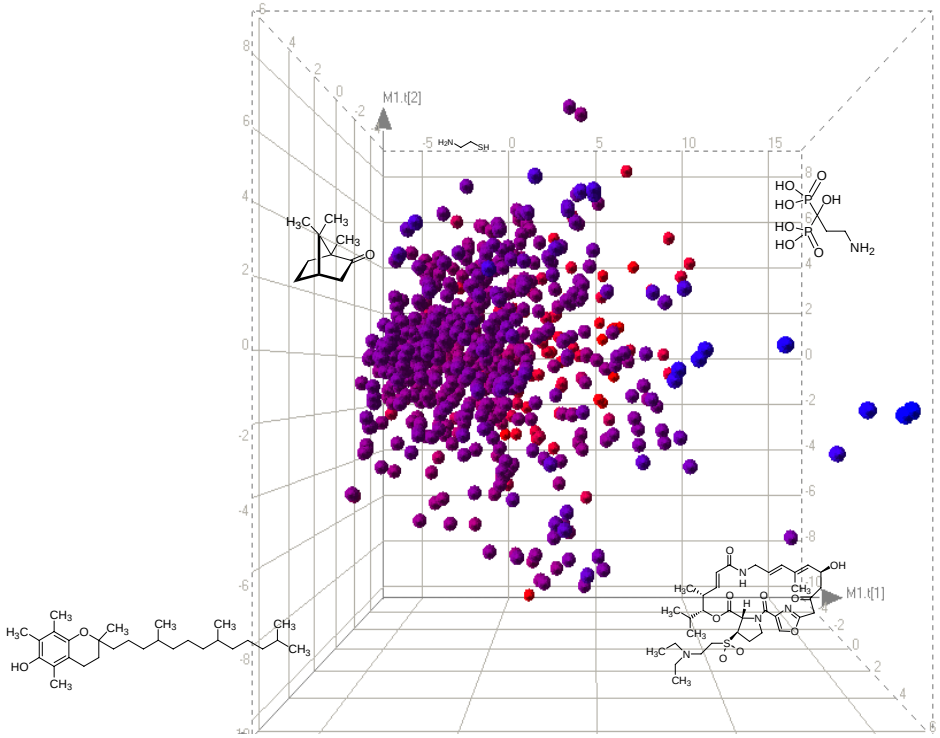
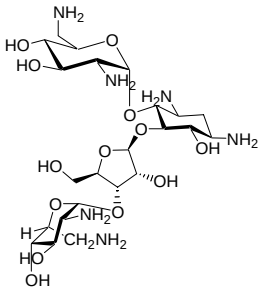
UPPSALA
UNIVERSITET

Presentation of a Structurally Diverse and Commercially Available Drug Data Set for Correlation and Benchmarking Studies

Anders Karlén
Uppsala University





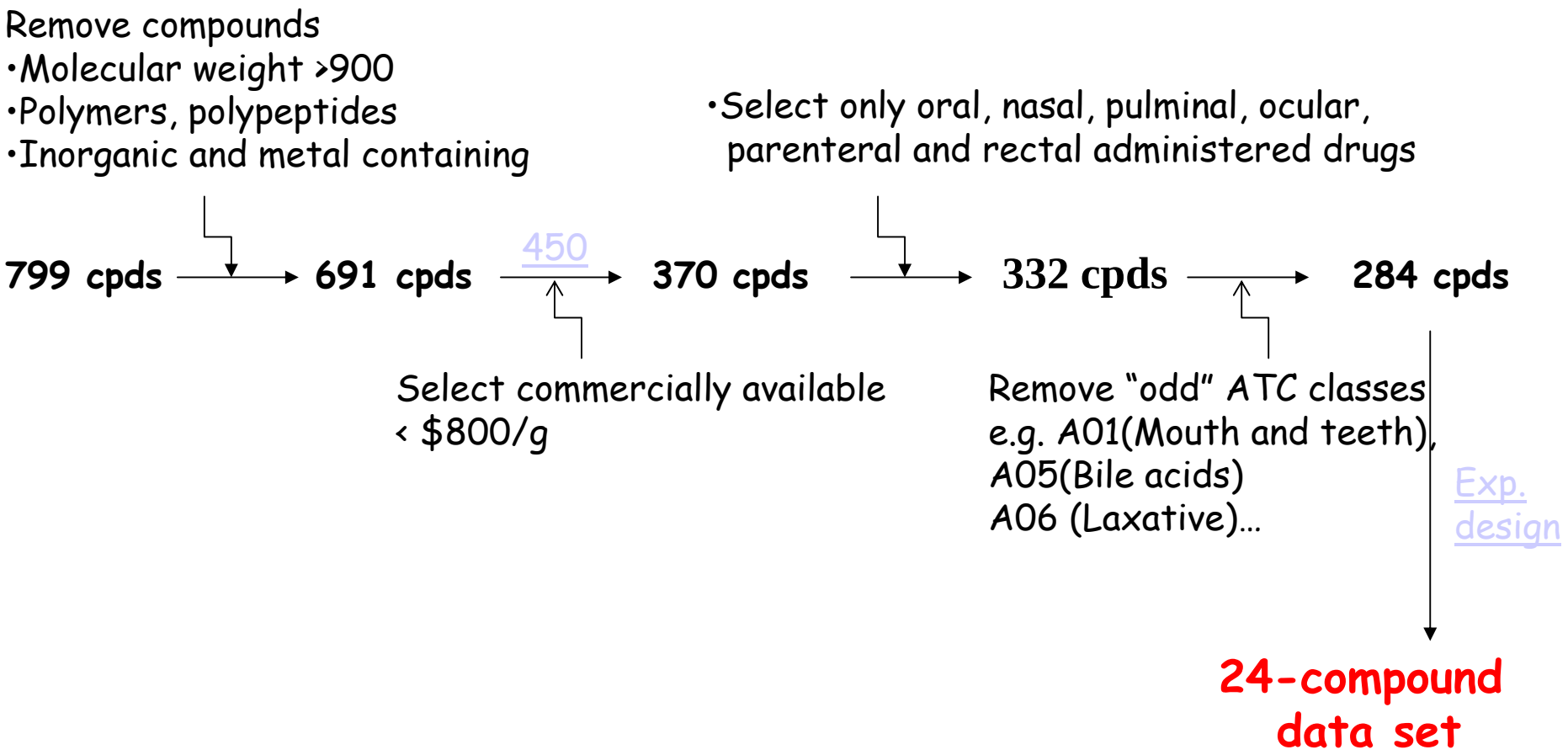
Aim of study

- Derive a "benchmark data set"
  - Drug-like
  - Physicochemically diverse
  - Commercially available and inexpensive
  - Amenable to analytical measurements
- Start the generation of benchmark data
  - Derive good-quality data from the same lab

