Chem 162/262 (2005) Assignment #6:

# QSAR

1. In your own words, explain how each of the statistical techniques listed below differs from the ordinary linear least-squares regression that you are already familiar with. Briefly explain how each of these techniques is relevant in the QSAR-based drug design:
   a. *Multiple linear regression*
   b. *Multiple non-linear regression*
   c. *Principal component analysis*

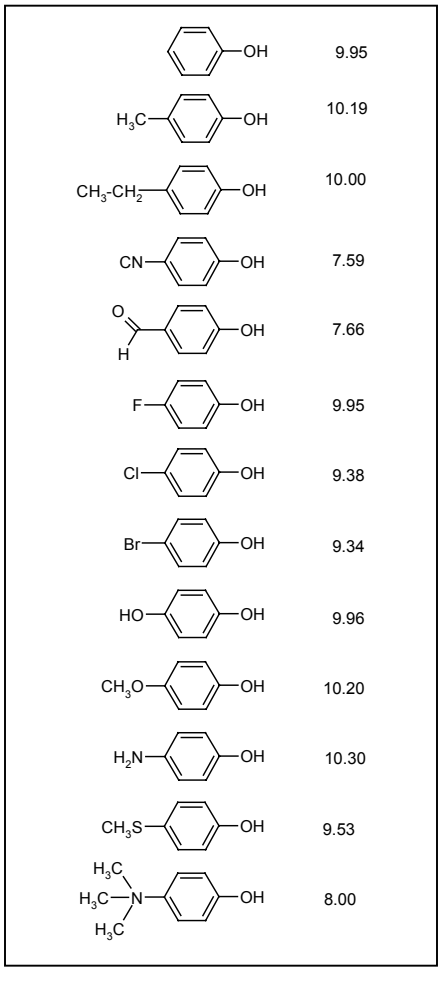
2. QSAR analysis of a series of drugs indicated that the potency of drug is well described by the Hansch equation $\log\frac{1}{C} = -k\pi^2 + k'\pi + \rho\sigma + k''$. Based on this QSAR analysis, a cyano derivative was chosen as it displayed the highest potency. However, toxicology studies with the cyano derivative established that it is toxic by a mechanism that does not involve the therapeutic target protein. Suggest a replacement for the cyano substituent that would be almost as potency as the cyano derivative without being toxic.

3. Table 1 lists Hammett sigma constants for various substituents. These values were determined by measuring dissociation constants of meta- and para-substituted benzoic acids in water. Table 2 lists $pK_a$ values of several para-substituted phenols in water.

Table 1

| **Substituent** | **$\sigma_{meta}$** | **$\sigma_{para}$** |
|---|---|---|
| H | 0.00 | 0.00 |
| $CH_3$ | -0.07 | -0.17 |
| $CH_2CH_3$ | -0.07 | -0.15 |
| CN | 0.56 | 0.66 |
| CHO | 0.35 | 0.42 |
| $COCH_3$ | 0.38 | 0.50 |
| F | 0.34 | 0.06 |
| Cl | 0.37 | 0.23 |
| Br | 0.39 | 0.23 |
| OH | 0.12 | -0.37 |
| $OCH_3$ | 0.12 | -0.27 |
| $NH_2$ | -0.16 | -0.66 |
| S-CH3 | 0.15 | 0.00 |
| $N(CH_3)_3^+$ | 0.88 | 0.82 |

Table 2



**Answer the following questions:**

1. Which benzoic acid derivative in Table 1 has the lowest p$K_a$? Briefly explain your reasoning.

2. Rewrite the Hammet relationship to give an expression for p$K_a$ of 4-acetophenol in terms of p$K_a$ of phenol and appropriate sigma and rho constants.

3. Based only on the data provided, calculate the estimate for the p$K_a$ value of 4-acetophenol. Show details of your work; include any graphs and calculations that you might need to answer the question.

