of characterizing these higher-order structures, a definition that can be accomplished by using the equations and protocols presented in this chapter. Recent studies on lesion-containing duplexes[55,56] illustrate how one also can use the equations presented here to combine both optical and calorimetric data in a manner that takes advantage of the strengths of both methods while avoiding a systematic error in the van't Hoff analysis which is observed with this important class of molecules.[56]

In summary, an examination of the current literature reveals that numerous researchers are focusing on oligomeric modeling of biologically relevant structures. The equations and protocols described in this chapter will allow these researchers to characterize thermodynamically these new structural forms by analyzing equilibrium melting curves.

Acknowledgment

This work was supported by National Institutes of Health Grants GM23509, GM34469, and CA 47795, the Charles and Johanna Busch Memorial Fund, the Research Corporation, and the Rutgers Research Council.

[55] G. E. Plum and K. J. Breslauer, *Ann. NY Acad. Sci.* **726,** 45–56 (1994).
[56] D. S. Pilch, G. E. Plum, and K. J. Breslauer, *Current Biology* (1995), in press.

[11] Predicting Thermodynamic Properties of RNA

By MARTIN J. SERRA and DOUGLAS H. TURNER

Biological molecules must fold into the correct three-dimensional shape to acquire their active, functional form. The folding of biological molecules is directed by the physical and chemical properties inherent in their molecular makeup. The blueprint for the construction of biological molecules is contained in the genetic material of an organism. It is possible to sequence the genetic information and determine the message present in the genes that code for the covalent structure (primary structure) of biological molecules. Since the advent of rapid methods for sequencing DNA,[1,2] there has been an exponential growth in nucleic acid sequence information. This progress will continue into the foreseeable future with the sequencing of model organisms and the 3 billion base pairs (bp) of the human genome as part

[1] A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74,** 560 (1977).
[2] F. Sanger and A. R. Coulson, *J. Mol. Biol.* **94,** 441 (1975).

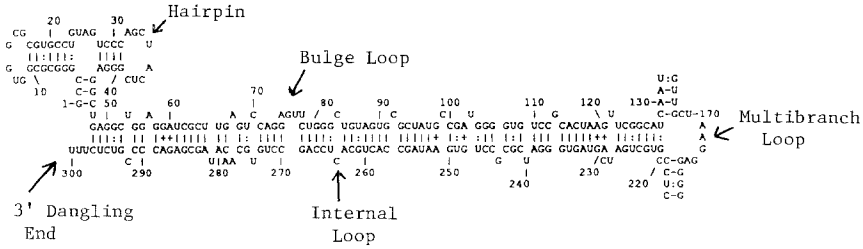


FIG. 1. Common RNA secondary structure motifs. Portion of human Signal Recognition Particle SRP-RNAB. (From Zweib.[2a]).

of the Human Genome Initiative. To make full use of this information, it will be necessary to determine the secondary and tertiary structures of the biological molecules encoded by this information.

The rules governing the folding of biological molecules into their three-dimensional structures are just beginning to be explored. RNA is an excellent candidate for determining such rules and relating them to molecular forces. The first step toward predicting the three-dimensional structure is to predict secondary structure. RNA secondary structure consists of a small number of structural motifs, illustrated in Fig. 1.[2a] RNA is also composed of a limited molecular vocabulary, the four nucleotides (adenosine, cytidine, guanosine, and uridine). The strong interactions between the nucleotides make it possible to design small model systems that fold spontaneously, permitting studies of the thermodynamics of folding. The results of such studies can then be used to predict the thermodynamics for folding other sequences. Here, we summarize these results and illustrate some typical applications.

Methods

The main methods for determining thermodynamic parameters for nucleic acid structural stability are calorimetry and temperature-dependent optical spectroscopy (melting curves). Calorimetric methods directly measure the heat lost or gained due to a structural change. Differential scanning calorimetry is the most commonly used calorimetric method for detecting the energy change associated with conformational changes in nucleic acids (see [9] in this volume).

The transition between the ordered conformation and the disrupted coil can also be easily detected by ultraviolet absorbance changes because loss of base pairing and stacking in RNA typically leads to absorbance increases. The most convenient method for converting between the ordered

[2a] C. Zweib, *Prog. Nucleic Acid Res. Mol. Biol.* **37,** 207 (1989).

TABLE I
THERMODYNAMIC VALUES FOR BASE PAIRS, TERMINAL MISMATCHES AND G·U PAIRS[a,b]

