Introduction

Because this lecture draws on material from several different sources, I will give you a fair amount of description to read. Cantor and Schimmel set up the important questions in Chpt 22:

- How do nucleic acids behave in aqueous solution?
- What kinds of ordered structures do they form?
- How can the observed physical and chemical properties be rationalized in terms of these structures?
- How does a nucleic acids convert from one structural form to another?

Consideration of nucleic acid folding principles versus protein folding principles

- There are only 4 side chains in nucleic acids versus 20 in proteins. These side chains are chemically very similar. Therefore the interactions are simpler.
- Most of the interactions occurring in intact nucleic acids also occur, in almost identical form, in the smallest fragments of nucleic acids, in dinucleotides, and in compounds one or two nucleotides longer. In contrast, short peptides are useless as models for proteins.
- The field of RNA folding has advanced to the point where tertiary structure folding can now be examined. However, we will focus on secondary structure formation for the time being.

Structure and stability of nucleic acid single strands

- We all know that dsDNA forms Watson-Crick helixes. But what about individual non-self-complementary strands? Synthetic homopolymers? Dinucleotides like CpC? Do these form structures?
- Construct limiting models (hypotheses):
  *** State 1: Random Coil?
  State 2: Helical (same geometry of strand as in duplex)?
  *** State 3: Helical (different geometry of strand as in duplex)?
- The answer is that State 3 is correct for most sequences, although it is in equilibrium with State 1 and the equilibrium depends on sequence, pH, temperature, solvent, ...
What is the experimental evidence for structure of single strands?

- When bases stack, the dipoles in the bases change. Any spectroscopic technique that depends on dipoles will show a different signal for the folded state.
- CD, UV hypochromism, and NMR are some of the methods that can detect stacking.
- A random coil (=State 1) should give optical or NMR signals approaching a mixture of the same monomers. This is seen experimentally at high T (≈100 °C) for single strands.
- However, at lower T the optical and NMR properties are much more intense for single strand molecules than for a mixture of the monomers. e.g. signs, positions and number of CD (=ellipticity) bands in the near UV change dramatically. These data suggest that single strand molecules are structured at low T (=State 3). See Fig 22-1 below, from Cantor and Schimmel.

Figure 22-1
Circular dichroism (per residue) of poly rA as a function of temperature. Data were obtained in 0.1 m NaCl (pH 7.4) at (1) 2 to 6 °C, (2) 17 °C, (3) 34 °C, (4) 42 °C, (5) 57 °C and (6) 70 °C. The CD of the monomer adenine is also shown (at 0 °C); it is essentially temperature independent. [After P. O. P. T'jo, in Basic Principles in Nucleic Acid Chemistry, vol. 2, ed. P. O. P. T'jo (New York: Academic Press, 1974).]

- How does single-strand stacking vary with sequence? See Table 22-1 below. You can see that polyA is a good stacker, as is polyC. Poly U is not. (see also L3p16 for $\Delta G^\circ$ for these sequences).
How does single-strand stacking vary with identity of the sugar? See Fig 22-2 below. Changing from ribose to deoxyribose has a profound effect. Since this doesn’t change the nature of the electronic interaction in the nucleobases, one concludes that the effect of the ribose is indirect. Changing the ribose ring changes the sugar pucker, which changes the stacking.

What are the global structural features of single strands?

- We can clearly see that poly A is changing conformation and unfolding as T is increased (Fig 22-1). But can we say anything about the nature of the structures populated? The shape of the CD curves in Fig 22-1 all look similar to each other. This suggests that the amount of structure formed is different, but type of structure formed is similar.
- Felsenfeld measured degree of stacking in poly A as a function of T by optical techniques to get fraction of stacking versus T (see later in this lecture for how).
- He then measured rigidity of the polymer as a function of T by hydrodynamic techniques.¹
- He then plotted rigidity versus degree of stacking (see Fig 22-3 below).

---

¹ This was done by sedimentation using the ultracentrifuge. The information content of this method is low. They extract the radius of gyration, $<R_g^2>$¹/², you can think of this as representing the shape of the molecule. Big radius of gyration means the molecule has an extended shape, small means compact shape. Other common methods are SAXS (small angle X-ray scattering).
Only where the bases are fully stacked (near 0 °C) does the polymer show a marked increase in extension and rigidity. At intermediate degrees of stacking (say 37 °C), ordered regions must be too short to extend the helix.

Poly U at room T has a radius of gyration that is close to poly A at high T. This confirms that poly U does not form ordered structure (except at high salt and very low T).

What are the detailed structural features of single strands?

One can imagine three possible structures of single stranded molecules: base stacked, base paired, and intercalated (Fig. 22-4).

Which of the structures in Fig 22-4 forms?

1. Base stacked? Supported by crystal structure of poly C (see Fig. 22-5). Supported by NMR of dinucleotides (see below).

2. Base paired? No evidence in non-self-complementary oligomers based on absence of imino protons in NMR, absence of hydrogen-bonded interactions in the IR, and based on crystal structures. Methylation of purine N1, N6 and N7 does not block single-strand structure in poly A. Chemical reagents modify N6 rapidly in poly A.

3. Intercalated? Unlikely due to close apposition of phosphates from opposite strands. No structural data to support.
Using NMR of to investigate single-strand stacked structures.

- What are the detailed arrangements of the bases in single-strand stacked structures? To investigate this, model dinucleotides have been used. These are good models of stacks in larger single-strand molecules based on similar trends for di- and polynucleotides in Table 22.1.
- NMR is well-suited to look at local structure since the chemical shift of a proton depends on its chemical environment.

Quick NMR review:

1. Coupling constants \( J \) values are a function of the dihedral angle between two interacting protons, and the value is independent of the chemical shift. \( J \) is measured in Hertz and is independent of the field strength; chemical shift is in ppm. Therefore, larger magnetic fields give better resolved sets of peaks.

   e.g.: H5 and H6 in pyrimidines must have the same coupling constant, \( J_{5,6} \), even though they have different chemical shifts. Since H5 and H6 are in a stiff aromatic nucleobases, \( J_{5,6} \) will not change as the dinucleotide stacks. Likewise, \( J_{1',2'} \) is likely to change upon stacking since sugar pucker is easy to change.

2. Large ring-current magnetic anisotropy causes chemical shift changes. If the proton is over the ring, then an upfield (=to the right) shift. If in the plane of the ring, then a downfield (=to the left) shift occurs. The ring current of A is much larger than C (=a highly aromatic purine versus a weakly aromatic pyrimidine).

- In Fig 22-9, compare the spectra of (CpA and ApC) to (A and C). Note the following features:
  1. In the dinucleotides, all C protons (H1’, H5 and H6) shift upfield. Therefore, these protons are near A.
  2. H5 shifts more dramatically in ApC than in CpA.
  3. A structure is given for CpA and ApC. Note that H5 is directly over A in ApC but not in CpA.
Figure 22-9
Conformational analysis of ApC and CpA by NMR spectroscopy. (a) Chemical structure of ApC, showing protons detected by NMR in color. (b) Time-average NMR spectra of 0.01 M solutions at neutral pH in D₂O at 29°C. Chemical shifts are indicated below each peak, in ppm; coupling constants are indicated above, in cps. (c) Possible stacked conformations consistent with the NMR data. [After W. Bangerter and S. Chan, J. Am. Chem. Soc. 91:3910 (1969).]
Thermodynamics of single strand stacked structures

- $\Delta H^\circ$ is proportional to the steepness of a melting (optical signal versus $T$) transition (more on this in a later lecture).
- Single-strand molecules have very broad melting transitions, and therefore small $\Delta H^\circ$ (average value is ca. –7 kcal/mol. See Table 8.1 in BCT).

Method 1: Differential Scanning Calorimetry (DSC)
- Some experts in this field: Breslauer (Rutgers), Record (Madison), Lohman (Wash. U), Makhatadze (Hershey).

Differential Scanning Calorimeter Schematic
(Microcal makes the best one—see reading list on class webpage for a link on the DSC theory):

- Raise temperature in sample and reference cells. Measure excess heat required to raise temperature in sample cell. Recall that heat capacity is the amount of heat required to raise the temperature of a sample 1 °C. Here, we get excess heat capacity, the extra heat required to heat your sample compared to your buffer.
- We only want the heat absorbed in the unfolding transition. A problem is that heat is absorbed by the random coil (=State 1) and folded (=State 3) molecules themselves. This can be corrected by a subtraction using baseline corrections (see Cantor and Schimmel insert from pp1086-1087 below).

\[
d\Delta H^\circ/dT = \Delta C_p^\circ
\]

or, integrating

\[
\Delta H^\circ = \int_{T_1}^{T_2} \Delta C_p^\circ dT
\]
Differential scanning calorimetry has gained widespread use in recent years and has emerged as an important tool for investigating a diversity of systems. The essential experimental purpose is to measure the heat absorption in a system as the temperature is raised from some initial to some final temperature, where the system under investigation has one or more transformations (such as conformational changes if the system consists of a protein dissolved in aqueous buffer media). Measurements are taken simultaneously with a suitable reference solution, so as to cancel out heat changes due to effects of no interest. Thus, measurements of a protein in buffer would be compared with those of buffer alone.

The resulting data often are plotted as $C_p$ (constant-pressure heat capacity) versus $T$. For a single transformation, such as a protein conformational change $N \rightleftharpoons D$, a curve such as that shown here might be obtained. The peak in $C_p$ occurs in the transition region as the molecule goes from the N to the D form.

To calculate $\Delta H$ for the transition, assume that the temperature is raised from an initial value $T_i$ (below the transition region) to a final value $T_f$ (above the transition region). The total heat change $\Delta H_{\text{tot}}$ in going from $T_i$ to $T_f$ is

$$
\frac{\Delta H}{\Delta T} = \Delta C_p
$$

$$
\Delta H_{\text{tot}} = \int_{T_i}^{T_f} C_p \, dT
$$

where $C_p$ is the apparent (measured) heat capacity at constant pressure. The $\Delta H_{\text{tot}}$ value contains not only the enthalpy change associated with the transition from N to D, but also the "background" heat absorption that comes as a result of heat absorption by N and D molecules that are present in the region between $T_i$ and $T_f$.

To compute just $\Delta H(T)$—the enthalpy change associated with the transition at temperature $T$—we must subtract from $\Delta H_{\text{tot}}$ the contribution to the total heat absorption that the N species makes between $T_i$ and $T$ and that the D species makes between $T$ and $T_f$:

$$
\Delta H(T) = \Delta H_{\text{tot}} - \int_{T_i}^{T} C_p^N \, dT - \int_{T}^{T_f} C_p^D \, dT
$$

where $C_p^N$ and $C_p^D$ are the heat capacities of the N and D forms, respectively. Referring to the figure, we note that $\Delta H(T)$ is given by the shaded area under the curve. The equation can be used to calculate $\Delta H(T)$ at any temperature between $T_i$ and $T_f$.

In some systems, more than one peak may be seen in a plot of $C_p$ versus $T$. This is evidence for more than one structural transition, occurring at distinctly different temperatures.

Method 2: Optical Melting (=temperature dependence of spectroscopic properties)

A. For folding transitions of single strand dinucleotides.

(N.B.: We will consider optical melting of double stranded structures in a future lecture.)

- The folding of dinucleotides, while perhaps not the most biologically interesting process you might think of, is very useful to consider. It provides a nice introduction to the formalism for interpreting optical melting curves, and the thermodynamic values we obtain will be useful for thinking about folding of larger nucleic acids, both ss and ds.

- Consider the process of stacking (st) and unstacking (u) of a dinucleotide

\[
\text{NpN}_u \xrightarrow{k_u} \text{NpN}_{st} \xleftarrow{k_{st}} \text{NpN}_u
\]

- You can consider this a simple two-state process. Therefore simple thermodynamic relationships apply.

\[
K_{st} = \frac{\text{(NpN}_{st})}{\text{(NpN}_u)}
\]

\[
\Delta G^\circ_{st} = -RT \ln K_{st} = \Delta H^\circ_{st} - T \Delta S^\circ_{st}
\]

- If \( K_{st} \) is known as a function of temperature, then \( \Delta H^\circ \) can be determined from the slope of a van’t Hoff plot of \( \ln K \) versus \( 1/T \).

\[
\frac{\partial \ln K}{\partial (1/T)} = -\frac{\Delta H^\circ}{R}
\]

- And \( \Delta S^\circ \) can be determined from

\[
\Delta S^\circ = (\Delta H^\circ - \Delta G^\circ)/T
\]

- Note that \( \Delta H^\circ \) determined by methods 1 and 2 should be the same. Also, the \( \Delta H^\circ \) obtained by method 2 should not depend on the optical method used. (i.e.: UV melts, CD melts, and NMR melts should all give the same \( \Delta H^\circ \) if the melt is truly two-state.) This is an important point.

---

2 A two-state process says that all the molecules exist in either of two possible states. When the \( T_{TM} \) is reached, 50% of the molecules are in state 1 and 50% in state 2; this is not the same as 100% of the molecules in a state that is 50% state 1 and 50% state 2.
In practice, How do we get the thermodynamic parameters from the optical melt?

- Let \( C_T \) be the total strand concentration.
- Let \( \alpha_s = \alpha \) be the fraction of strands stacked and \( \alpha_u = \) fraction of strands that are unstacked.
- Let \( s \) be \( N_p N_{st} \) and let \( u \) be \( N_p N_u \). Therefore, we have.

\[
C_T = [s] + [u]
\]

\[
\alpha_s = s / C_T
\]

\[
\alpha_u = u / C_T \quad \text{and} \quad \alpha_s + \alpha_u = 1
\]

Therefore,

\[
s = \alpha_s C_T
\]

\[
u = (1 - \alpha) C_T
\]

Therefore,

\[
K = \alpha / (1 - \alpha)
\]

- How do you get \( \alpha \) from spectroscopy? Consider absorbance, A.

\[
A(T) = \alpha_s A_s + \alpha_u A_u
\]

\[
A(T) = \alpha A_s + (1 - \alpha) A_u
\]

\[
\alpha(T) = [A(T) - A_u] / [A_s - A_u]
\]

- But, \( A_s \) and \( A_u \) are generally also \( f(T) \). This is manifested in sloping baselines for a folding transition.

![Diagram showing sloping baselines in absorbance spectrum]

- So, How do you get \( \alpha \) in this case? Assume that \( A_s \) and \( A_u \) are linear functions of temperature.

\[
A_s(T) = \varepsilon_s(T) C_T \ell \quad \text{(a similar equation applies for } A_u(T))
\]

\[
\varepsilon_s(T) = m_s T + b_s \quad \text{(a similar equation applies for } A_u(T))
\]
• So, we have

\[ \alpha(T) = \frac{[\varepsilon(T) - (m_u' T + b_u')]}/[(m_s' T + b_s') - (m_u' T + b_u')] \]

• This is a general and important equation that allows us to go from A vs T to \( \alpha \) vs T.

• See this sketch for going from A vs T to \( \alpha \) vs T.
• (x is the line segment from A_u to A and y is the line segment from A_u to A_s).

\[ \frac{\partial \ln K}{\partial (1/T)} = -\frac{\Delta H^o}{R}. \]

• In principle, you could now calculate \( K \) at each \( T \) from \( K = \alpha / (1 - \alpha) \). You could then use the van’t Hoff equation to get \( \Delta H^o \). If you do it this way, generally use \( \alpha \) between 0.15 and 0.85 to reduce error from small numbers.
• Generally, however, non-linear curve-fitting is used to fit six variables: \( m_s, m_u, b_s, b_u, \Delta H^o \) and \( \Delta S^o \), all in one fell swoop. There are programs available to do this (e.g. MeltWin).

• Thus, the equilibrium constant can be determined at any \( T \), provided that physical properties of the fully stacked and unstacked forms are available. Here lies the rub. Since \( \Delta H^o \) is so small for single-strand stacking of dinucleotides the fully unstacked form may not be found at 100 °C; otherwise the monomer properties can be used to estimate

---

3 In practice we usually solve for \( \Delta H^o \) and \( T_m \) as it is easier to converge in the fit. Can then get \( \Delta S^o \) from \( \Delta H^o \) and \( T_m \).
A$_s$. The fully stacked form may not be found at 0 °C. To get it, temperatures as low as –80 °C are needed (Fig 22-11). Such temperatures can be reached by using concentrated (25%) LiCl as a solvent. However, the thermodynamic properties derived for stacking under these conditions must somehow be extrapolated back to dilute salt conditions.

- In studying two-state conformational transitions, it is convenient to define a melting temperature, T$_M$. This is the temperature at which 50% of the strands are unstacked, or $\alpha = 0.5$.

