Thermal Methods for the Analysis of RNA Folding Pathways

RNAs form a remarkable variety of structures based on standard Watson-Crick helices, noncanonical pairings (e.g., as found in some internal and hairpin loops; Heus and Pardi, 1991; Correll et al., 1997), and tertiary interactions that join a loop or bulge to another part of the RNA. An objective of current RNA physical studies is to identify noncanonical and tertiary folding motifs and evaluate the factors responsible for their stability. Thermal melting analysis was used in the 1970s to establish the existence of a distinct set of tertiary interactions in transfer RNAs (Cole et al., 1972; Römer and Hach, 1975; Stein and Crothers, 1976) and has become a useful tool for examining the folding of unusual RNA structures up to ~100 nucleotides in length.

Suppose one has identified an RNA or RNA fragment that adopts a specific, functional conformation, e.g., the minimal sequence containing a ribozyme activity or a protein recognition site. The first task in determining the RNA structure is to devise a model of the secondary structure. A good approximation of canonical base pairing in an RNA is fairly easy to obtain from a combination of comparative sequence analysis (Gutell et al., 1992), computer prediction programs based on nearest-neighbor thermodynamic parameters (Zuker, 1989; Serra and Turner, 1995; *UNIT 11.2*), and "structure mapping" experiments (Ehresmann et al., 1987). Compensatory base changes may be able to establish the existence of specific helix segments required for function (Tang and Draper, 1989). At this point, it is worthwhile asking whether the deduced secondary structure accounts for all the intramolecular interactions of the RNA, or whether noncanonical and tertiary interactions might make the structure much more stable than predicted, or link parts of the structure in unexpected ways. One may also be interested in determining, as an aid to further structural or functional analysis, the pH, salt, and temperature ranges under which the RNA adopts a stably folded structure. Relatively simple melting experiments can answer these questions, and more extensive melting analysis can frequently provide a complete unfolding pathway for an RNA.

This unit describes procedures for applying melting analysis to an RNA. The discussion assumes that the secondary structure has already been established by a combination of prediction and experiment (see Chapter 6). After comments on sample preparation and instrumentation, a framework for describing the thermal denaturation of an RNA that unfolds in several steps is derived. An analysis of transfer RNA unfolding then illustrates the application of this framework to UV and calorimetry data sets and also shows how error analysis can point out uncertainties in the derived thermodynamic parameters. Lastly, comments are offered on experiments that can resolve ambiguities in the thermodynamic analysis and help identify unfolding transitions with specific structures in the unfolding pathway.

EXPERIMENTAL CONSIDERATIONS

RNA Samples

Transcription of DNA templates by T7 RNA polymerase (*UNIT 9.3*) is the easiest route to milligram quantities of RNAs longer than 15 to 20 nucleotides; methods have been reviewed elsewhere (Draper et al., 1988; Gurevich, 1996). Chemical synthesis is preferred for shorter RNAs, and may be necessary if modified nucleotides are required (Scaringe et al., 1990; Goodwin et al., 1996; *APPENDIX 3C*; also see Chapter 3).

Aggregation of RNAs is a common problem, particularly in the presence of Mg^{2+}, and samples should always be checked for the presence of multimers by nondenaturing gel electrophoresis (Chory and Pollard, 2000). A buffer of 100 mM potassium acetate and 30 mM Tris acetate (pH 7.3), with or without 5 mM magnesium acetate, is convenient (Gluick et al., 1997). Weak multimerization that is not detected by gel electrophoresis will appear as a concentration dependence of the melting temperatures. UV melting curves are typically obtained at RNA concentrations on the order of 5 to 10 µg/mL, while calorimetry requires about two orders of magnitude higher concentrations. Thus, comparison of melting temperatures in the two experiments should reveal problems with weak intermolecular interactions.

Buffers for melting experiments should have small ionization enthalpies to avoid large pH changes with temperature. Tris should be avoided on this account; phosphate or cacody-

late salts are preferred. Some of the Good buffers (Good and Izawa, 1972; *APPENDIX 2A*) are suitable for UV absorbance.

UV Absorbance

Commercial UV spectrophotometers are available with accessories for collecting UV absorbance as a function of temperature, or a programmable water bath and temperature probe may be added to instruments with suitably thermostatted cells. Refer to *UNIT 7.3* for experimental details of UV absorbance melting curves. Peltier devices may seem more convenient than circulating water baths for temperature control, but usually have a limited temperature range (e.g., 10° to 80°C) and a relatively short lifetime when repeatedly cycled over the full range. Instruments with a five- or six-cell carriage that can be automatically cycled for measurement will speed up data collection by a corresponding factor. The cells must be sufficiently surrounded by the heating block that large temperature gradients do not develop. Lastly, a temperature probe should be placed inside a dummy cuvette, and not in the heating block itself. The temperature difference between the block and the cuvette interior should not exceed ~5°C at the highest temperatures.

It is important to test for reversibility of the melting curve by running both heating and cooling experiments on the same sample. Folding kinetics at low temperatures can be slow (for example, see Gluick et al., 1997), and Mg^{2+} ion–induced hydrolysis of RNA at high temperatures is also a common problem. The latter can be minimized by using the lowest Mg^{2+} ion concentrations possible, and by working in a neutral to acid pH range (i.e., 6.0 to 7.0).

Renaturation of an RNA preparation before every experiment is necessary to avoid kinetic artifacts. Rather than devise renaturation protocols for each new RNA and salt concentration, the authors have used the following protocol for data collection. RNA samples are diluted into the desired buffers in stoppered cuvettes and then incubated for 15 min at 70°C in the spectrophotometer. The temperature is then ramped at 0.8° to 1°C/min to ~10°C, and then up to ~90°C, while temperature and absorbance at 260 nm are read every 0.4°C. In some cases, it may be necessary to slow the temperature ramp at temperatures less than ~20°C to achieve equilibrium. Example data are shown in Figure 11.3.1, panel A. The first unfolding transition in this RNA is a set of tertiary interactions whose stability is of interest. The coincidence of the heating and cooling curves demonstrate reversibility, and the two curves provide duplicate sets of data in the temperature range of most interest.

