

An Introduction to Protein-Ligand Binding for BindingDB Users

Michael K. Gilson

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of
California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0736,
mgilson@ucsd.edu

December 15, 2010

BindingDB collects measured binding affinities of proteins and small molecules that bind via noncovalent interactions, in order to help scientists design new medications. This document provides background information on protein-small molecule binding, drug design, and the measurement of binding affinities. Subsequent documents will discuss the data that BindingDB collects and how BindingDB may be used.

Pathways and targeted molecules



Figure 1. Foxglove flowers. (Wikipedia: Digitalis)

Once, medications were preparations from natural matter, whose therapeutic values were discovered by trial and error. A famous example is the plant [foxglove](#) (Figure 1), which was reported by [William Withering](#) in the late 18th century to be a treatment for dropsy (edema). We now know that foxglove contains compounds, the [cardiac glycosides](#), that strengthen the heartbeat and thereby alleviate heart-failure, which is a most common cause of edema.

Now, basic research is providing an increasingly detailed description of how living systems work at the molecular level, and of the biomolecular pathways involved in many diseases. For example, Figure 2 illustrates pathways in which signaling proteins (e.g. PKA and Ras) are activated by extracellular molecules, ultimately regulating effector proteins that control functions like gene activity and cell proliferation. Knowledge of such pathways now guides the development of new drugs, “targeted molecules” that work by binding to a specific protein and thus blocking its activity. The statin drugs are a good example: people with higher cholesterol levels are at greater risk for heart attacks, and the statins lower cholesterol levels by binding and blocking a protein involved in the cholesterol synthesis pathway. However, some drugs work in other ways, such as by binding to DNA or RNA molecules instead of to proteins.

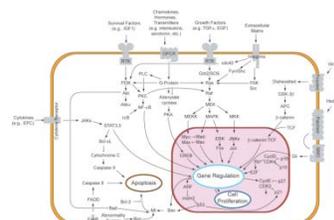


Figure 2. Highly simplified overview of cell signaling pathways. Outer rectangle: cell membrane. Pink rectangle: cell nucleus. (Wikipedia: Cell signaling.)

In order to serve as a drug, a compound must bind the targeted protein. It must also be sufficiently non-toxic and chemically and physically stable, and it must reach the targeted protein upon being administered by mouth (most convenient),

or by some other route (e.g., intravenous). Finding or designing a compound that meets all these requirements is an ongoing challenge to scientists trying to develop new medications. BindingDB helps in this process by collecting data from the medicinal and chemical genomics literature on what compounds bind what proteins, and how tightly they bind, making it possible to browse and query these data, and making it easier to use the data by providing them in machine-readable formats that can be downloaded for computational analysis.

The following sections provide more information regarding proteins and small molecules, how they bind, how their binding affinities are measured, and what data BindingDB collects.

Proteins: functions, sequences and folds

[Proteins](#) are varied molecules that execute many different cellular functions. [Enzymes](#), proteins that catalyze chemical reactions, are of particular medicinal interest. For example, the statins (above) bind and inhibit the cholesterol synthesis enzyme [HMG-CoA reductase](#); and the protease inhibitors used to treat HIV/AIDS bind and inhibit the viral enzyme [HIV Protease](#). Cell surface receptors (Figure 2) are another class of therapeutically relevant proteins. Floating in the cell membrane, they are able to bind specific extracellular signaling molecules, such as hormones, and respond by activating the appropriately corresponding intracellular processes, such as the signaling proteins in Figure 2. Many drugs work by binding to cell surface receptors and either activating them or blocking their activation by their natural signaling agents. For example, the [beta-blockers](#) bind beta-adrenergic cell surface receptors in place of epinephrine (adrenalin), thereby lowering heart-rate and blood pressure.

Chemically, a protein is basically a linear heteropolymer of the 20 major amino acids, so its chemical structure is defined primarily by its amino acid sequence. Protein sequences vary in length from tens to thousands with a mode probably in the hundreds. The enormous variety of protein functions is underlain by a combinatorial explosion of possible amino acid sequences. For example, there are $20^{200} \cong 10^{260}$ possible proteins of length 200. (By comparison, the number of atoms in the observable universe has been estimated at about 10^{80} . Clearly, only a miniscule fraction of all possible proteins could exist at any one time, or could ever even have existed in the history of life on Earth.) The chemical structure defined by a protein's sequence is furthermore subject to naturally occurring chemical modifications such as phosphorylation, the covalent addition of phosphate groups to some of the side-chains of serine or tyrosine amino acids in a protein. A protein's chemical properties can also be altered by changes in pH, due to the addition or removal of protons from its ionizable groups, such as histidine side-chains.

