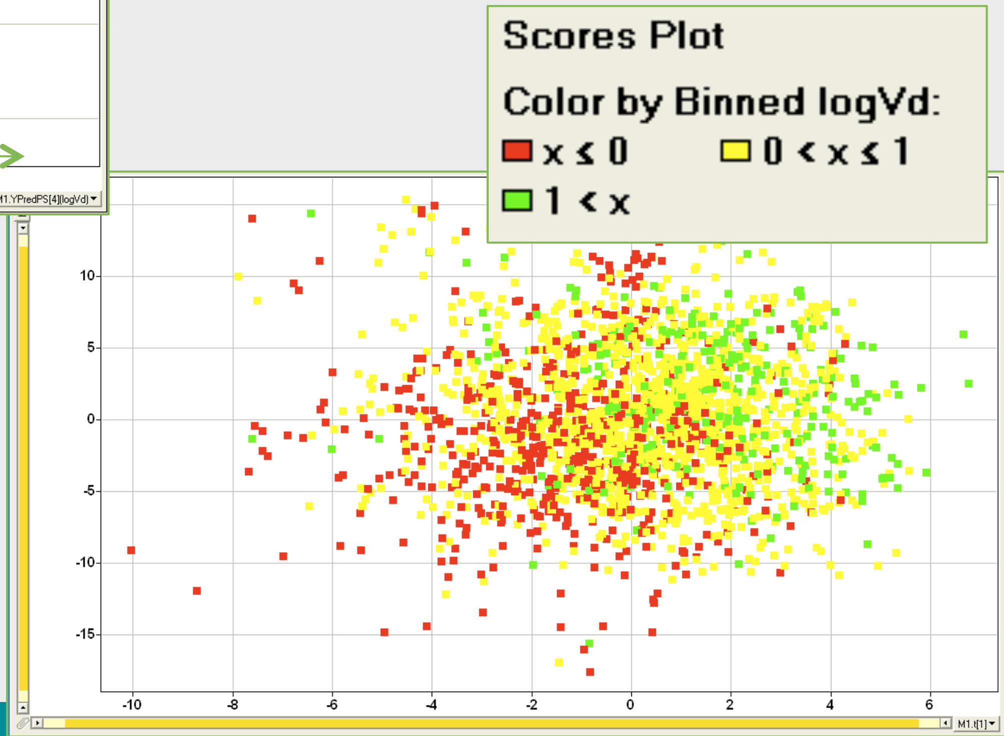




In-vivo Models from ChEMBL Database

Volume of Distribution (logVd) n=2227

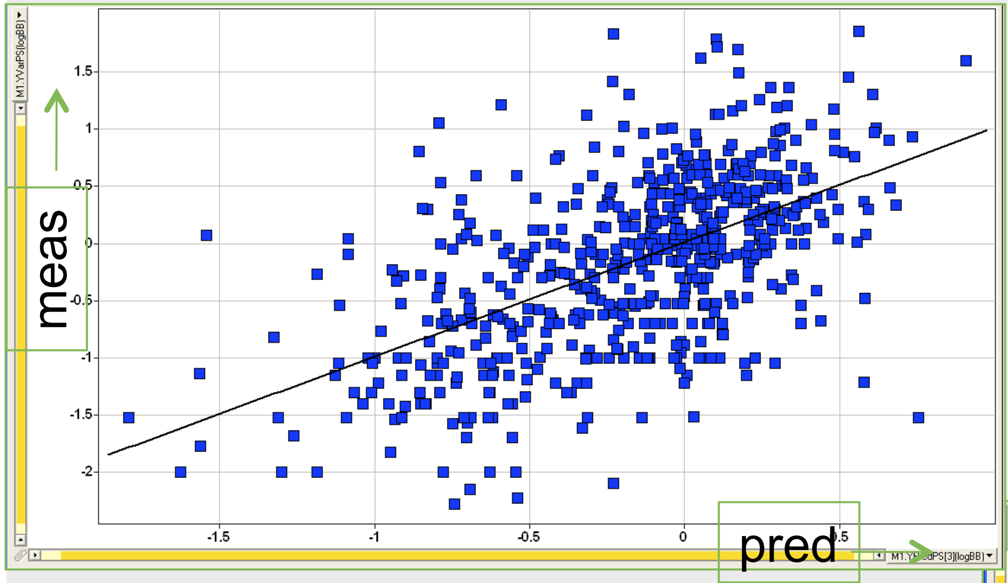


