Chemoinformatics in Drug Discovery - Quo Vadis?

Hugo Kubinyi

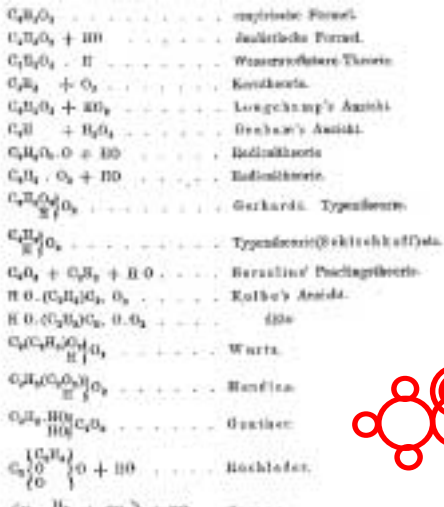
Germany

E-Mail kubinyi@t-online.de
HomePage www.kubinyi.de

2nd Strasbourg Summer School on Chemoinformatics, June 20-24, 2010

Hugo Kubinyi, www.kubinyi.de

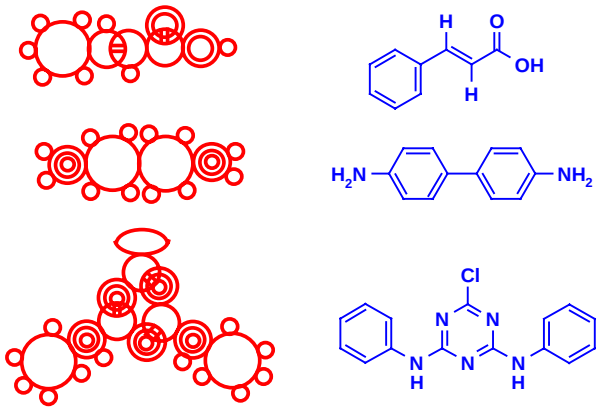
Historical Formulas of Acetic Acid (1860)



Joseph Loschmidt

"Chemische Studien. A. Constitutions-Formeln der organischen Chemie in geographischer Darstellung", Wien, 1861

Loschmidt Constitution Formulas (1861)



Kekulé Benzene Formula (1865)


3. Benzine.

QSAR

Corwin Hansch (* 1918)

(picture taken at the 5th EuroQSAR, 1984)

S. L. Carney (DDT 9, 158-160 (2004)):
Has there been a single development that, in your opinion, has moved the field of medicinal chemistry ahead more than any other?
Robin Ganellin (Professor of Medicinal Chemistry, University College, London, UK): I would go back to the 1960s to the work of Corwin Hansch on the importance of lipophilicity. ... that changed the way of thinking in medicinal chemistry. .... the application of physical organic chemical approaches to structure–activity analysis [has] been very important.

Is QSAR relevant to Drug Discovery?
A. M. Doweyko, Idrugs 11, 894-899 (2008)

QSAR: dead or alive?
A. M. Doweyko, J. Comput.-Aided Mol. Design 22, 81-89 (2008)

On outliers and activity cliffs - why QSAR often disappoints
G. M. Maggiora, J. Chem. Inf. Model. 46, 1535 (2006)

Beware of q^2!
A.Golbraikh and A. Tropsha, J. Mol. Graphics & Model. 20, 269-276(2002)

3D-QSAR illusions
A. M. Doweyko, J. Comput.-Aided Mol. Design 18, 587-596 (2004)

The trouble with QSAR (or how I learned to stop worrying and embrace fallacy)
S. R. Johnson, J. Chem. Inf. Model. 48, 25-26 (2008)

How not to develop a QSAR/QSPR relationship
J. C. Dearden et al., SAR and QSAR in Environ. Res. 20, 241-266 (2009)

How to recognize and workaround pitfalls in QSAR studies: a critical review
T. Scior et al., Curr. Med. Chem. 16, 4297-4313 (2009)

Hugo Kubinyi, www.kubinyi.de

QSAR: Problems in Statistical Analyses

- inappropriate biological data
- wrong scaling of biological data
- data from different labs
- different binding modes
- mixed data (e.g. oral absorption and bioavailability)
- different mechanism of action (e.g. toxicity data)
- too few data points
- too many single points
- lack of chemical variation
- clustered data
- small variance of y values
- systematic error/s in y
- too large errors in y values
- outliers / wrong values
- wrong model selection

QSAR: Problems in Statistical Analyses

inappropriate x variables
too many x variables (Topliss)
  a) in the model selection
  b) in the final model
x variable scaling in CoMFA fields
interrelated x variables
singular matrix
elimination of variables that are significant only with others
insignificant model (F test)
insignificant x variables (t test)
no qualitative (biophysical) model
no causal relationship (the storks)
extrapolation too far outside of observation space
no validation method applied
wrong validation method, ...

How the Trouble Started: Connectivity Indices $^{i}\chi$

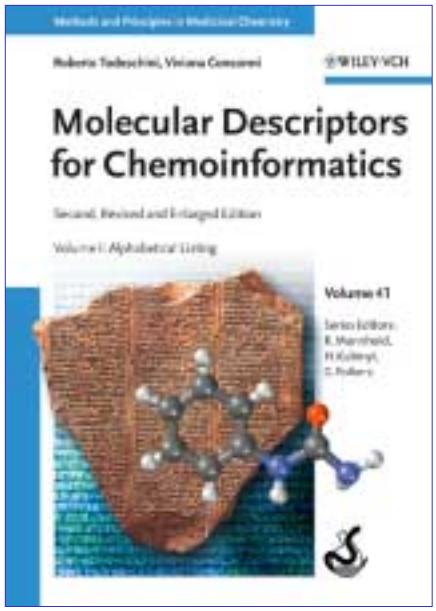
Connectivity indices
= electron-weighted
subgraph counts

$^{0}\chi = \Sigma\ (\#\ \sigma\text{-electrons of i})^{-0.5}$

$^{1}\chi = \Sigma\ (^{0}\chi(i).^{0}\chi(j))^{-0.5}$
(over all bonds ij)

... etc.

$^{0}\chi$ $^{1}\chi$ $^{2}\chi$ $^{3}\chi_{P,}$ $^{3}\chi_{C}$ ...



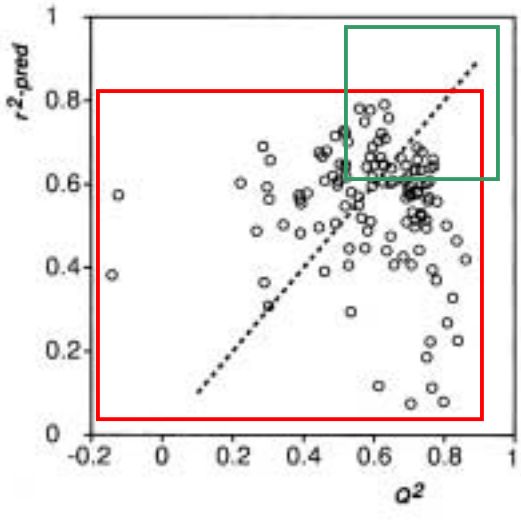
Program E-DRAGON
Roberto Todeschini

Web version of the DRAGON program at www.vcclab.org/lab/edragon/

E-DRAGON analyses up to 149 molecules and up to 150 atoms per molecule. Current version: Dragon 5.4 from March 28, 2006.

Calculates more than 1,600 molecular descriptors, organized in 20 blocks, from SMILES code, SDF, or MOL2 files.