A number of web-accessible databases collect and publish protein sequences, along with a range of information regarding, for example, known or predicted properties, roles in biomolecular pathways, and 3-dimensional structures. Such

databases include [Entrez Protein](#), [UniProtKB](#), [KEGG](#), and the [Protein DataBank \(PDB\)](#) and [others](#). However, it is important to be aware that the identity and properties of a protein are not always completely identified by the common name of the protein. Different organisms, such as humans and yeast, are often said to possess the “same protein”, although the human and yeast versions normally have somewhat different amino acid sequences, so their chemistry and properties are not identical. A given protein may also have point-mutations, in which one or a few sequence positions differ from their nominal amino acids; for example, an isoleucine might be replaced by a valine. If it resides in a critical part of a protein, such as the active site of an enzyme, such a substitution can strongly affect properties like binding affinity and catalytic power. On the other hand, some amino acid substitutions have little effect on function. Mutations are often introduced intentionally by experimentalists in order to produce a more stable and hence easily studied form of the protein. A nominal protein sequence can also be altered by cleavage of the backbone, so that a single amino acid chain is split into two chains that may, nonetheless, remain attached to each other in a functional complex. Moreover, entire chunks of protein sequence can be excised (“deleted”). Finally, as noted above, the chemical structure of a protein can be modified by covalent addition of chemical groups including phosphate groups, methyl and acetyl groups, carbohydrates, and lipids.

The physical state of a protein, especially its 3-dimensional conformation, is critically important to its function. As a linear polymer with many rotatable bonds, a protein is intrinsically flexible and, at high temperatures or in some solution conditions, does not adopt a well-defined conformation but instead is in constant thermal motion, randomly moving among many different conformations as time passes. However, biological proteins typically adopt well-defined and intricate 3-dimensional conformations which

are essential to function. For example, Figure 3¹ shows the shape of the polymer chain of HMG-CoA reductase (1DQ8²) (magenta ribbon). The resulting molecular surface (dark blue) forms a binding site for a small molecule (balls and sticks with white molecular surface).

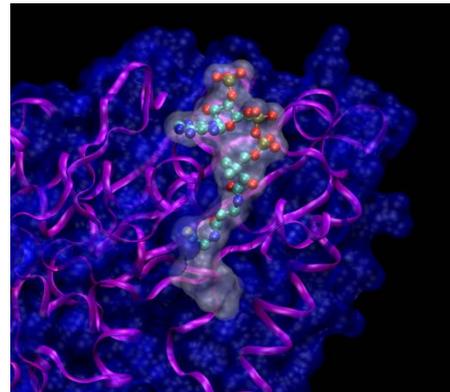


Figure 3. Structure of part of HMG CoA reductase (1DQ8) with a bound small molecule. Magenta: backbone ribbon, Blue: molecular surface of enzyme. Balls and sticks: small molecule in enzyme binding site. White: surface of small molecule. Rendered with VMD.

The properties of a protein can vary due to changes in its physical state, which usually result from changes in its environment. When present at sufficient concentration, some proteins attach to each other by noncovalent forces (below) to form dimers or larger multimers. At high, and sometimes low, temperatures, proteins can partly or completely lose their well-defined 3-dimensional structures. And co-solutes, especially metal ions but others as well, can bind to a protein and affect its structure and properties.

In summary, what we know as a specific protein can still have chemical and physical variations that influence its properties. This variability may help explain some cases where the literature, and hence BindingDB, reports more than one measured affinity for a given protein and small molecule.

