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

Historical Formulas of Acetic Acid (1860)

Joseph Loschmidt

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

On outliers and activity cliffs - why QSAR often disappoints

Beware of q2!

3D-QSAR illusions

The trouble with QSAR (or how I learned to stop worrying and embrace fallacy)

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

---

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 $\chi$

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.
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);

„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

Proper Validation of QSAR and 3D QSAR Models

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

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

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

Good and Poor Science

Sir Karl Popper
★ 1902 Vienna, ♿ 1998 London

[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:
  - \(-\text{SH},\ -\text{COOH},\ -\text{PO}_3\text{H}_2,\ >\text{PO}_2\text{H}\)
  - Functional groups: \(\text{C} / \text{O} / \text{COOH}\)
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.

LigandScout Pharmacophore Recognition (inte:ligand)

Biotin-Streptavidin Complex (2rtf, 1.47Å)
LigandScout Pharmacophore Recognition (inte:ligand)

Biotin
Streptavidin Complex
(2rtf, 1.47Å)
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 pKa Prediction

pKa model in ADMET Predictor 4.0

pKa = 6-9

Hugo Kubinyi, www.kubinyi.de

**Software for pKα Prediction**

pKα model in ADMET Predictor 4.0
courtesy of Robert Fraczkiewicz, Simulations Plus, Inc.

**Acceptor Potentials of Esters and Oxazoles**

**Esters**

**Oxazoles**

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?
Drug Research is ....

the Search for a Needle in a Haystack
Tools for Virtual Screening

- Garbage filter: 90%
- Druglike / Non-druglike: 75%
- Bioavailability: 60%
- Cytotoxicity:
- hERG channel inhibition:
- Antitargets:
  - $\alpha_{1a}$ (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

- 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
Performance of Different Scoring Functions

(n = 800)

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

R. Wang et al.,

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

n = 229; r = 0.875
n = 855; r = 0.770

Unrecognized Favorable Interactions

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


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


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

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

---

**Human Absorption and Polar Surface Area**

\[
\text{FDP} = \text{fraction of dose absorbed to the portal vein}
\]

Human Absorption and Polar Surface Area

\[ F_{Dp} = \text{fraction of dose absorbed to the portal vein} \]

triangles and circles:

data from

Rodent, Dog, Primate and Human Bioavailability

Rodents
green circles

Dog
red triangles

Primates
blue squares

data from:

**The Role of Transporters in Drug Absorption and Elimination**


---

**Oxidative Metabolism and Drug Design**

Prediction of Drug Metabolism: MetaPrint2D

predictions for human, dog, rat, all

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

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

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

figure: S. D. Krämer and B. Testa, Chemistry & Biodiversity 5, 2465-2578 (2008)
### Biological Activities of Metabolites

<table>
<thead>
<tr>
<th>Compound</th>
<th>monoamine uptake inhibition rat synaptosomes, IC₅₀ in nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sibutramine (racemate)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DAT</td>
</tr>
<tr>
<td></td>
<td>1200</td>
</tr>
<tr>
<td>(R)</td>
<td>12</td>
</tr>
<tr>
<td>(S)</td>
<td>180</td>
</tr>
<tr>
<td>(R)</td>
<td>9</td>
</tr>
<tr>
<td>(S)</td>
<td>12</td>
</tr>
</tbody>
</table>


---

### Biological Activities of Metabolites

- **Rat**: no toxicity
- **Human**: hepatotoxicity (via quinone imine)

Reasons for Failure in Drug Development

1964-1985
(n = 121; without antiinfectives)
T. Kennedy, Drug Discov. today 2, 436-444 (1997)

16%
ADME
7%
Lack of efficacy
7%
Animal toxicity
7%
Adverse effects in man
17%
Commercial reasons
46%
Miscellaneous

1992-2002
(reasons for market withdrawal, n = 16: toxicity 93%, efficacy 7%)

33%
Liberation + ADME
33%
Lack of efficacy
8%
Toxicity
11%
Economic
2%
Other
43%
Not published

Thank you

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