Base pair	$X\downarrow$	$Y\rightarrow$	A	C	G	U
			ΔH° (kcal/mol)			
$\overrightarrow{AX}$	A		−3.9	(+2.0)	(−3.5)	−6.6
$\underleftarrow{UY}$	C		−2.3	(+6.0)	−10.2	(−0.3)
	G		−3.1	−7.6	(−3.5)	−2.7
	U		−5.7	(+4.6)	−4.0	(−1.7)
$\overrightarrow{CX}$	A		−9.1	−5.6	−5.6	−10.5
$\underleftarrow{GY}$	C		(−5.7)	(−3.4)	−12.2	−2.7
	G		−8.2	−8.0	−9.2	−3.1
	U		−7.6	−5.3	−11.2	−8.6
$\overrightarrow{GX}$	A		−5.2	−4.0	−5.6	−13.3
$\underleftarrow{CY}$	C		−7.2	(+0.5)	−14.2	(−4.2)
	G		−7.1	−12.2	−6.2	−6.3
	U		−10.2	(−0.3)	−9.5	(−5.0)
$\overrightarrow{UX}$	A		−4.0	−6.3	−8.9	−8.1
$\underleftarrow{AY}$	C		−4.3	−5.1	−13.3	−1.8
	G		−3.8	−10.5	−8.9	−8.5
	U		−6.6	−1.4	−13.6	+1.4
			ΔS° (eu)			
$\overrightarrow{AX}$	A		−10.2	(+9.6)	(−8.7)	−18.4
$\underleftarrow{UY}$	C		−5.3	(+21.6)	−26.2	(+1.5)
	G		−7.3	−19.2	(−8.7)	−7.0
	U		−15.5	(+17.4)	−9.7	(−2.7)
$\overrightarrow{CX}$	A		−24.5	−13.5	−13.4	−27.8
$\underleftarrow{GY}$	C		(−15.2)	(−7.6)	−29.7	−6.3
	G		−21.8	−19.4	−24.6	−6.2
	U		−19.2	−12.6	−30.1	−23.9
$\overrightarrow{GX}$	A		−13.2	(−8.2)	−13.9	−35.5
$\underleftarrow{CY}$	C		−19.6	(+3.9)	−34.9	(−12.2)
	G		−17.8	−29.7	−15.1	−15.8
	U		−26.2	(+2.1)	−23.9	(−14.0)
$\overrightarrow{UX}$	A		−9.7	−17.7	−25.2	−22.6
$\underleftarrow{AY}$	C		−11.6	−14.6	−35.5	−4.2
	G		−8.5	−27.8	−25.0	−24.9
	U		−18.4	−2.5	−40.2	+6.0

TABLE I (*continued*)

Base pair	$X\downarrow$	$Y\rightarrow$	A	C	G	U
			ΔG°_{37} (kcal/mol)			
$\overrightarrow{AX}$	A		−0.8	(−1.0)	(−0.8)	−0.9
$\underleftarrow{UY}$	C		−0.6	(−0.7)	−2.1	(−0.7)
	G		−0.8	−1.7	(−0.8)	−0.5
	U		−0.9	(−0.8)	−1.0	(−0.8)
$\overrightarrow{CX}$	A		−1.5	−1.5	−1.4	−1.8
$\underleftarrow{GY}$	C		(−1.0)	(−1.1)	−2.9	−0.8
	G		−1.4	−2.0	−1.6	−1.2
	U		−1.7	−1.4	−1.9	−1.2
$\overrightarrow{GX}$	A		−1.1	(−1.5)	−1.3	−2.3
$\underleftarrow{CY}$	C		−1.1	(−0.7)	−3.4	(−0.5)
	G		−1.6	−2.9	−1.4	−1.4
	U		−2.1	(−1.0)	−2.1	(−0.7)
$\overrightarrow{UX}$	A		−1.0	−0.8	−1.1	−1.1
$\underleftarrow{AY}$	C		−0.7	−0.6	−2.3	−0.5
	G		−1.1	−1.8	−1.2	−0.8
	U		−0.9	−0.6	−1.1	−0.5

[a] Not listed are parameters for consecutive GU mismatches. The ΔH° (kcal/mol), ΔS° (eu), and ΔG° (kcal/mol), respectively, are as follows: $\begin{matrix}5'\,UG\,3'\\GU\end{matrix}$ (−11.0, −34.8, −0.2), $\begin{matrix}5'\,UU\,3'\\GG\end{matrix}$ (−20.0, −63.1, −0.4), $\begin{matrix}GU\\UG\end{matrix}$ (−9.3, −35.1, +1.5 or −18.3, −55.9, −1.0 for the context $\begin{matrix}GGUC\\CUGG\end{matrix}$). Most parameters result from single measurements and may not be general. Parameters in parentheses are estimated by (1) summing parameters for the constituent 5′ and 3′ dangling ends for AC and pyrimidine–pyrimidine mismatches, or (2) averaging measured AA and AG mismatches with the same adjacent base pair for purine–purine mismatches. Parameters for non-hydrogen-bonded Watson–Crick and GU pairs may be estimated similarly. Differences with previously published values and with the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ are due to round off errors. For initiation of a duplex from two strands, $\Delta H^\circ_i = 0$ kcal/mol, $\Delta S^\circ_i = -10.8$ eu, $\Delta G^\circ_{37} = 3.4$ kcal/mol. For self-complementary duplexes, add $\Delta S^\circ_{sym} = -1.4$ eu, $\Delta G^\circ_{37sym} = 0.4$ kcal/mol.

[b] Data from Refs. 5, 6, 9, 12–14a.

and coil forms of RNA is heating. As the temperature of the sample increases, the thermal energy denatures the RNA and the absorbance changes. The use of absorbance melting curves to derive thermodynamic parameters for RNA folding has been described in detail.[3,4]

[3] M. Petersheim and D. H. Turner, *Biochemistry* **22,** 256 (1983).