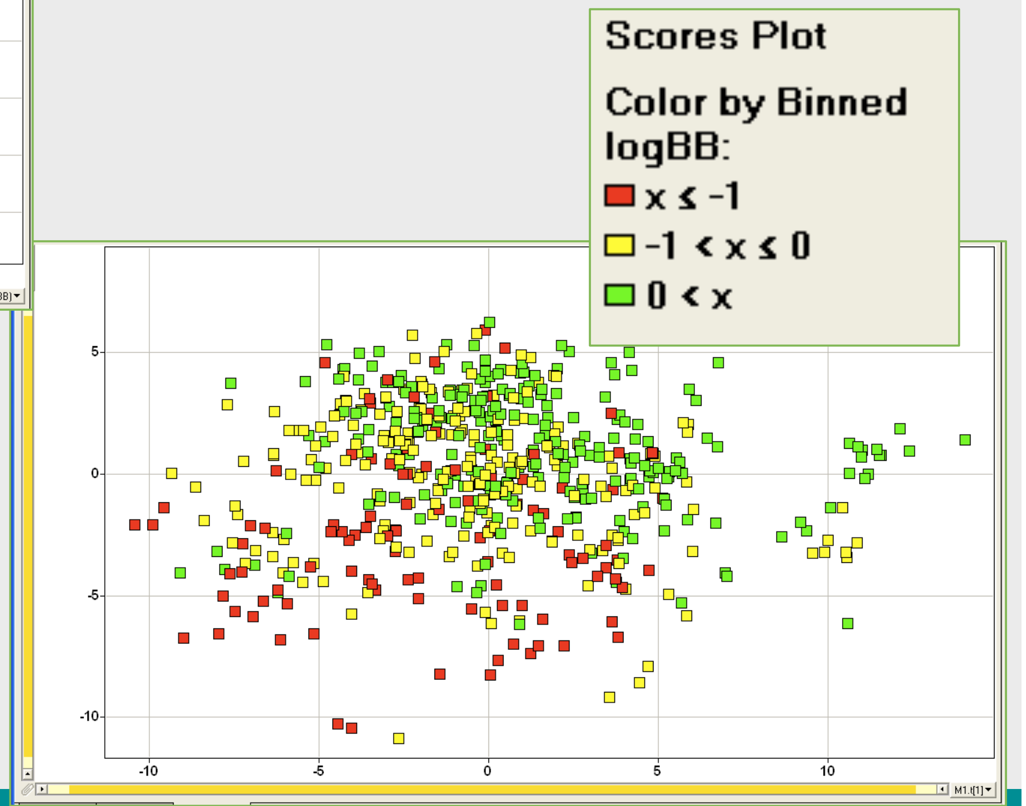


EMBL-EBI

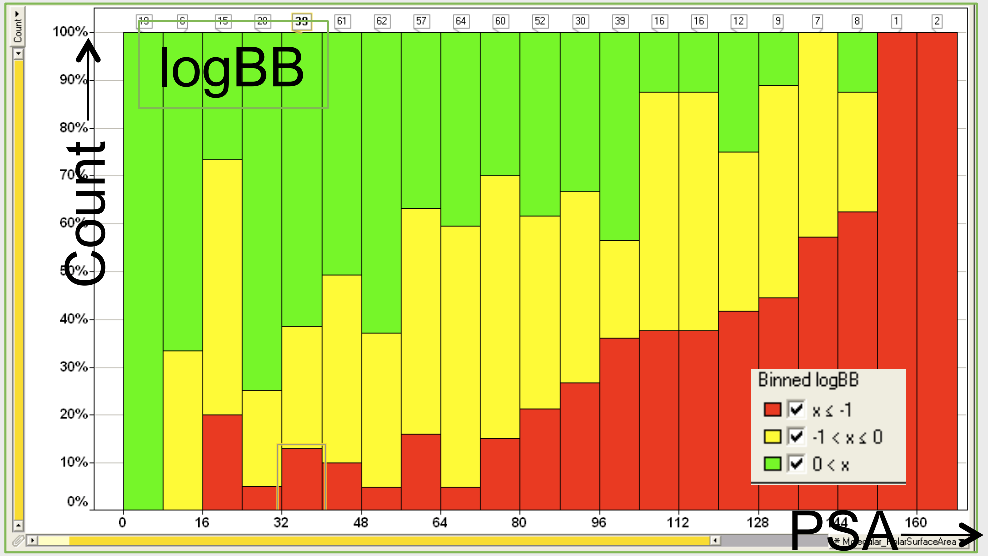
In-vivo Models from ChEMBL Database

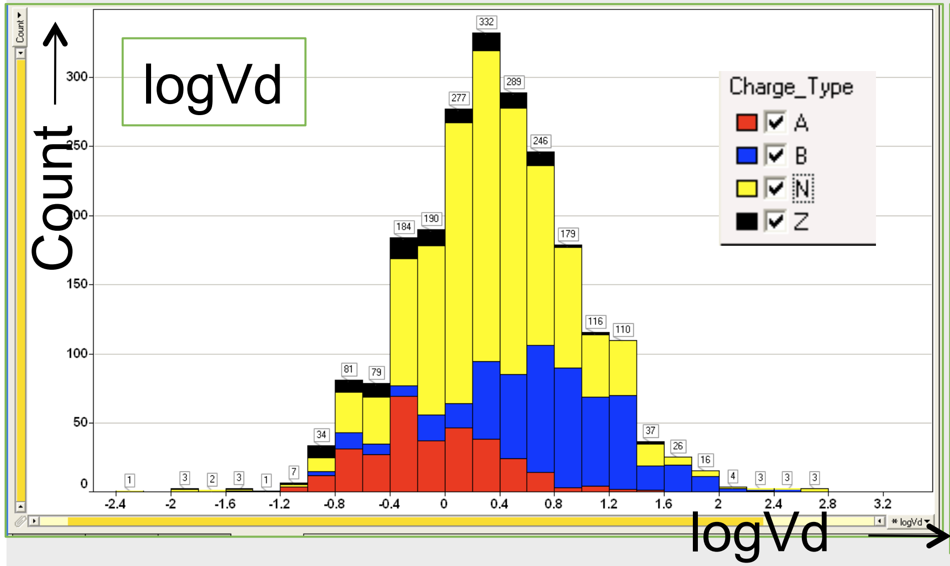
