Three Frontiers in the Thermodynamics of Protein Solutions

John Prausnitz and Loddie Hagar

Chemical Engineering Department, University of California, Berkeley; and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley CA 94720

Abstract: Three examples illustrate the versatility and usefulness of biothermodynamics. The first example concerns calculation of a phase diagram for aqueous lysozyme with a new potential of mean force that takes the Hofmeister effect into account; such calculations may be useful for design of a separation process where addition of a salt to an aqueous protein mixture precipitates a target protein. The second example concerns thermodynamic studies to elucidate the effect of an organic cosolvent on the mechanism of crystallizing aqueous insulin. The final example concerns a thermodynamic contribution to mitigating the AIDS epidemic; it indicates how isothermal-titration-calorimetry studies are helpful for choosing an optimum inhibitor that is effective not only for the wild-type HIV protease but also for at least some of its mutants.

INTRODUCTION

The great virtue of thermodynamics is its generality, its wide range of potential applicability for describing a large variety of natural phenomena. Regrettably, for historical reasons, the word "thermodynamics" is a misnomer; while the science of heat engines developed 150 years ago has been extended to numerous areas of physics and chemistry, the word "thermodynamics" has unfortunately not experienced a corresponding change. In the chemical sciences, when we talk about thermodynamics today, we rarely refer to heat engines. We refer, instead, to a quantitative description of nature but, nevertheless, we retain the word "thermodynamics" because we continue to use the general mathematical framework that was established by the thermodynamics pioneers of the nineteenth century.

In the eighteen seventies, when Gibbs showed how the framework of thermodynamics can be used for describing equilibria in physical chemistry, did he realize that his remarkable work could also be used for applications in biology? We don't know. But it is likely that if Gibbs could see the dramatically increasing literature on biothermodynamics, he would be pleased but not surprised.

This article presents three examples in recent biothermodynamics as discussed in the literature. They are:

1. Calculation of the phase diagram for an aqueous target protein in a separation process where addition of a salt causes protein precipitation (Tavares et al).
2. Effect of an organic co-solvent on the crystallization of aqueous insulin (Vekilov et al).
3. Design of a versatile inhibitor to prevent in-vivo reproduction of the HIV virus and its mutants (Ohtaka and Freire).

The first example concerns a classical phase-equilibrium calculation within the McMillan-Mayer framework. These calculations use a new potential of mean force to describe protein-protein interactions in an aqueous medium that also contains a salt. This new potential includes the Hofmeister effect.

The second example uses the thermodynamics of crystallization to interpret crystallization data; this interpretation provides information on how an organic co-solvent significantly increases the rate of crystallization of insulin from an aqueous solution.

The third example shows how experimental enthalpy and entropy data for binding the HIV protease help to guide development of a versatile drug for treating patients with AIDS.

Phase Equilibria for an Aqueous Protein

In biotechnology, initial separation of a target protein from a bioreactor broth is often achieved by adding a salt to induce precipitation. To provide guidance for attaining the desired separation, we require a pertinent phase diagram, i.e. a plot of temperature versus protein concentration or its equivalent, protein density in the solution, designated by ρ. A common way to express protein density is provided by the protein packing fraction designated by $\eta = \frac{\pi}{6}\rho\sigma_p^3$ where σ_p is the protein diameter.

For fluid-solid equilibria, the essential equations of equilibrium are

$$\mu^F(\eta^F, T) = \mu^S(\eta^S, T) \tag{1}$$

$$p^F = p^S \tag{2}$$

where μ^F is the chemical potential of the protein in the fluid phase and μ^S is the chemical potential of the protein in the solid phase. T is the temperature and p is the pressure. Similar equations hold for fluid-fluid equilibria.

At constant temperature, pressure p and chemical potential μ are related to ρ by well-known thermodynamics; both are derived from an expression for Helmholtz energy A as a function of ρ (or η), as discussed in numerous textbooks.