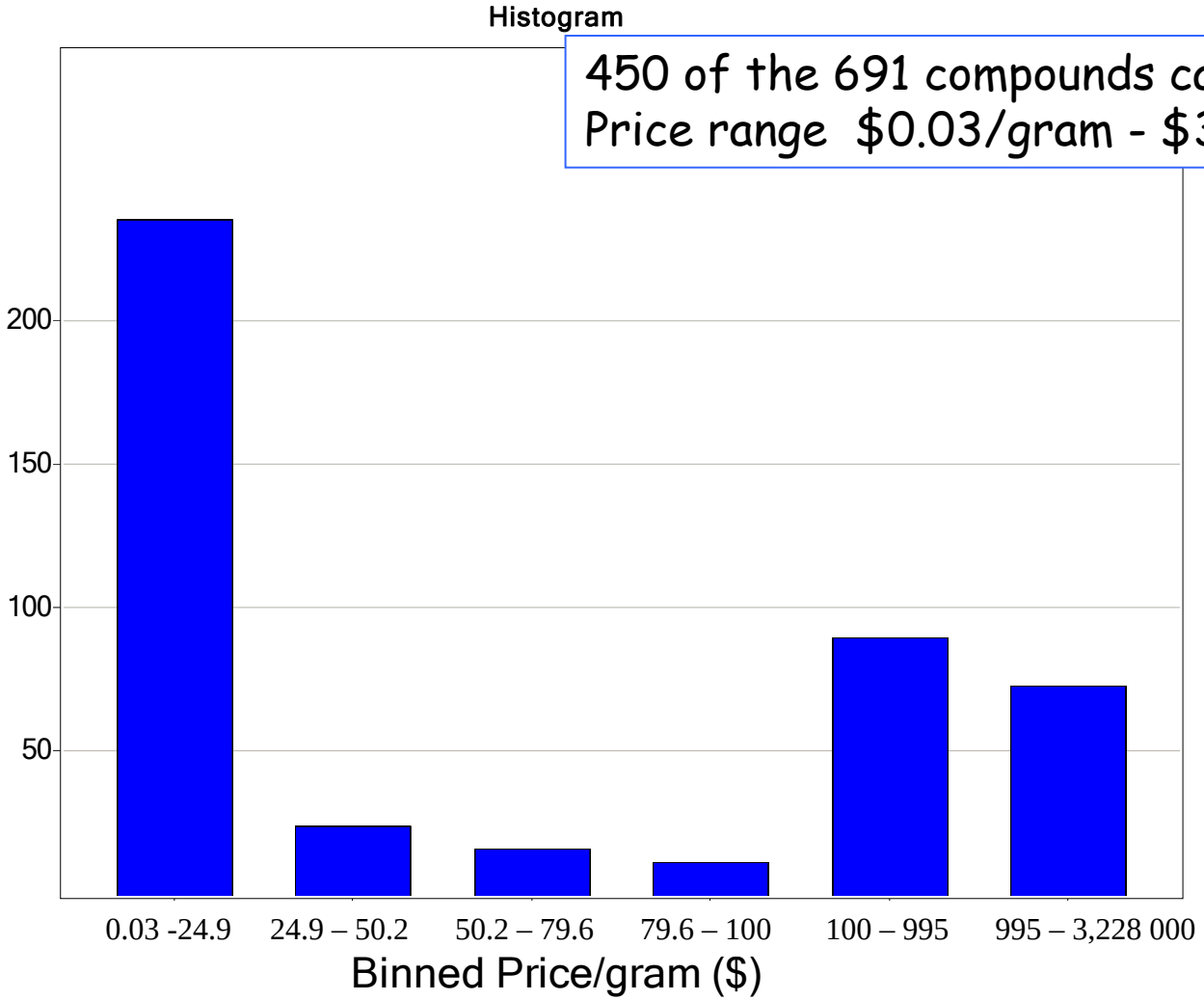
Possible use of the data set

- General description of drugs
- Developing ADME/TOX filters (permeability, solubility, plasma protein binding etc.)
- To validate novel experimental techniques

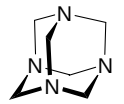
Generation of a "benchmark" data set based on the list of drugs in Sweden (FASS 2001)



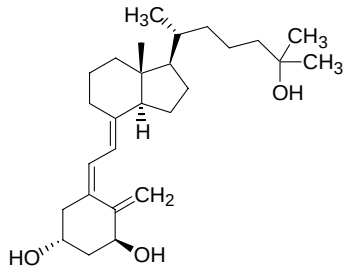
Cost and availability of the 691-compound data set



450 of the 691 compounds can be bought
Price range $0.03/gram - $3,228 000/gram (2001)



