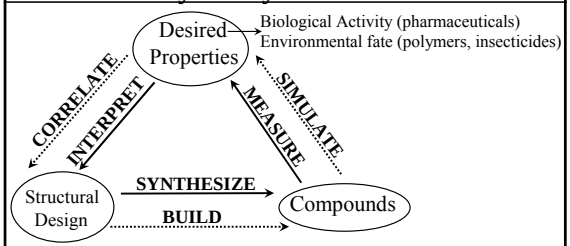
Session 8

QSAR/QSPR

Chemistry Today: A Different View



Old way: Synthesize, measure, repeat→ time & $$$

- Involved trial and error

New way: Build, calculate, repeat→ Save time & $$$

- Easier to do many steps computationally

New Tools

- Instead of trial and error, use computer to help
  - Models relate properties of interest to structure
  - New structures can be designed *computationally*
  - Synthesize only those with an increased probability of exhibiting the desired properties

→ A mathematical relationship must be found between the structures and the properties

- Once established, can "test" new structures
- Experimental work is more focused so chemists can be more productive

Definitions

Quantitative Structure Activity Relationship

- QSAR: Relates molecular structure to pharmacological activity

Quantitative Structure Property Relationship

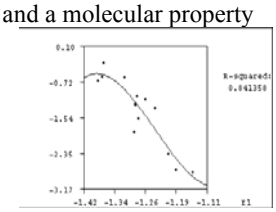
- QSPR: Relates structure to physical properties (boiling point, dipole moment, etc.)

→ Terms are sometimes used interchangeably

- Cheminformatics
  - Includes the design, creation, organization, management, retrieval, analysis, dissemination, visualization, and use of chemical information
    - Includes computational chemistry and QSAR/QSPR

Goal of QSAR/QSPR

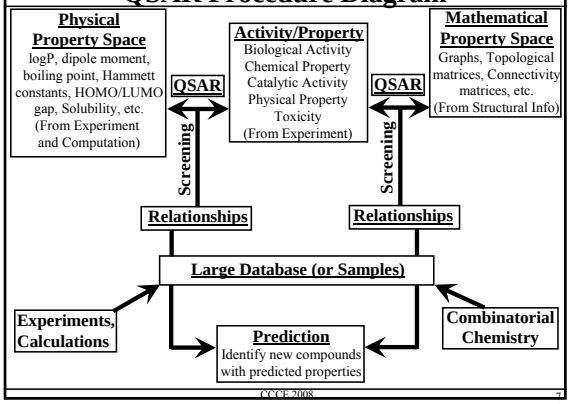
- Find a quantitative relationship between structure and observed activity or property
  - Empirical
- Multi-variant statistical predictions used
- Hope to find a good statistical correlation between an activity and a molecular property



Steps

1. Compile list of compounds with *experimentally determined* properties
   - Ideally want 10 times the number of compounds as parameters
2. Obtain geometries
3. Compute molecular descriptors
4. Calculate correlation coefficients
5. Perform a curve fit
   - Perhaps improve fit by including/excluding parameters

QSAR Procedure Diagram



Why QSAR/QSPR?

- Drug Design
  - Costs > $600M to bring a new drug to market
  - Patent lifetime is limited (generic drugs)
    - Wait until late as possible to file a patent
  - Synthesis/purification of compounds is expensive and time-consuming
  - Want to "zero in" on the best candidates quickly
  - QSAR provides a method for focusing on the group of most promising drug candidates
    - Now, spend time on only those compounds more likely to go forward in the process

Pros and Cons

- Pros: Detailed mechanistic understanding not required
  - Fast & easy screening of a large number of compounds
  - Can provide answers as to what type of molecular structures warrant further investigation
- Cons: Less insight than a mechanistic model
    - If mechanism were known, better candidates could be identified
  - Need experimental data to calibrate the regression line

Some Common Molecular Descriptors

→**Descriptor**: One or more things that explain properties (chemical, physical, biological) in a group of analogs

Constitutional: Molecular weight, number of rings, # of H atoms, # of heteroatoms, functional groups

Topological: Connectivity indices

Electrostatic: Polarizability, dipole moment

Geometrical: Molecular volume, surface area, shape indices

Statistical Mechanical: Vibrational frequencies, thermochemical parameters

Quantum Chemical: HOMO/LUMO energies, reactivity indices

Some Quantum Chemical Descriptors

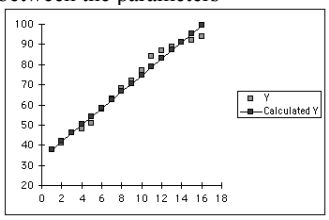
- Easier to determine than experimental descriptors
  - Atomic charges, frontier orbital densities
  - Molecular orbital energies: HOMO and LUMO
  - Susceptibilities and Superdelocalizabilities
    - Nucleophilic, Electrophilic, and Radical
  - Atom-atom polarizabilities
  - Molecular polarizabilities
  - Dipole (and higher order) moments
  - Polarity indices
  - Total energy

Quantum Chemical Considerations

- QM calculations offer an attractive source of new descriptors
  - Most researchers now include such descriptors since they are easily calculated
- Errors due to approximations:
  - Will tend to cancel out, since errors should be transferable within structurally similar molecules
- QSAR/QSPR is an aid, but cannot substitute for chemical intuition and experience
  - For example, structures with certain functional groups are not good drug candidates

Statistical Techniques

- Regression methods
  - Linear: $y = mx + b$
  - Polynomial: $y = c_0 + c_1x + c_2x^2 + c_3x^3 + ...$
  - Correlation coefficients: Predictors of the "goodness of fit" between the parameters



Some Examples from the Literature

- pK_a predicted from atomic charges on acidic H's
- Electrophilic aromatic substitutions predicted from activation hardness
- Octanol/water partition coefficents predicted from atomic charge densities
- Gas phase acidity of substituted benzoic acids predicted from AM1 calculated net atomic charges on O atoms
- Mutagenicity of quinolines predicted from LUMO energies
- GC retention indices predicted from multiple quantum mechanical descriptors

Tools

- CAChe, using Project Leader, automates the calculations, will perform statistical analyses, draw graphs, etc., etc.
- Spartan '06 includes a spreadsheet, with capabilities similar to that of CAChe
- Chem3D-Ultra provides the ChemSAR interface for Excel
- HyperChem will also output data to Excel

Lab Session

- QSPR investigation of the reaction rates for atmospheric oxidation (by hydroxyl radical) of hydrofluorocarbons (HFCs) and hydrofluoroethers (HFEs)
- Tool = Spreadsheet
  - Make sure the Data Analysis Tools in Excel are 'activated'