We calculate A^F and A^S using the first-order Barker-Henderson perturbation theory, one for the fluid phase and another for the solid phase, as discussed, for example, by Tavares et al (1, 2). For each phase, the chemical potential is obtained from two contributions, a hard-sphere (hs) term and a perturbation term (P): $\mu = \mu^{hs} + \mu^P$

For the fluid phase, μ^{hs} is obtained from the Carnahan-Starling equation of state. For the solid phase, μ^{hs} is obtained from the equation of state proposed by Velasco.

To calculate μ^P, we need an expression that provides information on the protein-protein interaction in the aqueous solution and in the precipitate. That expression is the potential of mean force W(r) where r is the center-to-center distance between two quasi-spherical protein particles. This potential of mean force contains three contributions:

$$W(r) = W(r)^{hs} + W(r)^{el} + W(r)^{att} \tag{3}$$

where el refers to electrostatic forces and att refers to attractive (dispersion) forces.

W^{hs} depends on the reduced distance r/σ_p.

W^{el} depends on r/σ_p, σ_s, q_p, q_s, D and T where σ_s is the salt diameter, q_p is the protein charge, q_s is the salt charge and D is the dielectric constant of water.

W^{att} depends on r/σ_p, H and on a new expression φ(ions). Here, H is the Hamaker constant that reflects the dispersion (van der Waals) forces between the aqueous protein particles, neglecting the effect of the salt ions. φ(ions) provides the contribution to protein-protein attractive forces due to dispersion forces arising from van der Waals interactions between the ions and between the ions and the protein. It is this function, φ(ions), that is responsible for the well-known Hofmeister effect. While this effect was discovered empirically more than 100 years ago, its quantitative (theoretical) formulation was derived only recently by Ninham and coworkers (3), and by Tavares et al (4). The function φ(ion) depends on the aqueous polarizabilities of the protein and the salt ions.

Figure 1 shows three calculated phase diagrams for lysozyme (5). Each diagram is for an aqueous lysozyme solution containing a sodium salt with concentration 0.2 M.

Figure 1 shows the important effect of the salt's anion. NaCl and NaI show qualitatively similar behavior; we see a fluid-fluid region with an upper critical point on the left side and a fluid-solid region on the right. However, the critical temperature for NaI is appreciably higher than that for NaCl.

For NaSCN, we obtain a phase diagram that is qualitatively different from those for NaCl and NaI. For NaSCN, we do not have a stable fluid-fluid region; only fluid-solid equilibria are thermodynamically stable. There is no upper critical point. However, in addition to fluid-solid equilibria, we also calculate metastable fluid-fluid equilibria, consistent with experiment. The calculations show a region (dashed line) where $\mu^{F'} = \mu^{F''}$ where ′ refers to the dilute aqueous lysozyme solution and ″ refers to the concentrated aqueous lysozyme solution.

For solutions containing NaCl or NaI, fluid-fluid equilibria are stable. However, for lysozyme solutions containing NaSCN, $\mu^{F'}$ and $\mu^{F''}$ are larger than the corresponding μ^{s} at the same temperature. For this case, the calculated (and observed) fluid-fluid equilibria are metastable.

Because the polarizability of the SCN anion is larger than those for chloride and iodide anions, SCN anions adsorb more to the positively charged lysozyme, reducing the repulsive forces between lysozyme particles. Thus, for lysozyme, NaSCN is a more effective precipitating salt than NaCl or NaI.

Thermodynamics of Insulin Crystallization

In the pancreas, insulin is stored in rhombohedral crystals. In vivo, organic molecules, soluble in water, affect the crystallization of insulin. Thermodynamics provides a useful tool toward a better understanding of the crystallization process.

Vekilov and coworkers (6, 7) have measured in-vitro kinetics and equilibria for crystallization of aqueous insulin. They found that addition of an organic cosolvent (acetone) to the aqueous solution dramatically increases the rate of crystallization by nearly one order of magnitude.

These studies are of interest not only because they may help to develop a better therapy for diabetes but also because fundamental studies on crystallization of biomacromolecules may be useful for pharmacology; crystalline pharmaceuticals are often preferred over amorphous forms because crystalline pharmaceuticals have a slower and more controlled rate of dissolution.