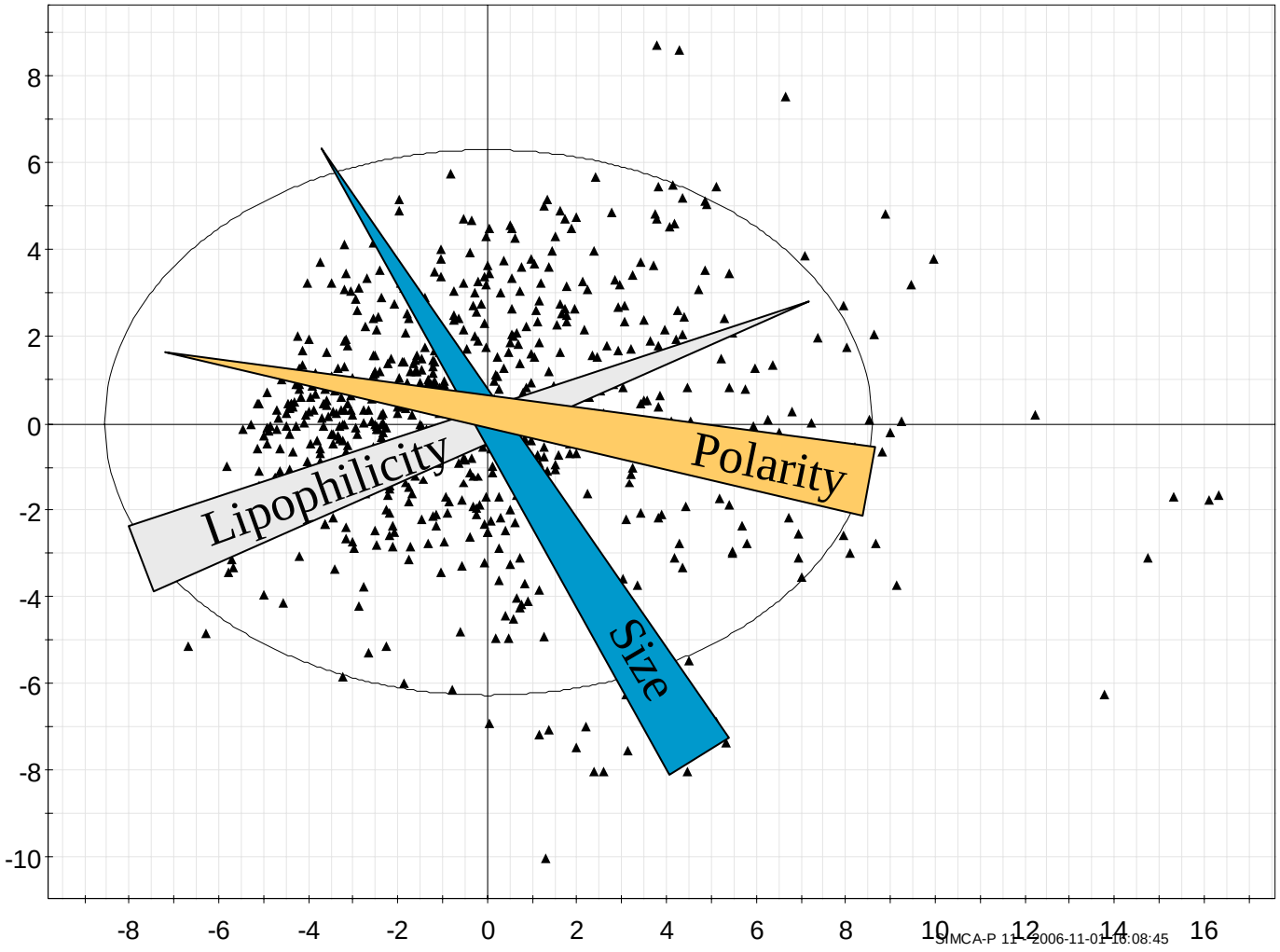
Methenamine



Calcitrol

Back

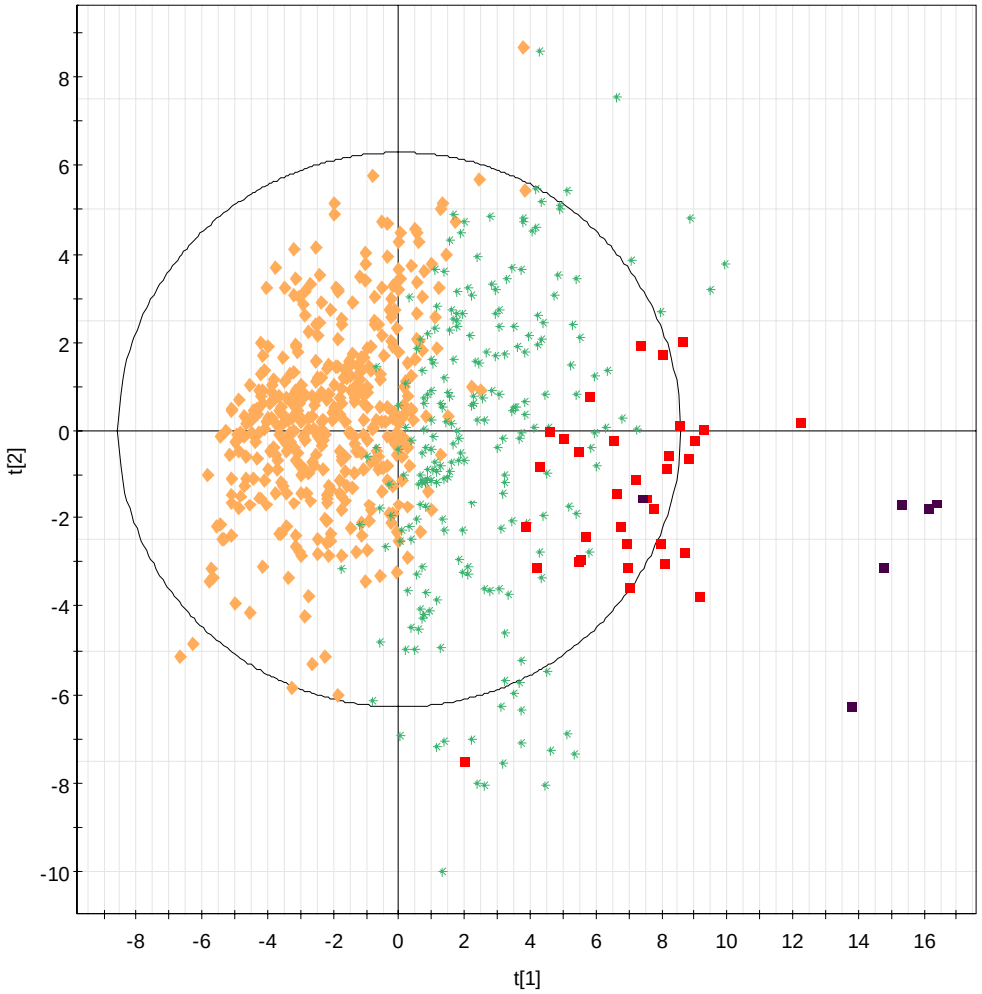
Principal component analysis

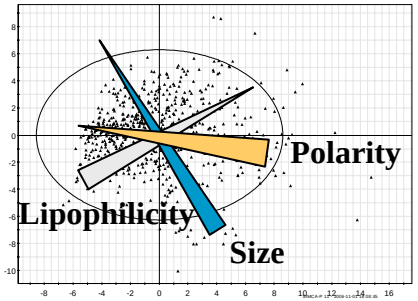


- General descriptors
- General hydrogen bonding descriptors
- Hydrogen bond donor descriptors
- Hydrogen bond acceptor descriptors

Σ28 molecular descriptors

Principal component analysis



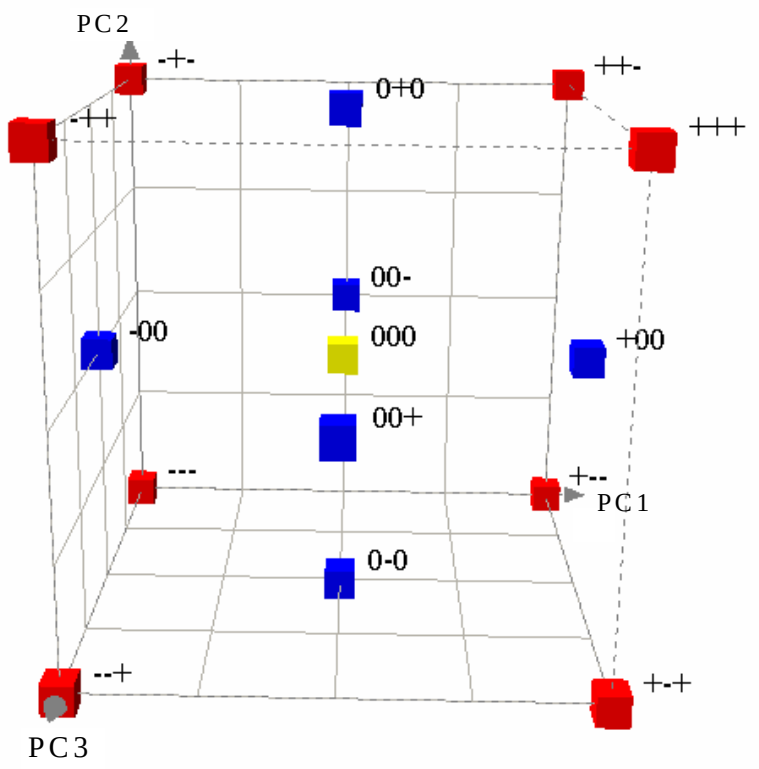


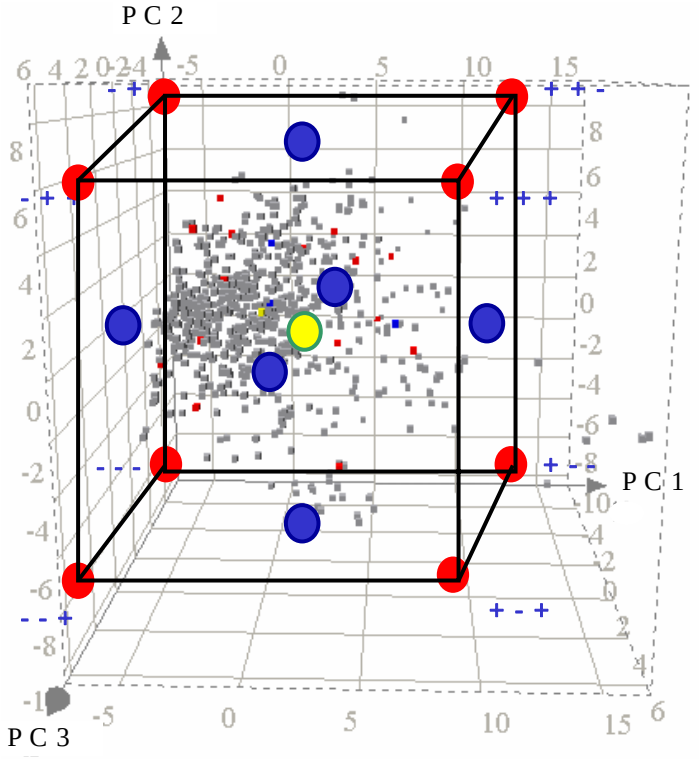
Series (Variable PSASAVOL)

