Drug Design - Problems in Prediction

Hugo Kubinyi

Germany

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

CCG User Group Meeting, Amsterdam, April 2013

Ascending and Descending (M. C. Escher, lithograph, 1960)

QSAR and Modelling: Living in Castalia?

In Castalia, intellectual efforts have no other purpose than the preservation and advancement of intellectual foundations of culture and humanity ... [they] engage in an intellectual exercise, the "Glass Bead Game", which aims at connecting scientific and cultural values within a formal framework of mathematics and music ...

Hermann Hesse
„The Glass Bead Game"

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)

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

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)

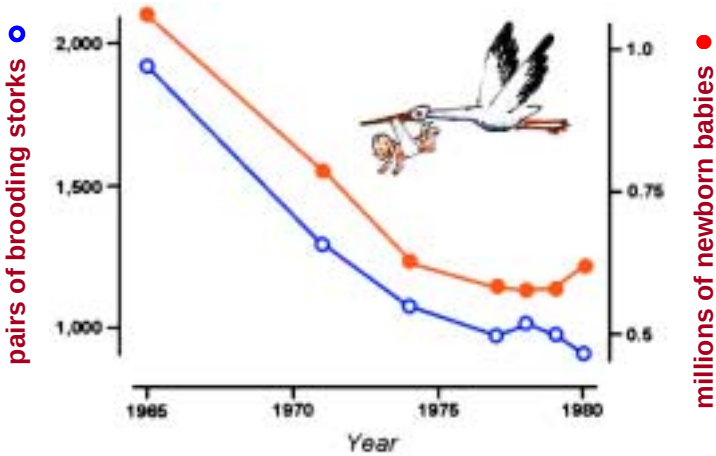
Is QSAR relevant to Drug Discovery?
A. M. Doweyko, Idrugs 11, 894-899 (2008)

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

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)

The Storks and the Babies



Sir – There is concern in West Germany over the falling birth rate. The accompanying graph might suggest a solution that every child knows makes sense. H. Sies, Nature 332, 495 (1988)

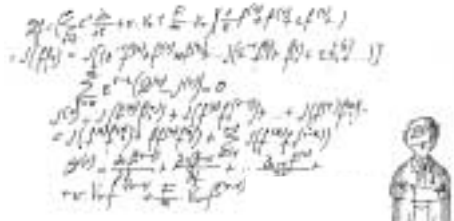
QSAR: The Texas Sharpshooter Fallacy

A Texan fires several shots at the door of a barn, then paints a target around the hits and claims to be a sharpshooter.

Information is interpreted or manipulated until it appears to have a meaning: cryptograms in the work of Shakespeare, Nostradamus predictions, more children in town A have leukemia than in town B …

http://en.wikipedia.org/wiki/Texas_sharpshooter_fallacy

"Good" QSAR
- parameters with biophysical relevance
- few variables to select
- few variables in the model
- leave-many-out crossvalidation



"Poor" QSAR
- artificial parameters
- too many variables to select
- many variables in the model
- no test set predictivity

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“

Hugo Kubinyi, www.kubinyi.de

Drug Research is ....

the Search for a Needle in a Haystack

Virtual Screening Reduces the Size of the Haystack by Selecting:

Compounds or libraries that are either
- lead-like, or
- drug-like, or have the
- potential of oral bioavailability, or are
- similar to a lead, or
- fit the binding site of a certain protein

by
- rules (e.g. Lipinski bioavailability rules),
- neural nets (e.g. drug-like character),
- similarity analyses,
- pharmacophore analyses,
- scaffold hopping, or
- docking and scoring