[4] J. D. Puglisi and I. Tinoco, Jr., this series, Vol. 180, p. 304.

Prediction of Thermodynamic Stability

Double Helix

The double helix is the main structural motif for RNA. Known secondary structures are at least 50% helical. Much of the initial characterization of RNA stability focused on duplexes containing Watson–Crick base pairs.[5] GU wobble base pairs also occur frequently in RNA and have stabilities similar to those of AU pairs.[6] Therefore, they may be considered as a normal part of the double-helical motif.

The base composition of an RNA molecule greatly influences its stability. Composition alone, however, is not sufficient to predict the stability of an RNA duplex. For example, $\begin{matrix}\text{GGCC}\\ \text{CCGG}\end{matrix}$ (top line is 5′ → 3′) and $\begin{matrix}\text{CGCG}\\ \text{GCGC}\end{matrix}$ have different stabilities.[7] The most popular model for predicting the stability of RNA is the nearest-neighbor model.[8] This model assumes that the stability of a given base pair depends on the identity of its adjacent base pair (nearest neighbor). For example, $\begin{matrix}\text{GC}\\ \text{CG}\end{matrix}$, $\begin{matrix}\text{CG}\\ \text{GC}\end{matrix}$, and $\begin{matrix}\text{GG}\\ \text{CC}\end{matrix}$ have different stabilities. Therefore, $\begin{matrix}\text{GGCC}\\ \text{CCGG}\end{matrix}$ and $\begin{matrix}\text{CGCG}\\ \text{GCGC}\end{matrix}$, containing different nearest-neighbor interactions, are expected to have different stabilities. For perfectly base-paired regions, the nearest neighbor model can usually predict within 10% the free energy change for duplex formation, ΔG°_{37}. Here $\Delta G^\circ_{37} = -R(273.15 + 37)\ln K_{eq}$ and K_{eq} is the equilibrium constant for duplex formation at 37°.

Thermodynamic parameters for all possible double-helical nearest neighbors of Watson–Crick and GU pairs have been determined.[5,6] Values for the free energy change (ΔG°_{37}), enthalpy change (ΔH°), and entropy change (ΔS°) are given in Table I. The thermodynamic parameters for a helical region are the sum of the individual nearest-neighbor values, plus values for initiation of the helix and a symmetry correction for duplexes composed of complementary strands. Calculations of thermodynamic parameters for the non-self-complementary duplex $\begin{matrix}\text{CGGC}\\ \text{GCCG}\end{matrix}$ are shown in Fig. 2. The thermodynamic parameters (ΔG°_{37}, ΔH°, and ΔS°) are predicted

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

[6] L. He, R. Kierzek, J. SantaLucia, Jr., A. E. Walter, and D. H. Turner, *Biochemistry* **30,** 11124 (1991).

[7] S. M. Freier, A. Sinclair, T. Neilson, and D. H. Turner, *J. Mol. Biol.* **185,** 645 (1985).

[8] 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).

Double Helix: Non-Self-Complementary

```
       5' CGGC
          ••••
       3' GCCG
```

$$\Delta H^\circ = \Delta H^\circ_{\substack{CG\\GC}} + \Delta H^\circ_{\substack{GG\\CC}} + \Delta H^\circ_{\substack{GC\\CG}}$$

$$\Delta H^\circ = (-8.0)+(-12.2)+(-14.2) = -34.4 \text{ kcal/mol}$$

$$\Delta S^\circ = \Delta S^\circ_{\substack{CG\\GC}} + \Delta S^\circ_{\substack{GG\\CC}} + \Delta S^\circ_{\substack{GC\\CG}} + \Delta S^\circ_{int}$$

$$\Delta S^\circ = (-19.4)+(-29.7)+(-34.9)+(-10.8) = -94.8 \text{ eu}$$

$$\Delta G^\circ_{37} = \Delta H^\circ - T\Delta S^\circ = -34.4 - (310.15)(-94.8)/1000 = -5.0 \text{ kcal/mol}$$

$$\Delta G^\circ_{37} = \Delta G^\circ_{37\,\substack{CG\\GC}} + \Delta G^\circ_{37\,\substack{GG\\CC}} + \Delta G^\circ_{37\,\substack{GC\\CG}} + \Delta G^\circ_{37int}$$

$$\Delta G^\circ_{37} = (-2.0)+(-2.9)+(-3.4)+3.4 = -4.9 \text{ kcal/mol}$$

Note: The value of ΔG°_{37} depends on how it is calculated due to round off errors.

$$T_M(^\circ C) = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln(C_T/4)} - 273.15 \quad \text{(non self-complementary duplex)}$$

$$T_M(\text{predicted } (10^{-4} \text{ M})) = \frac{-34400}{-94.8+1.987 \ln(10^{-4}/4)} - 273.15 = 23.8^\circ C$$

