From data production to hypothesis based project support

Extracting knowledge from data

10th PhysChem Forum, March 22-23 2011, Bracknell, UK

Bernard Faller

Novartis Institutes for BioMedical Research, Basel, Switzerland

Director, Discovery ADME

NOVARTIS

From data generation to hypothesis building

1990

2000

2010

NOVARTIS

What is a hypothesis ?

10th PhysChem Forum
March 22-23 2011

a testable idea

which may evolve as additional information becomes available

The mission of the scientist is to formulate relevant questions and design experiments to test it

NOVARTIS

What has happened ?

NOVARTIS

Acknowledgments

- Giorgio Ottaviani
- Giuliano Berellini
- Alfred Zimmerlin
- Caroline Rynn

NOVARTIS

Outline

- Getting started
- Multi dimensional optimization
- Local models to build hypotheses
- Extracting information in the absence of correlation
- Potential and limitations of in-vitro & in-silico approaches
- Conclusions/outlook

Target product profile

Defines the relevant assays and thresholds

- Dose, route of administration
- Target location
  - Periphery/central
  - Extracellular/intracellular
- PK/PD aspects
  - Cmax, free Cmax driven
  - AUC (free AUC) driven
  - Trough level
  - Duration of action

NOVARTIS

Gathering and connecting data

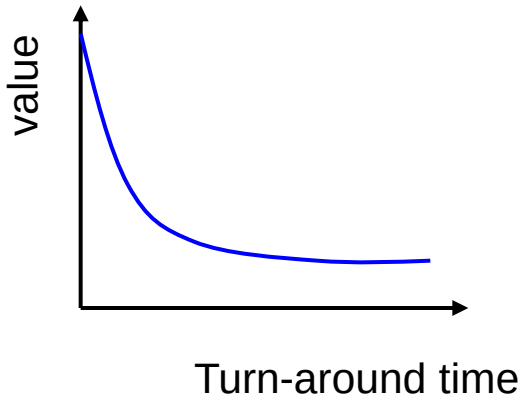
- Molecular Descriptor generators
  - Physicochemical properties
  - Structural
- Wet testing
  - In-vitro Primary – secondary – tertiary assays
  - Custom tailored assays to validate/invalidate a hypothesis
  - In-vivo PK (concentration vs. time profiles)
- Model building engine
  - Testing the “right” compounds
  - Generate hypotheses (PCA/PLS models, PBPK approaches)
  - Parameter sensitivity (PBPK, e-Numerics)

Assays packages and hierarchical testing

- Assay packages: addressing the same overall process
  - Solubility and permeability: absorption
  - Ionization and lipophilicity: logD
- Hierarchical testing
  - Primary assay: high-throughput, fast turn-around time
  - Follow-up assay: provides mechanistic understanding
    - Example of solubility: primary at pH6.8 followed by solubility pH-profile to separate solubility and ionization
    - Example of TDI: single concentration followed by K_i, K_{inact}

Properties of assay packages

- Primary assays
  - Thousands of compounds can be tested
  - “Undefined” compounds accepted
  - Need to spot check assay predictive power



- Secondary assays: build and test hypothesis
  - Question I want to answer and follow-up action need to be clear before I do the experiment