Scanning Calorimetry

Prior to calorimetry experiments (see *UNIT 7.4* for experimental details), it is advisable to thoroughly wash the cells to remove any trace of ribonuclease. A series of extensive rinses with RNase ZAP (Ambion), hot soapy water (using a commercial dishwashing soap such as Dove), and distilled water is adequate for this purpose. The amount of RNA needed for an experiment depends on the number of transitions and their enthalpies; typically ~1.0 mg/ml is used with the most sensitive calorimeters now available from MicroCal and CSC, though 0.2 mg/ml may be adequate for RNAs with single, sharp transitions. After thorough dialysis of the sam-
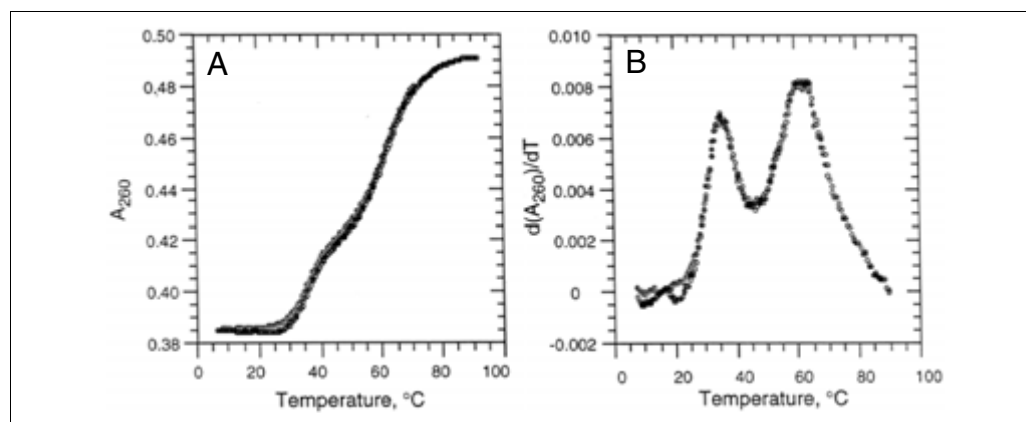

Figure 11.3.1 UV absorbance melting curve. (**A**) Cooling curve (starting at ~70°C, filled symbols) and heating curve (open symbols) for the same RNA sample. (**B**) First derivative calculated for the data set shown in panel A, using linear regression and a window size of 4°C.

ple against the desired buffer, two or three baseline scans using the dialysis buffer should be run. These ought to be reproducible. If not, check for degradation of buffer components. The sample should be renatured before being run; 65°C for 15 min, followed by slow cooling, is adequate for most RNAs, but higher temperatures or longer times may be needed. Renaturation conditions should be surveyed by UV experiments prior to calorimetry. A temperature ramp of 1°C/min is sufficiently slow for most RNAs; again, experience with UV melting experiments will indicate what ramp speed is appropriate.

Calculations of transition enthalpies depend on accurate measurements of RNA concentration in the calorimetry sample. It is therefore advisable to determine the extinction coefficient of the RNA under the buffer conditions used for the experiment. To do this, prepare an RNA sample and a buffer blank in stoppered cuvettes. After measuring the absorbance of the sample, add concentrated NaOH to a final concentration of 1 M and incubate at 37°C overnight. Check for evaporation by weighing the cuvettes before and after incubation. The extinction of the RNA can then be calculated from the absorbance of the hydrolyzed sample and extinction coefficients of the mononucleotides in alkali ($\varepsilon_{260} \times 10^3$ values are: AMP, 15.0; GMP, 11.2; CMP, 7.6; and UMP, 7.4).

THEORETICAL BACKGROUND

Sequential Unfolding Model

This section presents a framework for describing RNA unfolding as a function of temperature. To a good approximation, most short RNA hairpins unfold in a two-state manner, i.e., no partially unfolded intermediates are found in significant concentrations at equilibrium. The equilibrium between RNA in its fully folded, native structure (N) and fully unfolded RNA (U) is then simply:

$$N \overset{K}{\rightleftharpoons} U$$

Equation 11.3.1

Because the enthalpy change (ΔH) for unfolding is positive, an increase in temperature drives the reaction to the unfolded state. The T_m of the reaction is the temperature at which half of the molecules are unfolded, [N] = [U] and $K = 1$. From this definition and the van't Hoff relation for the temperature dependence of K, $d\ln(K)/d(1/T) = \Delta H/R$, where R is the ideal gas

constant, the expression for K can be written as:

$$K = \exp\left[\frac{\Delta H}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right]$$
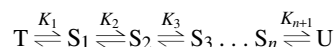
Equation 11.3.2

A melting curve is simply a plot of the fraction of molecules that are unfolded as a function of temperature, or:

$$F_U = \frac{[U]}{[U] + [N]} = \frac{K}{1 + K}$$

Equation 11.3.3

Thus, there are two variables, ΔH and T_m, that describe the melting curve for a simple hairpin, and from which the equilibrium constant K can be calculated at any temperature. The ΔH found from application of Equations 11.3.2 and 11.3.3 to a melting curve is called a two-state or van't Hoff enthalpy. Note that ΔH is assumed to be independent of temperature.

RNAs larger than hairpins unfold in several steps; for example, there may be several helical segments separated by internal loops or junctions loops, and, of particular interest, additional tertiary interactions that "cross-link" different loops and/or helical segments. In the latter case, it may be obligatory that a set of tertiary interactions unfold before helical segments:

$$T \overset{K_1}{\rightleftharpoons} S_1 \overset{K_2}{\rightleftharpoons} S_2 \overset{K_3}{\rightleftharpoons} S_3 \ldots S_n \overset{K_{n+1}}{\rightleftharpoons} U$$

Equation 11.3.4

In this sequential unfolding scheme, T is the completely folded RNA, with intact tertiary interactions, and the S states are RNAs with successively fewer helical segments. Studies of tRNA unfolding in the 1970s showed that its unfolding pathway was best analyzed by a sequential model, with a total of four steps—i.e., tertiary structure plus one of the cloverleaf stems in the first step, followed by the remaining three cloverleaf stems in order of their stabilities (Stein and Crothers, 1976).

The alternative to a sequential unfolding pathway is one in which each set of interactions unfolds independently, for example, a series of hairpins linked by single-stranded RNA and having no interactions with each other. It can be shown that use of a sequential pathway to analyze an independently unfolding RNA will, in most cases, yield an excellent approximation of the unfolding thermodynamics. Since most RNAs will probably unfold by a mixed pathway, in which the first, tertiary unfolding step

is obligatory, but subsequent melting of secondary structures may be largely independent, it is reasonable to use a sequential pathway for an initial analysis in most cases. This point has been considered in some detail by Draper and Gluick (1995), who found that the errors expected in using a sequential pathway to model independent unfolding are on the order of only 5%.