Problems in Pharmacophore Generation

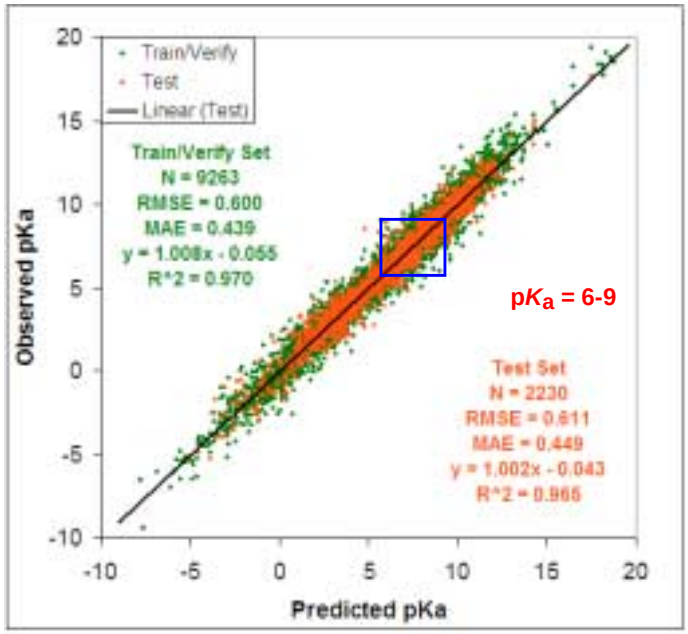
Isomers, enantiomers, diastereomers

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

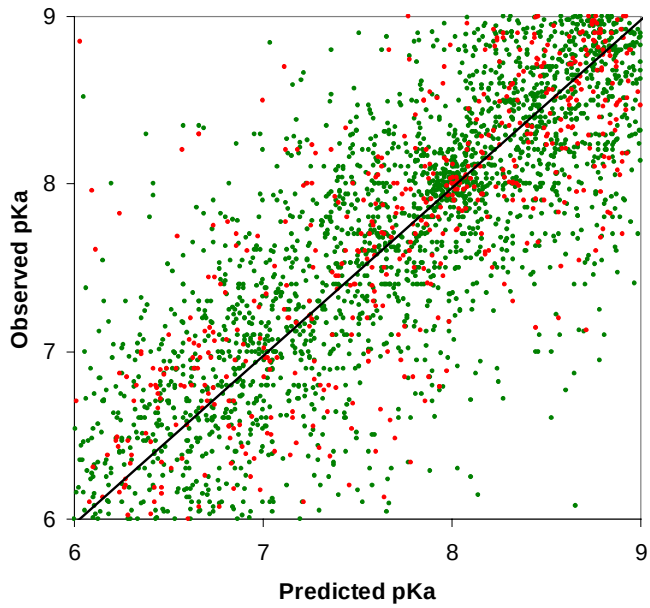


Software for pK_a Prediction

pKa model in ADMET Predictor 4.0

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

Hugo Kubinyi, www.kubinyi.de



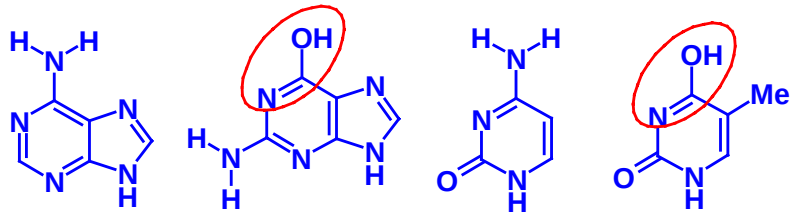
Software for pK_a Prediction

pKa model in ADMET Predictor 4.0

courtesy of Robert Fraczkiewicz, Simulations Plus, Inc.

The Discovery of the DNA Double Helix

Summer 1952: Erwin Chargaff critisizes that Francis Crick and James Watson are ignorant about the structures of the bases



adenine guanine cytosine thymine

J. N. Davidson, The Biochemistry of Nucleic Acids, London, 1950

early 1953: Pauling publishes a DNA model with a phosphate core
February 27, 1953: Jerry Donohue corrects the formulas of the bases
February 28, 1953: Watson and Crick derive the correct DNA model
April 02, 1953: Manuscript sent to Nature; published April 25, 1953

cited from: J. Watson and A. Berry, DNA. The Secret of Life, 2003

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

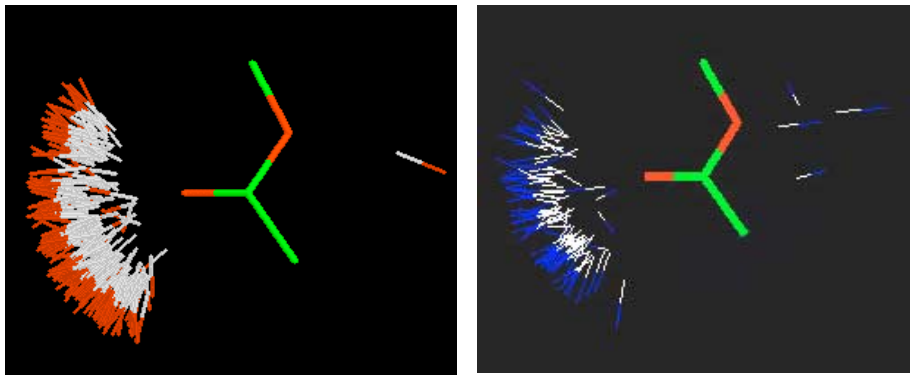
J. D. Watson and F. H. C. Crick, Nature 171, 737-738 (April 25, 1953)

Pharmacophore Analyses Must Consider Correct Donor and Acceptor Properties of Ligands

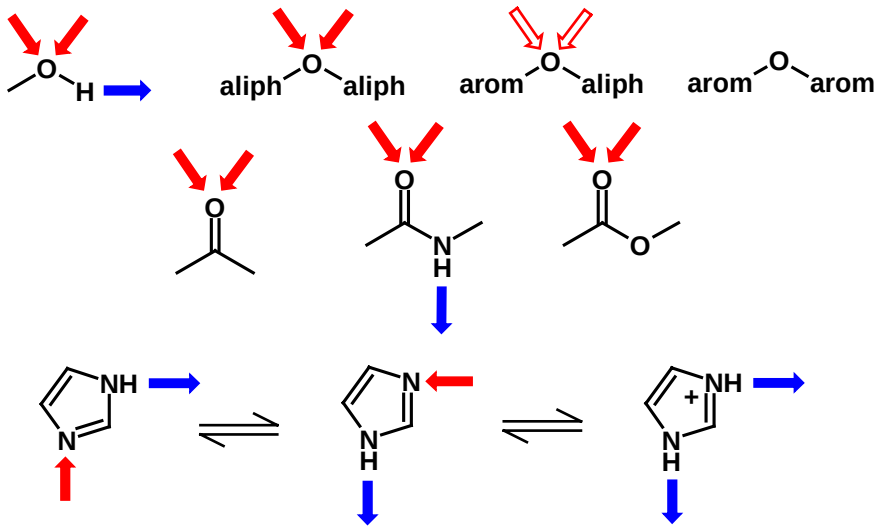
The billion dollar question:

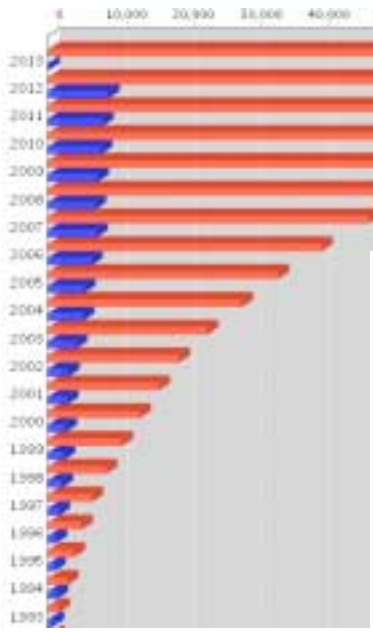
how many acceptor positions has an ester group ?

