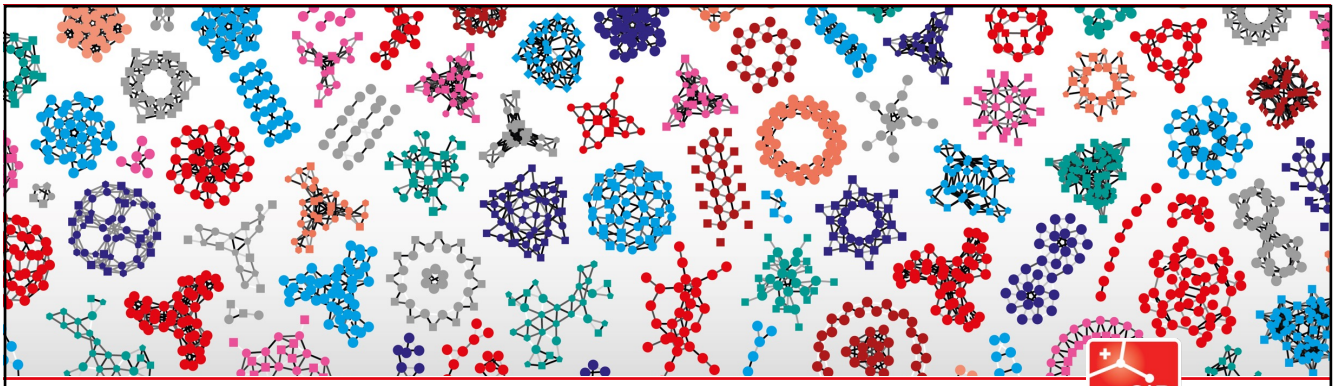


SIB
Swiss Institute of Bioinformatics

Computer-Aided Drug Design

Session 5. ADME, pharmacokinetics, druglikeness

Antoine Daina, Vincent Zoete
Molecular Modeling Group
SIB | Swiss Institute of Bioinformatics, Lausanne

1

Lectures & Practices Agenda

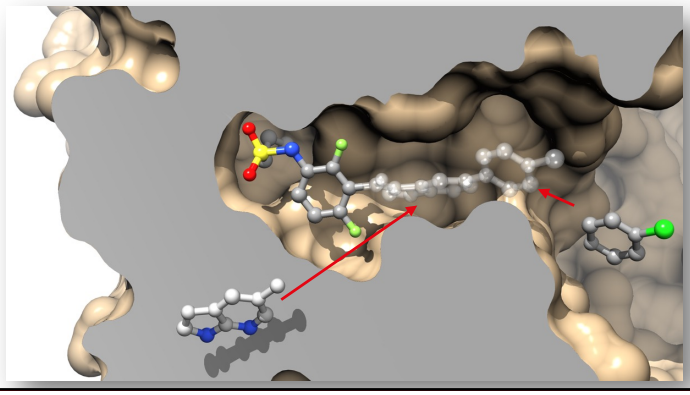
Session	Lecture	Practice
1	Prologue: molecular representation	
	Introduction to (computer-aided) drug design	
	Origin of 3D structures	
	Molecular recognition	Use of UCSF chimera to analyze protein-ligand complexes
2	Binding free energy estimation	
	Introduction to molecular docking	Ligand-protein docking with AutoDock Vina
3	Introduction to molecular (virtual) screening	Ligand-based virtual screening with **SwissSimilarity**
4	Short introduction on target prediction of small molecules	Use of **SwissTargetPrediction** to perform reverse screening.
5	Introduction to ADME, pharmacokinetics, druglikeness	Estimate physicochemical, pharmacokinetic, druglike and related properties with **SwissADME**
6	Short introduction to bioisosterism	Use of **SwissBioisostere** to perform bioisosteric design

2

Objectives of Drug Design

Objective: Create new chemical entities (NCE, “small molecules”) that activates or inhibits the function of a therapeutically relevant biomolecule target (mainly protein) in the organism.

To be effective as a drug, a potent molecule must reach its target in the body in sufficient concentration, and stay there in a bioactive form long enough for the expected biologic events to occur.



To address:

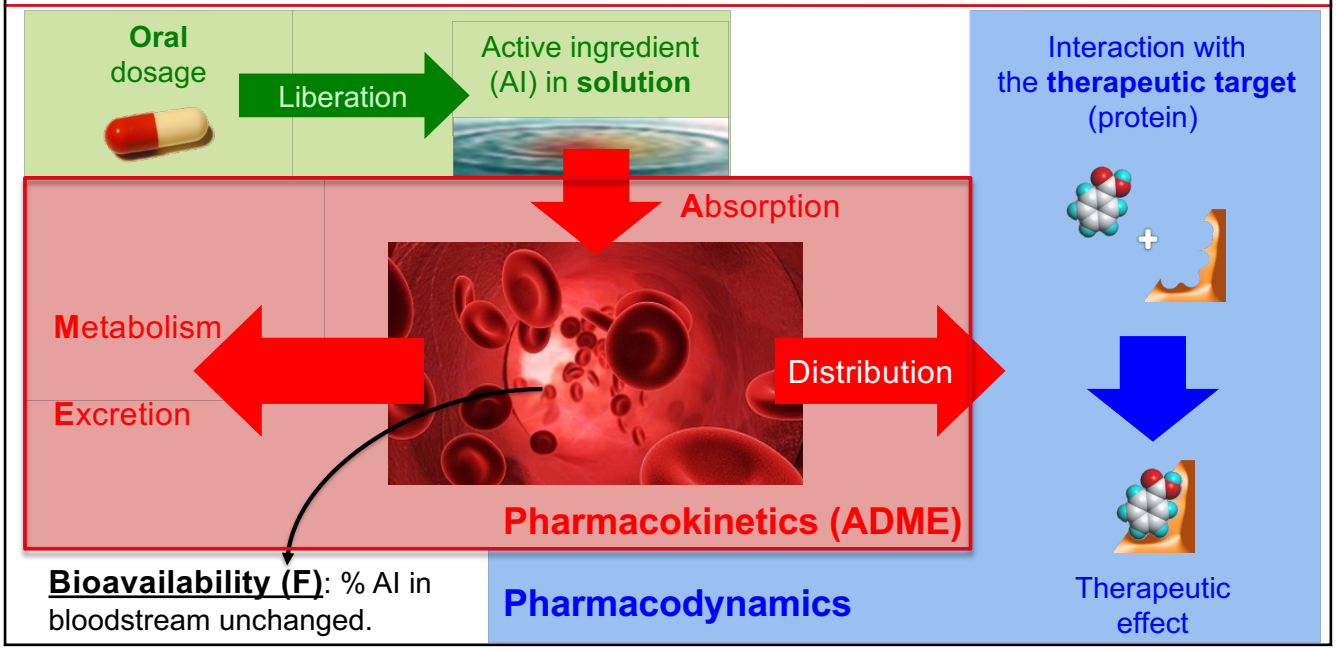
Molecular recognition; i.e. “Lock and key” (E. Fischer)

But also ADME,

- Absorption
- Distribution
- Metabolism
- Excretion
- (Toxicity)