Equation 11.3.3, which gives the fraction of unfolded RNA as a function of temperature for a single unfolding step, now must be modified to account for multiple transitions. It is convenient to use a partition function to do this. If the fully folded RNA in Equation 11.3.4, state T, is assigned as the reference state with a statistical weight of 1, the partition function Q for this sequential pathway is:

$$Q = 1 + K_1 + K_1K_2 + K_1K_2K_3 \ldots$$

Equation 11.3.5

From the definition of equilibrium constants, each successive term represents the ratio of the concentrations of states S_1, S_2...S_n to that of state T. Q is therefore a sum of the relative probabilities of finding all of the different possible conformations of the RNA. The fraction of RNAs in which any particular helices are unfolded can be easily calculated from Q. For example, the fraction (F) of RNAs with only the tertiary structure and the first set of secondary interactions unfolded would be:

$$F_2 = \frac{K_1K_2}{Q}$$

Equation 11.3.6

Each K in Equations 11.3.5 and 11.3.6 represents a single two-state transition, and has a temperature dependence with an associated ΔH and T_m (as in Equation 11.3.2). Readers unfamiliar with the use of partition functions to describe conformational equilibria may wish to consult Wyman and Gill (1990) for a detailed discussion.

Equations 11.3.2, 11.3.4, and 11.3.5 are all that are needed to describe the unfolding of many complex RNAs. The two simplifying assumptions that have been made are (i) the RNA unfolds in a sequential manner, and (ii) each transition is two-state. These assumptions were adequate for analysis of tRNA unfolding (Crothers et al., 1974; Stein and Crothers, 1976), and are a good starting place for the analysis of any RNA thought to have a well-defined structure. Specific application of these equations to UV hyperchromicity and cal-orimetry experiments are developed in the next two sections.

UV Melting Profiles

In a plot of absorbance versus temperature, it is usually difficult to pick out individual melting transitions or to see changes in transition T_m and sharpness between different plots. For this reason, the first step in analyzing UV melting data is to take the first derivative of the absorbance with respect to temperature; individual transitions then appear as peaks that are more easily distinguished. Two methods are available to do this. The simplest is to define a temperature "window," W, usually between 4° and 6°C. For each absorbance data point, a linear least-squares line is fit to all the points within ±W/2°C. The slopes of the lines are plotted as the derivative; an example curve is shown in Figure 11.3.1, panel B. The potential disadvantage of this method is that very sharp melting transitions may be flattened and the associated ΔH will be underestimated. A method that avoids this problem is the Savitsky-Golay convolution (Press et al., 1992). If the data points are evenly spaced in temperature (or have been made evenly spaced by interpolation), then a polynomial fit to the points in a temperature window can be found by multiplying the points by a matrix of appropriate weights. Using a second-order polynomial, derivatives of very sharp transitions are accurately calculated. A disadvantage of this method is that broad transitions are smoothed much less effectively than by linear regression. The authors' experience has been that RNA melting transitions are rarely sharp enough to warrant fitting a second-order polynomial.

The authors usually normalize the absorbance and derivative curves, so that melts done with different RNA concentrations can be directly compared. This is simply done by dividing all the absorbance readings by the absorbance at some arbitrarily chosen temperature. A low temperature at which the RNA is completely folded is preferable, as the absorbance readings at very high temperatures can be variable due to RNA hydrolysis. All the points within ±W/2°C of the chosen normalization temperature are averaged.

In order to analyze an absorbance melting curve, Equation 11.3.6 of the preceding section (which gives the fraction of molecules with a specific structure unfolded) must be modified to relate fraction unfolding to hyperchromicity. First, each unfolding transition must be assigned a hyperchromicity (i.e., the ratio of un-

folded- to folded-state absorbance), and the contribution of each transition to the overall hyperchromicity summed. The complete equation is:

$$\Delta A_t = \frac{\Delta A_1 K_1}{Q} + \frac{(\Delta A_1 + \Delta A_2)K_1 K_2}{Q}$$
$$+ \frac{(\Delta A_1 + \Delta A_2 + \Delta A_3)K_1 K_2 K_3}{Q} + \ldots$$

Equation 11.3.7

where ΔA_t is the total fractional hyperchromicity (the same as the normalized absorbance) and ΔA_1, $\Delta A_2 \ldots \Delta A_n$ are the fractional hyperchromicities associated with each specific transition K_1, $K_2 \ldots K_n$. Q is the partition function for sequential unfolding, Equation 11.3.5. Each term in the numerator represents the hyperchromicity associated with a successive unfolded state in Equation 11.3.4.

Upper- and lower-temperature baselines must be added to Equation 11.3.7, as the absorbances of completely folded or completely unfolded RNA are usually weakly temperature-dependent. After all base pairs are broken at high temperature, the absorbance continues to increase with temperature as bases unstack further. This increase is approximately linear over normally accessible temperature ranges. There is usually a similar increase in absorbance at low temperatures as base-pair stacking within helices decreases with temperature. (RNAs with stable tertiary structures tend to have temperature-independent absorbances at low temperatures, perhaps because the additional structure reduces motions within helices.) In the case of a single unfolding transition, the baselines are taken into account by the following equation (Albergo et al., 1981; Petersheim and Turner, 1983):

$$A = B_F \left(\frac{1}{1+K} \right) + B_U \left(\frac{K}{1+K} \right)$$
$$B_F = A_{F,0} + m_F(T); \ B_U = A_{U,0} + m_U(T)$$

Equation 11.3.8

B_F and B_U are the low- and high-temperature baselines, respectively. $A_{F,0}$ and $A_{U,0}$ are the absorbances of folded and unfolded RNA extrapolated to a reference temperature (e.g., 0°C) and it is assumed that the baselines are linear, with slopes m_F and m_U. With multiple unfolding transitions, it is not as clear how to devise appropriate baselines for intermediate states. Fortunately, it is only necessary to fit the high- and low-temperature baselines, because calculated values for enthalpies are quite insensitive to the choice of baselines for folding intermediates, as long as the baseline values are in between the high and low temperature values. The authors have found the following formula convenient in carrying out least-squares fits; it uses the enthalpy of a transition to calculate a weighted average of the high- and low-transition baselines (ΔH_{tot} is the total enthalpy of unfolding for the RNA):

$$B = B_F +$$
$$\left(\frac{B_U - B_F}{\Delta H_{tot}} \right) \left(\frac{\Delta H_1 K_1 + (\Delta H_1 + \Delta H_2)K_1 K_2 \ldots}{Q} \right)$$

Equation 11.3.9

The sum of Equations 11.3.7 and 11.3.9, ($\Delta A_t + B$), can be fit to a melting profile using standard nonlinear regression methods; the authors' programs use the Levenburg-Marquardt algorithm (Press et al., 1992). Only the three variables associated with each transition (ΔA, ΔH, and T_m) are fitted, and it is left to the user to provide high- and low-temperature baselines, either from linear least-squares analysis of the absorbance data or from experience. It should be noted that, when fitting curves to normalized first-derivative melting profiles, only the two slopes, $m_F(T)$ and $m_U(T)$, are needed, and not $A_{F,0}$ or $A_{U,0}$.