External vs. Internal Predictivity

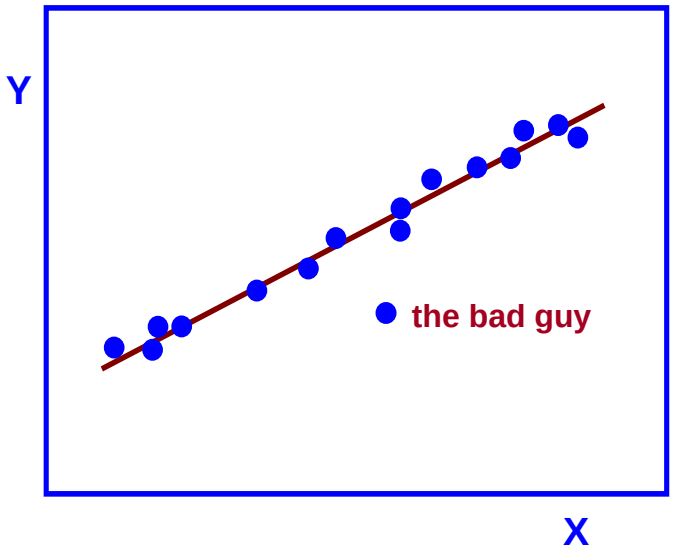


The „Kubinyi Paradox“

J. H. van Drie, Curr. Pharm. Des. 9, 1649-1664 (2003); J. H. van Drie, in: Computational Medicinal Chemistry for Drug Discovery, P. Bultinck et al., Eds., Marcel Dekker, 2004, pp. 437-460.

Data from H. Kubinyi et al., J. Med. Chem. 41, 2553-2564 (1998).

„Good“ and „Bad“ Guys in Regression Analysis



bad guy in the test set:
r^2, Q^2 good
r^2_{pred} poor

bad guy in the training set:
r^2, Q^2 poor
r^2_{pred} good

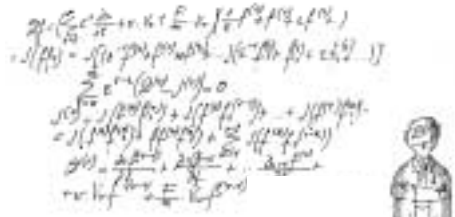
Hugo Kubinyi, www.kubinyi.de

Proper Validation of QSAR and 3D QSAR Models

Validation Method	Effect
Crossvalidation, using the original variables (LOO CV, LMO CV)	insufficient for model validation
Y scrambling, using the original variables	misleading
Y scrambling with new variable selection	may be misleading
Leave-one-out crossvalidation with new variable selection in every CV run	misleading in larger data sets
Leave-many-out (up to 30%) cross-validation with new variable selection in every CV run	the only reliable validation procedure

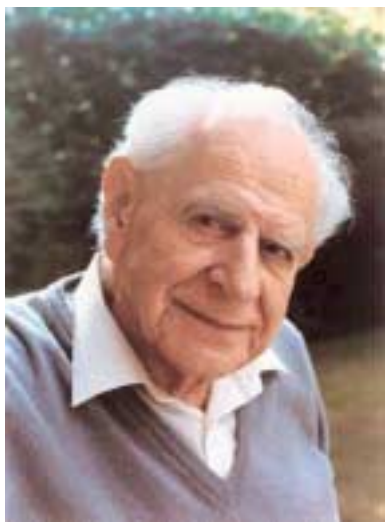
see also T. Scior et al., Curr. Med. Chem. 16, 4297-4313 (2009)

“Good” QSAR
- **parameters with biophysical relevance**
- **few variables to select**
- **few variables in the model**
- **validation by LOO, LMO, y scrambling**



“Poor” QSAR
- **artificial parameters**
- **too many variables to select**
- **many variables in the model**
- **no test set predictivity (“Kubinyi paradox”)**

Hugo Kubinyi, www.kubinyi.de



Sir Karl Popper
★1902 Vienna, ✞ 1998 London

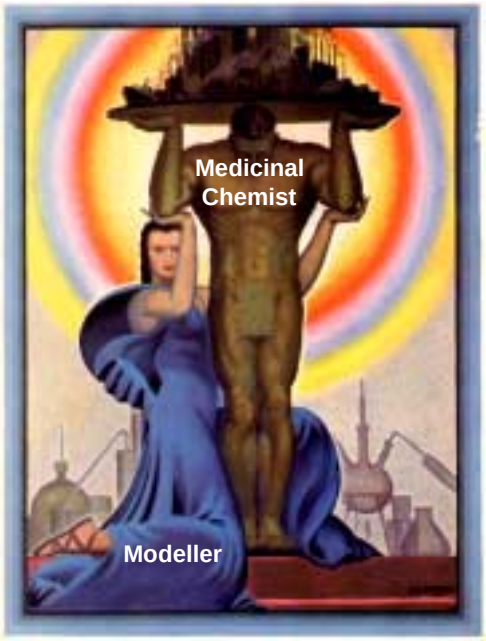
Good and Poor Science

[one has to] „differentiate between science and pseudoscience, knowing very well that science often errs and that pseudoscience may happen to stumble on the truth“

„it is easy to obtain confirmations - if one looks for them“

„a theory which is not refutable ... is non-scientific“

„some theories, when found to be false, are still upheld by their admirers - for example by introducing some auxiliary assumption, or by reinterpreting the theory *ad hoc* in such a way that it escapes refutation“



A. Cressy Morrison

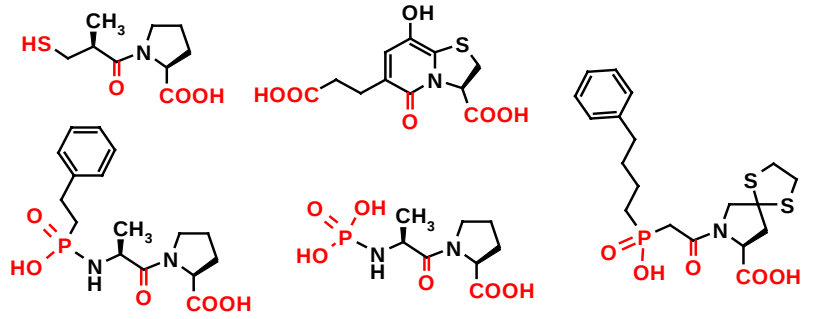
Man in a Chemical World
The Service of Chemical Industry

Ch. Scribner's Sons, NY, 1937

„Chemical Industry, Upheld by Pure Science, Sustains the Production of Man's Necessities"

Historical Pharmacophore Definition:

ACE Inhibitors

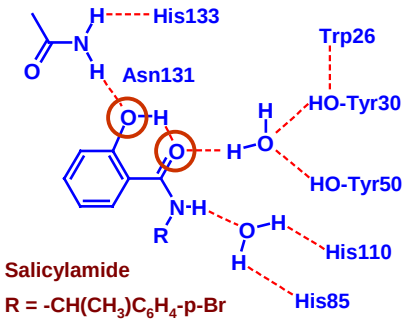


defined by functional groups

-SH, -COOH, $-PO_3H_2$, $>PO_2H$

C=O **COOH**

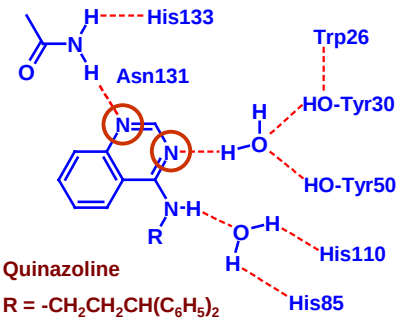
Receptors Just Recognize Properties



Salicylamide

R = -CH(CH$_3$)C$_6$H$_4$-p-Br

K_i = 0.14 nM



Quinazoline

R = -CH$_2$CH$_2$CH(C$_6$H$_5$)$_2$

K_i = 0.15 nM

A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger its biological response.