3

The fate of an oral drug in the body

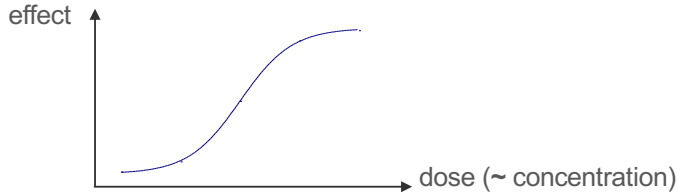


4

Clinical impact of ADME: in vitro vs. in vivo

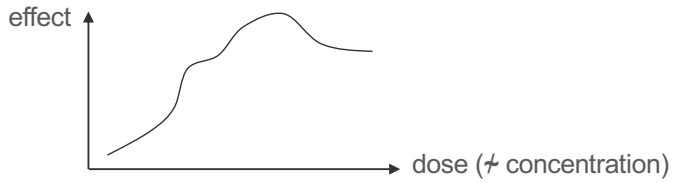
In vitro:

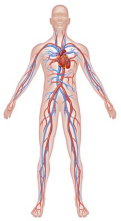
dose → concentration → effect



In vivo:

dose → ADME → concentration → effect

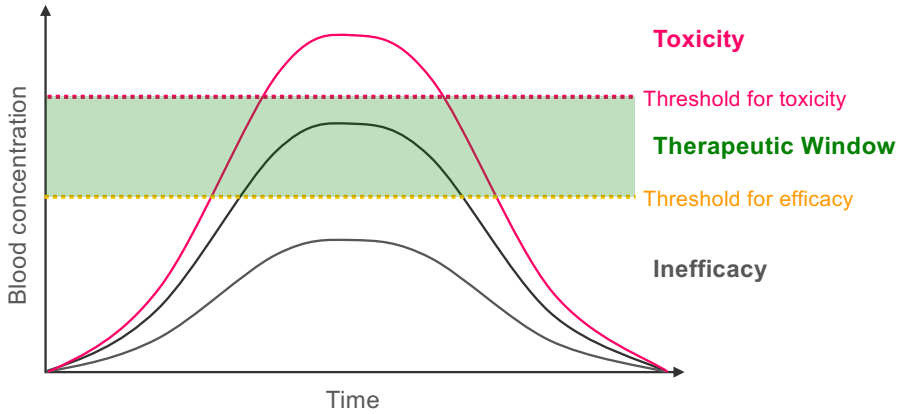




5

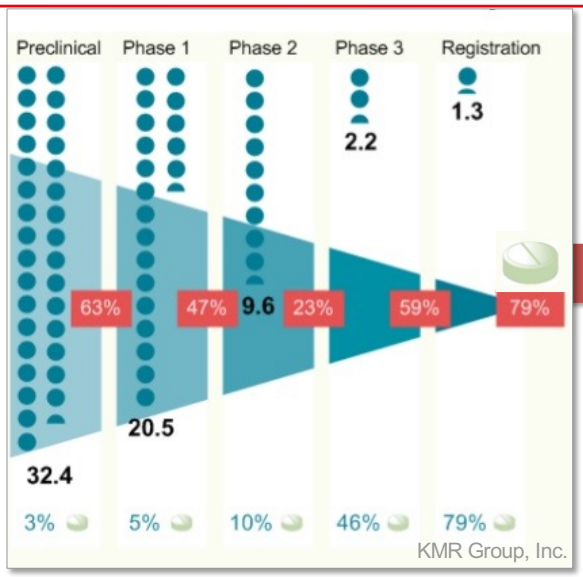
Clinical impact of ADME: efficacy vs. toxicity

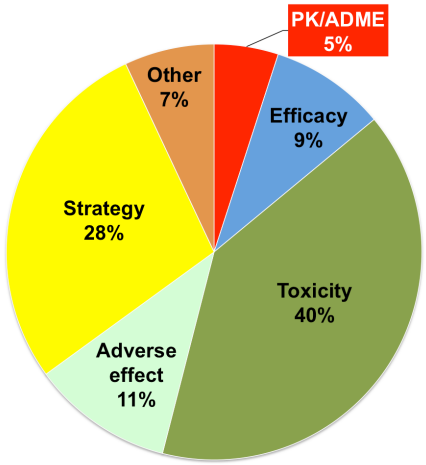
- Variability in ADME (e.g. metabolism) can have a dramatic impact on the efficacy and the toxicity of a drug.
- Schematic example, oral administration:



6

Failure for oral drugs in clinical development





2000 - 2010 (phases 2/3)

M.J. Waring et al. *Nat. Rev. Drug Discov.* **2015**
J. Arrowsmith & P. Miller *Nat. Rev. Drug Discov.* **2013**

- Estimation of ADME at early steps of drug discovery for PK profiles
- Development of computer models to predict ADME from chemical structures

7

Prediction of ADME *in silico*

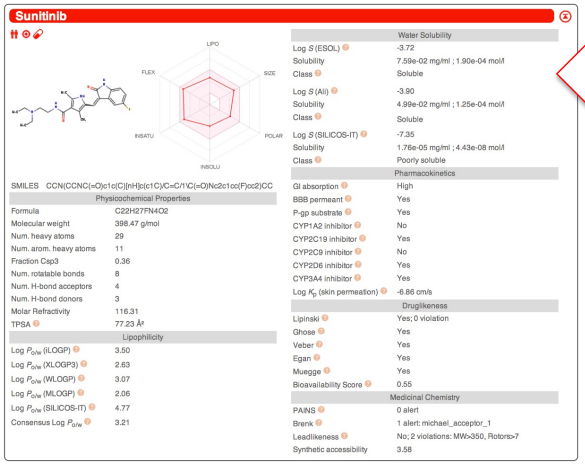
- To estimate ADME behaviors and pharmacokinetic parameters for drug design, **computational models** should be:
  - **robust** → reliable prediction
  - **fast** → numerous structures to be handled (also some « virtual » molecules).
  - **descriptive** enough and easy to **interpret in chemical** terms → main objective: modify molecule to target specific properties.

- **SwissADME**[1] web-based tools: access to multiple models for prediction of :
  - **Physicochemical parameters**
    (because ADME is directly related to the properties of the molecule)
  - individual **pharmacokinetic** behaviors
    (also graphical output)
  - **Medicinal chemistry** friendliness
    ... related properties

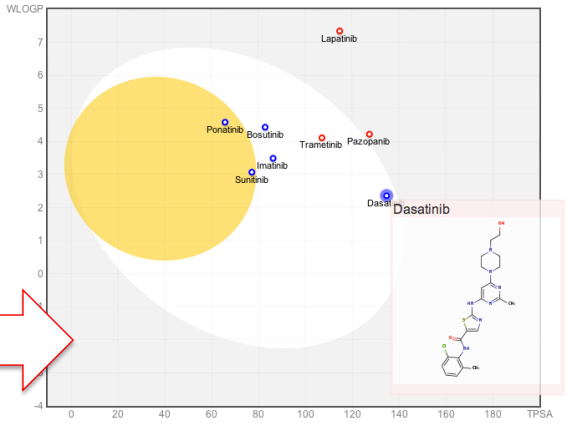
Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.*, **2017**, 7:42717

8

SwissADME: Output



One-panel-per-molecule
for clear output and export

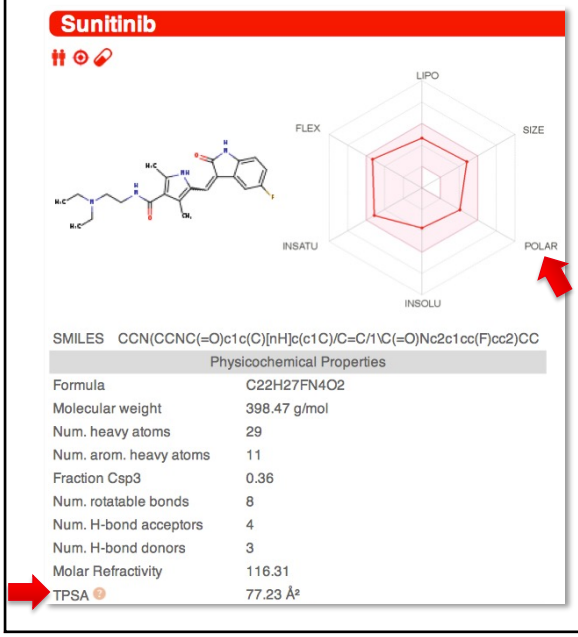


Graphical model for
individual ADME parameter
(e.g. BOILED-Egg[1]).

