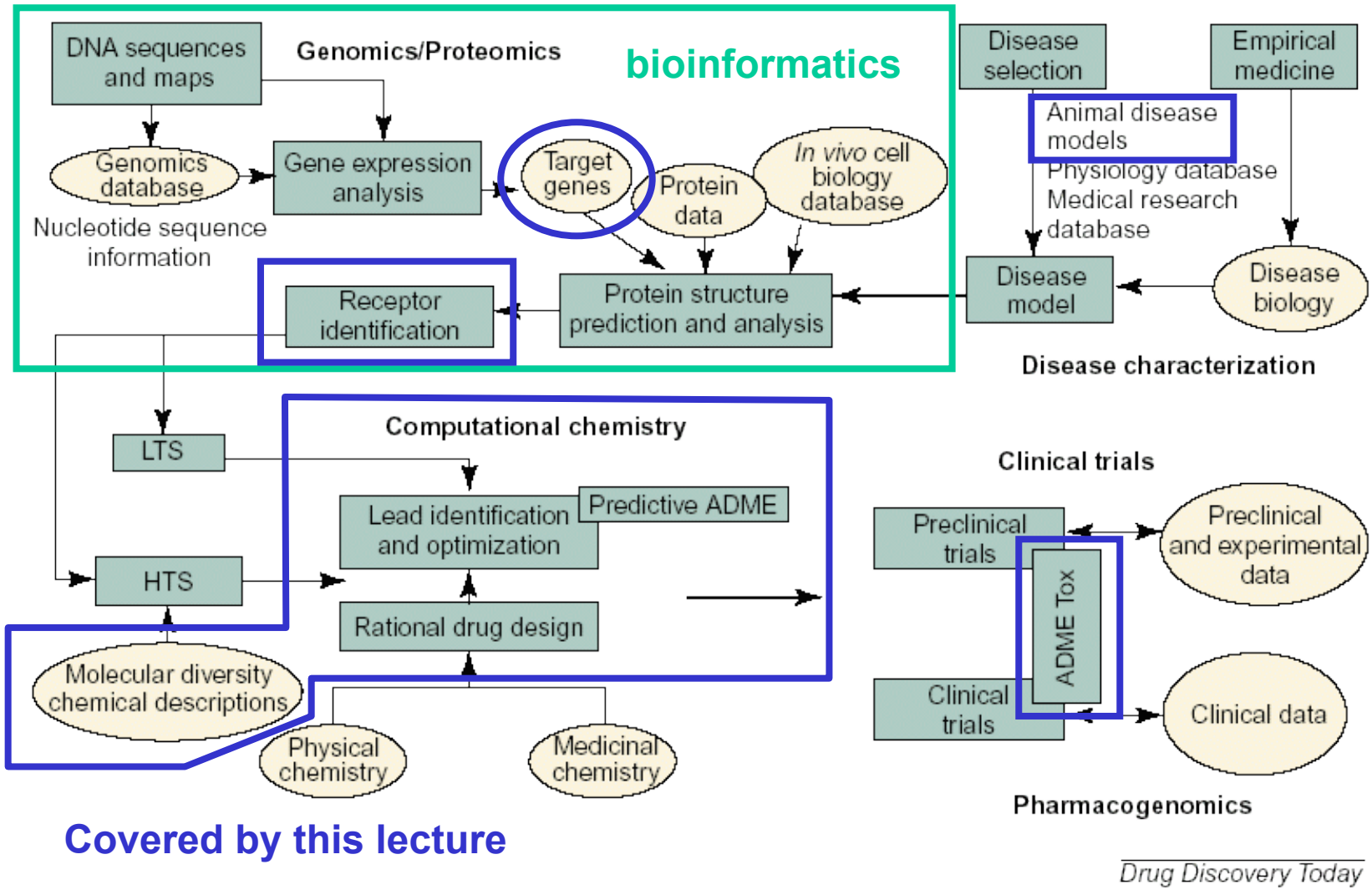
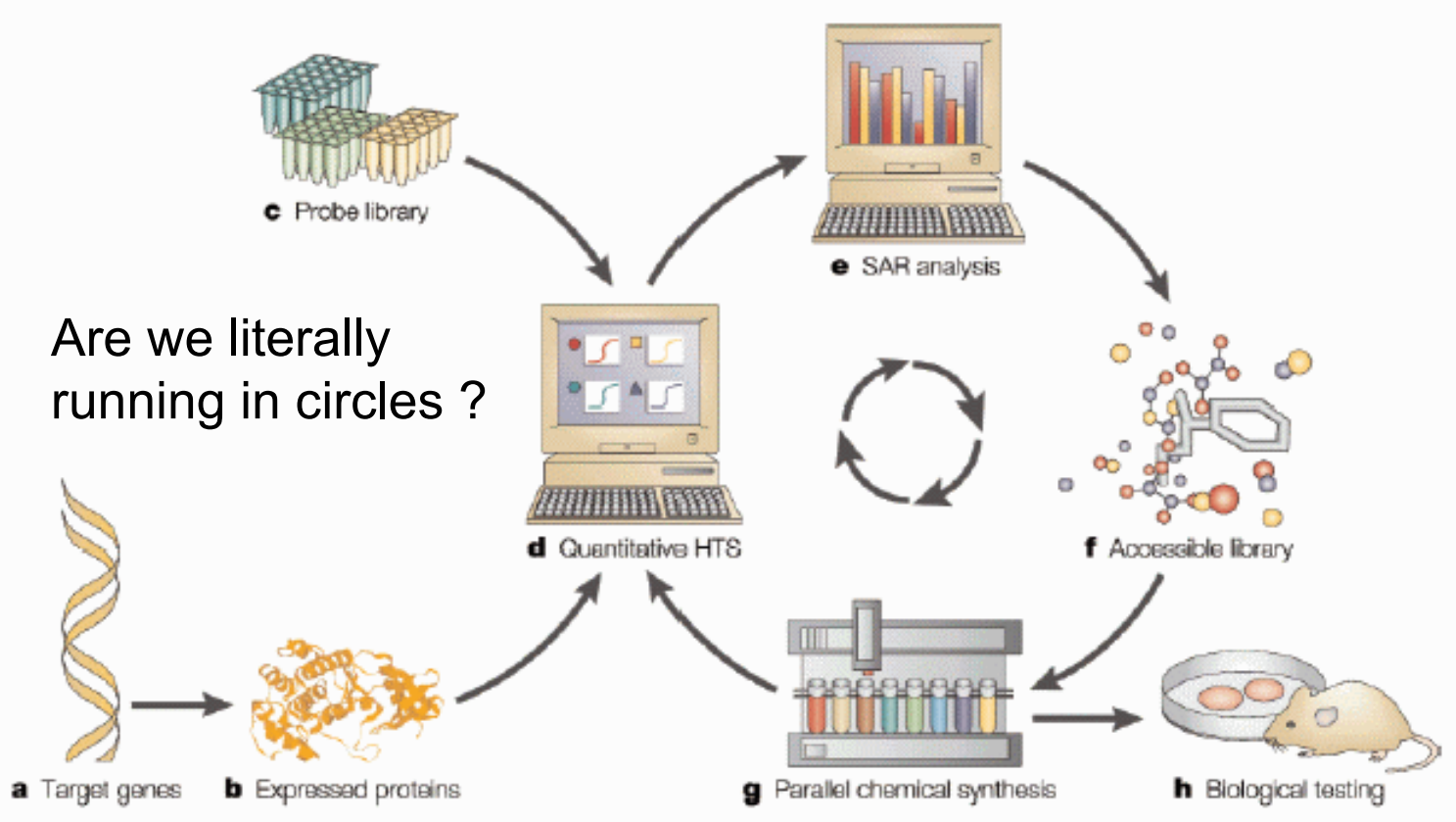
Current Trends



Cycle of optimization in the drug discovery pipeline



Source: D.K. Agrafiotis et al. *Nature.Rev.Drug.Discov.* **1** (2002) 337.

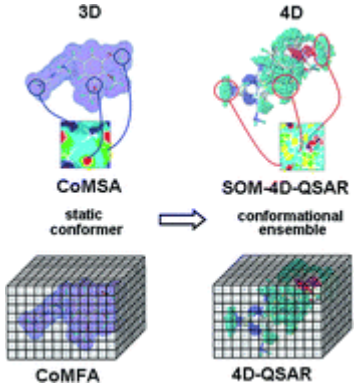
4D and 5D QSAR

3D QSAR: Information from the 3D structure is used

→ 3D descriptors, pharmacophore models

4D and 5D QSAR: multiple conformations

→ use of multiple docking results for one compound

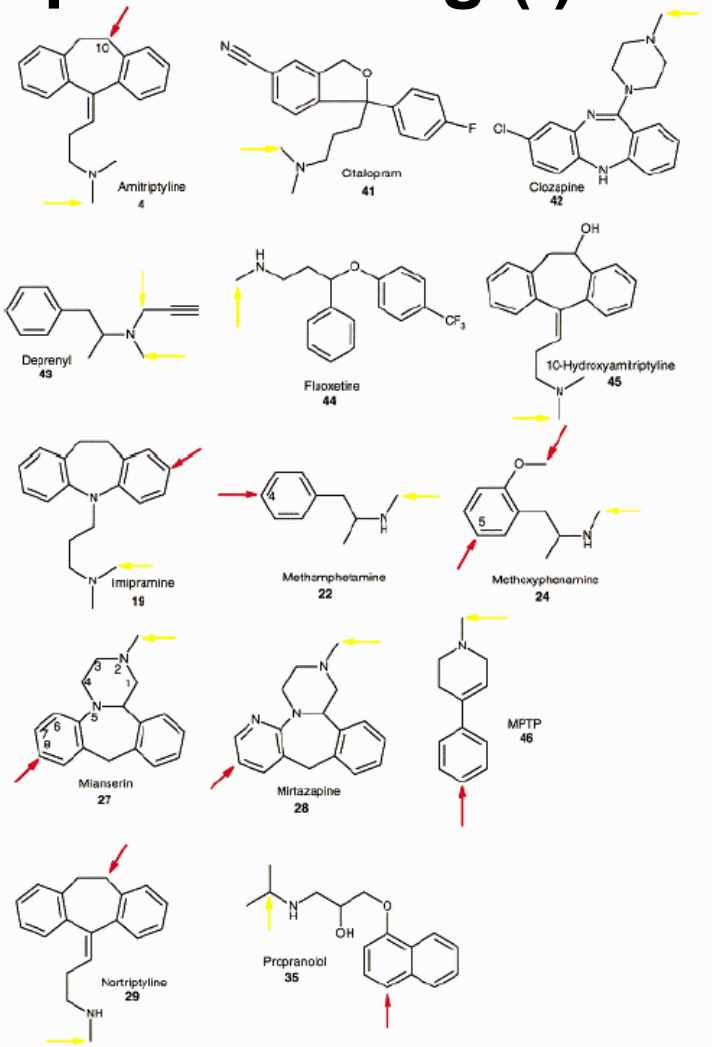


Lit: M. Dobler et al. *Helvetica Chim. Acta* **86** (2003) 1554.
A. Bak et al. *RCS Adv.* **6** (2016) 76183.

