Application of Computer in Chemistry SSC 3533

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS

Prof. Mohamed Noor Hasan
Dr. Hasmerya Maarof
Department of Chemistry

QSAR

- A quantitative structure-activity relationship (QSAR) correlates measurable or calculable physical or molecular properties to some specific biological activity in a form of an equation.

$$A = k_1 D_1 + k_2 D_2 + k_3 D_3 + k_n D_n$$

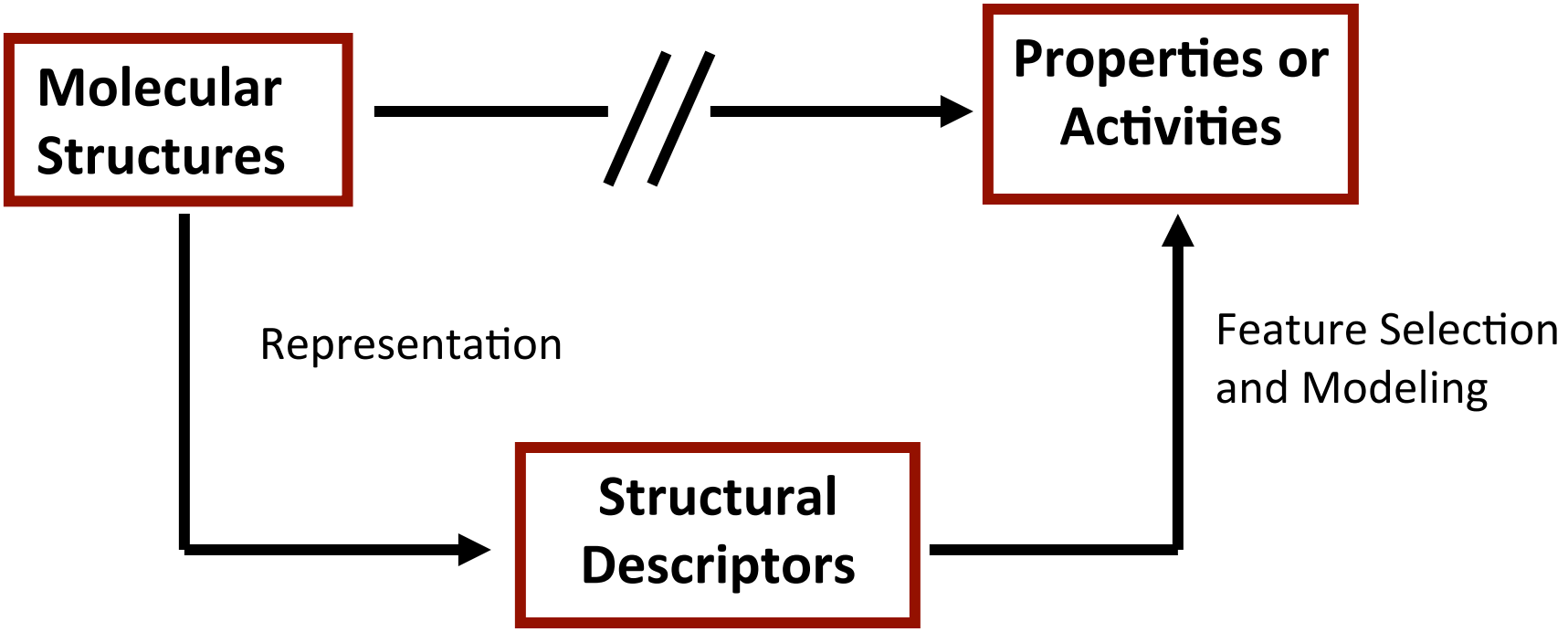
A – Biological activity

D – structural properties (descriptors)

k – regression coefficient

- Once a valid QSAR has been determined, it is possible to predict the biological activity of new compounds
- Sometimes the activity can be physical properties - QSPR

The QSAR Approach



The Motivation

- Main objectives of QSAR studies:
  - To understand the mechanism of the biological activity. Sometimes the type of descriptors selected in the model provides a clue to the mechanism
  - To predict activity of unknown compounds. This could save cost and time on expensive biological tests on animals.