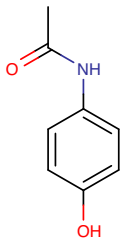
[1]Daina A. & Zoete V. ChemMedChem, 2016, 11(11), 1117-1121

9

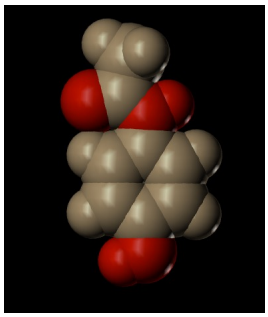
SwissADME: Physicochemical description of the molecule



e.g. **Polar Surface Area** (PSA, measure of **apparent polarity**)



Paracetamol

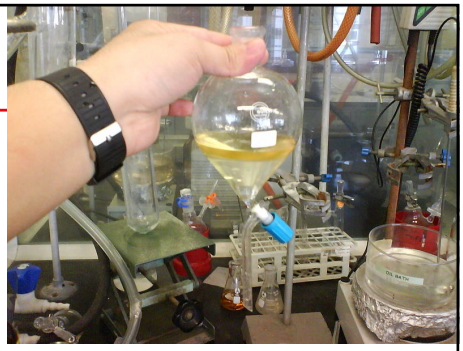


- TPSA, fast 2D, fragment-based[1].
- Necessary for **guidelines**:
  TPSA < 140 Å² good GI absorption.
  TPSA < 70 Å² good brain penetration.
- Necessary to build **rules** and more complex **models**.

[1]Ertl, P., Rohde, B. & Selzer, P., 2000. *J. Med. Chem.*, 43(20), pp.3714–3717.

10

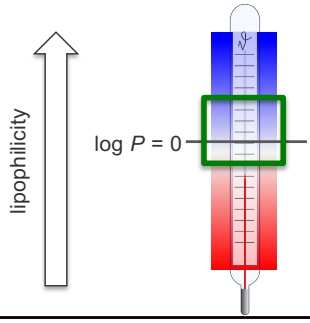
SwissADME: Lipophilicity



- Important physicochemical propertiy quantified by the partition coefficient **log $P_{o/w}$** between **water** and ***n*-octanol** (neutral form of the molecule).

$$\log P_{o/w} = \log \frac{C_o}{C_w}$$

- Best solvent system to mimic recognition of molecules with:
  - phospholipidic membranes ➔ ADME, pharmacokinetics,
  - macromolecules (proteins) ➔ Protein binding, pharmacodynamics.



Optimum of lipophilicity for an oral drug (druglikeness)

Hydrophobic (lipophilic) molecules are:
- well absorbed (A)
- BBB permeant (D)
- not released from plasma proteins (D, blood)
- rapidly metabolized (M)
- less excreted by kidney (E).

Polar (hydrophilic) molecules are:
- not well absorbed (but suitable for injection)
- well distributed (except in the brain)
- metabolically stable but rapidly excreted.

11

SwissADME: Prediction of (log *P*) Lipophilicity

- Multiple predictions by methods as different as possible
  - Each method has strongpoints and weaknesses

  ➔ Possibility of averaged **consensus prediction**

Lipophilicity	
Log $P_{o/w}$ (iLOGP)	3.50
Log $P_{o/w}$ (XLOGP3)	2.63
Log $P_{o/w}$ (WLOGP)	3.07
Log $P_{o/w}$ (MLOGP)	2.06
Log $P_{o/w}$ (SILICOS-IT)	4.77
Consensus Log $P_{o/w}$	3.21

- Classes of log *P* methods:
  1. **Fragmental** (*eg.* WLOGP)
  2. **Topological** (*eg.* MLOGP)
  3. **3D Physics-based** (iLOGP)

12

Log P estimation method 1: fragmental

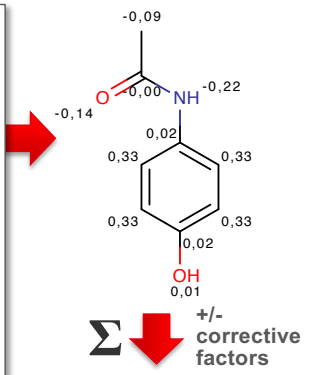
Principle:

1. **cutting** molecule into fragments
2. retrieving **contributions** in established **fragmental system**
3. **Addition** of fragmental values

fragmental system (by training):

```
for atom in openbabel.OBMolAtomIter(mol):
    if atom.MatchesSMARTS("[CH4]"):
        wlogp+= 0.1441
    elif wtom.MatchesSMARTS("[CH3]C"):
        wlogp+= 0.1441
    elif atom.MatchesSMARTS("[CH2](C)C"):
        wlogp+= 0.1441
    elif atom.MatchesSMARTS("[CH](C)(C)C"):
        wlogp+= 0
    elif atom.MatchesSMARTS("[C](C)(C)(C)C"):
        wlogp+= 0
    elif atom.MatchesSMARTS("[CH3][O,N,F,Cl,Br,#15,#16,#53;!a]"):
        wlogp+= -0.2035
    elif atom.MatchesSMARTS("[CH2X4][O,N,F,Cl,Br,#15,#16,#53;!a]"):
        wlogp+= -0.2035
    elif atom.MatchesSMARTS("[CHX4][O,N,F,Cl,Br,#15,#16,#53;!a]"):
        wlogp+= -0.2051
    elif atom.MatchesSMARTS("[CH0X4][O,N,F,Cl,Br,#15,#16,#53;!a]"):
        wlogp+= -0.2051
    elif atom.MatchesSMARTS("[C]=[A!#6]"):
        wlogp+= -0.2783
    elif atom.MatchesSMARTS("[CH2]=C"):
        wlogp+= 0.1551
    elif atom.MatchesSMARTS("[CH1](=C)A"):
        wlogp+= 0.1551
    elif atom.MatchesSMARTS("[CH0](=C)(A)A"):
        wlogp+= 0.1551
```

atomic contributions:



