

---

## Chapter I

# Global transitions in proteins

The relation between structure and function is central to molecular biology. But molecular structure can mean different things, especially when dealing with complex biological molecules. One can know the chemical structure of a molecule, how the atoms are connected by covalent bonds, but have no idea of the conformational state, how the atoms are arranged in space. The conformational state of a molecule has a profound impact on what it does, and much of the work in molecular biophysics deals with understanding molecular conformations, both what they are and how they perform biological functions.

The conformational state of a molecule can be studied at different levels. One might have a vague notion of its general shape, or one might have a detailed picture with the position of every atom specified. Thinking in terms of detailed structure is more difficult but more powerful. Some approaches to this problem will be taken up in later chapters. Here, we will start with something simple, introducing an approach to protein conformations that does not depend on all the structural details.

The approach of this chapter is based on the idea that proteins have “global” states. Global states are defined in terms of a protein’s functional capability. We will assume that a protein has a few – perhaps just two – of these global states. Global states can interconvert, in what we will call global transitions. In terms of structure the global state is a black box. Without dealing explicitly with structure, the simplifying assumption of global states and transitions generates a robust quantitative framework for treating functional transitions of proteins.

These ideas can be applied to virtually any area of molecular biology. Conformational states of proteins are the basic building blocks in mechanistic models, and interconversions between these states are the basic molecular signaling events. Examples include the activation of membrane receptors, the regulation of gene expression, the control of cell division, the gating of ion channels, and the generation of mechanical force. This chapter focuses on two well-defined types of transitions, one induced by temperature and the other by voltage. These transitions are very different from an experimental point of view, and only rarely have they been studied in the same protein. However, from the

theoretical point of view we can see striking parallels, and studying these two cases together provides a deeper understanding of the general nature of functional transitions in proteins.

## 1.1 | Defining a global state

The general strategy for now is to play down structural details, but we still need to define a global state rigorously. This will help make us aware of how the global state can at least in principle be related to a protein's structure. Global states are a coarse-grained view but they can be related to fine-grained views. The fine-grained view is based on what will be called microstates. A microstate has a conformation that is defined in great detail. We might know the positions of all of the atoms, or the dihedral angles of all of the rotating bonds (Chapter 3). A global state is envisioned as encompassing a large number of microstates. The microstates interconvert rapidly and the global state reflects the average behavior of these microstates. This view is taken directly from statistical mechanics. A collection of microstates forms an ensemble, and statistical mechanics provides the conceptual tools for understanding a global state in terms of its constituent microstates.

The free energy of a global state takes into account the internal potential energy of each of the microstates as well as the entropy arising from the conformational disorder of interconversions between microstates. We can express the free energy of a system containing  $N$  independent molecules as

$$G = -kT \ln(Q^N) \quad (1.1)$$

where  $Q$  is the partition function of the molecule and  $kT$  is Boltzmann's constant times temperature. The partition function is a sum over a set of microstates included in one global state

$$Q = \sum_i^{n_{GS}} e^{-E_i/kT} \quad (1.2)$$

Here  $E_i$  is the energy of the  $i$ th microstate and  $e^{-E_i/kT}$  is its Boltzmann weight. When we focus on one particular global state, we limit this sum to a selected subset, or subensemble of microstates. This is indicated by the subscript GS; so  $n_{GS}$  is the total number of microstates comprised by a given global state. We can thus distinguish different global states formally by summing over different, non-overlapping subsets of microstates.

The Boltzmann distribution can be used to obtain the probability of finding a particular microstate,  $j$ , with energy  $E_j$ , among all of the possible microstates of a given global state

$$P(j) = \frac{e^{-E_j/kT}}{\sum_i^{n_{GS}} e^{-E_i/kT}} \quad (1.3)$$

The sum in the denominator is the partition function of Eq. (1.2). Here, it normalizes the probability function so that the probabilities all add up to one. In this context, as a sum over probabilities, we can see how the partition function embodies the notion of thermodynamic stability. A global state with a greater partition function has a higher probability of occurring, and so will have a lower free energy.

The definition of a microstate is very flexible and can be expanded to include many important features. For example, each microstate has an entropy resulting from the disordering effect of bond vibrations. To take this into account we can extend the sum in Eq. (1.2) to include vibrational energy levels. Microstates of a protein can be further distinguished by different positions and orientations of the surrounding water molecules, and possibly by ions in solution. If these contributions are included in the summation in Eq. (1.2), then the free energy in Eq. (1.1) will be more accurate. These are formal considerations that help us visualize in broad terms how different levels of detail can be incorporated into the picture. The free energy of a global state does indeed depend on all of these complex features. We will recognize these dependencies, but for now we will not deal with them explicitly.