Small molecules

Currently, the term “small molecule” typically implies a nonpolymer, organic compound whose molecular weight is less than roughly 1000 Da. Such compounds have a reasonable chance of being chemically stable, having uncomplicated physical properties, and spreading through the body to reach a targeted protein after being ingested or injected. Most drugs are small molecules. As examples, Figure 3 shows the analgesic naproxen and the heart-failure drug digoxin (above). Digoxin is a natural product produced by the foxglove plant, while naproxen is completely synthetic. Digoxin’s greater size and complexity are common among natural products; nonetheless, some natural products, such as salicylic acid (Figure 4), are rather simple compounds. A number of other drugs also are natural products or modifications thereof. However, many drugs, and most of the compounds in BindingDB, are

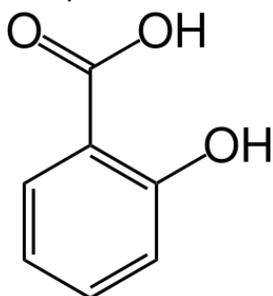


Figure 4. Salicylic acid.
(From http://en.wikipedia.org/wiki/Salicylic_acid.)

entirely synthetic.

The search for a small molecule that binds a targeted protein is often done conceptualized as a two-step approach. First, one identifies a “[lead compound](#)” that binds the target with moderate affinity. Then, one makes many variants of the lead compound in a search for higher-affinity compounds. This second step leads to a series of compounds with a constant chemical scaffold. This may be termed a congeneric series (Figure 5).

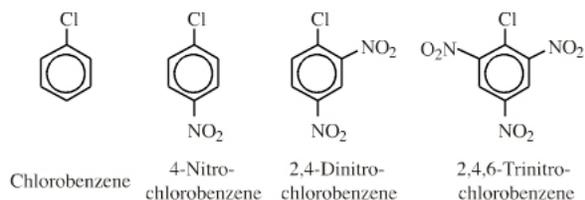


Figure 5. A congeneric series. (From http://www.meta-synthesis.com/webbook/12_lab/lab.html)

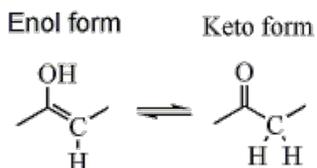


Figure 6. Tautomerism (Wikipedia).

Much like a protein, a small molecule can have variants. Some compounds are readily ionized via protonation or deprotonation, so their charge states depend upon solution pH. To complicate matters, a molecule's ionization state can change when it binds a protein. Similarly, the location of a proton on a molecule can shift, leading to alternate [tautomers](#) (Figure 6), and, again, the main tautomer state can change on binding. BindingDB stores and displays only a single protonation state, the neutral one, and a single tautomer state of a

given compound; it is up to the user to recognize possible protonation variants. Finally, some compounds with a given composition and arrangement of bonds may be able to exist in various [stereoisomers](#) that do not interconvert. It is important to be aware that different stereoisomers can have very different affinities for a given protein, and also that experiments sometimes are done with mixtures of stereoisomers, in which case it may be impossible to determine the binding affinities of any of the individual stereoisomers. When an article specifies a given stereoisomer, BindingDB captures and stores this information. When an article explicitly states that a mixture of stereoisomers was studied, BindingDB captures the stereochemistry and uses appropriate wedge bonds in displaying the compound (Figure 7). When an article reports on a compound that was studied as a mixture of stereoisomers, BindingDB displays the molecule without wedge bonds and annotates it as racemic.

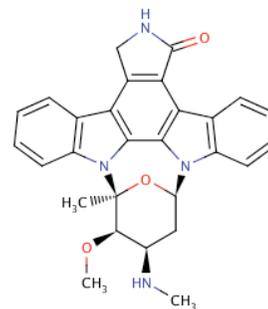


Figure 7. The natural product staurosporine, with wedge bonds indicating stereochemistry.

Small molecules may be represented on the computer in a number of machine-readable formats. These include various versions of the [MDL Molfile](#), which holds a single compound; the MDL structure-data file ([SDFFile](#)), which holds many molecules along with associated data; the [SMILES string](#), a much more compact format; and the [InChI](#) string, a compact, comprehensive format developed by the chemical informatics community. Compound downloads from BindingDB are currently provided as SDFfiles; their content and format is detailed [here](#).

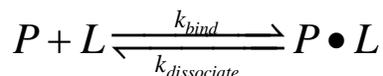
In addition to BindingDB, there are a number of databases and catalogs with small molecules and associated information, including, alphabetically, [ChEBI](#), [ChemSpider](#), [CSLS](#), [DrugBank](#), [Maybridge](#), [PubChem](#), [SigmaAldrich](#), and [Zinc](#).