Predicted log *P*
WLOGP = 1.16

13

SwissADME: Water solubility

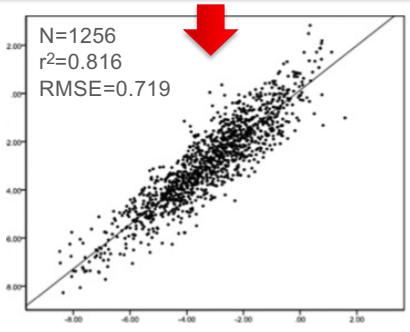
- Governs **absorption** and impacts many processes in **formulation**.
- Historical model (GSE, general solubility equation)[1]:

$$log\ S = 0.5 - 0.01 \times (m.p.\ ^{\circ}C - 25) - log\ P$$

- Ali[2]: replacement of melting point (difficult to predict) by TPSA:

$$log\ S = -1.0377 \times log\ P - 0.0210 \times TPSA + 0.4488$$

Solubility in water	
Log *S* (ESOL)	-5.73
Solubility	9.83e-04 mg/ml ; 1.85e-06 mol/l
Class	Moderately soluble
Log *S* (Ali)	-5.20
Solubility	3.38e-03 mg/ml ; 6.35e-06 mol/l
Class	Moderately soluble
Log *S* (SILICOS-IT)	-6.95
Solubility	6.00e-05 mg/ml ; 1.13e-07 mol/l
Class	Poorly soluble



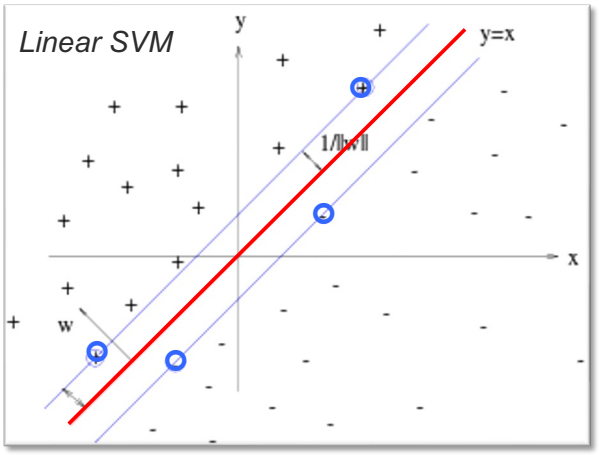
- SwissADME: 3 multiple linear models linking log *S* with physicochemical descriptors. ➔ **consensus** approach.
- **Qualitative** classes, log *S* scale:
  insoluble < -10 < poorly < -6 < moderately < -4 < soluble < -2 < very < 0 < highly

[1]Ran, Y. et al., 2001. J. Chem. inf. model., 41(5), pp.1208-1217
[2]Ali, J. et al., 2012. J. Chem. inf. model., 52(2), pp.420–428.

16

SwissADME: PK Classification Models

- **Support Vector Machine** (SVM)
  Machine learning patterns recognition algorithm to perform:
  - **Classification**: two (or more) groups of objects divided by a gap as wide as possible.
  - **Prediction**: new objects mapped onto that same space and predicted to which group they belong.



- To define the optimal **hyperplane** showing the **maximal margin** with the **support vectors** O.
- For prediction, often not optimal to separate all member of classes ➔ **permissivity** by **soft margin** controlled by C.
- In most case, better to map objects in a transformed space. Every dots are replaced by *eg*. a **Gaussian kernel function**
  ➔ **non-linear SVM** controlled by γ.

17

SVM classification model for Pharmacokinetics

Pharmacokinetics	
GI absorption	High
BBB permeant	Yes
P-gp substrate	Yes
CYP1A2 inhibitor	No
CYP2C19 inhibitor	Yes
CYP2C9 inhibitor	No
CYP2D6 inhibitor	Yes
CYP3A4 inhibitor	Yes
Log K_p (skin permeation)	-6.86 cm/s

- **P-gp** (permeability glycoprotein-1)
  - Actively **pumps out a substrate** across cellular membranes
  - ↘intestinal **absorption**, ↘**Brain** concentration
- **CYP** (cytochrome P450)
  Major family of metabolism enzymes
  - Mainly oxidation ↗water solubility, ↗elimination
  - **inhibition**: potential **drug-drug interactions**

Method to build the P-gp classification model:

- Objects / groups from litterature[1,2]:
  training set (564 P-gp substrates, 469 P-gp non-substrate [n=1033])
  test set (215 P-gp substrates, 200 P-gp non-substrates [n=415])
- Available descriptors: **55 SwissADME descriptors** (descriptive enough, orthogonal)
- Exhaustive grid search to find the **best combination of 15 descriptors and parameters C and γ**
  SVM built with libsvm3.20 library, RBF kernel (Gaussian).

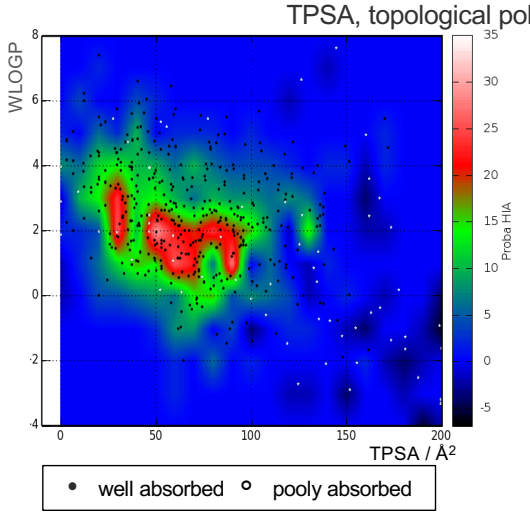
➔ **Classification accuracy: 72%** (10-fold CV), **89%** (external).

1. J. Mak *et al. J. Cheminfor.* **2015**
2. V. Poongavanam, *Bioorg. Med. Chem.* **2012**

18

Intuitive graphical model for passive absorption

- Dataset[1]: 567 well- and 93 poorly **absorbed** molecules by human GI (HIA)
- Structures : Cured for **passive** absorption, neutralized and translated into SMILES
- Descriptors : WLOGP, in-house implementation Wildman and Crippen[2] Log *P*
  TPSA, topological polar surface area from Ertl *et al.*[3]



- clear signal ✓
- region of elliptic shape[4] ✓

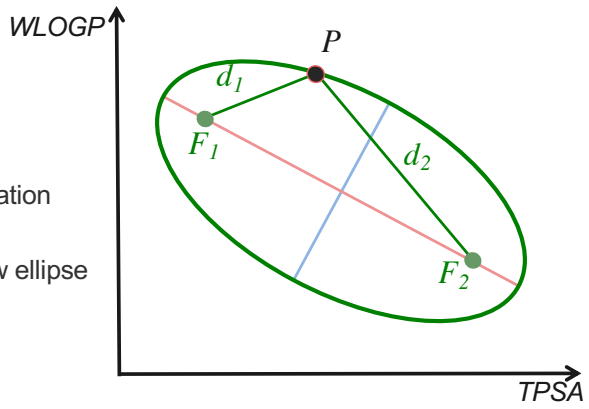
- build the predictive model
  i.e. finding the classifying ellipse.