C. G. Wermuth et al., Pure Appl. Chem. 70, 1129-1143 (1998)

Hugo Kubinyi, www.kubinyi.de

Pharmacophore Generation and 3D Searches

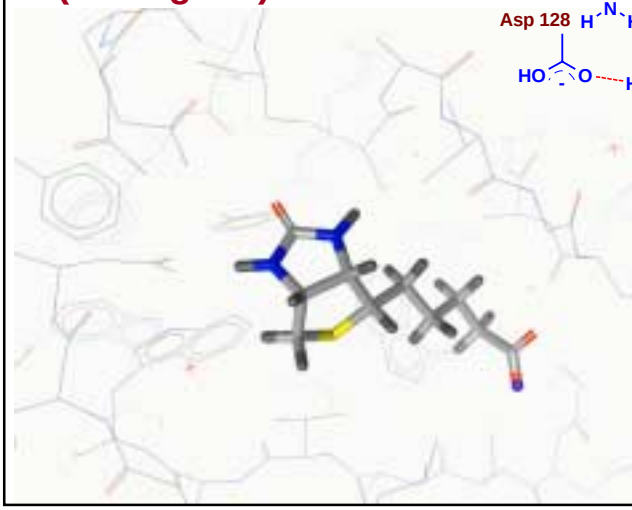
Catalyst (Accelrys)
established tool for hypothesis generation and 3D searches

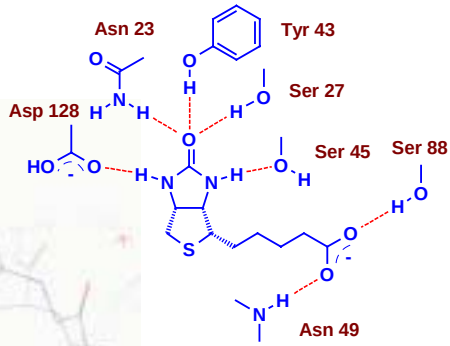
CATS (Roche) topological pharmacophores - no 3D structures required

FTree (feature trees; BioSolveIT)
no 3D structures required, ultrafast searches

LigandScout (inte:ligand)
automated generation of pharmacophores from protein 3D structures; ligand-based pharmacophore generation; 3D searches

LigandScout Pharmacophore Recognition (inte:ligand)

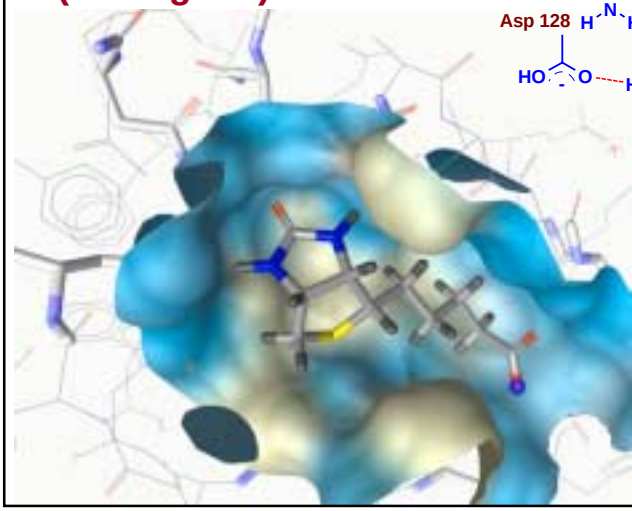


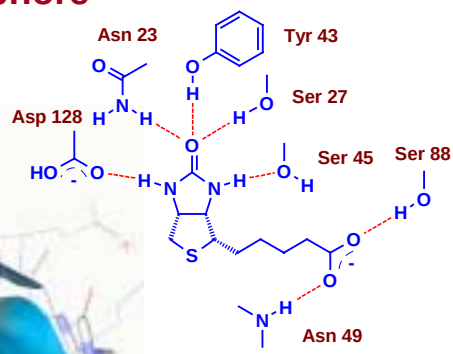


Biotin Streptavidin Complex (2rtf, 1.47Å)

Hugo Kubinyi, www.kubinyi.de

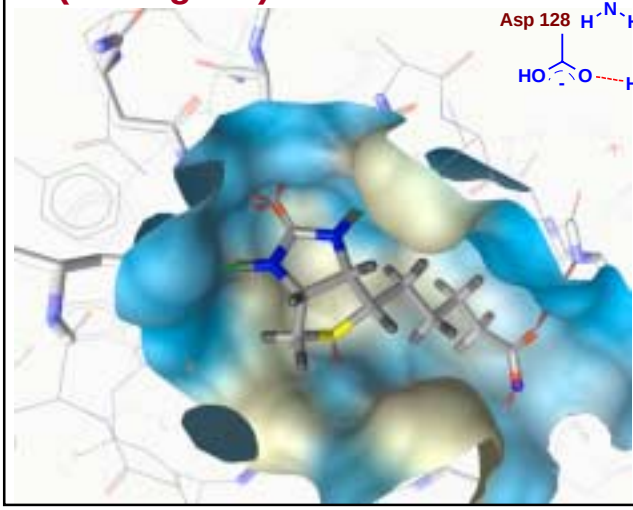
LigandScout Pharmacophore Recognition (inte:ligand)

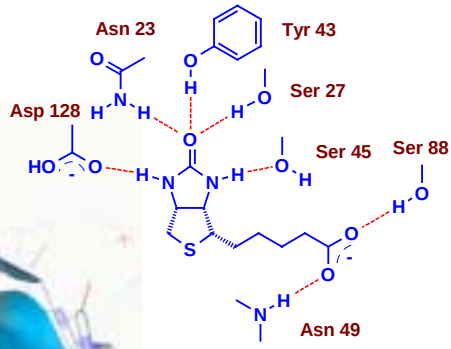




Biotin Streptavidin Complex (2rtf, 1.47Å)

LigandScout Pharmacophore Recognition (inte:ligand)

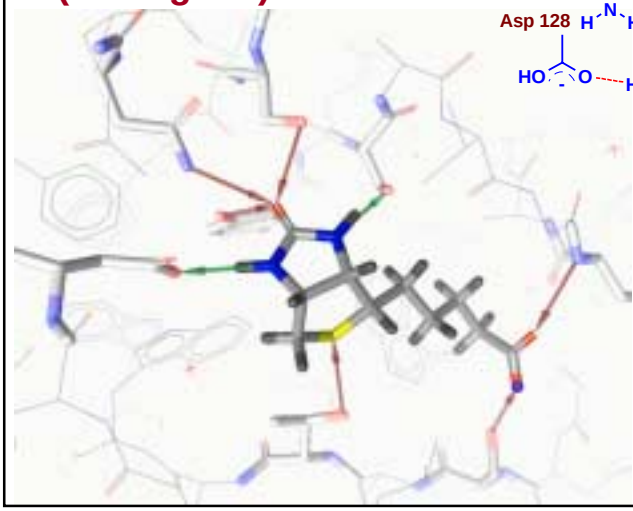


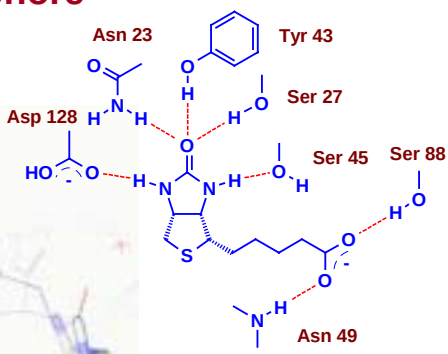


Biotin Streptavidin Complex (2rtf, 1.47Å)

Hugo Kubinyi, www.kubinyi.de

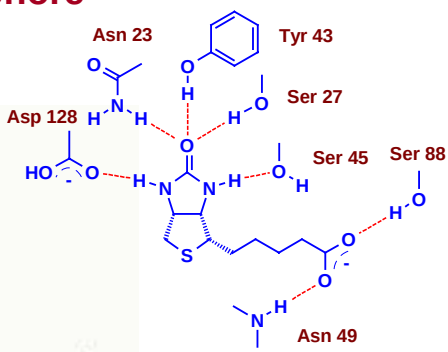
LigandScout Pharmacophore Recognition (inte:ligand)