Double Helix: Self-Complementary

```
       5' GGAUCC
          ••••••
       3' CCUAGG
```

$$\Delta H^\circ = \Delta H^\circ_{\substack{GG\\CC}} + \Delta H^\circ_{\substack{GA\\CU}} + \Delta H^\circ_{\substack{AU\\UA}} + \Delta H^\circ_{\substack{UC\\AG}} + \Delta H^\circ_{\substack{CC\\GG}}$$

$$\Delta H^\circ = (-12.2)+(-13.3)+(-5.7)+(-13.3)+(-12.2) = -56.7 \text{ kcal/mol}$$

$$\Delta S^\circ = \Delta S^\circ_{\substack{GG\\CC}} + \Delta S^\circ_{\substack{GA\\CU}} + \Delta S^\circ_{\substack{AU\\UA}} + \Delta S^\circ_{\substack{UC\\AG}} + \Delta S^\circ_{\substack{CC\\GG}} + \Delta S^\circ_{int} + \Delta S^\circ_{sym}$$

FIG. 2. Predictions of thermodynamic properties for double helices using nearest-neighbor model.

from the parameters for the three nearest-neighbor interactions and the initiation contribution. Calculation of thermodynamic parameters for the symmetric duplex, $\begin{matrix}GGAUCC\\CCUAGG\end{matrix}$, are done in an analogous manner, also illustrated in Fig. 2. The thermodynamic parameters are predicted from the five nearest-neighbor interactions, initiation, and the symmetry contribution.

The enthalpy and entropy changes, ΔH° and ΔS°, are considered to be temperature independent and allow prediction of the free energy of duplex

$\Delta S^\circ = (-29.7)+(-35.5)+(-15.5)+(-35.5)+(-29.7)+(-10.8)+(-1.4)$

$= -158.1$ eu

$\Delta G^\circ_{37} = \Delta H^\circ - T\Delta S^\circ = -56.7-(310.15)(-158.1) / 1000 = -7.7$ kcal/mol

$$\Delta G^\circ_{37} = \Delta G^\circ_{37}{}^{GG}_{CC} + \Delta G^\circ_{37}{}^{GA}_{CU} + \Delta G^\circ_{37}{}^{AU}_{UA} + \Delta G^\circ_{37}{}^{UC}_{AG} + \Delta G^\circ_{37}{}^{CC}_{GG} + \Delta G^\circ_{37}\text{int} + \Delta G^\circ_{37}\text{sym}$$

$\Delta G^\circ_{37} = (-2.9)+(-2.3)+(-0.9)+(-2.3)+(-2.9)+3.4+0.4 = -7.5$ kcal/mol

$$T_M(^\circ C) = \frac{\Delta H^\circ}{\Delta S^\circ + R\ln(C_T)} - 273.15 \quad \text{(self-complementary duplex)}$$

$$T_M(\text{predicted } (10^{-4}\text{M})) = \frac{-56700}{-158.1+1.987 \ln 10^{-4}} - 273.15 = 48.3^\circ C$$

3' Terminal Unpaired Nucleotides

```
5'  GGAUCCA
    ......
3' ACCUAGG
```

$$\Delta H^\circ = \Delta H^\circ \text{ (Core Duplex)} + 2\Delta H^\circ{}^{CA}_{G} = -56.7 + 2(-9.0) = -74.7 \text{ kcal/mol}$$

$$\Delta S^\circ = \Delta S^\circ \text{ (Core Duplex)} + 2\Delta S^\circ{}^{CA}_{G} = -158.1 + 2(-23.4) = -204.9 \text{ eu}$$

$$\Delta G^\circ_{37} = \Delta H^\circ - T\Delta S^\circ = -74.7 - 310.15\ (-204.9)/1000 = -11.2 \text{ kcal/mol}$$

$$\Delta G^\circ_{37} = \Delta G^\circ_{37} \text{ (Core Duplex)} + 2\Delta G^\circ_{37}{}^{CA}_{G} = -7.5 + 2(-1.7) = -10.9 \text{ kcal/mol}$$

$$T_M(\text{predicted } (10^{-4}\text{M})) = \frac{-74700}{-204.9 + R \ln 10^{-4}} - 273.15 = 61.5^\circ C$$

5' Terminal Unpaired Nucleotides

```
5' AGGAUCC
   ......
   CCUAGGA
```

$$\Delta H^\circ = \Delta H^\circ \text{ (Core Duplex)} + 2\Delta H^\circ{}^{AG}_{C}$$

$\Delta H^\circ = -56.7 + 2(-1.6) = -59.9$ kcal/mol

FIG. 2. (*continued*)

formation at any temperature: $\Delta G^\circ_T = \Delta H^\circ - T\Delta S^\circ$. This allows calculation of the equilibrium constant for the reaction: $K = e^{-\Delta G^\circ/RT}$. Additionally, the ΔH° and ΔS° allow the melting temperature, T_M, for the duplex to be calculated:

$$T_M(^\circ C) = \frac{\Delta H^\circ}{\Delta S^\circ + R\ln(C_T)} - 273.15 \quad \text{(self-complementary)} \qquad (1)$$

$\Delta S^\circ = \Delta S^\circ$ (Core Duplex) + $2\Delta S^\circ_{C}^{AG}$

$\Delta S^\circ = -158.1 + 2(-4.5) = -167.1$ eu

$\Delta G^\circ_{37} = \Delta H^\circ - T\Delta S^\circ = -59.9 - 310.15(-167.1)/1000 = -8.1$ kcal/mol