Binding affinity and equilibrium

A small molecule that binds a protein is often termed a “ligand”. (Note that this term is also used more broadly to mean any molecule that binds another one, not just a small molecule, so it is actually somewhat ambiguous.) Some ligands form a covalent bond to the protein, leading to what is sometimes called irreversible

binding. However, noncovalent or reversible binding is more common, and BindingDB focuses on ligands that bind noncovalently.

If one considers a solution with fixed total concentrations of dissolved ligands and proteins, $[L]_{\text{tot}}$ and $[P]_{\text{tot}}$, respectively, then, in the case of noncovalent binding, the ligands are constantly binding and dissociating from the proteins, as indicated by the following chemical reaction:



where the rate constant for binding is k_{bind} and that for dissociation of the protein-ligand complex is $k_{\text{dissociate}}$. If one imagines averaging over some time-period, there is a well-defined fraction of time, f , that each protein has a bound ligand. (Roughly speaking, the time required to establish this equilibrium is several times the smaller of the two rate constants, k_{bind} and $k_{\text{dissociate}}$.) The equilibrium concentration of protein with bound ligand, $[P \bullet L]_{\text{eq}}$, is then given by $[P \bullet L]_{\text{eq}} = f [P]_{\text{tot}}$, and the equilibrium constant for the forward reaction, also known as the binding constant, is given by

$$K_B = \frac{k_{\text{bind}}}{k_{\text{dissociate}}} = \frac{[P \bullet L]_{\text{eq}}}{[P]_{\text{eq}} [L]_{\text{eq}}}$$

Here square brackets indicate a concentration in moles/liter, and $[\]_{\text{eq}}$ indicates a concentration at equilibrium.

To help make sense of this equation, we can rewrite it as

$$\frac{[P \bullet L]_{\text{eq}}}{[P]_{\text{eq}}} = K_B [L]_{\text{eq}}$$

This says that the ratio of bound protein to free protein (left-hand side) is proportional to the binding constant and the concentration of free ligand (right-hand side). The role of $[L]$ reflects the principle of [mass action](#): the more ligand is present, the greater the odds that a given protein molecule has a bound ligand. The binding constant, K_B , is a measure of affinity: if it is large, then even a low concentration of ligand generates high odds that a given protein molecule has a bound ligand. A large value of K_B is helpful for a ligand that one wishes to use as a drug: if only a low concentration is enough to put most of the targeted protein molecules into the bound state, then one can treat a patient just a small dose of the drug. This is usually a good thing, because it minimizes side effects and reduces costs.

Another way that people often look at the binding equilibrium is to focus on the dissociation constant, K_D , instead of the binding constant, K_B . The dissociation constant is simply the reciprocal of the binding constant

$$K_D = \frac{1}{K_B} = \frac{[P]_{eq} [L]_{eq}}{[P \bullet L]_{eq}}$$

so it provides exactly the same information. However, it has a pleasing interpretation: it is the concentration of ligand that gives even odds that a given protein molecule has a ligand bound. That is, for a one-to-one ratio of $[P]_{eq}$ to $[P \bullet L]_{eq}$, $[L]_{eq} = K_D$. Thus, if one imagines gradually adding more and more ligand and monitoring the increasing fraction of ligand-bound protein, then $[L] = K_D$ will be the concentration of ligand for which half the protein is bound. Thus, the *lower* the value of K_D , the *higher* the affinity of the ligand. In the drug discovery process, a drug-like ligand with K_D in the millimolar range is considered rather uninteresting; K_D in the low micromolar range might indicate a lead compound (above) worth further chemical optimization; and K_D of low nanomolar or below is good enough for a viable drug candidate.

