JCTC Journal of Chemical Theory and Computation

A Note on the Standard State's Binding Free Energy

Ignacio J. General*

Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received May 14, 2010

Abstract: The relation between the equilibrium constant of a given chemical reaction and the associated free energy is an issue well studied in chemistry books, but when the reaction involves changes in the number of components in the system, as is the case in binding, things become a little more obscure since one needs to define the so-called standard state. This is reflected in the literature, especially in computational studies of binding, where contradicting approaches are followed when treating this problem. In this work, we present a detailed and unifying explanation of the concepts involved and derive the necessary relations to convert a binding free energy from an arbitrary state to some given standard state. This is done in three independent ways, from the point of view of (1) the dimensions of the quantities involved, (2) the energy and entropy of the molecules, and (3) their chemical potentials.

Introduction

Equilibrium constants are generally defined, in chemistry books, in terms of ratios of concentrations of reactants to products. This presents a problem, often ignored, when the number of reactants is not the same as the number of products, since the ratio will not be dimensionless. Many times, one needs to calculate the free energy associated with that equilibrium constant, K, and the usual way of doing so is through the well-known formula[1,2]

$$F = k_B T \ln(K) \tag{1}$$

where k_B stands for the Boltzmann constant, T is the temperature of the system under consideration, and F is the corresponding free energy. In this paper, we will be working in the canonical ensemble, so F represents the Helmholtz free energy, and we assume a constant volume in all calculations. This is not a big drawback since, in binding reactions where the system's volume is allowed to fluctuate, its change, ΔV, at normal values of concentration, pressure and temperature, is negligible.

Equation 1 shows that having units in the equilibrium constant forbids the use of the formula, since the argument of a logarithm must be dimensionless. In particular, when talking about the experimental binding of two molecules (e.g., protein A + ligand B), K^{exp}, defined by

$$K^{exp} = \frac{[A][B]}{[AB]} \tag{2}$$

is typically given in units of molar. The solution to this units problem is to first convert K^{exp} to a new K^0, without units. This can be achieved by doing the following:[3]

$$K^0 = \frac{\dfrac{[A]}{C^0}\dfrac{[B]}{C^0}}{\dfrac{[AB]}{C^0}} = \frac{K^{exp}}{C^0} \tag{3}$$

where C^0 is a constant with the same units as K^{exp}. We call this the *standard state equilibrium constant*. But note that C^0 is arbitrary; it is precisely here where the standard state is defined. In other words, in order to be able to calculate a free energy from the equilibrium constant, we must first define a standard state. Any free energies that we calculate later, via theory, simulation, or experiment, should also be referred to the same standard state, so that we can make a meaningful comparison between them.

This problem has already been correctly treated by other authors,[4,5] as related to computational studies of binding, but nevertheless, it is still easy for somebody who is confronted with this issue for the first time to get confused about it, since one can find some authors comparing results without taking care of the reference state at all,[6] some others considering corrections based on completely different grounds

* E-mail: ijgeneral@gmail.com.

10.1021/ct100255z
Published on Web 07/15/2010

and, thus, apparently contradicting each other (see, e.g., ref 7 for a correction based on the unbound sampled volume,[8] for a correction based on the binding site volume,[4,5] for one based on the chemical potential, and[9] for a model that contains its own correction), and yet some others explicitly denying the existence of such corrections.[10] It is the purpose of this work to review this issue in a comprehensive and rigorous way, trying at the same time to be as clear as possible, especially in those points where confusion is seen in the literature.

So the key question is: how do we convert a computational (or experimental or theoretical) free energy, F^{comp}, to a free energy in a given standard state, F^0 (usually, the standard state in which the experimental value is already known)? In the following sections, we will derive this transformation in three different ways: the first is a simple *dimensional* approach which, through the consideration of units, leads to the desired result; the second is an intuitive *physical* approach, based on the consideration of energy and entropy; and the last one is a more *chemical* approach, which considers the chemical potential of the molecules in the binding process. The first derivation is very illustrative, as it shows that the problem in comparing free energies is ultimately related to the matching of their units. The second derivation is, in a way, the best derivation since it makes very clear the physical meaning of the correction, from the point of view of physical magnitudes. The last derivation is presented for the sake of completeness, but different versions of it can be found in the literature.[4,5]