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



Donor and Acceptor Properties of O and N



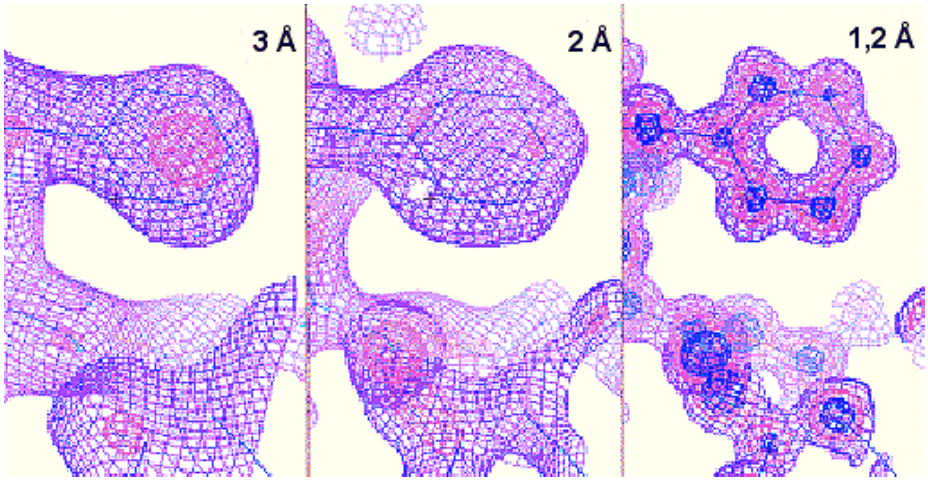


Annual and Cumulative Growth of the Protein Data Bank

76,706 X-ray structures (proteins: 71,626)
9,741 NMR structures (proteins: 8,529)
489 electron microsc. (proteins: 337)

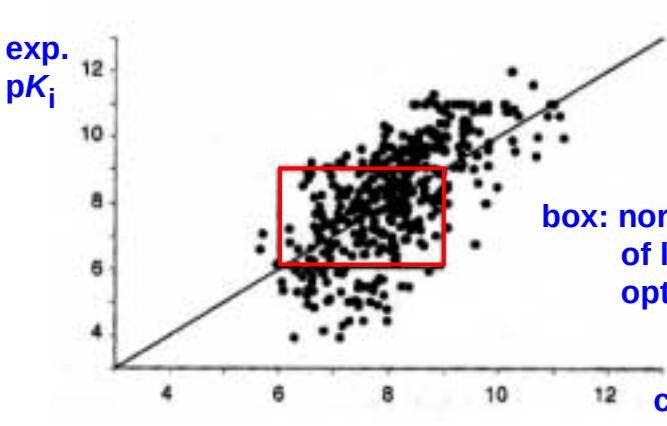
Date: January 02, 2013
Source: www.rscb.org/pdb/statistics

Problems of Resolution of Protein 3D Structures



lacking hydrogens, no differentiation between C, O and N (orientation of thr, amide groups in asn and gln, imidazole, etc.)

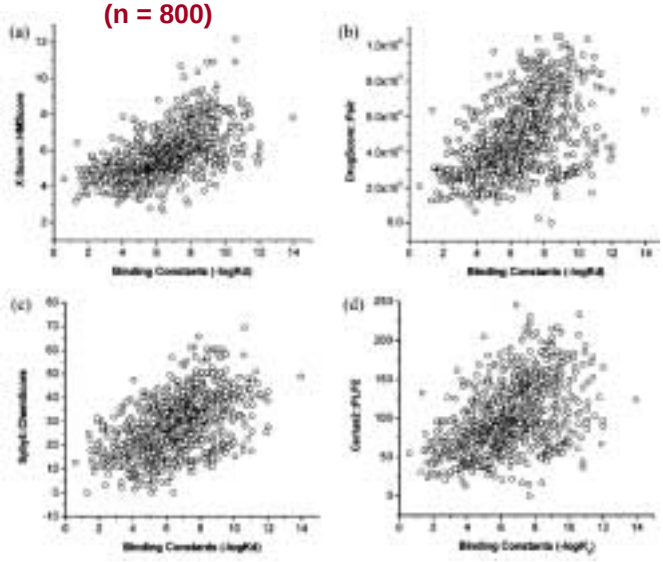
pK_i values of HIV Protease Inhibitors: VALIDATE II Predictions



A. M. Davis et al., Angew. Chem. Int. Ed. Engl. 42, 2718-2736 (2003); Angew, Chem. 115, 2822-2841 (2003)

Performance of Different Scoring Functions

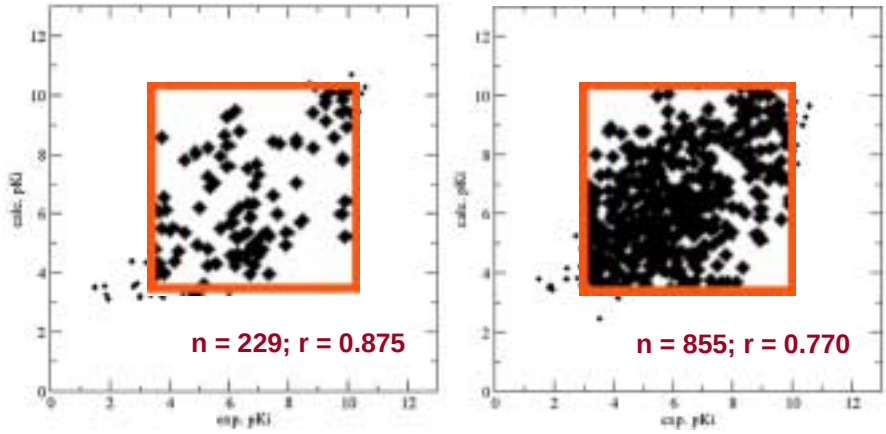
(n = 800)