History of QSAR

- The first application of QSAR is attributed to Hansch (1969),
- Developed an equation that related biological activity to electronic characteristics and the hydrophobicity

$$\log (1/C) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

C = minimum effective dose

P = octanol - water partition coefficient

σ = Hammett substituent constant

k_x= constants derived from regression analysis

Hansch's Approach

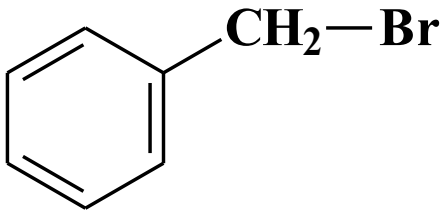
- Log P is a measure of the drug's hydrophobicity, which was selected as a measure of its ability to pass through cell membranes.
- The log P (or log $P_{o/w}$) value reflects the relative solubility of the drug in octanol (representing the lipid bilayer of a cell membrane) and water (the fluid within the cell and in blood).
- Log P values may be measured experimentally or, more commonly, calculated.

Calculating Log P

$$\text{Log P} = \text{Log K}_{(o/w)} = \text{Log } ([X]_{octanol}/[X]_{water})$$

- Most programs use a group additivity approach:

1 Aromatic ring	0.780
7 H's on Carbon	1.589
1 C-Br bond	-0.120
1 alkyl C	0.195

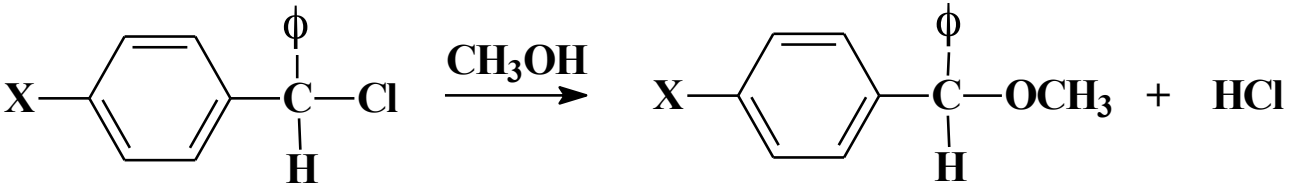


Sum = 2.924 = calc. log P

- Some use more complicated algorithms, including factors such as the dipole moment, molecular size and shape.

Hansch's Approach...

- The Hammett substituent constant (s) reflects the drug molecule's intrinsic reactivity, related to electronic factors caused by aryl substituents.
- In chemical reactions, aromatic ring substituents can alter the rate of reaction by up to 6 orders of magnitude!
- For example, the rate of the reaction below is ~10^5 times slower when X = NO_2 than when X = CH_3



Hammett Equation

- Hammett observed a linear free energy relationship between the log of the relative rate constants for ester hydrolysis and the log of the relative acid ionization (equilibrium) constants for a series of substituted benzoic esters & acids.

$$\log (k_x/k_H) = \log (K_x/K_H) = \rho\sigma$$

- He arbitrarily assigned ρ, the reaction constant, of the acid ionization of benzoic acid a value of 1.