1. D. Newby *et al. Eur. J. Med. Chem.* **2013**
2. S.A. Wildman & G.M. Crippen *J. Chem. Inf. Comput. Sci.* **1999**
3. P. Ertl *et al. J. Med. Chem.* **2000**
4. W.J. Egan *et al. J. Med.Chem.* **2000**

19

Methodology for the optimization of the ellipse

- **Parameters** geometric definition of the **ellipse**
  - $F_1(x_1,y_1)$; $F_2(x_2,y_2)$
  - $d_1 + d_2 = d$
- **Monte Carlo (MC)** optimization :
  1. generate initial ellipse and **evaluate** for classification
  2. random change ➔ new ellipse to evaluate
  3. use Metropolis criterion to keep or reject the new ellipse
  4. steps 2. and 3. repeated 100'000 times
- **Evaluation** : $Score = MCC - 0.05 \times surface$



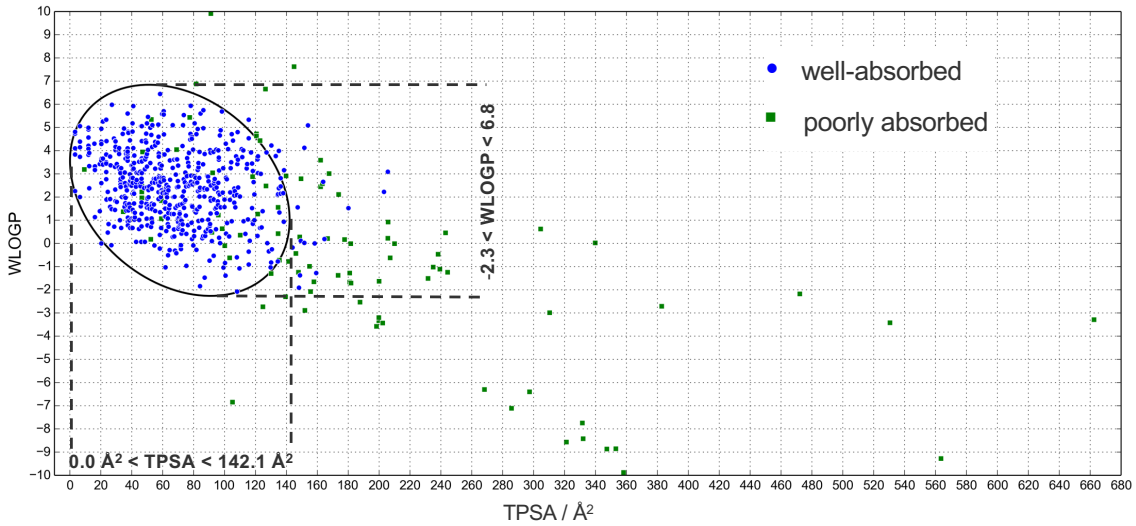
Matthew's Correlation Coefficient

from -1 to +1 (perfect prediction), near 0 for random classification.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

20

Best classification ellipse for passive absorption



MCC = 0.70

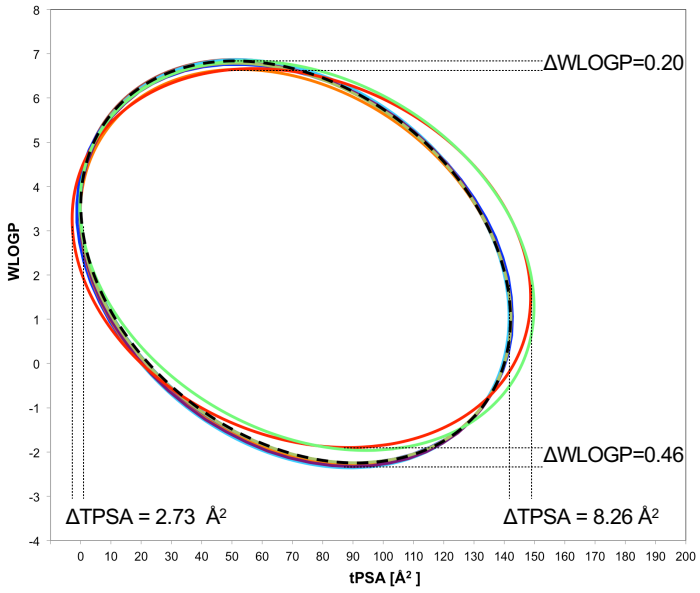
Accuracy = 93 %

In accordance with usual guidelines for good absorption

21

10-fold crossvalidation

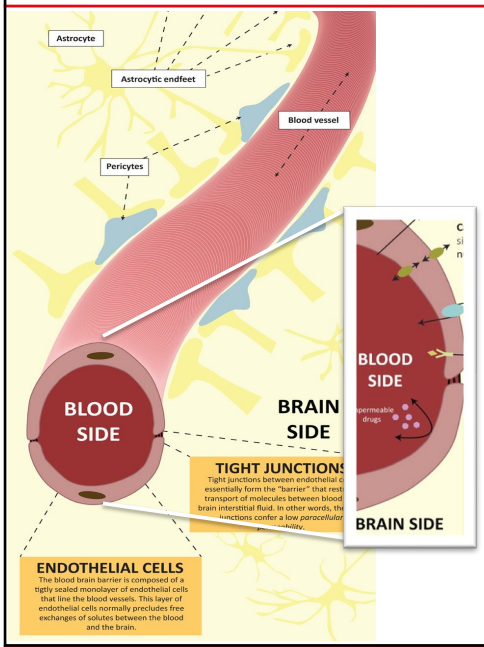
	training / validation	MCC_{TR}	ACC_{TR}	MCC_{VA}	ACC_{VA}
1	595/65	0.6957	0.9265	0.7421	0.9385
2	595/65	0.7064	0.9296	0.6948	0.9254
3	594/66	0.6981	0.9280	0.6948	0.9254
4	593/67	0.6924	0.9263	0.7649	0.9403
5	593/67	0.7064	0.9296	0.6474	0.9104
6	593/67	0.7091	0.9296	0.5298	0.8806
7	594/66	0.7012	0.9281	0.6940	0.9242
8	594/66	0.7233	0.9331	0.4854	0.8788
9	594/66	0.7041	0.9314	0.6140	0.9091
10	594/66	0.6977	0.9298	0.6641	0.9242
	Cross-validation			**0.6531**	**0.9156**



- MCC_{CV} = 0.65 / ACC_{CV} = 0.92 ➔ Internal robustness ✓

22

Another important physiological barrier : BBB



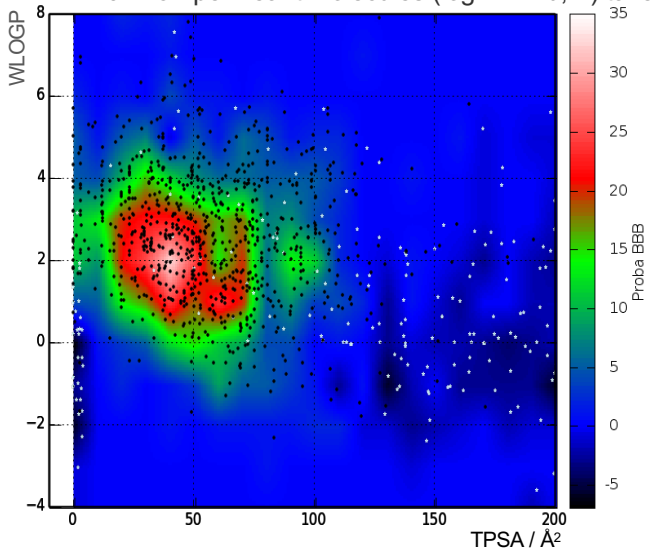
- **Blood-Brain Barrier** (BBB) is an effective **shield** protecting the brain :
  - **physical** barrier
    e.g. tight junctions preventing paracellular penetration
  - **biochemical** barrier
    enzymatic activities and active efflux
    e.g. P-glycoprotein 1 (P-gp) pumping out
- **Passive diffusion** through BBB is the **major route** for drugs to access the brain from the bloodstream.[1]
- Fundamental for the **distribution** of central-acting drugs or reversely for limited **unwanted effects** of peripheral drugs (or any molecule)