$\Delta G^\circ_{37} = \Delta G^\circ_{37}$ (Core Duplex) + $2\Delta G^\circ_{37}{}^{AG}_{C}$

$\Delta G^\circ_{37} = -7.5 + 2(-0.2) = -7.9$ kcal/mol

$$T_M(^\circ C(1\times10^{-4}\text{m})) = \frac{-59900}{-167.1 + R \ln 10^{-4}} - 273.15 = 49.9^\circ C$$

Terminal Mismatch

```
          5' AGGAUCCA
             ******
          3' ACCUAGGA
```

$\Delta H^\circ = \Delta H^\circ$ (Core Duplex) + $2\Delta H^\circ_{GA}^{CA}$

$\Delta H^\circ = -56.7 + 2(-9.1) = -74.9$ kcal/mol

$\Delta S^\circ = \Delta S^\circ$ (Core Duplex) + $2\Delta S^\circ_{GA}^{CA}$

$\Delta S^\circ = -158.1 + 2(-24.5) = -207.1$ eu

$\Delta G^\circ_{37} = \Delta H^\circ - T\Delta S^\circ = -74.9 - 310.15(-207.1)/1000 = -10.7$ kcal/mol

$\Delta G^\circ_{37} = \Delta G^\circ_{37}$ (Core Duplex) + $2\Delta G^\circ_{37}{}^{CA}_{GA}$

$\Delta G^\circ_{37} = -7.5 + 2(-1.5) = -10.5$ kcal/mol

$$T_M(\text{predicted } (10^{-4}\text{M})) = \frac{-74900}{-207.1 + R \ln 10^{-4}} - 273.15 = 59.1^\circ C$$

FIG. 2. (*continued*)

$$T_M(^\circ C) = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln(C_T/4)} - 273.15 \quad \text{(non-self-complementary, each strand at concentration } C_T/2) \qquad (2)$$

Terminal Unpaired Nucleotides

Most helical regions in RNA tend to be short, less than 8 bp. Thus unpaired nucleotides adjacent to helical regions can have a significant effect on stability. These effects have also been studied in model systems and the thermodynamic values for unpaired terminal nucleotides are given in Table

II.[9] Comparison of Tables I and II indicates that some 3′ terminal nucleotides can stabilize a helix as much as some base pairs. This is an indication of the importance of stacking interactions to the stability of an RNA helix. Unpaired nucleotides at the 5′ end of the helix contribute little to the stabilization of the helix. This has been attributed to the fact that in A-form secondary structure, the 5′ nucleotide does not stack on the helix.[10] The prediction of thermodynamic parameters for a helix with a 3′ or 5′ terminal dangling end is determined in a manner analogous to the method described above, treating the dangling end as an additional nearest-neighbor interaction. For example, $\begin{matrix} \text{GGAUCCA} \\ \text{ACCUAGG} \end{matrix}$ would be $2 \times 1.7 = 3.4$ kcal/mol more stable than the parent helix at 37° (see Fig. 2). As noted, the 3′ dangling A contributes nearly the same stabilization as an AU base pair. The enthalpy and entropy changes for this duplex formation would increase by −18 kcal/mol and −46.8 eu, respectively, relative to the core fully base-paired duplex.

The presence of the terminal adenosines on the 5′ ends of $\begin{matrix} \text{AGGAUCC} \\ \text{CCUAGGA} \end{matrix}$ only makes the helix more stable by $2 \times 0.2 = 0.4$ kcal/mol at 37° (see Fig. 2). Likewise, the enthalpy increases by −3.2 kcal/mol and the entropy increases by −9.0 eu, relative to the core duplex.

Terminal Mismatches

Terminal mismatches contain both a 3′ and 5′ dangling end. The thermodynamic values for the nearest neighbor interactions of terminal mismatches are presented in Table I.[11-14a] Comparison of Tables I and II indicates that for terminal mismatches, the stability increments are equal to or less than the sum of increments for the constitutive dangling ends. Thus, nucleosides in terminal mismatches do not interact synergistically to increase helix stability further. The prediction of the thermodynamic stability for a helix

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

[10] S. M. Freier, D. Alkema, A. Sinclair, T. Neilson, and D. H. Turner, *Biochemistry* **24,** 4533 (1985).

[11] S. M. Freier, N. Sugimoto, A. Sinclair, D. Alkema, T. Neilson, R. Kierzek, M. H. Caruthers, and D. H. Turner, *Biochemistry* **25,** 3214 (1986).

[12] D. R. Hickey and D. H. Turner, *Biochemistry* **24,** 2086 (1985).

[13] N. Sugimoto, R. Kierzek, and D. H. Turner, *Biochemistry* **26,** 4559 (1987).

[14] M. J. Serra, T. J. Axensen, and D. H. Turner, *Biochemistry* **33,** 14289 (1994).

[14a] S. M. Freier, R. Kierzek, M. H. Caruthers, T. Neilson, and D. H. Turner, *Biochemistry* **25,** 3209 (1986).