- T$_M$ is therefore where K=1, so $\Delta G^\circ_{Tm} = 0$, so

$$T_M = \frac{\Delta H^\circ_{st}}{\Delta S^\circ_{st}}$$

- How do you find where $\Delta G^\circ_{Tm} = 0$? One way is to use the maximal slope in an A versus T plot; which is also maximum in the first derivative of the melting curve (dA/dT vs T); which is also where $d^2A/dT^2 = 0$. However, this works only for sharp transitions where $\Delta H^\circ$ is 5 to 10 x larger than RT. (see Box 22-1)

- However, T$_M$ is the temperature where the slope is maximal in a plot of A versus 1/T (in Kelvins$^{-1}$). (see Box 22-1)
B. For folding transitions of single strand oligonucleotides or polynucleotides:  
The effect of length.
Nucleic Acids: Properties, Structures, and Functions

April 13, 2016

Philip Bevilacqua

Lecture 37b:  Helix Coil Transitions in Nucleic Acids (II)

Reading:  BCT Chpt 8
          Cantor & Schimmel Chpt 20, 22, 23 (selected portions)

Lecture Review

- In the previous lecture notes we saw that single stranded nucleic acids have structure.
  The experimental evidence came from the temperature dependence of CD, UV, and NMR spectra.  poly A and C stack, poly U does not; poly G forms G-quartets.  Stacking is NOT strongly salt dependent.  The sugar indirectly influences the stacking.  Extensive (>90%) stacking is needed to stiffen the polymer.  The thermodynamics of the stacking transition can be measured by calorimetry (DSC), or by optical melting.  For dinucleotides, we measured $\Delta H^\circ$ by calorimetry without a model but had the problem of poorly defined baselines; we also measured $\Delta H^\circ$ by optical melting with a two-state model.  Now, let’s consider optical melting of nucleic acids longer than two nucleotides and the need for non-two-state models and application of statistical thermodynamics.

B. For folding transitions of single-strand oligonucleotides or polynucleotides: The effect of length.

i.) Qualitative Treatment:
- Stacking is measured on a per residue basis.  Thus, the middle base in a trimer has the potential for two stacks/residue.  In a tetramer, the middle two bases have the potential for two stacks/residue.  In a 100-mer, almost all (98) of the bases have the potential for two stacks/residue.  Thus, we expect the optical changes associated with stacking each residue in a polymer to be more intense as length increases.  This is true, and a 10-mer behaves very similar to a 100-mer on a per residue basis.  (See Fig 6-11 on next page).
- So, if think that internal bases with two stacks are more favorable than end bases with one stack, how can we express this molecular idea mathematically?  Consider the one-dimensional Ising Model, 1D Ising Model:

  1D Ising Model:
  Let $s$ be the equilibrium constant for stacking next to a stack.  Assume $s=f(T)$
  Let $\sigma$ be the constant for starting a stack, where $\sigma$ is the nucleation parameter.  Assume $\sigma$, a multiplier on $s$,¹ is not a $f(T)$ (i.e. it is entirely entropic in origin).  See BCT Table 8.1.

  Limits on $\sigma$:  
  $\sigma>1$  Then it is easier to start stacks than continue them.  =ANTICOOPERATIVE
  $\sigma<1$  Then it is harder to start stacks than continue them.  =COOPERATIVE
  $\sigma=1$  Then starting stacks is same as continuing stacks.  =NON-COOPERATIVE

¹ We prefer to use a multiplier for several reasons: it is easier to think about it in terms of a relative effect.  e.g. initiation is 0.4x less favorable than propagation; it also makes the math (summations) easier—see below.
• If $\sigma<1$ for a trimer, then upon melting, the apparent van’t Hoff enthalpy/residue will be larger than for a dimer. Since transition sharpness is proportional to $\Delta H^0$ (more in a later lecture), the trimer melt will have a steeper transition.

• In practice, $\sigma$ values for melting of poly A are 0.5 to 0.8, which is only SOMEWHAT COOPERATIVE. For comparison, $\sigma$ is small for $\alpha$-helix in proteins ($\approx 10^{-4}$), and is small for dsDNA too.

• See BCT Table 8.1 (p266) for more $\sigma$ values for ssRNA. (BCT uses $\beta$ for $\sigma$.)
ii.) Quantitative Treatment

- **Introduction:** There are three different models to describe Helix-Coil transitions in single-stranded polymers

1. All-or-None Model
2. Zipper Model
3. Zimm-Bragg Model

1.) The **All-or-None Model** assumes that each chain is either all helix or all coil. It does not allow a chain that has say 30% helicity.

E.g. For a tetramer, with helix (h) or coil (c), the only allowed states are:

\[
\text{cccc} \quad \text{hhhh}
\]

This is a two-state model, just like we used with dinucleotides last lecture. Note that the two more complex models that follow will reduce to the All or None Model in the limit that all propagations are highly favorable and the only state with ‘h’ residues that populates is the state with ALL the residues in the ‘h’ conformation.

If we return briefly to earlier in the notes, recall that

\[
A = \alpha_h A_h + \alpha_c A_c \quad \text{eq (1a)}
\]

Also, note that

\[
C = C_4 + H_4 \quad \text{where } C \text{ is the total concentration of strands.}
\]

Since

\[
K = H/C
\]

Therefore,

\[
\alpha_c = 4C_4/(4C_4 + 4H_4)
\]

Let’s choose \(C_4\) as the reference state—we often choose the unfolded state as the reference state. Now, let’s divide the top and bottom of the equation by the reference state, \(C_4\).

\[
\alpha_c = 1/(1+K) \quad 2
\]

And so

\[
\alpha_H = K/(1+K) \quad 3
\]

Therefore

\[
A = \frac{K}{1+K} A_h + \frac{1}{1+K} A_c \quad \text{eq (1)}
\]

Again, \(A_h\) and \(A_c\) are typically lines and so have slopes and intercepts, and \(K\) is a function of \(T\). In fact, you typically use the following relationship for \(K (T_M, \Delta H^o)\). [see homework]

\[
K = \exp \left[ \frac{\Delta H^o}{R} \left( \frac{1}{T_M} - \frac{1}{T} \right) \right] \quad \text{eq (2)}
\]

\(\Delta S^o\) and \(\Delta G^o\) come from the standard thermodynamic relationships.

\[
\Delta S^o = \Delta H^o / T_M \quad \text{and } \Delta G^o = \Delta H^o - T \Delta S^o
\]

---

2 \(\alpha_c\) is fraction of bases that are coiled. Here I am doing something a bit unusual in multiplying by 4 and dividing it back out. This is because in the two following, more complex, models, the number of bases is important and doesn’t cancel out.

3 Note that \(\alpha_c\) and \(\alpha_H\) have the same denominator and one has the first term of the denominator in the numerator and the other has the second term of the denominator in the numerator. The denominator here is a very simple example of a partition function, and this is the way partition functions are turned into fractional values. Note that an experimentally useful parameter, a fraction population, is defined in terms of thermodynamic parameters. This allows us to transition between the experimental world and the theoretical world. More later....

4 You can show this equation easily from the van’t Hoff equation at the \(T_M\).
These are the equations you use to fit optical melting data for a two-state A=B system. I’ve programmed Kaleidagraph to take equations (1), (2) and linear baselines and give back $\Delta H^\circ$, $T_M$, $\Delta G^\circ_{37}$, and $\Delta S^\circ$. Let me know if you ever want to use this.

We will now go to more complex models for single-strand stacking. Note that the only change in the above equations in going to more complex models will be to $\alpha$.

2.) The **Zipper Model** assumes that each chain has only one helix on it. However, that helix may involve all the bases or only some of the bases. An incomplete helix may exist anywhere in the chain.

   e.g. For a tetramer, the *allowed* states (all of them) are as follows:
   - cccc
   - hccc, chcc, cchc, ccch
   - hhcc, chhc, cchh
   - hhhc, chhh
   - hhhh

   *Disallowed* states include:
   - hchc
   - hhch...

   N.B.: This is not a two-state model. *For a tetramer, this is an 11-state model.*

3.) The **Zimm-Bragg Model** allows each chain to have any mix of helix and coil in it. *i.e.* there are no disallowed states.

   e.g. For a tetramer, any state you can think of is allowed. These states include:
   - cccc
   - hchc
   - chch etc...

   *For a tetramer, you can show that this is a 16 state model.*

Derivation of Model 2: The Zipper Model for helix-coil transitions:

- Let’s consider a partition function, $q$. $q$ is a *summation* of the Boltzmann weights for all of the states; e.g. for a 16 state model, the partition function will have 16 terms, all added together (although some of the terms may be the same and therefore grouped). The term for state $i$ can be thought of as the likelihood (=probability; = statistical weight) of state $i$.

\[
q = \sum_{i=0}^{n} P_i = P_0 + P_1 + P_2 + P_3 + ..., \text{ where } P_i \text{ is the probability of state } i
\]

Let state 0 be the reference state.

\[
\therefore q = 1 + \frac{[\text{state 1}]}{[\text{state 0}]} + \frac{[\text{state 2}]}{[\text{state 0}]} + ...
\]
• Compare what we did here to eq 1.

• Another way to write the partition function is in terms of energies of the microstates. Using simple thermodynamic relationships you can show that [see homework]

\[ [\text{state}_i] = e^{-\varepsilon_i/kT}, \text{where } \varepsilon_i \text{ is the energy of the ith level.} \]

\[ \therefore q = \sum_{i=0}^{n} g_ie^{-\varepsilon_i/kT} = \sum_{i=0}^{n} g_iq_i = g_0q_0 + g_1q_1 + g_2q_2 + \ldots \]

Where,

\[ q_i = e^{-\varepsilon_i/kT} \]

\[ g_i \] is the number of states with \( \varepsilon_i \). \( g_i \) is also known as the degeneracy of state \( i \).

\[ q_i = \] is known as a Boltzmann factor. It is the same as an equilibrium constant.

• Combining the two sections above, take the energy representation and use state 0 as the reference state.

It follows that \( q_i/q_o = e^{\left(-\varepsilon_i+\varepsilon_o\right)/kT} \)

Since, \( \varepsilon_i - \varepsilon_0 \) is a change in free energy, \( q_i/q_o \) is an equilibrium constant.

Let \( q_i/q_o = k_i \)

Example. Using Zimm Bragg to explicitly enumerate \( q \) for a tetramer.

• Let’s take a concrete example. Consider a tetramer with each monomer having a helix state (h) and coil state (c). **Write down the partition function for a tetramer according to the Zimm-Bragg model.**

• The possible conformations are

<table>
<thead>
<tr>
<th></th>
<th>( i )</th>
<th>( g_i )</th>
<th>( q_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>cccc</td>
<td>0</td>
<td>1</td>
<td>( q_0 )</td>
</tr>
<tr>
<td>hccc, chcc, cchc, ccch</td>
<td>1</td>
<td>4</td>
<td>( q_1 )</td>
</tr>
<tr>
<td>hhcc, hchc, hhch, chhc, chch, cchh</td>
<td>2</td>
<td>6</td>
<td>( q_2 )</td>
</tr>
<tr>
<td>hhhc, hhch, hhch, chhh</td>
<td>3</td>
<td>4</td>
<td>( q_3 )</td>
</tr>
<tr>
<td>hhhh</td>
<td>4</td>
<td>1</td>
<td>( q_4 )</td>
</tr>
</tbody>
</table>

\[ q = g_0q_0 + g_1q_1 + g_2q_2 + g_3q_3 + g_4q_4 \]
\[ q = q_0 + 4q_1 + 6q_2 + 4q_3 + q_4 \]
\[ q = q_0 + 4k_0q_0 + 6k_2q_0 + 4k_3q_0 + k_4q_0 \]
\[ q = q_0\left(1 + 4k_1 + 6k_2 + 4k_3 + k_4\right) \]

• By themselves, partition functions have no physical meaning. However, they are useful for calculating fraction of molecules with a given property.
• For example, the fraction of molecules with two residues with helical units, \( f_2 \), is

\[
f_2 = \frac{6q_2}{(q_o + 4q_t + 6q_2 + 4q_3 + q_4)}
\]

\[
f_2 = \frac{6k_2q_o}{[q_o(1 + 4k + 6k_2 + 4k_3 + k_4)]}
\]

\[
f_2 = \frac{6k_2}{(1 + 4k + 6k_2 + 4k_3 + k_4)}
\]

• Thus, \( q_o \) cancels out and is not important. This makes sense since thermodynamics is always relative to some reference state.

• To simplify matters, assume (for now) that the \( \Delta G^\circ \) for c to h is independent of its sequence and location.\(^5\)

\[
k_1 = \exp(-\Delta G / RT) = s
\]

\[
k_2 = \exp(-\Delta G / RT) = s^2
\]

*If this last term bothers you, remember that \( k_2 = [\text{state2}] / [\text{state0}] = ([\text{state2}] / [\text{state1}])([\text{state1}] / [\text{state0}])

\[
q = 1 + 4s + 6s^2 + 4s^3 + s^4
\]

\[
q = \sum_{i=0}^{4} \Omega_{4,i} s^i
\]

Where \( \Omega_{4,i} \) is the number of states with \( i \) helical bases in our tetramer.

Let’s return to the Zipper Model and derive an exact, general equation.

• Review our definitions of \( s \) and \( \sigma \) on L21p1.

• We will ignore end effects (i.e. equilibria on ends are same as in middle of polymer.

\[
s = \text{[...cchhhhcc...]} \\
\text{[...ccchhhcc...]} \\
\alpha s = \text{[...ccchcccc...]} \\
\text{[...cccccccc...]} \\
q = 1 + \sum_{k=1}^{n} \Omega_k \alpha s^k
\]

\( \Omega_k = n - k + 1 \), where \( \Omega_k \) is the number of ways of having \( k \) helical residues contiguous in a molecule of \( n \) residues

[see homework for \( \Omega_k \) derivation]

\(^5\) We will see shortly that these \( s \)'s must have a \( \sigma \) prefactor for initiation, and Zimm-Bragg allows multiple initiations and so has multiple \( \sigma \)'s.
We now have a partition function for the Zipper Model for a strand n-bases long, with k of them helical. This means we can calculate a probability for each possible state as the statistical weight for a state, which is its term in q, divided by all of itself.

\[
\therefore \text{If } q = 1 + \sum_{k=1}^{n} \Omega_k \sigma^k
\]

then, \( P(k) = \frac{\Omega_k \sigma^k}{1 + \sum_{k=1}^{n} \Omega_k \sigma^k} \)

and, if \( \Omega_k = n - k + 1 \)

\[
P(k) = \frac{(n - k + 1)\sigma^k}{1 + \sum_{k=1}^{n} (n - k + 1)\sigma^k}
\]

The denominator here is q. (see footnote below)

Now, we know the probability of each state. e.g. For a trimer, we know the individual probabilities of having each of the states: ccc, hcc, chc, cch, hhc, chh, and hhh. (Remember that hch is not allowed in the Zipper model.)

Now, let’s ask the question, “What fraction (=\( \alpha \)) of the bases are helical (=stacked)?” This would be the probability of each state, \( P(k) \), multiplied by the number of helical bases in that state, \( k \), summed over all values of \( k \), divided by the total number of bases, \( n \).

We now define \( \langle k \rangle \) as the average number of helical bases, and \( \alpha \) as \( \langle k \rangle / n \).

\[
\alpha = \frac{\langle k \rangle}{n} = \frac{\sum_{k=1}^{n} kP(k)}{n} = \frac{\sum_{k=1}^{n} k(n - k + 1)\sigma^k}{nq}
\]

It is useful to take the derivative of the partition function with respect to \( s \):

\[
\frac{\partial q}{\partial s} = \sum_{k=1}^{n} k(n - k + 1)\sigma^{k-1}
\]

Therefore, we can reduce \( \alpha \) to the following:

\[
\alpha = s = \frac{\partial \ln q}{\partial \ln s} = \frac{1}{n} \frac{\partial \ln q}{\partial \ln s}
\]

---

6 Why is the 1 from q missing in the numerator of \( P(k) \)? Because this will not be needed when multiplying \( P(k) \) by \( k \) in the equation below. (i.e. \( 0P(0)=0 \)).

7 Remember from above that \( \alpha \) is all we need in order to fit melting data and that stacking is measured on a per residue basis.
• This demonstrates an elegant feature of the partition function method: A physically measurable parameter, \( \alpha \), (see eq 1a) is related to the derivative of the partition function.

• The next step would be to actually take the derivative of \( q \) w.r.t. \( s \). This requires taking a derivative of a summation. When the dust settles, you end up with \( \alpha = f(\sigma, s, n) \).

• Just so you have it here is \( \alpha \) for the Zipper Model

\[
\alpha = \frac{\alpha s}{(s-1)^3} \left( \frac{ns^n - (n+2)s^{n+1} + (n+2)s - n}{n(1 + [\alpha s/(s-1)^2][s^{n+1} + n - (n+1)s])} \right)
\]

\( \sigma \) and \( s \) are the variables in the fit of a melt, since \( n \) can be determined by mass spec. or gel electrophoresis. \( s \) has the same T-dependence as for a two-state melt, and \( \sigma \) is generally assumed to be independent of temperature. Thus we have added one more variable to the fit (see L24p11—we are now fitting for \( m_s, m_a, b_s, b_u, \Delta H, T_m, \) and \( \sigma \), with only the last three parameters of real interest). Again, see p266, Table 8.1 for values of \( \sigma \) (=\( \beta \)) and see that single strand stacking is slightly cooperative (i.e. \( \beta \) is somewhat < 1); of course, this means that multiple initiations are a reality for long polymers, and we really should be allowing them, and therefore using the Zimm-Bragg Model.