An alternative way to think about the partition function of a global state is to view the discrete states in Eq. (1.2) as a continuum. The sum then becomes an integral over a specified range of all the internal coordinates of the molecule

$$Q = \int_{GS} e^{-E(\mathbf{r})/kT} d\mathbf{r} \quad (1.4)$$

Here  $\mathbf{r}$  is a vector containing the positions of all of the atoms, and  $E(\mathbf{r})$  is the potential energy of the molecule as a function of these positions. Equation (1.4) is referred to as the classical configuration integral (McQuarrie, 1976). The range of integration, defined as GS, specifies a global state by limiting the region of coordinate space. This is analogous to limiting the total number of microstates to a subset of  $n_{GS}$  microstates in Eq. (1.2). Limiting the range of the integration of Eq. (1.4) and limiting the range of the summation in Eq. (1.2) are equivalent ways of dividing the vast state space of a protein into distinct regions corresponding to different global states.

It must be mentioned that Eqs. (1.2) and (1.4) leave out the contribution made by kinetic energy. This does not matter for our purposes because in classical physics (i.e. no quantum mechanical effects) every atom has an average kinetic energy of  $3/2 kT$ . As long as classical physics is obeyed this will be the same for all microstates. For strong covalent bonds, quantum effects are important and the vibrational kinetic energy for an atom can be less than  $3/2 kT$ . However, this contribution is not likely to change much during global transitions because the covalent bonds are not broken. These fortunate circumstances make Eqs. (1.2) and (1.4) especially well-suited for studying macromolecules in biophysical problems.

For now we have no intention of actually calculating the partition function of a protein. We will accept the existence of the free energy of a global state, defined in these terms. The free energy includes all of the energetic terms that make one global state more stable than another for a particular set of conditions. Without a complete accounting of all these terms we can still develop useful theories for what influences global transitions in proteins.

Models based on the idea of global states are very popular in molecular biology, but the conceptual basis and underlying assumptions often go unappreciated. In fact, we do not actually have a good reason to expect a priori that a protein will have just a few discrete global states, or that the behavior of a protein can be described well in these terms. One could imagine that functional changes in proteins involve shifts in the distribution of microstates. Each functional state could have access to all the same microstates, but would visit some more frequently than others. Thus, we have two extreme views of functional transitions in proteins. Global transitions at one end of the spectrum represent transitions between distinct nonoverlapping populations of microstates. The other extreme is a transition entailing a shift in the distribution within a single well-connected population of microstates. Deciding which of these extremes better approximates reality requires experiments, and some of these will be discussed as the theories are developed.

The variable  $E_i$  in the above expressions is the energy of a single molecule. However, we will often find it more convenient to use the energy of one mole of molecules. We can use these units provided that we replace Boltzmann's constant,  $k$ , with the gas constant,  $R$ . One often has to switch between molecular and molar energy units, and this is usually easy if we keep track of whether we are using  $k$  or  $R$ . To have some sense for energy magnitudes on the molecular versus the molar scale, consider the value of  $kT$  and  $RT$  for a physiological temperature of 300 K:  $kT = 4.11 \times 10^{-21}$  J and  $RT = 2.47 \times 10^3$  J ( $RT$  also equals 590 calories, and these units are more commonly used with moles). Energies are divided by these numbers when the Boltzmann distribution is used to determine the probability of a state, so they are well worth remembering. Therefore  $kT$  and  $RT$  are the fundamental reference points for energy. The energy of a state relative to these reference points determines how likely we are to encounter that state.

## 1.2 | Equilibrium between two global states

Consider an equilibrium between two global states, denoted as A and B, as shown in Scheme (1A) below



This is just like a chemical isomerization. Global state A has a molar free energy  $G_a$  and global state B has a molar free energy  $G_b$ .