NOVARTIS

Multidimensional optimization

reconciling different properties within ONE molecule

stability

selectivity

solubility

potency

h-ERG

reactive metabolites

permeability

Half-life

metabolic CL

VdSS

PPB

CYP-450 inhibition

PKPD

NOVARTIS

Metadata mask information

Getting started

Multidimensional optimization

Local models to build hypotheses

Extracting information
In the absence of correlation

Potential and limitations

Conclusions

	In-vivo				In-vitro								In-silico						
cmpd	**VSS (L/kg)**	Cl	t1/2	exp %FU 7.4	HTSol6.8	logPAMPA	logPGI	FA%	logPe4	logPe6.8	logPe8	RLM	mw	psa	clogp	flex_index	AlphaH2 (A)	BetaH2 (B)	Pi2 (S)
1	4.4	157.7	0.99	100	0.004	-4.2	-4.2	96	-4.2	-4.3	-4.6	21.8	452.5	97.8	4.3	17.7	0.13	1.9	3.9
2	2.29	157	0.28	95	0.004	-5.1	-5.3	50	-5.3	-6.1	-6.4	27.7	547.7	106.1	4.7	14.6	0.13	2.6	4.6
3	23.69	116.2	4.88	1	0.42	-4.2	-4.2	96	-7.1	-5.4	-4.2		461.6	82.6	4.3	10.8	0.13	2.0	3.5
4	1.25	17.9	1.12	5	0.805	-4	-4	98	-6.2	-4.6	-4	248.5	462.6	85.8	3.7	10.8	0.13	2.2	3.8
5	4.8	35	1.69	5	0.014	-4.3	-4.3	95	-9.3	-4.9	-4.3	48.7	462.6	85.8	3.7	10.8	0.13	2.3	3.8
6	0.63	12.6	1.23	100	0.015	-5.2	-5.4	46	-5.6	-5.4	-5.7	129.8	477.6	108.9	2.3	10.5	0.13	2.3	4.4
7	1.97	51.6	0.51	69	0.004	-4.2	-4.2	96	-6.2	-4.4	-4.2	513.3	493.6	101.1	3.7	16.2	0.13	2.4	3.9
8	9.64	67.6	2.02	2	0.284	-4.5	-4.6	88	-5.9	-5.3	-4.6		476.6	85.8	4.0	12.6	0.13	2.5	3.7
9	25.96	108.7	1.81	5	0.004	-4.9	-5.1	66	-6	-5.7	-5.1	214.8	509.7	102.5	5.8	15.7	0.62	2.5	3.4
10	3.64	64	0.89	20								331.9	490.6	102.9	2.9	10.2	0.13	2.6	4.4
11	2.91	21.8	1.57	100	0.004	-3.8	-3.8	99	-3.8	-3.9	-4.1		427.5	74.8	4.8	11.7	0.13	1.8	3.5
12	22.36	121.3	1.14	100	0.004	-3.9	-3.9	99	-7.3	-3.9	-4	334.2	463.6	91.8	3.4	12.9	0.13	2.2	3.7
13	1.38	40.6	0.6	100	0.004	-3.9	-3.9	99	-3.9	-3.9	-4	826.2	461.5	84.0	4.3	13.0	0.18	1.9	3.6
14	7.84	62.2	2.5	6	0.825	-5.3	-5.9	35	-6.2	-7.2	-5.9	75.2	478.6	111.7	2.8	16.7	0.4	2.5	4.3
15	9.09	66.3	2.34	100	0.004	-4.5	-4.5	91	-4.5	-4.5	-4.8	373.4	481.5	74.8	5.3	12.5	0.18	1.6	3.3
16	6.13	80.5	1.57	100	0.007	-4.6	-4.7	85	-4.7	-4.9	-5.2	107.5	422.5	88.6	4.4	14.2	0.13	1.7	3.4
17	3.97	69.2	1.26	100	0.004	-4	-4	98	-4.1	-4	-4.2	124.8	401.5	65.5	4.5	10.0	0.18	1.5	3.3
18	11.23	111.6	2.51	100	0.004	-4.6	-4.7	86	-4.7	-5.1	-5.3	108.9	346.4	62.3	4.9	11.6	0.18	1.2	2.9
19	7.02	73	2.51	94	0.046	-3.9	-3.9	99	-3.9	-4.1	-4.3		473.6	76.1	5.6	19.0	0.13	2.0	3.4
20	4.51	66.1	1.3	25	0.023	-4	-4	98	-6.9	-4.1	-4	604.1	474.5	78.0	4.7	12.6	0.18	2.2	3.6
21	8.5	54	2.9	24	0.004							4.6	454.6	68.8	5.8	11.0	0.13	2.0	3.4
22	4.72	34	1.94	100	0.005	-4	-4	98	-4.1	-4	-4.1	196.6	465.5	84.0	4.0	10.7	0.18	1.9	3.6
23	6.1	22	4.3	100	0.004	-4.5	-4.5	91	-4.5	-4.7	-5	100.1	445.5	74.8	5.0	9.0	0.13	1.8	3.4
24	13.7	35		99	0.004	-3.8	-3.8	99	-3.8	-5	-5.2	163.6	446.5	81.5	4.8	11.2	0.18	1.7	3.2
25	6	49		100	0.007	-3.6	-3.6	99	-3.6	-4.6	-4.8		494.5	68.8	5.5	10.1	0.18	2.0	3.4
26	2.7	18		98	0.019	-4	-4	98	-5.5	-4	-4.2	59.9	414.5	87.6	3.0	9.7	0.13	2.0	3.6
27	2.6	66		100	0.004	-4.3	-4.3	95	-4.4	-4.3	-4.3	202.5	439.4	95.0	2.4	9.1	0.13	2.0	3.6
28	5	91		16	0.01	-4.6	-4.7	85	-6.5	-5.1	-4.7	924	550.6	97.3	3.9	14.5	0.36	3.0	3.5
29	12.1	49		1	0.054	-4.5	-4.6	88	-6.1	-5.5	-4.6	364.7	534.6	85.8	4.9	11.2	0.29	2.6	3.3
30	4.2	16	2.7	99	0.004	-4.4	-4.4	93	-4.6	-4.4	-4.4	110.9	551.6	90.9	3.8	16.3	0	2.5	4.0
31	2.9	31	1.5	98	0.004	-4.1	-4.1	97	-4.8	-4.1	-4.4	185.2	506.6	70.8	4.9	11.8	0.13	2.0	3.4
32	4	18	3.5	97	0.004	-4	-4	98	-4	-4	-4.1	155.6	528.6	58.6	5.0	9.5	0	2.0	3.3
33	8	27	5.1	5	0.093	-4.7	-4.8	79	-5.8	-5.5	-4.8	115.5	597.7	91.2	4.7	11.7	0.13	2.7	4.0
34	8.8	18	7.2	2	0.605	-4.5	-4.6	88	-6.3	-5	-4.6	199.7	490.6	70.6	3.1	10.2	0.13	2.4	2.3
35	8.9	32	3.9	14	0.006	-4.8	-4.9	75	-5.9	-5.9	-4.9	39.2	549.6	82.9	5.2	16.4	0.4	2.5	3.6
36	5.4	39	2.6	100	0.004	-4.3	-4.3	95	-4.3	-4.5	-4.5	254.4	509.6	91.8	4.9	13.7	0.13	2.2	3.7
37	8.3	42	5.5	98	0.006	-3.5	-3.5	100	-4.5	-3.5	-3.8	520.9	422.4	70.6	3.9	11.8	0.13	1.9	2.9
38	7.2	87	2.1	45	0.004	-4.6	-4.7	85	-6.2	-4.9	-4.7	70.9	484.6	89.0	5.2	14.4	0.36	2.4	3.6
39	3.4	31	1.5	100	0.008	-3.9	-3.9	99	-4.6	-3.9	-4.2	94.2	436.4	87.6	2.8	9.2	0.13	2.0	3.6
40	6.7	101	1.5	99	0.004	-3.9	-3.9	99	-3.9	-5.6	-6	46.7	479.5	66.9	6.0	14.6	0.13	1.8	2.9
41	8.5	42	2.2	98	0.004	-3.9	-3.9	99	-3.9	-5.3	-5.9	152.3	449.5	57.7	6.2	11.1	0.13	1.6	2.6
42	26.7	46	9.8	4	0.036	-4.5	-4.5	90	-7.6	-5	-4.5	165	547.7	73.0	6.0	12.8	0.29	2.6	3.1
43	2.2	54	0.5	99	0.032	-5.2	-5.6	44	-6	-5.6	-5.9	200.9	533.6	90.0	4.7	13.1	0.4	2.7	3.6
44	15.9	58	5.2	9	0.893	-3.9	-3.9	99	-7.3	-4.6	-3.9	309.9	534.6	85.8	4.9	11.2	0.29	2.6	3.3
45	9.5	46	3.7	7	0.004	-5.1	-5.6	49	-6.2	-7.3	-5.6	95.5	477.6	107.5	3.3	10.5	0.29	2.5	3.9
46	2.4	31	1.3	100	0.036	-5.5	-6.6	19	-7.9	-6.6	-6.8	124.1	543.6	107.1	3.9	11.0	0.4	2.9	4.3
47	12.3	46	4.4	1	0.118	-4.4	-4.4	93	-5.8	-5.1	-4.4	50.8	533.6	73.0	5.5	13.1	0.29	2.5	3.1
48	3.4	16	3	98	0.27	-4.6	-4.7	85	-6	-4.7	-5	330.4	547.6	90.0	5.1	12.8	0.4	2.7	3.6
49	11.9	50	3.7	1	0.766	-5.3	-5.6	37	-5.6	-6.8	-7.2		547.6	105.7	4.6	11.0	0.56	3.0	3.9
50	60.9	70	13.2	1	0.807	-5	-5.2	56	-6.3	-6.5	-5.2	154	524.7	108.5	6.5	21.0	0.62	2.3	3.3
51	9.6	45	3.6	1	0.004	-5	-5.2	60	-6.7	-7.2	-5.2	130.8	534.7	119.5	5.2	13.1	0.65	2.8	3.8
52	10.8	47	2.7	61	0.877	-4	-4	98	-5.3	-4.3	-4	577.5	575.7	90.0	5.5	10.4	0.29	2.7	3.8
53	24.4	32	10.7	0	1	-4.3	-4.3	95	-7.3	-4.4	-4.3	49.5	561.7	73.0	6.5	12.5	0.29	2.6	3.1
54	7.5	38	3.4	11	0.036	-3.3	-3.3	100	-5.9	-3.3	-3.4	91.2	435.5	73.4	5.2	9.2	0.29	1.9	2.9
55	9.3	39	4.3	0	1	-4.7	-4.7	83	-5.6	-5.7	-4.7	96.9	557.7	90.0	5.3	10.8	0.29	2.7	3.8
56	9.5	44	4.4	9	1	-5	-5.2	60	-5.2	-5.9	-5.2	554.4	532.6	93.7	4.1	11.3	0.29	2.6	3.7
57	15.2	10	18.2	0	0.988	-5.3	-5.9	37	-5.9	-6.9	-7	19.4	520.6	85.8	3.6	13.5	0.29	2.8	3.3
58	12.9	93	2.8	0	0.425	-4.6	-4.7	83	-5.3	-6.5	-4.7	171.1	545.6	80.8	5.2	12.8	0.29	2.6	3.5
59	1.7	22	2.2	1	0.966	-4.9	-5	68	-5.1	-5.3	-5	213.2	531.6	80.8	4.6	13.2	0.29	2.5	3.4
60	17.5	28	9.8	0	0.01	-4.6	-4.6	87	-6.6	-6	-4.6	97.6	533.7	106.7	5.7	15.0	0.65	2.7	3.5
61	8.4	50	2	0	0.372	-4.3	-4.3	95	-7.1	-5.7	-4.3	68.6	577.7	80.3	4.8	15.6	0.13	3.1	3.5
62	4.6	78	1.5	86	0.004	-5	-5.2	60	-6.4	-5.6	-5.2	213.2	548.6	94.1	3.4	10.9	0.13	3.0	4.0
63	18.8	63	5.2	80	1	-4	-4	98	-5.1	-4.8	-4	171.1	548.7	85.8	5.4	10.9	0.29	2.6	3.3
64	11	54	6.5	97	0.004	-3.6	-3.6	99	-3.6	-4.8	-5	61.6	540.6	60.9	6.0	13.0	0.13	2.0	2.8
65	35.6	20	27.3	2	0.035	-5.4	-9.1	31	-9.1	-9.1	-9.2	144.4	505.6	81.7	4.3	11.9	0.42	2.5	3.1
66	13.7	72	3.2	28	0.209	-4.8	-5	72	-7.1	-5.2	-5	101.2	520.6	81.2	5.2	13.5	0.33	2.5	3.1
67	5.2	55	1.7	98	0.031	-4.1	-4.1	97	-4.7	-4.1	-4.1	195.2	561.6	81.2	5.8	12.5	0.13	2.8	3.6
68	2.7	63	1.1	67	1	-5.2	-5.5	46	-7	-5.8	-5.5	3.4	548.6	94.1	4.0	12.8	0.13	3.1	3.8
69	17.1	58	5	7	0.016	-3.8	-3.8	99	-5	-3.8	-3.9	69	569.1	85.8	5.7	10.5	0.36	2.5	3.4
70	22.2	66	5	17	1	-4	-4	98	-6	-4.2	-4	374.6	559.6	109.6	4.4	10.7	0.36	2.7	3.8
71	7.5	134	1.3	64	0.013	-4.5	-4.6	88	-5.6	-4.6	-4.8	198	547.6	90.0	5.0	12.8	0.28	2.7	3.6
72	1.8	43	0.8	70	0.037	-4.3	-4.3	95	-6.4	-4.5	-4.3	195.2	548.6	102.9	4.5	10.9	0.28	2.9	3.6
73	3.5	112	0.7	80	1	-5.4	-6.9	28	-6.9	-7.8	-7.5	75.7	534.6	102.9	3.6	11.2	0.28	2.9	3.6

Extracting value from data

NOVARTIS

Different dimensions, different dynamic range...