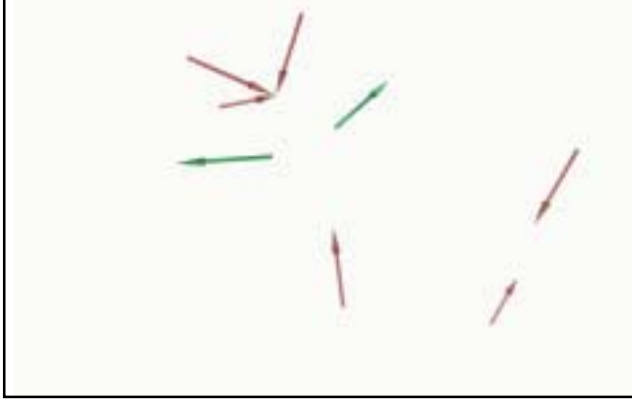




Biotin Streptavidin Complex (2rtf, 1.47Å)

LigandScout Pharmacophore Recognition (inte:ligand)

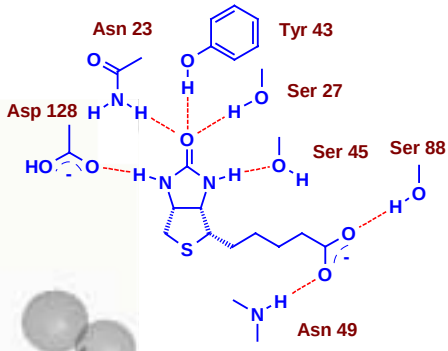


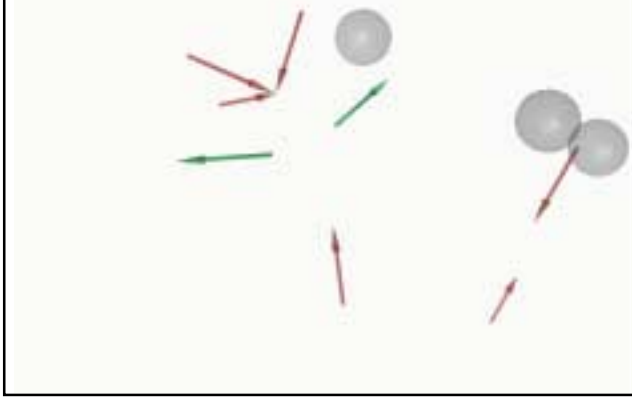


Biotin Streptavidin Complex (2rtf, 1.47Å)

Hugo Kubinyi, www.kubinyi.de

LigandScout Pharmacophore Recognition (inte:ligand)





Biotin Streptavidin Complex (2rtf, 1.47Å)

Problems in Pharmacophore Generation

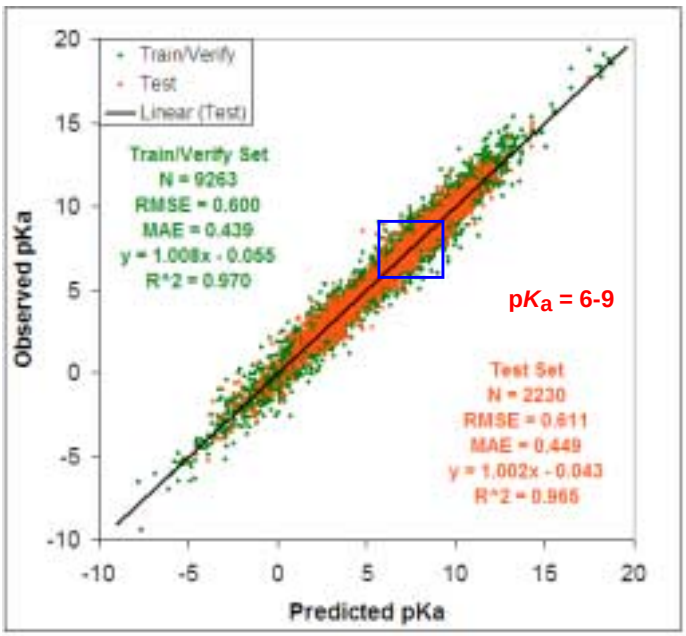
Isomers, enantiomers, diastereomers

Ionisation and Dissociation
(Sadowski rules, ACS Boston, 2002)

Tautomeric and protomeric forms
(program AGENT, ETH Zurich; ChemoSoft, ChemDiv; LigPrep, Schroedinger; and several others)

Acceptor properties of oxygen and sulfur atoms
(esters, aromatic ethers, oxazoles, isoxazoles, thiazoles, etc.)

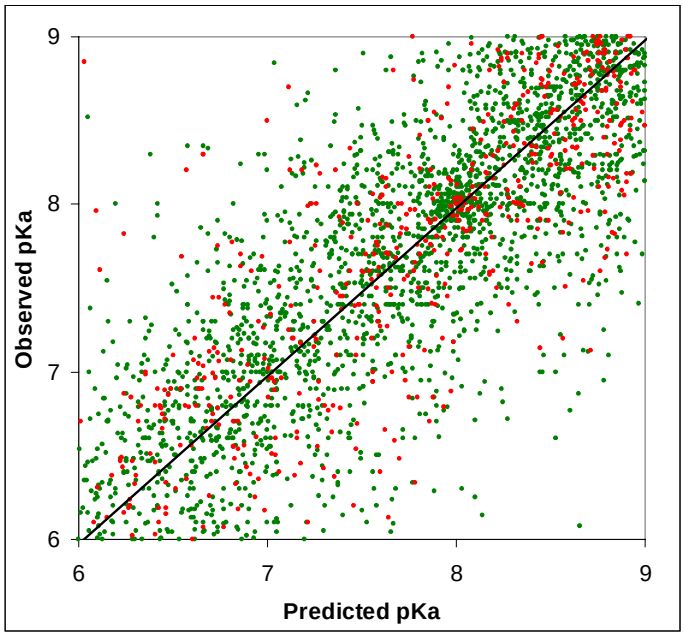
Superposition of flexible molecules



Software for pK_a Prediction

pKa model in ADMET Predictor 4.0

www.simulations-plus.com/Definitions.aspx?IID=55



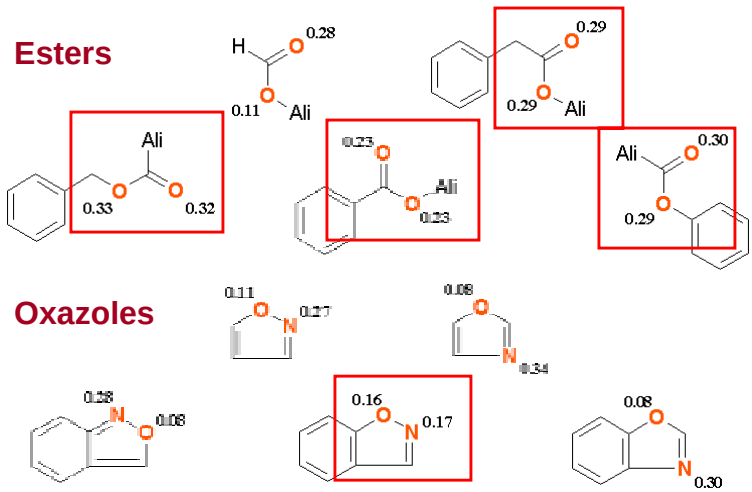
Software for pK_a Prediction

pKa model in ADMET Predictor 4.0

courtesy of Robert Fraczkiewicz, Simulations Plus, Inc.