The molar free energies of solutions with certain concentrations of A or B are then

$$G_a = G_a^\circ + RT \ln[A] \quad (1.5a)$$

$$G_b = G_b^\circ + RT \ln[B] \quad (1.5b)$$

where  $G_a^\circ$  and  $G_b^\circ$  are the molar free energies of the standard state (by convention a one molar solution). The free energy change for conversion from A to B is then

$$\begin{aligned} \Delta G &= G_b - G_a \\ &= G_b^\circ - G_a^\circ + RT \ln \frac{[B]}{[A]} \end{aligned} \quad (1.6)$$

At equilibrium  $\Delta G = 0$ . Taking the equilibrium concentrations as  $[A]_{eq}$  and  $[B]_{eq}$ , we can relate these to the free energy difference between the two standard states as

$$\Delta G^\circ = -RT \ln \frac{[B]_{eq}}{[A]_{eq}} = -RT \ln K_{eq} \quad (1.7)$$

where  $\Delta G^\circ = G_b^\circ - G_a^\circ$ , and  $K_{eq}$  is the equilibrium constant for inter-conversions between A and B. Taking the exponential of this equation gives

$$e^{-\Delta G^\circ/RT} = \frac{[B]_{eq}}{[A]_{eq}} \quad (1.8)$$

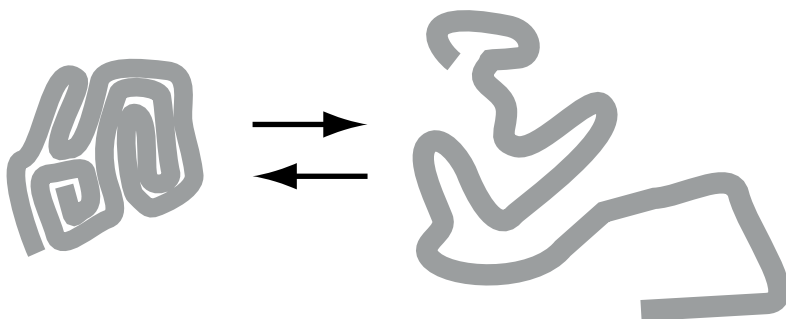
We will drop the subscripts “eq” after this point because all concentrations are equilibrium values for the rest of the chapter.

Equation (1.8) looks very much like a Boltzmann distribution, which gives the ratio of the relative probabilities of two microstates as an exponential function of the difference in their energies. A corresponding relation for the relative probabilities of two microstates can be derived from Eq. (1.3). Likewise, Eq. (1.8) can be derived directly from Eq. (1.3) by summing over the microstates for each of the two global states and then taking the ratio of the two sums. This connection emphasizes the point that the equilibrium represented by Eq. (1.8) reflects the relative stability of two distinct collections of microstates. Equation (1.8) is a useful starting point for introducing assumptions about energies into models for global transitions. The general strategy in much of what follows is to make assumptions about  $\Delta G^\circ$  and use Eq. (1.8) to explore the consequences.

## 1.3 | Global transitions induced by temperature

The thermal denaturation of a protein provides an excellent example of a global transition. At low temperatures a folded state is favored in which a relatively small number of microstates have a similar structure corresponding to a compact well-defined configuration (Fig. 1.1). This

**Fig. 1.1.** In the thermal denaturation of a protein, a compactly folded native state undergoes a global transition to a randomly coiled denatured state.



configuration is maintained by a large number of specific contacts between the residues (Section 2.13). At high temperatures the protein unfolds (Fig. 1.1). This denatured state has a very large number of microstates corresponding to all the different configurations of a randomly coiled chain (Chapter 3). The transition is reversible, so that cooling restores the folded state.

It will be seen shortly that the difference in number of microstates in these two global states is what makes the transition temperature sensitive. In keeping with the spirit of this chapter in which structural details are played down, we simply accept the situation as given. Chapter 3 uses statistical mechanics to develop models for this process.

To understand how temperature influences thermal denaturation, we divide the free energy change into enthalpy and entropy

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (1.9)$$

The enthalpy can be loosely connected with the contacts that stabilize the native, folded state. The entropy represents the contribution arising from the greater disorder of the unfolded state. This is an oversimplification as the contact energies are not purely enthalpic. More will be said about this in Chapter 2, but the separation of enthalpy and entropy along these lines is useful and appropriate for the present discussion of thermal unfolding. It is also important to mention that contacts can stabilize the unfolded state. The unfolded protein has no real structure to stabilize, but stronger contacts between buried amino acids and water or another solvent in the unfolded state factor into  $\Delta H^\circ$ . This is the mechanism of denaturant action (Section 1.7).