Parameter	Unit	Range (for drug like molecules)
MW	$g.mol^{-1}$	100-1000
PSA	A^2	10 to 200
Potency	M	10^{-10} to 10^{-6}
Dose	$g.kg^{-1}$	0.0001 to 1
Solubility	M	10^{-9} to 10^{-2}
Tm	K	320-580
ΔHf,m	$kJ.mol^{-1}$	10-100
LogP		-2 to +8
pKa*		0 to 12
permeability	$cm.s^{-1}$	10^{-7} to 10^{-2}
ER		0 to 10 (Caco), 0 to 100 (MDCK)
Vss	$L.kg^{-1}$	0.1 to 100
CL (int)	$\mu l.min^{-1}.mg^{-1}$	20-200
PPB	%	0-100

*: requires transformation to FI

NOVARTIS

...and its consequence

- Different dimension and dynamic range
  - Normalization required before statistical analysis
- Parameters of similar dimension can be +, -, x or /
  - Potency + solubility (M)
  - Potency / solubility (dimensionless)
- Parameters of different dimension can only be x or /
  - Solubility x permeability (flux factor) M. $cm.s^{-1}$
  - $\log P_o = \log P_{mem} + \log(D/h)$
  - $P_o = P_{mem} \times D/h$

    $cm.s^{-1}$ $cm^2.s^{-1}/cm$

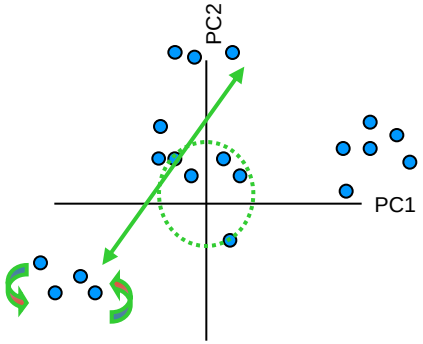
P_o = intrinsic membrane permeability, P_{mem} = membrane partition coefficient , D = diffusion coefficient within the membrane, h = membrane thickness

NOVARTIS

Compressing information

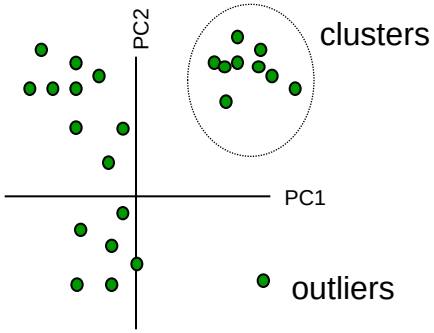
PCA produces two useful and straightforward diagrams

• The loading plot contains information about the variables: it is composed of few vectors (Principal Components, PCs) which are obtained as linear combinations of the original X-variables



to highlight the variables
which contain
similar/independent information

• The score plot contains information about the objects: each object is described in terms of its projection onto the PCs, (instead of the original variables)

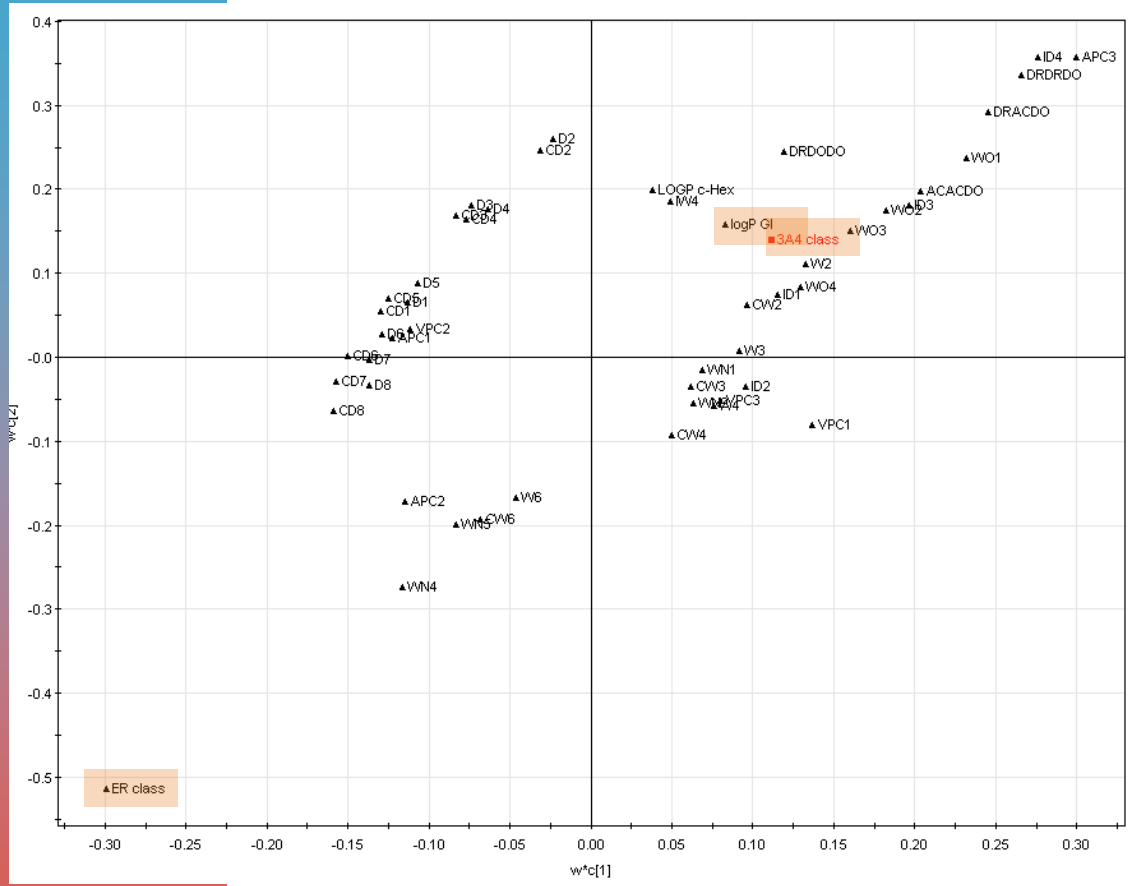


to understand the
distribution of the objects

NOVARTIS

Going beyond plotting X vs. Y

Analysis of chemotype A



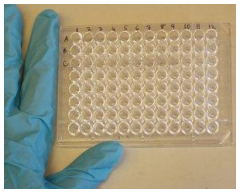
- Difficult to reconcile 3A4 activity and brain exposure with this chemotype
- Need to find optimal balance/compromise

- 3A4 activity and ER and negatively correlated
- 3A4 and Passive perm. are positively correlated

NOVARTIS

What are models good for ?

+



- Prediction of in-vivo response
- Link readout with chemical descriptors and...
- ...Formulate hypothesis
- **Global models**: to get started
  - Useful to manipulate large number of compounds
  - Rarely going beyond text book knowledge
- **Local models**: more information rich but restricted to local chemical space