Brian/Blood Ratio (logBB) n=596

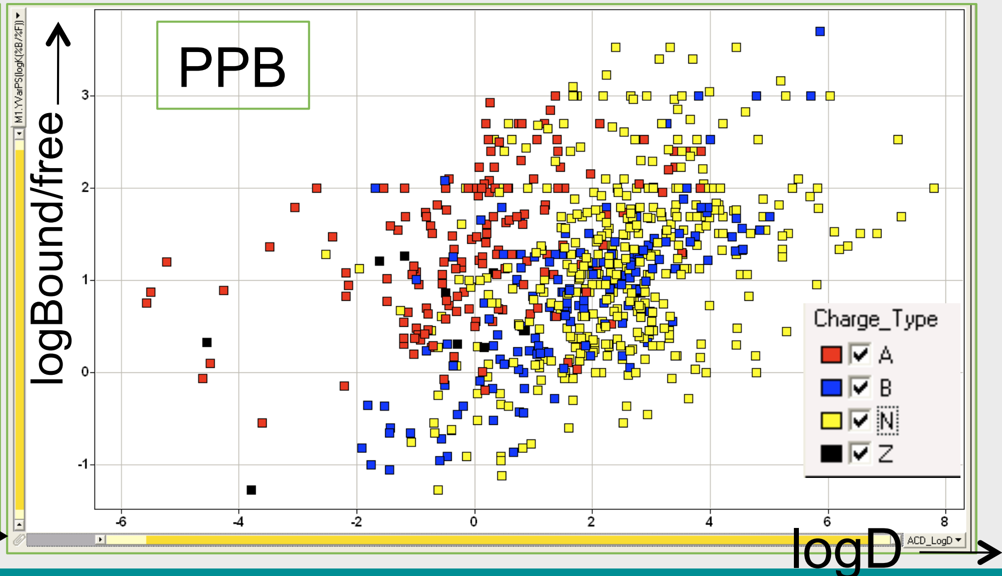




Descriptor Trends







EMBL-EBI

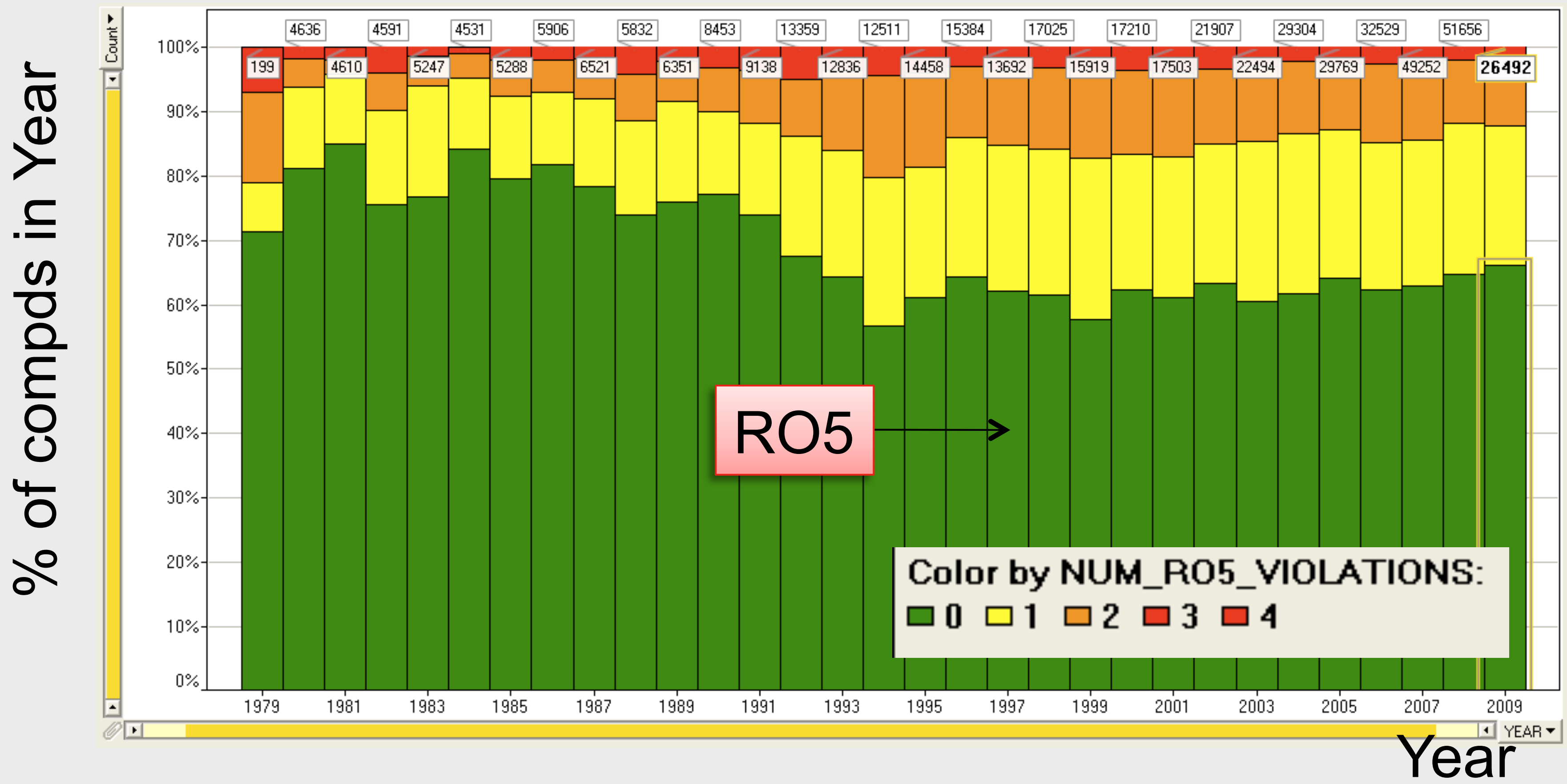
Reality of ADMET Prediction