• For Model 3, the Zimm-Bragg Model with more than one helical region allowed, matrix algebra also leads to a value for \( \alpha \). It is slightly less messy in the end.

\[
\alpha = 0.5 + 0.5(s - 1)(1 - s)^2 + 4\alpha s \]^{-1/2}
\]

• Note that the Zimm-Bragg results is independent of \( n \); allowing multiple initiations decoupled the length dependence of \( \alpha \).

• You will get a chance to show the \( \alpha \) has limits of 1 (\( s \) is large) and 0 (\( s \) is small) for both the Zipper model and the Zimm-Bragg model.

• In closing, we have found a way to model multi-state behavior of polymer unfolding (i.e. non-two-state helix-coil transitions), and to extract parameters that are important for the 1D-Ising model: the probability of propagating a helix and the probability of initiating the helix. For single strands it is not that hard to initiate (\( \sigma \) isn’t too small), so melting is not very cooperative. WHAT ABOUT DOUBLE-STRANDED HELICES???
Nucleic Acids: Properties, Structures, and Functions

Philip Bevilacqua

Lecture 37c: Helix Coil Transitions in Nucleic Acids (III)
April 13, 2016

Reading: BCT Chpt 8
Cantor & Schimmel Chpt 20, 22, 23 (selected portions)

Lecture Review
- Last time, we moved from single-stranded dinucleotides to single-stranded polynucleotides. We saw that proper descriptions of the helix-coil transition required us to move from a two-state model to a multi-state model, although each monomer was allowed to be in one of two states only (h,c). We used the 1D Ising Model (=statistical thermodynamics approach) in which propagating each bp occurs with the same equilibrium constant $s$, but initiation events have an equilibrium constant $\sigma$, which differs from propagation by $\sigma$. $\sigma$ values were 0.5-1, which are not as small as for protein $\alpha$-helix folding with $\sigma \approx 10^{-4}$. We considered the zipper model (only one helix/strand) and Zimm-Bragg model (OK to have >1 helix/strand) in some detail and used these to define $\alpha$ and then extract both $\sigma$ and $s$ from optical melts.

Duplex Formation: Introduction
(I have tried, to the extent possible, to parallel this section with that on single-strand stacking.)

- We know that DNA typically exists as a double-stranded molecule. However, DNA needs to replicate, and cruciforms may form and regulate gene expression. Also, many of the protons in dsDNA can exchange rapidly with tritiated water. This all means that the structure of DNA is dynamic and the base pairs are fluctuating all the time. Thus, it is worth thinking about the thermodynamics and kinetics of duplex formation of DNA.

- RNA is almost always a mix of ds and ss structure, so models of duplex formation—both local and long-range—are very biologically relevant and important.

- Both RNA and DNA are drug targets, and invading or capturing ss intermediates is a distinct therapeutic strategy.

Structure and stability of nucleic acid double strands (no helical imperfections yet)
- What are the relevant states for duplex formation?
  State 1: Random coil?
  State 2: Single-strand stacked?
  State 3: Partial duplex? (Use Zipper or Zimm-Bragg?)
  State 4: Full duplex?
- We will consider only states 1, 3, and 4; however, state 1 is likely a combination of random coil and ss stacked. Recently, there has been one laudable attempt to detangle states 1 and 2 in duplex formation by Tom Record.1

What is the experimental evidence for structure of double strands?

---

• As we saw with ss stacking, folding of ds structures leads to hypochromicity in the UV region of the absorbance spectrum.

• Hypochromicity (=the change in extinction coefficient of the molecule upon folding) is caused by changes in the base dipoles upon stacking, which leads to changes in the transition dipole moment, which is proportional to the probability of absorption. Tinoco developed the theory of hypochromicity of DNA in the early 1960s.

• %Hypochromicity is defined as \((\Delta A_{ss}/A_{ss})/A_{ss}\), where \(A_{ss}\) is the absorbance of the ss state and \(A_{ds}\) is the absorbance of the ds state. %h is typically 15-25%.

• %Hypochromicity values depend on the wavelength monitored, and further varies depending on the GC and AU content of the duplex under consideration. As the following figure shows, for AU-rich duplexes, %h is maximal near 260 nm and \(\approx 0\) near 280 nm; for GC-rich duplexes %h is maximal near 280 nm, but still substantial near 260 nm. This is an experimentally handy observation, as you can assign a transition in the unfolding of a large nucleic acid to an AU-rich feature if the hypochromicity at 260 is large and that at 280 is very small.

• Strand stoichiometry can be determined by mixing plots. These are a.k.a “Job plots”—see p11227 of paper from L14 for another case with two proteins. These curves assume infinitely tight binding of the two (or three strands). If binding is not tight, then the plots will show curvature near the intersection, but linear extrapolation from the endpoints to the intersection should still work OK.

![Graph of mixing plots](image)

Figure 22-13
Absorbance mixing curve for poly A and poly U. Relative absorbance is plotted as a function of the fractional composition of poly A + poly U mixtures at constant total concentration. (a) \(A_{259}\) under conditions such that only a 1:1 poly A-U complex forms. (b) \(A_{280}\) under conditions where both poly A 2U and poly A - U can form. [After results obtained by Gary Felsenfeld and colleagues.]

• At this point in the course, we are familiar with the global and local features of DNA and RNA structure. I will now discuss the statistical thermodynamics of duplex formation and then look at two-state approximations and optical melting curves. (Note that for ss stacking, we did two-state first followed by statistical thermodynamics.)

**Statistical thermodynamics consideration for duplex formation**

• Let’s examine the steps for duplex formation from 2 complementary strands, which are present in solution at equal concentrations. Many of the features of this process are shared with the helix-coil transition in a single strand. The equilibrium constant for initiating duplex will be called \(\kappa\), and it will be different from the propagation constant \(s\) for adding a base pair to a preexisting duplex. These equilibrium constants are defined below.

\[
\kappa = \begin{bmatrix}
\text{[structure]}
\end{bmatrix}^2 \\
\]

\[
S = \begin{bmatrix}
\text{[structure]}
\end{bmatrix}
\]

\(\kappa\) is analogous to \(\sigma_s\) for ss stacking. Different notation is used here since initiation of a duplex is a concentration dependent event and therefore fundamentally different than propagation. In other words, initiation of a single-strand stack has the same molecularity as propagation so the two can be related by a constant.
• \(\kappa\) is determined primarily by entropic effects for bringing two strands together. Clearly, initiating a duplex will be considerably harder than initiating a ss stack.

• \(s\) is determined by entropic (primarily conformational entropy--more on this later) and enthalpic effects (hydrogen bonding and stacking).

• As for ss helices, there are three models to consider: All-or-None, Zipper, and Staggered Zippers.

• Let’s start with the Zipper Model for duplex formation: (You can find a discussion of this in Appendix A.1 in Chpt 8 of BCT.) As you will see, this will lead us to the All-or-None model. The Staggered Zipper is a nasty model mathematically that becomes important only in long helices that have independent initiating events--we won’t consider it here.

• If we return to \(q\) for ss stacking, first we will replace \(\sigma\)s with \(\kappa\), as per footnote on previous page. Next we will consider what is called a ‘conformational partition function’, \(q_c\), which is equivalent to \(q-1\). This comes about by choosing single strands as the reference state. In other words, we are considering the relative weights only of states with one or more base pairs in \(q_c\).

\[
q_c = \kappa \sum_{k=1}^{n} (n-k+1)s^{k-1}
\]

• Once again, we will skip the rest of the math here and jump to the punch line; fyi: to get here, one has to prepare geometric series summations and their derivatives (see appendix A.1), with appropriate taking of limits to arrive at…the conclusion that:

• if \(s\gg1\), and \(n\) is large (in other words, if the equilibrium constants for base pairing are favorable, meaning their \(\Delta G^o\), \(\gamma\) are \(<0\), which, as you will soon see, is reasonable for base pairs), \(q_c\) simplifies to:

\[
q_c \approx \kappa s^{n-1}
\]

• Hooray! This is just the last term in the full expression for \(q_c\). (i.e.: when \(k = n\)). And the last term in \(q_c\) is for the fully base paired species. This means that if \(s\) is favorable and \(n\) is large (in practice > 4 bp) the only significantly populated species is the fully base paired one, and the All-or-None model applies, and we can use a two-state approximation.

• Also, please note that

\[
q_c \approx \kappa s^{n-1} = \kappa \prod_{i=1}^{n-1} s_i = \kappa s_1 s_2 \ldots s_{n-1} = K
\]

\[
\therefore \Delta G^o_{\text{tot}} = \Delta G^o + \sum_{i=1}^{n-1} \Delta G^o_i
\]
This last equation for $\Delta G_{37}^{\circ}$ says the total free energy is just the initiation free energy plus the sum of the free energies for propagating ("zipping up") the helix. It turns out that the free energies for propagating the helix are determined by who your neighbor is (i.e. the important interactions are strong and very local). This leads to the nearest-neighbor (NN) approximation for calculating the free energy of duplex formation. It says that you can calculate the total free energy by summing up the interaction free energies for each base pair as you propagate up the helix. There has been extensive experimental work, mainly by the Turner lab for RNA and the SantaLucia lab for DNA, to determine the free energy terms for simple secondary structure interactions (see Table 8.4, p276 of BCT).

**CAUTION: When might the NN approximation fail?**

- Poland & Scheraga (1970s).
- For weak base pairs at end of helices (= “fraying” = non-two-state behavior)
- For very long duplexes (>20 bp, certainly >100 bp), wherein multiple initiation (folding) events and multiple dissociation (unfolding) events occur, leading to bubbling, budding, and multi-step melting.
- For RNA with a large number of helical defects in it, which effectively decouples pairing regions, allowing them to fold and unfold independently.
- e.g. Fig 8-11 of BCT.

Next lecture we will look at optical melting of duplexes (and higher order complexes) and how to interpret their melting curves. Then we will look at the kinetics of duplex formation. We will consider values of $\kappa$ and $s$, and conformational and bonding models that explain their values. We will find that, unlike single-strand stacking, it is very hard to initiate duplex formation.
Lecture 37d: Helix Coil Transitions in Nucleic Acids (IV)
April 13, 2016

Reading: BCT Chpt 8
Cantor & Schimmel Chpt 20, 22, 23 (selected portions)

Lecture Review

- In the last lecture, we derived the partition function for stacking of oligo- and polynucleotides according to the Zipper Model. We then obtained an expression for fraction of bases stacked involving terms for propagation, $s$, and initiation, $\sigma$. Next, we looked at duplex formation, including hypochromicity and mixing curves. We considered a statistical thermodynamic description of duplex formation involving the Zipper Model. We found that, if local interactions are strong (i.e., $s > 1$), $\Delta G^{\circ,\text{tot}}_{37}$ could be estimated from the sum of nearest neighbor interactions.

Nearest-Neighbor expressions for $\Delta H^o$ and $\Delta S^o$

- We ended the last lecture with the following simplification of the partition function, in which

$$q_c = K s^{n-1} = K \prod_{i=1}^{n-1} s_i = K$$

$$\therefore \Delta G^{\circ}_{\text{tot}} = \Delta G^{\circ}_K + \sum_{i=1}^{n-1} \Delta G^{\circ}_i$$

- Note that one can write analogous expression for $\Delta S^o$ and $\Delta H^o$.

$$\therefore \Delta S^{\circ}_{\text{tot}} = \Delta S^{\circ}_K + \sum_{i=1}^{n-1} \Delta S^{\circ}_i$$

$$\therefore \Delta H^{\circ}_{\text{tot}} = \Delta H^{\circ}_K + \sum_{i=1}^{n-1} \Delta H^{\circ}_i$$, where $\Delta H^{\circ}_K$ is generally assumed to be zero.

Interpreting Melting Curves of Duplexes: Three ways of obtaining $\Delta H^o$ and $\Delta S^o$

- A reliable way to obtain thermodynamic parameters for nucleic acid folding for the NN (nearest-neighbor) approximation involves optical melting.

- Optical melting was introduced for dinucleotides. The same concept applies here. Absorbance is monitored as a function of $T$ at a wavelength that gives a good change in Absorbance (see earlier lecture for suggested wavelengths). N.B.: The hypochromicity (Absorbance change) is proportional to the fraction of strands folded. This last sentence is very important. In other words, the $T_M$ will be found when half the strands are paired, meaning half the bases are paired. This does not necessarily mean the $T_M$ will be found when half the molecules are paired, or when the paired complex is one half the total strand concentration. The definition is centered on the strand since this is what our observable (hypochromicity) is sensitive too. This definition holds for dinucleotide
stacking, hairpin formation, duplex formation, triplex formation, everthing! You will feel its effects in the following derivation.

- Let’s look at the equilibrium constant (=q for this case) for some specific duplex cases.
- Consider an equilibrium between non-self-complementary oligomers, A and B, and mix them together in equal concentrations. (This last condition is not necessary, but simplifies the derivation.)

\[
A + B \leftrightarrow AB
\]

\[
K = \frac{[AB]}{[A][B]}
\]

- Let \( C_T \) be the total strand concentration, and define \( \alpha \) as just described.

\[
C_T = [A] + [B] + 2[AB] = 2[A] + 2[AB]
\]

\[
\alpha_{AB} = \frac{2[AB]}{C_T} \equiv \alpha
\]

\[
\therefore [AB] = \frac{\alpha C_T}{2}
\]

\[
2[A] = C_T - \alpha C_T
\]

\[
[A] = C_T (1 - \alpha)/2
\]

\[
\therefore K = \frac{\frac{\alpha C_T}{2}}{\left(\frac{C_T (1 - \alpha)}{2}\right)^2} = \frac{2\alpha}{C_T (1 - \alpha)^2}
\]

In addition, as mentioned above, we define \( T_M \) where \( \alpha = 1/2 \).

\[
\therefore K_{T_M} = \frac{2(1/2)}{C_T (1 - 1/2)^2} = \frac{4}{C_T}
\]

\[
\Delta G_T^o = -RT_M \ln(4 / C_T) = \Delta H^o - T_M \Delta S^o
\]

\[
\Delta H^o = T_M [\Delta S^o - R\ln(4 / C_T)]
\]

\[
T_M = \frac{\Delta H^o}{\Delta S^o + R \ln(C_T / 4)}
\]

- This \( T_M \) is concentration-dependent, meaning the \( T_M \) increases as the \( C_T \) increases.\(^1\)

\(^1\)This may seem surprising, but is due to \( \Delta S^o \) and \( \Delta H^o \) being defined for duplex formation, rather than duplex unfolding. This is opposite how protein folders often defined the transition, namely for protein unfolding. In our case here, \( \Delta S^o \) is negative (unfavorable) and \( \Delta H^o \) is negative (favorable), leading to the effect of increasing \( C_T \) increasing the \( T_M \).
• You could use this form of the \( T_M \) equation to predict \( T_M \) from calculation of \( \Delta H^o \) and \( \Delta S^o \) from NN parameters.\(^2\)

• To see the concentration-dependence of the \( T_M \), one must use a wide-range of concentrations (typically 100- to 1000-fold). The larger the \( \Delta H^o \) and \( \Delta S^o \) (i.e.: the more base pairs in the duplex), the larger the \( C_T \) needed to see the \( T_M \) change. The range of concentrations needed can be estimated from the equations above and values of \( \Delta G^o_{37} \), \( \Delta S^o \), and \( \Delta H^o \) from tables.

• A very reliable way to extract \( \Delta H^o \) and \( \Delta S^o \) is by rearranging to give:

\[
\frac{1}{T_M} = \frac{\Delta S^o}{\Delta H^o} - \frac{R \ln(4/C_T)}{\Delta H^o}
\]

or,

\[
\frac{1}{T_M} = \frac{R}{\Delta H^o} \ln(C_T/4) + \frac{\Delta S^o}{\Delta H^o} \quad (*a \text{ common form})
\]

\[
\frac{1}{T_M} = \frac{\Delta S^o}{\Delta H^o} - \frac{R}{\Delta H^o} \ln 4 + \frac{R}{\Delta H^o} \ln C_T
\]

\[
\frac{1}{T_M} = \frac{R}{\Delta H^o} \ln C_T + \frac{\Delta S^o - R \ln 4}{\Delta H^o} \quad (*a \text{ common form})
\]

\[
\frac{1}{T_M} = 2.303R \log C_T + \frac{\Delta S^o - R \ln 4}{\Delta H^o} \quad (*a \text{ common form})
\]

• I mentioned that there were three ways to get \( \Delta H^o \) for duplexes. **Here is method one:** Measure \( 1/T_M \) as a function of \( \log C_T \),\(^3\) where \( T_M \) can be reliably obtained from a plot of \( dA/dT \) or \( dA/dT^{-1} \) versus \( T \) (see L22). At high \( C_T \), use short pathlength cells (0.1 to 0.01 cm) to get \( A \) in a linear range for Beer’s law.\(^4\)

• The form of the equation for self-complementary oligomers (i.e. \( 2A\text{=}A_2 \)) can be shown (by similar methods in HW 5) to be

\[
\frac{1}{T_M} = \frac{R}{\Delta H^o} \ln C_T + \frac{\Delta S^o}{\Delta H^o}
\]

• **Method two** is to use non-linear curve fitting of each melt at each \( C_T \) to obtain \( \Delta H^o \) and \( \Delta S^o \), as for dinucleotides. You can then average \( \Delta H^o \) and \( \Delta S^o \) from curve fit parameters at the different concentrations. (Even though \( T_M \) changes with concentration, \( \Delta H^o \) and \( \Delta S^o \) do not). The \( \Delta H^o \) and \( \Delta S^o \) obtained by Methods 1 and 2 should be in good agreement; if not, the melt is probably not two-state, and re-examination of the system is in order. If

---


\(^3\) I like plotting \( T_M^{-1} \) versus \( \log C_T \) since I can easily visualize the range of concentrations used in the experiments. However, it is equally valid to plot versus \( \ln C_T \) or as \( \ln (C_T/4) \) as long as you use the right version of the melt equation to fit your data.

