Advanced

in silico Drug Design

KFC/ADD

QSAR and ADMET

Karel Berka, Ph.D.
Jindřich Fanfrlík, Ph.D.
Martin Lepšík, Ph.D.

UP Olomouc, 1.-3.2. 2016
QSAR
QSAR

• **Quantitative Structure-Activity Relationship**

• Mathematical regression function of linearized (biological) activity from description of the molecule (MW, size, number of atoms...)

• \( f(\text{activity}) = A \cdot \text{descriptor}_A + B \cdot \text{descriptor}_B + ... \)

Active compounds

New molecules with predicted activity
3D-QSAR Assumptions

- The effect is produced by modeled compound and not its metabolites.
- The proposed conformation is the bioactive one.
- The binding site is the same for all modeled compounds.
- The biological activity is largely explained by enthalpic processes.
- Entropic terms are similar for all the compounds.
- The system is considered to be at equilibrium, and kinetics aspects are usually not considered.
- Pharmacokinetics: solvent effects, diffusion, transport are not included.
General Procedure of QSAR

• Select a set of molecules interacting with the same receptor with known activities.
• **Calculate** features – descriptors (e.g. physicochemical properties, etc., 2D, 3D)
• Divide the set to two subgroups: one for **training** and one for **testing**.
• Build a **model**: find the relations between the activities and properties (regression problem, statistic methods, machine learning approaches, etc).
• **Test** the model on the testing dataset.
Advantages of QSAR

• Quantifying the relationship between structure and activity => an understanding of the effect of structure on activity (SAR).

• It is also possible to make predictions leading to the synthesis of novel analogues.

• The results can be used to help understand interactions between functional groups in the molecules of greatest activity, with those of their target
**Statistical Concept**

- Input: $n$ descriptors $P_1,..,P_n$ and the value of biological activity in linearizable form ($EC50$ for example is usually changed for $pEC50$) for $m$ compounds

<table>
<thead>
<tr>
<th></th>
<th>Bio</th>
<th>P1</th>
<th>P2</th>
<th>......</th>
<th>..</th>
<th>..</th>
<th>..</th>
<th>..</th>
<th>Pn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cpd 1</td>
<td>0.7</td>
<td>3.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpd2</td>
<td>3.2</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>......</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpd m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Molecular Properties

**MOLECULAR STRUCTURE**

**INTRINSIC PROPERTIES**
- Molar Volume
- Connectivity Indices
- Charge Distribution
- Molecular Weight
- Polar surface Area...

**CHEMICAL PROPERTIES**
- pKa
- Log P
- Solubility – log S
- Stability

**BIOLOGICAL PROPERTIES**
- Activity
- Toxicity
- Biotransformation
- Pharmacokinetics
Molecular Descriptors

- Molecular descriptors are numerical values that characterize properties of molecules.

- The descriptors fall into 4 classes:
  a) Topological
  b) Geometrical
  c) Electronic
  d) Hybrid or 3D
Classification of Descriptors

Topological Descriptors

- derived directly from the connection table representation of the structure:

  a) Atom and Bond Counts
  b) substructure counts
  c) molecular connectivity Indices (Weiner Index, Randic Index, Chi Index)
  d) Kappa Indices
  e) path descriptors
  f) distance-sum Connectivity
  g) Molecular Symmetry
Classification of Descriptors

Geometrical Descriptors

- derived from the 3D representations:
  a) principal moments of inertia
  b) molecular volume
  c) solvent-accessible surface area
  d) hydrophilic/hydrophobic partial surface area
  e) Molecular Surface area
Classification of Descriptors

Electronic Descriptors

- derived from electronic distribution within molecule:

  a) dipole moment
  b) quadrupole moment
  c) polarizibility
  d) HOMO and LUMO energies,
  e) dielectric energy
  f) molar refractivity
Classification of Descriptors
Hybrid and 3D Descriptors

a) geometric atom pairs and topological torsions
b) spatial autocorrelation vectors
c) WHIM indices
d) BCUTs
e) GETAWAY descriptors
f) Topomers
g) Pharmacophore fingerprints
h) Eva Descriptors
i) Descriptors of Molecular Field
# Group Additive Properties, GAPs