In RNAs with several unfolding transitions, the breadth and spacing of the transitions are frequently such that some of the peaks are not individually resolved. In that case, it may not be obvious how many transitions should be fit to the melting profile because, unless the T_m of each transition is fixed by a separate maximum in the melting profile, there are usually too many degrees of freedom (three variables per transition) to obtain a unique fit. This problem can sometimes be circumvented by collecting data at two different wavelengths. RNA structures frequently differ in their hyperchromicities at different wavelengths; G-C rich helices, for instance, have much larger hyperchromicities at 280 nm than do A-U rich helices (Puglisi and Tinoco, 1989). One then has to fit two melting curves simultaneously, using the same set of ΔH and T_m values for each curve but allowing for different hyperchromicities. While this approach may still result in ambiguities, the benefits of a two-wavelength analysis can be substantial and the experiment should always be tried. An example of two-wavelength analysis applied to transfer RNA melting is shown in a following section, and more extensive methods for ascertaining

and improving the reliability of a fit are also discussed below.

Calorimetry

Calorimetry experiments detect enthalpy directly. The equations used to describe RNA unfolding are similar to those discussed above for absorbance, with ΔH replacing hyperchromicity. Thus, Equation 11.3.7 for sequential unfolding becomes:

$$\Delta H(T) = \frac{\Delta H_1 K_1}{Q} + \frac{(\Delta H_1 + \Delta H_2)K_1 K_2}{Q}$$
$$+ \frac{(\Delta H_1 + \Delta H_2 + \Delta H_3)K_1 K_2 K_3}{Q} + \cdots$$

Equation 11.3.10

where $\Delta H(T)$ is the total enthalpy change in going from fully folded RNA to the collection of partially unfolded molecules at temperature T. The actual measurement made is the heat capacity of the RNA, $\Delta C_P(T) = [\partial \Delta H(T)/\partial T]_P$, so it is the first derivative of Equation 11.3.10 that is fit to the data.

The elimination of hyperchromicities as unknown parameters in Equation 11.3.10 simplifies its use for fitting data. Unfortunately, this advantage is partially offset by new parameters that must be introduced to describe the calorimetry baseline. To see this, first consider a single, two-state unfolding reaction. The completely general equation for the heat capacity curve is:

$$C_p = C_F \left(\frac{1}{1 + K} \right) + C_U \left(\frac{K}{1 + K} \right) + \left(\frac{\Delta H}{RT^2} \right) \frac{K}{(1 + K)^2}$$

Equation 11.3.11

The last term is the temperature derivative of Equation 11.3.10 for a single transition. C_F and C_U are the intrinsic heat capacities of the native and unfolded states, and generally have small temperature dependencies:

$$C_F = C_{F,0} + \left(\frac{dC_F}{dT} \right)(T - T_0)$$
$$C_U = C_{U,0} + \left(\frac{dC_U}{dT} \right)(T - T_0)$$

Equation 11.3.12

where $C_{F,0}$ and $C_{U,0}$ are the heat capacities at reference temperature T_0. Equations 11.3.11 and 11.3.12 are similar to Equation 11.3.8 for an absorbance melting curve, with heat capacity replacing absorbance in the description of the baseline. In both sets of equations, six variables need to be specified (for a heat capacity curve, the variables are ΔH, T_m, $C_{F,0}$, $C_{U,0}$, and the temperature dependencies of the last two variables). It is possible to extract all of these parameters from a heat-capacity curve that extends over a large enough temperature range.

Problems arise when the overall change in heat capacity between native and unfolded RNA, $\Delta C_P = C_{U,0} - C_{F,0}$, is significant and there are several overlapping transitions. In principle, each folding intermediate has an associated heat capacity, $C_{I,0}$, and a temperature dependence of the heat capacity, dC_I/dT. The latter value is usually small, and can be assigned an arbitrary value in between those of the native and unfolded states without introducing significant error. The difficulty comes in determination of $C_{I,0}$, as illustrated in Figure 11.3.2 for a heat-capacity curve with overlapping transitions. Very different baselines and total enthalpies of unfolding are calculated, depending on whether $C_{I,0}$ is close to $C_{F,0}$ (giving a larger total ΔH) or $C_{U,0}$ (smaller total ΔH). This is a similar problem to that encountered with absorbance melting curves, i.e., three variables associated with each unfolding transition (in this case, ΔH, T_m, and $C_{I,0}$) is, in many cases, too many degrees of freedom to fit a unique curve to a data set.

A small ΔC_p is expected for RNA unfolding, due to the temperature dependence of base stacking; the fully unfolded, single-stranded state of an RNA will have less base stacking (and less enthalpy) at higher temperatures. Thus, RNA unfolding reactions will yield progressively larger ΔH values as the temperature of the measurement is raised; this temperature dependence of ΔH gives a positive ΔC_P to any RNA unfolding. The effect is small, on the order of 0.1 kcal/mol per base pair in short helices (Freier et al., 1983; Petersheim and Turner, 1983). However, the authors find the heat capacity baseline problem can be a serious problem for larger RNAs with multiple unfolding transitions, for two reasons: (i) when melting takes place over a wide temperature range, small adjustments in the C_p values associated with each transition can have a significant effect on the total enthalpy, even if the overall ΔC_P is relatively small; (ii) in some situations, the overall ΔC_P of unfolding an RNA tertiary structure is quite substantial, i.e., much larger than ΔC_P for unfolding a protein of similar molecular weight (Laing and Draper, 1994). Though the curve in Figure 11.3.2 might seem an exaggerated case, it is in fact based on actual data and is a realistic illustration of problems that may be encountered. Why some RNAs have
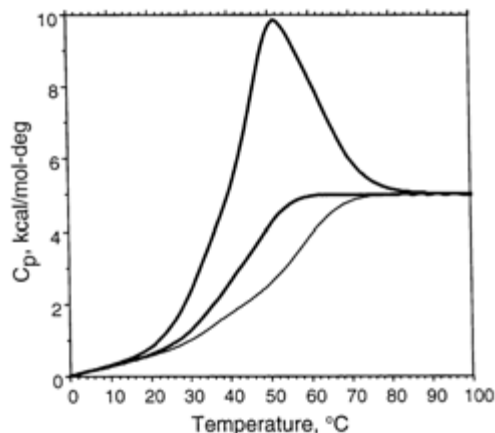
Figure 11.3.2 The baseline problem in fitting heat-capacity data. The heat-capacity curve was generated using Equation 11.3.10 and three transitions with total enthalpy 150 kcal/mol and overall ΔC_p 5 cal/mol·K. The two different baselines were obtained from fits of the heat capacity curve with three transitions (ΔH = 135 kcal/mol) or four transitions (thinner line; ΔH = 173 kcal/mol); there was no significant difference in the quality of the fits to the data.

dramatically larger ΔC_P compared to others is not presently understood.

