4

From:
Principles of Physical Biochemistry
(2004) van Holde, Johnson, Ho

Statistical Thermodynamics

4.1 GENERAL PRINCIPLES

Most of our observations of macromolecular properties in solution and in cells reflects the average behavior of a population. Even the native structure of a protein represents an ensemble of different conformations, all with the common property of showing biological activity. Statistical mechanics allows us to describe the distribution of molecular conformations that contribute to this population under a given thermodynamic state. The average properties of a macromolecule can be represented as either the time-average behavior of a single molecule or the behavior of an ensemble of many molecules at any instant in time. One of the basic postulates in a statistical-mechanical treatment of thermodynamic properties, or *statistical thermodynamics*, is that the time-average and ensemble average behaviors are identical. A time-average conformation of a structure can be modeled by carrying out a molecular dynamics simulation over a very long period of time (Chapter 3). In this chapter, we will discuss the behavior of molecular ensembles as described by statistical-mechanical methods.

Statistical mechanics is most often applied to the study of macromolecular systems that involve multiple thermodynamic states. This includes the transitions between different conformations in biopolymers, the assembly of multiple subunits into multicomponent complexes, and the binding of multiple ligands to macromolecules. In this chapter, we will focus on the application of statistical mechanics to structural transitions in well-defined biopolymers and see how this leads to approaches for predicting the conformations of proteins and nucleic acids. The application of these same statistical concepts to model the mechanisms of ligand binding are treated in Chapter 15.

4.1.1 Statistical Weights and the Partition Function

The simplest multiple-state system is one in which a molecule can exist in one of two forms, A or B. If the two forms are in equilibrium, with the reaction $A \leftrightarrow B$, their relative concentrations are expressed as the equilibrium constant $K_{eq} = [B]/[A]$. A and

B are the two possible states of the molecule in this system. From this, we can ask: What is the probability of observing the molecule as B relative to the probability of observing it as A? This is by definition K_{eq}. However, for multistate systems, the more general problem is to determine the probability of finding B versus all possible states ($P_B = [\mathrm{B}]/\Sigma[\text{all forms}]$). In the simple example above, this is specifically $P_B = [B]/([A]+[B])$, which is redefined as a function of K_{eq}.

$$P_B = \frac{[B]}{[A]+[B]} = \frac{\dfrac{[A]}{[A]}}{\dfrac{[A]}{[A]} + \dfrac{[B]}{[A]}} = \frac{K_{eq}}{1+K_{eq}} \tag{4.1}$$

In this simple two-state system, A serves as the reference state to which B is compared. As we expand upon this, it will generally be useful to compare all states of a system to a reference state. We therefore define the *statistical weight* ω as the concentration of all species at some state relative to that of a reference state (in this case, B as compared to the reference state A). The two states of this particular example can be represented by the statistical weights $\omega_A = [A]/[A] = 1$ and $\omega_B = [B]/[A]$, so that Eq. 4.1 can be rewritten as

$$P_B = \frac{\omega_B}{\omega_A + \omega_B} = \frac{\omega_B}{1+\omega_B} \tag{4.2}$$

This may appear to be a trivial redefinition of the equilibrium constant. However, for systems with multiple states, ω_j can be treated as a microequilibrium constant, referring to all species in any thermodynamic state j of a system to those of the reference state (which we will now designate as ω_o). In contrast, K_{eq} in such cases is a measure of the macroscopic behavior of the overall system, relating the concentrations of all products to all reactants.

When treated as an equilibrium constant, ω is related to the difference in the free energy between any state j (G_j°) and the reference state (G_o°). If each state of the system is unique, with a particular energy level, then

$$G_j^\circ - G_o^\circ = \Delta G_j^\circ = -RT \ln \omega_j \tag{4.3}$$

or

$$\omega_j = e^{-\Delta G_j^\circ/RT} \tag{4.4}$$

For the specific case of $A \leftrightarrow B$, P_B is given by Eq. 4.5.

$$P_B = e^{-\Delta G_B^\circ/RT}/(1 + e^{-\Delta G_B^\circ/RT}) \tag{4.5}$$

It is clear from this relationship that the state having the lowest energy is the most probable state of a system. Thus, if $\Delta G_B^\circ < 0$, then $\omega_B > 1$ and $P_B > 0.5$, or a majority of the molecules will be in form B. This is in accord with chemical intuition.

For a system having N possible states, the sum over all states (the denominator of this relationship) is simply the sum of the statistical weights for all states j. How the energy is distributed or partitioned across all the possible thermodynamic states in the system is defined as the *partition function* Q.

In our simple example, each state represents a single form of the molecule, A or B. However, if a particular state represents more than one isoenergetic form of the molecule, it is *degenerate,* and there is a higher statistical probability of observing that state. Consider for example two tossed coins that can land with either their head side (H) or tail side (T) up. The three possible states for the two coins are HH, TT, and HT. Both the HH and TT states are unique and thus the number of different combinations that result in either state (the degeneracy of each state, g_{HH} and g_{TT}) is 1. The HT state, however, represents two possible combinations with one coin landing heads up and the other tails up ($g_{HT} = 2$). The HT combination is thus twice as probable as either HH or TT. This is consistent with the second law of thermodynamics, which favors more degenerate states. The degeneracy of each state j therefore is reflected in the intrinsic entropy of that state ($S_j = R \ln g_j$). The degeneracy of each state is incorporated into the partition function as

$$Q = \sum_{j=0}^{N} g_j \omega_j \tag{4.6}$$

The probability P_j of observing a particular state j of a system therefore depends on the degeneracy and the statistical weight of that state relative to all possible states, where Q is now defined as in Eq. 4.6.

$$P_j = \frac{g_j \omega_j}{Q} \tag{4.7}$$

This result is identical to the Boltzmann distribution, as given in Chapter 2 by Eq. 2.9.

These are the basic relationships for the statistical-mechanics treatment of thermodynamic processes. We will describe applications of statistical mechanics to a number of structural transitions in polypeptides and polynucleotides in order to illustrate how ω and Q are derived and utilized. Before discussing the specifics of each system, we need to define in each case a model that describes the transitions between structures of the particular biopolymers.

4.1.2 Models for Structural Transitions in Biopolymers

The question of the probability of observing one of two possible states is the simplest case of a broader statistical-mechanics treatment of transitions between two or more defined forms of a molecule. A complete description must consider all the possible states of a system. A question that is often asked concerning the structure of a biological macromolecule is: What is its secondary structure? This limits the problem to transitions between a set of well-defined possible states, such as between a random coil and an α-helix or between a β-strand and an α-helix. For polynucleic acids, the two states can be, for example, the duplex and single-stranded forms of DNA or RNA (i.e., the melting and annealing of the double helix), or the standard right-handed (B-form) and the left-handed (Z-form) duplexes of DNA. We will therefore present a set of simple *two-state models* for structural transitions in biopolymers.