multiple conformations upon docking (I)

The binding pocket of many cytochrome P450 enzymes (esp. CYP3A4 and CYP2D6) is large enough to accomodate the same substrate in different orientations, which leads to different products.

The reactivity of a certain spot of the molecule can be estimated by quantum chemical calculations.

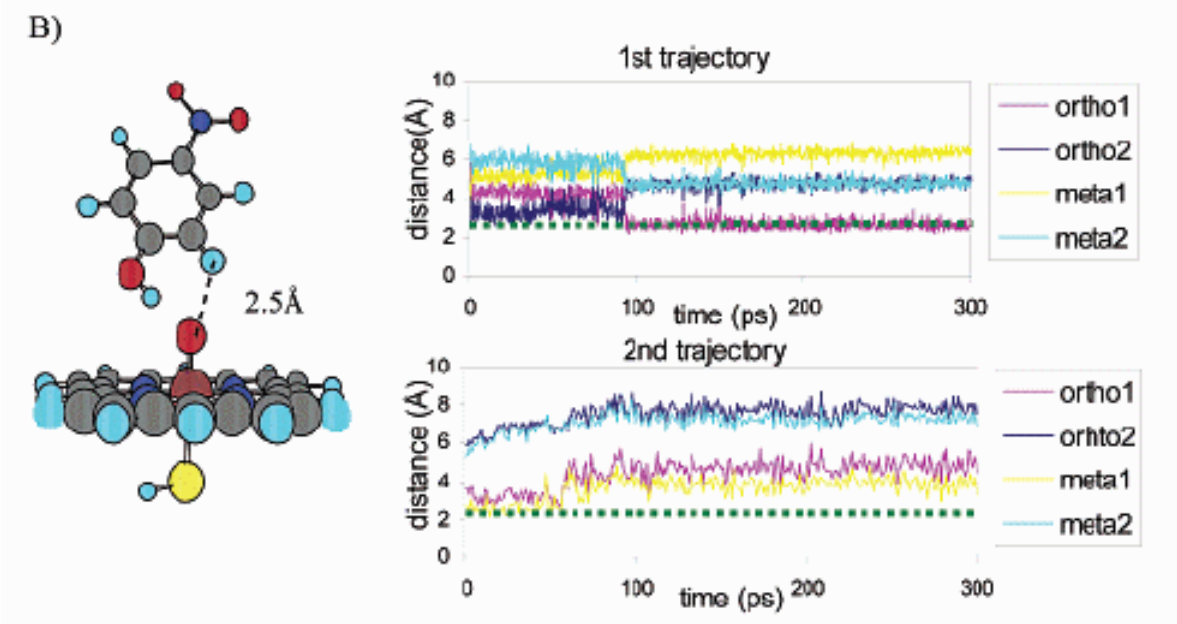


Lit: M. de Groot et al. *J.Med.Chem.* **42** (1999) 4062
S.B. Singh et al. *J.Med.Chem.* **46** (2003) 1330

multiple conformations upon docking (II)

Besides information about the reactivity at a certain spot of the molecule, also the propability of the according binding position in the enzyme is required.
Can be obtained from statistical analysis (clustering) of

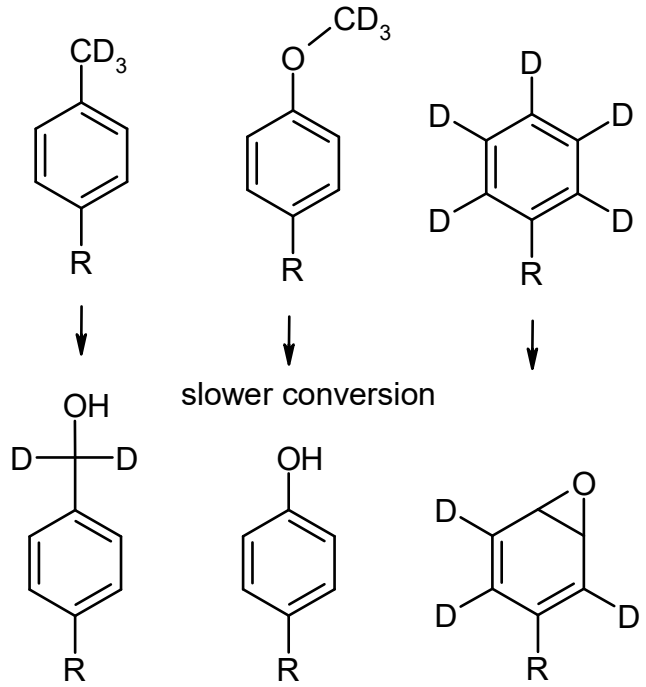
- a large number of docking results, or by
- molecular dynamics simulations

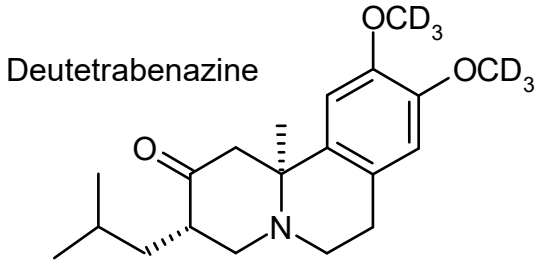


Lit: Park & Harris *J.Med.Chem.* **46** (2003) 1645

Increased Metabolic Stability by Deuteration

Replacing hydrogen(s) by deuterium (D) causes bonds to become more stable (primary deuterium kinetic effect) leading to reduced metabolic conversion (esp. by CYP P450 enzymes). This effect can thus be exploited to avoid unwanted or toxic metabolites.





Lit: Harbeson & Tung *Med.Chem.News* **24** (2014) 8-22.

Drug / Non-Drug Separation (1)

Is it possible to predict the potential suitability of a compound from typical properties of drugs?

approaches:

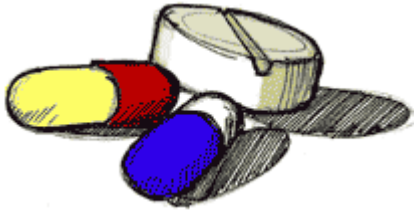
Reckognition of typical properties in data bases that (almost) exclusively contain drugs

For example:

- World Drug Index (WDI)
- Comprehensive Medicinal Chemistry (CMC)
- MACCS-II Drug Report (MDDR)
- Merck Index (filtering required)
- Drugbank
- ChEMBL (filtering required, e.g. clinical phase)

Drug / Non-Drug Separation (2)

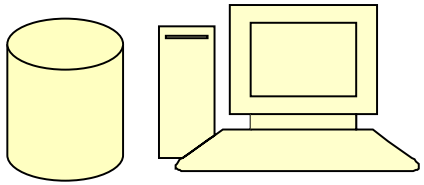
Previous data base analyses:



1997 Christopher Lipinski‘s rule of 5 (Pfizer)

Orally administered drugs typically have

molecular weight < 500
ClogP < 5
less than 5 hydrogen-bond donors (O-H, N-H)
less than 10 hydrogen-bond acceptors (N, O, S)



2000 Tudor Oprea (AstraZeneca)

Typical drugs (70% of all) have

less than 3 hydrogen-bond donors
between 2 and 9 hydrogen-bond acceptors
between 2 and 9 rotatable bonds
between 1 and 4 rings

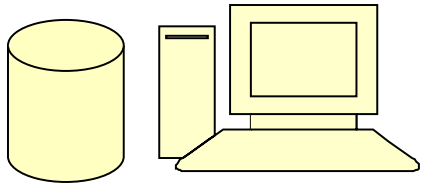
Lipinski‘s rule of 5 refers to oral bioavailability but not neccessarily drug-likeness !