EXAMPLE ANALYSIS: tRNA UNFOLDING

Analysis of melting data requires considerable effort to ensure that a useful thermodynamic description of the unfolding has been obtained. As an example of problems that may be encountered, this section describes the analysis of transfer RNA melting data and calculations that test the reliability of a proposed fit. The next section describes experiments that can further test a proposed melting pathway.

Initial Fitting of The UV Melting Profile

Figure 11.3.3, panel A, shows the UV melting profile of yeast tRNA[Phe] at 260 and 280 nm. The data clearly distinguish three transitions with different ratios of hyperchromicities at the two wavelengths, and a least-squares fitting program readily finds three sets of transition parameters that give an excellent fit to the data. Note that the low- and high-temperature baselines were fixed manually, and not by the fitting program; the transition parameters are relatively insensitive to the exact baseline values chosen.

The first test of a fit is whether the total unfolding enthalpy is consistent with the RNA secondary structure. From compiled nearest-neighbor base-stacking enthalpies (Serra and Turner, 1995), the four tRNA[Phe] cloverleaf helices alone are predicted to have a total unfolding enthalpy of 190 kcal/mol. "Dangling bases" at the ends of helices and coaxial stacking of helices could add as much as 43 kcal/mol (Serra and Turner, 1995), and further stacking of bases within the anticodon loop and tertiary structure could also add a significant amount of enthalpy. The total enthalpy change associated with the three transitions in Figure 11.3.3, panel A, ~140 kcal/mol, is thus much too small to be consistent with unfolding of the tRNA.

When four transitions are used, the fitting program converges on the set of transitions shown in Figure 11.3.3, panel B. The total enthalpy is now 232 kcal/mol, more in line with the expected value, and the fit has improved significantly (2.7-fold smaller value for χ^2). Further tests are now needed to see if this fit is unique and if all variables are well determined.

Determining if an optimum fit has been found

With a large number of variables, it is quite possible for a least-squares fitting routine to converge on a local minimum, rather than the actual optimum set of parameters. To circumvent this problem, the fitting program can be run a large number of times (>100) using randomized initial values of the parameters. Starting with the transition parameters of the fit in Figure 11.3.3, panel B, a random number generator was used to increase or decrease ΔH and ΔA values by factors of up to 4, and varied T_m values by ±5°C. Forty-seven of the 100 trials converged; one had a larger value for χ^2, and the rest had χ^2 values not significantly different from that of the fit shown in Figure 11.3.3, panel B. From several such sets of trials, using different ranges of the transition parameters, it is clear that a better fit (smaller χ^2 value) has not been missed; however, some of the parameters varied substantially. ΔH_3, for instance, has a range of 14 to 167 kcal/mol. This suggests that the two melting profiles do not have sufficient information to constrain all parameters for four melting transitions.

Determining if the data support unique values for all of the parameters

The problem of whether all the transition parameters are well constrained by a set of melting profiles can be approached by "bootstrap" analysis (Press et al., 1992). This is a Monte Carlo method, in which synthetic data
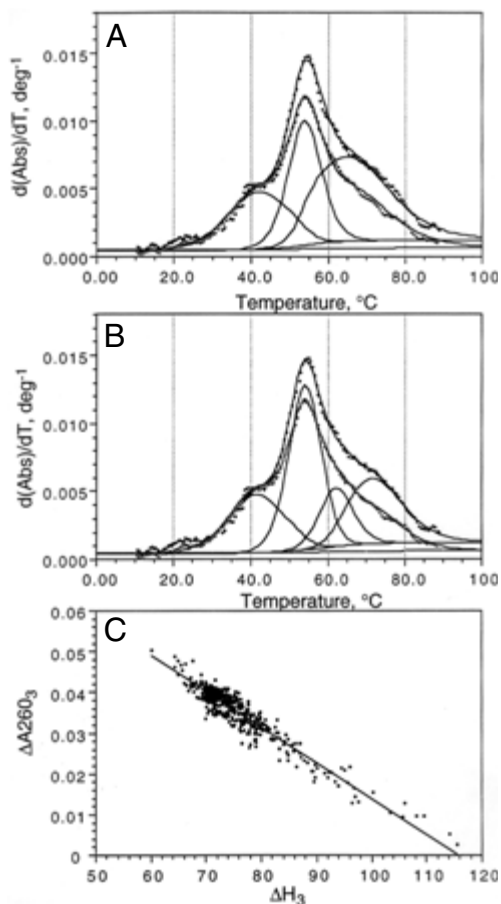
Figure 11.3.3 UV analysis of yeast tRNA^{Phe} melting (buffer contained 100 mM KCl and 5 mM potassium cacodylate). (**A**) Fit obtained assuming three sequential transitions (Equation 11.3.7). Data were obtained at 260 nm (open symbols) and 280 nm (closed symbols). (**B**) Fit of four sequential transitions to the same data set as in panel A. (**C**) Correlation plot, ΔH_3 versus A260$_3$, for the four transition fit (Table 11.3.1).

sets are generated by randomly drawing the same number of points from the original data set as contained in that set. Some points will be duplicated, and other points will be missing; different optimum fits will be found for each synthetic set, as the data points will be differently weighted. Statistics compiled on a large number (200 to 1000) of such trials give the confidence intervals for any one variable. A useful aspect of this analysis is the ability to ask whether different variables are correlated. If changes in one fit parameter are paralleled by changes in another parameter, then the data contain insufficient information to specify unique values of both parameters.

In this analysis, 500 synthetic data sets were generated from the two melting profiles in Fig-

ure 11.3.3B, and the linear correlation coefficients of each fit parameter versus all other parameters were obtained. These values are summarized in Table 11.3.1. All coefficients with absolute values greater than 0.8 are in boldface. The last two lines of the table give the average and standard deviation (1σ) for each parameter. There are a large number of strong correlations. Particularly worrisome are ΔH_3, $A260_3$, and $A280_3$, which strongly correlate with a number of parameters and also have unusually large standard deviations. A plot of ΔH_3 versus $A260_3$ is in Figure 11.3.2C; it is clear from the wide range of values and strong correlation ($R = 0.95$) that the enthalpy and hyperchromicity of transition 3 are not uniquely determined. The authors conclude that the UV melting data sets have enough information to constrain three melting transitions, but not enough information to determine four transitions. Additional data are needed to interpret the unfolding thermodynamics.