\(^4\) You can use Table 6.1 of BCT to estimate \( \varepsilon \) from high temperature absorbance measurements.
they are in agreement, method one is generally preferred since $T_M$ can be obtained reliably even in the absence of good baselines by using the $dA/dT$ or $dA/dT^{-1}$ methods.

- **Method three** is to look at the slope of the melting curve itself. Recall, that we discussed how to make a species plots ($\alpha$ vs $T$) from melting data. From a species plot, you can determine a slope at any temperature (i.e.: you can determine $d\alpha/dT$ at any $T$). Let’s see if we can find a relationship between $\Delta H^o$ and $d\alpha/dT$. We will just consider a unimolecular melt since the math is a little easier. However, equivalent expressions apply for duplex and high-order oligomer formation (see Puglisi and Tinoco handout).

Recall the van't Hoff relationship:

$$\frac{\partial \ln K}{\partial (1/T)} = -\frac{\Delta H^o}{R},$$

$$\frac{\partial \ln K}{\partial T} = \frac{\Delta H^o}{RT^2}$$

For a unimolecular transition,

$K = \alpha / (1 - \alpha)$  (see L20p10)

$$\therefore \Delta H^o = RT^2 \frac{\partial \ln [\alpha/(1-\alpha)]}{\partial T}$$

$$= RT^2 (1-\alpha)/\alpha \frac{\partial [\alpha/(1-\alpha)]}{\partial T}$$

$$= RT^2 (1-\alpha)/\alpha \left\{ \frac{1}{1-\alpha} \frac{\partial \alpha}{\partial T} + \alpha \frac{\partial (1-\alpha)^{-1}}{\partial T} \right\}$$

$$= RT^2 (1-\alpha)/\alpha \left\{ \frac{1}{1-\alpha} \frac{\partial \alpha}{\partial T} - \alpha \frac{1}{(1-\alpha)^2} \frac{\partial \alpha}{\partial T} \right\}$$

$$= RT^2 / \alpha \left\{ \frac{\partial \alpha}{\partial T} + \alpha \frac{1}{(1-\alpha)} \frac{\partial \alpha}{\partial T} \right\}$$

at $T_M$, $\alpha = 1/2$

$$\therefore \Delta H^o = RT_M^2 / 2 \left\{ \frac{\partial \alpha}{\partial T} \right\}_{T=T_M} \left\{ 1 + \frac{1/2}{(1-1/2)} \right\}$$

$$\Delta H^o = 4RT_M^2 \left\{ \frac{\partial \alpha}{\partial T} \right\}_{T=T_M}$$

- For unimolecular transitions, only Methods two and three are available!

- The most important point here is that the sharper the transition, the larger the $\Delta H^o$ for folding.
• You can find similar $\Delta H^0$ formulas for melts of duplexes, triplexes, etc in Table II of Puglisi and Tinoco handout. *In all cases, $\Delta H^0$ is proportional to $d\delta/dT$ at the $T_M$."

**Calculating thermodynamic values for duplex formation: NN theory.**

• Recall, after Zipper-Model treatment of duplex formation, we ended up with the equations:

$$q_c = \kappa S^{n-1} = \kappa \prod_{i=1}^{n-1} S_i = \kappa S_1 S_2 \ldots S_{n-1} = K$$

$$\therefore \Delta G^0_{tot} = \Delta G^0 + \sum_{i=1}^{q_c} \Delta G^0_i$$

• To a first approximation, free energy to propagate a helix depends only on the adjacent base pair in the helix. There are 10 unique nearest neighbors (NN) for duplex formation (HW5).

$$\Delta G^0_{tot} = \Delta G^0_{nuc} + \Delta G^0_{sym} + 2\Delta G^0_{GA} + 2\Delta G^0_{GU} + \Delta G^0_{CG} + \Delta G^0_{GC} + \Delta G^0_{GG}$$

• Let’s take some examples. (see p280 of BCT for more examples, and Serra & Turner review).

**EXAMPLE 1:** Consider two self-complementary oligonucleotides that are the reverse of each other. You can find the values for these terms in BCT Table 8.4 pp276-277.

5’GUCGAC3’/3’CAGCUG5’ = $\Delta G^0_{nuc} + \Delta G^0_{sym} + 2\Delta G^0_{GA} + 2\Delta G^0_{GU} + \Delta G^0_{CG} = -7.00$ kcal/mol

5’CAGCUG3’ = $\Delta G^0_{nuc} + \Delta G^0_{sym} + 2\Delta G^0_{CA} + 2\Delta G^0_{CU} + \Delta G^0_{GC} = -7.28$ kcal/mol

• Note that these two sequences (the reverse of each other) do not have the same NNs, and therefore do not have the same $\Delta G^0_{37}$.

**EXAMPLE 2:** Consider two non-self-complementary strands. (Has no $\Delta G^0_{sym}$ term.)

CA$_6$G + CU$_6$G = $\Delta G^0_{nuc} + 5\Delta G^0_{AA} + \Delta G^0_{CA} + \Delta G^0_{CU} = -4.75$ kcal/mol

• Note that $\Delta G^0_{sym} = -RT \ln (1/2) = +0.427$ kcal/mol at 37 °C. This arises due to a gain in rotational symmetry of the self-complementary duplex not present in the free strands. This gain in symmetry is equivalent to loss of a degree of freedom, and entropy loss is unfavorable.

**EXAMPLE 3:** Consider two oligomers with identical NN terms.

• 5’GCCGCCG3’3’CGGCCG5’ and 5’GGGCCG3’3’CCGCCG5’. $T_M$’s at $10^{-4}$ M $C_T$ are 67.2 and 65.2 °C. Suggests NN approximation is pretty good.
Lastly, note that there is a 0.45 kcal/mol penalty for each terminal AU base pair. The molecular basis for this is that AU base pairs tend to fray at the ends. See p280 of BCT and Fig 8-6.
Lecture 37e: Helix Coils Transitions in Nucleic Acids (V)
April 13, 2016

Reading: BCT Chpt 8
Cantor & Schimmel Chpt 20, 22, 23 (selected portions)

Lecture Review

- In the last lecture, we took the simplified partition function for duplex formation and saw how that led to the nearest-neighbor model for duplex formation. We were able to estimate $\Delta G^\circ$, $\Delta H^\circ$, and $\Delta S^\circ$ as sums of terms for individual interactions, plus a symmetry penalty, plus an initiation penalty. Note, that as mentioned, the initiation penalty makes $\kappa <<1$ and duplex formation highly cooperative (cf. to single strand stacking); this makes the zipper model pretty good for duplex formation. We then looked at three methods for extracting $\Delta H^\circ$ from melt curves: the concentration dependence of the $T_M$, non-linear curve fitting of individual melts, and analysis of the steepness of the melt in a species-plot. We then looked at some examples of setting up calculations for $\Delta G^\circ_{37}$, including self-complementary, non-self-complementary, and identical NN cases.

Further consideration of $\Delta G^\circ_{37}$, $\Delta S^\circ$, $\Delta H^\circ$, and $T_M$ predictions

- One of the questions I'm asked most frequently is, “What do you think the $T_M$ is for duplex X or hairpin Y or motif Z?” We talked about how to predict $T_M$ at the end of last lecture, but let’s look a little more deeply into this and consider different motifs.

- Let’s start with a helix defect (very common in RNA but also important in DNA as insertion or deletion mutations). The simplest defect would be a single nucleotide bulge.

```
5' A 3'
GCG GCG
3' CGC CGC 5'
```

- When we estimate $\Delta G^\circ_{tot}$ for this case, we have to decide whether or not to put in a favorable term for the stacking of the two GC base pairs that sandwich the A-bulge. For our purposes here, we will assume that the bulge is flipped out and that these base pairs can stack. However, as you might imagine, whether a bulged nucleotide stacks into a helix or not will be case-dependent; moreover, you might have a bulge that is flipped out, but for which the neighboring bases still cannot stack well, especially if the bulge is large.\(^1\) If you know the details of the case at hand, perhaps from structural data, then you should adjust your equation appropriately.

---

\[ \Delta G^\circ_{\text{tot}} = \Delta G^\circ_{\text{nuc}} + 2\Delta G^\circ_{\text{GC}} + 2\Delta G^\circ_{\text{CG}} + \Delta G^\circ_{\text{bulge}} + \Delta G^\circ_{\text{GG}} \]

- There are of course many other types of small motifs involving helix defects that are possible for the secondary structure RNA. For example, in addition to bulge loops, there are hairpin loops, internal loops, and multibranch loops. An excellent review (see handout) is available showing how to estimate thermodynamic parameters and \( T_M \) for these systems.\(^3\) We will discuss some of these examples in class during this lecture, and you should read over some of the examples on your own. Note that the NN model can also have terms for more complex structural features, such as a stable hairpin tetraloop; in this case, the loop parameter is treated as a ‘cassette’ free energy, i.e. as an independent additive free energy term.

- Since most of these defects involve loops, let’s look at the theoretical basis for the nucleation of a large loop.

**Calculating the nucleation entropy of a large loop**

- Let’s start with our definition of a big loop. A big loop is one in which there is not extensive hydrogen bonding and stacking interactions in the loop. OK, based on this definition, once a loop gets big enough (say 7 or 8 nts), we don’t expect any more change in \( \Delta H^o \) for closing or initiating the loop. That is to say, the nucleotides in the middle of a large loop aren’t forming bonding interactions, as per our definition above. Therefore, the length dependence of the free energy of formation of a large loop should show up entirely as entropy.

- Let’s calculate the entropy change in closing an internal or hairpin loop (see Fig below).

  ![Diagram of nucleation and growth of loops](image)

  - Closure is modeled here as the process of bringing the pairing bases together in some volume \( V_i \). (see the Figure)

\(^2\)“\( \Delta G^\circ_{\text{GC}} \)” is the term that assumes stacking of the adjacent GC base pairs.


\(^4\) Strictly speaking, \( T_M \) is not a thermodynamic parameter. For one thing, thermodynamic parameters are not concentration dependent. Moreover, \( T_M \) is not used in the predicting or modeling of nucleic acid structure (\( \Delta H^o \) and \( \Delta S^o \) are used). However, one can predict \( T_M \) from \( \Delta H^o \) and \( \Delta S^o \); one can easily measure \( T_M \) in the lab and use it to extract \( \Delta H^o \) and \( \Delta S^o \) (e.g.: L27); and \( T_M \) is biologically relevant, since it tells you at what temperature half the strands are folded.
• The internal entropy per mole of chain, $S$, is related to the number of conformations available to it, $\Omega$:
\[ S = R \ln \Omega \]
This is also known as the statistical entropy. For 1 molecule, $S = k_B \ln \Omega$, which is Boltzmann’s equation, and $k_B$ is Boltzmann’s constant.

• Thus, the change in internal entropy is related to the difference in the number of states available when the ends are in closure, $\Omega_{cl}$ versus all other states.
\[ \Delta S_{\text{loop}} = R \ln (\Omega_{cl} / \Omega) \]

• The trick now is to define $\Omega$ and $\Omega_{cl}$ according to some reasonable model.

• We turn to polymer physics (developed by Flory) and look at the normalized distribution function, $W_N(r)dr$, of end-to-end distances for a coil. $N$ is the number of monomer chains in the loop, and $l$ is the length of a nucleotide.

\[ W_N(r)dr = 4\pi r^2 \left( \frac{3}{2 \pi Nl^2} \right)^{3/2} \exp\left( -\frac{3r^2}{2Nl^2} \right) dr \]

- This is a Gaussian distribution multiplied by the surface of a sphere, $4\pi r^2$, which leads to a Gaussian that has a longer tail at high $r$. We are interested in the likelihood of the two ends being close to each other, say within the range of a separation of $r = 0$ (ends occupying same region in space) to $a$, where $a$ is the radius of the sphere with volume $V_i$.
\[ \therefore \frac{\Omega_{cl}}{\Omega} = \int_0^a W_N(r)dr \]

5 This is the same idea introduced earlier. You can see it nicely on L21p6, where $\Omega$ is the number of ways of arranging a helix on a strand, and $\Omega$ multiplies the propagation constant $s$, leading to $-RT \ln \Omega$, as the contribution to $\Delta G^0$, all of which is entropic, making $\Delta S^0 = +RT \ln \Omega$.

6 This equation really has an integral in the denominator too, which is from 0 to infinity. But $W_N(r)$ is normalized so this integral is 1. \[ \Omega = \int_0^\infty W_N(r)dr = 1. \]
• If N is large and r is small (i.e. we are only interested in close approach), then \( r^2 / N \) is very small, and the exponential term is the exponential of a very small number, which we will approximate as unity.

\[
\frac{\Omega_{cl}}{\Omega} = \left( \frac{3}{2\pi N l^2} \right)^{3/2} \int_0^a 4\pi r^2 \, dr
\]

But, \( \int_0^a 4\pi r^2 \, dr = (4/3)\pi a^3 = V_i \)

\[
\therefore \frac{\Omega_{cl}}{\Omega} = \left( \frac{3}{2\pi N l^2} \right)^{3/2} V_i
\]

• This expression allows us to write the entropy change as the Jacobson-Stockmayer equation.

\[
\Delta S_{\text{loop}} = R \ln \left[ \left( \frac{3}{2\pi N l^2} \right)^{3/2} V_i \right]
\]

\[
= -(3/2)R \ln N + R \ln \left[ \left( \frac{3}{2\pi l^2} \right)^{3/2} V_i \right]
\]

\[
= -(3/2)R \ln N + R \ln C
\]

• The last term is just a constant since \( l \) and \( V_i \) are fixed for all sizes of loops.

• How can we determine the value of C? Consider a loop of size n. (Note this is lowercase n).

For a loop of size n,

\[
\Delta S_{\text{loop}}(n) = -\frac{3}{2} R \ln n + R \ln C
\]

or, \( R \ln C = \Delta S_{\text{loop}}(n) + \frac{3}{2} R \ln n \)

• Thus, for a loop of size N, we can replace \( R \ln C \) with terms for loop of size n, which we can determine experimentally. In other words, we can first measure \( \Delta S^o \) for size n to get the term with C, and then use the following equation to estimate \( \Delta S^o \) for size N.

For a loop of size N,

\[
\Delta S_{\text{loop}}(N) = -\frac{3}{2} R \ln N + \Delta S_{\text{loop}}(n) + \frac{3}{2} R \ln n
\]

\[
\Delta S_{\text{loop}}(N) = -\frac{3}{2} R \ln (N/n) + \Delta S_{\text{loop}}(n)
\]
• Again, we assume that $\Delta G_{\text{nuc}}$ is purely entropic, so we have $\Delta G_{\text{nuc}} = -T\Delta S_{\text{loop}}$. Here I use $\Delta G_{i,\text{loop}}^\circ$ in place of $\Delta G_{\text{nuc}}^\circ$, since we are initiating loop formation.

$$\Delta G_{i,\text{loop}} (N) = \frac{3}{2} RT \ln(N/n) + \Delta G_{i,\text{loop}} (n)$$

• This is an important equation, as it gives us a way to predict the free energy penalty to make a loop.

• Since we have gone to the trouble of estimating the free energy for forming large loops, let’s look at some of the interesting phenomena that can be described by it. Large loops are possible in RNA and DNA as they melt out. Let’s ask the following question, “During melting, do multiple bubbles form along the helix, OR do those bubble fuse to make one big bubble?” In other words, what can these equations tell us about whether melting is likely to be two-state or multi-state for a given molecule?

• Consider the following equilibrium in which a conformation with two 8 nt loops separated by two AU pairs is in equilibrium with a state with one 20 nt loop. Fusion is defined as the process of fusing together the two smaller loops to make the one bigger loop. $\Delta G_{37}^\circ$ is defined such that if it is negative, fusion is favored.
• We need a few rules for large loop formation. We will use $\Delta G_{i,\text{loop}}(7) = +3$ kcal/mol. (From experiments by Ninio.)

• For this example, we will work at 25°C, where $1.5RT = 0.9$ kcal/mol

one big loop: $\Delta G^\circ_{\text{tot}} = \Delta G^\circ_{\text{loop}}(20) = 0.9\ln(20/7) + 3 = +3.9$ kcal/mol

two small loops: $\Delta G^\circ_{\text{tot}} = \Delta G^\circ_{\text{AA}} + 2\Delta G^\circ_{\text{loop}}(8) = -1.2 + 2[0.9 \ln(8/7) + 3] = +5.0$ kcal/mol

• Interestingly, the big loop is more stable than the two small loops, and fusion of the small loops to form the big loop is favored by $\Delta G^\circ_{\text{fus}} = -1.1$ kcal mol$^{-1}$ ($= +3.9 - (+5.0)$).