Relatively easy to build ADMET Models

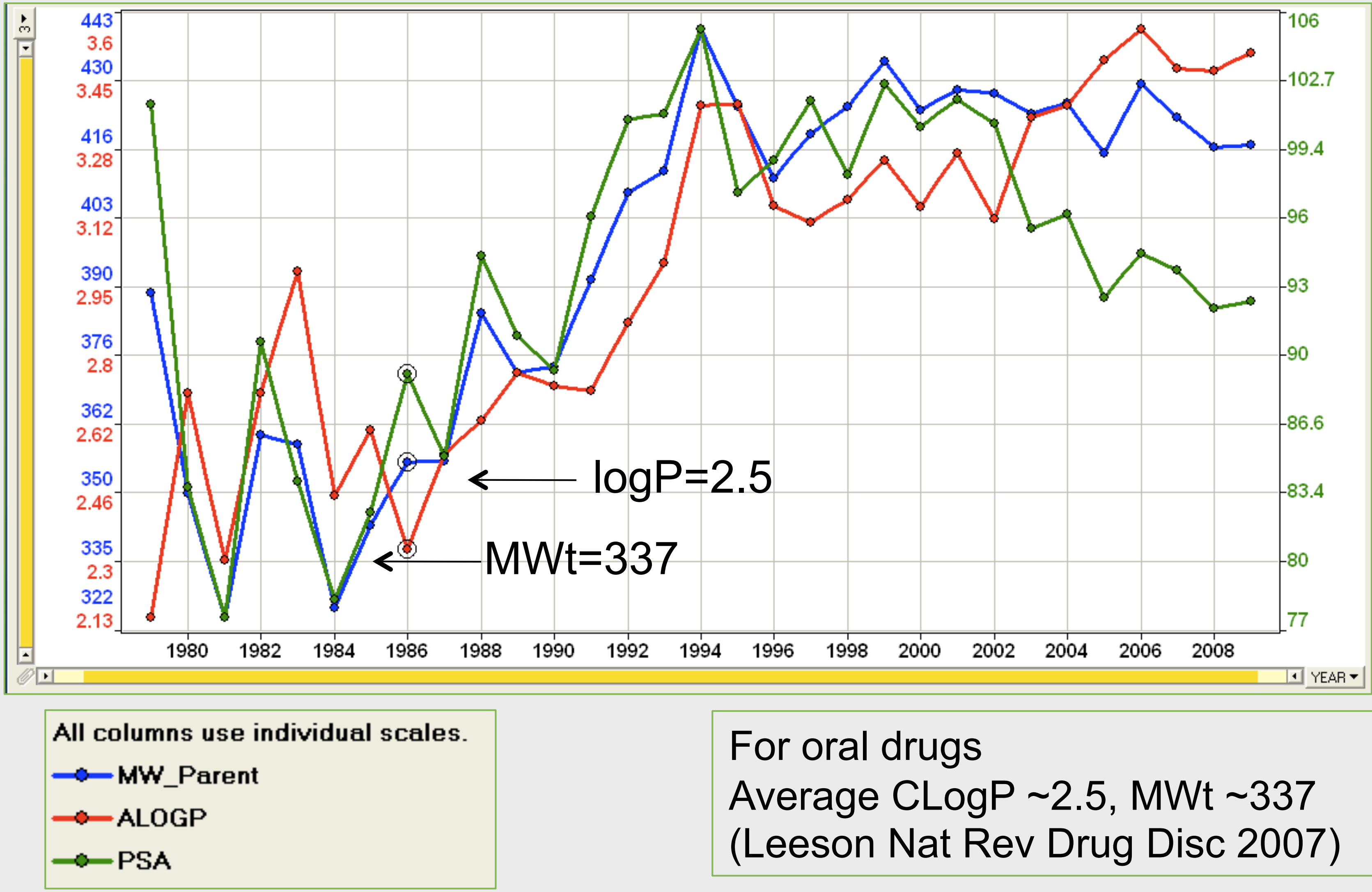
- Do we use them?
- How good are they?
- What are the issues?
- Are they useful?

Do we use them?

- ChEMBL extracts data from peer reviewed MedChem Journals e.g JMedChem 1980 onwards, EurJMedChem from 2007
- 485K Compounds (MWT>1000 not included)