Drug / Non-Drug Separation (3)

1999 Ghose, Viswanadhan & Wendoloski

Analysis of the Comprehensive Medicinal Chemistry database:

80% of all drugs have



160 < molecular weight < 480
–0.4 < logP < 5.6
20 < number of atoms < 70
40 < molar refractivity < 130

The preferred range covering 50% of all drugs shows

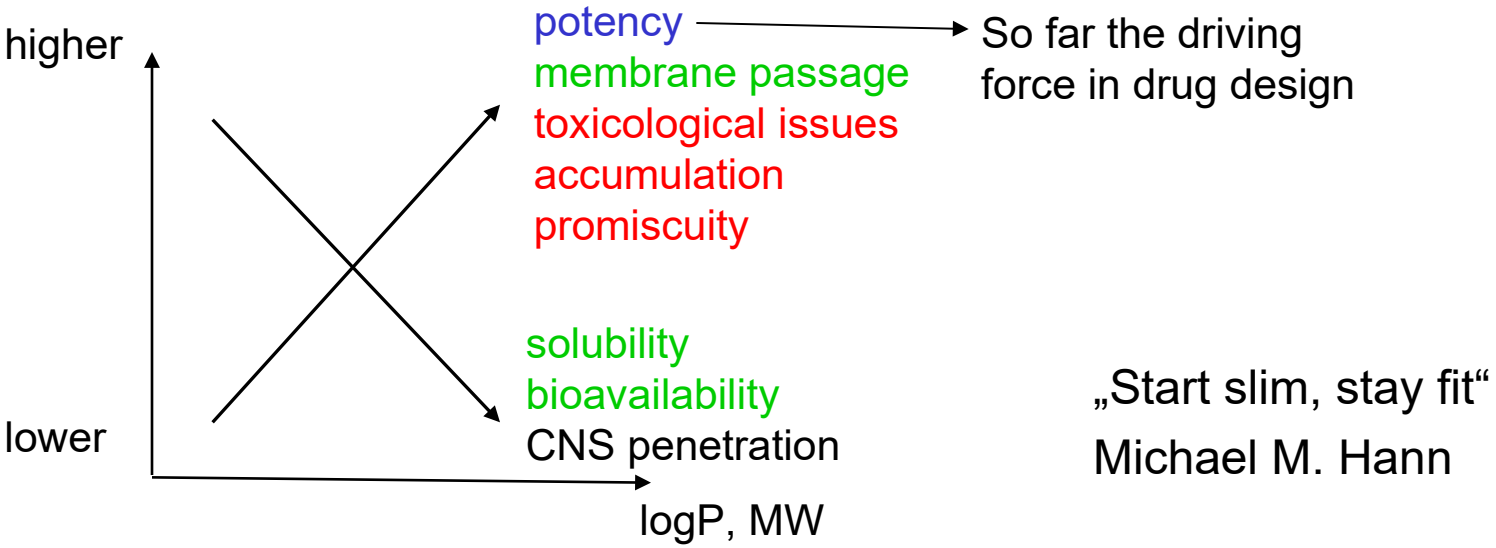
230 < molecular weight < 390
1.3 < logP < 4.1
30 < number of atoms < 55
70 < molar refractivity < 110

Lit: A. Ghose et al. *J.Comb.Chem.* **1** (1999) 55-68.

Drug / Non-Drug Separation (4)

Even tighter restrictions required to avoid adverse effects?

Molecular weight < 400 and ClogP < 4 (GSK 4/400 rule)



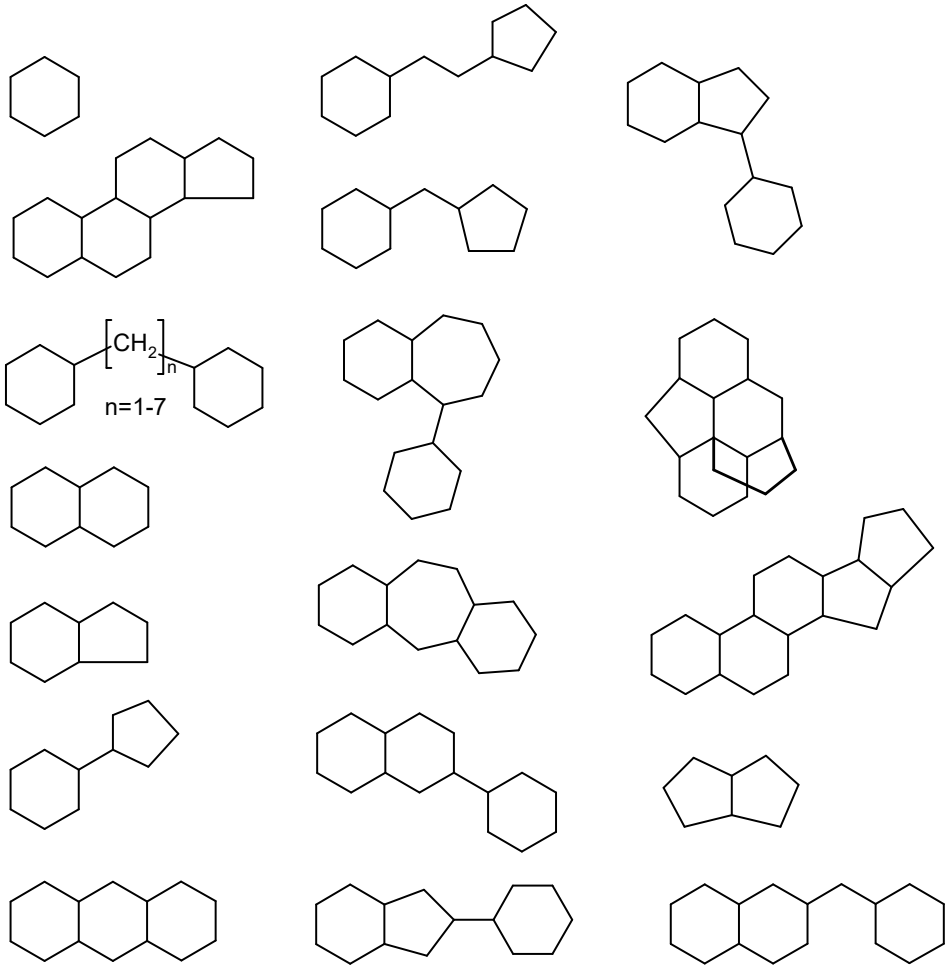
„Start slim, stay fit“
Michael M. Hann

Find smallest crucial parts of molecules → fragments

Lit: M.M. Hann „Molecular Obesity, Potency and Other Addictions in Drug Discovery“ *Med.Chem.Commun.* **2** (2011) 349-355.

Drug / Non-Drug Separation (5)

The most common (ring) fragments of drugs



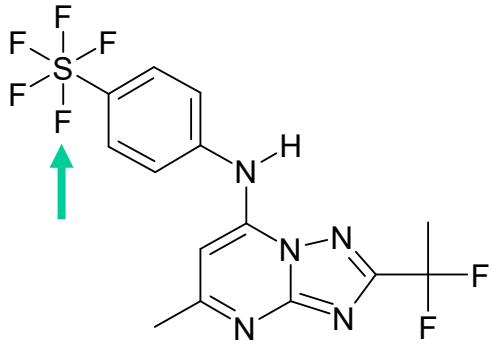
Lit: G.W.Bemis & M.M.Murcko, *J.Med.Chem.* **39** (1996) 2887

Drug / Non-Drug Separation (6)

Rare appearance of certain fragment or side chains does not necessarily mean that it is unsuitable or negligible.

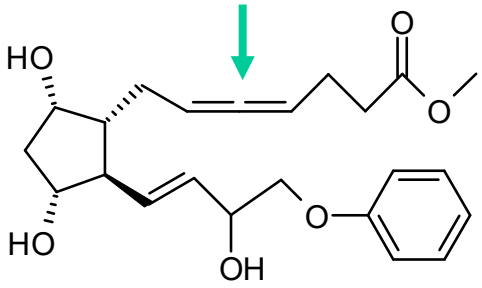
Such fragments can rather

- be difficult to synthesize, or
- be newly introduced, or
- possess unsuitable properties