• Let’s generalize this, and generate a general equation for fusion.

\[
\Delta G^\circ_{\text{fus}} = \Delta G^\circ_{\text{loop}}(2N+2m) - [2\Delta G^\circ_{\text{loop}}(N) + (m-1)\Delta G^\circ_{\text{gr}}] \\
= 0.9 \ln [2(N+m)/7] + 3 - 1.8 \ln (N/7) - 6 - (m-1)\Delta G^\circ_{\text{gr}}
\]

An average value for $\Delta G^\circ_{\text{gr}}$ is $-2.5$ kcal/mol, but this is very sequence dependent (see below).

\[
= 0.9 \ln 2 + 0.9 \ln [(N+m)/7] + 3 - 1.8 \ln (N/7) - 6 + 2.5m - 2.5
\]

• We are dealing with large loops compared to joining regions, $m$ (i.e. $N \gg m$)

\[
=-4.9 - 0.9 \ln (N/7) + 2.5 m
\]

Or, $\Delta G^\circ_{\text{fus}} = -3.1 - 0.9 \ln N + 2.5 m$ (kcal mol$^{-1}$)

Can also show that $\Delta G^\circ_{\text{fus}} = \Delta G^\circ_{\text{gr}} - 0.6 - 0.9 \ln N - \Delta G^\circ_{\text{gr}} m$ (kcal mol$^{-1}$)

• This last equation is a very interesting one. It predicts how many base pairs in the joining region will be necessary to keep the loops from fusing. For example, if you ask for conditions where $\Delta G^\circ_{\text{fus}} = 0$ (i.e. evenly balance the big loop/small loop equilibrium) and let $\Delta G^\circ_{\text{gr}}$ vary (as it does based on NN theory) and then solve for $m$ (the no. of bp in the joining region to stop the small loops from fusing) you get the following relationships.

When $\Delta G^\circ_{\text{fus}} = 0$,

\[
m = (\Delta G^\circ_{\text{gr}} - 0.6 - 0.9 \ln N) / \Delta G^\circ_{\text{gr}}
\]

This leads to the following simulation

---

7 $\Delta G^\circ_{\text{gr}}$ is for growth of the joining segment between the loops. It is $m-1$ rather than $m$, because the first base pair is used for starting the joining segment.
$\Delta G_{gr} = -2 \quad \Delta G_{gr} = -1 \quad \Delta G_{gr} = -0.5 \quad \Delta G_{gr} = -0.2$

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>m</td>
<td>N</td>
<td>m</td>
<td>N</td>
<td>m</td>
<td>N</td>
<td>m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.9</td>
<td>4</td>
<td>2.8</td>
<td>4</td>
<td>4.7</td>
<td>4</td>
<td>10.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.1</td>
<td>6</td>
<td>3.2</td>
<td>6</td>
<td>5.4</td>
<td>6</td>
<td>12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.2</td>
<td>8</td>
<td>3.5</td>
<td>8</td>
<td>5.9</td>
<td>8</td>
<td>13.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.3</td>
<td>10</td>
<td>3.7</td>
<td>10</td>
<td>6.3</td>
<td>10</td>
<td>14.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2.4</td>
<td>12</td>
<td>3.8</td>
<td>12</td>
<td>6.7</td>
<td>12</td>
<td>15.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2.5</td>
<td>14</td>
<td>4.0</td>
<td>14</td>
<td>7.0</td>
<td>14</td>
<td>15.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2.5</td>
<td>16</td>
<td>4.1</td>
<td>16</td>
<td>7.2</td>
<td>16</td>
<td>16.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>2.6</td>
<td>18</td>
<td>4.2</td>
<td>18</td>
<td>7.4</td>
<td>18</td>
<td>17.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.6</td>
<td>20</td>
<td>4.3</td>
<td>20</td>
<td>7.6</td>
<td>20</td>
<td>17.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>2.7</td>
<td>22</td>
<td>4.4</td>
<td>22</td>
<td>7.8</td>
<td>22</td>
<td>17.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2.7</td>
<td>24</td>
<td>4.5</td>
<td>24</td>
<td>7.9</td>
<td>24</td>
<td>18.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- As expected, as $\Delta G_{gr}^\circ$ becomes less favorable, the number of bp (m) required to halt fusion increases. Now, you might ask “When does $\Delta G_{gr}^\circ$ ever become as small as $-0.2$ kcal/mol?” The answer is, “Near the $T_M$.” Since $\Delta G_{gr}^\circ$ has a strong and favorable enthalpic component, this leads to a temperature-dependent equilibrium constant, which leads to a very temperature sensitive $\Delta G_{gr}^\circ$. $\Delta G_{fus}^\circ$ on the other hand is entropic, so $\Delta G_{37}^\circ$ has a very small temperature dependence (i.e.: on a percentage basis $-T\Delta S^\circ$ doesn’t change much with over the melting temperature range in Kelvins). Therefore, small $\Delta G_{gr}^\circ$ are a reality as the $T_M$ is approached! At small $\Delta G_{gr}^\circ$, you need larger and larger joining regions to prevent fusion. Therefore fusion happens, preventing bubbling near the $T_M$, leading to cooperative melts. You can’t fight entropy!
Nucleic Acids: Properties, Structures, and Functions

Lecture 37f: Helix Coil Transitions in Nucleic Acids (VI)
April 13, 2016

Reading: BCT Chpt 8
Cantor & Schimmel Chpt 20, 22, 23 (selected portions)

Lecture Review

- Earlier, we looked into how to carry out predictions of helix stability by applying the NN theory. We also looked at the effect of helix defects on helix stability. We saw that we could calculate $\Delta H^\circ$, $\Delta S^\circ$, and $\Delta G^\circ$ for a variety of secondary structural motifs. We considered the process of nucleating formation of a large loop in some detail. We derived an equation that related the equilibrium for fusing together smaller loops into a big loop to the size of the loops, and the number and free energy of base pairs joining the loops. We saw that as $T$ approached the $T^\text{M}_37$ that loops will tend to fuse together, minimizing the number of terms in the partition function and moving towards an all-or-none case.

Predicting RNA secondary structure

- There is a vast amount of sequence information for RNA and DNA. The human genome has been sequenced, with ca. 3 billion base pairs of DNA, and many (30+) other genomes are sequenced as of the year 2002. Despite enormous progress in obtaining high-resolution RNA structures, especially solving the structure of the ribosome in the last two years, the amount of structural data cannot keep pace with the sequence data. Thus, accurate prediction of RNA structure from sequence is an important project. We will focus our attention on efforts to predict RNA secondary structure; prediction of tertiary structure is more problematic and generally requires additional experimental and comparative results.

$$F_2 \leftrightarrow U \leftrightarrow F_1$$

$$K_i = e^{-\Delta G_i / kT}$$

$$\Delta G_{tot} = \sum \Delta G_{nuc} + \sum \Delta G_{i,loop} + \sum \Delta G_{gr}$$

- Ideally, we would calculate $\Delta G^\circ_{tot}$ for all reasonable conformations. In practice, must use approximations.

- mfold, from Zuker and Turner, predicts the lowest free energy structure.

- On average, Turner lab gets $\approx 85\%$ of the helices correct\(^1\) when you look at structures within a 10% window of the lowest free energy structure.

- This accuracy increases sharply (and is in a smaller % window) when experimental constraints are used in the prediction (see below).

---

\(^1\) This means that helix is predicted correctly within 2 bp's.
• The Zuker algorithm, however, but does not allow pseudoknots.

![A pseudoknot](image)

• In nature, occurrence of pseudoknots is $\approx$1 out of 10 helices

• Recently, Rivas & Eddy has written a program that allows pseudoknots to be considered. It runs very slowly, but is an important contribution.\(^2\)

**An mfold example:**

• Go <http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi> and follow instructions to get to appropriate part of the Zuker home page. You can choose RNA or DNA to fold. I am not giving you all the detailed explanations of what each parameter means. However, you can find this conveniently located on the html form on the web page. Simply click on the name, such as “Window Size”, and you will find its definition.

• Input the name and sequence into the appropriate boxes. (It is convenient to cut and paste the latter out of a pre-existing text file to save time and avoid errors.) Numbers in sequence are OK, as program ignores them.

    Name: example1
    Sequence: aaagggacuccaguacgcugaguaccaaa
    Window Size=0
    Percent Suboptimality=10

• You should get back three structures with $\Delta G^\circ \approx 37$ of $–7.8$, $–7.8$, and $–7.3$. The first two differ only by a migrating C-bulge. (see in-class demo, and try yourself).

• It is possible to calculate the fractional population of each of these 3 states using a simple partition function analysis. See in-class.

• Now, suppose you had structure mapping data from enzymes or chemicals that position 11 was single-stranded and position 4 was paired (but you didn’t know to whom).

• You could use these constraints in the prediction by entering the following information (‘F’ = force to pair; ‘P’ = prohibit pairing). *A few experimental constraints can greatly improve the overall accuracy of the prediction.*

---

• You now get back just one structure, which has the C-bulge at position 11.

• I have also included a link on the course page to SantaLucia’s site for predicting nucleic acid (DNA/DNA, RNA/RNA, DNA/RNA or RNA/DNA) hybridization thermodynamic parameters, including $T_M$’s.

• Please remember that all of these programs use NN theory, which means they use the simplified partition function for duplex formation involving the last term in the expression.

---------------------------

Kinetics of Nucleic Acid Folding Transitions

A.) Kinetics of single-strand stacking

$$ U \xrightarrow{k_1} S \xleftarrow{k_{-1}} $$

• This process is fast (nano to microseconds). No concentration dependence. Therefore cannot use mixing methods.

• $T$-jump is necessary

• $T$-jump is a perturbation method. The system starts at equilibrium and then the temperature is rapidly jumped by 2-10 °C (in a nanosecond or so using an IR laser). This is the perturbation. The system finds itself in a non-equilibrium situation (i.e. if $\Delta H^0$ is nonzero, the temperature change has changed $K$—see below) and must relax to this new $K$. Obviously, the temperature change must be fast relative to the rate of relaxation.

Recall

$$ \ln K = \frac{\Delta H^0}{RT} + \frac{\Delta S^0}{R} $$

3

If $\Delta H^0 \neq 0$, then $K = f(T)$.

---

3 Please do not confuse the temperature dependence of $K$ with that of $\Delta G^\circ$. For $K$ to change with temperature, $\Delta H^0$ must be nonzero; for $\Delta G^\circ$ to change with temperature $\Delta S^0$ must be nonzero (assuming a negligible heat capacity).
• Where T=temperature, and t=time. ΔU is the displacement from equilibrium.

• Results for ss stacking
  \[ k_1 = 10^7 \text{ s}^{-1} \]
  \[ k_1 = 10^7 \text{ s}^{-1} \]
  Therefore, \( \tau \approx 100 \text{ ns} \) (THIS IS SLOW. READ ON.)

• Laser T-jump is useful. Capacitor-discharge can be used (\( \tau > 0.1 \mu\text{s} \), but aligns molecules).

• OK, so what meaning do the rate constant have? To extract a meaning, let’s compare values of related systems, and then use a model to try to understand the differences.

• Example 1: \( \text{A}-(\text{CH}_2)_{3}-\text{A} \) stacks \( 1/\tau = 10^8 \text{ s}^{-1} \)

• Example 2: 2 adenines stack \( 1/\tau = 10^8 \text{ M}^{-1}\text{s}^{-1} \)

• High ‘effective concentration’ (10-100 M) of bases attached to the backbone, naively suggests that \( k_1 \) should be \( > 10^9 \text{ s}^{-1} \). WHY NOT?

\[
k = \frac{eRT}{hN} \exp\left[ -\frac{E_a}{RT} + \frac{\Delta S^\dagger}{R} \right]
\]

The Eyring Equation.

\( e = 2.72, \ h = \text{Plank's constant, } N = \text{Avogadro's #},^4 \)
\( R = \text{gas constant} \)

---

^4 This equations come from transition state theory, which considers the ground state to be in ‘equilibrium’ with the transition state. Because of this, \( \Delta S^\dagger=\Delta S^\dagger_1-\Delta S^\dagger_-1 \) and \( \Delta H^\dagger=\Delta H^\dagger_1-\Delta H^\dagger_-1=\Delta E_a^\dagger_1-\Delta E_a^\dagger_-1 \). Note that \( \Delta H^\dagger = E_a - RT \).
• Can make an Arrhenius Plot of ln k vs T^{-1} to graphically obtain the parameters Ea and \(\Delta S^\ddagger\).

• Table 8.2 gives typical parameters for ss stacking. Notice that \(\Delta S^\ddagger\) is \(-7.5\) to \(-30\) eu. This is an unfavorable entropy change upon moving from the ground state to the transition state. Suggests that stacking involves constraining bonds and reducing available states.

• Chemists Heed: EVEN NON-COVALENT REACTIONS HAVE TRANSITION STATES AND CAN BE TREATED WITH TRANSITION STATE THEORY! Meaning the Eyring equation can be used.

Conformational Entropy:

• See Fig 2-3 in BCT for the seven torsion angles in a nucleotide.

• Available conformations for the dinucleotide monophosphate

\[
\begin{align*}
\alpha & = 1 \\
\beta & = 3 \\
\gamma & = 3 \\
\delta & = 2 \times 2 \ (1 \text{ for each nucleoside}) \\
\epsilon & = 3 \\
\xi & = 2 \\
\chi & = 2 \times 2 \ (1 \text{ for each nucleoside})
\end{align*}
\]

• Entropic contribution to free energy.

\[
\Delta G^\circ = -RT \ln K = \Delta H^\circ - T \Delta S^\circ \quad \text{gives}
\]

\[
\Delta S^\circ = R \ln K = R \ln ([S]/[U]), \text{ where } [\] \text{ means number of conformations here.}
\]

\([U] = 1 \times 3 \times 3 \times (2 \times 2) \times 3 \times 2 \times (2 \times 2) = 864 \text{ different conformations}
\]

\([S] = 1
\]

\(\Delta S^\circ = 0.001987 \text{ kcal K}^{-1} \text{ mol}^{-1} \times 10^3 \text{ cal/kcal ln 864}^{-1} = -13.4 \text{ eu for dinucleoside stacking.}
\]

• Entropic contribution to \(\Delta G^\circ_{37} = -T \Delta S^\circ = -13.4 \text{ eu x 310.15 K/1000} = +4.15 \text{ kcal mol}^{-1}
\]

• This is very similar to the \(\Delta G^\circ_{\text{init}}\) in Table 8.4, consistent with idea that initiation of helix formation is dominated by fixing the conformation of the initiating two nucleotides, aka “entropic fixation.” (Suggests initiating a duplex may be similar to the process of stacking to nucleotides.)

• Negative values for \(\Delta S^\circ\) are inconsistent with the hydrophobic effect (where \(\Delta S^\circ = +\)).
• See also Table 2.2 (p40 BCT and Gellman paper earlier in semester).

B.) Kinetics of duplex formation

\[
\begin{array}{c}
\alpha & 1 \\
\beta & 3 \\
\gamma & 3 \\
\delta & 2 \\
\varepsilon & 3 \\
\xi & 2 \\
\chi & 2 \\
\end{array}
\]

• $\Delta S^o = [1.987 \text{ eu ln } \{1 \times 3 \times 3 \times 2 \times 3 \times 2 \times 2\}^{-1}] \times 2 = -21.4 \text{ eu for base pair formation/bp}$
  (where the last $x2$ is for 2 bases in a pair)

• This value of $\Delta S^o$ is similar to (or somewhat less unfavorable than) NN $\Delta S^o$ (see p276. Table 8.4). Suggests that conformational entropy presents an enormous barrier to folding and dominates the folding issue. For example, this $\Delta S^o$ leads to a $\Delta G^o_{37}$ of $+6.6/\text{bp}$ formed. For a catalytic RNA of 400 nts with maybe 100 bp, this is an unfavorable $\Delta G^o_{37}$ effect of close to 700 kcal/mol!

• This conformational entropy loss must be offset by enthalpic stabilization.

• Note that the addition of a 5'-phosphate restricts conformation of ribose leading to small stabilization and observation of terminal bp in NMR imino experiments. (Sundaralingam 1969, 1973).

• 3'-dangling end stacking: See L3p16-17. 3'-end bases can stack to give $\Delta G^o_{37}$ of $-1.8$ kcal/mol. (Table 8.5)

• Entropic penalty is estimated at $+1.9$ kcal/mol at 37 °C (Freier et al., 1986). Therefore, the intrinsic stacking energy$^5$ may be larger than $-1.7$ at at least $-1.7 - 1.9$ kcal mol$^{-1} = -3.6$ kcal mol$^{-1}$

• Hydrogen bonding is estimated to be worth a $\Delta G^o_{37}$ of $-0.5$ to $-2.0$ kcal mol$^{-1}$. See BCT pp285-287 for calculations estimates. Again, the intrinsic energy of hydrogen bond is probably worth much more than this.