- 0 - 100
- 100 - 200
- 200 - 300
- 300 - 400

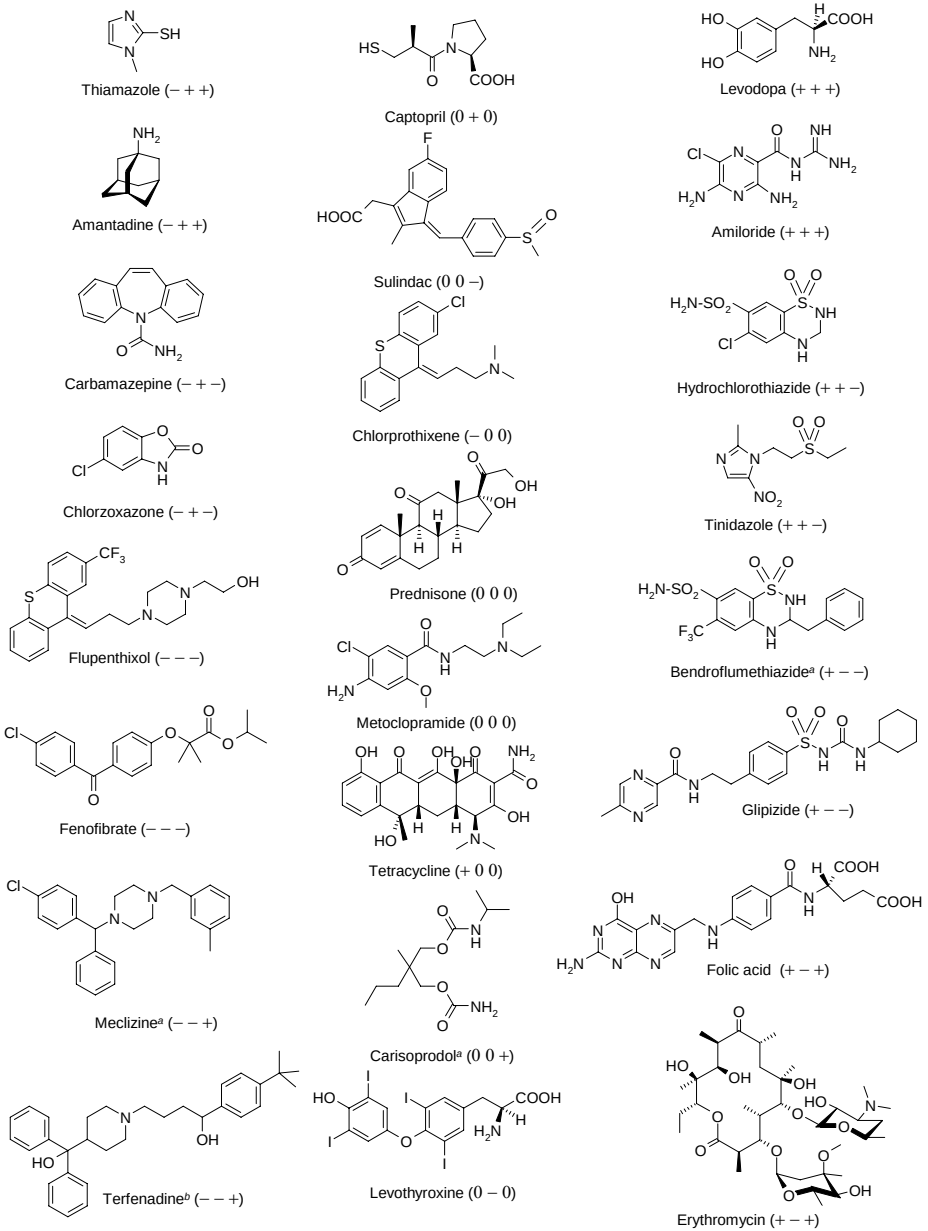
SIMCA-P+ 11 - 2006-11-10 10:34:12

The factorial design
"A face-centered central composite design"

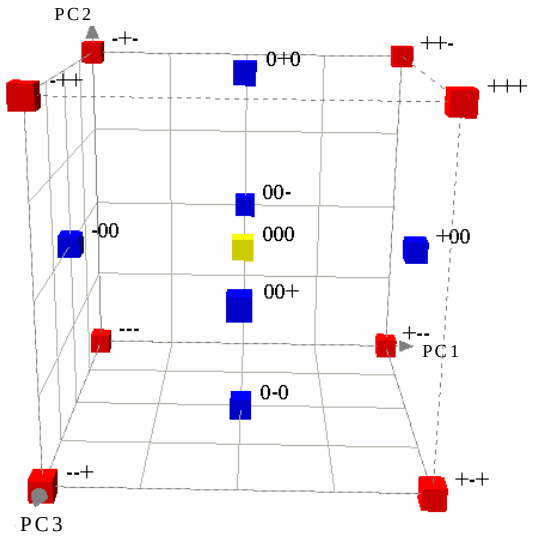




24-compound data set



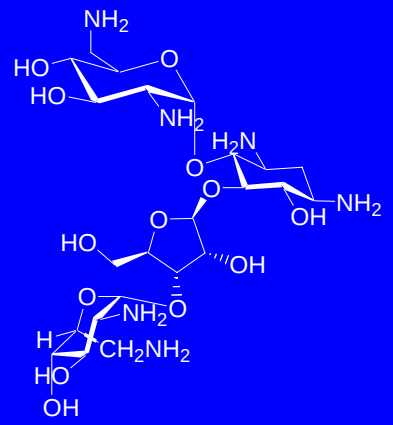
20 proteolytes
4 nonproteolytes



The cost of buying the entire data set (at least 1 gram of each compound) is less than $1,500

Comparison of the data sets with respect to some common molecular descriptors

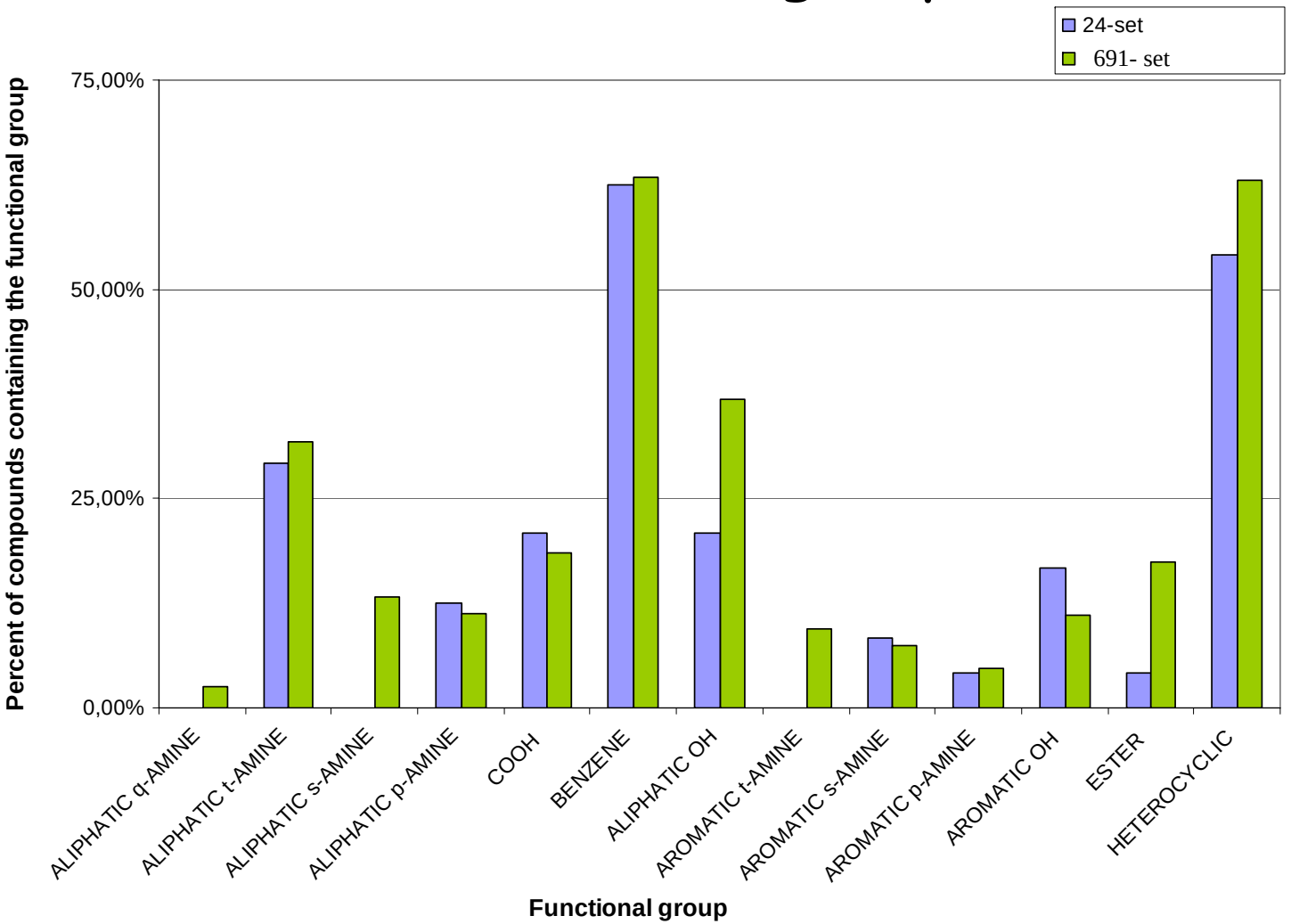
	691-compound data set			24-compound data set		
	Min	Max	Mean	Min	Max	Mean
MW	60	854	347	114	777	349
PSA	0	373	93	8	246	99
$\log P_{\mathrm{Mor}}$	–6.4	7.6	1.9	–2.0	5.3	1.9
$\log D_{\mathrm{ACD\_6.5}}$	–10.6	12.3	0.74	–5.0	4.8	0.94
HBD	0	19	2.4	0	8	2.7
HBA	0	19	4.9	1	14	4.7