Calorimetry Analysis

Calorimetry provides additional information that can be used to interpret the UV melting profile. Figure 11.3.4 shows a heat capacity curve for tRNA^{Phe} under the same buffer conditions as used for the UV melting experiment in Figure 11.3.3. As discussed above, the uncertainty in analyzing these data is interpolation of a baseline. For this curve, the authors used the following procedure:

1. Low- and high-temperature baselines were determined by linear least-squares fitting of data over 10°C ranges (10° to 20°C and 90° to 100°C).

2. An initial set of transition parameters (including C_p values that determine the baseline) was obtained by a fitting routine that apportions the overall ΔC_p between the transitions according to their enthalpies ($\chi^2 = 6.05$; total $\Delta H = 239.6$ kcal/mol). A minimum of four transitions was required to fit the data reasonably.

3. Fitting was continued after relaxing the constraint in step 2 (above) and substituting the constraint that ΔC_p from one transition to the next must be positive; a slightly better fit was obtained ($\chi^2 = 5.72$; total $\Delta H = 241.9$ kcal/mol).

4. The fit parameters were randomly varied (as described above for UV melting profiles) to see if significantly different baselines could be drawn, subject to the constraint that ΔC_p values must be positive. Of 107 converged fits, none was a significantly improved fit (minimum $\chi^2 = 5.71$). Among these fits, the maximum and

Table 11.3.1 Correlation Coefficients from Bootstrap Analysis[a]

	ΔH_1	T_{m1}	$\Delta A260_1$	$\Delta A280_1$	ΔH_2	T_{m2}	$\Delta A260_2$	$\Delta A280_2$	ΔH_3	T_{m3}	$\Delta A260_3$	$\Delta A280_3$	ΔH_4	T_{m4}	$\Delta A260_4$	$\Delta A280_4$	
1.000	**-0.860**	-0.912	-0.926	-0.558	-0.148	0.665	0.612	-0.074	0.288	-0.034	-0.006	0.081	0.108	-0.102	-0.116	ΔH_1
	1.000	**0.906**	**0.894**	0.696	0.190	-0.690	-0.624	0.152	-0.354	-0.015	-0.045	-0.096	-0.176	0.166	0.179	T_{m1}
		1.0000	**0.930**	0.715	0.162	-0.705	-0.633	0.141	-0.373	-0.005	-0.037	-0.116	-0.175	0.167	0.184	$A260_1$
			1.0000	0.727	0.115	-0.721	-0.683	0.114	-0.387	0.0264	0.0003	-0.090	-0.150	0.143	0.158	$A280_1$
				1.0000	-0.268	-0.757	-0.710	0.111	-0.642	0.104	0.0716	-0.047	-0.179	0.170	0.178	ΔH_2
					1.0000	0.510	0.556	0.642	0.398	-0.712	-0.701	-0.335	-0.460	0.469	0.463	T_{m2}
						1.0000	**0.972**	0.432	0.526	-0.579	-0.540	-0.182	-0.270	0.283	0.259	$A260_2$
							1.0000	0.491	0.488	-0.632	-0.611	-0.253	-0.342	0.351	0.336	$A280_2$
								1.000	-0.224	-**0.950**	-0.940	-0.590	-0.896	0.898	**0.879**	ΔH_3
									1.0000	0.132	0.152	0.440	0.499	-0.448	-0.493	T_{m3}
										1.0000	**0.989**	0.703	**0.908**	-**0.914**	-**0.900**	$A260_3$
											1.0000	0.7423	**0.922**	-**0.917**	-**0.921**	$A280_3$
												1.0000	**0.813**	-**0.827**	-**0.873**	ΔH_4
													1.0000	-**0.975**	-**0.981**	T_{m4}
														1.0000	**0.981**	$A260_4$
															1.0000	$A280_4$
42.1	41.5	0.0826	0.0777	71.8	53.9	0.103	0.133	75.2	62.3	0.0357	0.0409	46.96	72.2	0.0576	0.0936	A$\varpi\varepsilon$
1.3	0.3	0.0027	0.0032	1.3	0.1	0.004	0.005	7.14	0.3	0.0066	0.0088	1.42	0.5	0.0046	0.0061	$\Sigma\Delta$

[a]The UV melting data from Figure 11.3.3 were used in 500 cycles of bootstrap analysis, with four sequential transitions; both 260 and 280 nm wavelength data sets were used simultaneously. The table shows the results of statistical analysis of the 492 cycles that converged on a fit; average and standard deviation for each of the variables are shown in the bottom two lines of the table, and linear correlation coefficients for each pair of variables are in the body of the table. Correlation coefficients with absolute values greater than 0.8 are in bold face type. ΔH units are kcal/mol, and T_m values are in °C.

RNA Folding Pathways

11.3.9

minimum total ΔH differed by only 9 kcal/mol (236.3 to 245.2); these curves have the most extreme possible positions of the baseline (Fig. 11.3.4). This source of error is probably comparable to the reproducibility of the heat capacity curve.

5. Bootstrap analysis showed no significant correlations between parameters, and small confidence intervals.

The relatively small ΔC_p for tRNA melting is well within the range expected for unfolding of the four cloverleaf helical segments, and is small enough that uncertainties in the C_p values contribute less than $\pm 2\%$ to the overall error in measuring unfolding ΔH. Note that the assumption that $\Delta C_p > 0$ for any one transition is essential; otherwise, there is not enough information in the curve to determine the parameters of all four transitions. Since the source of ΔC_p in this example is the temperature-dependent stacking of bases in the unfolded state, it is reasonable to assume $\Delta C_p > 0$. In another RNA examined by the authors, ΔC_p is on the order of 5 kcal/mol-K. The assumption of positive ΔC_p values is then made with less confidence, since the origin of the large ΔC_p is unknown.

Agreement of UV Melting and Calorimetry Data Sets

A simple test for consistency between UV and heat capacity data sets is to refit the UV data using fixed enthalpy and T_m values from the calorimetry analysis. When the calorimetric parameters obtained in Figure 11.3.4 were applied to the UV data in Figure 11.3.3, the hyperchromicities adjusted to give an excellent fit. A more general method is simply to fit both UV and heat capacity data sets simultaneously. Bootstrap error analysis can then be run on the combined data sets. With the tRNA^{Phe} data, the parameters obtained are very close to those found by analysis of the calorimetry data alone, and the only significant differences from the UV analysis are the enthalpies of transitions 2 and 3.