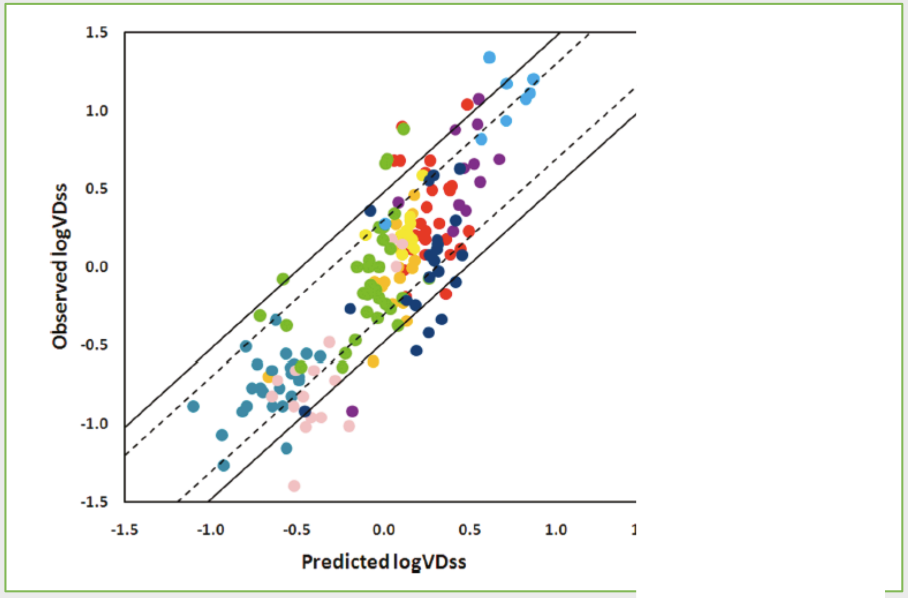
More Detail



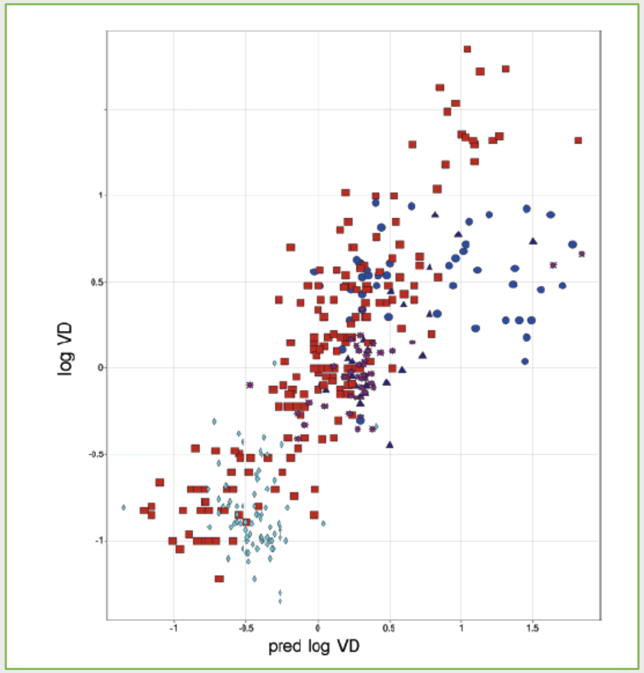
For oral drugs
Average CLogP ~2.5, MWt ~337
(Leeson Nat Rev Drug Disc 2007)

How Good are they?

Example – Vd Models



- Lombardo J.Med Chem., 2009



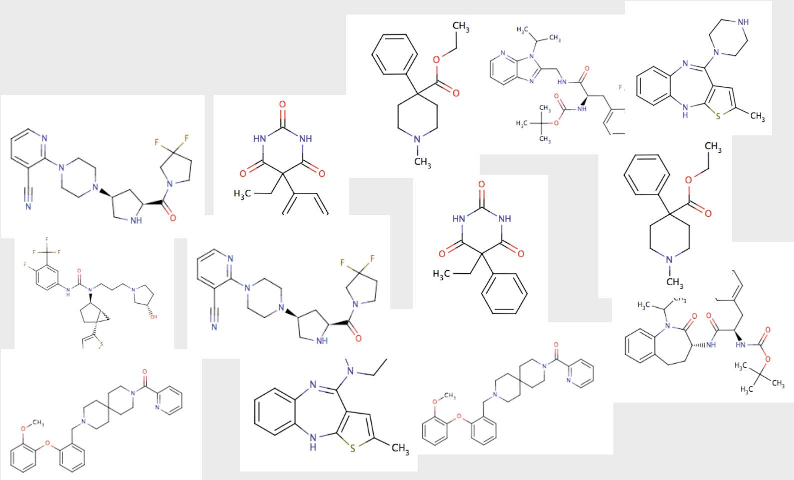
Valko et al, J MedChem 2006

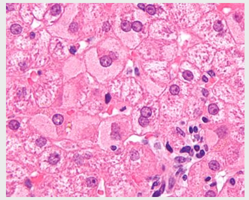
Good models at identifying trends
Less useful within a chemotype

Why is this?

EMBL-EBI

Why is it Difficult?





Ki IC50

Rate

%Bound

Vd

F

frag1 HBA PSA

logP Charge

?

Issues - Molecular Descriptors

logP

		Star set (N=223)		
		% of Molecules within Error Range		
	RMSE	<0.5	0.5–1	>1
ACD/logP	0.5	75	17	7
CLOGP	0.52	74	20	6
ALOGP	0.69	60	25	16

	Non-Star set (N=43)			
		% of Molecules within Error R		
	RMSE	<0.5	0.5–1	>1
ACD/logP	1	44	33	23
CLOGP	0.91	47	28	26
ALOGP	0.92	28	40	33

	Nycomed (N=882)			
	RMSE	<0.5	0.5–1	>1
ACD/logP	0.87	46	34	21
CLOGP	1.01	46	28	22
ALOGP	0.72	52	33	15

	Pfizer set (N=95809)			
	RMSE	<0.5	0.5–1	>1
ACD/logP	1.28	35	27	38
CLOGP	1.23	37	28	35
ALOGP	1.12	39	29	32

Mannhold et al J Pharm Sci 2008

Most descriptors there is no independent way of measuring them

Model Prediction Space

- Helps to identify which compounds are well predicted
- Molecules outside training set property space are poorly predicted
- Models get worse with time as “property space” of molecules changes
- Difficult to do for properties not represented in training set
- Simple example - BBB models
- Abraham (1994) n=57 diverse compounds
- New compounds – model descriptors in original range but obvious outliers