TABLE II

THERMODYNAMIC PARAMETERS FOR UNPAIRED TERMINAL NUCLEOTIDES[a]

	X = A			X = C			X = G			X = U		
Nucleotides	ΔH°	ΔS°	ΔG°_{37}	ΔH°	ΔS°	ΔG°_{37}	ΔH°	ΔS°	ΔG°_{37}	ΔH°	ΔS°	ΔG°_{37}
3′ Unpaired												
$\overrightarrow{AX}$ / U ←	−4.9	−13.2	−0.8	−0.9	−1.2	−0.5	−5.5	−15.0	−0.8	−2.3	−5.4	−0.6
$\overrightarrow{CX}$ / G ←	−9.0	−23.4	−1.7	−4.1	−10.7	−0.8	−8.6	−22.2	−1.7	−7.5	−20.4	−1.2
$\overrightarrow{GX}$ / C ←	−7.4	−20.0	−1.1	−2.8	−7.9	−0.4	−6.4	−16.6	−1.3	−3.6	−9.7	−0.6
$\overrightarrow{UX}$ / A ←	−5.7	−16.4	−0.7	−0.7	−1.8	−0.1	−5.8	−16.4	−0.7	−2.2	−6.8	−0.1
5′ Unpaired												
$\overrightarrow{XA}$ / U ←	1.6	6.1	−0.3	2.2	7.9	−0.3	0.7	3.4	−0.4	3.1	10.6	−0.2
$\overrightarrow{XC}$ / G ←	−2.4	−6.0	−0.5	3.3	11.8	−0.3	0.8	3.4	−0.2	−1.4	−4.3	−0.1
$\overrightarrow{XG}$ / C ←	−1.6	−4.5	−0.2	0.7	3.1	−0.3	−4.6	−14.8	−0.0	−0.4	−1.2	−0.0
$\overrightarrow{XU}$ / A ←	−0.5	−0.7	−0.3	6.9	22.8	−0.1	(0.6)	(2.7)	(−0.2)	(0.6)	(2.7)	(−0.2)

[a] In 1 *M* NaCl. (From Turner *et al.*[9])

containing AA terminal mismatches, $\begin{matrix}\text{AGGAUCCA}\\ \text{ACCUAGGA}\end{matrix}$, is shown in Fig. 2. The terminal AA mismatches increase the stability of the duplex relative to the core duplex by $2 \times 1.5 = 3.0$ kcal/mol at 37°. Notice that the terminal AA mismatch helix is slightly less stable than the corresponding helix with only 3′ dangling A residues, and considerably more stable than the corresponding helix with only 5′ dangling A residues.

Loops

Loops and single-stranded regions have not been investigated as fully as base-paired regions. Thus, the sequence dependence of their stabilities is not well characterized or understood, and parameters used for predictions are still changing. Hairpins have been studied the most extensively.

Hairpin Loops

Hairpins are a prominent RNA structural motif. For example, nearly 70% of *Escherichia coli* 16S rRNA is found in small hairpin structures. A hairpin can be represented in the following way:

A A	
U U	loop
X *Y*	first mismatch
B · P	closing base pair
G · C	stem
G · C	

A reasonable model for predicting the free energy for hairpin loop formation is to consider the loop size, closing base pair, interactions of the first and last unpaired nucleotides of the loop with the closing base pair, and an additional stabilization for UU or GA first mismatches.[14]

$$\Delta G^\circ_{37L}(n) = \Delta G^\circ_{37i}(n) + \Delta G^\circ_{37MM} \quad \begin{matrix}+0.4 \text{ (if closed by AU or UA) and}\\ -0.7 \text{ (if first mismatch is GA or UU)}\end{matrix} \qquad (3)$$

Here $\Delta G^\circ_{37i}(n)$ is the free energy for initiating a loop of n nucleotides. Table III lists ΔG°_{37} values for initiating hairpin loops as a function of loop size. Not much is known about the factors determining the ΔH° and ΔS° of initiation for hairpin loops. Therefore it is usually assumed that $\Delta H^\circ_i = 0$ kcal/mol for loop initiation and that $\Delta S^\circ_i = -\Delta G^\circ_{37i}/310.15$. Given this approximation, the thermodynamic parameters for hairpin loops are predicted as illustrated in Fig. 3.

The predicted thermodynamic parameters for hairpin folding are calcu-

TABLE III

FREE ENERGY PARAMETERS FOR LOOP INITIATION[a]

Size (nucleotides)	$\Delta G^\circ_{37,i}$ (kcal/mol)		
	Hairpin[b]	Bulge[c]	Internal[d]
1		3.9	
2		3.1	0.8[e]
3	4.1[e]	3.5	5.1
4	4.9	(4.2)	4.9
5	4.4	4.8	5.3
6	5.0	(5.0)	5.7
7	5.0	(5.2)	(5.9)
8	(5.1)	(5.3)	(6.0)
9	5.2	(5.4)	(6.1)

[a] At 37° in 1 *M* NaCl. For large loop sizes, *n*, use $\Delta G^\circ(n) = \Delta G^\circ(n_{max}) + 1.75\ RT \ln (n/n_{max})$, where n_{max} is 9, 5, and 6 for hairpin, bulge, and internal loops, respectively. Parameters not derived from experimental measurements are listed in parentheses. All parameters and models are derived from limited experimental data and therefore subject to change without further notice.