Assuming that  $\Delta H^\circ$  and  $\Delta S^\circ$  themselves are independent of temperature allows us to focus on the term  $T\Delta S^\circ$  as the drive for a temperature-induced shift in the equilibrium between A and B. There is a transition temperature,  $T^*$ , where the folded and unfolded states are present in equal concentrations. At that point  $\Delta G^\circ = 0$ , so by Eq. (1.9)  $\Delta H^\circ = T^*\Delta S^\circ$ . Below  $T^*$ ,  $\Delta H^\circ > T\Delta S^\circ$ , and  $\Delta G^\circ$  is positive so that A is more stable. Above  $T^*$ ,  $\Delta H^\circ < T\Delta S^\circ$  and  $\Delta G^\circ$  is negative so B is more stable. This is readily seen by combining Eqs. (1.8) and (1.9).

$$\frac{[B]}{[A]} = e^{(-\Delta H^\circ + T\Delta S^\circ)/RT} \quad (1.10)$$

Practical applications often proceed by defining a variable  $x$  as the fraction of B or equivalently, the extent of the conversion from A to B.

$$x = \frac{[B]}{[B] + [A]} \quad (1.11)$$

With the aid of Eq. (1.10),  $x$  can be expressed in terms of  $\Delta G^\circ$ . Factoring  $[B]$  from the numerator and denominator gives

$$x = \frac{1}{1 + ([A]/[B])} \quad (1.12)$$

Substituting Eq. (1.10) then gives

$$x = 1 / (1 + e^{\frac{\Delta H^\circ}{RT} - \frac{\Delta S^\circ}{R}}) \quad (1.13)$$

At low temperatures the positive enthalpy term is larger, so the exponential is large, and  $x$  is close to zero. At high temperatures the negative entropy term dominates, so the exponential is small and  $x$  is close to one. Thus, Eq. (1.13) expresses the conversion from  $x \sim 0$ , where  $[A] \gg [B]$ , to  $x \sim 1$ , where  $[B] \gg [A]$ .

## 1.4 | Lysozyme unfolding

Thermal denaturation has been studied extensively in an effort to determine how different forces contribute to the stability of the native, folded state of a protein. These forces will be discussed in more detail in Chapter 2, but for the present purposes, these kinds of experiments provide a good illustration of how global transitions are modified by structural changes.

The enzyme lysozyme from T4 bacteriophage was subjected to a screen to select temperature-sensitive mutants. Lysozyme unfolds or denatures as the temperature is raised, and the unfolded state has no enzymatic activity. In temperature-sensitive mutants denaturation occurs at lower temperatures than in the wild type enzyme. Mutants with lower melting points identified residues that are critical to thermal stability. One of these was threonine 157. Using site-directed mutagenesis, this threonine was replaced with a number of different amino acids (13 in all) (Alber *et al.*, 1987).

In temperature-sensitive mutants the free energy difference between the folded and unfolded states is altered. This can occur through changes in the strength of the contacts of the amino acid side chain with its neighboring residues, leading to a change in enthalpy. Alternatively, the new amino acid side chain may have more or less conformational freedom, thus changing the entropy. It was found that  $T^*$  dropped for every amino acid replacement of threonine 157. When the mutations were studied by calorimetry (Connelly *et al.*, 1991) it was found that the destabilization reflected both reductions in  $\Delta H^\circ$  and increases in  $\Delta S^\circ$ .

**Fig. 1.2.** (a) A plot of enthalpy versus temperature for wild type lysozyme illustrates the basic features of a melting curve. The total change gives the  $\Delta H^\circ$  of the transition;  $T^*$  is the half-way point. (b) Mutant forms of T4 lysozyme melt at lower temperatures. Isoleucine (I), valine (V), and alanine (A), each replaced threonine 157 of the wild type (WT). Enthalpy was calculated as  $x\Delta H^\circ$  using Eq. (1.13). The measurements of  $\Delta H^\circ$  and  $\Delta S^\circ$  were from Connelly *et al.* (1991). (c) The sketch shows the threonine 157 side chain hydroxyl hydrogen bonded with a hydroxyl from threonine 155 and the backbone amide of aspartate 159 (see Alber *et al.*, 1987).