Hammett Plot

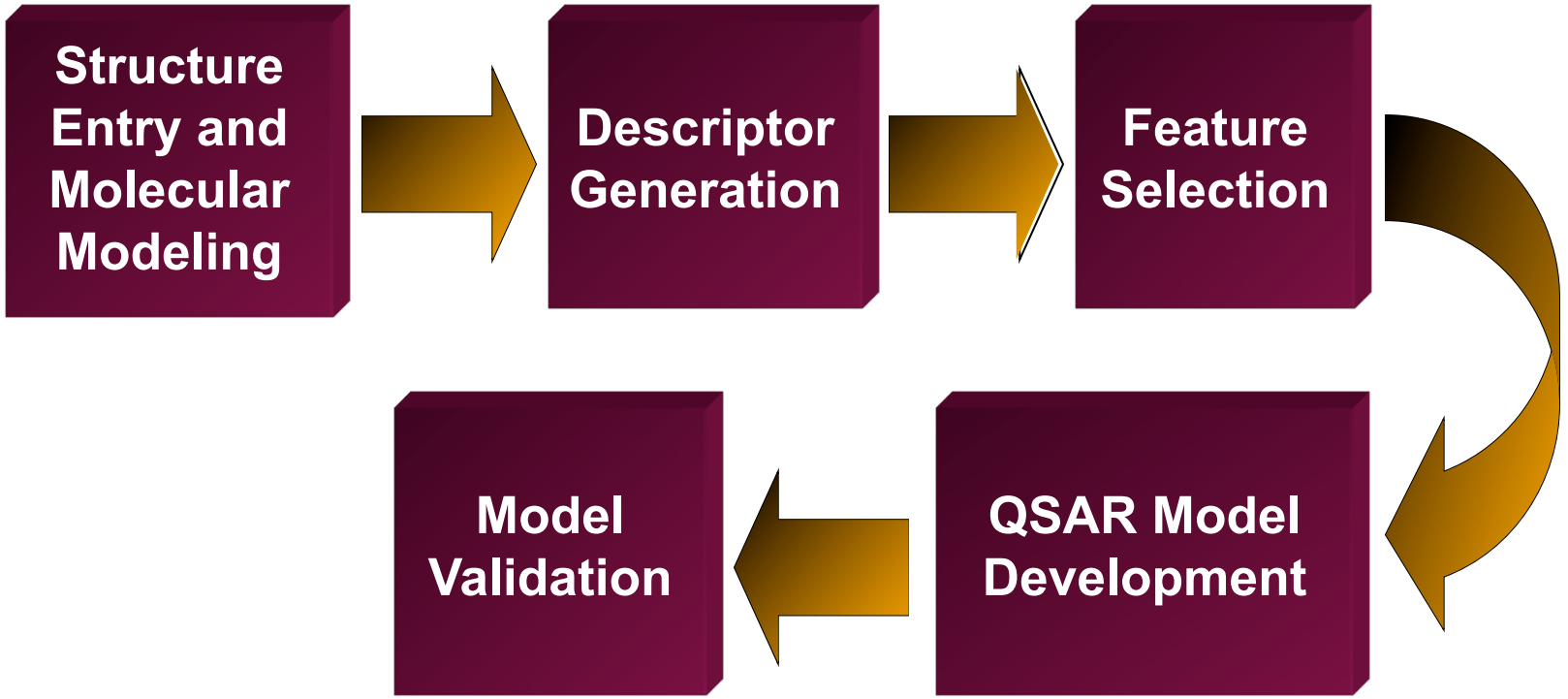
- Aryl substituent constants (σ) were determined by measuring the effect of a substituent on a reaction rate (or K_{eq}). These are listed in tables, and are constant in widely different reactions.
- Reaction constants (ρ) for other reactions may also be determined by comparison of the relative rates (or K_{eq}) of two differently substituted reactants, using the substituent constants described above.
- Some of these values (σ and ρ) are listed on the following slide.

Molecular Properties in QSAR

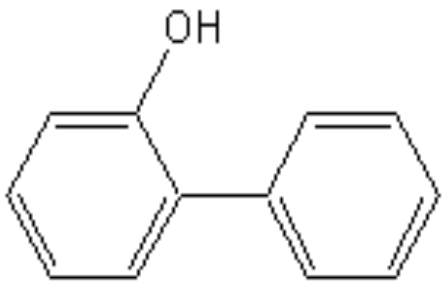
- Many other molecular properties have been incorporated into QSAR studies; some of these are measurable physical properties, such as:

  - density
  - ionization energy
  - boiling point
  - refractive index
  - molar refractivity
  - molecular weight
  - dipole moment (μ)
  - Surface area
  - Molecular volume
  - polarizability