- Descriptors can be optimized to common groups

<table>
<thead>
<tr>
<th>Substituent</th>
<th>Volume (SA)</th>
<th>MR</th>
<th>$\pi$</th>
<th>Rot Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>-H</td>
<td>1.48</td>
<td>0.10</td>
<td>0 (reference)</td>
<td>0</td>
</tr>
<tr>
<td>-CH$_3$</td>
<td>18.78</td>
<td>0.57</td>
<td>0.56</td>
<td>0</td>
</tr>
<tr>
<td>-CH$_2$CH$_3$</td>
<td>35.35</td>
<td>1.03</td>
<td>1.02</td>
<td>1</td>
</tr>
<tr>
<td>-CH$_2$CH$_2$CH$_3$</td>
<td>51.99</td>
<td>1.5</td>
<td>1.55</td>
<td>2</td>
</tr>
<tr>
<td>-CH(CH$_3$)$_2$</td>
<td>51.33</td>
<td>1.5</td>
<td>1.53</td>
<td>1</td>
</tr>
<tr>
<td>-CH$_2$CH$_2$CH$_2$CH$_3$</td>
<td>68.63</td>
<td>1.96</td>
<td>2.13</td>
<td>3</td>
</tr>
<tr>
<td>-C(CH$_3$)$_3$</td>
<td>86.99</td>
<td>1.96</td>
<td>1.98</td>
<td>1</td>
</tr>
<tr>
<td>-C$_6$H$_5$</td>
<td>72.20</td>
<td>2.54</td>
<td>1.96</td>
<td>1</td>
</tr>
<tr>
<td>-F</td>
<td>7.05</td>
<td>0.10</td>
<td>0.14</td>
<td>0</td>
</tr>
<tr>
<td>-Cl</td>
<td>15.85</td>
<td>0.60</td>
<td>0.71</td>
<td>0</td>
</tr>
</tbody>
</table>

- Final value is sum of group additions
Limit Of Descriptors

- The data set should contain at least 5 times as many compounds as number of descriptor in QSAR.
- The reason for this is that too few compounds relative to the number of descriptors will give a falsely high correlation:
  - 2 point exactly determine a line.
  - 3 points exactly determine a plane (etc.)
Examples

• Hammett Relationships
• log P : Octanol-water partition coefficients
  – uses in Pharmaceutical Chemistry
  – uses in Environmental Chemistry
  – uses in Chromatography

ADMET
Hammett Relationships

• pKa of benzoic acids
• Effect of electron withdrawing and donating groups
• based on $\Delta_r G = -RT \ln K_{eq}$
pKa Substituted Benzoic Acids

• $\log K_a - \log K_{aH} = \sigma$

• $K_{aH}$ is the reference compound (unsubstituted)
# Hammett $\sigma$ Constants

<table>
<thead>
<tr>
<th>Group</th>
<th>$\sigma_p$</th>
<th>$\sigma_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-NH$_2$</td>
<td>-0.57</td>
<td>-0.09</td>
</tr>
<tr>
<td>-OH</td>
<td>-0.38</td>
<td>0.13</td>
</tr>
<tr>
<td>-OCH$_3$</td>
<td>-0.28</td>
<td>0.10</td>
</tr>
<tr>
<td>-CH$_3$</td>
<td>-0.14</td>
<td>-0.06</td>
</tr>
<tr>
<td>-H</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-F</td>
<td>0.15</td>
<td>0.34</td>
</tr>
<tr>
<td>-Cl</td>
<td>0.24</td>
<td>0.37</td>
</tr>
<tr>
<td>-COOH</td>
<td>0.44</td>
<td>0.35</td>
</tr>
<tr>
<td>-CN</td>
<td>0.70</td>
<td>0.62</td>
</tr>
<tr>
<td>-NO$_2$</td>
<td>0.81</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Octanol-Water Partition Coefficients