Conversion of a Free Energy to Some Standard State

Dimensional Approach. Let us define K as a dimensionless equilibrium constant, related to $\tilde{K}$, its equivalent with units, i.e.,

$$K = \frac{\tilde{K}}{C} \tag{4}$$

C is some concentration that we arbitrarily choose as a standard (typically 1 M), but there are some times when we are not able to choose it, but it is chosen automatically, instead. This is the case when we calculate F by simulation; there we get the free energy, which is related to some already dimensionless equilibrium constant, via the inverse of eq 1:

$$K = \exp\left(\frac{F}{k_B T}\right) \tag{5}$$

So now we could ask: In the experiment we obtain $\tilde{K}^{exp}$, and get K^{exp} by dividing by some standard concentration C^0, in order to get rid of the units. In the simulation, we are getting K^{comp}. Then, by what factor did the computer (or the theory) divide to get rid of the units? The only concentration available to the theory is that of the system we are simulating, so we must assume that we have

$$K^{comp} = \frac{\tilde{K}^{comp}}{C^{comp}} \tag{6}$$

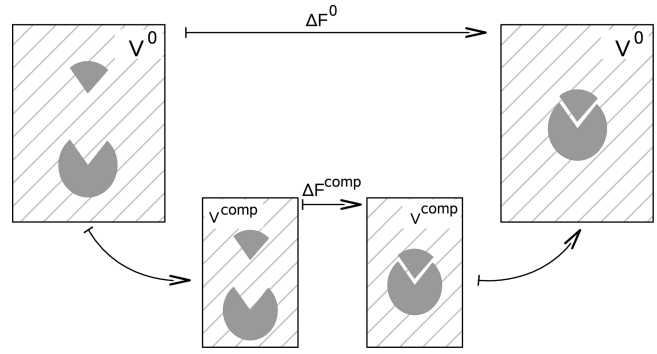

Figure 1. Relation between computational and standard state binding free energies. V^0 and V^{comp} represent the volumes of the corresponding systems.

where C^{comp} is the concentration we use in the simulation. But, of course, we want the experimental and computational constants to be the same (within some error) when they are expressed in the same units, so

$$\tilde{K}^{exp} = \tilde{K}^{comp} \Rightarrow \frac{\tilde{K}^{exp}}{C^0} = \frac{\tilde{K}^{comp}}{C^{comp}}\frac{C^{comp}}{C^0} \tag{7}$$

which, by taking $k_B T \ln(...)$, leads to

$$F^0 = F^{comp} + k_B T \ln\left(\frac{C^{comp}}{C^0}\right) \tag{8}$$

where we renamed F^{exp} as F^0, since it is in the standard state. This expression tells us, by way of dimensional arguments, that the computational free energy needs a correction term in order to be compared to a free energy calculated by other methods. This is necessary if the units of the different equilibrium constants are to match.

The magnitude of the correction is a function of the concentration of the molecules we have in the study. Typical ligand concentrations in computational binding studies are between 1 per 15 Å radius sphere and 1 per 100 Å radius sphere. Using a standard state of 1 M (1 ligand per 1661 $Å^3$), the corresponding corrections at 300 K fall between 1.3 and 4.7 kcal/mol. So, although these corrections will not, in general, drastically change the results, they are significant, especially if one is interested in getting precise values of free energy.

Equation 8 also suggests a way of checking the results of computational calculations. Note that F^0, being in a given standard state, must be independent of C^{comp} (it should not matter how it was calculated, as long as the standard state in which it is expressed is the same). This means that F^{comp} has to be a function of C^{comp}, so that, together with the last term in eq 8, this dependence on the computational concentration is canceled. Thus, the lack of dependence of F^{comp} on C^{comp}, sometimes observed in computational studies, indicates a problem, typically a convergence issue, such as the ligand not sampling the full available phase space.

Physical Approach. This approach is vaguely mentioned in the literature,[7,8,11] and to the author's knowledge, it has never been put on a solid foundation. Figure 1 shows the relation between the binding free energies of two molecules, as calculated in two different systems. We will call them

standard and computational, but the discussion below is correct for any two systems (standard−experimental, computational−experimental, even computational−computational, when each is calculated using different methods). The volumes there represented, V^{comp} and V^0, are not in general the same, so we should now find a way to convert F^{comp} to F^0. We can achieve this by considering the physical changes in the figure.