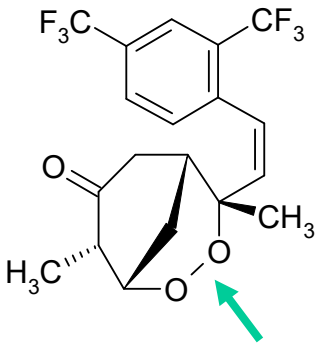
DSM265

rare SF_5 fragment



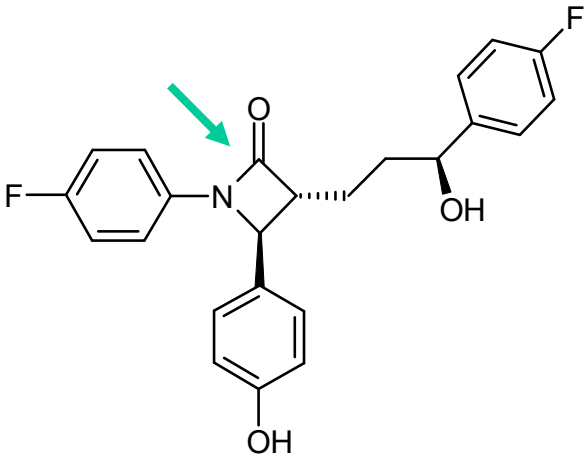
fenprostalene

rare ketene fragment



arteflene

difficult dioxygen bridge

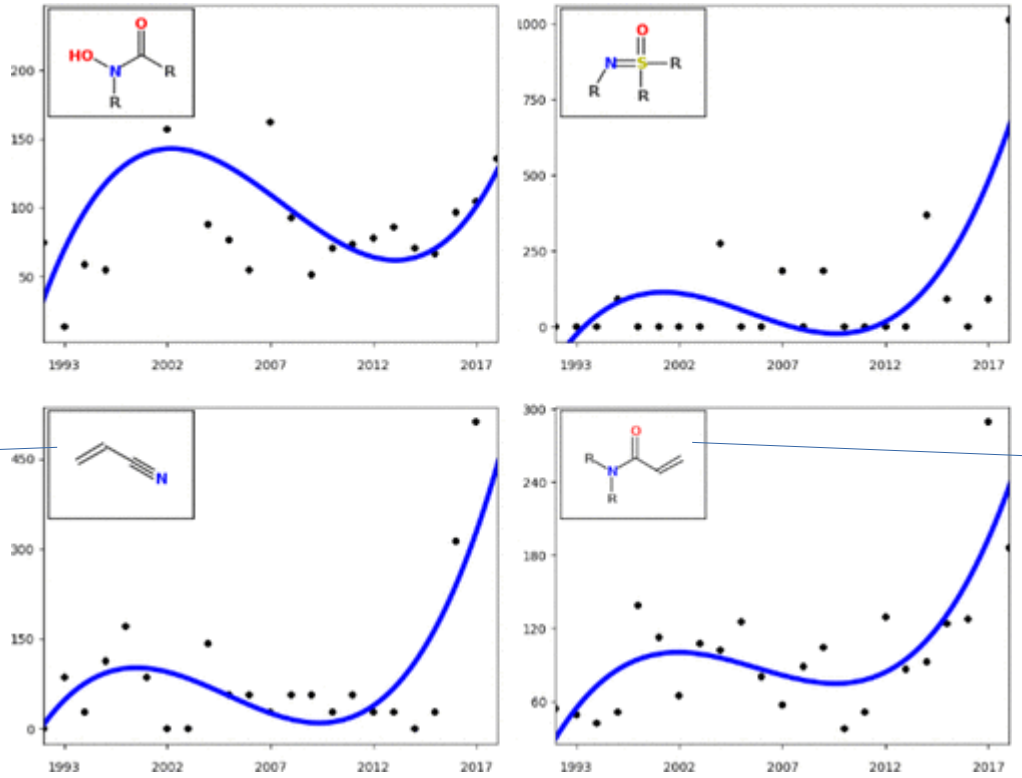


ezetibime

new subsitution pattern of lactam

Use of fragments over time

While some functional groups seem to indispensable, a number of new and increasingly used fragments have been identified from the ChEMBL database:



e.g. rilzabrutinib, atuzabrutinib

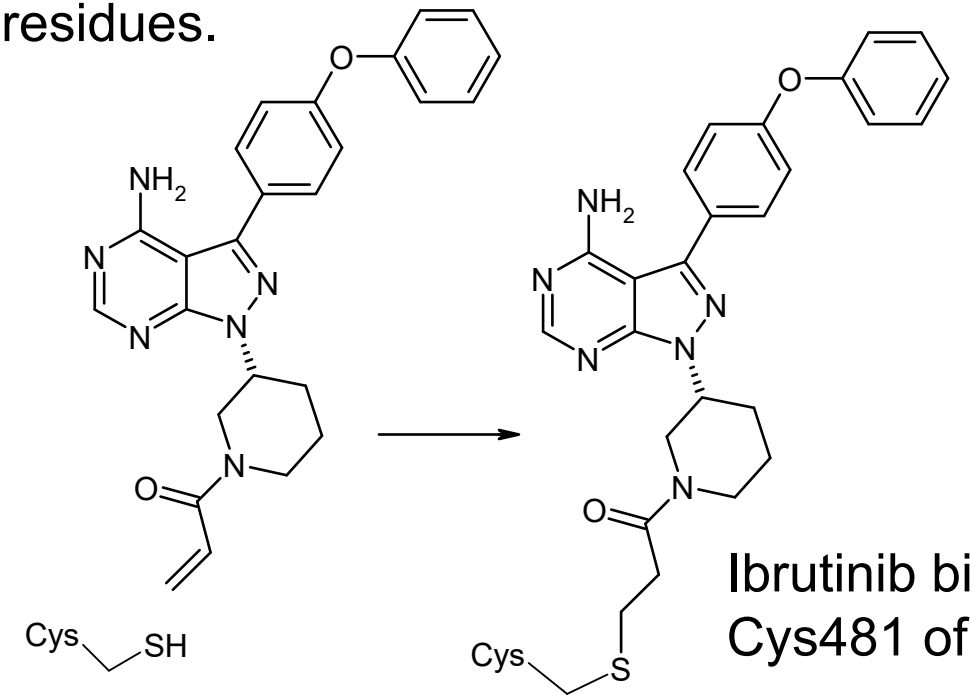
reactive fragment as in ibrutinib

Lit. P. Ertl et al. *J. Med. Chem.* **63** (2020) 8408.

Reactive Groups for Irreversible Inhibitors

Covalent binding of inhibitors to enzymes renders these to become unfunctional.

Typical reactive groups are electrophilic, such as acrylamides and α,β-unsaturated carbonyls that bind to noncatalytic cysteine residues.

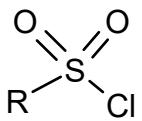


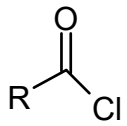
Ibrutinib binds to the side chain of Cys481 of Bruton‘s Tyrosine Kinase.

Lit: M.Gehringer & S.A.Laufer *Chem. Chem. Biol.* **62** (2019) 5673.

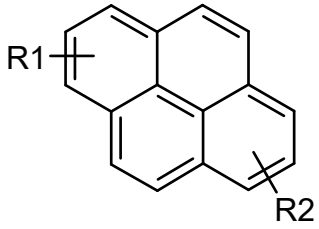
Drug / Non-Drug Separation (7)

Examples of groups that possess wellknown unsuitable properties

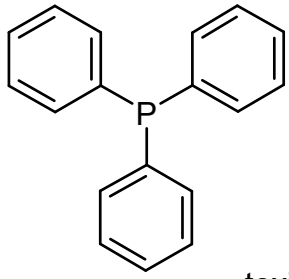




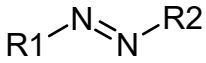
reactive



mutagenic



toxic



cancerogen

Such groups should not be present in clinical drugs, but may be important during synthesis

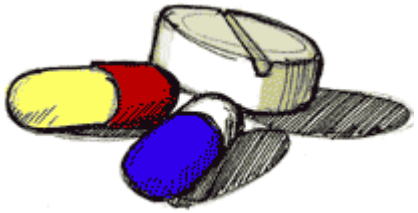
Lit: D.R.Flower, *J.Mol.Graph.Model.* **16** (1998) 239.

M.Hann et al. *J.Chem.Inf.Comput.Sci.* **39** (1999) 897.

Drug / Non-Drug Separation (8)

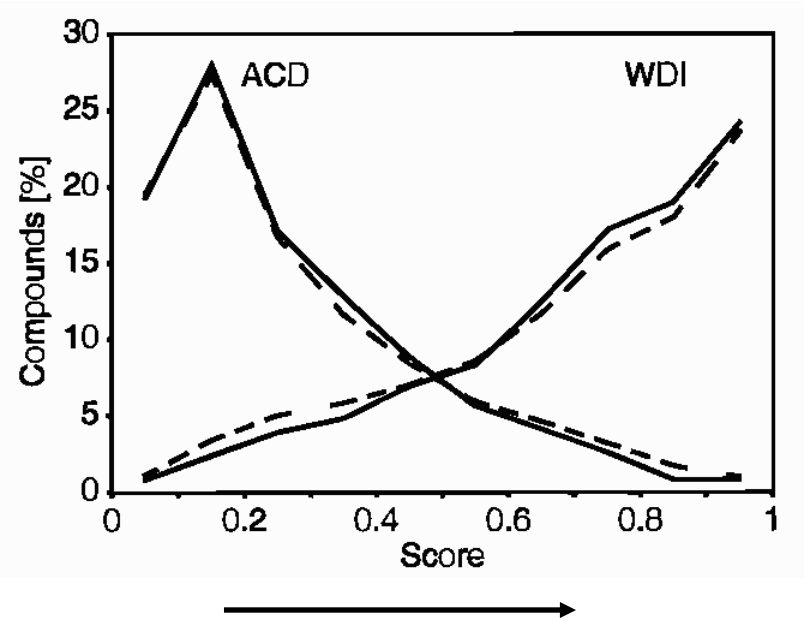
Further approach:

Comparison of compounds in a data base containing solely drugs (World Drug Index) to substances from a data base that predominately consists of non-pharmaceutical compounds
(Available Chemical Directory).



Drug / Non-Drug Separation (9)

Classification of compounds according to their atom types using a neural net



Training set 10000 compounds

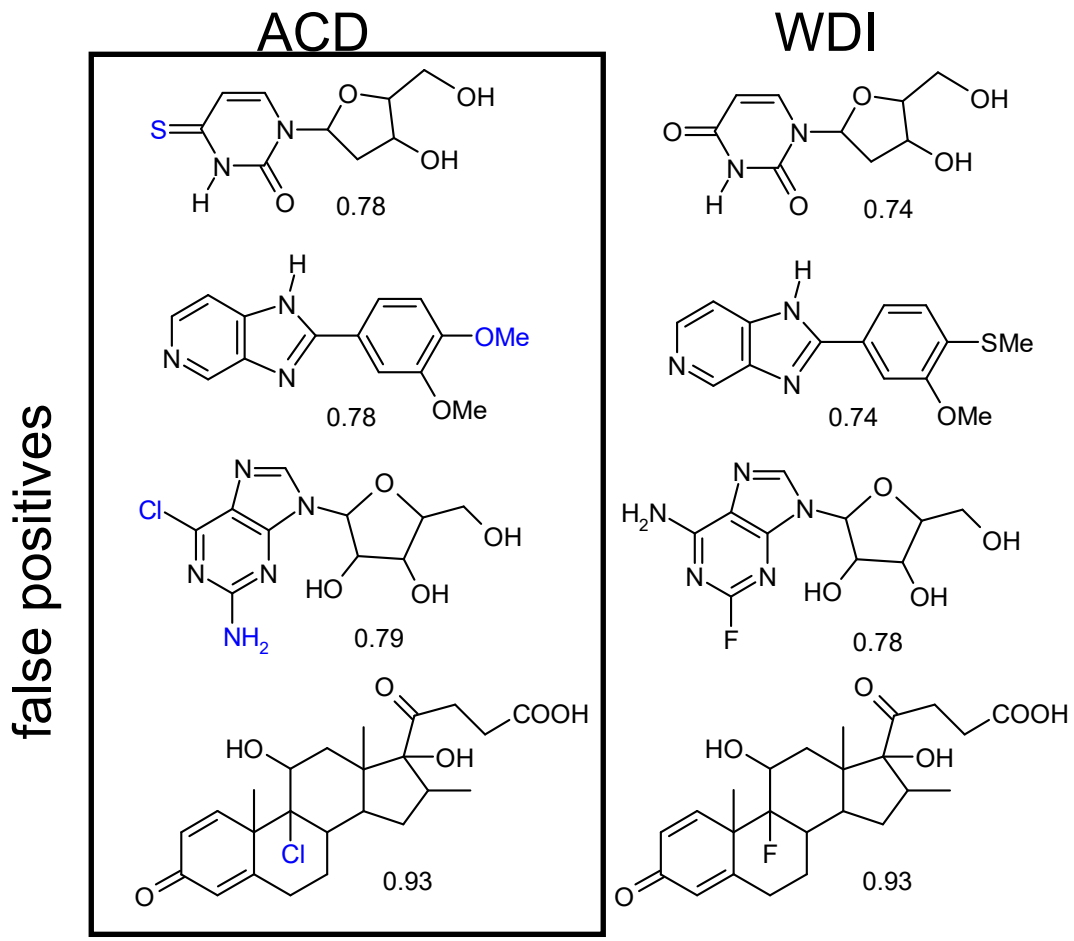
Test set 207747 compounds

Increasing drug-likeness,
but where is a reasonable threshold of the score?