Computer Programs

Global Melt Fit, a program for simultaneous curve fitting of UV absorbance and calorimetry data sets, and for applying bootstrap analysis, is available on the Web (*http://www.jhu.edu/~chem/draper/*) or from the authors.

EXPERIMENTAL TESTS OF AN UNFOLDING PATHWAY

An analysis of UV and calorimetry data will invariably yield a set of two-state unfolding transitions whose total enthalpy is at least as large as the predicted nearest-neighbor base-stacking enthalpy. The next question is the physical significance of the transitions, i.e., can individual transitions be identified with specific structures? Though it might seem that each transition should correspond to the unfolding of a single helical segment or set of tertiary interactions, there is no reason why this has to be true. Secondary structure units may melt in different orders by several pathways, so that the unfolding pathway is branched instead of strictly sequential. A single "state" in the sequential unfolding analysis may then represent a mixture of partially unfolded states that have about the same stability (Laing and Draper,
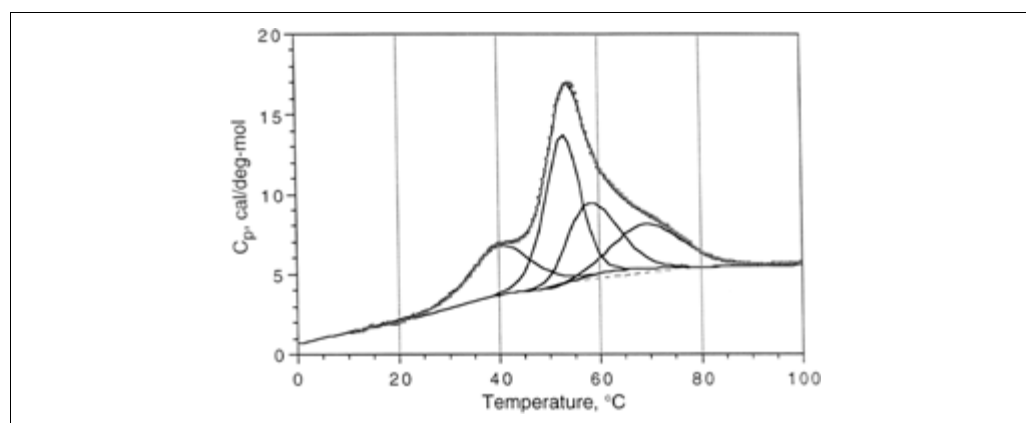

Figure 11.3.4 Calorimetry analysis of tRNA^{Phe} under the same buffer conditions as in Figure 11.3.3. The solid curves represent the total C_p, baseline, and individual transitions for the fit with the smallest total enthalpy: $\Delta H_1 = 48.7$; $T_{m1} = 40.5$; $C_{p1} = 3.95$; $\Delta H_2 = 82.9$; $T_{m2} = 53.2$; $C_{p2} = 4.70$; $\Delta H_3 = 56.7$; $T_{m3} = 59.1$; $C_{p3} = 5.37$; $\Delta H_4 = 49.0$; $T_{m4} = 70.0$; $C_{p4} = 5.39$ (ΔH in kcal/mol, T_m in °C, and C_p in kcal/mol-K). The dashed line is the baseline for a fit with the largest overall enthalpy, and differs principally in its smaller value for C_{p3} (4.83).

1994; Draper and Gluick, 1995). An additional problem is whether the two-state assumption is warranted for a transition. This section suggests additional experiments that can test a hypothesized unfolding mechanism.

Substitution of Base Pairs

Substitution of a helix A-U base pair with a G-C pair, or vice versa, destabilizes or stabilizes the helix by a predictable amount (Serra and Turner, 1995). The resulting change in helix T_m is usually several degrees, which can be easily detected in a melting experiment. If the unfolding pathway is simple (i.e., branches are not significantly populated) then the ΔH and T_m of only a single transition will change significantly in the mutant. Doing this kind of experiment for each helix in the secondary structure is a rigorous test of the unfolding pathway.

Similar substitution experiments can be done for sets of tertiary interactions, if specific bondings are known or suspected. Too little is known about tertiary structure to predict the outcome of such an experiment quantitatively, but it can be useful as a way to confirm the existence of an interaction within a tertiary structure and to probe the range of substitutions that can be accommodated by a structure. Such experiments have been done for a base pair linking D and variable loops within tRNA tertiary structure (Hou et al., 1995), and for a putative base triple interaction in a ribosomal RNA fragment (Conn et al., 1998).

Any time a variant RNA is made, the new sequence should be run through a program that predicts secondary structure, such as Mulfold (Zuker, 1989; see also *http://www.rpi.edu/~zukerm/*). It is not uncommon that a one- or two-base sequence change favors an alternative secondary structure with very different unfolding properties.

Melting of Subfragments

Larger RNAs can usually be divided into several smaller fragments that should retain the same secondary structure as the larger fragment under study (Liang and Draper, 1994; Gluick and Draper, 1997). The melting behavior of such small fragments, with only one or two transitions, is usually easy to analyze and should correspond to transition(s) in the melting of the larger RNA. This is a particularly useful way to find out if the two-state approximation is valid for a specific helix. Note that coaxial stacking and simple entropic effects may alter the stability of a helix when it is taken out of context of a larger, extended hairpin (Draper and Gluick, 1995).

Deviations from Two-State Behavior

The model developed above for RNA unfolding assumes that the RNA behaves as a collection of two-state unfolding events. This assumption is difficult to test rigorously. The usual criterion for two-state behavior is that ΔH calculated from van't Hoff analysis of the unfolding (as described above, Equation 11.3.3) agrees with the calorimetric ΔH measured from the area under the heat-capacity curve. Only in the case of a small RNA with a single transition, or an RNA with very well-resolved transitions, is it possible to carry out this quantitative test. But frequently an RNA can be divided into several segments which can be individually tested (see above).

Deviations from two-state behavior can be anticipated from the sequence of a helix. Using nearest-neighbor base-stacking parameters and loop entropies (Serra and Turner, 1995), a separate equilibrium constant can be calculated for removal of each successive base pair from either end of a helical segment (the so-called "zipper" model of duplex unwinding). The terms for the fully folded state and all the partially unfolded states are then included in a partition function (Equation 11.3.5), and the melting curve is predicted. If the two-state approximation is valid, the calculated curve will not deviate much from a curve predicted using only a single equilibrium constant for folded and unfolded forms. Long helices (more than 6 to 8 base pairs), or helices containing runs of A-U or G-U pairs, tend to "fray" significantly. Fraying may be substantially reduced when the helix is adjacent to loops or helices in a larger structure.