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

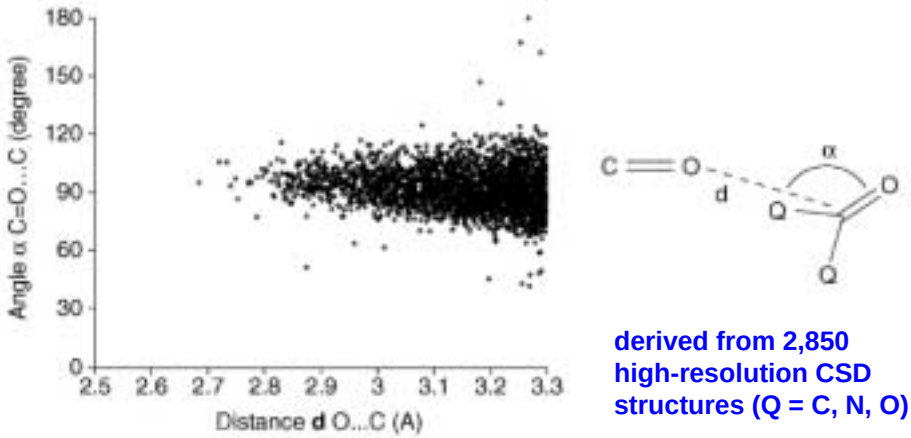
R. Wang et al., J. Chem. Inf. Model. 44, 2114-2125 (2004)

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)

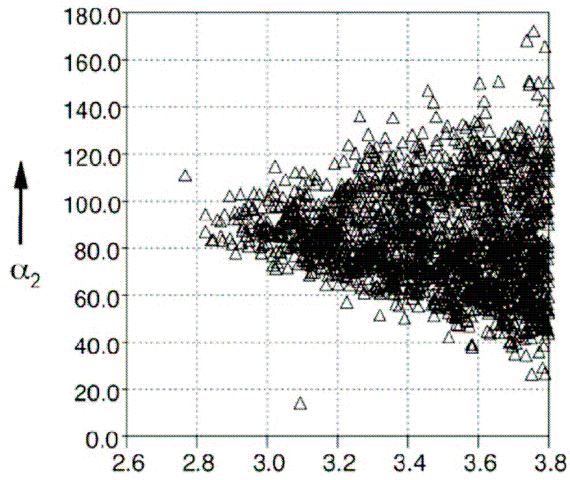
Unrecognized Favorable Interactions

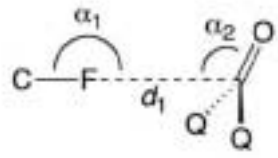


derived from 2,850 high-resolution CSD structures (Q = C, N, O)

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

Unrecognized Favorable Interactions

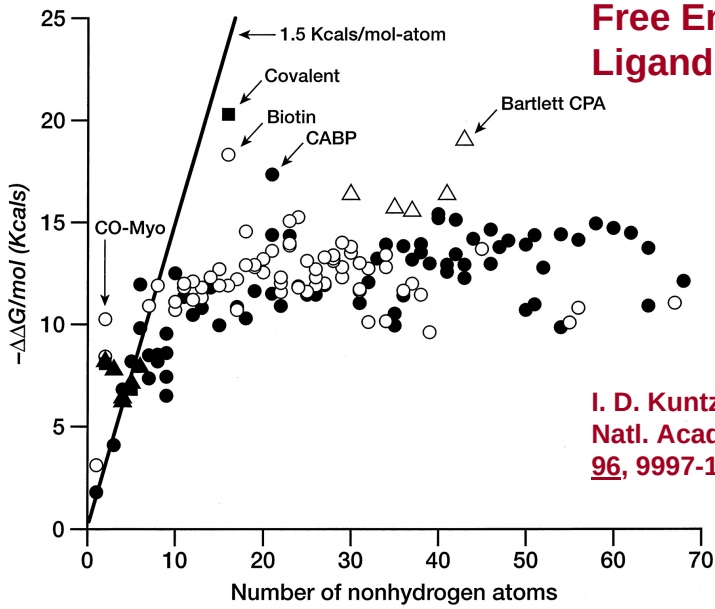




derived from 1,087 high-resolution CSD structures (Q = C, N, O)

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)