Lit: J. Sadowski & H. Kubinyi *J.Med.Chem.* **41** (1998) 3325.

Drug / Non-Drug Separation (10)

Compounds for which a high drug-likeness score was predicted:



Lit: J. Sadowski & H. Kubinyi *J.Med.Chem.* **41** (1998) 3325.

Drug / Non-Drug Separation (11)

Classification of compounds using their ISIS *fingerprint* (set of 73 descriptors which indicate the presence of structural and topological features, and encode chemical properties)

$\rightarrow$ Allow comparison of the compounds by their similarity using the Tanimoto index.

These 73 binary descriptors were used as input layer of a neural net, which was trained with compounds from drug data bases and non-drugs from the ACD.

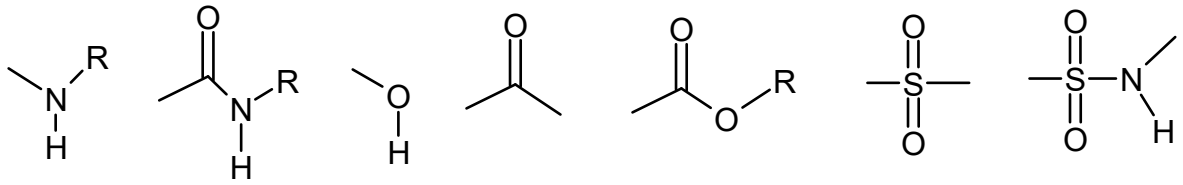
result: about 80% of all compounds were classified correctly.

Lit: Ajay et al. *J.Med.Chem.* **41** (1998) 3314.

Drug / Non-Drug Separation (12)

Classification of compounds according to atom types that represent so-called *pharmacophoric points*:

Among these functional groups are



preselection: a compound is potentially drug-like, if it contains at least one ring and between 2 and 7 of such functional groups.

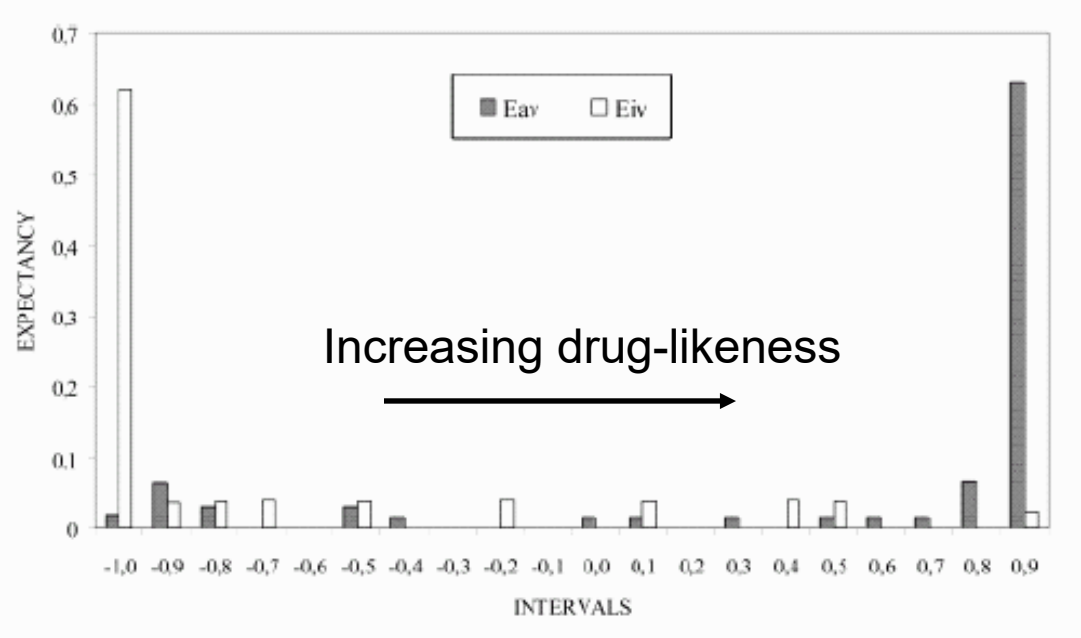
The atoms of the molecule are classified according to the affiliation to one of these *pharamacophoric points* and used as input layer of a neural net.

Here again compounds of the ACD were compared to drug data bases.

Lit: J.Muegge et al. *J.Med.Chem.* **44** (2001) 1841.

Drug / Non-Drug Separation (13)

Classification of compounds according to topological descriptors using a neural net.



680 compounds of the Merck Index, of which about 76 % were classified correct.

Lit: M.Murcia-Soler et al. *J.Chem.Inf.Comput.Sci.* **43** (2003) 1688.

Drug / Non-Drug Separation (14)

Classification of compounds using a *decision tree*. Used were atom types that represent certain functional groups.

Advantages of a decision tree compared to a neural net:

The criteria for classification at each branching point can be traced easily and a corresponding error can be assigned.

results:

- ¾ of all compounds can be assigned based on the presence of 6 chemical groups.
- Non-drugs typically contain not enough of these functional groups

Lit: M.Wagner et al. *J.Chem.Inf.Comput.Sci.* **40** (2001) 280.

Drug / Non-Drug Separation (15)

Preliminary resume:

Neither the presence of atom types, nor that of (sub-) structure fragments or functional groups, allows to classify a substance precisely as drug-like (> 95% accuracy)

Seemingly an (even) larger variety of descriptors, e.g. those that account for electronic properties are required.

→ use of quantum chemical descriptors?

Lit: N.Schneider et al. *J.Chem.Inf.Model.* **48** (2008) 613.
M.C.Hutter *Curr.Med.Chem.* **16** (2009) 189.

Drug / Non-Drug Separation (16)

Principal component analysis (PCA) of 26 descriptors of compounds from the Maybridge data base yielded the numerical value of the 3rd principal component as most significant separation criteria.

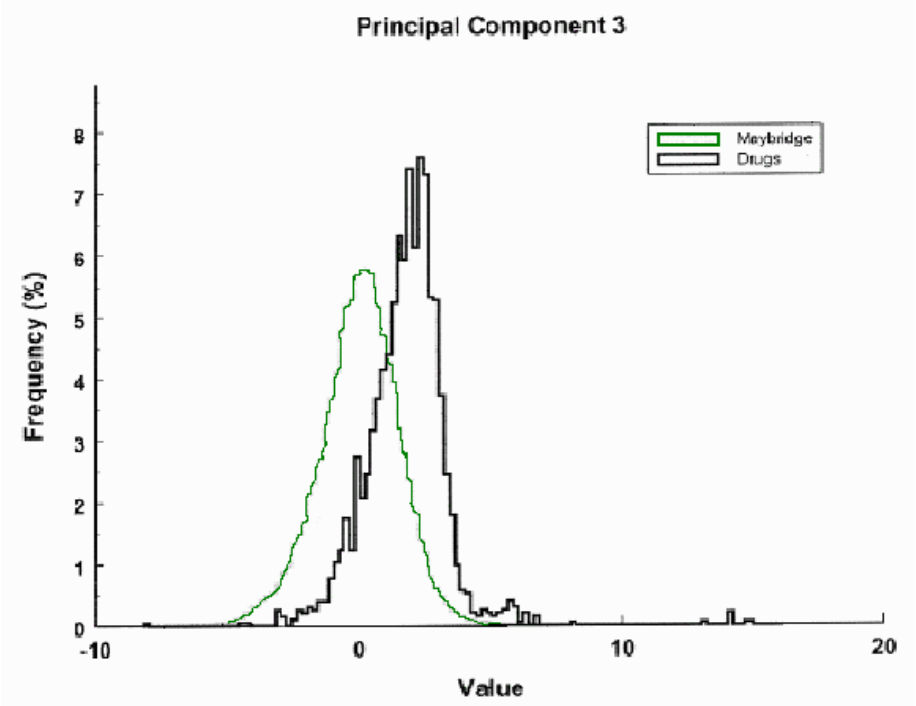
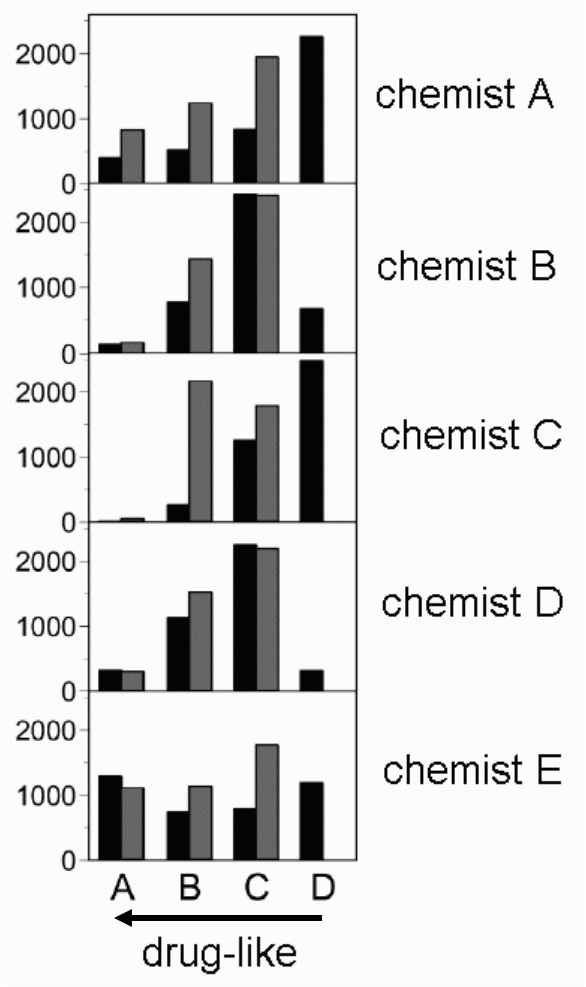


Figure 3. Frequency histogram for PC number 3 (MEP-) for drugs and nondrugs.