• ΔF^{0-comp}: change in F when the volume of the system varies from V^0 to V^{comp}. Assuming the molecules in the system to be electrically neutral, we can see that the energies of the two molecules, A and B, remain constant, while the entropies of the two do change; increasing (decreasing) the volume will produce an increase (decrease) in the entropy of these molecules, since they will have more (less) possibilities to move.

$$\Delta F^{0-comp} = (\Delta E - T\Delta S)_A + (\Delta E - T\Delta S)_B = (-T\Delta S)_A + (-T\Delta S)_B = -2Tk_B \ln \frac{V^{comp}}{V^0} \quad (9)$$

Above, A and B typically represent a protein−ligand pair.

• ΔF^{comp}: change in F as calculated computationally.

• ΔF^{comp-0}: change in F when the volume is changed back from V^{comp} to V^0. As in the first step, the only difference here is an entropic one. But this time, molecule B can only move inside the binding site. We can think of this AB complex as only one molecule with some extra freedom because the ligand can slightly move in the complex. But the amount of this extra freedom is the same in the simulation and the standard state boxes. So the ligand does not contribute to this change.

$$\Delta F^{comp-0} = (\Delta E - T\Delta S)_{AB} = -Tk_B \ln \frac{V^0}{V^{comp}} \quad (10)$$

So, finally, adding the three terms above we find the conversion term for going from the simulation to the standard state:

$$\Delta F^0 = \Delta F^{comp} - k_B T \ln \frac{V^{comp}}{V^0} \quad (11)$$

or, using $C_i = N_i/V_i$, we recover eq 8.

The above approach to the standard state correction, also provides a different kind of computational correction that we may sometimes need to consider, depending on the particular calculation we are doing. We are talking specifically about (alchemical) thermodynamic integrations. Over a decade ago, it was found[4,12,13] that it is computationally convenient to calculate binding free energies using thermodynamic cycles of the type shown in Figure 2, but performing the integration of the right-hand side leg of the cycle with a restraint applied on the ligand, so that it cannot leave the binding pocket, even when the interaction between the ligand and the rest of the system is decreased to zero (the so-called *double decoupling method*,[4] as opposed to the older *double annihilation method* of Jorgensen et al.[14]). This brings the problem that the change in free energy on the bottom of the cycle (where the ligand is altogether decoupled from the rest of the system) is not zero anymore (see Figure 2). But if we realize that the difference between those two systems is just the volume accessible to the ligand, then we can easily apply the correction derived above:

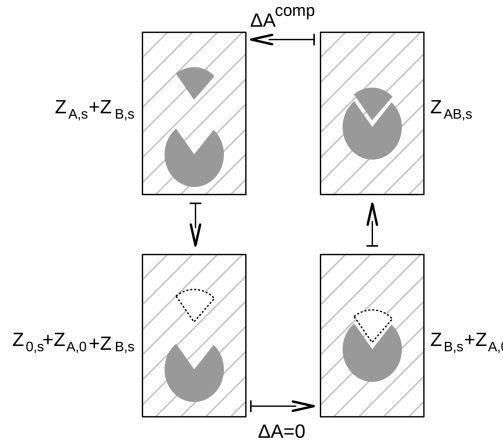

Figure 2. Thermodynamic cycle for protein−ligand binding. Each step in the cycle shows the partition functions associated with it. The white triangular shape, representing the ligand (A), is decoupled from the rest of the system, protein (B) and water, as it is usually done in computational simulations (e.g., thermodynamic integration). The hatched lines in the background represent a solvent.

$$k_B T \ln \frac{V^{comp}}{V^{pocket}} \quad (12)$$

Adding this term to the free energy corrects also for having a restraint applied on the ligand. Note that there could be an analogous correction, in angular space, if the restrained ligand is not able to freely rotate. But we should keep in mind that eq 12 is a different correction than the one studied above. Their forms are equal since both arise from the consideration of a change in volume, but they should not be confused as the same, something that often happens in the literature, because then there is going to be a missing contribution to the binding free energy. Sometimes, this second correction is put together with the standard state one:[8,15,16]

$$-k_B T \ln \frac{V^{comp}}{V^0} + k_B T \ln \frac{V^{comp}}{V^{pocket}} = -k_B T \ln \frac{V^{pocket}}{V^0} \quad (13)$$

but the authors[8] may fail to mention that there are two different contributions here, and they just call it the standard state correction. This has no numerical consequences, as the result is correct, but it may confuse the reader.