$^5$ Intrinsic energy was an idea developed by Bill Jencks. It goes something like this. There is an incredible amount of bonding energy innate in most interactions, however most of this is never realized because it must be “paid” to achieve the interactions. Here, being paid in conformational entropy to align the bases properly for stacking; in the case of hydrogen bonding, paid for entropic fixation and desolvation of the hydrogen bond donors and acceptors. I like to think of intrinsic energy as “How much money would I have if I never had to pay any bills” — fanciful but amusing and insightful.
Drug–DNA Interactions

April 13, 2016

Reading: BCT Chpt 12

Drug-DNA Interactions

- A large variety of drugs have been synthesized or occur naturally (including antitumor and antibiotic compounds) that interact strongly and specifically with DNA. Many drugs also interact with RNA, especially on the ribosome, including a wide range of antibiotics. In recent years, a number of companies have started up to try to develop small molecules (<10K MW) that bind to RNA. Antibiotic resistance of organisms is an increasingly important problem, and a great amount of resources are currently being invested into developing novel antibiotics.

- Study of Drug-DNA interactions is a multidisciplinary problem. Of course, structure is of vital importance, but the binding thermodynamics are equally important to a complete description of the process. We will look at both aspects.

BCT Fig 12-1 shows the two main classes of DNA binding compounds.

I. Intercalators.

- These molecules are often aromatic cations. As you might guess, these compounds like to stack between the bases, and the positive charge provides favorable electrostatic interaction with the backbone.

- These compounds are often heteroaromatic (i.e.: they have N in the aromatic ring), and are fused (i.e.: multi-ring, rigid beasties) and often have an unfused ring too (e.g. ethidium bromide).

- Intercalation often lengthens the helix by ≈1 bp, and unwinds the helix.

- See Fig 12-2 for examples of ethidium bromide and daunomycin intercalated into DNA

II. Non-Intercalators (Groove Binders)

- Often prefer the minor groove of AT stretches

- Can have hydrophobic interaction with H2 of A in the minor groove.

- Drugs have curved shapes that follow the groove. (See Fig 12-1, 12-3 for Berenil bound to DNA)

- These are often unfused and cationic, with cationic nature often coming from terminal basic groups.

Equilibrium binding models
• Consider a DNA with N bp; n = bp/drug and $B_{ap} =$ binding sites/bp (=1/n).

• N.B. $B_{ap}$ is a constant for a given drug. $B_{ap}$ never changes during an experiment.

• r is the occupancy ratio of [bound drug]/[bp], which changes during an experiment and has a maximum value of $B_{ap}$.

I. **Independent sites model (Figure 12-4a) aka Scatchard Model.**

• potential binding sites do not overlap

• gaps allowed, if sizes are n, 2n, 3n, ...

• maximum drug/DNA is N/n

II. **Neighbor exclusion model (Figure 12-4b)**

• potential binding sites can overlap

• gaps of any size are allowed

• number of potential binding sites is almost N (for N>>n and low r)

• maximum drug/DNA is N/n

important to consider any sequence–specificity of a given drug.
Scatchard Binding Isotherm
I think of this as the drug being in equilibrium with a concentration of binding sites, rather than in equilibrium with a concentration of DNA.

\[
DNA + drug \iff Complex
\]

\[
K_{ap} = \frac{[\text{complex}]}{[\text{drug}][\text{sites}]}
\]

\[
C_N^o \equiv [bp]_{total} = [bp] + [bp]_{bound}
\]

\[
C_B \equiv [\text{bound drug}] = [\text{complex}]
\]

\[
C_F \equiv [\text{drug}]
\]

\[
r \equiv [\text{bound drug}]/[bp]_{total} = C_B/C_N^o
\]

or \( C_B = rC_N^o \)

\[
[\text{sites}] = \frac{[bp]}{n} = \frac{C_N^o - [bp]_{bound}}{n} = \frac{C_N^o - C_B n}{n}
\]

\[
= \frac{C_N^o - rC_N^o n}{n} = C_N^o \left( \frac{1}{n} - r \right) = C_N^o \left( B_{ap} - r \right)
\]

\[
\therefore K_{ap} = \frac{rC_N^o}{C_F C_N^o (B_{ap} - r)}
\]

or \( \frac{r}{C_F} = K_{ap} (B_{ap} - r) \)
Absorbance Melting Curves of RNA

By Joseph D. Puglisi and Ignacio Tinoco, Jr.

The transition between an ordered, native structure and a disordered, denatured state in a nucleic acid can be conveniently monitored using ultraviolet (UV) absorbance. As the ordered regions of stacked base pairs are disrupted, the UV absorbance increases. The increase in absorbance is called hyperchromicity; the absorbance in the disordered state approaches the sum of the absorbances of the constituent nucleotides. Thus, the absorbance of a native nucleic acid is hypochromic relative to its nucleotides; the amount of hypochromicity is a measure of the base pairing and stacking of the secondary structure. The easiest way to denature a nucleic acid is by heating. The resulting profile of absorbance versus temperature is called a melting curve, due to its similarity in appearance to a phase transition. From the absorbance data, a curve of the fraction of one component versus temperature can be constructed; the midpoint of the transition is defined as the melting temperature, $T_m$. Qualitative and quantitative information about conformations of RNA molecules can be obtained from measurement of UV absorbance melting curves. The percentage of hyperchromicity on melting at a chosen wavelength depends on the number and type of base pairs broken. The dependence on RNA strand concentration of the $T_m$ of a melting transition yields information on the molecularity of a transition (unimolecular hairpin to coil, bimolecular duplex to single strand, etc.). This information can be obtained from a qualitative analysis of the melting curves. Further analysis can yield quantitative thermodynamic data for the melting transition. This chapter will deal with the experimental methods needed to acquire a melting curve and the analysis and interpretation of the data.

Experimental Methods

Any standard commercial UV spectrophotometer can be equipped to measure melting curves. A useful instrument is a single-beam Gilford (Oberlin, OH) spectrophotometer (Model 2530) with an automated reference compensator that allows melting curves to be obtained on three separate samples simultaneously. Four cuvettes are placed in a thermistor-controlled thermoelectric cell holder, which is connected to a thermoprogrammer that controls the heating rate; three cuvettes are for samples, the fourth cuvette contains the reference solvent. The temperature also...
can be controlled using a circulating water-ethylene glycol bath with equal effectiveness. Data can be collected on any microcomputer, which can be interfaced to the spectrometer through a RS-232 port. Absorbance data can be collected over the temperature range from ~0 to 90° at a heating rate of 1-0.25°/min. Absorbance data should be acquired about every 0.2°. The sampling rate, controlled by the Gilford spectrophotometer, should be set to measure at least 1 sec at each temperature point. Each data point can be the average of several absorbance readings and temperature readings per point; this gives smoother data at lower absorbance.

A heating rate should be chosen that is consistent with the rates of the processes being observed. The Gilford thermoprogrammer can heat at rates of 1, 0.5, and 0.25°/min. For the normal melting of short duplexes to single strands, the rates of the forward and reverse reactions in the transition region are fast enough to justify a heating rate of 1°/min. Since a correct melting curve is an equilibrium measurement, this assumption should be tested by measuring the melting curve at a lower heating rate; the two curves should be identical. Some processes have very slow kinetics. Sequences that form hairpins can exist in either a hairpin (monomer) form or in a duplex with an internal loop (dimer). At higher strand concentrations, the dimer may be favored at low temperature, with the dimer first melting to hairpins and then to single strands. The equilibrium melting curve for these processes should be measured at a very low heating rate, since the dimer to hairpin transition can be very slow. Melting curves measured at too high a heating rate will give erroneous results. A decade divider circuit can be added to the Gilford thermoprogrammer, which allows heating rates as low as 0.025°/min to be used. Once the heating rate has been set, the data acquisition rate should be adjusted to acquire data every 0.1-0.4°. Normally, a melting curve data set consists of 200-400 data points. After a melting experiment, the sample should be cooled to the starting temperature, and the final absorbance should be compared to the initial absorbance. Any evaporation or hydrolysis of the sample will result in a rise in absorbance. Differences of ~1% or less are acceptable.

A wide range of cell path lengths are available for UV melting experiments. The standard 1-cm path-length cell is usually the longest path length used; 5-, 2-, and 1-mm cells are commercially available, and path lengths of 0.1 mm are obtainable with spacers. This means a concentration range of about a factor of 1000 can be measured by using an absorbance of 0.2 in a 1-cm cell and an absorbance of 2 in a 0.1-mm cell (absorbance $A = -\log(I/I_0) = ecl$, where $c$ is the strand concentration, $l$ is the path length, and $e$ is the extinction coefficient per strand). Of course, one
must realize that an absorbance of 2 means only 1% of the light is transmitted, and thus a poor signal-to-noise ratio results. Although the absorbance is usually measured relative to a reference cell containing only solvent, the absolute absorbance of all sample cells relative to air should not be above 2.

Cells that are 0.5 cm wide, rather than the normal 1-cm-wide cells, are preferred, since only 250–300 μl of sample are required; these cells also allow faster and more even heating of the sample. Shorter path-length cells should be used in conjunction with aluminum adapters, such that the cells are in thermal contact with the sample holder. The 10- and 5-mm path-length cells are sealed with Teflon stoppers. Since 1- or 2-mm path-length cells are usually stopperless, they are sealed with a small amount (10 μl) of Dow silicone oil (Corning, 200-fluid 20-centipoise viscosity). This oil completely prevents evaporation. Sealing shorter path-length cells requires careful attention. For a 0.1-mm path, about 7 μl of sample is placed at the bottom of a 2-mm cuvette, and the spacer is carefully slid in. Any bubbles in the sample should be removed before adding the oil. This can be done by briefly (30 sec) centrifuging the cuvette placed in an Eppendorf tube. A small amount of oil is then added; if excess oil is added, the oil can creep down the sides of the cell during an experiment and cause spurious results. If the sample is to be recovered, the sample must be separated from the silicone oil. This can be accomplished by pipetting the sample–oil mixture onto a Teflon dish. The oil will stick to the Teflon, as the bead of sample is rolled around the dish.

One major advantage of using UV spectroscopy is the high sensitivity of the method. Normally, the absorbance of the sample used should be between 0.2 and 2.0. The value of the extinction coefficient for RNA molecules at their absorbance maximum is typically 10^5–10^6, so for an absorbance of 0.2, concentrations as low as 0.5 μM in RNA strands can be studied. A typical volume for a sample is 0.3 ml, so that nanomoles of RNA are needed.

Sample preparation for UV melting studies is straightforward. The RNA stock solution is prepared by dialysis against the desired buffer, and different concentrations are made by dilution. Samples should then be degassed in preparation for a melting experiment. Oxygen dissolved in the sample will form bubbles at higher temperatures, which will scatter light and affect absorbance measurements. Simple degassing procedures are to either bubble N₂ or argon through the sample for about 10 min to saturate the sample with these gases or to subject the sample to a vacuum for about 5 min. Care should be taken by moderating the vacuum to avoid vigorous bumping of the sample. Degassing is especially important for short path-length (2 mm or less) samples; bubbles seem to form easier in these cells. Each sample should be heated above its melting point (to 80–
90°) and allowed to equilibrate at the starting temperature (normally 0–5°) for a minimum of 15 min before the melting curve is determined. After heating, check for air bubbles. For equilibria with slow kinetics, such as hairpin-duplex transitions, longer equilibration times are necessary. Equilibration can be monitored by the change of absorbance as a function of time. For measurements below 20°, the sample compartment should be purged with N₂ to prevent moisture condensation on the cells.

Since the melting behavior will depend on the solution conditions, choice of solvent is very important. Most work on model compounds¹⁻² was done in 1 M NaCl and 10 mM sodium cacodylate or phosphate, pH 7.0 (with 0.1 or 1 mM ethylenediaminetetraacetic acid (EDTA); note that 1 mM EDTA has a high A₂₆₀). The high salt concentration was chosen to minimize electrostatic repulsion between strands and to avoid divalent ions, which catalyze hydrolysis of RNA and favor triple-strand formation. This solvent provides a standard condition for measuring melting curves and for comparing results with previously published data. However, other salt conditions may be needed depending on the sequence and structure being studied. For example, the Tₘ of a structure may be too high (>80°) in 1 M NaCl to allow analysis of the melting curve (see Fig. 2b).

The formation of more complex secondary and tertiary structures often requires the addition of divalent ions, which bind very specifically to stabilize the structure (a good example is tRNA).³ Lower monovalent ion concentrations are also used in these situations to avoid competition with divalent ion binding. The buffer of choice is usually phosphate, which has a very low temperature coefficient. Cacodylate is sometimes used because of its lower binding constant to divalent ions. Traditional biochemical buffers, such as Tris and HEPES, should not be used because of their high temperature variability (ΔpKₐ/° = -0.031 and -0.014, respectively).⁴ Ideally, RNA samples should be dialyzed into the desired buffer, especially for melting curves in low salt (<100 mM). A flow dialysis apparatus (Microdialysis System, BRL, Gaithersburg, MD) is ideal for the dialysis and recovery of small volumes (approximately 300 μl). The use of 1000 molecular-weight cutoff dialysis tubing (Spectrum Medical Industries, Los Angeles, CA) allows the safe dialysis of oligonucleotides as small as 8 nucleotides. A typical dialysis sequence for the preparation of a sample in 50 mM NaCl, 10 mM sodium phosphate, and 0.1 mM

---

³ A. Stein and D. M. Crothers, Biochemistry 15, 157 (1976).
⁴ Calbiochem, "Buffers" Behring Diagnostic, La Jolla, California, 1985.
EDTA, pH 7, is as follows: dialyze once against 50 mM NaCl, 10 mM sodium phosphate, and 10 mM EDTA, pH 7 (12 hr) and once against final buffer. The dialysis against a high EDTA concentration, in the presence of NaCl, is crucial for the removal of any divalent impurities.

In general, salt will bind preferentially to the state with the greatest charge density. For nucleic acids, the double-stranded state has a greater charge density than the single strands. We can write the equilibrium as follows:

$$M_n \cdot D \rightleftharpoons M_{n'} \cdot S_a + M_{n'} \cdot S_b + \Delta n \ M$$

where M is an ion or any small molecule that binds differentially to the states and $\Delta n$ is the net number of salt molecules released or bound per mole of double strands (D) melted to single strands (S) ($\Delta n = n - n' - n''$). The equilibrium constant for this dissociation is

$$K = \frac{[(M_{n'} \cdot S_a)(M_{n'} \cdot S_b)(M)^{\Delta n}]/(M_n \cdot D)}{1}$$

and $d \ln K/d \ln(M) = \Delta n$, which is the change in the number of ions bound. Usually we want to know how the $T_m$ changes with salt concentration. Using the van't Hoff equation in the form

$$d \ln K/dT = \Delta H^o/RT^2$$

one obtains

$$dT_m/d \ln(M) = -\Delta n \ RT_m^2/\Delta H^o$$

where $\Delta H^o$ is the standard enthalpy change of melting the duplex; $\Delta H^o$ is a positive quantity. Since, $\Delta n$ is positive (ions are released on melting), the $T_m$ will increase with the salt concentration. Experimentally, up to about 1 M salt concentration, the $T_m$ increases with increasing salt. For higher salt concentrations the $T_m$ may decrease; salts, such as NaClO₄ at 6 M or above can lower the $T_m$ by 30–40°. Below 0.1 M salt, $T_m$ is linear in the logarithm of the salt concentration for DNA, RNA, and hybrid polynucleotide helices. For each factor of 10 increase in NaCl, the $T_m$ increases 17–19°, but depends on base composition; the increase reaches a plateau by 1 M salt. The increase for oligonucleotides is smaller, but the increase is still substantial. The melting of triple strands will be more dependent on salt concentration than double strands; the increase of $T_m$ with 10-fold salt increase is about 30° for triple-stranded polynucleotides.

The wavelength of UV light that is most useful for a melting curve varies between 240 and 280 nm. The absorbance maximum for most RNAs is 260 nm, and this is the wavelength of maximum hyperchro-

5 M. T. Record, Jr., Biopolymers 5, 975 (1967).
micity. This is the wavelength normally monitored in a melting experiment. However, RNA structures with a preponderance of A·U or G·C base pairs should be monitored at different wavelengths. As shown in Fig. 1, A·T base pairs in DNA show a maximum hypochromicity at 240 nm, while G·C base pairs have a maximum hypochromicity at 280 nm; similar effects occur for RNA. A comparison of melting curves (the percentage of hypochromicity) at these different wavelengths can give information about the base composition of the double-stranded structures that are melting (A·U rich versus G·C rich).

Data Analysis

Typical absorbance versus temperature profiles are shown in Fig. 2a and b. The following sections will describe how to analyze and interpret these data. The first important parameter is the total strand concentration. This can be determined spectrophotometrically using $A = e c l$. The extinction coefficient, $e$, for any native structure can be determined experimentally by hydrolyzing the RNA to nucleotides and by measuring the $A_{260}$ of the resulting mixture. This gives the molar concentration of the nucleotides (the base composition of the RNA must be known), and thus the extinction coefficient of the RNA can be determined. The extinction coefficient for a single-stranded molecule can be estimated, if one assumes only nearest-neighbor interactions among the bases in the sequence. For

---

**Fig. 1.** Change in extinction coefficient ($\Delta \varepsilon$) for melting of G·C and A·T base pairs in DNA as a function of wavelength. (Data from Felsenfeld and Hirchman.)