1. L. Di *et al. Drug Discov. Today* **2012**

23

Heatmap for passive diffusion through BBB

- 156 permeant molecules (log *BB* > *0*, •) and
- 104 non-permeant molecules (log *BB* < *0*, °) taken from Brito-Sanchez *et al.*[2]



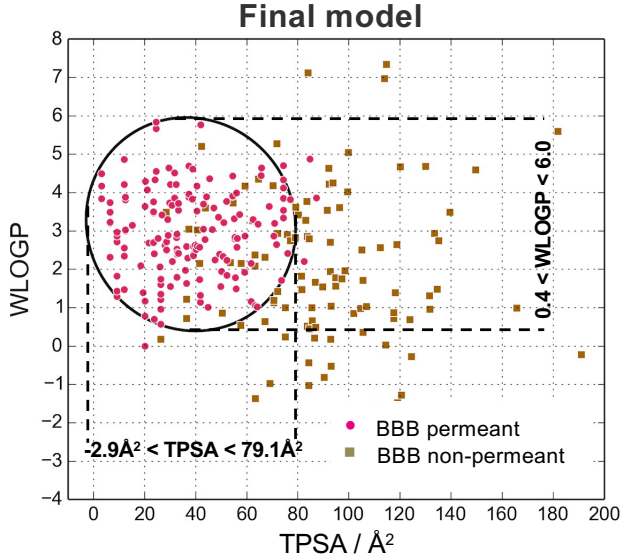
- clear signal ✓
- narrower space ✓
- optimization of a different ellipse
  same method
  same descriptors

1. L. Di *et al. Drug Discov. Today* **2012**
2. Y. Brito-Sánchez *et al. Mol. Inf.* **2015**

24

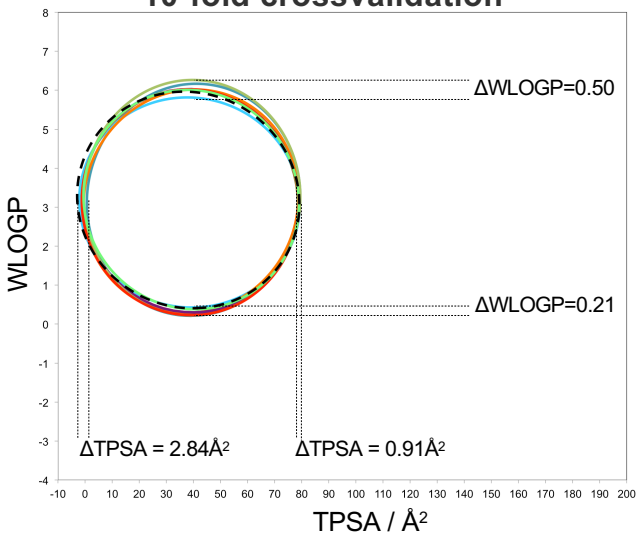
Best classification ellipse for passive BBB permeation

Final model



- **MCC = 0.79**
- **Accuracy = 90 %**

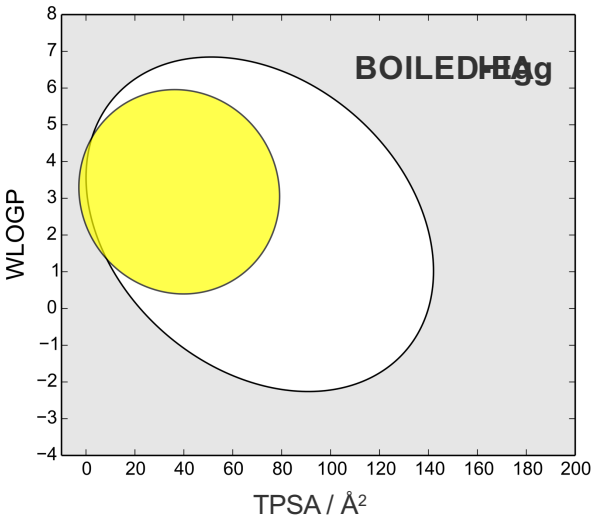
10-fold crossvalidation

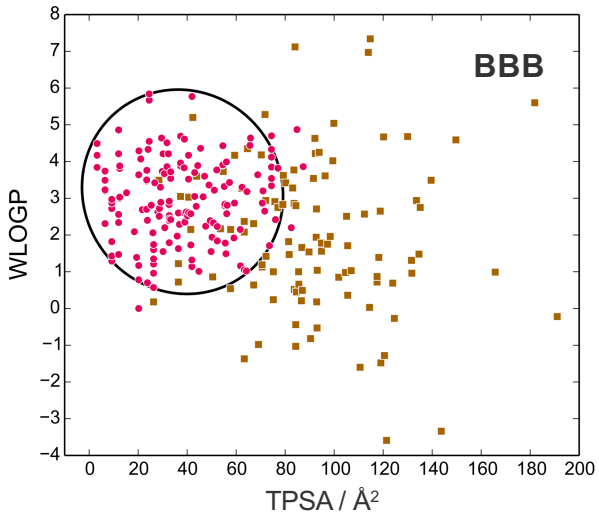


- Consistent with common guidelines for brain penetration
- MCC_{CV} = 0.75 / ACC_{CV} = 88%
- Internal robustness ✓

25

BOILED-Egg: merging both models

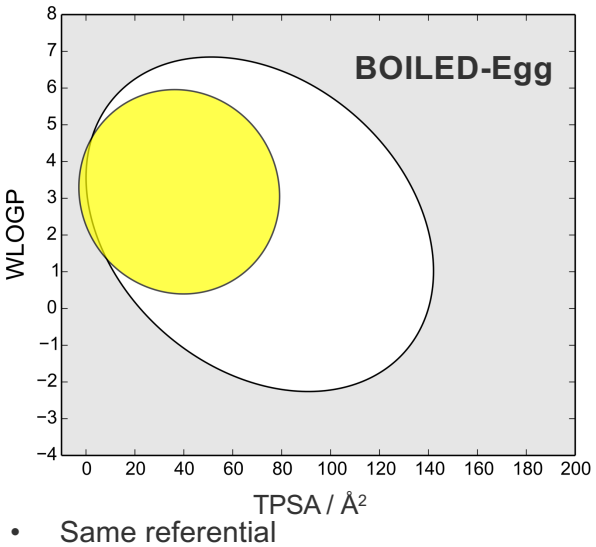


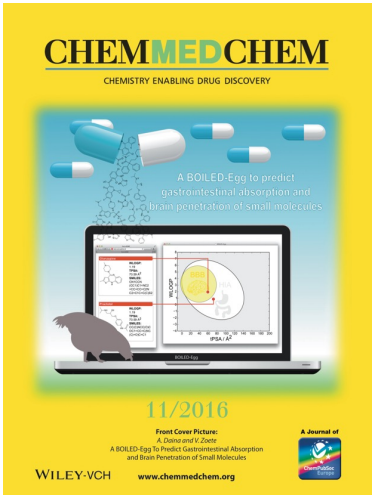


- Same referential
- Brain Or IntestinaL EstimateD permeation / Yolk and white not mutually exclusive

