[22] 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 $A = -\log(I/I_0) = \varepsilon cl$, where c is the strand concentration, l is the path length, and ε 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 N_2 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 N_2 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 compounds[1,2] was done in 1 *M* NaCl and 10 m*M* sodium cacodylate or phosphate, pH 7.0 [with 0.1 or 1 m*M* ethylenediaminetetraacetic acid (EDTA); note that 1 m*M* EDTA has a high A_{260}). 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 T_m 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 ($\Delta pK_a/° = -0.031$ and -0.014, respectively).[4] Ideally, RNA samples should be dialyzed into the desired buffer, especially for melting curves in low salt (<100 m*M*). 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 m*M* NaCl, 10 m*M* sodium phosphate, and 0.1 m*M*

[1] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, and D. H. Turner, *Proc. Natl. Acad. Sci. U.S.A.* **83,** 9373 (1986).

[2] I. Tinoco, Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla, *Nature (London) New Biol.* **246,** 40 (1973).

[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 m*M* NaCl, 10 m*M* sodium phosphate, and 10 m*M* 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 Δn 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

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

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

$$d \ln K/dT = \Delta H^\circ/RT^2 \quad (2)$$

one obtains

$$dT_m/d \ln(M) = -\Delta n\, RT_m^2/\Delta H^\circ \quad (3)$$

where ΔH° is the standard enthalpy change of melting the duplex; ΔH° is a positive quantity. Since, Δn 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 $NaClO_4$ 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.[5] 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).

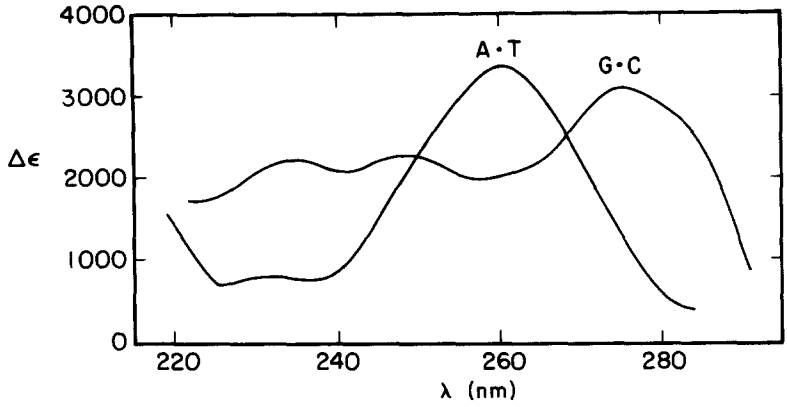

FIG. 1. Change in extinction coefficient ($\Delta\varepsilon$) for melting of G–C and A–T base pairs in DNA as a function of wavelength. (Data from Felsenfeld and Hirchman.[6])

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[6]; 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 $A = \varepsilon cl$. The extinction coefficient, ε, 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

[6] G. Felsenfeld and S. Z. Hirschman, *J. Mol. Biol.* **13**, 407 (1965).