Lit: M.Brüstle et al. *J.Med.Chem.* **45** (2002) 3345

Drug / Non-Drug Separation (17)

Classification of compounds based on chemical intuition



3980 compounds were classified by 5 chemists according to their **drug-likeness** and the according **synthesic efford**

Table 2. Correlation Coefficient between Scores Assigned by Each Chemist on Drug-Likeness

	chemist A	chemist B	chemist C	chemist D	chemist E
chemist A	1.00	0.55	0.63	0.57	0.58
chemist B		1.00	0.51	0.50	0.50
chemist C			1.00	0.52	0.54
chemist D				1.00	0.54
chemist E					1.00

Table 3. Correlation Coefficient between Scores Assigned by Each Chemist on Ease of Synthesis

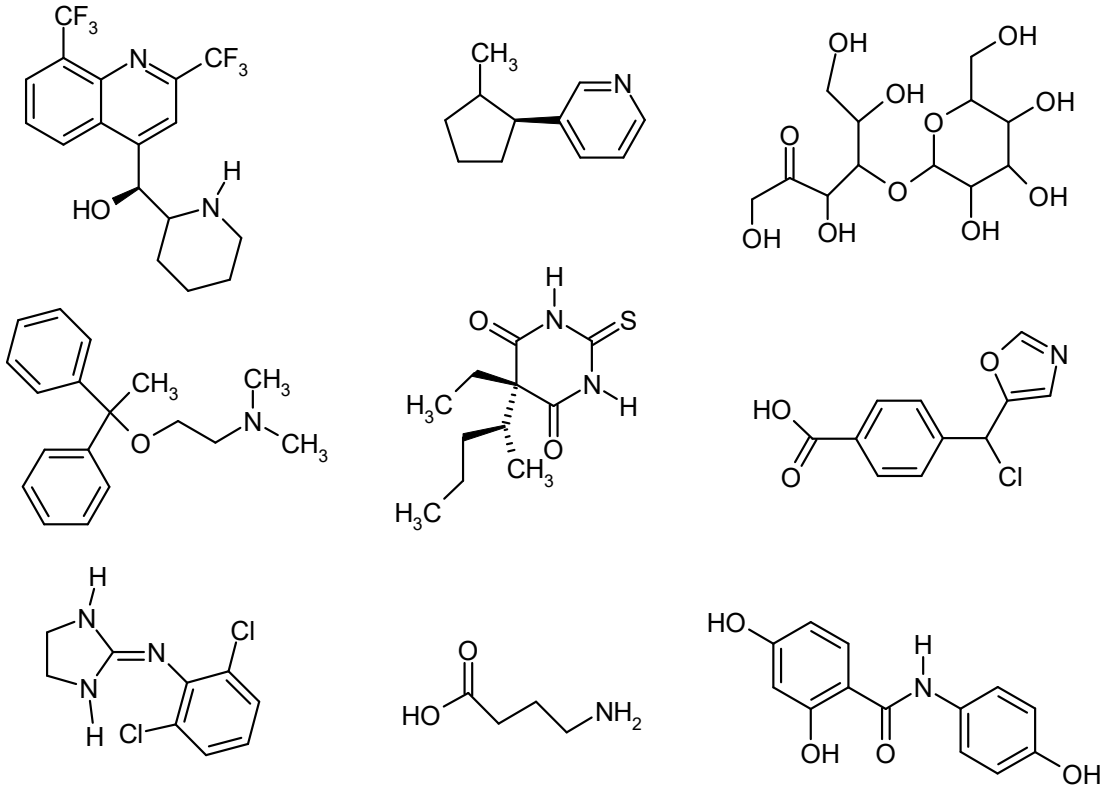
	chemist A	chemist B	chemist C	chemist D	chemist E
chemist A	1.00	0.50	0.40	0.40	0.56
chemist B		1.00	0.42	0.47	0.52
chemist C			1.00	0.40	0.48
chemist D				1.00	0.48
chemist E					1.00

Lit: Y.Takaoka et al. *J.Chem.Inf.Comput.Sci.* **43** (2003) 1269

Drug / Non-Drug Separation (18)

try yourselves !

Classify these compouds into drug or non-drug



Compare your results to that of the property prediction module at

http://www.organic-chemistry.org/prog/peo/index.html

Drug / Non-Drug Separation (19)

Back to the basics:

So far it has been only assumed that there is an unequal feature distribution between drugs and non-drugs.
How can we statistically prove this assumption ?

Idea: Certain combinations of atom types are found with a different frequency among drugs compared to non-drugs

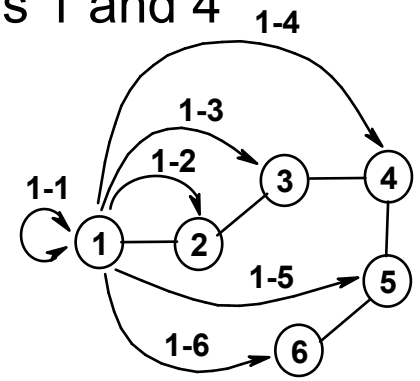
1-1 Interaction: the atom itself

1-2 Interaction: bond between two atoms 1 and 2

1-3 Interaction: angle between atoms 1 and 3

1-4 Interaction: dihedral angle between atoms 1 and 4

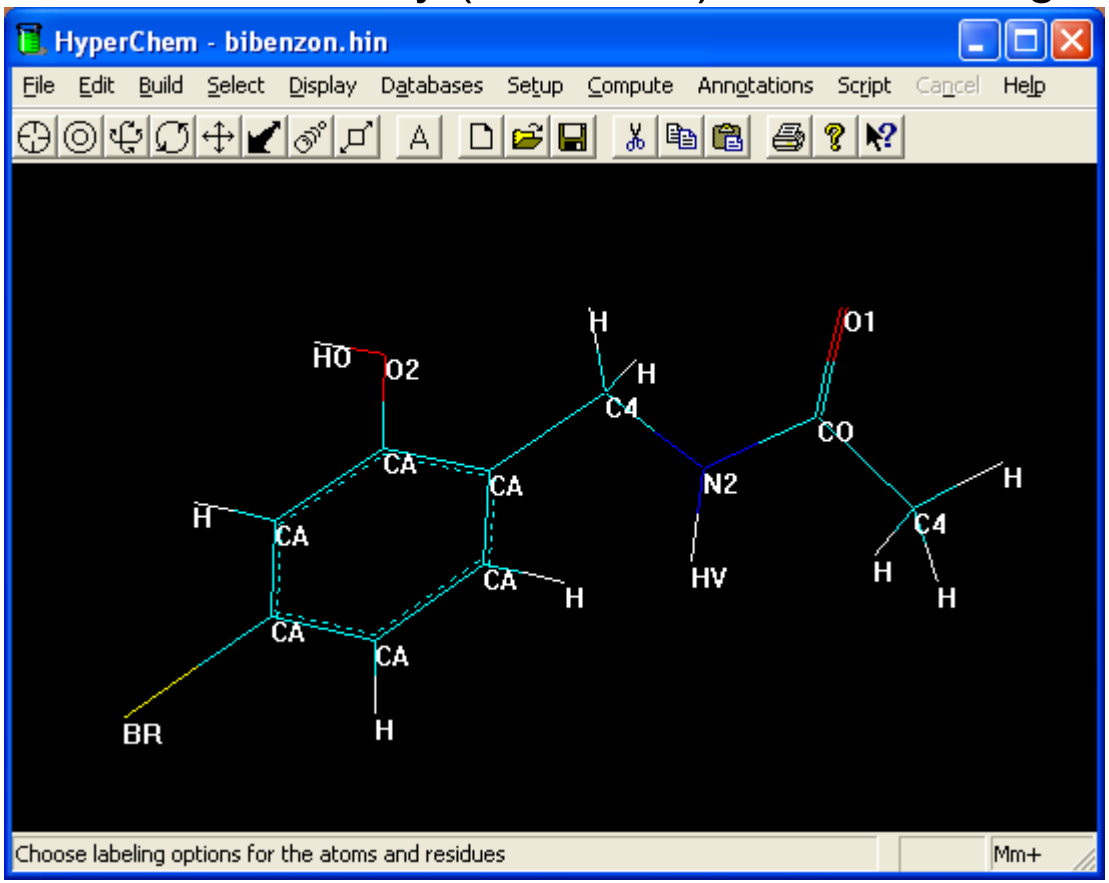
1-5 Interaction: between atoms 1 and 5



Drug / Non-Drug Separation (20)

What atom types and how many should be used ?