NOVARTIS

The power of local models

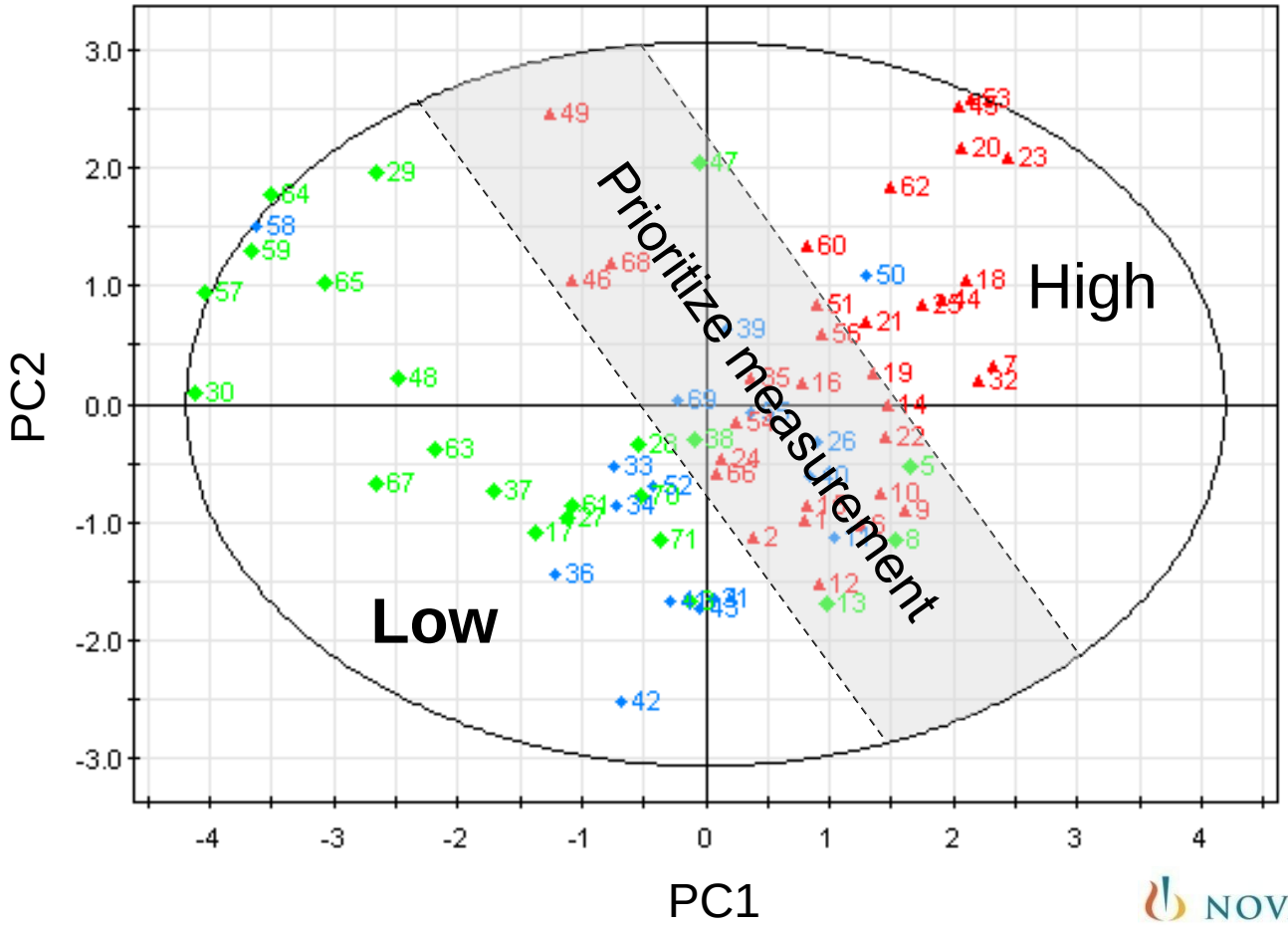
- Prioritize wet testing
- Uncover local opportunities

NOVARTIS

Targeting the right compounds for wet testing

MDCK-MDR1 efflux within a chemical series

ER >10 **ER 5-10** **ER <5**



Uncover local opportunities

Reducing Vss w/o compromising solubility

- General guideline: reduce logP and basicity
- Problem
  - Lowering logP may lead to poor potency
  - Reducing basicity leads to low soluble compounds
- Are there other opportunities within the local chemistry space ?

73 compounds, 16 dimensions

Reducing Vss w/o compromising solubility

Getting started

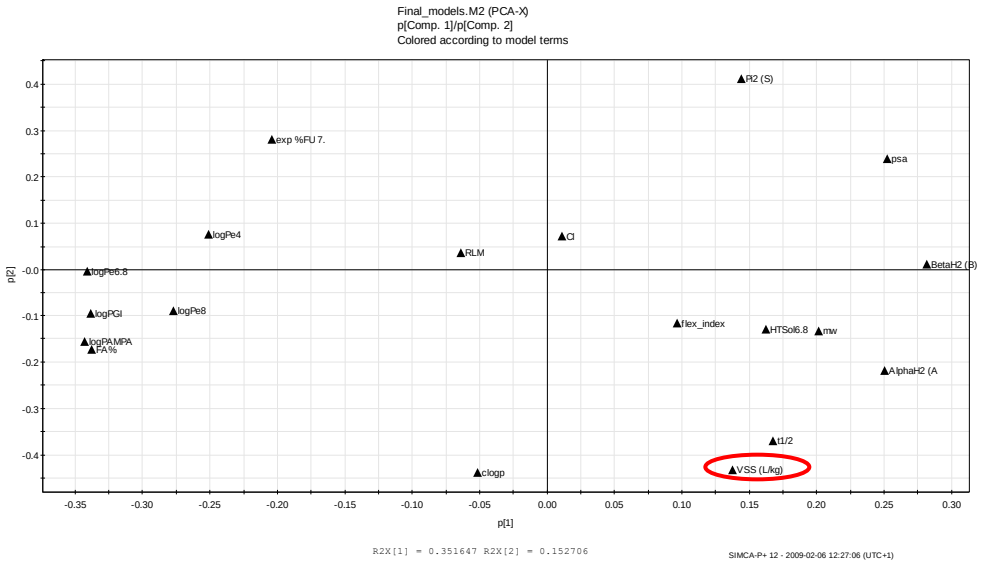
Multidimensional optimization

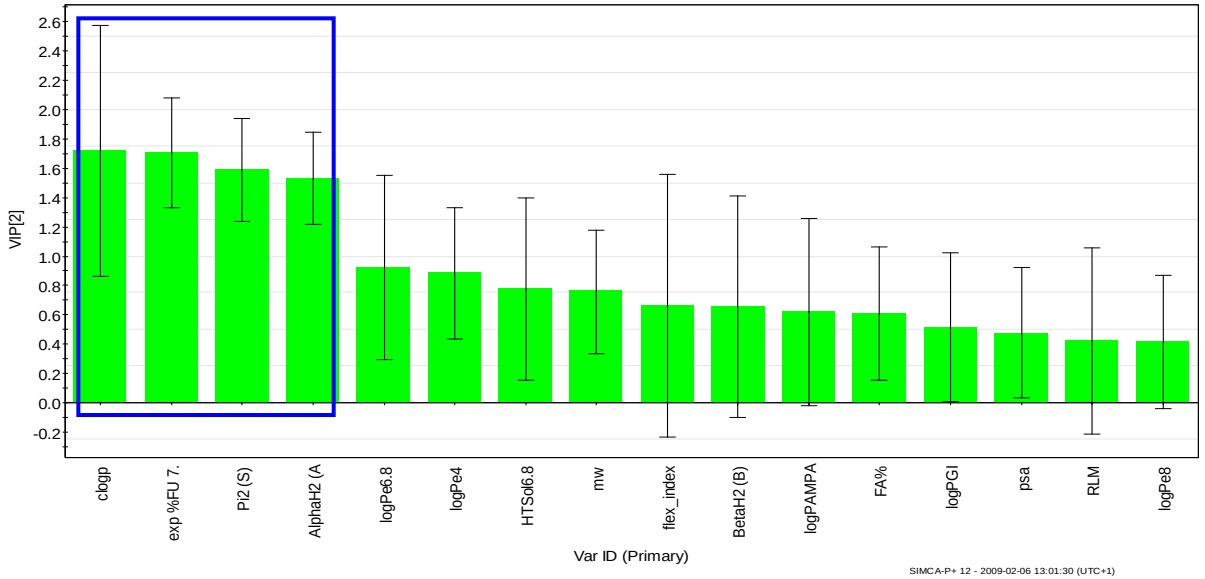
Local models to build hypotheses

Extracting information
In the absence of correlation

Potential and limitations

Conclusions



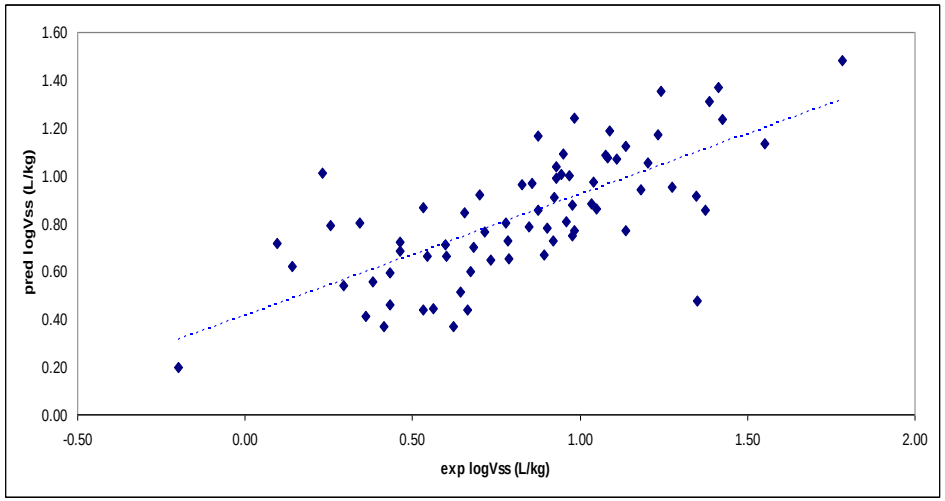


N=73

- 2PCs model
- R^2= 0.53
- Q^2 = 0.42

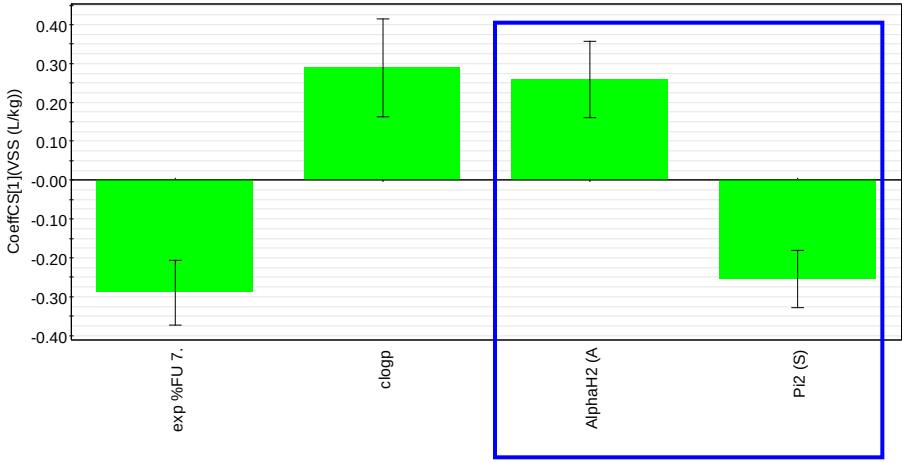
NOVARTIS

Local model for Vss: going beyond generic rules



- Y = log(Vss)
- Xs: %FU 7.4, clogP, HBD, polarizability
- 1PC model
- R^2= 0.51
- Q^2 (LOO) = 0.49
- 85% within 2-fold