---

FIG. 2. (a) A melting curve with well-defined base lines; 5'-GCGAUUUCUGACCGC-3' in 50 mM NaCl, 10 mM sodium phosphate, and 0.1 mM EDTA, pH 6.5. Normalized absorbance at 260 nm is plotted versus temperature. (b) A melting curve with undefined upper base line. 5'GGGAGUUUCUGCUCUC-3' in 1 M NaCl, 10 mM sodium phosphate, and 0.1 mM EDTA, pH 7.0. Normalized absorbance (----) at 260 nm is plotted versus temperature with the approximate upper and lower base lines (-----) used for data analysis. (c) Fraction double strands (f) versus temperature (T) curves corresponding to a (-----) and b. The two f versus T profiles for melt (b) correspond to unsubtracted base lines (-----) and subtracted base lines (----).

example, the calculation of the extinction coefficient for 5'-ApCpGpUp ...

\[
\varepsilon(\text{ApCpGpU...}) = 2[\varepsilon(\text{ApC}) + \varepsilon(\text{CpG}) + \varepsilon(\text{GpU}) + \ldots] \\
- [\varepsilon(\text{Cp}) + \varepsilon(\text{Gp}) + \varepsilon(\text{Up}) + \ldots]
\]
TABLE I

EXTINCTION COEFFICIENTS FOR NUCLEOTIDES AND DINUCLEOSIDE PHOSPHATES TO CALCULATE EXTINCTION COEFFICIENTS OF SINGLE STRANDS

<table>
<thead>
<tr>
<th>Nucleotide or dinucleoside phosphate</th>
<th>( \varepsilon(260) \ M^{-1} \text{cm}^{-1} \times 10^{-3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ap</td>
<td>15.34 15.34</td>
</tr>
<tr>
<td>Cp</td>
<td>7.60 7.60</td>
</tr>
<tr>
<td>Gp</td>
<td>12.16 12.16</td>
</tr>
<tr>
<td>Up(Tp)</td>
<td>10.21 8.70</td>
</tr>
<tr>
<td>ApA</td>
<td>13.65 13.65</td>
</tr>
<tr>
<td>APC</td>
<td>10.67 10.67</td>
</tr>
<tr>
<td>ApG</td>
<td>12.79 12.79</td>
</tr>
<tr>
<td>CpA</td>
<td>10.67 10.67</td>
</tr>
<tr>
<td>CpC</td>
<td>7.52 7.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nucleotide or dinucleoside phosphate</th>
<th>( \varepsilon(260) \ M^{-1} \text{cm}^{-1} \times 10^{-3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpG</td>
<td>9.39 9.39</td>
</tr>
<tr>
<td>CpU(CpT)</td>
<td>8.37 7.66</td>
</tr>
<tr>
<td>GPA</td>
<td>12.92 12.92</td>
</tr>
<tr>
<td>GpC</td>
<td>9.19 9.19</td>
</tr>
<tr>
<td>GpG</td>
<td>11.43 11.43</td>
</tr>
<tr>
<td>GpU(GpT)</td>
<td>10.96 10.22</td>
</tr>
<tr>
<td>UpA(TpA)</td>
<td>12.52 11.78</td>
</tr>
<tr>
<td>UpC(TpC)</td>
<td>8.90 8.15</td>
</tr>
<tr>
<td>UpG(TpG)</td>
<td>10.40 9.70</td>
</tr>
<tr>
<td>UpU(TpT)</td>
<td>10.11 8.61</td>
</tr>
</tbody>
</table>

\[ a \] At 260 nm, 25°C, 0.1 ionic strength, pH 7. Data calculated by K. H. Johnson and D. M. Gray, Program in Molecular Biology, The University of Texas at Dallas, based on nucleotide extinction coefficients from the Ph.D Thesis of M. Alexis, University of London, 1978, provided by E. G. Richards, and hypochromicity data from the Ph.D. Thesis of M. M. Warshaw, University of California, Berkeley, 1966. The extinction coefficients are estimated to be valid to ±0.10.

where \( \varepsilon(\text{ApC}) \), etc. are the extinction coefficients for the component dinucleoside phosphates per mole of nucleotides (this is the reason for the factor of two), and \( \varepsilon(\text{Cp}) \ldots \) are the extinction coefficients for the nucleotides (note that the two end nucleotides are not subtracted). Shown in Table I are the extinction coefficients for the mononucleotides and dinucleoside phosphates needed to calculate the single-strand extinction coefficient at 25°C. These calculations are good to within 10% of the true values. As can be seen in Fig. 2a, the extinction coefficients for the double-stranded and single-stranded states depend on temperature. This dependence is usually approximately linear, so that the single-strand absorbance can be extrapolated to 25°C for use with the calculated single-strand extinction coefficient.

Melting curves can be analyzed either semiquantitatively to determine hypochromicity and \( T_m \) or quantitatively to determine thermodynamic parameters. First, one must determine the type of transition represented by the melting curve; melting curves can be normalized to the same absorbance to allow comparison of different curves. Table II gives the relevant equations for the common transitions that can be studied by
<table>
<thead>
<tr>
<th>Reaction type</th>
<th>Equilibrium constants</th>
<th>$\Delta H^0$ from slope of $f$ versus $T$</th>
<th>Concentration dependence of $T_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomolecular</td>
<td>$K = \frac{[H]}{[S]} = \frac{(f)}{(1 - f)}$</td>
<td>$\Delta H^0 = 4RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{R}{\Delta H^0} \ln c_t + \frac{\Delta S^0}{\Delta H^0}$</td>
</tr>
<tr>
<td>Bimolecular (self-complementary)</td>
<td>$K = \frac{[D]}{[S]^2} = \frac{f}{2(1 - f)^2 c_t}$</td>
<td>$\Delta H^0 = 6RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{R}{\Delta H^0} \ln c_t + \frac{(\Delta S^0 - R \ln 4)}{\Delta H^0}$</td>
</tr>
<tr>
<td>Bimolecular (non-self-complementary)</td>
<td>$K = \frac{[D]}{[S_A][S_B]} = \frac{2f}{(1 - f)^2 c_t}$</td>
<td>$\Delta H^0 = 6RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{2R}{\Delta H^0} \ln c_t + \frac{[\Delta S^0 - R \ln 4/3]}{\Delta H^0}$</td>
</tr>
<tr>
<td>Trimolecular (identical strands)</td>
<td>$K = \frac{[T]}{[S]^3} = \frac{f}{3c_t^2(1 - f)^3}$</td>
<td>$\Delta H^0 = 8RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{2R}{\Delta H^0} \ln c_t + \frac{(\Delta S^0 - 2R \ln 6)}{\Delta H^0}$</td>
</tr>
<tr>
<td>Trimolecular (nonidentical strands)</td>
<td>$K = \frac{[T]}{[S_A][S_B][S_C]} = \frac{9f}{c_t^2(1 - f)^3}$</td>
<td>$\Delta H^0 = 8RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{3R}{\Delta H^0} \ln c_t + \frac{(\Delta S^0 - R \ln 2)}{\Delta H^0}$</td>
</tr>
<tr>
<td>Tetramolecular (identical strands)</td>
<td>$K = \frac{[Q]}{[S]^4} = \frac{f}{4c_t^2(1 - f)^4}$</td>
<td>$\Delta H^0 = 10RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{3R}{\Delta H^0} \ln c_t + \frac{(\Delta S^0 - 3R \ln 8)}{\Delta H^0}$</td>
</tr>
<tr>
<td>Tetramolecular (nonidentical strands)</td>
<td>$K = \frac{[Q]}{[S_A][S_B][S_C][S_D]} = \frac{64f}{c_t^2(1 - f)^4}$</td>
<td>$\Delta H^0 = 10RT_m^2 \frac{df}{dT}_{T=T_m}$</td>
<td>$\frac{1}{T_m} = \frac{3R}{\Delta H^0} \ln c_t + \frac{(\Delta S^0 - 3R \ln 8)}{\Delta H^0}$</td>
</tr>
</tbody>
</table>
melting experiments. Unimolecular (hairpin to coil) and multimolecular transitions (duplex or triple strand to coil) can be distinguished by varying the concentration of RNA strands. The $T_m$ and the shape of the melting curve [fraction of species versus temperature ($T$)] should be independent of strand concentration for a unimolecular transition. For multimolecular transitions, $T_m$ and the shape of the melting curve will depend on concentration. The concentration should be varied over a factor of about 100 to test the concentration dependence; a smaller concentration variation may miss the change in $T_m$. It should be noted that the melting of long RNA duplexes, such as poly[(rA)·(rU)], is concentration independent, although bimolecular; the concentration-dependent initiation event is negligible for a long enough polynucleotide.

It is difficult to distinguish between types of multimeric transitions, since the $T_m$ values for all the transitions will depend on strand concentration. Varying the stoichiometry of the reacting strands or determining the size of the constituent species by electrophoresis should be done to determine the molecularity of a structure. One should be especially careful when working with sequences that are partly self-complementary, since these sequences can exist as either duplex or hairpin structures. However, once the type of transition has been identified, the melting curve can be analyzed using the theory of helix-to-coil transitions.

Standard helix-to-coil theory describes transitions between a native structure and a melted structure (single strands). A helix-to-coil transition can be induced by changing any of a number of thermodynamic variables—pressure, concentration of reactants, concentration of salts, and most commonly, temperature. The transition can be monitored as a function of temperature by any physical property that is dependent on the number of base pairs formed. For example,

$$A = fA_D + (1 - f)A_S$$

where $f$ is the fraction of bases paired and $A_D$ and $A_S$ are the values of the property for the single-stranded (S) and double-stranded (D) species, respectively. The physical property $A$ is usually UV absorbance, but can also be circular dichroism, NMR chemical shift, etc. The origin of hypochromicity is the electronic interactions between neighboring stacked bases. Theoretical calculations and experimental data show that per-

---

Hypochromicity (%)$= \frac{(A_S - A_D)}{A_S}$

is approximately a linear function of the number of stacked bases. UV absorbance therefore monitors the fraction of bases that are stacked as a nucleic acid molecule is melted.

A melting curve can be analyzed using Eq. (4); this curve relates the absorbance (or other property) to a profile of fraction of bases paired ($f$) versus temperature. The $T_m$ is the temperature where $f = 0.5$. So far, we have assumed no explicit model for how the native double-stranded state is in equilibrium with single strands. In order to derive thermodynamic parameters for the transition ($\Delta H^o$, $\Delta S^o$, and $\Delta G^o$), the absorbance melting curve must be translated into the concentrations of the single-stranded and double-stranded states. This is done most simply by assuming a two-state (all-or-none) model. This model assumes that single strands are in equilibrium with only one base-paired native structure; there are no partially base-paired intermediates in the melting process. This approximation is most appropriate for short (<12 bp) duplexes. If the two-state model is accurate, $f$ is the fraction of fully base-paired strands. Expressions are given in Table II for equilibrium constants as a function of $f$ and $c_t$, where $c_t$ is the total concentration of RNA strands (equimolar amounts of complementary strands are assumed). Expressions are also given for obtaining $\Delta H^o$ from the slope of $f$ versus $T$ at the $T_m$ and from the concentration dependence of $T_m$. Equations can easily be derived for reactions of molecularity greater than three (for example, four-stranded cruciform structures). If we assume a two-state model, a melting curve can be converted to a fraction native structure versus temperature profile, which in turn gives an equilibrium constant at each temperature. These data can then be treated using the van’t Hoff relation

$$d \ln K/d(1/T) = -\frac{\Delta H^o}{R}$$ (5)

and standard thermodynamic equations

$$\Delta G^o = -RT \ln K$$ (6)

$$\Delta S^o = \frac{(\Delta H^o - \Delta G^o)}{T}$$ (7)

to obtain the standard enthalpies, entropies, and free energies per mole of the reaction: $\Delta H^o$, $\Delta S^o$, and $\Delta G^o$ (see Thermodynamic Parameters for details).

For longer base-paired sequences (>12 bp), the helix-to-coil transition is usually not two state.\textsuperscript{14} RNA molecules with complex structures often melt in stages,\textsuperscript{16} with separate regions of structure melting independently (see Complex RNA Molecules). This type of non-two-state transition may produce a shoulder in the melting curve. However, even melting curves that appear two state can involve intermediates in the melting process.\textsuperscript{14} The analysis of transitions, where the double-stranded (native) and single-stranded states are in equilibrium with a significant number of partially base-paired intermediates, usually requires the use of statistical mechanics.\textsuperscript{12}

**Thermodynamic Parameters**

Once the type of transition (unimolecular, bimolecular, etc.) is known, thermodynamic quantities can be derived using several different methods; Table II gives the relevant equations. The methods are (1) plot $\ln K$ versus $1/T$ at a single concentration, (2) plot $(df/dT)$ versus $T$ for a single concentration, (3) fit an absorbance versus $T$ profile at a single concentration, and (4) plot $\ln c_t$ versus $1/T_m$. Clearly, the last method cannot be used for unimolecular transitions. The advantages and disadvantages of each method are evaluated below; all methods, except for the third method above, require that absorbance curves be converted to $f$ versus $T$ curves.

The absorbance versus temperature profile is converted into the fraction of molecules base paired ($f$) versus $T$, using Eq. (4). The temperature dependence of the extinction coefficients of the double strands and single strands must be taken into account. The sloping upper and lower base lines in absorbance versus $T$ plots (the so-called base-line problem) is caused by the single-stranded and double-stranded states changing with temperature. The temperature dependence is usually approximated by assuming a linear dependence of the extinction coefficients on temperature where $m$ and $b$ are the slope and the intercept, respectively.

\[
\varepsilon_D = m_D T + b_D
\]

\[
\varepsilon_S = m_S T + b_S
\]

If the base lines are well defined (10–15° of linear absorbance versus $T$) as in Fig. 2a, $\varepsilon_D$ and $\varepsilon_S$ can be determined using a linear least-squares fit of the absorbance data in the base-line region. However, many times either one or both base lines are not well defined. If the $T_m$ is too high (>80°) or too low (<20°), the upper or lower base lines may not be sufficiently defined for a good least-squares fit. In these situations, the data should be analyzed using different approximations for the base lines. Figure 2c

shows the fraction versus temperature profiles for data from Fig. 2b, generated using two different choices for the base lines. Recognizing which base line is best will be discussed below. A poor choice of base line can be the largest source of error in determining thermodynamic parameters from UV melting data.

Method 1: \( \ln K \) versus \( 1/T \). Values of \( K \) are calculated at each temperature from \( f \), using the appropriate equation from Table II. Normally, only points with \( 0.15 \leq f \leq 0.85 \) are used in the van’t Hoff plot, because this is the region where \( K \) values are most precise. From Eqs. (5–7), one sees that a plot of \( \ln K \) versus \( 1/T \) yields \(-\Delta H^0/R\) as the slope and \((\Delta S^0/R)\) as the intercept; this is a van’t Hoff plot. If \( \Delta H^0 \) is independent of temperature, the plot should be linear. A nonlinear van’t Hoff plot can result from several factors: temperature dependence of \( \Delta H^0 \), poor choice of base lines, or a non-two-state transition.\(^\text{17} \) The most likely candidate is a poor choice of base lines.\(^\text{14} \) In this case, base lines should be adjusted, and the data reanalyzed. It is important that the van’t Hoff data actually be plotted, so that a poor least-squares fit due to errors in the data can be distinguished from a nonlinear plot.

Method 2: \( df/dT \). This is a variation of the van’t Hoff analysis that involves numerical differentiation of the \( f \) versus \( T \) data. The differentiation can be done using a nonlinear least-squares fit to a quadratic equation at each data point.\(^\text{18} \) \( \Delta H^0 \) can be obtained from \( df/dT \) at the \( T_m \) (at \( f = 0.5 \)), as shown in Table II; note that in general, the \( T_m \) is not the maximum in the derivative \( (df/dT) \) versus \( T \) plot. Expressions relating \( df/dT \) to \( \Delta H^0 \) at any value of \( f \) have been derived by Gralla and Crothers\(^\text{19} \) and discussed by Marky and Breslauer.\(^\text{20} \) Thus, \( \Delta H^0 \) can be determined from any portion of the derivative curve, such as the full-width, or half-width, of the derivative curve at half-height. For example, \( \Delta H^0 \) can be determined from only the upper half of a melting curve (i.e., for a molecule with very low \( T_m \)). The advantage of this method is that the results are less sensitive to the choice of base lines than the first method. A good test of the choice of base lines is comparison between the \( \Delta H^0 \) determined using the van’t Hoff plots and the derivative method. If the results using both methods do not agree to within \( \sim 5\% \), then the base lines should be readjusted, and the data reanalyzed. For longer duplexes, the transitions are sharper and the base lines are better defined; in these cases, disagreement between the two methods is more likely due to a non-two-state transition.