Hugo Kubinyi, www.kubinyi.de

Acceptor Potentials of Esters and Oxazoles



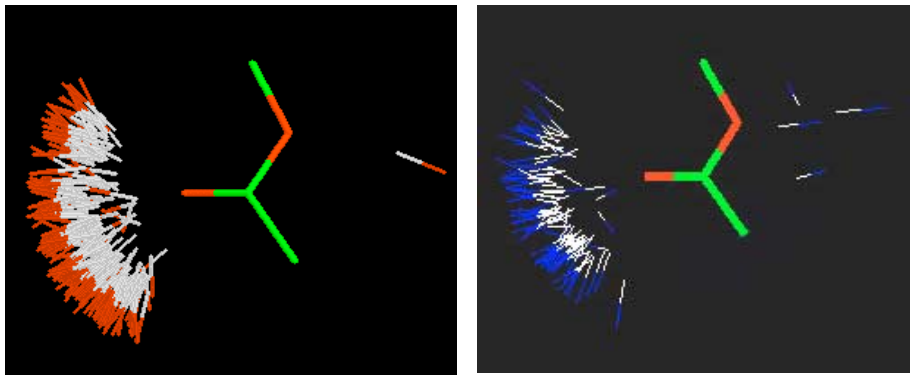
S. Rey et al., J. Mol. Graphics Model. 19, 521-535 (2001)

Pharmacophore Analyses Must Consider Correct Donor and Acceptor Properties of Ligands

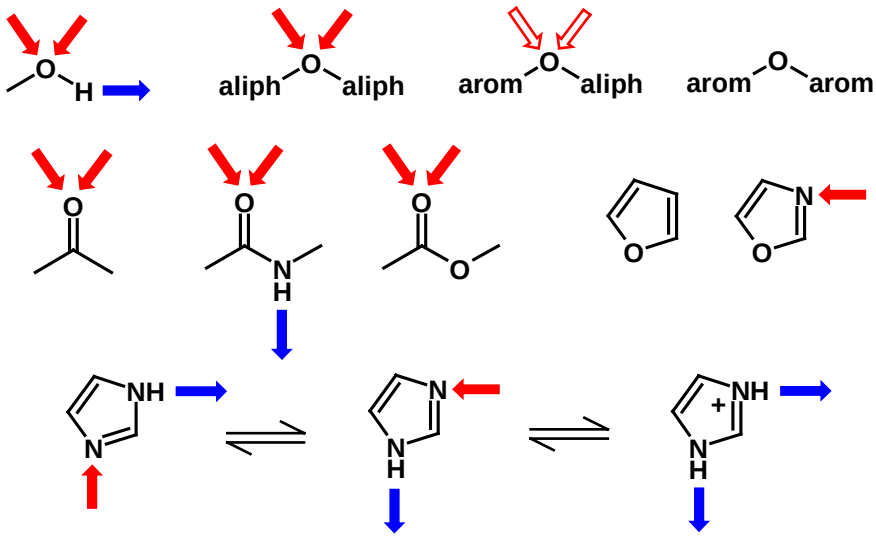
The billion dollar question:

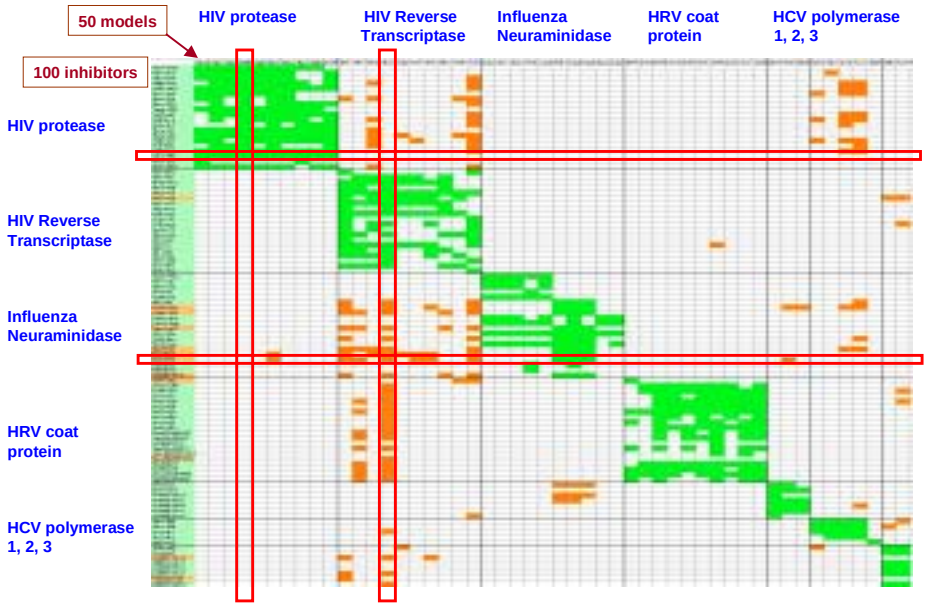
how many acceptor positions has an ester group ?

Correct answer: Two ...., but why?



Donor and Acceptor Properties of O and N





T. Steindl et al., J. Chem. Inf. Model. 46, 2146-2157 (2006)

Hugo Kubinyi, www.kubinyi.de

Drug Research is ....

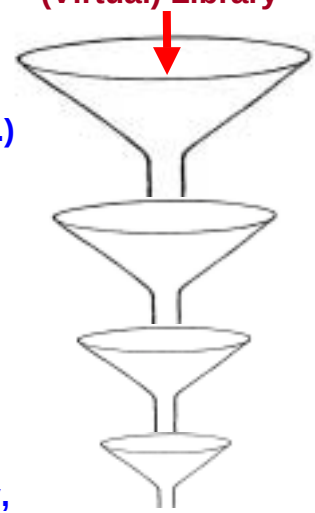
the Search for a Needle in a Haystack

Tools for Virtual Screening

	remaining
Garbage filter	90%
Druglike / Non-druglike	75%
Bioavailability	60%
Cytotoxicity	:
hERG channel inhibiton	:
Antitargets	:
α_{1a} (orthostatic hypotension)	:
D2 (extrapyramidal syndrome)	:
5-HT_{2c} (obesity)	:
musc. M1 (hallucinations, memory)	:
CYP inhibition (3A4, 2C9, 2D6)	:
Pharmacophore searches	:
Docking and scoring	0% ?

Stepwise Virtual Screening

(Virtual) Library



Property Filters
(MW, rule of 5, nRot, drug-like, ...)

1D Pharmacophore and
3D Pharmacophore Searches

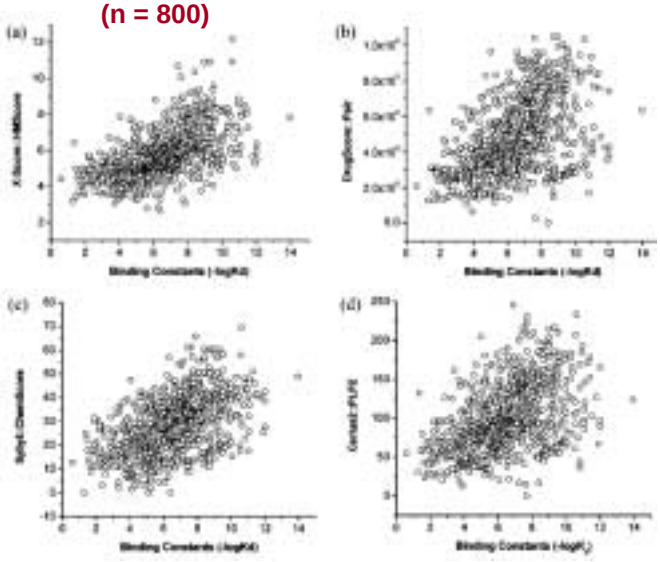
Docking and Scoring

Selection by Diversity, Similarity,
and Visual Inspection

Leads / Candidates

Performance of Different Scoring Functions

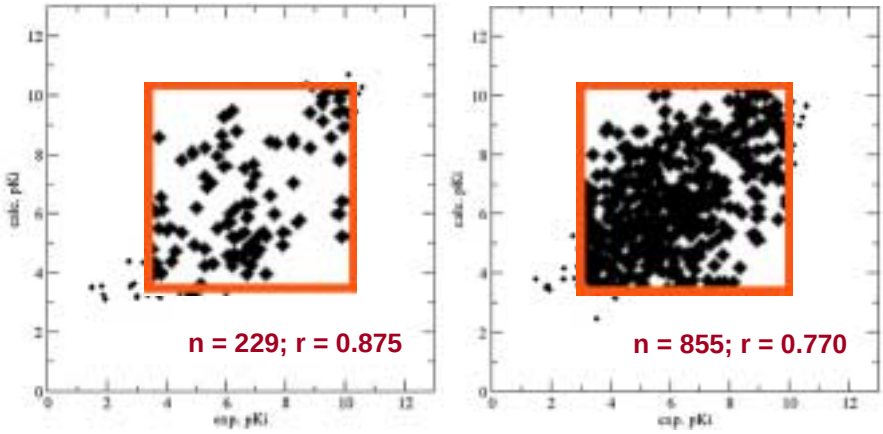
(n = 800)



