Property-Based Design Methodology I: Parameters Influencing Solubility

Robert DeWitte ACD/Labs, Toronto, ON, Canada

Fedor Gorohov, Eduard Kolovanov ACD/Labs, Moscow, Russian Federation

Abstract

Many biological barriers between dose and exposure have been shown to correlate with the log*D* (pH) of a compound. Such is certainly the case for aqueous solubility (see below), permeability, plasma protein binding, and tissue partitioning. To a degree, and for certain classes of compounds, the same is true of blood brain barrier permeability, though in this case, it is generally observed that polar surface area is the dominant molecular physical property.

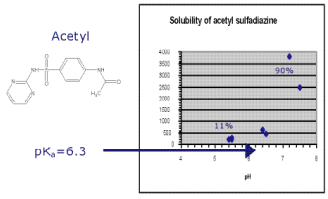
Given these observations, one would like to ask the question: how can I change my compound in order to improve the property of interest? Specifically, how can I decrease the log*D* value at a specific pH?

This series of posters develops an approach for addressing these questions using the drug acetyl sulfadiazine as an example. We begin by reviewing the evidence supporting the claim that solubility correlates with log*D*, and use this to develop working hypotheses for improving solubility. These hypotheses are then translated into a general approach to improve the drug-like properties of medicinal compounds.

Solubility Correlates with Log*D*: A Theoretical Argument

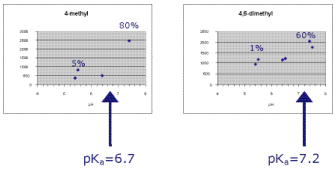
This dependence can be argued as follows: consider solubility as a coupled pair of phenomena, where the first transition moves from an ordered solid state to a disordered liquid state. One can consider this intermediate state as either a super-cooled liquid, or a solution of solute in itself. The second transition moves from this ideal solution to a solution of solute in water. Therefore, the first transition corresponds to isothermal melting, whereas the second corresponds to partitioning. The magnitude of the isothermal melting contribution to aqueous solubility will depend on the architecture of the solid state, which is largely a property of the sample, rather than the isolated chemical structure (e.g., crystalline vs. amorphous, polymorphs, salts, hydrates, etc. will have a pronounced affect on this term). The magnitude of the second transition will depend on the partitioning between ideal solution and aqueous solution, and log*D* is a reasonable model for this. [Note that log*D* does not correctly take into account the influence of the solubilities of various salts, and so the solubility of the fully ionized form in the GI could be quite different from that modeled with log*D*]. Finally, the relative magnitude of these terms can be gleaned by an evaluation of Yalkowsky's model for solubility, where log*S*~-log*P*-0.01mp. In short, an increase of 100 degrees in melting point is required to decrease solubility by one order of magnitude, whereas this same degree of change can be accomplished with a one log unit change in lipophilicity.

Solubility Correlates with Log*D*: Experimental Evidence



Acetyl Sulfadiazine is a weak acid with pK_a of 6.3, having an intrinsic solubility of 146 µg/mL. Low solubilities are therefore observed at relatively low pHs, such as 5.5. In this case, the compound is only 11% ionized. At higher pHs, however, the solubility increases greatly: at pH 7.4 the compound is 90% ionized and the solubility is in the range of 2.5-3.6 mg/mL.

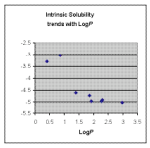
Because the pH of 5.5 is more relevant than 7.4 for the purposes of achieving appreciable oral absorption, we will consider ways to increase the solubility at 5.5 by making small changes to the chemical structure of acetyl sulfadiazine.



Further experimental data for related compounds (methyl and dimethyl substitution on the pyrimidine ring) provide further illustration of the influence of pK_a on the pH profile of solubility. In the case of 4-methyl substitution, the pK_a has raised to 6.7 from 6.3. This small change is sufficient to alter the fraction ionized significantly—down from 11% to 5% at pH 5.5 and down from 90% to 80% at pH 7.4. This effect is more extreme in the case of the dimethyl analogue: a further half log unit increase in pK_a reduces the ionized fraction to 1% and 60% at pHs 5.5 and 7.4, respectively.