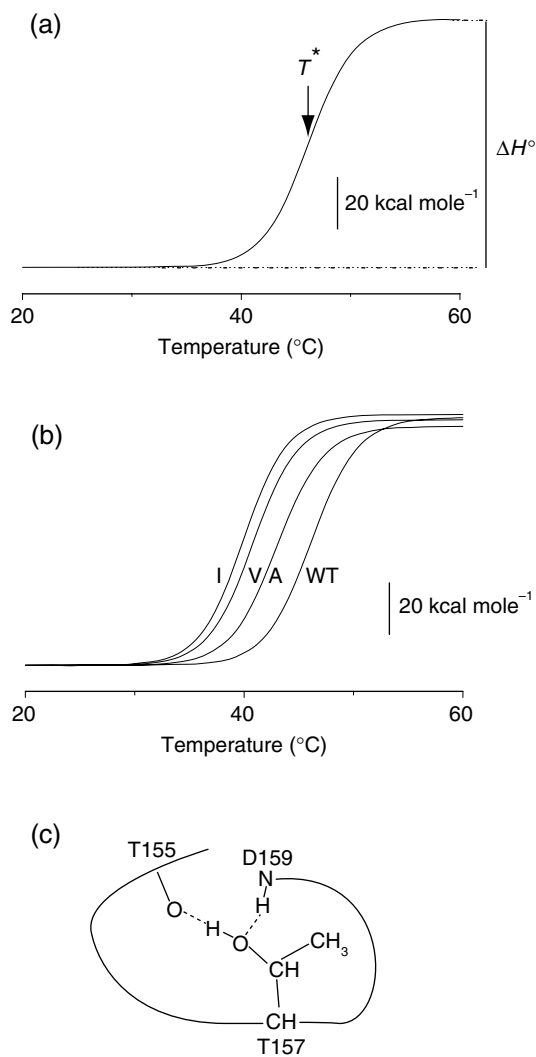


Figure 1.2 shows calculated melting curves for the wild type protein and a few mutants. As the temperature is raised the protein absorbs heat. This can be measured as enthalpy, using a calorimeter that scans temperature and records the heat absorption. These plots are based on Eq. (1.13), using experimentally measured values of  $\Delta H^\circ$  and  $\Delta S^\circ$ . The variable  $x$  computed in this way from Eq. (1.13) is then multiplied by  $\Delta H^\circ$  to obtain the excess enthalpy, which is plotted. Baseline heat absorption by the solvent and by the global states of the protein unrelated to the transition are subtracted out to leave the heat of the transition. So the heat absorbed as a result of the unfolding transition is easily resolved.

The wild type melting curve (Fig. 1.2a) indicates that the total heat absorbed is the  $\Delta H^\circ$  for the protein. The half-way point is the melting temperature,  $T^*$ , where  $x = 1/2$ , and the heat is  $\Delta H^\circ/2$ . Once  $\Delta H^\circ$  and  $T^*$  have been read off such a plot,  $\Delta S^\circ$  is determined as  $\Delta H^\circ/T^*$ .



The mutant melting curves are shown in Fig. 1.2b, along with the wild type melting curve. Figure 1.2c shows a sketch of the structure around the mutated site (threonine 157). The alanine mutation shifts  $T^*$  down by 4 °C, and the high temperature plateau is reduced. Both of these changes in the melting behavior are consequences of the lower  $\Delta H^\circ$  of this mutant. The reduction in  $T^*$  follows from  $T^*\Delta S^\circ = \Delta H^\circ$ , for a mutation where  $\Delta S^\circ$  is the same as in the wild type protein. The reduction in  $\Delta H^\circ$  for this mutation reflects the loss of the two hydrogen bonds formed by the hydroxyl group of threonine 157 (Fig. 1.2c). Putting a valine in this position reduces  $T^*$  even more, but the final plateau is close to that of the wild type protein, so this destabilization is primarily caused by an increase in the  $\Delta S^\circ$  of the transition, rather than a decrease in  $\Delta H^\circ$ . This may reflect a restriction in the motion of neighboring side chains caused by the larger valine side chain. Finally, with isoleucine  $\Delta H^\circ$  is higher than wild type. If  $\Delta S^\circ$  were the same then  $T^*$  would go up. But this mutation produces the greatest reduction in  $T^*$ , indicating that the increase in  $\Delta H^\circ$  is offset by a much larger increase in  $\Delta S^\circ$ . The isoleucine side chain forms contacts with other parts of the protein in a structure stabilizing manner, but it is too big for the pocket occupied by threonine, so the motions of the side chains are more severely restricted. Furthermore, isoleucine can be more disordered in the unfolded state because of the flexibility of its large side chain, so  $\Delta S^\circ$  increases owing to changes on both sides of the transition.