Neomycin
HBD = 19

Candesartan cilexetil
$\log P_{\mathrm{Mor}}$ = 7.6

Comparison of the data sets with respect to functional groups



Comparison of the data sets with respect to ATC classes

Distribution in ATC					
		Number of substances		Percent of dataset	
ATC	Description	24-set	691-set	24-set	691-set
A	GI	1	69	4,2%	9,99%
B	Blood	0	21	0,0%	3,04%
C	Cardio	2	89	8,3%	12,88%
D	Topical	0	36	0,0%	5,21%
G	Gen.hormones	1	38	4,2%	5,50%
H	Hormones	3	14	12,5%	2,03%
J	Infection	5	89	20,8%	12,88%
L	Tum.,immuno	1	53	4,2%	7,67%
M	Muscle,mov.	3	37	12,5%	5,35%
N	Nervous	6	134	25,0%	19,39%
P	Antiparasite	0	13	0,0%	1,88%
R	Respiration	1	52	4,2%	7,53%
S	Eye,ear	1	24	4,2%	3,47%
V	Various	0	22	0,0%	3,18%

The Anatomical Therapeutic Chemical (ATC) classification system is the most commonly used classification system for drug substances

Start the generation of benchmark data. Derive good-quality data from the same lab

1. Measurment of pKa by pH-metric or pH-UV technique (n=20)

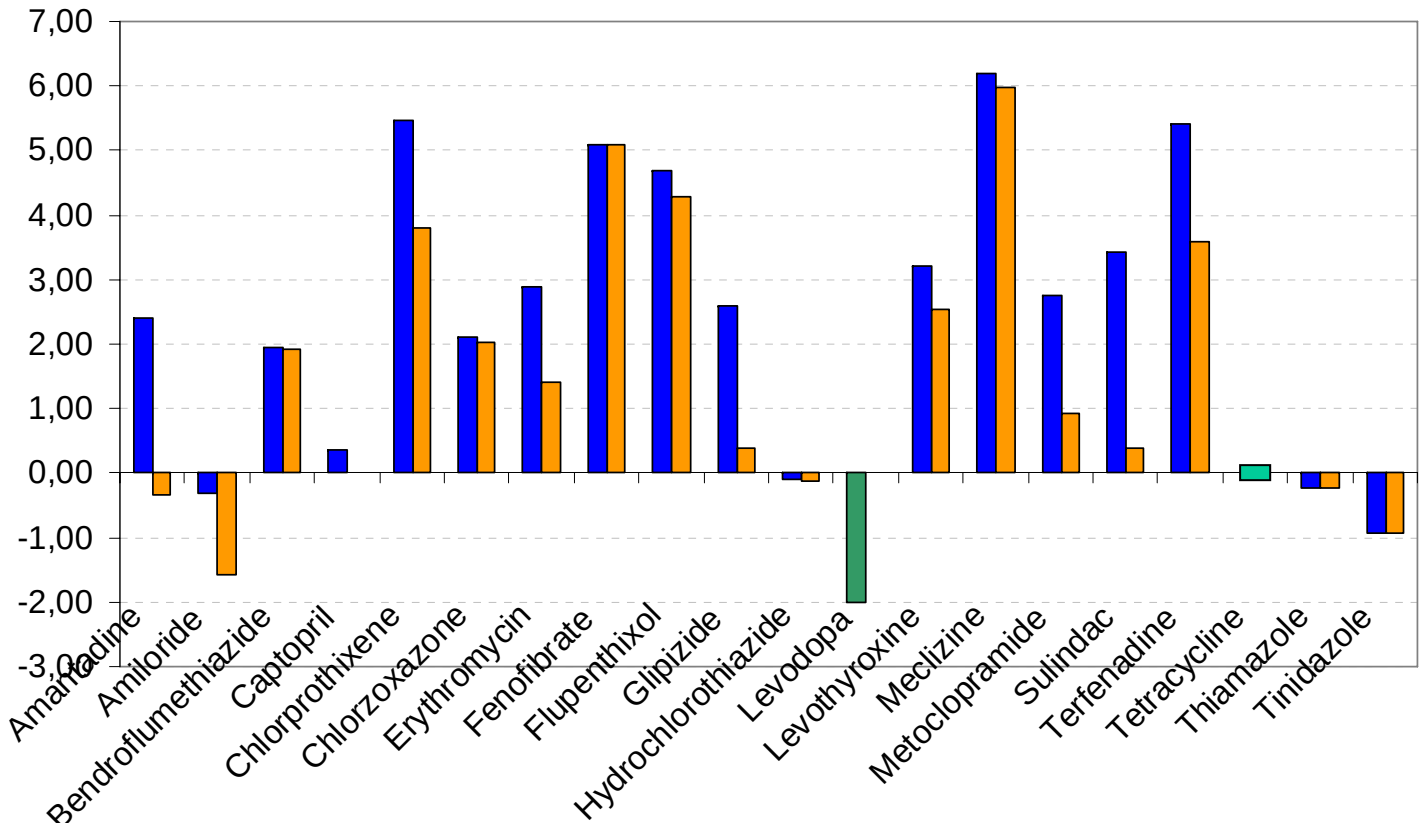
2. Measurment of lipophilicity
   (a) pH-metric logP (n=18)
   (b) capacity factors by RP-HPLC (n=21)

3. Measurment of intrinsic and kinetic solubility
   pH-metric solubility (CheqSol technique) or shake-plate solubility (n=17)

4. Measurment of permeability across Caco-2 Cells. A to B direction (n=22)

2. Lipophilicity

pH-metric measurment of logP and logD

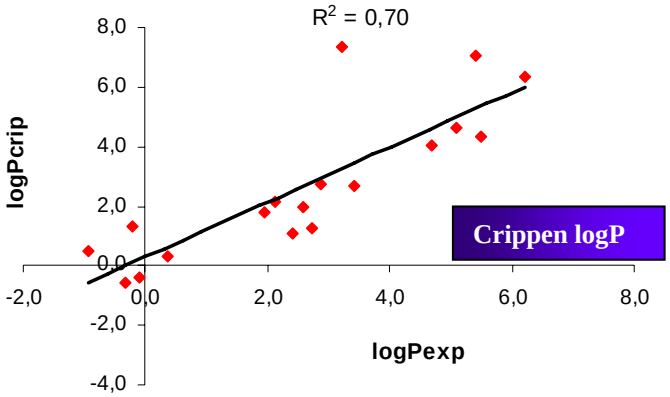


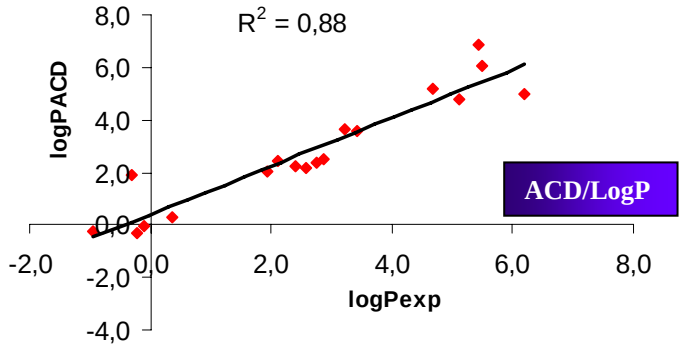
logP missing for;