Atom types should account for the chemical diversity
Thus, elements only (C, N, O,..) are not enough

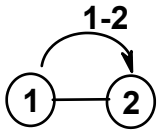


Here, atom types from the MM+ force field are used (total of 47)

Drug / Non-Drug Separation (21)

When is an atom pair combination i-j statistically overrepresented ?

→ If its frequency q_{ij} is higher than that by chance
(= relative probability S‘)



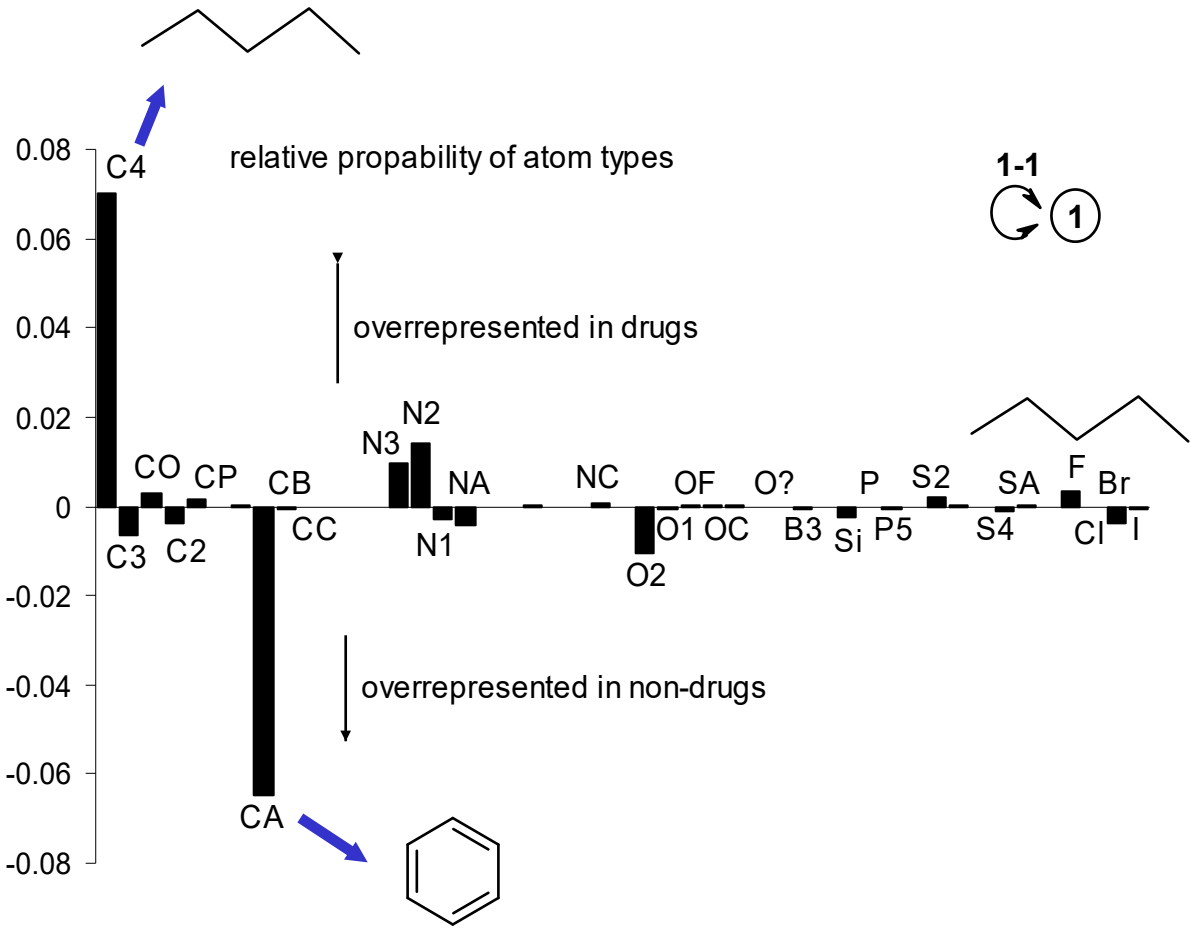
$$S'_{ij} = \frac{q_{ij}}{p_i \cdot p_j}$$

where p_i is the individual frequency of an atom of type i

For better handling we use the logarithmic value
= log odds score

$$S_{ij} = \ln \frac{q_{ij}}{p_i \cdot p_j} \qquad \begin{cases} >0 & \text{overrepresented} \\ <0 & \text{underrepresented} \end{cases}$$

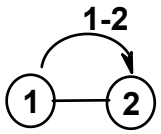
Drug / Non-Drug Separation (22)

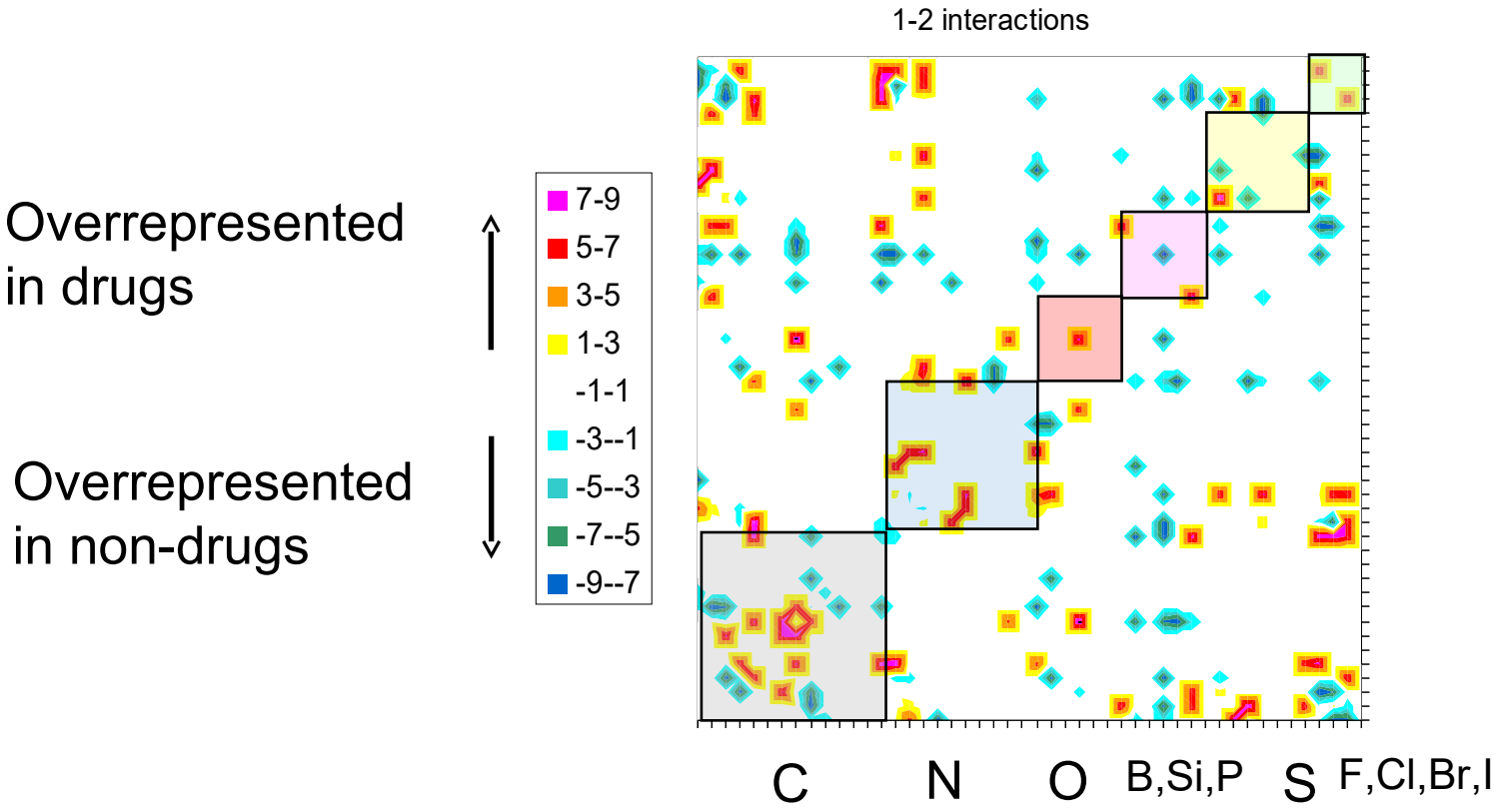


Distribution of atom types (1-1 interaction) alone is not sufficient

Drug / Non-Drug Separation (23)

Visualized using a difference matrix





Similar to amino acid exchange matrix!

Drug / Non-Drug Separation (24)

But how to calculated the drug-*likeliness* from the atom type distribution ?