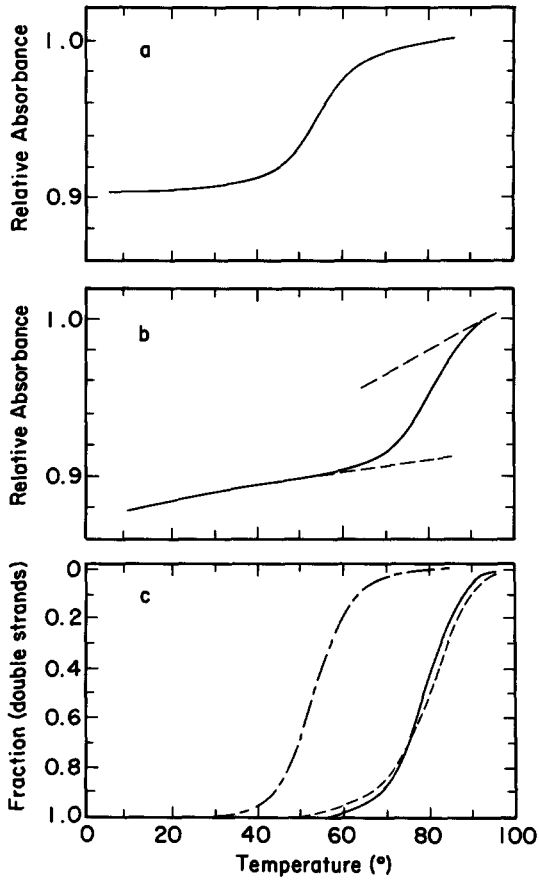

FIG. 2. (a) A melting curve with well-defined base lines; 5′-GCGAUUUCUGACCGC-3′ in 50 m*M* NaCl, 10 m*M* sodium phosphate, and 0.1 m*M* EDTA, pH 6.5. Normalized absorbance at 260 nm is plotted versus temperature. (b) A melting curve with undefined upper base line. 5′GGGAGUUUCCGCUCCC-3′ in 1 *M* NaCl, 10 m*M* sodium phosphate, and 0.1 m*M* 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 . . . is

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

TABLE I
EXTINCTION COEFFICIENTS FOR NUCLEOTIDES AND DINUCLEOSIDE PHOSPHATES TO CALCULATE EXTINCTION COEFFICIENTS OF SINGLE STRANDS[a]

Nucleotide or dinucleoside phosphate	$\varepsilon(260)$ M^{-1} cm^{-1} $\times 10^{-3}$		Nucleotide or dinucleoside phosphate	$\varepsilon(260)$ M^{-1} cm^{-1} $\times 10^{-3}$	
	RNA	DNA		RNA	DNA
Ap	15.34	15.34	CpG	9.39	9.39
Cp	7.60	7.60	CpU(CpT)	8.37	7.66
Gp	12.16	12.16	GPA	12.92	12.92
Up(Tp)	10.21	8.70	GpC	9.19	9.19
ApA	13.65	13.65	GpG	11.43	11.43
APC	10.67	10.67	GpU(GpT)	10.96	10.22
ApG	12.79	12.79	UpA(TpA)	12.52	11.78
ApU(ApT)	12.14	11.42	UpC(TpC)	8.90	8.15
CpA	10.67	10.67	UpG(TpG)	10.40	9.70
CpC	7.52	7.52	UpU(TpT)	10.11	8.61

[a] At 260 nm, 25°, 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 ε(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 ε(Cp) . . . 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°. 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° 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 II
TWO-STATE ANALYSIS OF NUCLEIC ACID TRANSITIONS

Reaction type	Equilibrium constants	ΔH° from slope of f versus T	Concentration dependence of T_m
Monomolecular S = H	$K = \frac{[H]}{[S]} = \frac{(f)}{(1-f)}$	$\Delta H^\circ = 4RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	
Bimolecular (self-complementary) 2S = D	$K = \frac{[D]}{[S]^2} = \frac{f}{2(1-f)^2 c_t}$	$\Delta H^\circ = 6RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	$\frac{1}{T_m} = \frac{R}{\Delta H^\circ}\ln c_t + \frac{\Delta S^\circ}{\Delta H^\circ}$
Bimolecular (non-self-complementary) $S_A + S_B = D$	$K = \frac{[D]}{[S_A][S_B]} = \frac{2f}{(1-f)^2 c_t}$	$\Delta H^\circ = 6RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	$\frac{1}{T_m} = \frac{R}{\Delta H^\circ}\ln c_t + \frac{(\Delta S^\circ - R\ln 4)}{\Delta H^\circ}$
Trimolecular (identical strands) 3S = T	$K = \frac{[T]}{[S]^3} = \frac{f}{3c_t^2(1-f)^3}$	$\Delta H^\circ = 8RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	$\frac{1}{T_m} = \frac{2R}{\Delta H^\circ}\ln c_t + \frac{[\Delta S^\circ - R\ln 4/3)]}{\Delta H^\circ}$
Trimolecular (nonidentical strands) $S_A + S_B + S_C = T$	$K = \frac{[T]}{[S_A][S_B][S_C]} = \frac{9f}{c_t^2(1-f)^3}$	$\Delta H^\circ = 8RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	$\frac{1}{T_m} = \frac{2R}{\Delta H^\circ}\ln c_t + \frac{(\Delta S^\circ - 2R\ln 6)}{\Delta H^\circ}$
Tetramolecular (identical strands) 4S = Q	$K = \frac{[Q]}{[S]^4} = \frac{f}{4c_t^3(1-f)^4}$	$\Delta H^\circ = 10RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	$\frac{1}{T_m} = \frac{3R}{\Delta H^\circ}\ln c_t + \frac{(\Delta S^\circ\ R\ln 2)}{\Delta H^\circ}$
Tetramolecular (nonidentical strands) $S_A + S_B + S_C + S_D = Q$	$K = \frac{[Q]}{[S_A][S_B][S_C][S_D]} = \frac{64f}{c_t^3(1-f)^4}$	$\Delta H^\circ = 10RT_m^2\left(\frac{df}{dT}\right)_{T=T_m}$	$\frac{1}{T_m} = \frac{3R}{\Delta H^\circ}\ln c_t + \frac{(\Delta S^\circ - 3R\ln 8)}{\Delta H^\circ}$

