A Primer on Molecular Similarity in QSAR and Virtual Screening

Part II – How reliable are experimental measurements (endpoints) in QSAR studies?

By Andreas Bender, PhD

1. Introduction

In the first part of this three-part primer on how to construct "better" QSAR models we have seen that not always the more complex descriptors should be chosen. Indeed, in most cases the simplest descriptors that are still able to describe a certain phenomenon should be chosen to generate a model. Reasons are, among others, the risk of over fitting if too many variables are present; and also interpretation problems of more complex descriptors, rendering rational improvement of a compound difficult.

In the current work, we will investigate reliability of the data we are attempting to model. Naively, one might just take the numbers from experimental measurements and apply some kind of statistical or machine learning approach in order to model the data points. In the following paragraphs we will see that in the real world, life is not so easy. Experimental data ("endpoints") depend heavily on the conditions under which they were measured, making the comparison of data as difficult as the creation of models based on data from different sources. We will discuss two different endpoints often used for quantitative modeling, namely solubility and bioactivity. The aim here is not to provide a comprehensive guide to experimental measurements, but to sharpen your eye for problems associated with any kind of experimental data points used for modeling.

2. Endpoint Measurements

a) Solubility

While solubility seems to be a trivial property of matter, once the attempt is made to produce reliable (and reproducible!) measurements disillusionment can occur[1]. Even for drugs on the market, such as diclofenac, published solubilities varied by a factor of around 100 (!)[2]. Commonly, two different types of measurements are used, called thermodynamic and kinetic solubility measurements. Both names are a little misleading, since "thermodynamic" measurements imply waiting until an equilibrium state between solid and solubilized phase has been achieved, a process, which is, in practice, stopped after a certain time (often 24 hours). "Kinetic" measurements are performed by adding a DMSO solution of the compound to the solvent of interest. While the speed of solution plays a role in these "kinetic" measurements, their nature is actually based on precipitation (instead of solubility) under particular conditions (see below)[1].

Which problems do we face when measuring "thermodynamic" solubilities? Firstly, the crystal form of the molecule has a huge influence on solubility, as shown recently for the example mentioned above, diclofenac[2]. During measurement, crystals may not convert into the most stable form, giving apparent solubilities of a particular salt form instead of the true solubility of the most stable form. Also, different ionization forms of a molecule change solubility quite considerably – the value of solubility is not meaningful without knowing the pKa of an acid or base as well as well as the protonation state and pH at equilibrium. In particular, attention should be paid to regions where the solubility is strongly pH-dependent. Often buffers are used during

© 2007 Strand Life Sciences Pvt. Ltd.

solubility measurements, either to keep conditions such as pH constant or to mimic certain environments such as the gut. Here the buffer as well as the final concentration of solid and solubilized form need to be documented properly, since otherwise the conditions of the experiment are not completely defined.

Even if attention to all the above points is paid, in practice often not enough crystal might be available for thermodynamic solubility measurements (which is usually in the order of 1mg of compound). In such cases, "kinetic" solubility measurements are often used which are based on the precipitation of compounds. The substance is first dissolved in DMSO as a near-perfect solvent, which is added to the solvent of interest (e.g. water) in which solubility has to be determined. After more and more compound is added, at a certain stage precipitation occurs, where the point of precipitation defines the solubility of the compound. Problems here are that the kinetic solubility is often determined on compounds that have not been purified sufficiently, often (due to standard thermodynamic laws) leading to higher solubilities than that of the pure substance. Also, supersaturation may occur since the compound was previously dissolved in a "perfect" solvent, potentially leading to overoptimistic solubility assumptions.

b) Bioactivity

One of the endpoints routinely modeled are bioactivity data, also giving rise to the area of QSAR (quantitative structure-activity relationships). There are at least three different classes of bioactivity endpoints:

1. Affinity assays measuring the drug-receptor dissociation constant (commonly abbreviated as K_i values),
2. Functional assays which measure, for example, inhibition constants (IC_{50} values) and
3. Cell-based assays that give some type of phenotypic readout.

While cell-based assays are generally more complex and have potentially more sources of error, they offer the huge advantage of taking intra- (and inter-) cellular signaling into account[3]. Affinity assays are a good choice for the quick measurement of binding energies, but they do not necessarily give much information about how effective a compound actually is. Two compounds with identical K_i values can (but don't have to) have vastly different IC_{50} values[4]. Functional assays are in some sense located between these two assay types – they are less complex than cell-based assays (at the expense of omitting the signaling network in the cell), but they provide some information about the efficacy of the compound against the target. (Under the assumption that the compound will reach the target in later stages such as animal models – an assumption that needs to be thoroughly validated separately.)

The error sources in biological assays are manifold. Solubility in the solvent may be limited, as shown in the previous paragraph, leading to different than expected concentrations in solution (and thus less likelihood of observing activity). In addition, "frequent hitters" may form aggregates in solution[5], leading to unspecific inhibition (with potentially misleading structure-activity relationships that can be inferred from those "activities"). DMSO may influence the target protein at too high a concentration, and evaporation of solvent leads to unwanted compound concentrations. Also, from the technical side, high-throughput readouts such as those based on optical detection, often do not show consistent sensitivity over the wells of a screening plate, leading to "edge effects" which need to be normalized in a second step.

3. Summary and Conclusions

While experimental measurements are the “gold standard” against which every hypothesis and model needs to be validated, we have shown that not every number obtained from an experiment can be taken at face value. For two examples, bioactivity and solubility data, we have given reasons why values may not be comparable between different laboratories (or even among different conditions used in the same laboratory). The conclusions are two-fold: Firstly, merging data points from different sources is a tricky process. And, secondly and most importantly, before any modeling attempt is started on a particular data set, the user is advised to have a thorough look at the underlying data quality – since every model can only be as good as the input data provided.

(To be followed by 3. Connecting descriptors and experimental measurements – model generation.)

References

1. Bhattachar, S. N.; Deschenes, L. A.; Wesley, J. A., Solubility: it's not just for physical chemists. *Drug Discov Today* **2006,** 11, (21-22), 1012-8.
2. Llinas, A.; Burley, J. C.; Box, K. J.; Glen, R. C.; Goodman, J. M., Diclofenac solubility: independent determination of the intrinsic solubility of three crystal forms. *J Med Chem* **2007,** 50, (5), 979-83.
3. Clemons, P. A., Complex phenotypic assays in high-throughput screening. *Curr Opin Chem Biol* **2004,** 8, (3), 334-8.
4. Knight, Z. A.; Shokat, K. M., Features of selective kinase inhibitors. *Chem Biol* **2005,** 12, (6), 621-37.
5. McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K., A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **2002,** 45, (8), 1712-1722.

© 2007 Strand Life Sciences Pvt. Ltd.