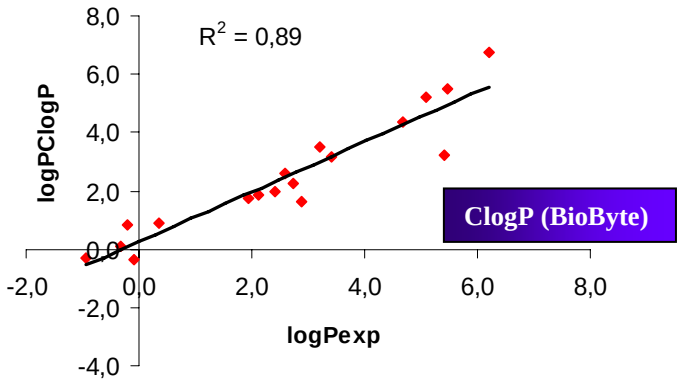
- Folic acid
- Carbamazepin
- Prednisone
- Carisoprodol

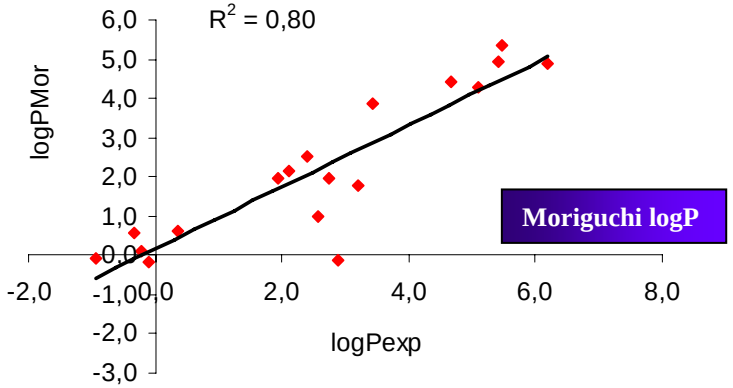
2. Lipophilicity

Experimental logP *vs* calculated logP



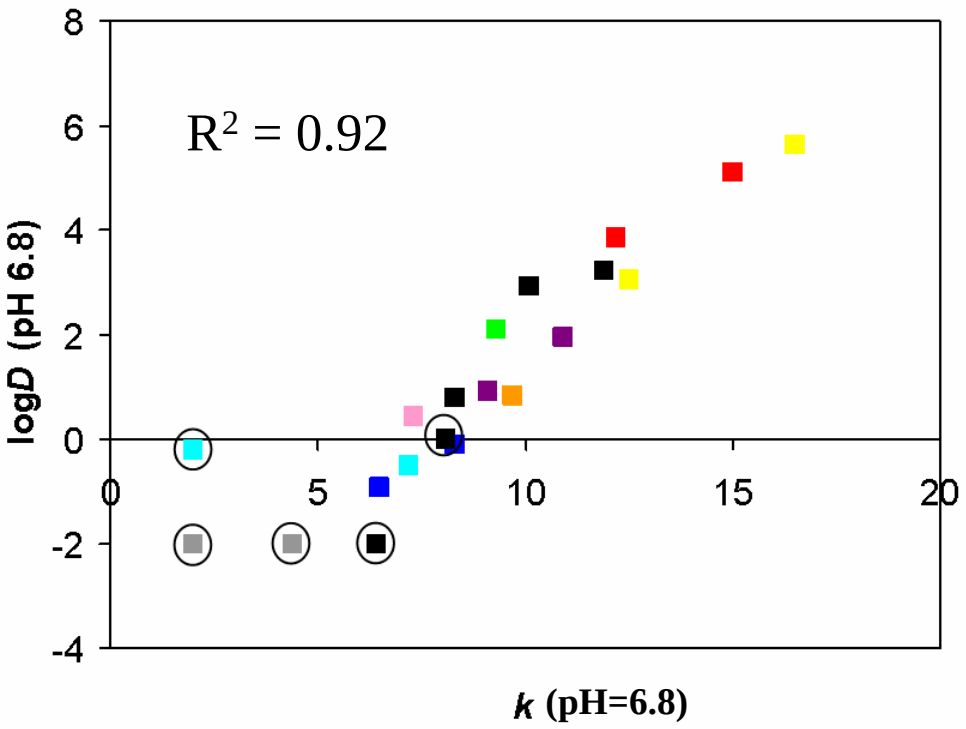






2. Lipophilicity

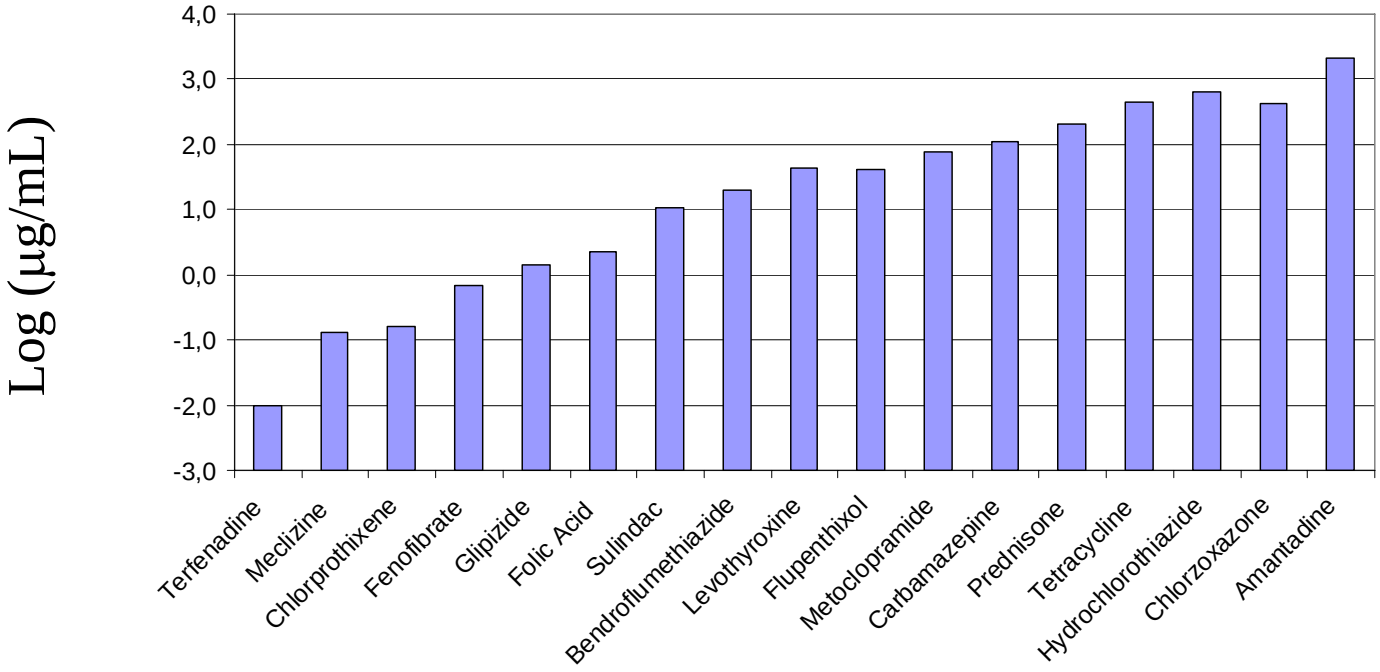
Correlation between the measured HPLC capacity factor (*k*) and pH-metric log*D* (pH 6.8)