scaled



Vss

log(Vss) = 1.13 - 0.24Fu7.4 + 0.11clogP + 0.67alpha - 0.24Pi2

NOVARTIS

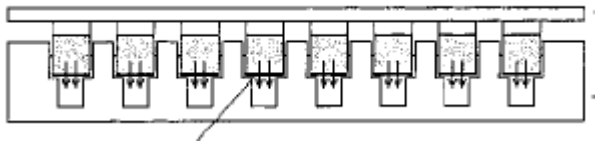
Extracting information in absence of correlation

- The value of orthogonal assays
- In-vivo/in-vitro Clearance

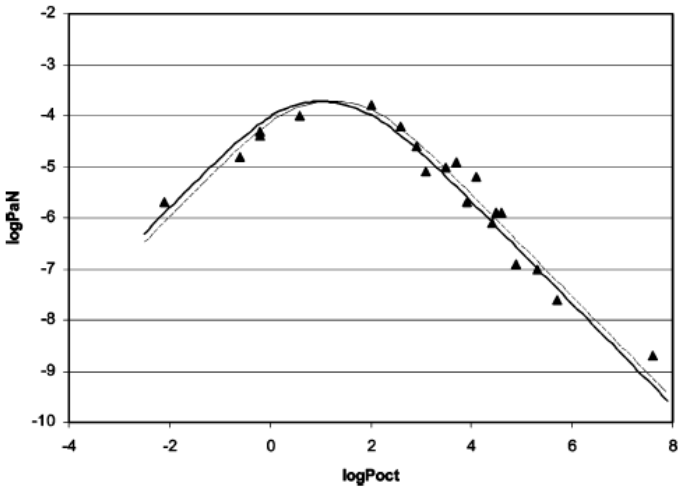
NOVARTIS

Low [C] logP > high [C] logP: what does it mean?

Example 1

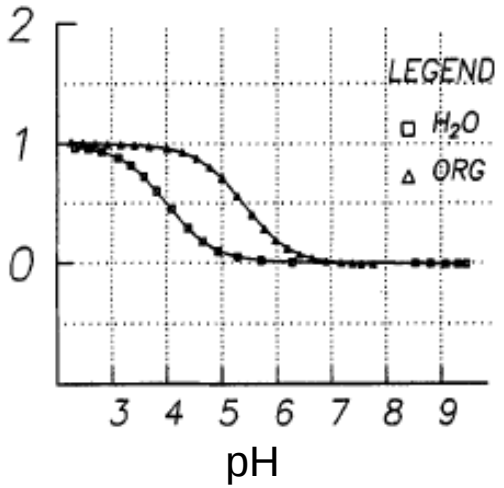


Octanol layer



from Faller-B et al , J. Med. Chem, 2005, 48(7), 2571

Compound 218



from Avdeef-A , J. Pharm. Sci, 1993, 82(2), 183

Loading [C]: 5 uM

Compound 218 : logP = 5.8

Loading [C]: 500 uM

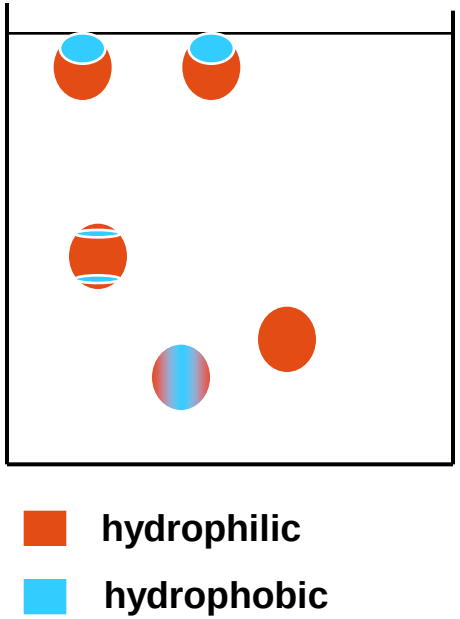
Compound 218 : logP = 4.2

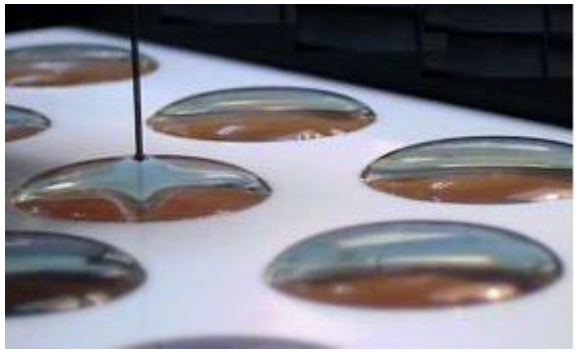
Are both assays measuring the same thing?

NOVARTIS

Surface pressure and amphiphilicity

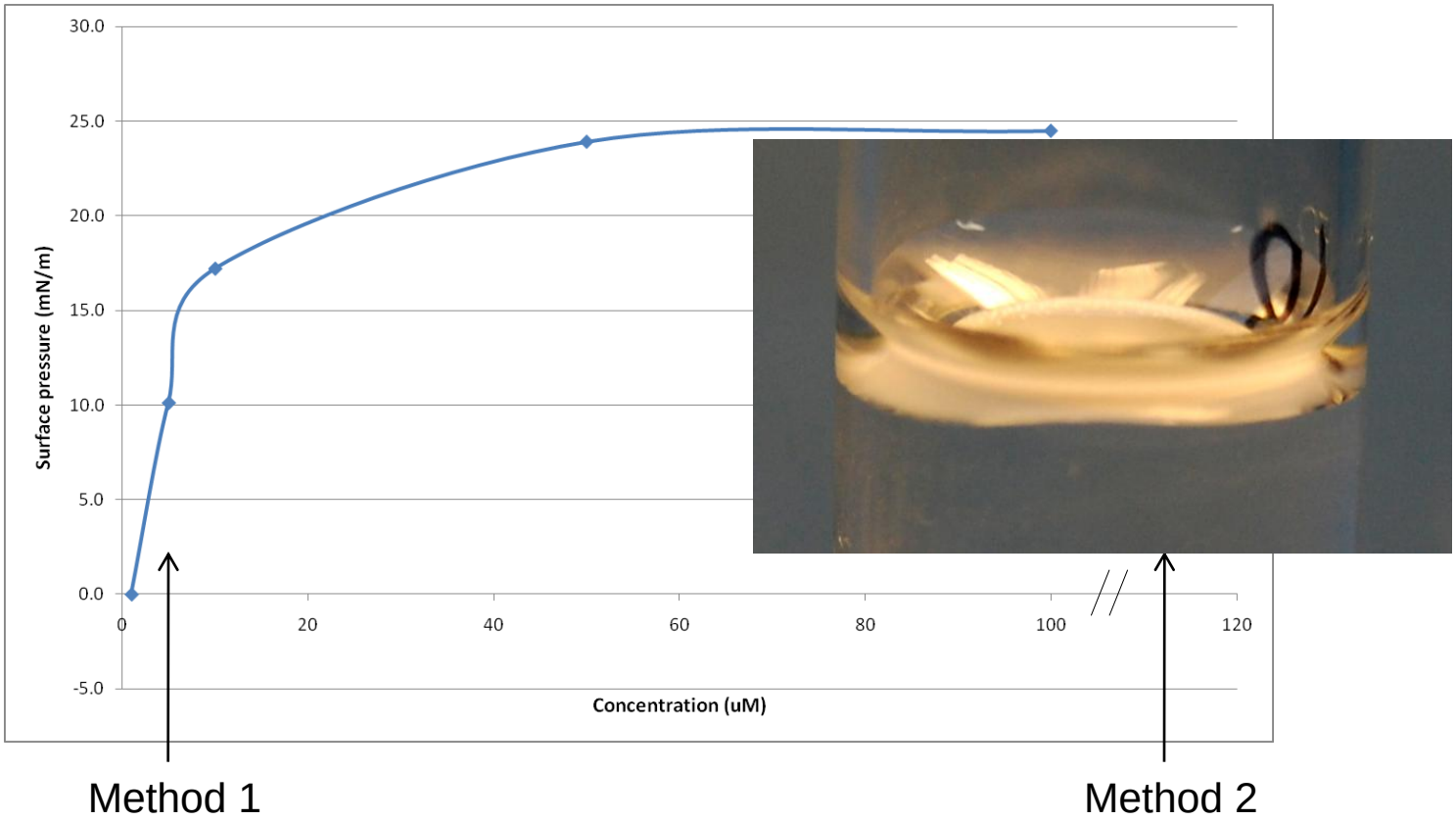
- Amphiphilic molecules accumulate at the surface (air-water interface)
- Causes a drop in surface tension (this is what is measured)
- At higher concentrations, micelles are being formed in the solution