Compound	Meas logBB	Calc Old eqtn
Indomethacin	-1.26	-0.1
Ibuprofen	-0.18	0.39

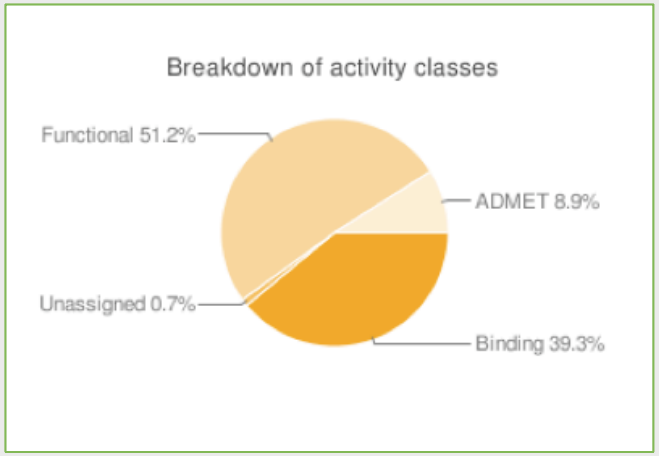
- Platts (2001) n=148 new compounds

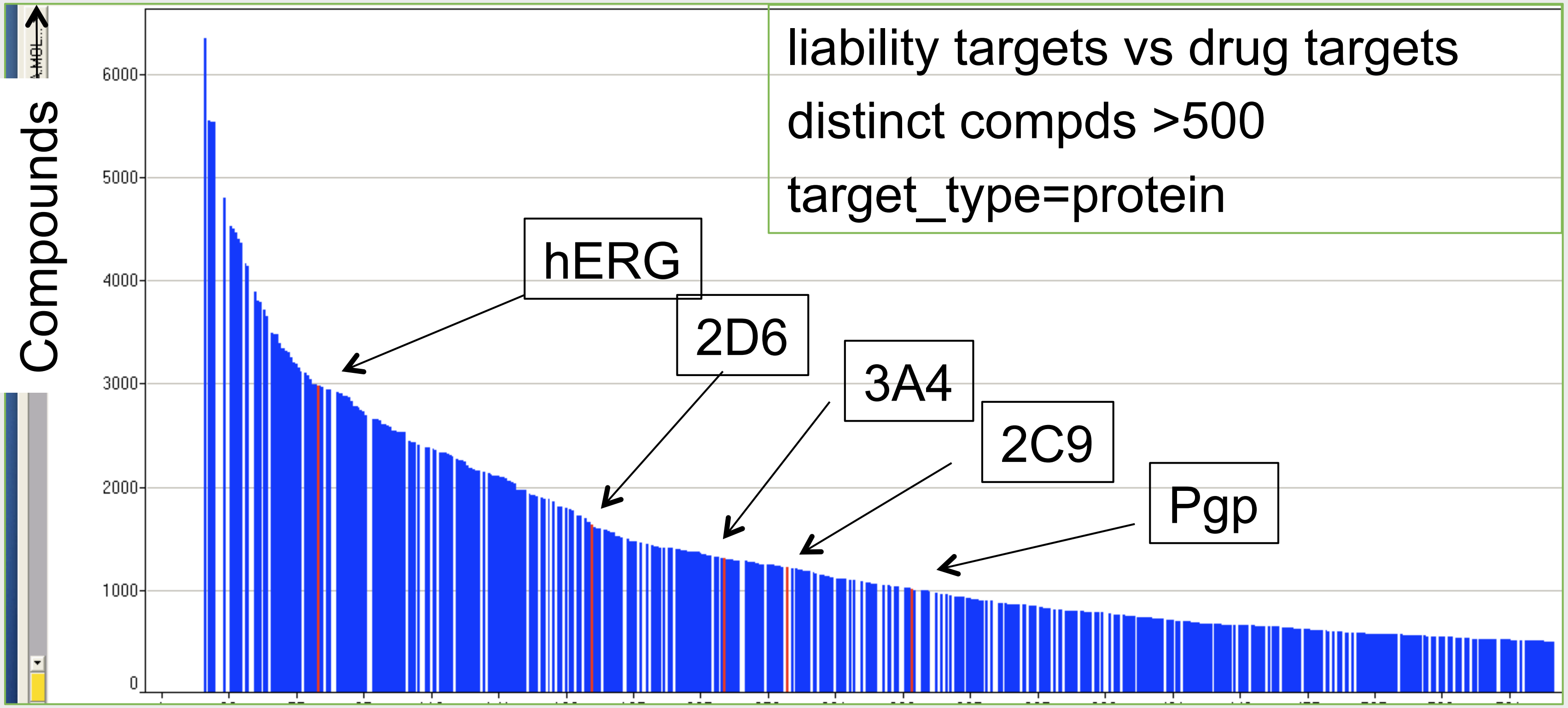
Compound	Meas logBB	Calc Old eqtn	Calc New eqtn
Indomethacin	-1.26	-0.1	-0.92
Ibuprofen	-0.18	0.39	-0.23