- Compounds from the 8 corner points have different colors
- The 2 compounds at each corner point have the same color
- The axis points are colored black
- Center point pink

3. Solubility

Measurment of intrinsic solubility using CheqSol (24-compound data set)



Solubility ranges from 0.009 µg/ml to 2119 µg/ml

3. Solubility

	Name	Equilibrium solubility CheqSol	Shake-Flask	Literature	Kinetic Solubility Chaser	Kinetic Solubility non-chaser
		All results in µg/mL				
●	1 Phthalic Acid	5330	5950		8462	
	2 Quinine	363	201	491		391
●	3 Trazodone	134.6	138.0		435	
	4 Nitrofurantoin	112.5	109.5	78.9	319	
	5 Nortriptyline	27.0	49.3	20.0		27.3
	6 Verapamil	48.5	48.5	9.7		47.8
●	7 Niflumic Acid	9.53	29.5		59	
	8 Imipramine	17.2	21.7	18.1		17.3
●	9 Flumequine	34.2	20.7		121	
	10 Furosemide	19.7	20.4	5.9	96	
	11 Maprotiline	5.80	8.05	3.49	77	
	12 Piroxicam	5.92	5.95	3.16	233	
	13 Warfarin	5.30	5.25	5.60	120	
	14 Chlorpromazine	2.70	2.41	1.71		2.70
	15 Lidocaine	3500		3810	4600	
	16 Famotidine	740		1100	5900	
	17 Hydrochlorothiazide	630		700	2400	
●	18 Chlorpheniramine	608.3		615.2		668
●	19 Sulfamerazine	200.3		203.0	701	
	20 Ketoprofen	130.6		178.0	336	
	21 Propranolol	81.0		70.0	340	
	22 Ibuprofen	50.0		49.0	180	
	23 Pindolol	41.7		32.7	1424	
	24 Miconazole	1.00		0.67		
	25 Diclofenac	0.90		0.80	45	
●	26 Amodiaquin	0.41			8.8	
●	27 Pamoic acid	0.0003			0.019	

● Compound not present in the 691 data set

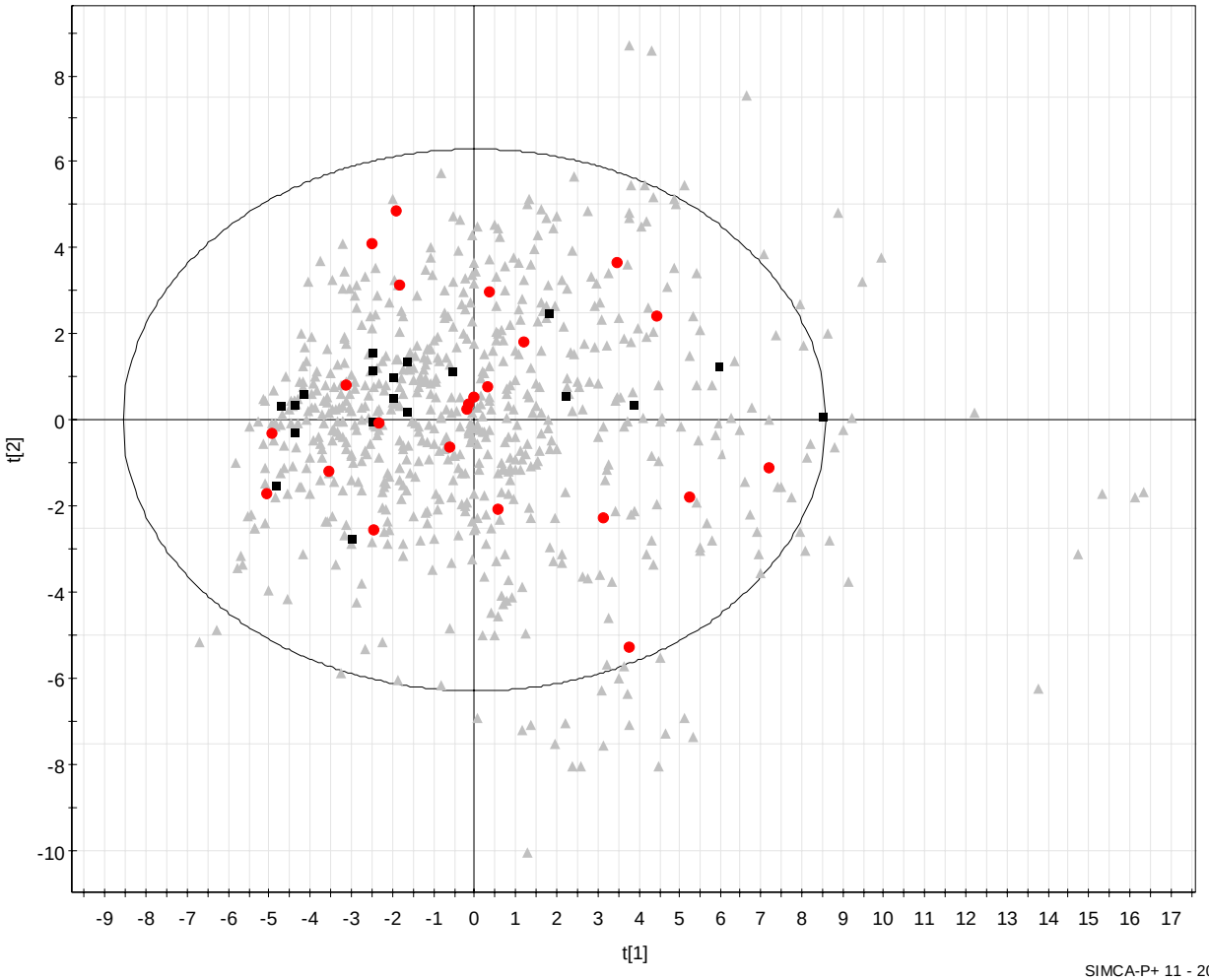
19 of the compounds studied also present in the 691-compound data set

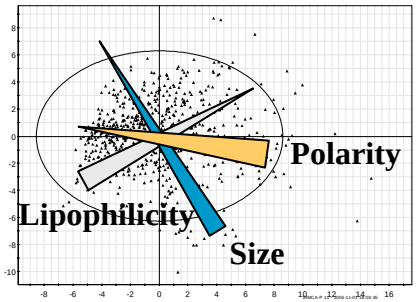
CheqSol solubility ranges from 0.9 µg/mL to 3500 µg/mL in these 19 compounds