Consider a saturated solution of insulin in water in equilibrium with crystalline insulin. Vekilov and coworkers measured the (small) aqueous equilibrium concentration of insulin designated by c_e.

The standard Gibbs energy of crystallization is

$$\Delta G^\circ = RT\ln(c_e/c^\circ) \qquad (4)$$

where c° is the standard-state concentration (1 mol/kg solvent).

Further, $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

From well-known thermodynamics,

$$\frac{\partial\left(\frac{\Delta G^\circ}{\mathrm{T}}\right)}{\partial\frac{1}{T}} = \Delta H^\circ \tag{5}$$

and

$$\frac{\partial \Delta G^\circ}{\partial T} = -\Delta S^\circ . \tag{6}$$

Vekilov and coworkers measured c_e as a function of temperature from 5 to 30 degrees Celsius. To obtain information on the effect of a representative organic cosolvent, Vekilov et al also measured c_e as a function of acetone concentration. Figure 2 shows results. We see a discontinuity when the acetone concentration is near 12 volume percent.

Because a crystal is more ordered than a liquid, we expect the entropy of crystallization to be negative. However, in the absence of acetone, the observed entropy of crystallization is positive. By contrast, with acetone, the observed entropy of crystallization is negative. Vekilov et al conclude that in aqueous solution, hydrophobic insulin is surrounded by a shell of adsorbed oriented water molecules; upon crystallization, these water molecules are released. Because the released water molecules gain vibrational, translational and rotational degrees of freedom, the observed entropy of crystallization is positive.

For good crystallization, a positive entropy charge is favorable, as observed for insulin crystallization from acetone-free water. In the presence of acetone, the observed entropy charge is negative. Nevertheless, with acetone, good crystallization is obtained because the desired negative change in enthalpy is more negative than that when insulin is crystallized from acetone-free water.

When the concentration of acetone exceeds about 12 volume percent, the measured rate of crystallization is about seven times larger than that at lower acetone concentrations. The mechanism of insulin crystallization from water requires removal of adsorbed water from the hydrated insulin molecule. However, as suggested by the observed entropy of crystallization, at a sufficiently large acetone concentration, such removal is not necessary; because acetone prevents the formation of a shell of adsorbed water molecules, the rate of crystallization increases.

Thermodynamic interpretation of the experimental thermodynamic results shows that the mechanism of insulin crystallization is significantly affected by the presence of an organic cosolvent like acetone. Thermodynamic interpretation suggests that the organic co-solvent prevents hydration of the insulin molecule.

This example shows that when experimental solubility data are subjected to thermodynamic analysis, we can obtain insight toward a molecular interpretation of crystallization of a biomacromolecule. Such insight may contribute to improved processing of pharmaceuticals and, perhaps, to development of improved therapy for diabetes. While thermodynamics alone is not sufficient, it can contribute toward progress in pharmacology.

Thermodynamics for Drug Development

Much continuing effort has been directed toward finding drugs to reduce the AIDS epidemic caused by HIV. Thermodynamics makes a contribution to that effort through thermodynamic interpretation of experimental calorimetric-titration data.

As shown in Figure 3, after HIV infects a human cell, a series of steps produces replication of HIV. The infected cell expresses a polyprotein and an HIV protease; the role of the protease is to cleave the polyprotein. The cleaved proteins assemble to form a new HIV virus.

An effective strategy to prevent formation of a new virus is provided by introducing a drug that deactivates the HIV protease. This drug, called a protease inhibitor, prevents cleavage of the polyprotein, as indicated in Figure 4.

Deactivation of a protease inhibitor can be described by a traditional lock-and-key mechanism indicated in Figure 5. The inhibitor must have the correct geometric shape to fit into the active site of the HIV protease; here the inhibitor is the "key" that must fit into the protease "lock".