19 J. Gralla and D. M. Crothers, J. Mol. Biol. 73, 497 (1973).
Method 3: A versus T. The experimental absorbance versus temperature curve is fit directly to six parameters: $\Delta H^0$, $\Delta S^0$, and the four parameters that specify the slopes and intercepts of the upper and lower base lines. The raw melting curve is fit$^{17}$ by the Marquardt nonlinear least-squares method$^{18}$ to the following equations

$$
A = fA_D + (1 - f)A_S
$$
$$
A_D = \varepsilon_D c_t l = (m_D T + b_D)c_t l
$$
$$
A_S = \varepsilon_S c_t l = (m_S T + b_S)c_t l
$$
$$
K = \exp(- \frac{\Delta H^0}{R} + \frac{\Delta S^0}{RT})
$$

Petersheim and Turner$^{17}$ report differences of only 0.5% between the raw data and the calculated curve. This method is similar to the $\ln K$ versus $1/T$ method, except that the whole curve is used in the fit, not just the central part of the transition curve, and the base line parameters are fit simultaneously with $\Delta H^0$ and $\Delta S^0$. One must be careful in using this method, when either base line is not well defined or has an anomalous shape. The six parameters fit the experimental curve well enough, but the thermodynamic parameters are not meaningful.

Method 4: $\ln c_t$ versus $1/T_m$. For all of the above methods, thermodynamic data should be calculated from replicate experiments at more than one concentration. A very easy and effective method to obtain thermodynamic data is from the concentration dependence of the $T_m$. The relevant equations relating $T_m$ and total strand concentration $c_t$ are listed in Table II; a van’t Hoff plot of $\ln c_t$ versus $1/T_m$ yields $\Delta H^0$ and $\Delta S^0$. All that needs to be determined precisely for each experiment is the strand concentration and $T_m$. Normally, the melting curves are measured at a minimum of 10 different strand concentrations over 2–3 orders of magnitude (micromolar–millimolar). In principle, as $c_t$ is raised, only the $T_m$ should change, with no change in the percentage of hypochromicity. Experimentally, melts at higher concentrations often show a greater percentage of hypochromicity than at lower concentrations; this is usually ascribed to end-to-end aggregation of the RNA double strands.$^{21}$ A nonlinear plot of $\ln c_t$ versus $1/T_m$ indicates possible non-two-state behavior, of which aggregation is a specific example. Nevertheless, this method of measuring thermodynamic parameters has advantages over the other averaging methods, since $\Delta H^0$ and $\Delta S^0$ just depend on measuring $c_t$ and $T_m$, which are less sensitive to the choice of base lines.

Table III summarizes the enthalpies for the melting of d(GC)$_3$, determined by the different methods outlined above.$^{22}$ The $\Delta H^0$ values deter-

---


TABLE III
ENTHALPIES FOR COIL-TO-HELIX TRANSITION OF d(GC)$_3$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Method 1 (ln $K$ versus $1/T$)</th>
<th>Method 2 (slope at $T_m$)</th>
<th>Method 3 (curve fit)</th>
<th>Method 4 (ln $c_i$ versus $1/T_m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No base lines subtracted</td>
<td>$-39.0$</td>
<td>$-42.1$</td>
<td>$-33.0$</td>
<td>$-58.1$</td>
</tr>
<tr>
<td>Upper base line subtracted</td>
<td>$-37.0$</td>
<td>$-41.4$</td>
<td>$-$</td>
<td>$-57.6$</td>
</tr>
<tr>
<td>Both base lines subtracted</td>
<td>$-58.9$</td>
<td>$-56.8$</td>
<td>$-56.6$</td>
<td>$-57.4$</td>
</tr>
</tbody>
</table>

* Data in kilocalories per mole from Albergo et al.$^{22}$

mined using all four methods agree with each other only if base lines are properly subtracted. All these values agree with the enthalpy determined calorimetrically, which indicates that the two-state model is valid for this molecule. Only the ln $c_i$ versus $1/T_m$ is in agreement with the calorimetric enthalpy if the base lines are not subtracted; this demonstrates the insensitivity of this method to base lines.

One should keep in mind the implicit and explicit assumptions made in deriving thermodynamic parameters from melting curves. Activities are replaced by molar concentrations in the equilibrium constants; this determines the standard state. The standard values of $\Delta H^0$, $\Delta S^0$, and $\Delta G^0$ refer to a reaction at 1 M concentration of each species, but with each species having the properties corresponding to an infinitely dilute solution (an ideal solution). Although we are interested in the thermodynamic properties of the reaction extrapolated from infinite dilution, we must specify a concentration (1 M), because $\Delta S$ and $\Delta G$ depend on concentration even for ideal solutions. $\Delta H$ is independent of concentration for ideal solutions. The standard thermodynamic values obtained are valid only for the solvents used in the experiments. Other concentrations of salt, or other kinds of salt, may produce different values. Second, as discussed earlier, the absorbance is assumed to be linearly related to the fraction of bases paired (an assumption based on the theory of hypochromicity). Third, the native to single-strands transition is assumed to proceed in a two-state (all-or-none) manner. The validity of thermodynamic data, derived from melting curves, depends on the validity of this model. A good test is agreement between data derived using the different methods outlined above. In general, any assumption of a two-state model can be experimentally tested by observing the melting by a different technique [calorimetry, nuclear magnetic resonance (NMR), circular dichroism, etc.].
the transition is indeed two state, the thermodynamic data should be independent of the measuring technique. Fourth, the \( \Delta H^\circ \) and \( \Delta S^\circ \) values, derived from this analysis, are assumed to be independent of temperature. Since

\[
d \Delta H / dT = \Delta C_p
\]

where \( \Delta C_p \) is the difference of heat capacities (at constant pressure) of the native and melted states, this assumption can be checked by plotting \( \Delta H^\circ \) values obtained using methods 1, 2, or 3 versus \( T_m \); the slope is \( \Delta C_p \). The errors introduced by this assumption are small, but Petersheim and Turner\textsuperscript{17} have introduced a simple correction for \( \Delta H^\circ \) based on Eq. (11).

### Prediction of Melting Behavior

The previous sections have described how to determine experimentally the thermodynamic stability and melting behavior of a RNA molecule. One can also calculate the approximate \( \Delta G^\circ \) and \( T_m \) for melting a RNA secondary structure. In order to predict the melting behavior of a complex RNA molecule, one needs to know the free-energy contributions of regions of duplexes, bulges, hairpin loops, and internal loops relative to the single strands. These structural elements are shown schematically in Fig. 3. Contributions from G·U base pairs and from dangling single strands must also be included. Prediction of secondary structure has been reviewed recently.\textsuperscript{23} It is assumed that the free energy for a secondary structure is the sum of the free energies for the separate regions.\textsuperscript{2} The further approximation is made that the free energy of a duplex region is the sum of nearest-neighbor interactions. According to the nearest-neighbor model, the free energy for duplex formation consists of an initiation free energy for formation of the first base pair plus a sum of propagation free energies for formation of the subsequent base pairs. There is also a small symmetry term, which arises for double-strand formation of self-complementary sequences. For example, the free energy of formation for the following non-self-complementary duplex

\[
\begin{array}{cccc}
5' & G & C & G \\
& C & G & U \\
3' & G & C & A \\
& C & G & C \\
3' & C & G & A \\
& C & G & U \\
5' & G & C & G
\end{array}
\]

Fig. 3. A representation of the various types of secondary structures that a RNA molecule can adopt.

is calculated as the sum of nearest-neighbor contributions with zero-symmetry correction

$$\Delta G_{\text{total}}^{\circ} = \Delta G^{\circ}_{CG} + \Delta G^{\circ}_{GC} + \Delta G^{\circ}_{CA} + \Delta G^{\circ}_{AC}$$

$$+ \Delta G^{\circ}_{CU} + \Delta G^{\circ}_{UC} + \Delta G^{\circ}_{U'A} + \Delta G^{\circ}_{A'U} (=0) + \Delta G^{\circ}_{\text{init}}$$

$\Delta S^{\circ}$ and $\Delta H^{\circ}$ values can be calculated in an analogous manner. Standard free energies, entropies, and enthalpies for all nearest neighbors at 37° in 1 M NaCl have been determined by Turner and co-workers. The oligonucleotides used for the determination of these parameters were chosen such that their $T_m$ values were near 37°; thus, the parameters required little extrapolation. The free energies depend on temperature [see Eq. (7)]; the enthalpies and entropies are assumed temperature independent, but are most useful near 37°. The $T_m$ of any oligonucleotide can be predicted using the expressions in Table II. As noted by Freier et al., the
difference between measured and predicted parameters are due not only to the inaccuracy of the data set, but also to the imperfection in the nearest-neighbor model, i.e., contributions due to next nearest-neighbor interactions.

Since RNA secondary structure consists of duplexes, loops, bulges, and single-stranded regions (Fig. 3), one needs to know the thermodynamic contributions of loops and bulges, as well as duplex regions, in order to calculate the free energy. The free-energy contributions of non-bonded loops and bulges are simply added to the free energy for the duplex region. Duplex initiation free energies are contained in the $\Delta G^\circ$ for a loop. Bulge free energies are just the destabilization of the perfect duplex. Published free-energy parameters for loops and bulges are very limited; these data assume that $\Delta G^\circ$ of a loop only depends on its size, with a maximum stability at a loop size of six or seven. However, recent data show that loop free energy depends both on size and sequence. 24

In view of the limited data on these structures, for precise comparison of RNA sequences, model compounds should be studied. A good example of this approach are the studies by Tuerk et al. of hairpin loop sequences, which occur with very high frequency in intercistronic regions in RNA produced by T4 bacteriophage infection. 25 These investigators showed that certain four-base loops can have stabilities much greater than predicted by the present parameters. Similarly, Groebe and Uhlenbeck studied a number of variants of the R17 coat protein binding site sequence. 26 This sequence contains a bulged A in the stem and a four-base hairpin loop. In this study, the loop sequence was kept constant, while the bulged nucleotide or flanking base pairs were varied. These variants were compared to the hairpin with a perfect duplex stem.

In order to evaluate, or to predict, a folded structure for a RNA molecule, one must consider other factors in addition to the secondary structural elements discussed previously. Specific ion binding may stabilize a particular secondary or tertiary structure. Tertiary interactions may favor, or may prevent, the formation of certain secondary structures. The free-energy contributions from tertiary structural elements, such as pseudoknots and base triplets, are not known. Thus, well-designed model compounds must be studied to provide the missing free-energy values.

Other Techniques

Temperature-Jump Methods. Derivatives of melting curves can be obtained directly by using a temperature-jump apparatus. The temperature is rapidly (microseconds) increased by a few degrees, and the corresponding absorbance change (ΔA) is measured. There are two major advantages of this method over normal UV melting studies. First, the base-line problem is eliminated. The molecular processes, which give rise to sloping base lines (most likely, unstacking of single or double strands), are very fast compared to the actual melting processes. Thus, the ΔA due to these very fast steps can be resolved from the ΔA due to the slower melting. This advantage can also be used to resolve overlapping transitions, which occur at different rates. Crothers et al. and Riesner et al. have used this method to study tRNA melting and the stability of hairpin and internal loops. Differential melting curves can also be measured directly using a double-beam spectrometer with two identical RNA samples at slightly different (0.1°) temperature.

Nuclear Magnetic Resonance (NMR). Melting transitions can be followed using the chemical shift of the nonexchangeable protons, usually aromatic protons. The chemical shift versus temperature profile can be analyzed using a two-state model

\[ \delta_{\text{obs}} = f\delta_{\text{native}} + (1 - f)\delta_{\text{coil}} \]  

where \( \delta_{\text{obs}} \) is the observed chemical shift of a given proton, \( \delta_{\text{native}} \) and \( \delta_{\text{coil}} \) are the chemical shifts of that proton in the native and coil forms, respectively, and \( f \) is the fraction of the native form. From the calculated fraction versus temperature profile, thermodynamic parameters can be derived. The advantage of using NMR to study melting is that, in principle, each residue in the sequence can be monitored once the NMR spectrum has been assigned. This allows the validity of the two-state model to be tested: in a two-state transition, melting curves for different residues should yield the same thermodynamic data. Also, for complex molecules like tRNA, which have multiphasic melting profiles, NMR can assist in determining what is melting in each transition. This was done in tRNA by monitoring the exchangeable imino protons for each stem. However, NMR has major disadvantages. The assumption of a two-state model in analyzing a chemical shift versus temperature curve is questionable. This is because kinetically fast exchange between the two states (duplex and

single strand) is assumed for Eq. (12) to be valid

\[ k_{\text{ex}} \gg 2\pi \Delta \nu \]  \hspace{1cm} (13)

where \( k_{\text{ex}} \) is the rate constant for exchange between the two states and \( \Delta \nu \) is the difference in resonant frequencies for a proton in the two states. This often is not the case; if the rate constant is too slow, intermediate exchange (\( k_{\text{ex}} \approx 2\pi \Delta \nu \)), resulting in line broadening, or slow exchange (\( k_{\text{ex}} \ll 2\pi \Delta \nu \)), resulting in two lines for the two states, can occur. Another obvious difficulty with NMR melting experiments is the high concentration required for NMR in general (millimolar strand concentrations). NMR melting curves should be compared to optical melting curves taken on the same sample at the high NMR concentration and at the lower optical concentrations to ensure that the same phenomenon is being studied.

Calorimetry. Unlike the other methods presented in this section, calorimetry allows the determination of \( \Delta H \) directly; the method of differential scanning calorimetry (DSC)\(^{30,31}\) is most often used. In DSC, the excess heat capacity of the transition (\( \Delta C_p \)) is measured as the temperature is varied. Since \( \Delta H = \int \Delta C_p \, dT \), the integrated area under a DSC versus \( T \) curve is the transition enthalpy. The major advantage of DSC over optical melting is that no assumptions about the transition need to be made; DSC values of \( \Delta H \) are model independent.\(^2\) However, it is important to remember that the \( \Delta H \) measured calorimetrically is the enthalpy of whatever process occurs between the two states used to define the base lines. Unlike the van't Hoff standard enthalpy, \( \Delta H^\circ_{\text{vH}} \), which is obtained from the temperature dependence of the equilibrium and refers to the standard-state conditions (1 M strand concentrations, but with the properties of infinitely dilute solutions), the calorimetric \( \Delta H \) refers to the conditions of the experimental measurement (strand concentration measured in millimolar). These values will be equal only (for the same solvents) if the enthalpy of the reaction is independent of the concentration of oligonucleotides over the concentration range studied optically and calorimetrically.

A schematic DSC transition curve is shown in Fig. 4. In order to determine the area under the transition curve, upper and lower base lines (states of approximately constant heat capacity) must be defined, and initial and final temperatures for integration must be chosen. If the base lines are not colinear, the initial and final states have different \( C_p \). Incomplete transitions, or very broad transitions, present the same problems as


in the van't Hoff methods. However, comparison of $\Delta H$ from an optical melting experiment to $\Delta H$ from calorimetry can be very helpful. For example, if $\Delta H_{\text{vH}} < \Delta H_{\text{CAL}}$, this indicates the presence of intermediates in the melting process. The major drawback of DSC is the large amount (0.5 ml at millimolar concentrations) of material needed. The small concentrations and small amount of material needed make optical melting the current method of choice for studies of RNA secondary structure.

Complex RNA Molecules

We will outline briefly in this section some of the information that can be obtained from melting studies of complex RNA molecules. Examples of such molecules studied to date include tRNA, 5 S RNA, viroids, and viral RNA. The melting transition of a complex RNA will not be two state; the melting curve is often multiphasic. Comparison of hypochromcities at 260 and 280 nm will allow an estimate of $A \cdot U$ and $G \cdot C$ base pairs
broken for each transition in the melting curve; this will aid in assigning transitions. The $T_m$ values of the transitions usually are affected differently by addition of Mg$^{2+}$ or Na$.^+$ Cole et al. constructed a phase diagram of tRNA$^{Phe}$ as a function of salt.$^{16}$ Each transition may be interpreted as an independent melting of a portion of the molecule. The temperature-jump method is especially useful for resolving slow ($t >$ milliseconds) and fast ($t <$ milliseconds) transitions. For example, the melting of the tertiary structure to an extended, base-paired structure is slow, while the melting of individual hairpin regions is fast. The $\Delta H^o$ for each transition can be determined from the differentiated melting curve, assuming the transition is independent of others.

Acknowledgments

We are very grateful to Professor Donald Gray, The University of Texas at Dallas for providing the data in Table I and for making helpful comments. Dr. Susan Freier (Molecular Biosystems, Inc., La Jolla, CA) carefully read the manuscript and made valuable suggestions; her help is greatly appreciated. The work was supported by grant GM 10840 from the National Institutes of Health and grant DE-FG03-86ER60406 from the Department of Energy.


By Shelby L. Berger

Genomic exons are characterized by means of nuclease protection assays. Cloned, single-stranded genomic DNA probes are hybridized with cellular RNA; the hybrids are treated with a single-strand-specific nuclease to remove unreacted probe together with irrelevant RNA and unprotected stretches of DNA; and finally, the protected hybrids are characterized by means of gel electrophoresis. The number of surviving fragments equals the number of exons. The mobility of each fragment is a measure of the size of that exon.

The nuclease digestion method presented in this chapter is the inverse of nuclease protection; exons are degraded, while introns are preserved. The method is as follows. Synthetic, genomic sense-strand RNA is first hybridized to an excess of single-stranded antisense cDNA. Then, the purified hybrids are treated with ribonuclease H (RNase H), an enzyme that degrades only RNA in DNA–RNA hybrids. As a consequence, the exons in the RNA moiety are degraded; the RNA fragments that survive are the introns. These fragments can be analyzed electrophoretically,