- $P_{ow} = \frac{C(\text{octanol})}{C(\text{water})}$
- $\log P$
  
  like $\Delta_r G = -RT \ln K_{eq}$

- Hydrophobic - hydrophilic character
- $P$ increases then more hydrophobic
log P

hydrophillic

methanol -1.27
isopropanol -0.36
ethanol -0.75
n-propanol -0.23

hydrophobic

pentanol 0.81
butylamine 0.85
pyridine 0.64
diethylamine 0.45
imidazole -0.08
phenylalanine -1.38
tetraethylammonium iodide -2.82
alanine -2.85

benzene 2.13
QSAR and log P

- Isonarcotic Activity of Esters, Alcohols, Ketones, and Ethers

<table>
<thead>
<tr>
<th>Compound</th>
<th>log(1/C)</th>
<th>log P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_3$OH</td>
<td>0.30</td>
<td>-1.27</td>
</tr>
<tr>
<td>C$_2$H$_5$OH</td>
<td>0.50</td>
<td>-0.75</td>
</tr>
<tr>
<td>CH$_3$COCH$_3$</td>
<td>0.65</td>
<td>-0.73</td>
</tr>
<tr>
<td>(CH$_3$)$_2$CHOH</td>
<td>0.90</td>
<td>-0.36</td>
</tr>
<tr>
<td>(CH$_3$)$_3$COH</td>
<td>0.90</td>
<td>0.07</td>
</tr>
<tr>
<td>CH$_3$CH$_2$CH$_2$OH</td>
<td>1.00</td>
<td>-0.23</td>
</tr>
<tr>
<td>CH$_3$COOCH$_3$</td>
<td>1.10</td>
<td>-0.38</td>
</tr>
<tr>
<td>C$_2$H$_5$COCH$_3$</td>
<td>1.10</td>
<td>-0.27</td>
</tr>
<tr>
<td>HCOOC$_2$H$_5$</td>
<td>1.20</td>
<td>-0.38</td>
</tr>
<tr>
<td>C$_2$H$_5$COC$_2$H$_5$</td>
<td>1.20</td>
<td>0.59</td>
</tr>
<tr>
<td>(CH$_3$)$_2$C(C$_2$H$_5$)OH</td>
<td>1.20</td>
<td>0.59</td>
</tr>
<tr>
<td>CH$_3$(CH$_2$)$_3$OH</td>
<td>1.40</td>
<td>0.29</td>
</tr>
<tr>
<td>(CH$_3$)$_2$CHCH$_2$OH</td>
<td>1.40</td>
<td>0.16</td>
</tr>
<tr>
<td>CH$_3$COO$_2$H$_5$</td>
<td>1.50</td>
<td>0.14</td>
</tr>
<tr>
<td>C$_2$H$_5$COC$_2$H$_5$</td>
<td>1.50</td>
<td>0.31</td>
</tr>
<tr>
<td>CH$_3$(CH$_2$)$_4$OH</td>
<td>1.60</td>
<td>0.81</td>
</tr>
<tr>
<td>CH$_3$CH$_2$CH$_2$COCH$_3$</td>
<td>1.70</td>
<td>0.31</td>
</tr>
<tr>
<td>CH$_3$COOCH$_2$C$_2$H$_5$</td>
<td>2.00</td>
<td>0.66</td>
</tr>
<tr>
<td>C$_2$H$_5$COO$_2$H$_5$</td>
<td>2.00</td>
<td>0.66</td>
</tr>
<tr>
<td>(CH$_3$)$_2$CHCOO$_2$H$_5$</td>
<td>2.20</td>
<td>1.05</td>
</tr>
</tbody>
</table>
QSAR and log P

• Isonarcotic Activity of Esters, Alcohols, Ketones, and Ethers

\[ \log(1/C) = 0.73 \log P + 1.22 \]
\[ R^2 = 0.7767 \]
\[ R = 0.881 \]
\[ n = 20 \]
Isonarcotic Activity

• Esters, Alcohols, Ketones, and Ethers
  \[ \log(1/C) = 0.73 \log P + 1.22 \]
  \[ n = 20 \quad r = 0.881 \]

• subset of alcohols (overfitted):
  \[ \log(1/C) = 1.49 \log P - 0.10 \left( \log P \right)^2 + 0.50 \]
  \[ n = 10 \quad r = 0.995 \]
Calculation of clogP