Chemical Approach. *Preliminaries.* We start by recalling the derivation of a basic result in thermochemistry. The Helmholtz free energy can be defined, in terms of temperature and volume, by $dF = -SdT - pdV$. Now, if we consider an ideal gas undergoing an isothermal volume change from V^0 to V, we have

$$dF = -pdV \Rightarrow F(V) = F(V^0) - \int_{V^0}^{V} \frac{Nk_B T}{V} dV = F(V^0) - Nk_B T \ln \frac{V}{V^0} \quad (14)$$

$$\Rightarrow \mu = \mu^0 - k_B T \ln \frac{V}{V^0} \tag{15}$$

In the last step, we used $\mu = F/N$. This expression, the relation between the chemical potential of a gas under two different conditions, can be generalized to nonideal gases and ideal and nonideal solutions. For example, for the case of some component i in a solution, the expression becomes

$$\mu_i = \mu_i^0 + k_B T \ln \frac{\gamma_i V_i^0}{V_i} \tag{16}$$

where γ_i is the activity coefficient of the substance we are considering (this coefficient contains the nonideality information) and V_i and V_i^0 are the volumes covered by component i of the solution, under the two external conditions.

The important point to note about this relation is that it expresses just a change in the conditions of the systems, as a change in volume or concentration. There is nothing special about the "0-state" or "standard state"; it is arbitrary, and thus we can choose any state that makes things easier for us.

With this relation, we can start the derivation of the conversion term.

Derivation. The following derivation of the conversion term is based on the ones given by Gilson et al.[4] and Boresch et al.[5] Consider a binding reaction, $A_s + B_s \leftrightarrow AB_s$, where the subscript s indicates that the molecules are immersed in a solvent. In the following, we are going to assume there is only one molecule of each component, i.e., $N_A = N_B = N_{AB} = 1$. We know that in equilibrium $\mu_A + \mu_B = \mu_{AB}$; thus, according to eq 16

$$\begin{aligned} &\mu_A^0 - k_B T \ln \frac{V_A}{V^0} + \mu_B^0 - k_B T \ln \frac{V_B}{V^0} = \mu_{AB}^0 - k_B T \ln \frac{V_{AB}}{V^0} \\ &\Rightarrow \Delta F^0 \equiv \mu_{AB}^0 - \mu_A^0 - \mu_B^0 = -k_B T \ln \frac{V_A V_B}{V^0 V_{AB}} \end{aligned} \tag{17}$$

where we have used $\gamma_i = 1$, the very low concentrations approximation. Since $C_i = N_i/V_i$, we can rewrite this as

$$\Delta F^0 = -k_B T \ln K;\ K \equiv \frac{C^0 C_{AB}}{C_A C_B} \tag{18}$$

recovering eq 1. But notice this relation is telling us that when we choose the value of C^0, so that we can calculate ΔF from K, we are determining the standard state. Consequently, in the equations above, when we write μ^0, V^0, or C^0, we are referring to the same standard state. In conclusion, the current derivation, started with eq 16, is just a conversion of some quantity from some standard state to a nonstandard state.

Let us now turn our attention to statistical mechanics. In the canonical ensemble

$$\mu_{i,s} = F_{i,s} - F_{0,s} = -k_B T \ln \frac{Z_{i,s}}{Z_{0,s}} \tag{19}$$

is the chemical potential of component i in the solution, s. $Z_{i,s}$ and $Z_{0,s}$ represent the partition functions of one and zero molecules of type i, respectively, in solution. Replacing this in eq 16

$$\begin{aligned} -k_B T \ln \frac{Z_{i,s}}{Z_{0,s}} &= \mu_i^0 + k_B T \ln \frac{\gamma_i C_i}{C_i^0} \\ \mu_i^0 &= -k_B T \ln\left(\frac{\gamma_i C_i}{C_i^0}\frac{Z_{i,s}}{Z_{0,s}}\right) \end{aligned} \tag{20}$$

And inserting this expression in eq 17:

$$\Delta F^0 = -k_B T \ln\left(\frac{\gamma_{AB}}{\gamma_A \gamma_B}\frac{C_{AB}}{C_A C_B}\frac{C_A^0 C_B^0}{C_{AB}^0}\frac{Z_{0,s} Z_{AB,s}}{Z_{A,s} Z_{B,s}}\right) \tag{21}$$