a) X-Score
b) DrugScore
c) ChemScore
d) PLP2

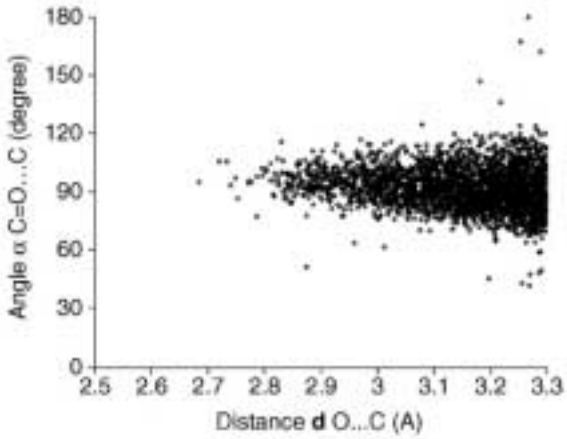
R. Wang et al., J. Chem. Inf. Model. 44, 2114-2125 (2004)

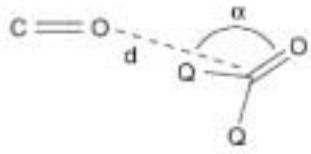
SFCscore (Scoring Function Consortium): Affinity Prediction of Protein-Ligand Complexes



C. A. Sotriffer et al., Proteins 73, 395-419 (2008); cf. A. M. Davis et al., Angew. Chem. Int. Ed. Engl. 42, 2718-36 (2003)

Unrecognized Favorable Interactions



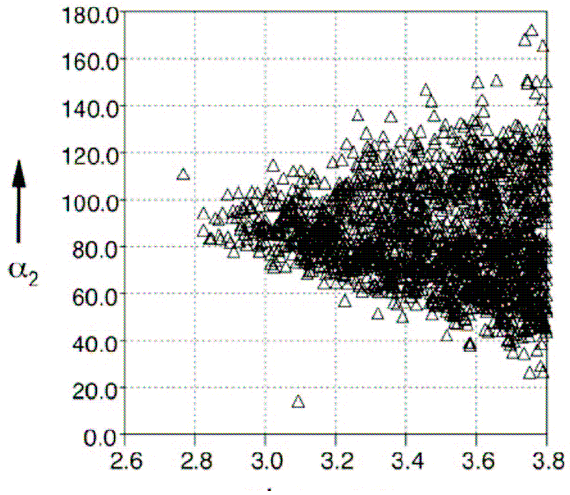


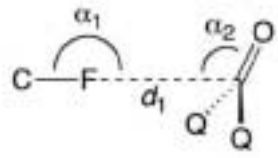
derived from 2,850 high-resolution CSD structures (Q = C, N, O)

T. Schulz-Gasch and M. Stahl, Drug Discov. Today: Technologies 1, 231-239 (2004)

Hugo Kubinyi, www.kubinyi.de

Unrecognized Favorable Interactions

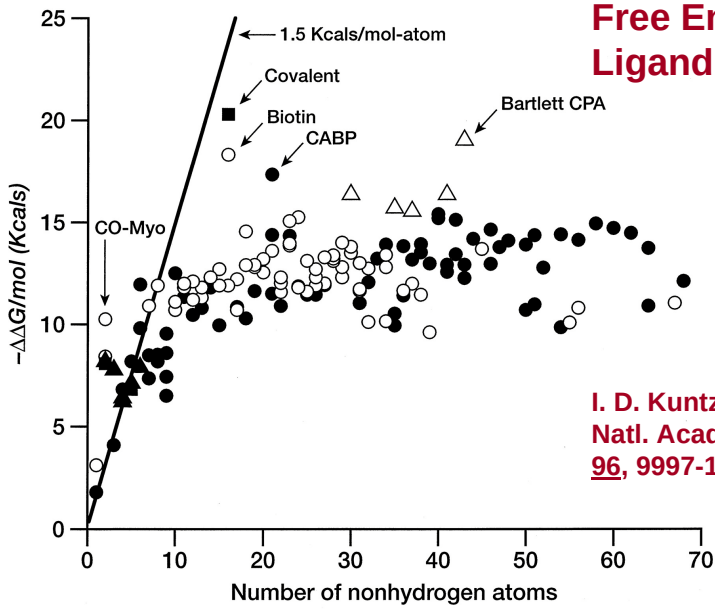




derived from 1,087 high-resolution CSD structures (Q = C, N, O)

M. Zürcher and F. Diederich, J.Org. Chem. 73, 4345-4361 (2008)

Free Energy of Ligand Binding



I. D. Kuntz et al., Proc. Natl. Acad. Sci. USA 96, 9997-10002 (1999)

Hugo Kubinyi, www.kubinyi.de

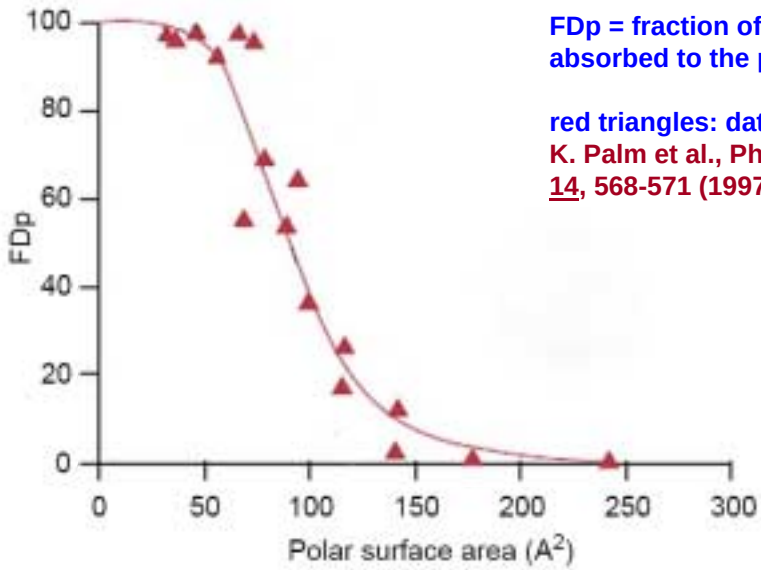
Factors to be Considered in Scoring Functions

- Desolvation enthalpy and entropy (ligand and protein)
- Protonation state of the ligand and the binding site
- Distortion energy of the ligand and its binding site
- Loss of translational and rotational degrees of freedom of the ligand
- MEP + dielectric constant at the binding site
- Dipole moment of the ligand and local dipole moment at the binding site
- Binding enthalpy of the ligand-protein complex
- Repulsive effects (e.g. -O···O-)
- Inserted water molecules
- Solvation enthalpy and entropy of the complex

Drug Discovery Bottlenecks of the Past

Problem	Solution
Target search	genome information
Target validation	knock-outs, RNA silencing
Lead search	in vitro test models, HTS, VS
Lead optimization	automated parallel syntheses, chemogenomics
Absorption, permeability	Lipinski rules, Caco cells, formulation, prodrugs
Metabolism	MetaSite, MetaPrint2D, liver microsomes, hepatocytes
Toxicity	Ames test, hERG models, etc.
Drug-drug interactions	CYP inhibition/induction