NOVARTIS

logP below and above the CMC



Apparent logP drops as compound does not homogeneously distribute in the water phase

NOVARTIS

Power and limitations of in-silico molecular descriptors

NOVARTIS

When in-silico does better than in-vitro

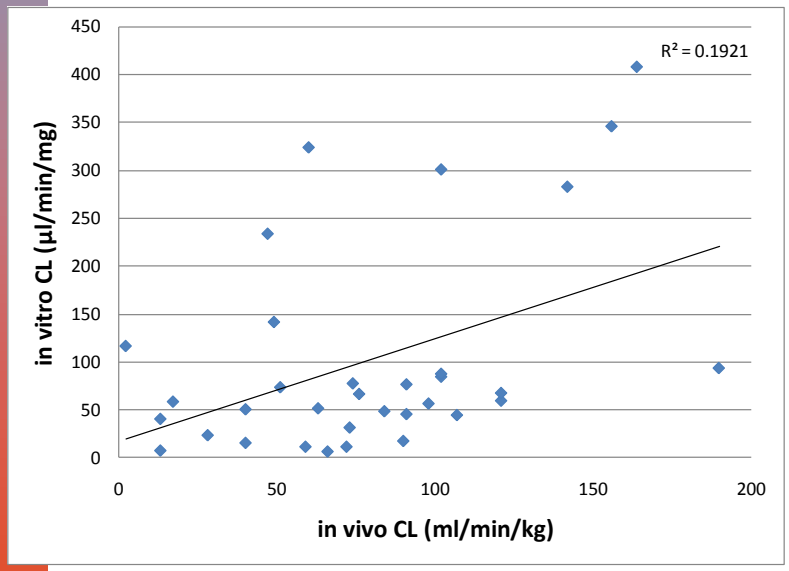
Example of clearance ivivc

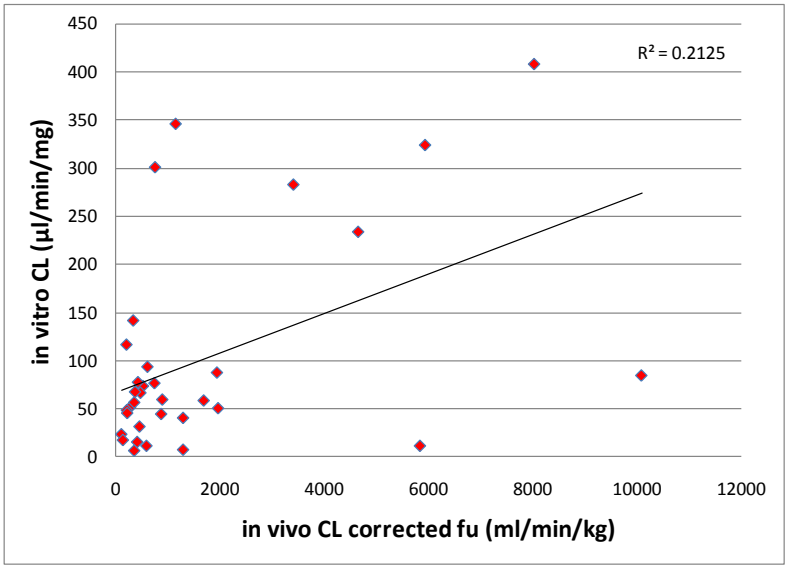
NOVARTIS

When in-silico does better than in-vitro

Example of clearance ivivc

- Poor ivivc (RLM, in-vivo CL)
- Correction by fraction unbound (f_u) did not improve correlation





PLS model and hypothesis building

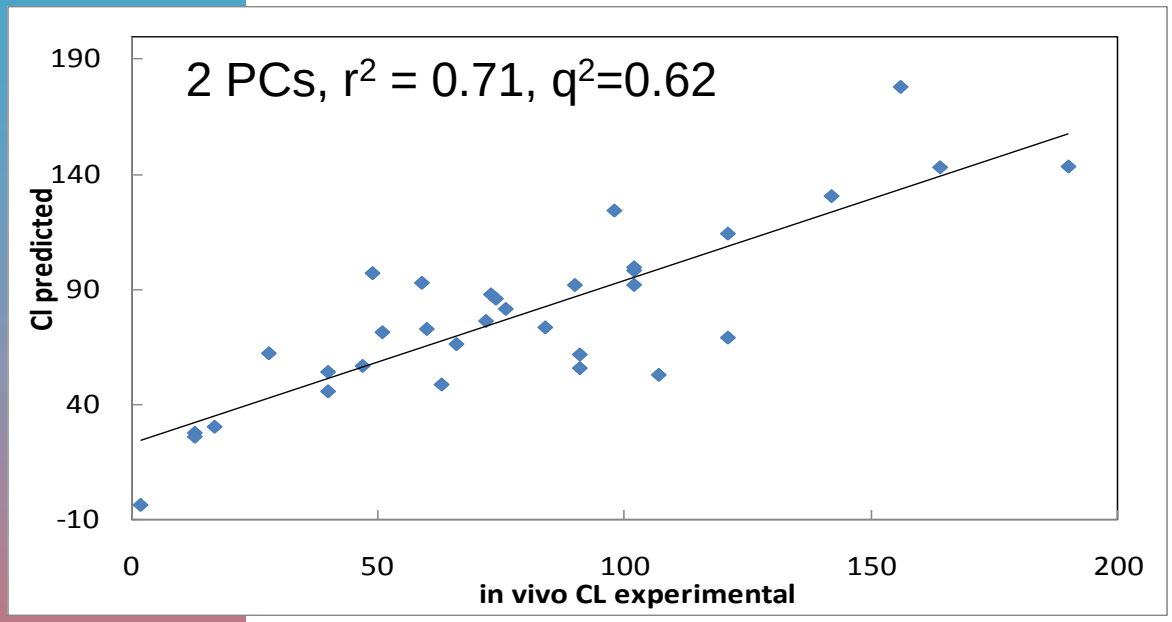
The addition of 3 in silico descriptors led to a fairly predictive PLS model

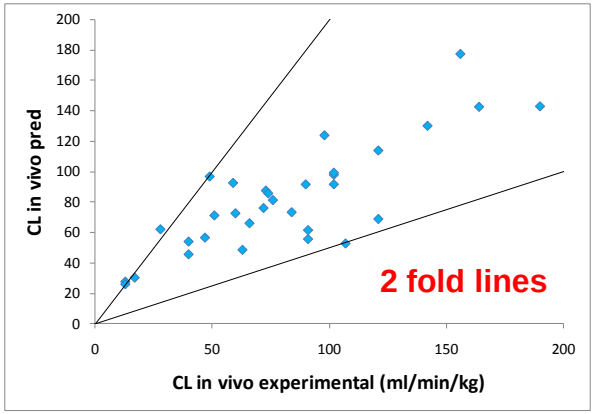
CL in vivo decreases

- The most important descriptors related to in vivo CL are in vitro CL (CL_{int}), hydrophobic descriptors (D4, D5) and neutral fraction at pH 7.4 (AUS7.4).
- plasma fraction unbound (f_u) had a low coefficient.