LogP for a molecule can be calculated (clogP) from a sum of fragmental or atom-based terms plus various corrections.

\[ \text{logP} = \sum \text{fragments} + \sum \text{corrections} \]

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Class} & \text{Type} & \text{Log(P) Contribution} & \text{Description} & \text{Value} \\
\hline
\text{FRAGMENT} & \# 1 & 3,5-pyrazolidinedione & -3.240 \\
\hline
\text{ISOLATING} & \text{CARBON} & 5 & \text{Aliphatic isolating carbon(s)} & 0.975 \\
\hline
\text{ISOLATING} & \text{CARBON} & 12 & \text{Aromatic isolating carbon(s)} & 1.560 \\
\hline
\text{EXFRAGMENT} & \text{BRANCH} & 1 & \text{chain and 0 cluster branch(es)} & -0.130 \\
\hline
\text{EXFRAGMENT} & \text{HYDROG} & 20 & \text{H(s) on isolating carbons} & 4.540 \\
\hline
\text{EXFRAGMENT} & \text{BONDS} & 3 & \text{chain and 2 alicyclic (net)} & -0.540 \\
\hline
\text{RESULT} & & & & 2.11 \\
\hline
\end{array}
\]

\[
\text{clogP for windows output} \\
\text{C: 3.16 M: 3.16 PHENYLBUTAZONE} \\
\text{clogP 3.165}
\]
What else does logP affect?

- Binding to enzyme / receptor
- Aqueous solubility
- Binding to P450 metabolising enzymes
- Absorption through membrane
- Binding to blood / tissue proteins – less drug free to act
- Binding to hERG heart ion channel - cardiotoxicity risk

So log P needs to be **optimised**
ADMET
Absorption, Distribution, Metabolism, Excretion (Elimination), Toxicity
Bioavailability

Absorption

Permeability

Lipophilicity

Hydrogen Bonding

Liver Metabolism

Transporters

Solubility

Molecular Size/Shape

Flexibility

Gut-wall Metabolism
Metabolism of Xenobiotics

A xenobiotic is a chemical compound which is found in an organism but which is not normally produced or expected to be present in it.

- natural compounds
- pollutants
- drugs

Detoxification and elimination of xenobiotics is mainly done in liver.

Phase I (mainly monooxygenases) convert hydrophobic chemicals into hydrophilic chemicals.

Phase II (UGTs, SULTs, GSTs, NATs) further convert these products into amphiphilic anionic conjugates.

Phase III (transporters) export products out of the liver.
Prediction of ADMET Properties

• Requirements for a drug:
  – Must bind tightly to the biological target \textit{in vivo}
  – Must pass through one or more physiological barriers (cell membrane or blood-brain barrier)
  – Must remain long enough to take effect
  – Must be removed from the body by metabolism, excretion, or other means

• ADMET - prior 2010 main cause of clinical trial failure!
Lipinski Rule of Five (Oral Drug Properties)

• Poor absorption or permeation is more likely when:
  – MW > 500
  – LogP >5
  – More than 5 H-bond donors (sum of OH and NH groups)
  – More than 10 H-bond acceptors (sum of N and O atoms)
ADMET Descriptors Calculation Tools

**PreADMET**  http://preadmet.bmdrc.org/
- **Molecular Descriptors Calculation** - 1081 diverse molecular descriptors
- **Drug-Likeness Prediction** - Lipinski rule, lead-like rule, Drug DB like rule
- **ADME Prediction** - caco-2, MDCK, BBB, HIA, plasma protein binding and skin permeability data
- **Toxicity Prediction** - Ames test and rodent carcinogenicity assay

**SPARC Online Calculator**  http://ibmlc2.chem.uga.edu/sparc/
- SPARC on-line calculator for prediction of pKₐ, solubility, polarizability, and other properties; search in the database of experimental pKa values is also available

**Daylight Chemical Information System**  www.daylight.com/daycgi/clogp
- Calculation of log P by the CLOGP algorithm from BioByte; also access to the LOGPSTAR database of experimental log P data

Chemicalize, ChemSpider, PubChem, ZINC ...