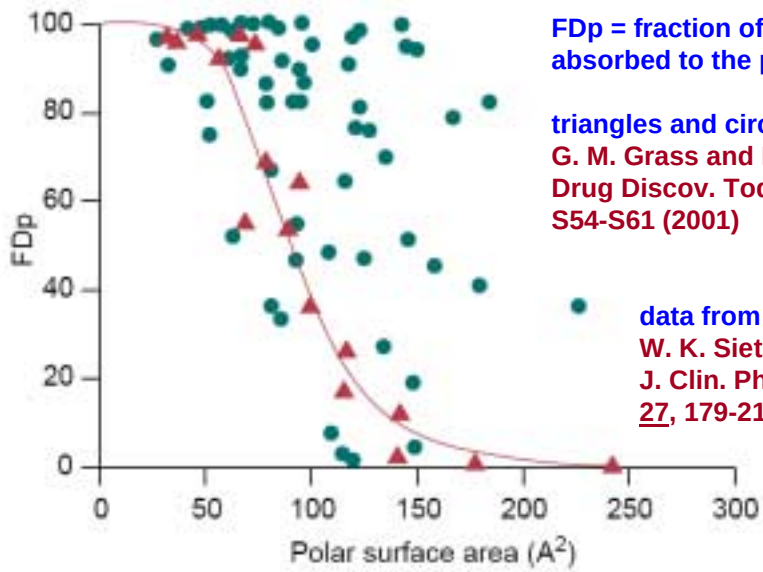
Human Absorption and Polar Surface Area



FDp = fraction of dose absorbed to the portal vein

red triangles: data from K. Palm et al., Pharm. Res. 14, 568-571 (1997)

Human Absorption and Polar Surface Area

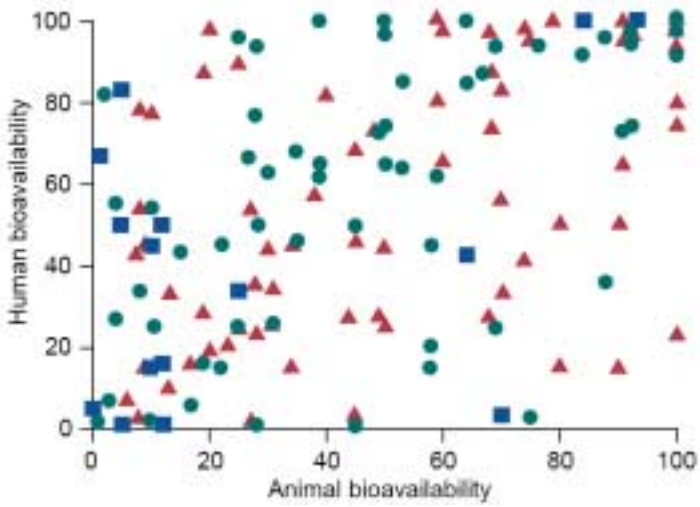


FDp = fraction of dose absorbed to the portal vein

triangles and circles:
G. M. Grass and P. J. Sinko, Drug Discov. Today 6, S54-S61 (2001)

data from
W. K. Sietsema, Int. J. Clin. Pharmacol. 27, 179-211 (1989)

Rodent, Dog, Primate and Human Bioavailability



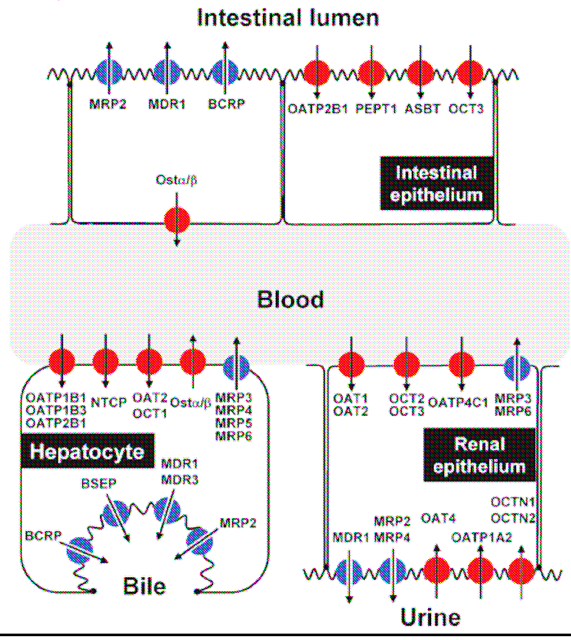
Rodents
green circles

Dog
red triangles

Primates
blue squares

data from:
W. K. Sietsema, Int. J. Clin. Pharmacol. 27, 179-211 (1989)

G. M. Grass and P. J. Sinko, Drug Discov. Today 6, S54-S61 (2001)

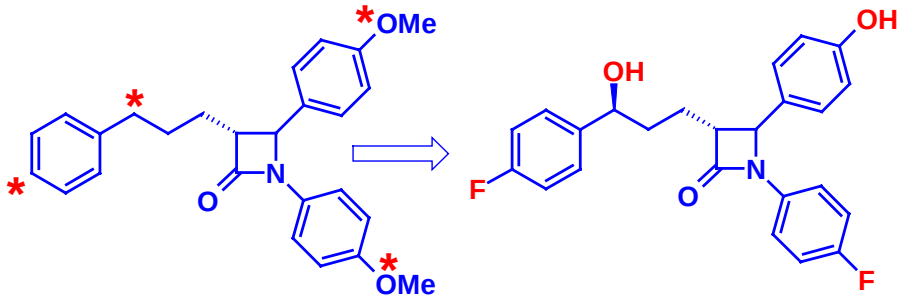


The Role of Transporters in Drug Absorption and Elimination

H. Gleaser et al., in R. J. Vaz and T. Klabunde, Antitargets, Wiley-VCH, 2008, pp. 341-366

Hugo Kubinyi, www.kubinyi.de

Oxidative Metabolism and Drug Design



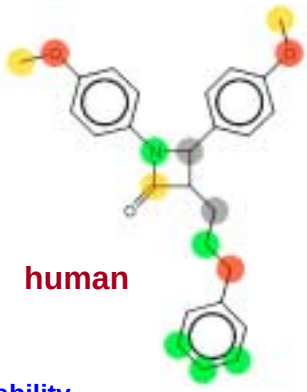
SCH 48461
ED_{50} (hamster) = 2.2 mg/kg

Ezetimib (SCH 58235, oral cholesterol absorption inhibitor)
ED_{50} (hamster) = 0.04 mg/kg

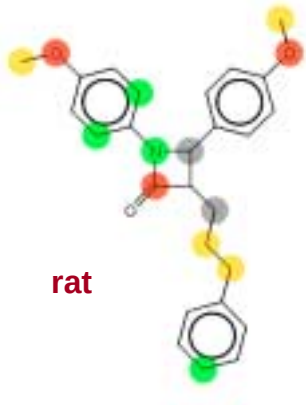
M. van Heek et al., J. Pharmacol. Exp. Ther. 283, 157-163 (1997);
D. A. Smith, H. van de Waterbeemd and D. K. Walker, Pharmacokinetics and Metabolism in Drug Design, Wiley-VCH, 2001, p. 85

Prediction of Drug Metabolism: MetaPrint2D

predictions for human, dog, rat, all



human



rat

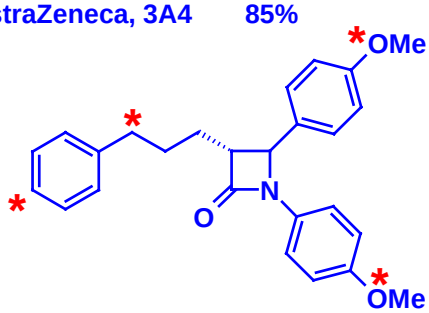
red = high probability
orange = medium proability
green = low probability
white = no probability