Stepwise Virtual Screening

(Virtual) Library

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

1D Pharmacophore and
3D Pharmacophore Searches

Docking and Scoring

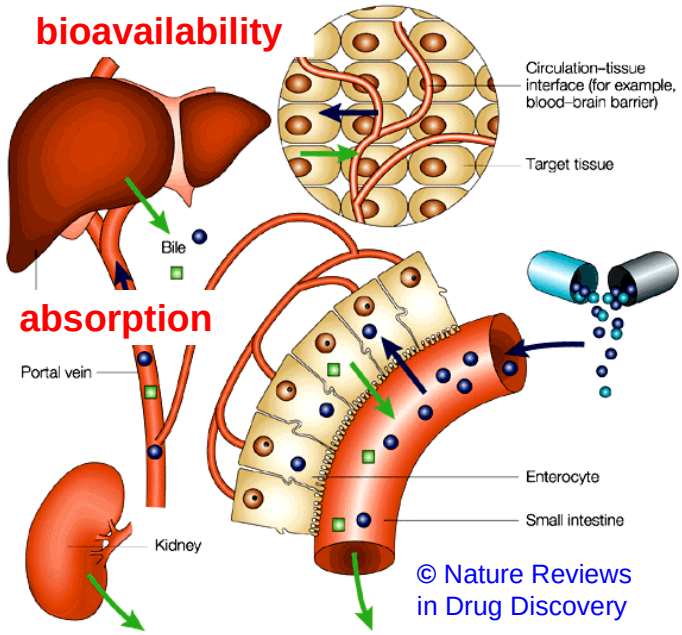
Selection (Diversity, Similarity, Inspection)

Leads / Candidates

Tools for Virtual Screening

	remaining
Garbage filter	90%
Druglike / Non-druglike	75%
Bioavailability	60%
Cytotoxicity	:
hERG channel inhibition	:
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% ?

bioavailability



Sites of Drug Metabolism:

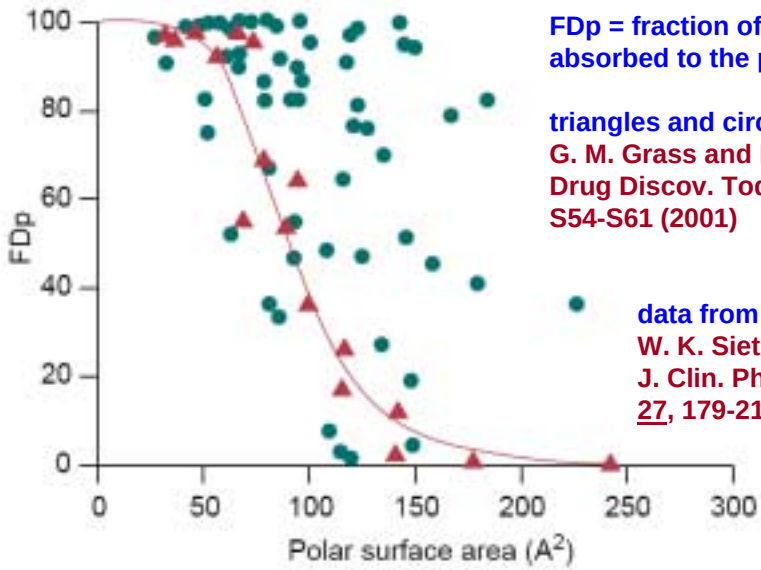
(intestinal wall), liver, (organs)

Sites of Drug Elimination:

kidneys (polar compounds), bile, feces (lipophilic analogs), lung

Hugo Kubinyi, www.kubinyi.de

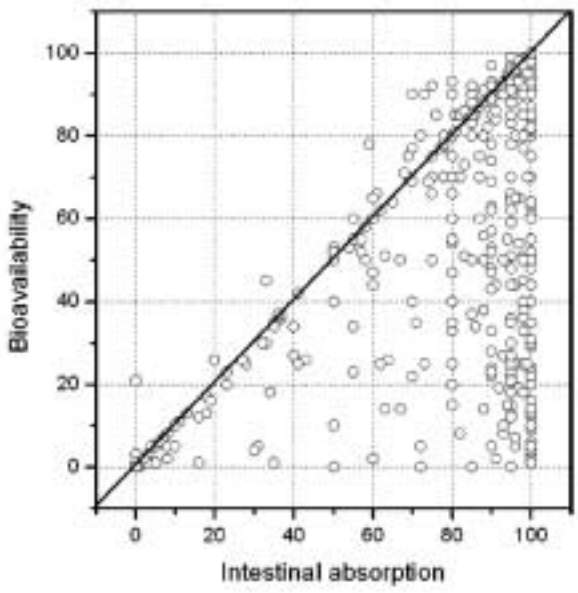
Human Absorption and Polar Surface Area



FDp = fraction of dose absorbed to the portal vein

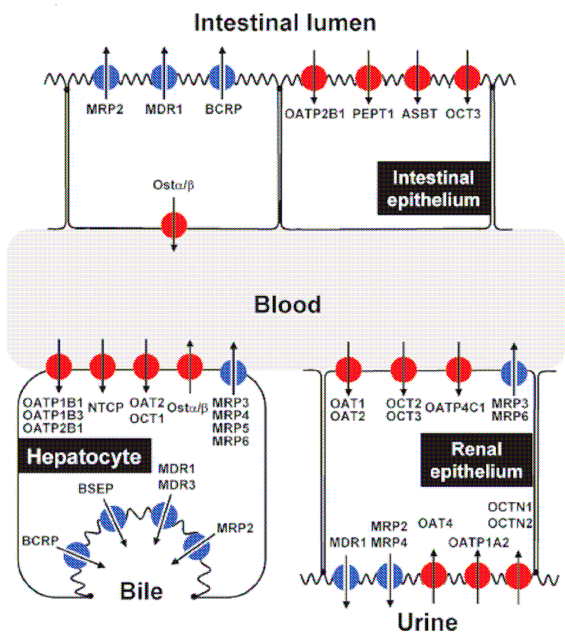
triangles and circles:
G. M. Grass and P. J. Sinko, Drug Discov. Today 6, S54-S61 (2001)

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



Bioavailability vs. Human Intestinal Absorption (n = 470)