26

BOILED-Egg: merging both models





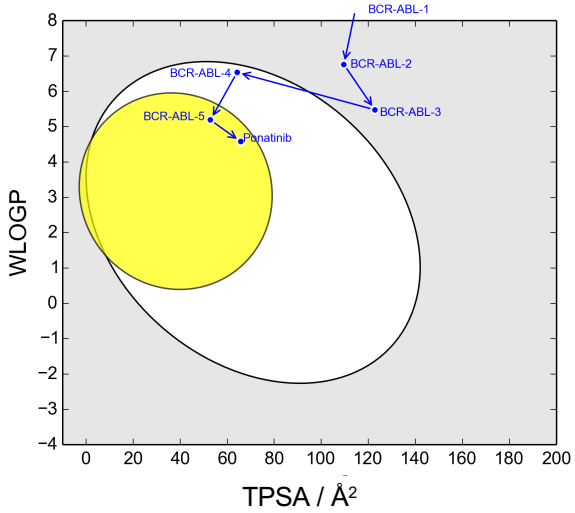
- Same referential
- Brain Or IntestinaL EstimateD permeation / Yolk and white not mutually exclusive

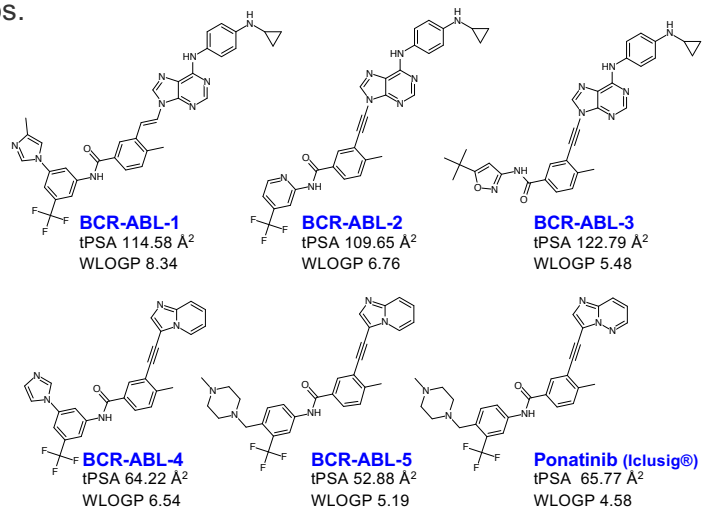
1. A. Daina & V. Zoete, *ChemMedChem* **2016**

27

BOILED-Egg to track drug optimization path

- Exercise: Lead optimization of BCR-ABL tyrosine kinase inhibitors to anti-cancer Ponatinib[1].
- Successive PD and PK optimization steps.





BCR-ABL-1
tPSA 114.58 Å²
WLOGP 8.34

BCR-ABL-2
tPSA 109.65 Å²
WLOGP 6.76

BCR-ABL-3
tPSA 122.79 Å²
WLOGP 5.48

BCR-ABL-4
tPSA 64.22 Å²
WLOGP 6.54

BCR-ABL-5
tPSA 52.88 Å²
WLOGP 5.19

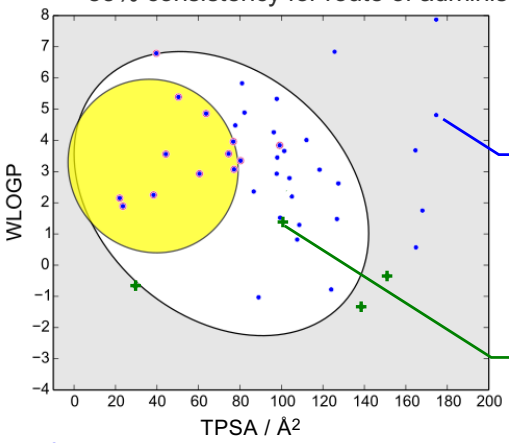
Ponatinib (Iclusig®)
tPSA 65.77 Å²
WLOGP 4.58

1. W.S. Huang *et al. J. Med. Chem.* **2010**

28

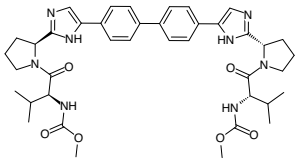
BOILED-Egg to map new FDA drugs

- Exercise: map 46 NCEs accepted by the FDA in 2014 and 2015
- Colored according to usage label
- 83% consistency for route of administration and in/out « white »



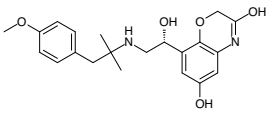
- Oral administration
- Other routes only (i.v., inhalation, transdermal, ...)
- CNS indication

- 6 false negatives (FN)



Daclatasvir
Good absorption
Active transport[1] OCT1

- 1 false positives (FP)

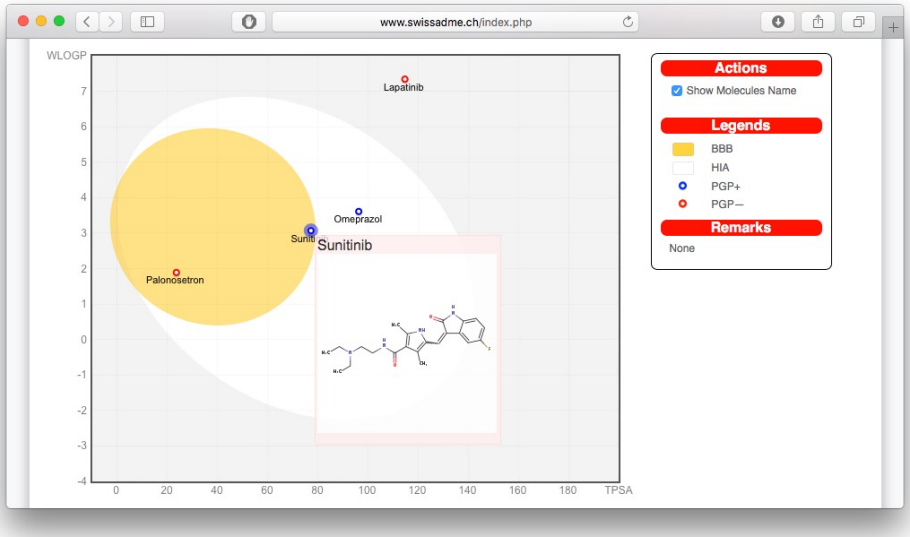


Olodaterol
FA 10-20%
P-gp efflux[2]

1. FDA, Clinical pharmacology reviews **2014**
2. Health Safety Regulation (Aus) **2014**

29

SwissADME graphical output: enhanced BOILED-Egg



SwissADME includes an enhanced version of the BOILED-Egg, including

- estimation of P-gp substrate (active efflux)
- interactive analysis capabilities

30

SwissADME: Druglikeness

Druglikeness	
Lipinski	Yes; 0 violation
Ghose	Yes
Veber	Yes
Egan	Yes
Muegge	Yes
Bioavailability Score	0.55

- **Lipinski Rule-of-five (Pfizer Ro5)[1]:**
  (fail not more than 1 criteria)
  - MW < 500 g/mol
  - CLOGP < 5 (MLOGP < 4.15)[2]
  - # H-bond donors ≤ 5
  - # H-bond acceptors ≤ 10
- **Egan filter (Pharmacopia)[3]:**
  - ALOGP98 < 6
  - TPSA < 132 Å²

- Druglikeness Rules:
  Qualitative estimation defined by ranges of specific physichochemical properties that make a molecule a **possible oral drug.**
- « *Like a drug* » from the PK (ADME) point of view (**bioavailable**, not necessarily bioactive).