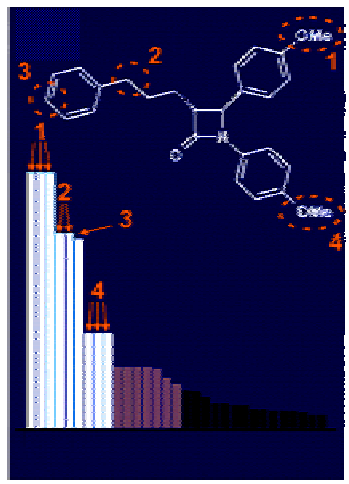
S. Boyer et al.,
www-metaprint2d.ch.cam.ac.uk/

Prediction of Drug Metabolism: MetaSite

correct predictions:

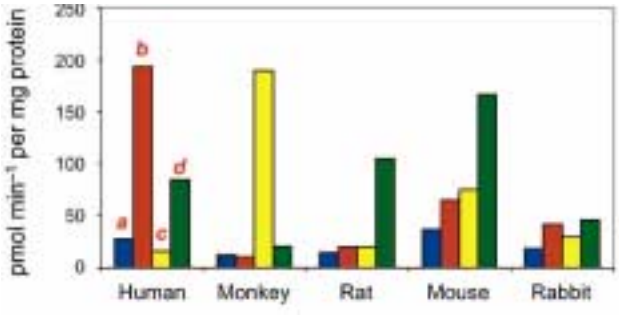
Sanofi-Aventis, 2C9	84%
Pfizer, 2D6	85%
3A4	86%
J&J, 2C9, 2D6, 3A3	85%
AstraZeneca, 3A4	85%

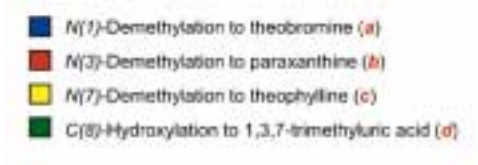




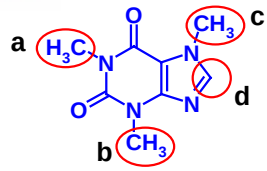
G. Cruciani et al., J. Med. Chem. 48, 6970-6979 (2005)

Species Differences of Caffeine Metabolism



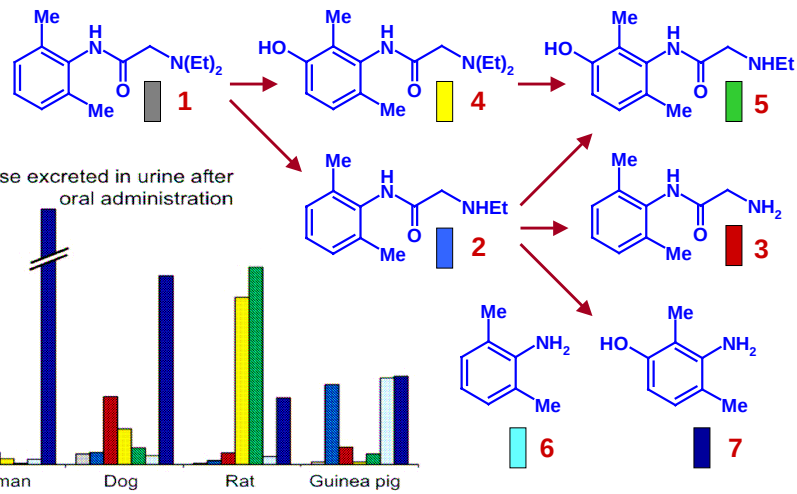


production of caffeine metabolites by liver microsomes of different species



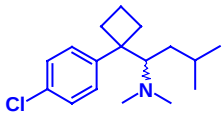
F. Berthou et al., Xenobiotica 22, 671-680 (1992)
figure: S. D. Krämer and B. Testa, Chemistry & Biodiversity 5, 2465-2578 (2008)

Species Differences of Lidocaine Metabolism



J. B. Keenaghan and R. N. Boyes, J. Pharmacol. Exp. Ther. 180, 459-463 (1972)
figure: S. D. Krämer and B. Testa, Chemistry & Biodiversity 5, 2465-2578 (2008)

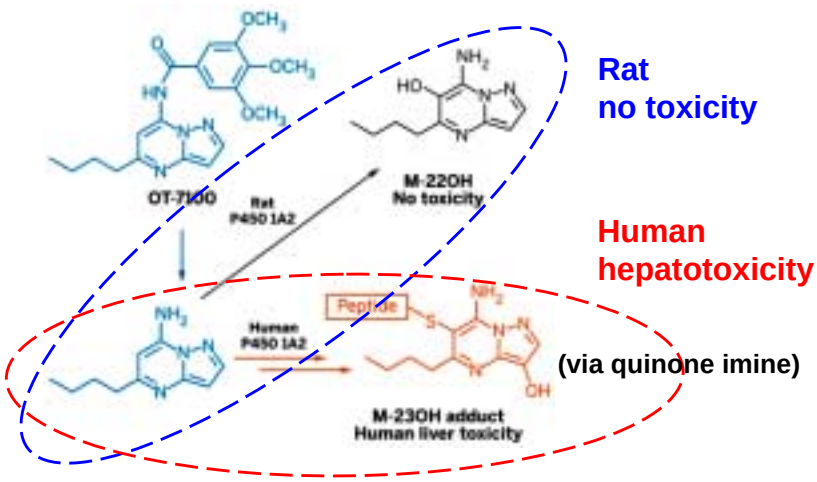
Biological Activities of Metabolites

Compound	monoamine uptake inhibition rat synaptosomes, IC_{50} in nM		
Sibutramine (racemate)	DAT	NET	SERT
	1200	350	2800
(R) 	12	4	44
(S) 	180	870	9200
(R) 	9	13	140
(S) 	12	62	4300

D. L. Nelson and D. R. Gehlert, Endocrine 29, 49-60 (2006);
data from S. D. Glick et al., Eur. J. Pharmacol. 397, 93-102 (2000)

Hugo Kubinyi, www.kubinyi.de

Biological Activities of Metabolites



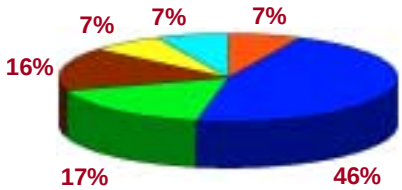
Rat
no toxicity

Human
hepatotoxicity

(via quinone imine)

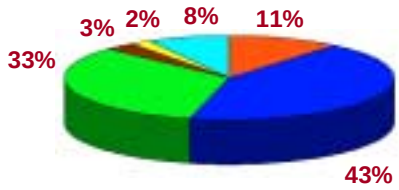
S. Kuribayashi et al., Chem Res. Toxicol. 22, 323-331 (2009);
cf. Chem. & Eng. News, August 31, 2009, p. 27

Reasons for Failure in Drug Development



- ADME
- Lack of efficacy
- Animal toxicity
- Adverse effects in man
- Commercial reasons
- Miscellaneous

Reasons for failure in clinical development, 1964-1985 (n = 121; without antiinfectives)
T. Kennedy, Drug Discov. today 2, 436-444 (1997)



- Liberation + ADME
- Lack of efficacy
- Toxicity
- Economic
- Other
- Not published

Reasons for failure in clinical development, 1992-2002 (n = 73) (reasons for market withdrawal, n = 16: toxicity 93%, efficacy 7%)
D. Schuster et al., Curr. Pharm. Design 11, 3545-3559 (2005)

Thank you

Pieter van Musschenbroek (1692-1761)
Tentamina Experimentorum Naturalium
(Museo di Storia Naturale dell'Accademia dei Fisiocritici di Siena)