## 1.5 | Steepness and enthalpy

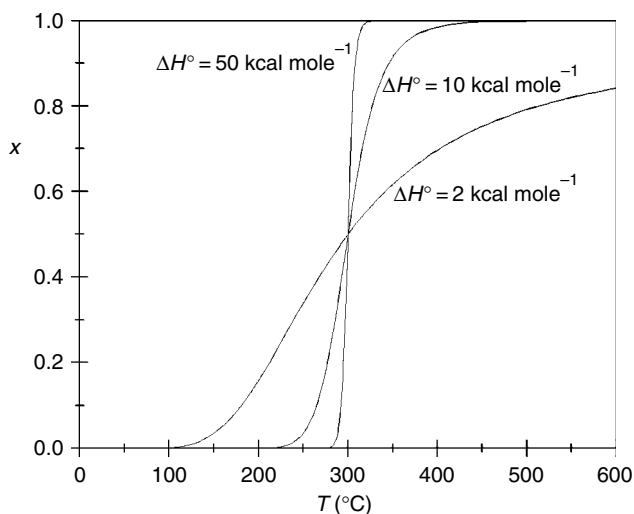
A thermal transition reflects competition between enthalpy and entropy, so that changing one or the other moves the transition temperature up or down. Now we ask, what if  $\Delta H^\circ$  and  $\Delta S^\circ$  change in parallel? The value of  $T^*$  then stays the same but the plots change and this reveals another interesting aspect of thermal transitions. Equation (1.13) is plotted in Fig. 1.3 for transitions with the same  $T^* = 300$  °C. Different values of  $\Delta H^\circ$  are used, and  $\Delta S^\circ$  was taken as  $\Delta H^\circ/T^*$  for each plot. Figure 1.3 shows how the transition becomes steeper as  $\Delta H^\circ$  and  $\Delta S^\circ$  increase together.

The increasing steepness with increasing  $\Delta H^\circ$  gives us a useful way of thinking about a thermal transition. The  $\Delta H^\circ$  (or equivalently,  $\Delta S^\circ$ ) dictates the sensitivity of the transition to temperature. If the  $\Delta H^\circ$  and  $\Delta S^\circ$  are larger, then the  $\Delta G^\circ$  changes more rapidly as  $T$  passes through  $T^*$ .

This relation between the temperature dependence of an equilibrium and its  $\Delta H^\circ$  and  $\Delta S^\circ$  is firmly rooted in classical thermodynamics. Start with the following basic relation

$$\frac{\partial G^\circ}{\partial T} = -S^\circ \quad (1.14)$$

**Fig. 1.3.** Plots of Eq. (1.13) with  $T^* = 300^\circ\text{C}$  and three values of  $\Delta H^\circ$  ( $\Delta S^\circ = \Delta H^\circ/T^*$ ).



This also applies to the changes in these quantities

$$\frac{\partial \Delta G^\circ}{\partial T} = -\Delta S^\circ \quad (1.15)$$

This amounts to a quantitative statement that  $\Delta S^\circ$  gives the temperature dependence of an equilibrium, and replacing  $\Delta S^\circ$  with  $\Delta H^\circ/T^*$  makes an analogous point for enthalpy.

Using Eq. (1.15) to substitute for  $\Delta S^\circ$  in Eq. (1.9) gives

$$\Delta G^\circ = \Delta H^\circ + T \frac{\partial \Delta G^\circ}{\partial T} \quad (1.16)$$

Rearranging and dividing by  $T^2$  leads to

$$\frac{1}{T} \frac{\partial \Delta G^\circ}{\partial T} - \frac{\Delta G^\circ}{T^2} = -\frac{\Delta H^\circ}{T^2} \quad (1.17)$$

from which we obtain what is known as the Gibbs-Helmholtz equation

$$\frac{\partial(\Delta G^\circ/T)}{\partial T} = -\frac{\Delta H^\circ}{T^2} \quad (1.18)$$

This can be checked by evaluating the derivative on the left. Using Eq. (1.7) takes us one step further

$$\frac{\partial \ln([B]/[A])}{\partial T} = \frac{\Delta H^\circ}{RT^2} \quad (1.19)$$

This expression is known as the van't Hoff equation, and provides an important way to interpret the temperature dependence of an equilibrium between two global states. The slope of a plot of  $\ln([B]/[A])$  versus temperature can be used to determine  $\Delta H^\circ$ . An enthalpy determined in this way is often referred to as a van't Hoff enthalpy, or  $\Delta H_{\text{vh}}^\circ$ .

Note that  $\Delta S^\circ$  and  $\Delta H^\circ$  cannot be viewed in isolation from one another when thinking about temperature effects. For the sake of