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[7] or determining the size of the constituent species by electrophoresis[8] 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[9] 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,

$$A = fA_D + (1 - f)A_S \tag{4}$$

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[10] and experimental data[11] show that per-

[7] C. Stevens and G. Felsenfeld, *Biopolymers* **2,** 293 (1964).
[8] N. R. Kallenbach, R.-I. Ma, and N. Seeman, *Nature (London)* **305,** 829 (1983).
[9] D. Poland and H. A. Scheraga, "Theory of Helix-Coil Transitions in Biopolymers." Academic Press, New York, 1970.
[10] I. Tinoco, Jr., *J. Am. Chem. Soc.* **82,** 4785 (1960).
[11] A. Rich, and I. Tinoco, Jr., *J. Am. Chem. Soc.* **82,** 6409 (1960).

centage of hypochromicity

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

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 ($\Delta H°$, $\Delta S°$, and $\Delta G°$), 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.[12,13] 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.[14] 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 $\Delta H°$ 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).[15] 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

$$d \ln K/d(1/T) = -\Delta H°/R \tag{5}$$

and standard thermodynamic equations

$$\Delta G° = -RT \ln K \tag{6}$$

$$\Delta S° = (\Delta H° - \Delta G°)/T \tag{7}$$

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

[12] J. Applequist and V. Damle, *J. Am. Chem. Soc.* **87**, 450 (1965).

[13] C. R. Cantor and P. R. Schimmel, "Biophysical Chemistry," Vol. 3. Freeman, San Francisco, California, 1980.

[14] K. J. Breslauer, J. M. Sturtevant, and I. Tinoco, Jr., *J. Mol. Biol.* **99**, 549 (1975).

[15] L. A. Marky, N. R. Kallenbach, K. A. McDonough, N. C. Seeman, and K. J. Breslauer, *Biopolymers* **26**, 1621 (1987).

For longer base-paired sequences (>12 bp), the helix-to-coil transition is usually not two state.[14] RNA molecules with complex structures often melt in stages,[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.[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.[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.

$$\varepsilon_D = m_D T + b_D \tag{8}$$

$$\varepsilon_S = m_S T + b_S \tag{9}$$

If the base lines are well defined (10–15° of linear absorbance versus T) as in Fig. 2a, ε_D and ε_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

[16] P. E. Cole, S. K. Yang, and D. M. Crothers, *Biochemistry* **11**, 4358 (1972).

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 $(-\Delta H^\circ/R)$ as the slope and $(\Delta S^\circ/R)$ as the intercept; this is a van't Hoff plot. If ΔH° is independent of temperature, the plot should be linear. A nonlinear van't Hoff plot can result from several factors: temperature dependence of ΔH°, poor choice of base lines, or a non-two-state transition.[17] The most likely candidate is a poor choice of base lines.[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.[18] ΔH° 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 ΔH° at any value of f have been derived by Gralla and Crothers[19] and discussed by Marky and Breslauer.[20] Thus, ΔH° 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, ΔH° 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 ΔH° determined using the van't Hoff plots and the derivative method. If the results using both methods do not agree to within ~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.

[17] M. Petersheim and D. H. Turner, *Biochemistry* **22,** 256 (1983).
[18] P. R. Bevington, "Data Reduction and Error Analysis for the Physical Sciences." McGraw-Hill, New York, 1969.
[19] J. Gralla and D. M. Crothers, *J. Mol. Biol.* **73,** 497 (1973).
[20] L. A. Marky and K. J. Breslauer, *Biopolymers* **26,** 1601 (1987).

Method 3: A versus T. The experimental absorbance versus temperature curve is fit directly to six parameters: ΔH°, ΔS°, 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

$$\begin{aligned} A &= fA_D + (1 - f)A_S \\ A_D &= \varepsilon_D c_t l = (m_D T + b_D)c_t l \\ A_S &= \varepsilon_S c_t l = (m_S T + b_S)c_t l \\ K &= \exp(-\Delta H^\circ/R + \Delta S^\circ/RT) \end{aligned} \tag{10}$$

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 ΔH° and ΔS°. 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 ΔH° and ΔS°. 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 ΔH° and ΔS° 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 ΔH° values deter-

[21] J. W. Nelson, F. H . Martin, and I. Tinoco, Jr., *Biopolymers* **20,** 2509 (1981).