In the 24-compound data set the solubility ranges from 0.009 µg/ml to 2119 µg/ml

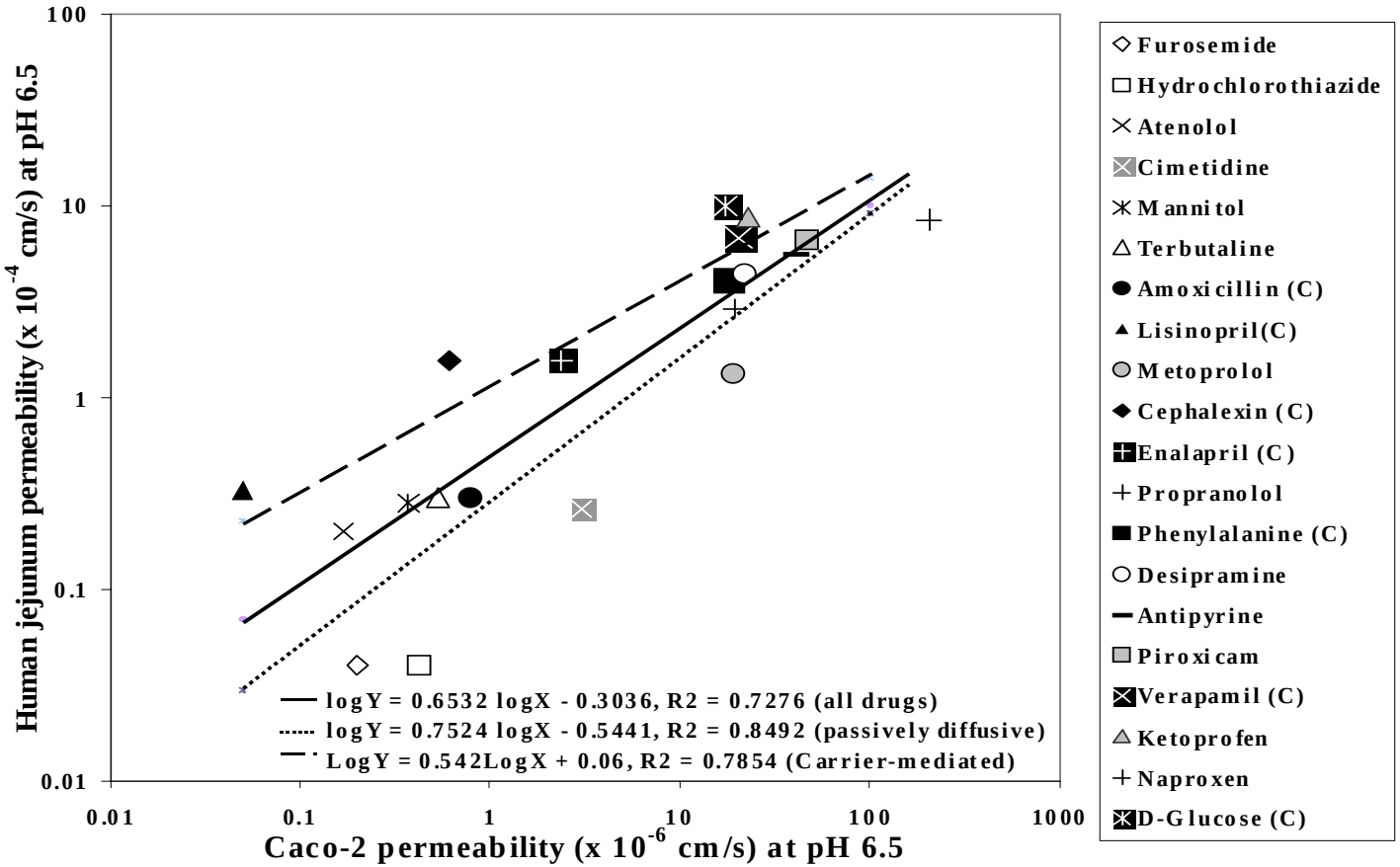
http://www.cheqsol.com/download%20files/download01.pdf

24-compound data set is structurally diverse





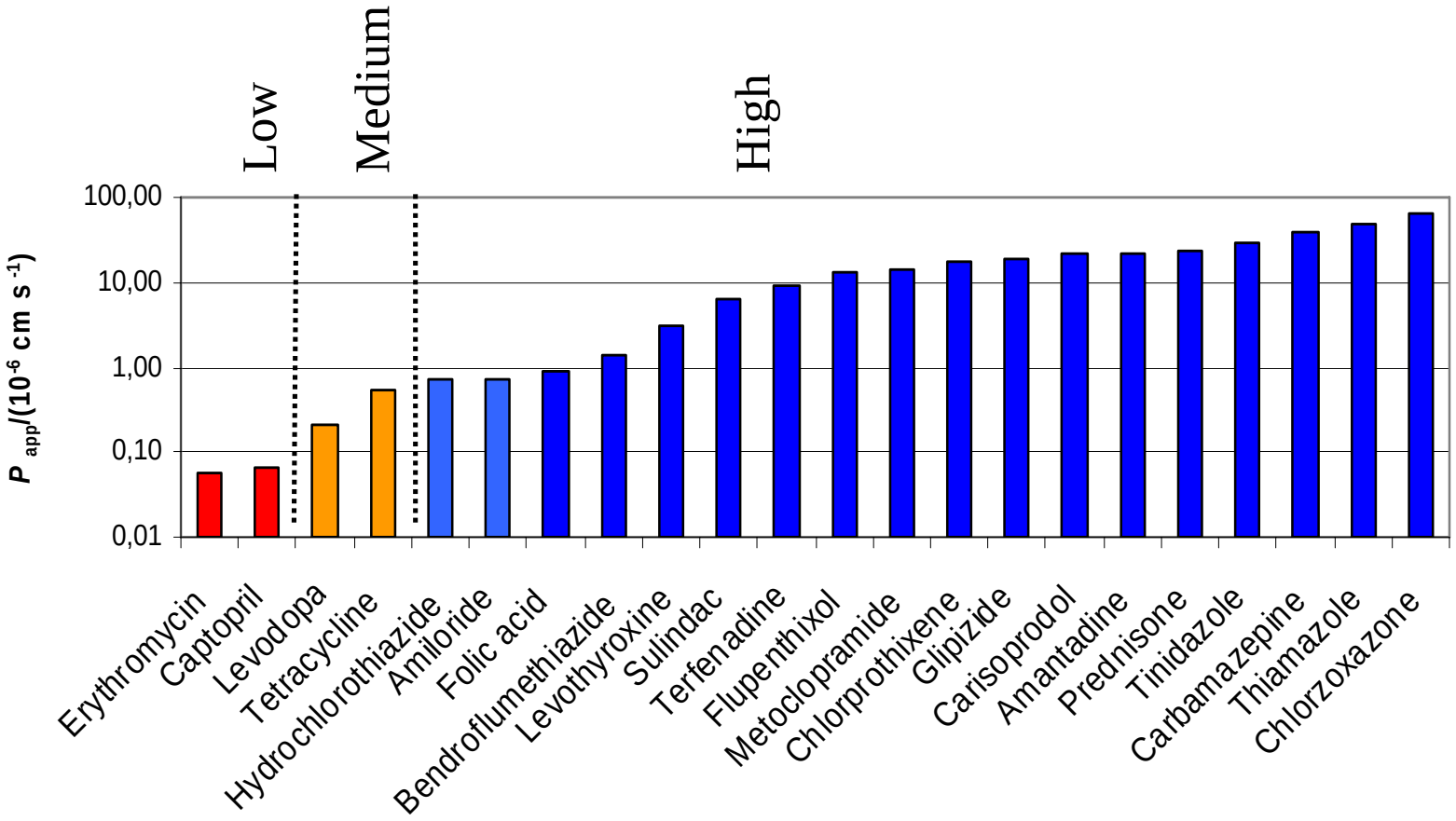
4. Permeability/absorption



Sun, D. et al. Comparison of Human and Caco 2 Gene Expression Profiles for 12,000 Genes and the Permeabilities of 26 Drugs in the Human Intestine and Caco 2 Cells. *Pharm Res* 2002, *19*, 1398-1413

4. Permeability/absorption

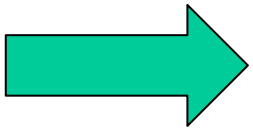
In vitro P_{app} values in human Caco-2 cells



Suggestions on the "Uppsala diverse data set" usage

- The 24 compounds can be used
  - as a test set for testing already derived models of permeability, lipophilicity, solubility etc.
  - as a validation set for new experimental techniques
  - on its own for building and validating models by dividing it into a training set and a test set

We hope that other groups are willing to help us to supplement the herein-started characterization



"Bench mark data set"

J. Med. Chem.; **(ASAP); 2006**; *49*(23); 6660-6671

UPPSALA
UNIVERSITET

Acknowledgements

Faculty of Pharmacy
Uppsala University
Christian Sköld
Torbjörn Lundstedt
Anders Hallberg
Hans Lennernäs

Sirius Analytical Instruments Ltd
John Comer
Karl Box
Ruth Allen
Jon Mole

AstraZeneca R&D Mölndal
Susanne Winiwarter
Anna-Lena Ungell
Johan Wernevik
Fredrik Bergström
Leif Engström