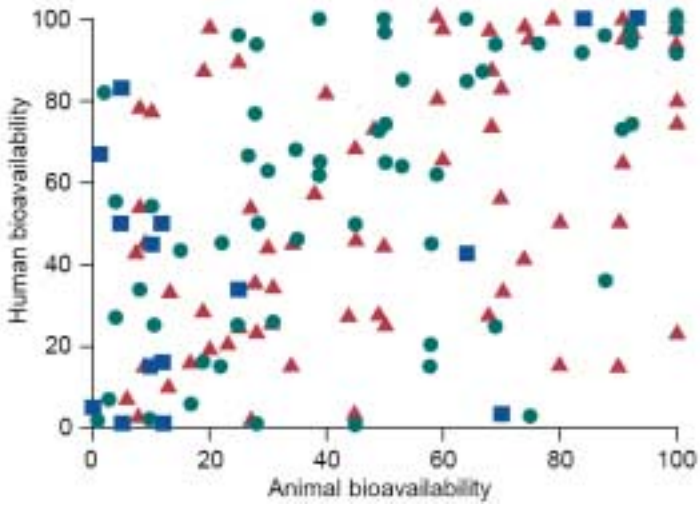
T. Hou et al., J. Chem. Inf. Model. 47, 208-218 (2007)



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

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)

A Set of Simple ADMET Rules

neutral molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4
solubility	average	lower
permeability*	higher	average/higher
bioavailability	average	lower
volume of Dist.**	average	average
plasma protein binding	average	higher
CNS penetration***	higher/average	average/lower
brain tissue binding	lower	higher
P-gp efflux	average	higher/average
in-vivo clearance	average	average
hERG Inhibition	lower	lower
P450 inhibition****	lower 2C9, 2C19, 2D6 & 3A4 inhibition	higher 2C9, 2C19 & 3A4 inhibition
P450 inhibition****	higher 1A2 inhibition	lower 1A2 inhibition
P450 inhibition****		average 2D6 inhibition

M. P. Gleeson, J. Med. Chem. 51, 817-834 (2008).

A Set of Simple ADMET Rules

basic molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4
solubility	higher/average	lower/average
permeability*	higher/average	average
bioavailability	average	lower
volume of Dist.**	higher/average	higher
plasma protein binding	lower	average
CNS penetration***	higher/average	average/lower
brain tissue binding	lower	higher
P-gp efflux	average	higher/average
in-vivo clearance	average	higher/average
hERG Inhibition	average/higher	higher
P450 inhibition****	lower 1A2, 2C9, & 2C19 inhibition	lower 1A2 inhibition
P450 inhibition****	average 2D6 & 3A4 inhibition	average 2C9, 2C19 inhibition
P450 inhibition****		higher 2D6 & 3A4 inhibition

M. P. Gleeson, J. Med. Chem. 51, 817-834 (2008).

Hugo Kubinyi, www.kubinyi.de

A Set of Simple ADMET Rules

acidic molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4
solubility	higher	average/higher
permeability*	lower	average/lower
bioavailability	average	average
volume of Dist.**	lower	lower
plasma protein binding	average/higher	higher
CNS penetration***	lower	lower
brain tissue binding	lower	higher
P-gp efflux	lower	lower
in-vivo clearance	lower/average	average
hERG Inhibition	lower	lower
P450 inhibition****	lower 1A2, 2C9, 2C19, 2D6 & 3A4 inhibition	lower 1A2, 2C19, 2D6 & 3A4 inhibition
P450 inhibition****		higher 2C9 inhibition

M. P. Gleeson, J. Med. Chem. 51, 817-834 (2008).

A Set of Simple ADMET Rules

(d) zwitterionic molecules	MWT < 400 and clogP < 4	MWT > 400 and/or clogP > 4
solubility	higher	average/higher
permeability*	lower	lower/average
bioavailability	lower	lower
volume of Dist.**	lower	average/lower
plasma protein binding	average/lower	higher
CNS penetration***	average/lower	lower
brain tissue binding	lower	higher
P-gp efflux	average	average
in-vivo clearance	average	average
hERG Inhibition	lower	average/lower
P450 inhibition****	lower 1A2, 2C9, 2C19, 2D6 & 3A4 inhibition	lower 1A2, 2C19 & 3A4 inhibition
P450 inhibition****		average 2C9, 2D6 inhibition

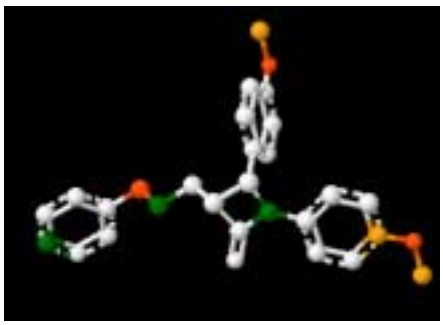
M. P. Gleeson, J. Med. Chem. 51, 817-834 (2008).