4. Comment on reliability of your p$K_a$ estimate based on statistical analysis of your data.

5. Using molecular modeling programs create 3D structures of phenol, *p*-methylphenol, *p*-ethylphenol, *p*-fluorophenol, *p*-chorophenol, and *p*-bromophenol. Programs that will work include Maestro or SYBYL in the SGI lab or PyMOL (get your free copy from http://pymol.sourceforge.net/ ) on your PC. When using Maestro, save structure files as SYBYL MOL2 files, when using PyMOL save the structure files as MOL file. Provide images of each structure you created. See hints below for how to accomplish this with PyMOL.

6. Using any software capable of predicting molecular properties based on molecular structure, calculate the following molecular descriptors for each molecule for which you created 3D structure. A suitable program is DRAGON, which is freely available from http://www.talete.mi.it/dragon.htm.

   a. Octanol –water partition coefficient (Moriguchi logP)
   b. Molar refractivity (Ghose-Crippen MR)
   c. Sum of atomic polarizabilities
   d. Radius of gyration

Organize your data in the following table:

| | **Log P** | **Molar refractivity** | **Sum of polarizabilitities** | **Radius of gyration** |
|---|---|---|---|---|
| **phenol** | | | | |
| ***p*-methylphenol** | | | | |
| ***p*-ethylphenol** | | | | |
| ***p*-fluorophenol** | | | | |
| ***p*-chorophenol** | | | | |
| ***p*-bromophenol** | | | | |

7. Which molecule is most lipophilic? Would you have predicted this based on your chemical intuition?

**PyMOL editing hints**

The current version of PyMOL (0.97) has rudimentary molecular editing capabilities accessible via the **Build** menu. The program allows adding atoms and functional groups to a selected atom, remove selected atoms, or change torsional angles (rotation around bonds). Basic editing is best done while the mouse is in the **(3-Button) Universal Cycle** mode while rotations around bonds need mouse in the **(3-Button) Editing Mode**.

To create a new molecule, it is best to use one of the templates provided in the **Fragment** or **Residue** submenu of the **Build** menu. For example, you can use tyrosine (select Build -> Residue -> Tyrosine) as a template for all the substituted phenols. The atom marked with banded sphere is ready for editing.

To create *p*-ethylphenol from tyrosine, you need to delete atoms that are not present in *p*-ethylphenol. You want to delete four atoms: the amide hydrogen and amide nitrogen, the carbonyl oxygen and the carbonyl carbon. To delete atoms you need to select them first. To select atoms while in the universal cycle mode, select one atom at a time by doing *Ctrl-Right click* on the desired atom. To delete this atom, select **Remove** from the **Build** menu, or press *Ctrl-D*. Last, you need to fix the methyl group; PyMOL can do it automatically when you select this carbon atom and then select **Cycle Bond Valence** from the **Build** menu.

To save the molecular structure as MOL file, select **Save Molecule** from the **File** menu, and make sure that the molecule name (e.g. tyr is highlighted). Select MOL as file format and give a unique MOL file name for each molecule saved. PyMOL saves molecules as MDL molfiles (not to be confused with SYBYL molfile)

**Project development**.

In this part of the project development you are expected to come up with a lead compound that could be developed into a drug to cure the disease you are working at. Your choice of approach depends on what is known about your target. If the structure of the target is known, you may want to take the structure-based approach. This could involve visualization of the active site, preparation of a small virtual library of compounds that are similar to known ligands or inhibitors and possibly carrying out docking. If the structure of the target is unknown but the target is an enzyme, deduce the structure of the transition state and design molecules that resemble the transition state. You can carry out conformational analysis and electrostatic potential calculations to compare your lead candidates with the transition state. If your disease is not an obscure one, there is probably good amount of published potency data and qou can do QSAR analysis to guide your search.

If you cannot do any of the above because very little is known about your target or the required data is not available use your (bio)chemical intuition to come up with a lead, then describe what steps need to be taken in order to characterize the interactions between your drug and the target.