NOVARTIS

Hybrid model to reconcile in vitro and in-vivo data





GMFE = 1.36

$$CL_{invivo} = 274 + 0.19CL_{int} - 2.1D4 - 2.4D5 - 32AUS7.4$$

NOVARTIS

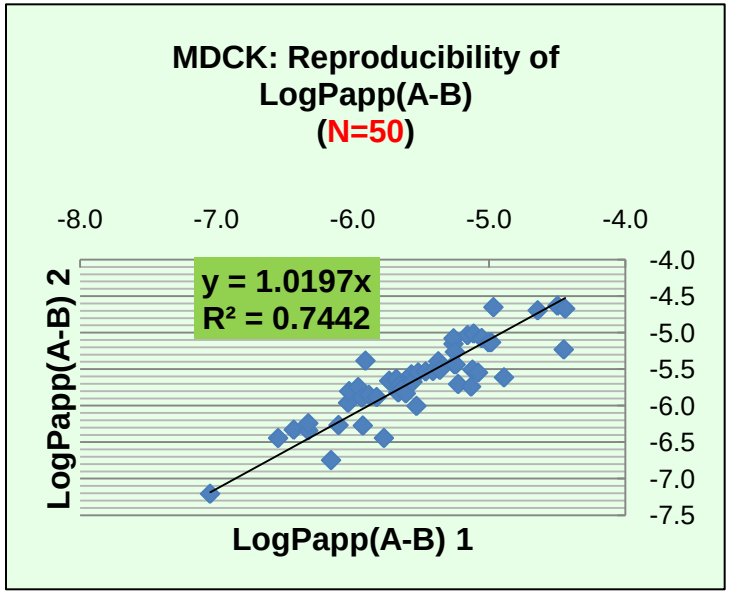
When in-vitro does better than in-silico

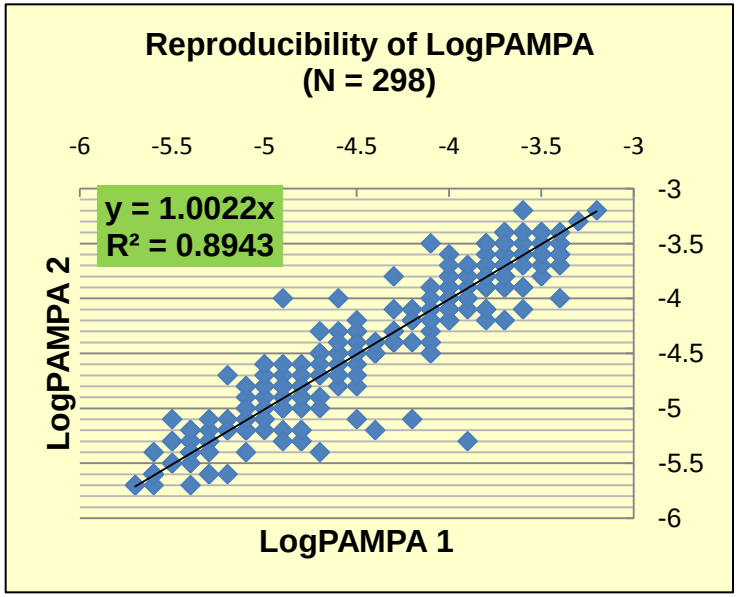
Example of passive permeability

NOVARTIS

Step 1: setting the expectations right

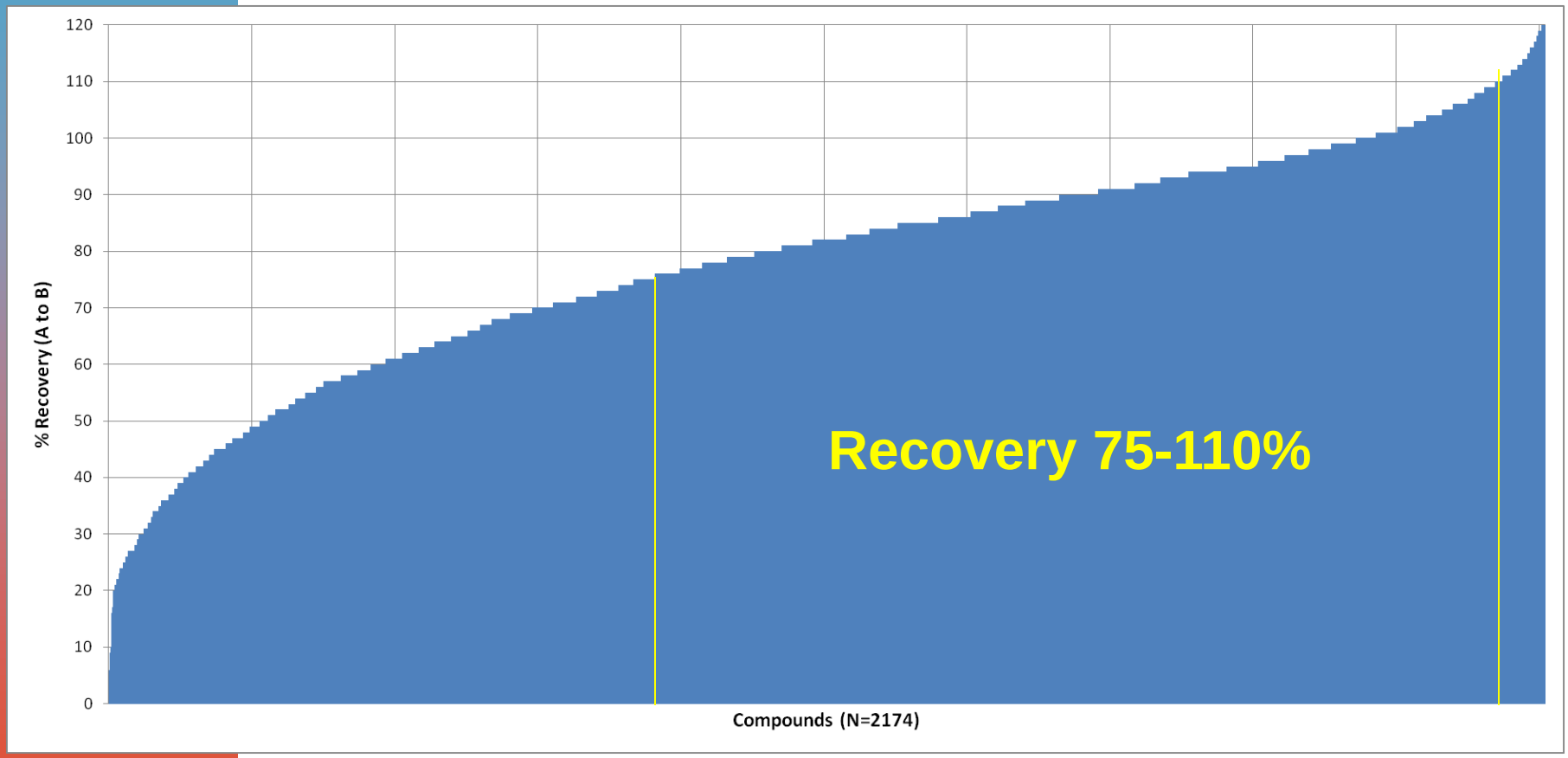
Correlation cannot be better than assay robustness