An example in which deviations from two-state behavior were identified is an RNA pseudoknot that contains two coaxially stacked helices (Gluick and Draper, 1997). This RNA melted in two well-resolved transitions. However, the heat capacity curve could only be fit with three two-state transitions; the two peaks of the melting curve were associated with large enthalpy transitions, and a third transition of much lower enthalpy was spread between the two main transitions. Several clues suggested that the low enthalpy transition might be an artifact of non-two-state behavior:

1. The total enthalpy calculated from two-state analysis of UV melting curves increased with salt concentration, even though the enthalpy of helix unfolding is usually independent of salt (Williams et al., 1985). Higher

salt concentrations can reduce the proportion of molecules with "frayed" helices, promoting two-state melting.

2. An RNA fragment representing hairpin 1 of the pseudoknot melted in a transition with lower T_m at 260 than at 280 nm. This suggested that some A-U pairs (with larger hyperchromicities at 260 nm) melted before G-C pairs.

3. The two-state enthalpies of both component hairpins increased with salt concentration.

The UV and heat-capacity melting profiles of the pseudoknot could be fit simultaneously with two transitions if the ratio of van't Hoff and calorimetric ΔH values, $\Delta H_{vH}/\Delta H_{cal}$, were included to allow the transitions to be broader than expected based on the total enthalpy of unfolding. This ratio is a measure of the deviation from two-state behavior, and approaches 1 for a perfectly two-state transition. For the two transitions of the pseudoknot, the ratios varied from 0.68 to nearly 1, depending on the salt concentration (Gluick and Draper, 1997).

LITERATURE CITED

Albergo, D.D., Marky, L.A., Breslauer, K.J. and Turner, D.H. 1981. Thermodynamics of (dG-dC)₃ double-helix formation in water and deuterium oxide. *Biochemistry* 20:1409-1413.

Chory, J. and Pollard, J.D. Jr. 2000. Separation of small DNA fragments by conventional gel electrophoresis. *In* Current Protocols in Molecular Biology (F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds.) pp. 2.7.1.-2.7.8. John Wiley & Sons, New York.

Cole, P.E., Yang, S.K., and Crothers, D.M. 1972. Conformational changes of transfer ribonucleic acid: Equilibrium phase diagrams. *Biochemistry* 11:4358-4368.

Conn, G.L., Gutell, R.R., and Draper, D.E. 1998. A functional ribosomal RNA tertiary structure involves a base triple interaction. *Biochemistry* 37:11980-11988.

Correll, C.C., Freeborn, B., Moore, P.B. and Steitz, T.A. 1997. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* 91:705-711.

Crothers, D.M., Cole, P.E., Hilbers, C.W. and Shulman, R.G. 1974. The molecular mechanism of thermal unfolding of *Escherichia coli* formylmethionine transfer RNA. *J. Mol. Biol.* 87:63-88.

Draper, D.E. and Gluick, T.C. 1995. Melting studies of RNA unfolding and RNA-ligand interactions. *Methods Enzymol.* 250:281-305.

Draper, D.E., White, S.A., and Kean, J.M. 1988. Preparation of specific ribosomal RNA fragments. *Methods Enzymol.* 164:221-237.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P., and Ehresmann, B. 1987. Probing the structure of RNAs in solution. *Nucl. Acids Res.* 15:9109-9128.

Freier, S.M., Burger, B.J., Alkema, D., Neilson, T. and Turner, D.H. 1983. Effects of 3′ dangling end stacking on the stability of GGCC and CCGG double helices. *Biochemistry* 22:6198-6202.

Gluick, T.C. and Draper, D.E. 1997. Folding of an mRNA pseudoknot required for stop codon readthrough: Effects of mono- and divalent ions on stability. *Biochemistry* 36:16173-16186.

Gluick, T.C., Gerstner, R.G. and Draper, D.E. 1997. Effects of Mg^{2+}, K^+, and H^+ on an equilibrium between alternative conformations of an RNA pseudoknot. *J. Mol. Biol.* 270:451-463.

Good, N.E. and Izawa, S. 1972. Hydrogen ion buffers. *Methods Enzymol.* 24:53-68.

Goodwin, J.T., Osborne, S.E., Scholle, E.J., and Glick, G.D. 1996. Design, synthesis, and analysis of yeast tRNAPhe analogs possessing intra- and inter-helical disulfide cross-links. *J. Am. Chem. Soc.* 118:5207-5215.

Gurevich, V.V. 1996. Use of bacteriophage RNA polymerase in RNA synthesis. *Methods Enzymol.* 275:382-397.

Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., and Stormo, G.D. 1992. Identifying constraints on the higher-order structure of RNA: Continued development and application of comparative sequence analysis methods. *Nucl. Acids Res.* 20:5785-5795.

Heus, H.A. and Pardi, R. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* 252:191-194.

Hou, Y.M., Sterner, T. and Jansen, M. 1995. Permutation of a pair of tertiary nucleotides in a transfer RNA. *Biochemistry* 34:2978-2984.

Laing, L.G. and Draper, D.E. 1994. Thermodynamics of RNA folding in a highly conserved ribosomal RNA domain. *J. Mol. Biol.* 237:560-576.

Petersheim, M. and Turner, D.H. 1983. Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, CCGGUp, and ACCGGUp. *Biochemistry* 22:265-263.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. Numerical Recipes in C. Cambridge University Press, Cambridge.

Puglisi, J.D. and Tinoco, I., Jr. 1989. Absorbance melting curves of RNA. *Methods Enzymol.* 180:304.

Römer, R. and Hach, R. 1975. tRNA conformation and magnesium binding: A study of yeast phenylalanine-specific tRNA by a fluorescent indicator and differential melting curves. *Eur. J. Biochem.* 55:271-284.

Scaringe, S.A., Francklyn, C. and Usman, N. 1990. Chemical synthesis of biologically active oligoribonucleotides using beta-cyanoethyl protected ribonucleoside phosphoramidites. *Nucl. Acids Res.* 18:5433-441.

Serra, M.J. and Turner, D.H. 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol.* 259:242-261.

Stein, A. and Crothers, D.M. 1976. Conformational changes of transfer RNA: The role of magnesium(II). *Biochemistry* 15:160-167.

Tang, C.K. and Draper, D.E. 1989. An unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell* 57:531-536.

Williams, A.P., Longfellow, C.E., Freier, S.M., Kierzek, R. and Turner, D.H. 1985. Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry* 28:4283-4291.

Wyman, J. and Gill, S. 1990. Binding and Linkage. Functional Chemistry of Biological Macromolecules. University Science Books, Mill Valley, Calif.

Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* 244:48-52.

Contributed by David E. Draper,
    Yury V. Bukhman, and Thomas C. Gluick
Johns Hopkins University
Baltimore, Maryland