However, because of mutations, the active site of an HIV protease can exist in a variety of forms, as shown schematically in Figure 5. In this schematic illustration, the shape of the active site of the wild-type (the lock) is represented by a cross; for the mutant, the shape of the lock is represented by a hexagon. We seek an inhibitor (the key) that is sufficiently flexible such that it "fits" into both types of "lock". We need a drug (the key) that can bind not only with the active site of the wild type protease but also with that of its mutant. Thermodynamics can help to find the best drug candidates.

To illustrate, Figure 6 shows two possible drug candidates, (1) and (2). Drug (2) has better adaptability than drug (1) because, compared to drug (1), it has asymmetric functionality: the toluene group has less symmetry than the tertiary butyl group, as indicated by oval circles. More important, drug (2) is more flexible because it has two rotating bonds while drug (1) has only one, as indicated by dashed arrows. Asymmetry and flexibility provide the drug with additional conformations that can adapt to a mutant HIV binding site.

Toward obtaining a quantitative measure of drug efficacy, we can use results from isothermal titration calorimetry (ITC) reported by Ohtaka and Freire (8).

Let A stand for the HIV protease and B for the inhibitor. We define a dissociation constant K_d and its reciprocal, association constant K_a, where subscript a stands for association and subscript d for dissociation.

$$K_d = \frac{[A][B]}{[AB]} \qquad K_a = \frac{[AB]}{[A][B]} \qquad K_a = (K_d)^{-1}$$

where [ ] stands for concentration. For good drug efficacy, we want K_d to be small or, equivalently, K_a to be large.

The binding of A (protease) and B (inhibitor) is determined by enthalpy $\Delta H°$ and entropy $\Delta S°$ through the thermodynamic relation

$$-RT\ln K_a = \Delta G° = \Delta H° - T\Delta S° \tag{7}$$

where superscript ° indicates a standard state.

ITC provides experimental data for measuring K_d (or K_a), $\Delta G°$, $\Delta H°$ and $\Delta S°$.

Figure 7 shows experimental results for $\Delta G°$, $\Delta H°$, and $-T\Delta S°$ for twelve drug candidates where the protease is the wild form.

As indicated in Eq(7), for a successful drug, we want $\Delta G°$ to be large and negative. We can achieve a large negative $\Delta G°$ by either a large and negative $\Delta H°$ or else by a large and negative $-T\Delta S°$ or both.

If $\Delta G°$ is large and negative, the extent of binding is large; that's good. But we also want $\Delta H°$ to be large and negative because, in that case, we increase the strength of interaction of inhibitor (B) with protease (A).

In Figure 7, the first generation of drugs contains molecules that are relatively rigid (not flexible). Within the active-site pocket of the protease, rigid molecules cannot interact strongly with the functional groups of the protease and therefore, for these drugs, $\Delta H°$ is not favorable; in some cases, while negative, the absolute value is small and in other cases, most unfavorable, $\Delta H°$ is positive. However, $-T\Delta S°$ is favorable because upon binding, the hydrophobic drug releases

hydrated water molecules with a subsequent desirable gain in entropy.

Because rigid drug molecules do not have as much adaptability to mutant proteases as flexible molecules, second generation inhibitors contain relatively flexible drugs. As shown in Figure 7, these drugs show a favorable $\Delta H°$ but $-T\Delta S°$ is now less favorable because of what is called the entropic penalty. When a flexible "key" enters a confining "lock", it loses not only translational degrees of freedom (as does a rigid "key") but, in addition, it loses rotational degrees of freedom. These losses in freedom produce a decrease in entropy; because this decrease is unfavorable, flexible drugs exhibit a greater entropic penalty (smaller $|-T\Delta S°|$) than rigid drugs. However, the tradeoff for this entropic penalty is a more favorable binding interaction, i.e. a large and negative $\Delta H°$. Figure 7 suggests that drugs 11 and 12 are best since $\Delta H°$ and $\Delta G°$ are both large and negative; the inhibitors are adaptive, the benefits of which we will see in a moment.