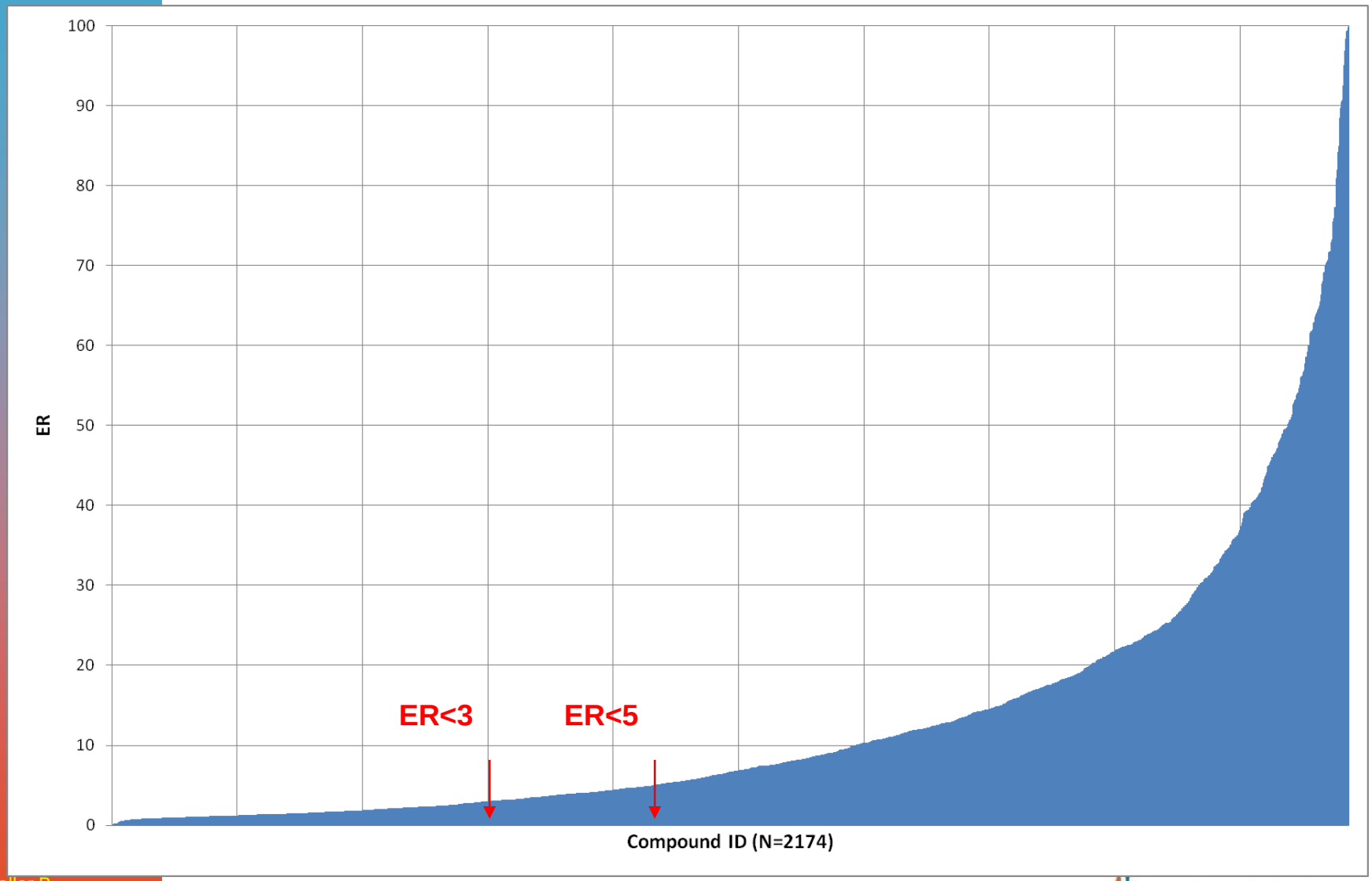
NOVARTIS

Filtering by recovery



NOVARTIS

Filtering by ER

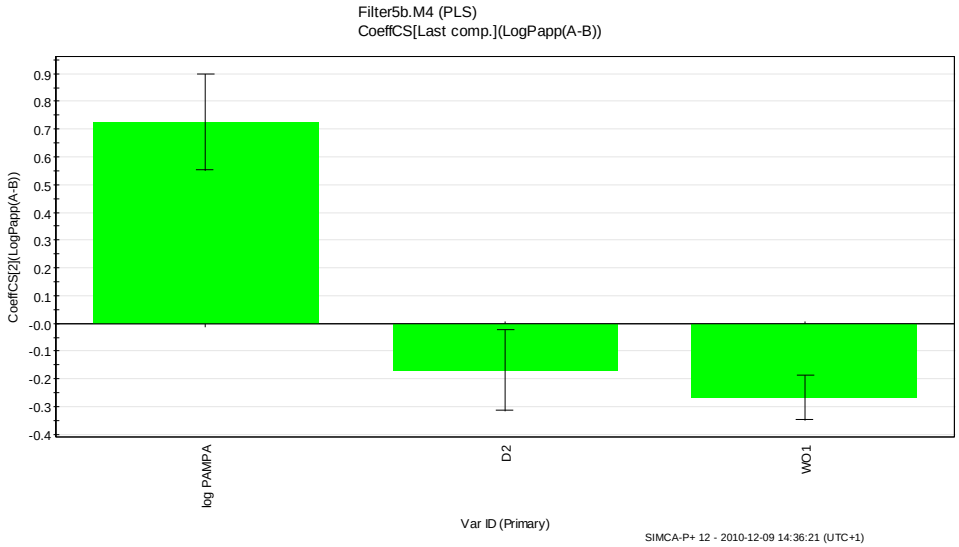


NOVARTIS

Impact of data filtering on correlation

Criteria	N	R2 PAMPA-MDCK	slope	R2 hyPLS	Q2 hyPLS
ER <5	535	0.38	0.55	0.51	0.5
ER<5, Rec >50%	436	0.43	0.65	0.60	0.57
ER<5 Rec.75-110%	207	0.51	0.73	0.70	0.66
ER<3 Rec.75-110%	116	0.68	0.95	0.72	0.67

hyPLS: hybrid PLS model based on VS+ descriptors **and** experimental PAMPA



NOVARTIS

Prediction without experimental data

- Exp. PAMPA came as a major descriptor in the hyPLS model
- Can we predict MDCK permeability based on calculated properties alone ?

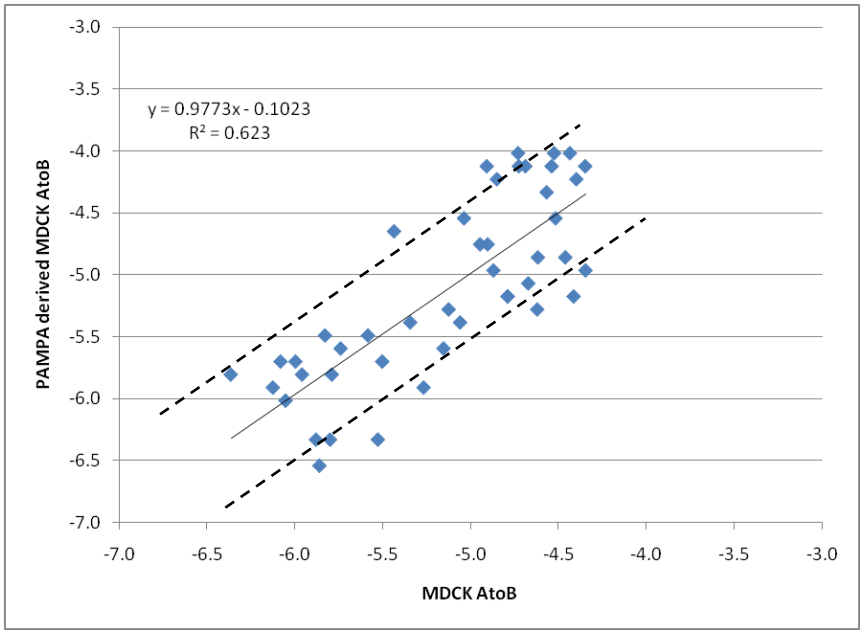
Criteria	N	R2 PAMPA-MDCK	R2 PLS	Q2 PLS
ER<3 Rec.75-110%	116	0.68	0.40	0.30

- Calculated molecular properties fail to replace exp. PAMPA in predicting MDCK passive permeability

NOVARTIS

External test set: 45 generic drugs

- Selection criteria
  - Rec. 75-110%
  - ER <3
- R2 = 0.62, slope = 0.97
- 70% predicted within 0.5 log unit



NOVARTIS

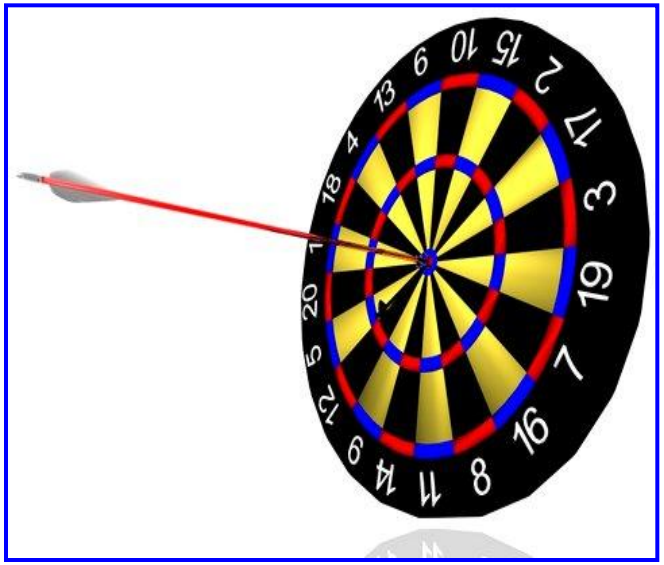
Conclusions -1

- Define clearly what you need to achieve
  - What is critically important
  - Where can I compromise
- Assay packages: science based as opposed to technology driven
- Mechanistic understanding needed for correct data interpretation
  - No variance, no information
  - Synthesize model compounds outside the potency race

Conclusions -2

- Different readout for same property: information rich
- What it takes to extract knowledge from data
  - Quality data
  - Reformatting / normalization
  - Build model
  - Use descriptors that are interpretable in MedChem terms
- Local models help to find opportunities within a chemical series and prioritize wet testing
- Models are best used to formulate and test hypothesis
- Go beyond plotting X vs. Y

Outlook



Fishing expedition

Hypothesis testing

NOVARTIS