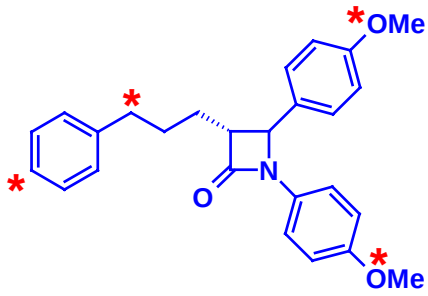
Metabolism Prediction Tools

	SOM (sites of metabolism)	Metabolite Structure	Enzyme Structure & Function	CYP P450 Inhibition	CYP P450 Induction
Scope:	Identification of atom positions with metabolic liability	Prediction of the chemical structure of potential metabolites	Fundamental insight on enzyme function and SARs	Elucidation of SAR of CYP substrates/ inhibitors	Rationalization of biological activity on inducing targets: AhR, CAR, PXR
Methods:	Reactivity Data mining Shape MIFs Docking	Expert systems Data mining Reaction modeling	Structure Reactivity MD: QM/MM Mechanisms	QSAR & ML 3D-QSAR Pharmacophores Docking MD	QSAR & ML 3D-QSAR Pharmacophores Docking

MetaSite, Metaprint2D, QMBO, CypScore, Meteor, META, Times,.....
J. Kirchmair et al., J. Chem. Inf. Model. 52, 617-648 (2012)

Prediction of Drug Metabolism



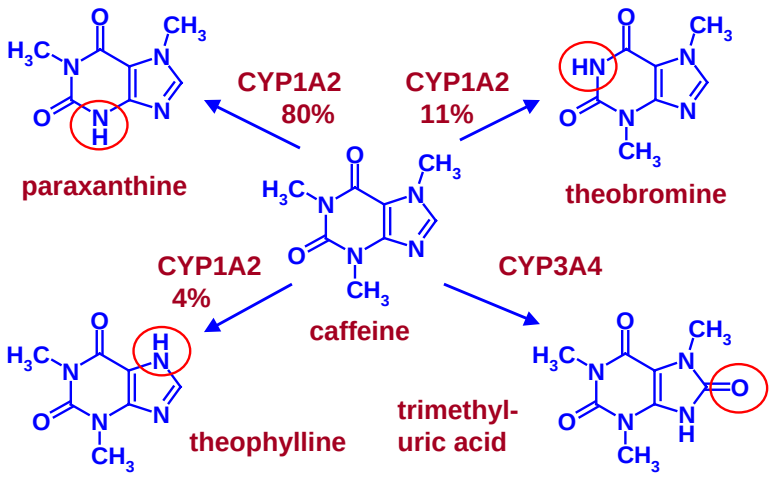


M. van Heek et al., J. Pharmacol. Exp. Ther. 283, 157-163 (1997)

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

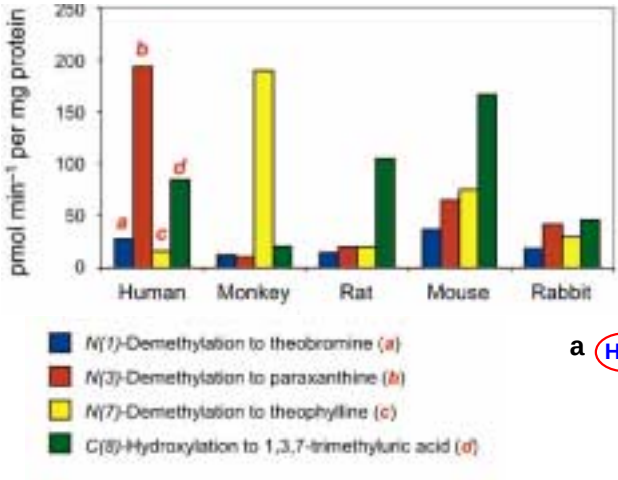
S. Boyer et al., J. Chem. Inf. Model. 47, 583-590 (2007)

Cytochrome P450 Oxidation of Caffeine

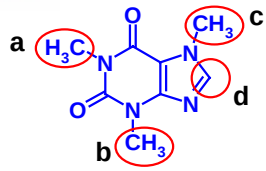


B. Testa and S. D. Krämer, The Biochemistry of Drug Metabolism. Principles, Redox Reactions, Hydrolyses, VHCA & Wiley-VCH, Zürich, 2008

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)

Alle Dinge ſind Gift
und nichts ohn Gift;
allein die Doſis macht,
daß ein Ding kein Gift iſt.

„All things are poison and nothing without poison; only the dose determines, whether a thing be no poison“

Salt, Fat, Alcohol ...
Aspirin, Corticoids ...
Phenacetin, Phenphen, Cerivastatin ...

Toxicity Prediction Tools

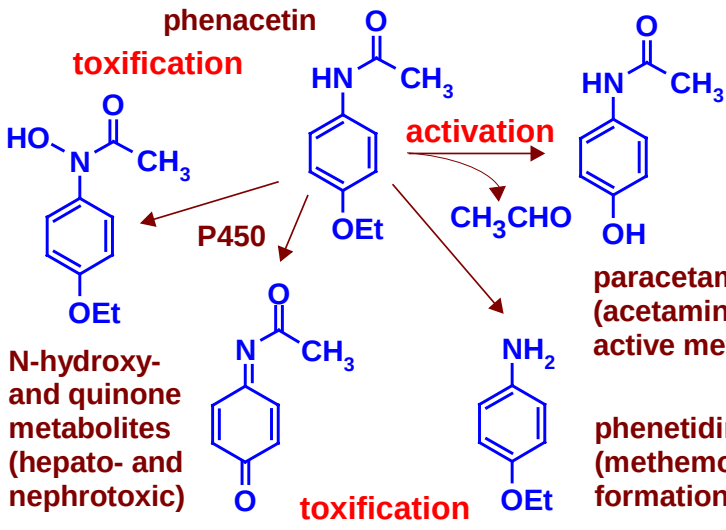
Expert Systems

DEREK: Expert system for the prediction of toxicity
OncoLogic: Rule-based expert system for carcinogenicity prediction
HazardExpert Pro: Rule-based system to estimate toxic symptoms
ToxTree: Places chemicals into categories and predicts toxic effects (open source)