[b] For hairpin loops consisting of more than three nucleotides, ΔG°_{37} is approximated as

$$\Delta G^\circ_{37}(\text{loop} > 3) = \Delta G^\circ_{37,i} + \Delta G^\circ_{37}\begin{pmatrix} BX \\ PY \end{pmatrix} - 0.7 \text{ (if } XY \text{ is GA or UU)} + 0.4 \text{ (if } BP \text{ is AU or UA)}$$

Here *BP* is the closing base pair and *XY* is the first mismatch in the loop.

[c] The parameter for bulge loops of one nucleotide only assumes additional stability is conferred by stacking of the adjacent base pairs as approximated by nearest-neighbor parameters in Table I. No stacking across bulges of two or more nucleotides is assumed.

[d] For internal loops consisting of more than two nucleotides with branches of N_1 and N_2 nucleotides, ΔG°_{37} is approximated as

$$\Delta G^\circ_{37}(\text{loop} > 2) = \Delta G^\circ_{37,i} + \Delta G^\circ_{37}(\text{mm1}) + \Delta G^\circ_{37}(\text{mm2}) + \text{minimum of } 3.0 \text{ or } 0.3|N_1 - N_2|$$

Here ΔG°_{37}(mm1) and ΔG°_{37}(mm2) are free energy increments for the first and last mismatch in the loop. They are approximated as −2.7, −2.5, and −1.5 kcal/mol, respectively, for GA, UU, and other mismatches adjacent to CG pairs, and −2.2, −2.0, and −1.0 kcal/mol, respectively, for these mismatches adjacent to AU pairs. For a loop of three nucleotides, one of the ΔG°_{37}(mm) is taken from free energy increments for dangling ends (Table II).

[e] $\Delta G^\circ_{37,i}$ values for hairpins of three nucleotides and internal loops of two nucleotides (single mismatches) are assumed to be independent of the sequence in the loop.

Hairpin

```
  A A
U   U
G   A
 C•G
 G•C
 G•C
```

$$\Delta H^\circ_{loop} = \Delta H^\circ_{loop(6)} + \Delta H^\circ_{GA}^{CG} + \Delta H^\circ_{A}^{G} = 0+(-8.2)+(-0.7) = -8.9 \text{ kcal/mol}$$

$$\Delta S^\circ_{loop} = \Delta S^\circ_{loop(6)} + \Delta S^\circ_{GA}^{CG} = (-16.1)+(-21.8) = -37.9 \text{ eu}$$

$$\Delta G^\circ_{37}loop = \Delta G^\circ_{37}loop(6) + \Delta G^\circ_{37}GA^{CG} + \Delta G^\circ_{37}\ {}^{G}_{A}$$

$$\Delta G^\circ_{37}loop = 5.0+(-1.4)+(-0.7) = +2.9 \text{ kcal/mol}$$

$$\Delta H^\circ_{stem} = \Delta H^\circ_{CC}^{GG} + \Delta H^\circ_{CG}^{GC} = (-12.2)+(-14.2) = -26.4 \text{ kcal/mol}$$

$$\Delta S^\circ_{stem} = \Delta S^\circ_{CC}^{GG} + \Delta S^\circ_{CG}^{GC} = (-29.7)+(-34.9) = -64.6 \text{ eu}$$

$$\Delta G^\circ_{37}stem = \Delta G^\circ_{37}CC^{GG} + \Delta G^\circ_{37}CG^{GC} = (-2.9)+(-3.4) = -6.3 \text{ kcal/mol}$$

$$\Delta H^\circ_{hairpin} = \Delta H^\circ_{loop} + \Delta H^\circ_{stem} = (-8.9)+(-26.4) = -35.3 \text{ kcal/mol}$$

$$\Delta S^\circ_{hairpin} = \Delta S^\circ_{loop} + \Delta S^\circ_{stem} = (-37.9)+(-64.6) = -102.5 \text{ eu}$$

$$\Delta G^\circ_{37} = \Delta H^\circ - T\Delta S = -35.3 - 310.15(-102.5)/1000 = -3.4 \text{ kcal/mol}$$

$$\Delta G^\circ_{37}hairpin = \Delta G^\circ_{37}loop + \Delta G^\circ_{37}stem = +2.9 + (-6.3) = -3.4 \text{ kcal/mol}$$

$$T_M(^\circ C) = \frac{\Delta H^\circ}{\Delta S^\circ} - 273.15$$

$$= \frac{35300}{102.5} - 273.15 = 71.2^\circ C$$

FIG. 3. Prediction of thermodynamic stability for hairpins using nearest-neighbor model.

lated by taking the sum of the loop plus the stem values. Note that the duplex initiation term is not included in determining the stem value because initiation is included in the loop values. The interaction of the first mismatch with the closing base pair is assumed to be the same for the hairpin as for a terminal mismatch on a duplex and contributes both an enthalpy and entropy term to the stability of the hairpin. The additional stability of the GA and UU (−0.7 kcal/mol) first mismatches is approximated as entirely due to a favorable enthalpy term. In a similar fashion, the penalty for closure of the hairpin loop by an AU or UA base pair is approximated as an unfavorable enthalpy (+0.4 kcal/mol). The melting of a hairpin is a