As a result, the increase in solubility as a function of pH is attenuated for these compounds. This observation provides a working hypothesis for identifying members of this structural class with improved solubility: if the pK_a can be lowered, so that a greater extent of ionization can be achieved at pH 5.5, then the solubility observed at that pH will be increased due to further ionization.

R-group	Log*P*	S(intrinsic)
H	0.4	146.2
4-methyl	0.9	275.7
4-ethyl	1.4	7.8
4-propyl	1.9	6.4
4-ethyl, 5-methyl	1.9	3.6
4-methylpropyl	2.3	3.8
4-pentyl	3.0	4.4
4-phenyl	2.3	3.6



Turning our attention now to the intrinsic solubilities of analogues within this chemical series, we can see that the logarithm of the intrinsic solubility trends negatively with log*P*. Recall that intrinsic solubility refers to the solubility of the neutral form of the compound. In terms of the pH profile of the solubility, the intrinsic solubility defines the plateau at the bottom of the curve. (Note that zwitterions do not manifest their intrinsic solubilities because there is no pH at which the neutral form is in 100% abundance).

Caveats

The above assertion that solubility correlates with log*D*, has its limitations. Several mitigating factors can obfuscate this trend:

1. The upper limit of solubility differs from the lower limit of log*D*. While the lower limit of log*D*(pH) is defined by ion-pair partitioning, this is not the case for the upper limit of solubility.
- **a.** Where there is no salt counter-ion present in solution, the upper-bound on ion solubility is limited by the ability of the solvent to support a very high ionic strength.
- **b.** Where there is ample counter-ion in solution, the solubility is limited by the salt product formed with this solution.

2. The lower limit of solubility depends on more than log*P*. Clearly, other factors, such as size, flexibility, symmetry, and other supramolecular factors, contribute strongly to the intrinsic solubility of a compound.

A General Procedure for Designing Improved Properties

In general, we could describe a suitable process in the following algorithmic terms

1. Consider your SAR, and its relationship to molecular structure.
- **a.** Ideally, there will be some region of the structure for which the SAR is relatively flat with respect to substitution. In such a case, attention will therefore be focused on how to adapt these parts of the molecule to achieve the maximal improvement in physicochemical properties. This is generally best accomplished through the addition or substitution of a small substituent.
- **b.** In some cases, however, there is no such region of flat SAR, and you are limited to working within the topology of the parent compound. In this case, heterocycle substitution may be the only recourse available. This case is taken up as the last example of this application where it is shown that heterocycle replacement can be as effective as the addition or substitution of a small substituent.

2. Predict the log*D*(pH) curve. Note the predicted values at the key pH, and also the shape of the curve around that point. One of three cases will be present:
- **a.** The molecule contains a pK_a within two log units of the pH of interest, which, if moved closer, will reduce the log*D* dramatically.
- **b.** The molecule contains a pK_a within two log units of the pH of interest, which, if moved further will reduce the log*D* dramatically.
- **c.** The molecule does not have a pK_a near the pH. In this case, introducing an ionizable group with a suitable pK_a will reduce the log*D* dramatically.

3. For cases a & b, predict the pK_a s of the parent molecule and inspect the relationship between chemical structure and the pK_a to be moved. Look for a relationship between the electron withdrawal strength of substituents and/or heterocycles and the acid/base strength of the ionization centre. This is reported as a Hammett equation in the calculation protocol window of ACD/pK_a DB.