Simply add up corresponding matrix entries and divide by the number of occuring atom pairs in the molecule:

Drug-likeliness score L

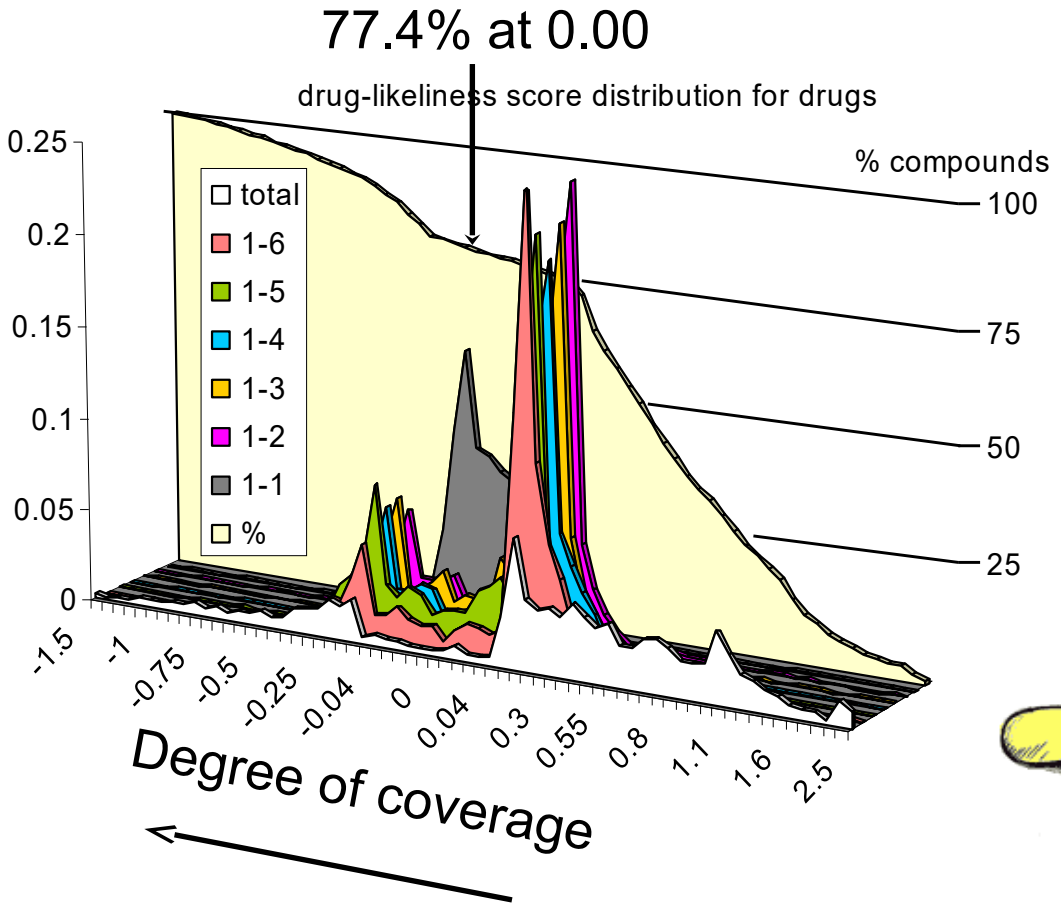
$$L = \sum_{1}^{6} \frac{\sum S_{ij}}{M}$$

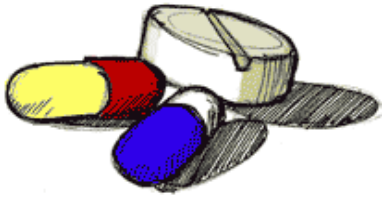
- >0 drug-like
- <0 non-drug-like

Timing:

Less than 5 minutes computing the difference matrices and scores for 4083 compounds

Drug / Non-Drug Separation (25)





Lit: M.C.Hutter *J.Chem.Inf.Model.* **47** (2007) 186-194.

https://github.com/michahutter/druglikeliness → perl code

Drug-likeness and Synthetic Accessability

Quantifying the chemical beauty of drugs (QED score)*

G.R. Bickerton, A.L. Hopkins et al. *Nat.Chem.* **4** (2012) 90-98.

Based on the distribution of molecular properties (MW, logP, hydrogen-bond acceptors/donors, polar surface area, rotatable bonds, unwanted fragments).

Synthetic Accessability (SAscore)*

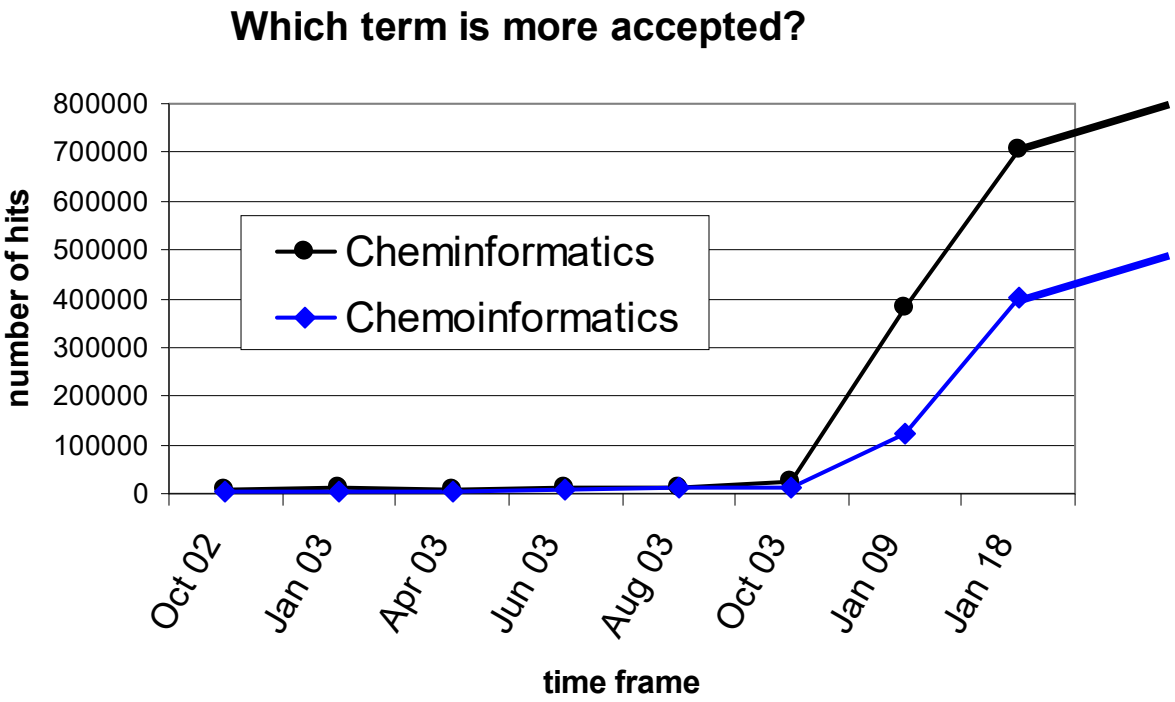
P. Ertl & A. Schuffenhauer *J.Cheminf.* **1** (2009) 8.

"Based on molecular complexity and fragment contributions"

...and heaps of more recent approaches

* Available within RDKit https://www.rdkit.org/docs/

Cheminformatics or **Chemo**informatics ?



Data source: http://www.molinspiration.com

http://www.google.de

Doping (I)

Illicit use of substances to achieve an increase in performance (in sport)

→ A definition is difficult, since there must be a causative link between cause and action, similar to drugs

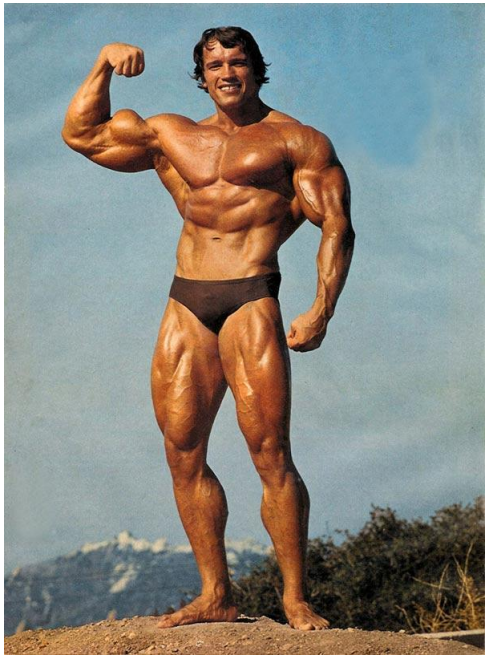
According substances are put together in doping lists by national and international sport committees (e.g. international olympic committee IOC) based on medical knowledge.

Doping list (I)

Illicit substance groups

- anabolic steroids (anabolics) lead to an increased building up of muscles

  naturally in the body occurring steroids such as testosterone, as well as totally artificial steroids e.g. tetrahydrogestrinone (THG)
  Partly not even allowed for fattening of porks!



- antiestrogenic compounds

  aromatase inhibitors
  tamoxifen, etc.

- hormons and related drugs

  erythropoietin (EPO): increased production of red blood cells

- insulin and insulin-like growth factors

→ substances that increase the oxygen transport capacity of the blood

Doping list (II)

Banned substance groups

- Stimulants increase the short term motivation

  amphetamines (cardiovasuclar and addiction risks)
  caffeine (until 2004 with a limit), due to newer results no limits any more

- narcotics and β-blocker show a calming down effect (pain reducing)

  (boxing, archery [Sportschießen])

Doping list (III)

- glucocorticoides (heart and circulation function)

- cannaboids
  hashish, marihuana

Mascing substances

- diuretica (increased elimination,
  reduction of body weight)
- inhibitors of the steroid-α-reductase (finasterid)
- plasmaexpanders (albumin, dextran) cause reduced drug concentration in the serum

Doping list (IV)

Substances with limits in certain sports

- alcohol (billard, tighter limits e.g. in racing)
- β-blocker (sports that require increased concentration)

- gene doping
  modification on the genetic level to increase performance
(nuclear receptors, mRNA, gene silencing)
feasibility, analytical proof?

Doping (V)

Doping lists are not comprehensive, which means that all similar compounds and those with a similar effect are included implicitly.
→ possibly not formulated precise enough for legal actions

Doping tests

Mainly urine samples, blood samples less frequent

problems: limits for metabolites of naturally occuring compounds, e.g. of testosterone and hematocrite
→ synthetic steroids show a different $^{12}C/^{13}C$ ratio
traceability of certain compounds (EPO)
new and formerly unknown compounds (e.g. THG)
→ give rise to unusual patterns in Mass-spectroscopy

Doping (VI)

Why doping tests ?

fairness, (self-)protection of the athletes

Risks of doping

- anabolic steroids: liver damage
- stimulants: addiction, lethal exhaustion
- known common adverse effects

Many drugs that are included in doping lists can administered with exception permits.
E.g. steroidal anti-inflammatories, anti-asthmatics