Binding energy and entropy

The binding reaction can also be analyzed in thermodynamic terms involving free energy, enthalpy, entropy, and so on. For a binding reaction at constant pressure (the typical situation for biochemical studies), the standard [Gibbs free energy](#) change is given by

$$\Delta G^\circ = -RT \ln(K_B)$$

where R is the [gas constant](#) and T is the [absolute temperature](#) (usually ~300 K for biochemical systems). This free energy change can be broken down into energetic and entropic contributions:

$$\begin{aligned} \Delta G^\circ &= \Delta H - T \Delta S^\circ \\ &\approx \Delta E - T \Delta S^\circ \end{aligned}$$

Here H is the enthalpy, S° is standard entropy, E is average energy, and the approximate equality indicated by “ \approx ” is based on the definition of enthalpy, ($H=E+PV$, where P and V are pressure and V volume, respectively), and the fact that PV hardly changes for a protein-ligand binding reaction in water at typical lab conditions.

The change in energy on binding, ΔE , is averaged over the range of molecular conformations that the free and bound molecules explore due to [thermal motion](#). The distribution of conformations changes with binding, so the mean energy changes as well. The energy E can be parsed into a number of contributions, including [electrostatic interactions](#) among charged and dipolar chemical groups,

[London dispersion forces](#), and the intrinsic energetics of rotatable bonds (e.g., a linear alkane prefers an all-trans conformer).

When a protein and ligand bind, they form new interactions with each other but lose interactions with the solvent (mainly water here). Thus, ΔE is a trade-off between protein-water and ligand-water interactions before binding and protein-ligand interactions after binding. Water molecules are small and mobile, so they can interact very intimately with the protein and ligand molecules. Therefore, in order for the new protein-ligand interactions after binding to outbalance the water interactions before binding, the protein and ligand must fit together snugly and form a large stabilizing interface. Thus, small molecules typically bind in protein clefts or cavities (“binding pockets”), where they can contact the protein on many sides, and it is difficult to find a small molecule that binds a relatively flat portion of protein surface with high affinity.

Thermal motion plays a second role in protein-ligand binding, through the entropy change on binding ΔS° . The operative principle is that [objects in motion tend to keep moving in a straight line](#), so that forming and maintaining well-defined structures, like proteins, ligands and protein-ligand complexes, requires restraining forces, and the greater the kinetic energy (i.e., the temperature), the more work is required to impose such restraints. Thus, all other things being equal, raising the temperature will tend to disrupt structures, such as protein-ligand complexes. In thermodynamic terms, forming structures implies lowering the entropy and thus raising the free energy. However, it is not a given that protein-ligand binding always lowers the entropy, because binding is associated with many other motional changes. For example, binding can “set free” water molecules associated with the free protein and ligand, producing an increase in entropy that favors binding. In some cases, binding may also “set free” parts of the protein away from the binding site, leading to a further increase in entropy³⁴⁵⁶⁷. Experimental studies (below) show that the net change in entropy on binding can be positive or negative, depending up the molecules involved and the experimental conditions.

The statistical thermodynamics of binding is discussed in detail elsewhere⁸⁹.

Measurement of binding affinity

There are many useful methods of determining protein-ligand binding affinities experimentally. Most or all involve measuring some indicator of binding as a function of the concentration of the ligand and or protein. For example, if the free (unbound) protein can be specifically detected in some way

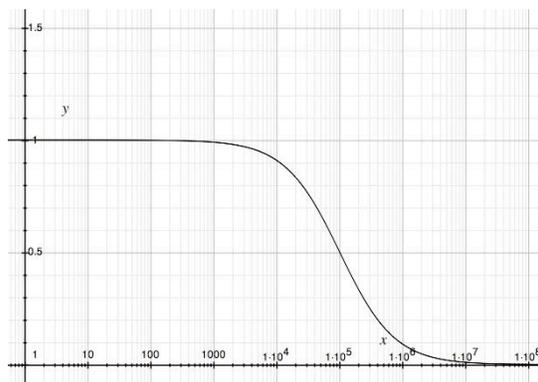


Figure 8. Fraction of free protein vs. concentration of the free ligand. RD=10⁴ M.

(e.g., spectroscopically or based on its enzymatic activity), then one can measure the fraction, f , of free protein as a function of the concentration of free ligand $[L]_{eq}$

$$f = \frac{[P]_{eq}}{[P \bullet L]_{eq} + [P]_{eq}} = \frac{1}{1 + K_B [L]_{eq}}$$

This function is graphed in Figure 8. Note that the fraction of free protein is identically 0.5 when $[L]_{eq} = K_D$, which is 10^{-5} M for this example. More complex equations and graphs are helpful for interpreting actual experiments where, for example, one may control the total amount of ligand $[L]_{eq} + [P \bullet L]_{eq}$ rather than the equilibrium concentration of free ligand $[L]_{eq}$.