[22] D. D. Albergo, L. A. Marky, K. J. Breslauer, and D. H. Turner, *Biochemistry* **20,** 1409 (1981).

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

Parameter	Method 1 (ln K versus $1/T$)	Method 2 (slope at T_m)	Method 3 (curve fit)	Method 4 (ln c_t versus $1/T_m$)
No base lines subtracted	−39.0	−42.1	−33.0	−58.1
Upper base line subtracted	−37.0	−41.4	—	−57.6
Both base lines subtracted	−58.9	−56.8	−56.6	−57.4

[a] 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_t 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 ΔH°, ΔS°, and ΔG° 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 ΔS and ΔG depend on concentration even for ideal solutions. ΔH 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.]. If

the transition is indeed two state, the thermodynamic data should be independent of the measuring technique. Fourth, the ΔH° and ΔS° values, derived from this analysis, are assumed to be independent of temperature. Since

$$d\,\Delta H/dT = \Delta C_p \tag{11}$$

where ΔC_p is the difference of heat capacities (at constant pressure) of the native and melted states, this assumption can be checked by plotting ΔH° values obtained using methods 1, 2, or 3 versus T_m; the slope is ΔC_p. The errors introduced by this assumption are small, but Petersheim and Turner[17] have introduced a simple correction for ΔH° 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 ΔG° 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.[23] It is assumed that the free energy for a secondary structure is the sum of the free energies for the separate regions.[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

```
5′          3′
  GCGUGAU
  CGCACUA
3′          5′
```

[23] D. H. Turner, N. Sugimoto, and S. M. Freier, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 167 (1988).

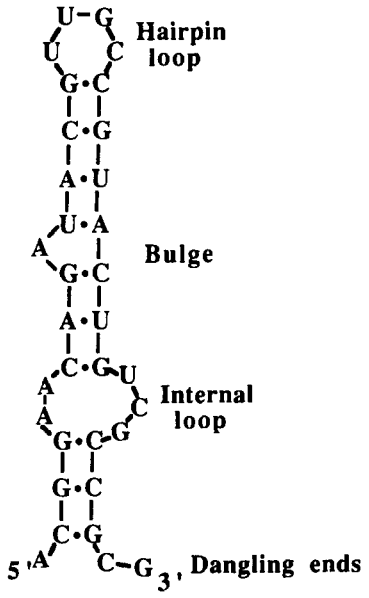

FIG. 3. A representation of the various types of secondary structures that a RNA molecule can adopt.

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

$$\Delta G^{\circ}_{\text{total}} = \Delta G^{\circ}\begin{matrix}\overrightarrow{GC}\\ \underleftarrow{CG}\end{matrix} + \Delta G^{\circ}\begin{matrix}\overrightarrow{CG}\\ \underleftarrow{GC}\end{matrix} + \Delta G^{\circ}\begin{matrix}\overrightarrow{GU}\\ \underleftarrow{CA}\end{matrix} + \Delta G^{\circ}\begin{matrix}\overrightarrow{UG}\\ \underleftarrow{AC}\end{matrix}$$

$$+ \Delta G^{\circ}\begin{matrix}\overrightarrow{GA}\\ \underleftarrow{CU}\end{matrix} + \Delta G^{\circ}\begin{matrix}\overrightarrow{AU}\\ \underleftarrow{UA}\end{matrix} + \Delta G^{\circ}_{\text{sym}}\,(=0) + \Delta G^{\circ}_{\text{init}}$$

ΔS° and ΔH° 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.[1] 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.*,[1] 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 nonbonded loops and bulges are simply added to the free energy for the duplex region. Duplex initiation free energies are contained in the ΔG° 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[1]; these data assume that ΔG° 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.

[24] D. R. Groebe and O. C. Uhlenbeck, *Nucleic Acids Res.* **16,** 11725 (1988).

[25] C. Tuerk, P. Gauss, C. Thermes, D. R. Groebe, N. Guild, G. Stormo, M. Gayle, Y. d'Auberton-Carafa, O. C. Uhlenbeck, I. Tinoco, Jr., E. N. Brody, and L. Gold, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 1364 (1988).