acids – class not represented in original dataset

Data

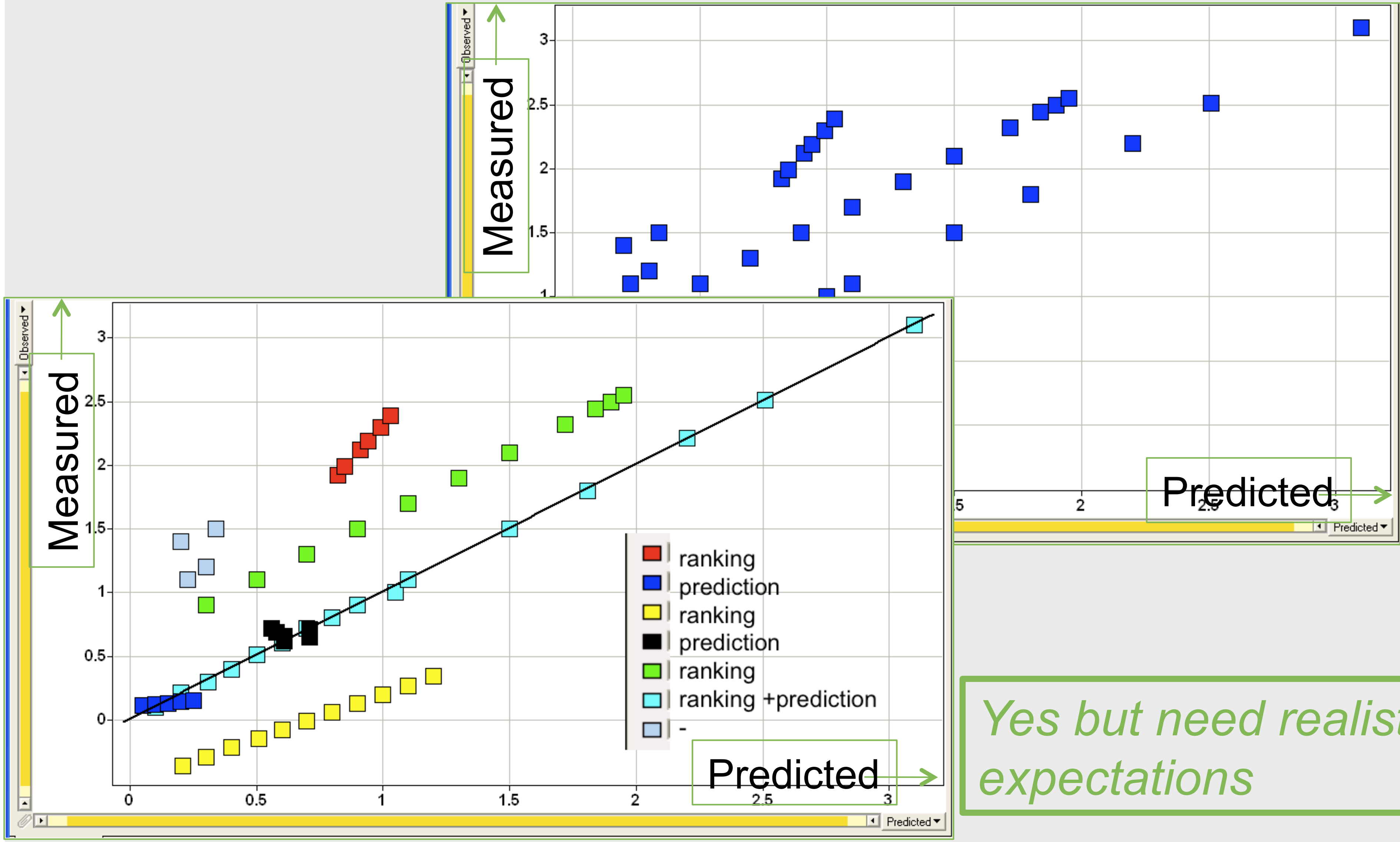
- Do we measure enough?
- Is it the right data (project cascade effect)
- More mechanistic information?





protein target

Are Predictions Useful?



Yes but need realistic expectations

EMBL-EBI

"How modern state of the art methods can accelerate the process of drug design?"

- Use the information from the simple predictions
- Reduce logP & MWT
- Focus more on designing molecules with good ADMET and less on increasing potency

- More comprehensive data
- Data sharing – publish more ADMET data
- Better descriptors

Acknowledgements

John Overington

Louisa Bellis
Patricia Bento
Jon Chambers
Mark Davies
Anna Gaulton
Kaz Ikeda

Felix Krueger
Yvonne Light
Shaun McGlinchey
Karen Bonner

Bissan Al-Lazikani

EMBL-EBI