Because protease mutations are common, we seek an inhibitor that is effective not only for the wild-type protease but, in addition, adaptive to some of its mutants.

Figure 8 shows ITC results for a series of inhibitors interacting with the wild type (WT) and with a particular mutant (M). The ordinate, on a logarithmic scale, shows the dissociation constant K_d; we want K_d to be small not only for the wild type but also for the mutant.

Figure 8 shows the importance of having a large, negative $\Delta H°$. The more adaptive second generation drugs, for example inhibitors 11 and 12, have the smallest dissociation constant for both the wild type and for the mutant HIV protease. In Figure 7, results for the wild type indicate that inhibitors 11 and 12 are best because they have two desirable features: both $\Delta G°$ and $\Delta H°$ are negative and large. However, Figure 8 shows that, while inhibitors 11 and 12 are good for the wild-type protease, inhibitor 11 is superior for inhibiting the mutant protease. Figure 8 suggests that drug 11 is better than drug 12.

We want $\Delta G°$ and $\Delta H°$ to be large and negative. We achieve large and negative $\Delta H°$ with a flexible inhibitor. Further, we prefer a flexible inhibitor because, unlike a rigid inhibitor, it has a better chance to be effective not only for the wild-type protease but also for a mutant. However, for a flexible inhibitor we must pay the price of an entropic penalty; this penalty is undesirable because it lowers the absolute value of a negative $\Delta G°$. In other words, we want an inhibitor that is flexible enough to give a large negative $\Delta H°$ and that is flexible enough to be useful for a protease mutant. However, the inhibitor should not be too flexible lest the entropic penalty become too large. Thermodynamic studies are useful for guidance toward designing the optimum inhibitor.

Conclusion

The examples shown here illustrate the utility of chemical thermodynamics. In each example, thermodynamics provides a theoretical framework for contributing toward obtaining an answer to a problem in biotechnology. Coupled with insights from molecular physics and with suitable experimental data, this framework can lead to useful results.

The generality and versatility of thermodynamics allows application to all materials, no matter how complex. We must not be afraid of biomacromolecules like proteins; they too are subject to the laws of thermodynamics.

More than 80 years ago, Gilbert Lewis and Merle Randall published a pioneering book (Thermodynamics and the Free Energy of Substances) that showed how the abstract chemical thermodynamics of Gibbs can be applied to real solids and fluids. Once again we are inspired by the book's first page:

> Let this book be dedicated to the chemists of the newer generation, who will not wish to reject all inferences from conjecture or surmise, but who will not care to speculate concerning that which may be surely known. The fascination of a growing science lies in the work of the

pioneers at the very borderland of the unknown, but to reach this frontier one must pass over well-traveled roads; of these one of the safest and surest is the broad highway of thermodynamics.

Acknowledgments:

For financial support, the authors are grateful to the National Science Foundation and to the Office of Basic Sciences, US Dept. of Energy.

References

1. F. W. Tavares, D. Bratko, A. Striolo, H. W. Blanch, J. M. Prausnitz. J. Chem. Phys. **120**, 9859 (2004).

2. F. W. Tavares, J. M. Prausnitz. Colloid 282, **620** (2004).

3. M. Boström, V. S. J. Craig, R. Albion, D. R. M. Williams, B. W. Ninham. J. Phys. Chem. B **107**, 2875 (2003).

4. F. W. Tavares, D. Bratko, H. W. Blanch, J. M. Prausnitz. J. Phys. Chem. B. **108,** 9228 (2004)

5. M. Boström, F. W. Tavares, B. W. Ninham, J. M. Prausnitz. J. Phys. Chem. B. article in press (2006)

6. L. Bergeron, L. F. Filobelo, O. Galkin, P. G. Vekilov. Biophys. J. **85**, 3935 (2003).

7. I. Reviakine, D. K. Georgiou, P. G. Vekilov. J. Am. Chem. Soc. **125**, 11684 (2003).

8. H. Ohtaka, E. Freire. Prog. In Biophys. & Mol. Bio. **88**, 193 (2005).