4. Note the sigma and Pi constants of any key substituents and/or heterocycles influencing the pK_a.

5. Consult the Substituents Databases within ACD/MedChem Advisor to identify suitable replacements.
- **a.** For Substituents, use "Neutral Substituents.CFD"
  - **i.** Query on a data range for the appropriate Sigma Constants,
  - **ii.** Restrict molecular weight to below 40 or so,
  - **iii.** Restrict Pi values to be less than zero,
  - **iv.** Limit the view to substituents that have appeared in at least one phase II compound.
- **b.** For Heterocycles, use "Heterocycles.CFD"
  - **i.** Begin with a substructure search to limit the search to heterocycles with similar topology to the one in the parent compound,
  - **ii.** Query on a data range for the appropriate Sigma Constants,
  - **iii.** Restrict Pi values to be less than zero,
  - **iv.** Limit the view to substituents that have appeared in at least one phase II compound,
  - **v.** Compute the pK_as of the heterocycle on the parent compound.
  - *In the case that the heterocycle includes the pK_a you are seeking to modify, query also on ranges of either acidic or basic pK_a to identify suitable substitutes.*
  - *In the case that the heterocycle does not include the pK_a you are seeking to modify, limit the ranges of acidic and basic pK_as of the heterocycles to those similar to the parent heterocycle.*
- **c.** Consider synthetic feasibility, molecular stability, and other factors
  - **i.** The features of ACD/ChemFolder will help you focus and limit your browsing.
- **d.** Select and assess key substituents. For each in turn,
  - **i.** Double-click in ChemFolder to bring the substituents into ChemSketch,
  - **ii.** Attach the substituents or heterocycle to the parent compound,
  - **iii.** Compute pK_a, log*D*(pH),
  - **iv.** Assess log*D* at the pH of interest.

6. In the case where you seek to introduce an acidic or basic centre, this can be achieved by consulting one of the following databases:
- **a.** "Heterocycles.CFD"
- **b.** "Acidic Substituents.CFD"
- **c.** "Basic Substituents.CFD"
- **d.** Note that you can query all of them simultaneously using the features of ACD/ChemFolder, and then work within the result set as a separate database.
- **e.** The above design principles can be applied similarly to identify a substituent or heterocycle that introduces the appropriate degree of ionization to your compound.

Note that it is always preferable to begin with experimental values for the compound, or a related compound, and use this to verify that the predictions from ACD/Log*D* Sol Suite are accurate, to within a reasonable confidence interval. In addition, it is a good idea to inspect the calculation protocol for log*P* and pK_a to determine if the algorithm is using secondary approximations (log*P* coefficients indicated as "Approximated", "Estimated", or "Assumed"), Ring-Breaking approximations (pK_a DB), or extensive correction factors in the pK_a calculation protocol. In any of these situations, or if the prediction error is not sufficiently low, it is recommended to invoke one of the system training capabilities, whereby experimental measurements are used to improve the calculation accuracy on related compounds. This is described in detail in the Application Note entitled "The Inner Workings of User Training."

In general, given two predicted values (**A** and **B**) and their confidence intervals (±a and ±b), the difference between the two of them (**B-A**) is only relevant if **B-A** is greater than the larger of **a** or **b**. Using this rule of thumb, you can be certain that the predicted values supporting the design approach illustrated in this application note will deliver recommendations that will be borne out to within rank order. This means that the molecular changes predicted to be improvements of the properties will indeed improve the properties, though the extent of the improvement may be greater or less than predicted.

Conclusions and Working Hypotheses

Nonetheless, the observations presented here provide a framework for making decisions about structural analogues:

1. The effect of ionization on solubility is to modulate the intrinsic solubility, providing a steep rise in solubility above the intrinsic level as a function of pH. Therefore, in considering structural changes within our chemical series, we need to select substituents that increase the fraction of ionization

2. The effect of lipophilicity is to reduce solubility. Therefore, the structural changes pursued in order to increase ionization should be done in such a manner as to not significantly increase the lipophilicity of the compounds.

90 Adelaide St. W., Suite 600
Toronto, Canada M5H 3V9
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com