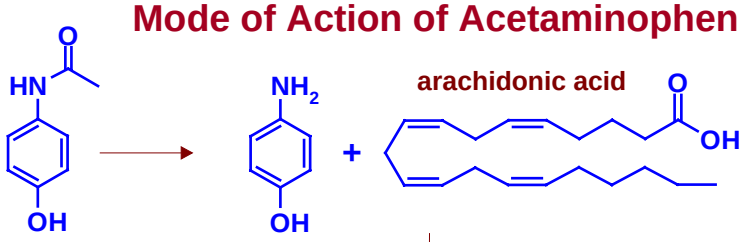
Data Driven Systems

Iazar: Open source database for the prediction of chemical toxicity
MC4PC: Structure-Activity Relationship automated expert system
PASS: predicts 900 pharmacological effects, moa's, mutagenicity, carcinogenicity, teratogenicity, and embryotoxicity
TOPKAT: Quantitative Structure Toxicity Relationship models

Phenacetin, a Pro-Prodrug



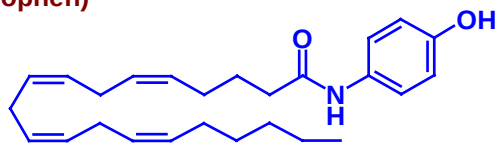
Mode of Action of Acetaminophen



paracetamol (acetaminophen)

p-amino-phenol

fatty acid amido hydrolase (FAAH)



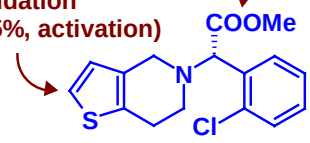
E. D. Högestätt et al., J. Biol. Chem. 280, 31405-31412 (2005)

N-arachidonoyl phenolamine, a potent TRPV1 (transient receptor potential vanilloid 1, vanilloid receptor) agonist, pEC_{50} = 7.80 (about 16 nM), binds also to the cannabinoid CB_1 receptor and inhibits cellular anandamide uptake.

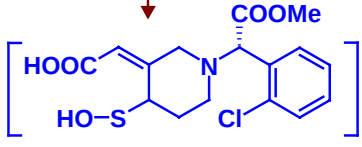
Clopidogrel, Mode of Action

hydrolysis (85%, deactivation)

oxidation (15%, activation)



CYP 3A4



unstable, highly reactive

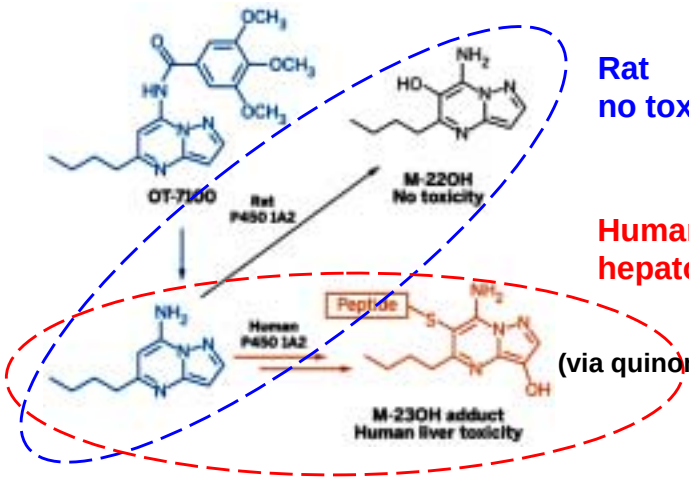
Plavix®, Iscover® (Sanofi-Aventis, BMS) worldwide net sales 2005: about 6 bio US-$

active metabolite, irreversibly reacts in the liver with platelet $P2Y_{12}$ (ADP) receptors

activation by CYP2C19 (absent in 30% Caucasians)

J.-M. Pereillo et al., Drug Metab. Dispos. 30, 1288-1295 (2002);
E. J. Topol, Nature Rev. Drug Discov. 8, 259 (2009);
cf. P. M. Dansette et al., Chem. Res. Toxicol. 22, 369-373 (2009)

Biological Activities of Metabolites



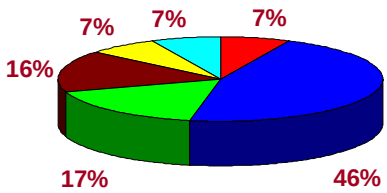
Rat
no toxicity

Human
hepatotoxicity

(via quinone imine)

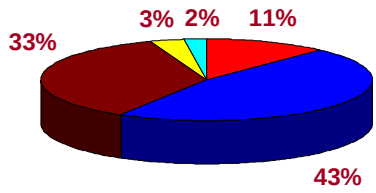
S. Kuribayashi et al., Chem Res. Toxicol. 22, 323-331 (2009);
cf. Chem. &Eng. News, August 31, 2009, p. 27

The Ultimate (Unavoidable) Bottlenecks?



- 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

Reasons for failure in clinical development, 1992-2002 (n = 73) (reasons for market withdrawal, n = 17: toxicity 93%, efficacy 7%) D. Schuster et al., Curr. Pharm. Design 11, 3545-3559 (2005)

Yes, We Can? No, We Can't

What we can	Estimation of lipophilicity
	Prediction of 3D structure/s
	3D pharmacophore generation
	3D pharmacophore searches
	Prediction of plausible metabolites
What we can't	Prediction of crystal lattices
	Prediction of melting points
	Prediction of (difficult) pK_a values
Where we fail	Prediction of solubility (pK_a, mp)
	ADME prediction (log S, transporters)
	Affinity prediction (scoring functions)
	Prediction of biological activities
	Prediction of selectivity and toxicity