- Druglike Filters:
  Ranges based on computed/predicted properties derived from **known oral drugs**.
- Industry: enrich, improve quality of proprietary collection
- Main use:
  Filtering large chemical librairies
  SwissADME:
  Multiple filters for consensus

[1]Lipinski, C., et al., *Adv. Drug Deliv. Rev.*, **1997**, 23, 3.
[2]Lipinski, C., et al. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3.
[3]Egan, W.J., Merz, K.M., Baldwin, J.J., *J. Med. Chem.*, **2000**, 43, 3867.

31

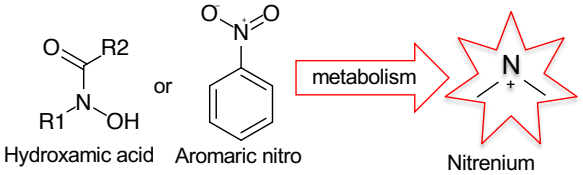
SwissADME: Medicinal Chemistry

Medicinal Chemistry	
PAINS	0 alert
Brenk	3 alerts: hydroxamic_acid, oxygen-nitrogen_single_bond, triple_bond
Leadlikeness	No; 3 violations: MW>350, Rotors>7, XLOGP3>3.5
Synthetic accessibility	3.39

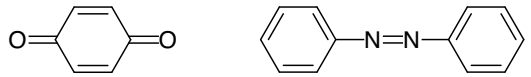
- PAINS: promiscuous fragments[2]:
  Tendency for compounds to be **unspecific frequent hitter** in many assays.

- Synthetic accessibility score[3], estimation of the ease of chemical synthesis, combination of:
  - **Similarity** with **existing molecules** in vendor catalogs (list of molecules immediately available).
  - Objective **estimation of the complexity** (macrocycles, chiral centers, ...)

- Brenk: problematic fragments[1]:
  - **toxic**, **reactive**, **unstable**
  - **aggregator** ➔ false positive
  - **dye** ➔ perturbation of assays



Hydroxamic acid | Aromaric nitro | Nitrenium highly reactive



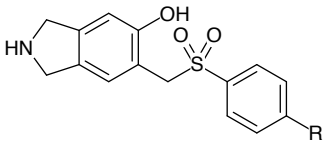
Quinones (promiscuous, unstable, colored) | Azoic dyes

[1]Brenk, R. et al., 2008. *ChemMedChem*, 3(3), p.435
[2]Baell, J.B. & Holloway, G.A., 2010. *J. Med. Chem.*, 53(7), p.2719
[3]Ertl, P. & Schauffenhauer, A., 2009. *J Cheminf.*, 1(1), p.1

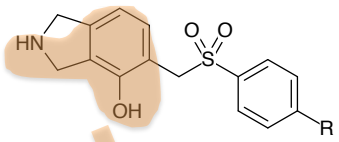
32

SwissADME: PAINS

- PAINS[1] (Pan Assay Interference compounds from Screening): **Statistical analysis** of **frequent hitters** (**promiscuous compounds**) and breakdown in 481substructures.
- Real-life usefulness:



Medicinal chemistry	
PAINS	0 alert
Brenk	0 alert



Medicinal chemistry	
PAINS	1 alert(s) mannich_A
Brenk	0 alert

Promiscuous (PAINS) fragment if enrichment ≥ 30%

Table S12. Phenolic mannich bases as problematic compounds

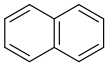
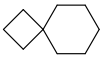
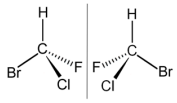
Substructure[a]	Number of AlphaScreen® assays hit							Total Cpds	Enrichment[b]
	6	5	4	3	2	1	0		
 mannich_A	2	4	13	15	59	57	146	296	64%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1
b Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

[1]Baell, J.B. & Holloway, G.A., 2010. J. Med. Chem., 53(7), p.2719.

33

Synthetic accessibility score (SAscore)

- Estimation of the **ease of synthesis**, based on:
  - **Similarity**: Cut the molecule into fragments (**fingerprints**) and compare to:
    - ECFP4 Fingerprints on "all-now" subset of ZINC (12,782,590 molecules).
    - 412,579 fragments were retrieved and ranked.
    - Contribution: log (frequency / # frg counting for 80% of database).
      - Frequent fragments ➔ positive contributions.
      - Rare fragments ➔ negative contributions.
  - **Complexity** penalties: Count of specific chemical moieties.
    - Fused rings:  / spiros:  / macrocycles: 
    - Stereocenter: 

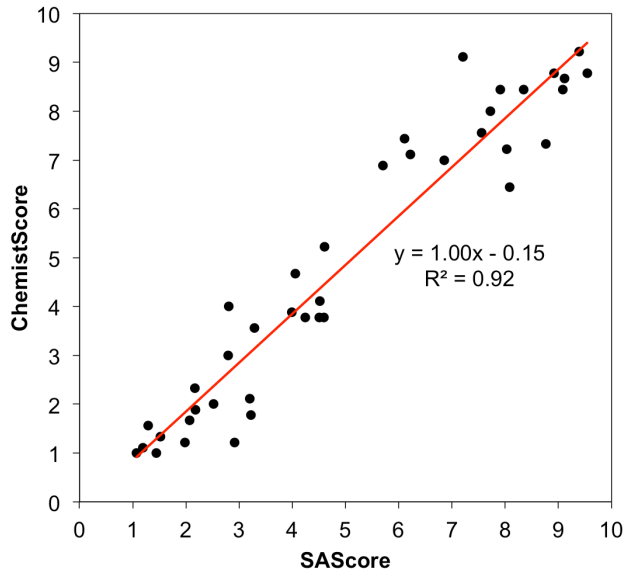
    - Number of atoms
- Finally normalized, so:

  1 (**easy**) ≤ SAscore ≤ **10** (**difficult**)

34

SAscore: performance

- Comparison with estimation given by 9 chemists (ChemistScore) for 40 molecules[1]

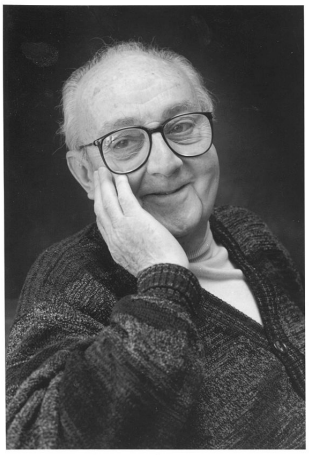


- Good linear correlation (R^2=0.92)
- Good ranking correlation (Spearman's r=0.94).

[1] Ertl, P. & Schuffenhauer, A., 2009. *J Cheminf.*, 1(1), p.8.

35

« *All models are wrong but some are useful*. »



George E. P. Box, 1979

36