Enzyme inhibition assays

Many proteins that are potential drug targets are [enzymes](#), and one wishes to discover small molecule drugs that bind and inhibit them; i.e., that are [enzyme inhibitors](#). Probably the most common type of enzyme inhibitor is a competitive inhibitor, which binds in the enzyme's active site and competitively blocks binding of the enzyme's physiological substrate. (A noncompetitive inhibitor is one that can bind along with substrate.) *For enzyme inhibitors, the dissociation constant K_D is usually termed the inhibition constant K_i .* It is important to be aware that data from enzyme inhibition is often reported not as K_i (or, equivalently, K_D), but instead as the [IC50](#), the concentration of ligand that reduces enzyme activity by 50%. The difference between IC50 and K_i results from the fact, in an enzymatic binding assay of a competitive enzyme inhibitor, the inhibitor is not the only molecule trying to bind the enzyme's active site. It is competing with the substrate, so the concentration of ligand needed to reduce the enzymatic activity by 50% depends on the concentration of substrate and how tightly it binds the enzyme¹. In general, then, the IC50 is expected to be greater than K_i , but when the concentration of substrate is very low, the IC50 should become essentially equal to K_i . (See the [Cheng-Prusoff equation](#).)

Enzyme inhibition assays can be very convenient because the action of the enzyme acts as an amplifier, since each enzyme molecule in solution can generate many reactions. On the other hand, some effort is needed to develop an enzyme inhibition assay, as some means of detecting the removal of substrate or the generation of product is needed. One common approach is to devise a substrate whose fluorescence properties change with reaction, and then to use a fluorimeter to monitor the activity of the enzyme as a function of inhibitor concentration.

¹ More properly, in the case of [Michaelis-Menten kinetics](#), the key quantity is the ratio of the substrate concentration to the Michaelis constant, K_M concentration of substrate is low relative to K_M ,

Isothermal titration calorimetry (ITC)

Binding of a protein with a drug-like molecule often leads to a small amount of heating or cooling of the solution; i.e., the reaction can be exothermic or endothermic. [Isothermal titration calorimetry](#) measures binding through detection of these changes. “Isothermal” means that the solution is kept at a constant temperature with a thermostat, and the chemical heat release or uptake is evaluated based on the energy demanded of the thermostat to maintain a constant temperature in the solution. One reactant is held in the thermostatted cell and the other is titrated in with small injections from a syringe holding a solution with the other reactant. Aliquots are added until there is basically no more heating or cooling with successive reactions, at which point the cell reactant is presumed to be fully saturated with the syringe reactant. The curve of heating or cooling vs amount of injected reactant is analyzed to extract not only the equilibrium constant for the binding reaction, but also the standard molar enthalpy (see “Binding energy and entropy”, above).

An ITC assay tends to require a larger quantity of sample (protein and ligand) than an enzyme inhibition assay, since it does not have the same property of amplification (see above). However, ITC has the advantage of not requiring the development of a specialized substrate whose reaction can be detected; the natural heat release of the binding reaction is that that is needed. Note, however, that some binding reactions neither take up nor release much heat, and these are not so easily monitored by ITC.

Other assays

Many other methods of measuring binding affinities are also available. [Functional and competitive displacement assays](#) are particularly important for membrane-bound receptor proteins. These are not enzymes, so enzyme inhibition assays cannot be used; and they are not readily solubilized in a functional form and at sufficiently high concentration for ITC to be suitable.

In one type of functional assay, a membrane bound receptor protein is expressed in a cell in such a way that the consequences of receptor activation can be detected. Then the cell is exposed to varying concentrations of inhibitor and its responses are monitored to generate a binding curve (see Measurement of Binding Affinity, above). The concentration of compound that generates a half-maximal response is then reported as the [EC50](#), the half maximal effective concentration.

In a competitive displacement assay, the receptor is bound to a ligand that is detectible on release. For example, it may be radioactive or fluorescent. Then the compound of interest is added, and the amount of released ligand is measured. The more effective the compound in question is at releasing the detectible ligand, the greater its affinity. Again, an EC50 is often reported, although if the affinity of the detectible ligand is known, then the K_i of the compound being assayed can be determined.