[26] D. R. Groebe and O. C. Uhlenbeck, *Biochemistry* **28,** 742 (1989).

Other Techniques

Temperature-Jump Methods. Derivatives of melting curves can be obtained directly by using a temperature-jump apparatus.[19] 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.*[27] and Riesner *et al.*[28] 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.[29]

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

$$\delta_{obs} = f\delta_{native} + (1 - f)\delta_{coil} \tag{12}$$

where δ_{obs} is the observed chemical shift of a given proton, δ_{native} and δ_{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.[27] 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

[27] D. M. Crothers, P. E. Cole, C. W. Hilbers, and R. G. Shulman, *J. Mol. Biol.* **87,** 63 (1974).
[28] D. Riesner, G. Maass, R. Thiebe, P. Philippsen, and H. G. Zachau, *Eur. J. Biochem.* **36,** 76 (1973).
[29] A. Wada, S. Yabuki, and Y. Husimi, *CRC Crit. Rev. Biochem.* **9,** 87 (1980).

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

$$k_{ex} \gg 2\pi\ \Delta\nu \tag{13}$$

where k_{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 ($k_{ex} \cong 2\pi\ \Delta\nu$), resulting in line broadening, or slow exchange ($k_{ex} \ll 2\pi\ \Delta\nu$), 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 ΔH directly; the method of differential scanning calorimetry (DSC)[30,31] is most often used. In DSC, the excess heat capacity of the transition (ΔC_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 DCS over optical melting is that no assumptions about the transition need to be made; DSC values of ΔH are model independent.[20] However, it is important to remember that the ΔH 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, $\Delta H°_{VH}$, 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 ΔH 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

[30] L.-H. Chang, S.-J. Li, T. L. Ricca, and A. G. Marshall, *Anal. Chem.* **56,** 1502 (1984).
[31] P. L. Privalov, *FEBS Lett.* **40,** S140 (1974).

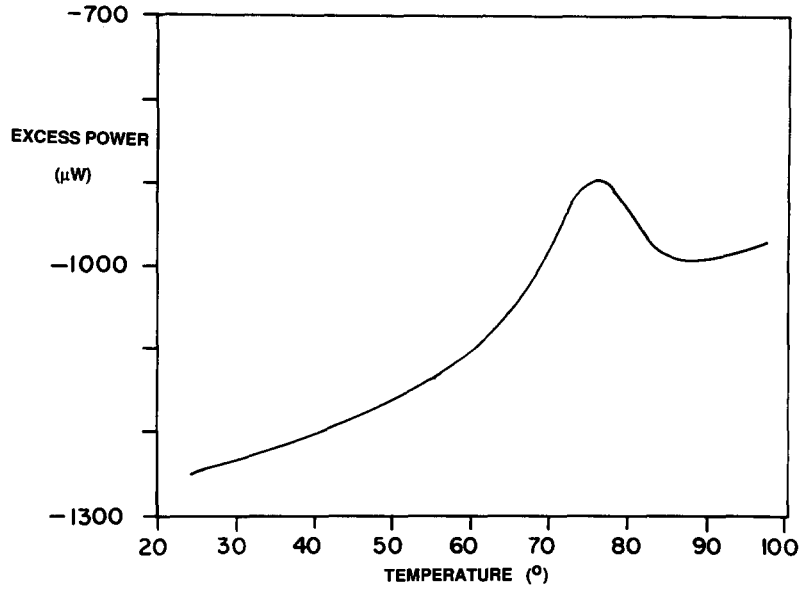

FIG. 4. A sample DSC trace for d(CGCGAATTCGCG)$_2$ in 1 *M* NaCl obtained on a Hart Scientific DSC. The vertical axis is the excess power required to maintain the same temperature in the sample and reference cells, which is proportional to the excess heat capacity due to the nucleic acid, and the horizontal axis is temperature. The data represent raw data before base-line correction. [Data courtesy of Dr. D. Kallick (Department of Chemistry, University of California, Berkeley).]

in the van't Hoff methods. However, comparison of ΔH from an optical melting experiment to ΔH from calorimetry can be very helpful. For example, if $\Delta H_{VH} < \Delta H_{CAL}$, 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 hypochromicities at 260 and 280 nm will allow an estimate of A · U and G · 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 ΔH° for each transition can be determined from the differentiated melting curve, assuming the transition is independent of others.

Acknowledgments

We are very greatful 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.

[23] Nuclease Digestion: A Method for Mapping Introns

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,