The QSAR Methodology

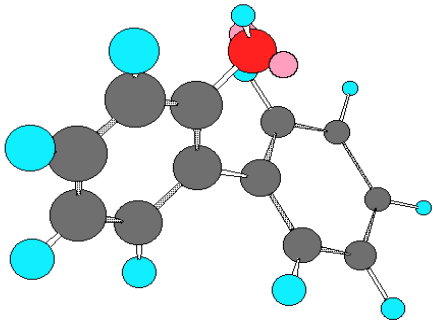


Structure entry and modeling



2-D Structure

Molecular Modeling



3-D Model

Descriptors

- Descriptors are generated numerical quantities that represent molecular structures
- Types of descriptors:
  - Topological
  - Geometrical
  - Electronic
  - Calculated physical properties

Topological Descriptors

- The structures of organic compounds can be represented as connection tables and graphs. Rules of graph theory can be applied to form topologial indices.
- Examples:
  - atom counts
  - ring counts
  - molecular connectivity indices
  - substructure counts
  - molecular weights

Electronic Descriptors

- Electronic descriptors are calculated to encode aspects of the structures that are related to the electrons.
- Examples of electronic descriptors:
  - Partial atomic charges
  - Sigma charges
  - HOMO or LUMO energies
  - Dipole moment

Geometrical Descriptors

- Geometric descriptors are calculated to encode the 3-D aspects of the structures
- Examples of geometrical descriptors:
  - moments of inertia
  - solvent-accessible surface area
  - Molecular volume
  - length-to-breadth ratios
  - shadow areas

Calculated Property Descriptors

- Physical properties which are estimated using computer method
- Log P
- Refractive index
- Polarizability

Descriptor Selection

- After descriptors have been calculated for each compound, the set must be reduced to a set of descriptors which is as information rich but as small as possible.
- Objective feature selection uses only the independent variables, and descriptors to discard are identified by
  - pairwise correlations
  - tests of identical values
- Subjective feature selection, which does use the dependent variable values, is applied to further reduce the descriptor set

QSAR Model Development

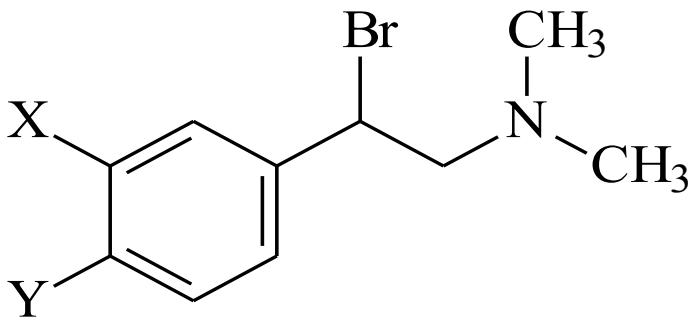
Statistical/computer methods used for model development:

1. Multiple Linear Regression (MLR)
   - Limitations: 5 to 1 ratio, no correlation among descs.
3. Partial Least Square Regression (PLS)
4. Neural Networks

Model Validation

- Developed QSAR models must be validated to ensure that the models will be useful
- Validation of the model is accomplished by predicting the activity of compounds in the external prediction set
- The errors produced in the predictions should be comparable to those achieved for the training set

Example of a QSAR



Anti-adrenergic Activity and Physicochemical Properties of *3,4-* disubstituted N,N-dimethyl-α-bromophenethylamines

π = Lipophilicity parameter
σ^+ = Hammett Sigma$^+$ (for benzylic cations)
$E_{s(meta)}$ = Taft's steric parameter

Example of a QSAR...

QSAR Equation a: (using 2 variables)

$$\log (1/C) = 1.151\,\pi - 1.464\,\sigma^{+} + 7.817$$

$$(n = 22;\ r = 0.945)$$

QSAR Equation b: (using 3 variables)

$$\log (1/C) = 1.259\,\pi - 1.460\,\sigma^{+} + 0.208\,E_{s(meta)} + 7.619$$

$$(n = 22;\ r = 0.959)$$