These assays in which the protein is a receptor in the membrane of a living cell can be quite sensitive, but are subject to uncertainties due to the complexity of the living system. For example, the cell may take up or even metabolize a compound, thereby reducing its effective or actual concentration.

Sources of uncertainty in affinity measurements

As discussed above, experimental determination of a binding affinity requires measuring some “signal” as a function of the concentration of the ligand and/or protein. These measured signals are then analyzed to extract a value of the binding affinity. Error can result from the measurements or from invalidity of assumptions used to extract the affinity from the measurements. Sources of error include:

- Instrumentation issues that lead to erroneous or inconsistent signal measurements, such as drift. Such errors generate uncertainty in the y-values of the data points that make up the binding curve illustrated in Figure 8.
- Errors in the actual or effective concentration of free protein and/or ligand. These can generate uncertainty in the x-values of the data in Figure 8. For example
 - if one uses the incorrect molecular weight for the protein when making up the protein solution, then the molar concentration of protein will be incorrect.
 - Even if the molarity of the solution is correct, some of the protein molecules may be chemically or physically damaged (e.g., unfolded) and therefore not competent to bind ligand. In this case, the effective concentration of protein is lower than the actual molarity.
 - Some of the protein may stick to the walls of the vessel, leaving a lower concentration in solution.
 - The ligand may aggregate into tiny droplets, lowering the effective concentration of free ligand. (This scenario is further complicated if the droplets stick to the proteins and generate a large binding signal that does not reflect the one-to-one-binding that one intends to measure¹⁰).
 - The ligand may have a chiral center, and the two enantiomers may have very different binding constants. If the solution contains a mixture of both forms (a racemic mixture) then one will probably notice primarily the stronger binding constant, but the effective ligand concentration will be lower than initially supposed, since only the tight-binding enantiomer is detected.
- The protein may have multiple binding sites for the ligand, making it incorrect to analyze the binding data with the math appropriate to one-to-one binding.

It is also worth reiterating that there are additional sources of variation among binding measurements for a given protein and ligand. These include the choice of enzyme substrate in enzyme IC₅₀ assays, the pH, temperature, and ionic strength, and potentially the choice of cosolutes such as buffers and redox agents. There is also a widespread impression (one that is somewhat troubling, if valid) that different laboratories may employ techniques that differ subtly yet enough to affect their affinity measurements significantly. As a consequence, if one wishes to compare the affinities of a range of ligands for a given protein, it is generally thought that the most comparable data are those obtained from one laboratory with one assay method.

References

1. Humphrey, W., Dalke, A. & Schulten, K. VMD -Visual Molecular Dynamics. *J. Molec. Graph.* **14**, 33-38 (1996).
2. Istvan, E., Palnitkar, M., Buchanan, S. & Deisenhofer, J. Crystal structure of the catalytic portion of human HMG-CoA reductase: insights into regulation of activity and catalysis. *EMBO J.* **19**, 819-830 (2000).
3. Homans, S.W. Probing the binding entropy of ligand-protein interactions by NMR. *ChemBioChem* 1591 (2005).
4. Zidek, L., Novotny, M.V. & Stone, M.J. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct.* **6**, 1118-1121 (1999).
5. Lee, A.L., Kinnear, S.A. & Wand, J.A. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat. Struct. Biol.* **7**, 72-77 (2000).
6. Arumugam, S. et al. Increased Backbone Mobility in beta-Barrel Enhances Entropy Gain Driving Binding of N-TIMP-1 to MMP-3. *J. Molec. Biol.* **327**, 719-734 (2003).
7. Killian, B.J. et al. Configurational entropy in protein-peptide binding. Computational study of TSG101 UEV domain with an HIV-derived PTAP motif.
8. Gilson, M.K., Given, J.A., Bush, B.L. & McCammon, J.A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **72**, 1047-1069 (1997).
9. Zhou, Huan-Xiang & Gilson, Michael K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **in press**,
10. Seidler, J., McGovern, S.L., Doman, T.N. & Shoichet, B.K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **46**, 4477-4486 (2003).