For the binding case, where we consider $C_A = C_B = C_{AB}$, we get

$$\Delta F^0 = -k_B T \ln\left(\frac{\gamma_{AB}}{\gamma_A \gamma_B}\frac{C_A^0}{C_A}\frac{Z_{0,s} Z_{AB,s}}{Z_{A,s} Z_{B,s}}\right) \tag{22}$$

To understand the meaning of this equation, let us assume infinite dilution ($\gamma_i = 1$), for simplicity, and rewrite it in the following way:

$$\begin{aligned} \Delta F^0 = -k_B T \ln\left(\frac{C_A^0}{C_A}\right) + [F_{AB,s} - F_{B,s} - F_{A,0}] + \\ [F_{0,s} + F_{A,0} + F_{B,s} - F_{A,s} - F_{B,s}] \end{aligned} \tag{23}$$

where F_i is the free energy associated with partition function Z_i, and we have added and subtracted a pair of terms, one involving $F_{A,0}$, which represents molecule A in no solution, meaning that it has been decoupled from the solution (it is in a gas phase), as it is done when using thermodynamic integration. Equation 23 can be easily interpreted in terms of the thermodynamic cycle in Figure 2. But that cycle is exactly the one used in computational simulations; consequently, eq 23 leads to

$$\Delta F^0 = -k_B T \ln\left(\frac{C_A^0}{C_A}\right) + \Delta F^{comp} \tag{24}$$

or

$$\Delta F^0 = -k_B T \ln\left(\frac{V_A}{V_A^0}\right) + \Delta F^{comp} \tag{25}$$

So we see that in order to convert the computational free energy to a given standard state, we need to add a correction term that takes into account the difference in volume between the two states (the simulation and the standard state volumes). And, of course, V^0 has to be the one used in eq 3, when the standardization of the experimental result is performed. Equation 25, with the subscripts A removed, coincides with eqs 8 and 11, derived using other approaches.

Summary and Conclusions

We have seen that binding free energies are defined in terms of some reference state, whose purpose is to make the equilibrium constant dimensionless. In this way, one can compare two constants calculated in different ways and obtain sensible results since they are in the same standard state.

This argument can also be put in terms of volumes and concentrations, instead of units, since the second derivation

showed that the correction appears as a consequence of considering different volumes, hence different entropies, when the free energy is calculated through different methods. The expression for the correction, given in eqs 8, 11, and 25, depends on the ratio of the volumes (or the concentrations) of the two states. In the binding case we use in the present work, they are the volume V^0 used to make the experimental equilibrium constant dimensionless (eq 3) and the volume V_A that was available to the type-A molecule, during the simulation, before binding (which was the same as that for type B: $V_A = V_B = V^{comp}$).

We also showed that the second formulation of the problem gives the correction that has to be added when using thermodynamic integration with restraints on the ligand. The reason for why the same type of correction can be used is that both problems, the standard state and the binding site problems, arise from comparing two systems with different volumes. Thus, the solution is the same, a factor that takes into account that difference.

Acknowledgment. The author would like to thank Prof. Jeffry Madura for helpful comments and Prof. Hagai Meirovitch for valuable discussions and for his support.

References

(1) Moore, W. J. *Physical Chemistry*, 5th ed.; Longmans: London; Prentice Hall: New York, 1972.

(2) Glasstone, S. Lewis, D. *Elements of Physical Chemistry*, 2nd ed.; Macmillan: London, 1964.

(3) Janin, J. *Proteins: Struct., Funct., Genet.* **1996**, *24*, i–ii.

(4) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.

(5) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

(6) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*, 084108.

(7) Doudou, S.; Burton, N. A.; Henchman, R. H. *J. Chem. Theory Comput.* **2009**, *5*, 909–918.

(8) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. *J. Chem. Phys.* **2006**, *125*, 084901.

(9) Fukunishi, Y.; Mitomo, D.; Nakamura, H. *J. Chem. Inf. Model.* **2009**, *49*, 1944–1951.

(10) Fujitani, H.; Tanida, Y.; Matsuura, A. *Phys. Rev. E* **2009**, *79*, 021914.

(11) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092–4107.

(12) Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670–681.

(13) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

(14) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742–3746.

(15) Rodinger, T.; Howell, P. L.; Pomès, R. *J. Chem. Phys.* **2008**, *129*, 155102–1.

(16) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

CT100255Z