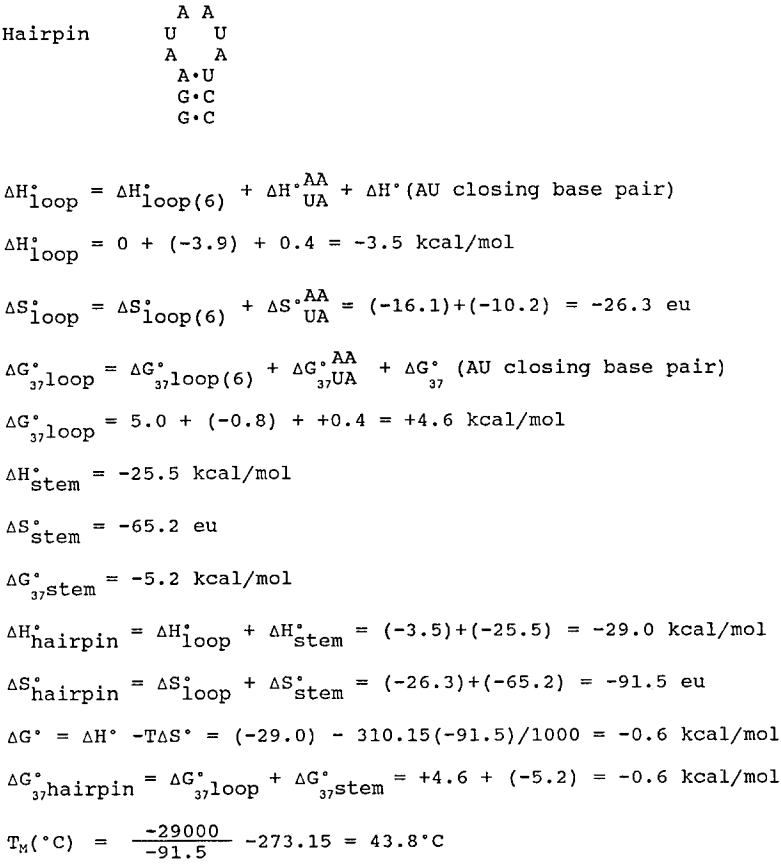


FIG. 3. (*continued*)

unimolecular reaction and the melting temperature is therefore concentration independent:

$$T_m(^\circ C) = (\Delta H^\circ/\Delta S^\circ) - 273.15 \quad (4)$$

This model predicts the melting temperature for a large number of hairpins within 5° of the actual melting temperature.

The model is undoubtedly oversimplified, and should therefore be used with caution. For example, the hairpin loop C(UUCG)G is more stable than expected partially due to an interaction between the amino group of the loop C and a phosphate.[15] Very small loops (fewer than four nucleotides)

[15] G. Varani, C. Cheong, and I. Tinoco, Jr., *Biochemistry* **30,** 3280 (1991).

or large loops (more than eight nucleotides) are not predicted well by this model and additional terms will need to be considered when more data become available.

Bulge Loops

Bulge loops contain unpaired bases on one side of a double helix. Little is known about the sequence dependence of bulge loops. Therefore, the stability of these loops is modeled as sequence independent. Base pairs adjacent to a single nucleotide bulge loop are assumed to maintain their stacking interactions. Bulge loops of two or more nucleotides disrupt the nearest neighbor interactions of the adjacent base pairs.[16,17] The $\Delta G^\circ_{37,i}$ values for bulge loops of different sizes are given in Table III. Examples for predicting the stability of duplexes containing bulge loops are presented in Fig. 4. Current data sets are not complete enough to allow partitioning of the bulge loop stability into enthalpic and entropic energy terms. Therefore ΔH°, ΔS°, and melting temperatures cannot be predicted reliably.

Internal Loops

Internal loops are regions within a double helix that contain base pairs other than AU, GC, and GU. Little is known about the determinants of stability for these regions. The limited data available can be fit reasonably well with a model that includes the loop size, closing base pairs, first mismatches, and loop asymmetry. Table III lists ΔG°_{37} values for initiating symmetric internal loops as a function of loop size. If an internal loop has a different number of nucleotides on the two sides, then an asymmetry penalty is added to ΔG°_{37} for initiation. This penalty is the minimum of 3.0 or $0.3|N_1 - N_2|$ kcal/mol for a loop with branches of N_1 and N_2 nucleotides.[18] For example, internal loops of three nucleotides ($N_1 = 2$, $N_2 = 1$) are given $\Delta G^\circ_{37,i}$ values of $5.1 + 0.3\ (2 - 1) = 5.4$ kcal/mol. Single mismatches are given $\Delta G^\circ_{37} = 0.8$ kcal/mol regardless of sequence, owing to lack of experimental data. For larger internal loops, terminal GA, UU, and other mismatches adjacent to CG pairs are assumed to make the loop more stable by -2.7, -2.5, and -1.5 kcal/mol, respectively.[19] These mismatches adjacent to AU pairs are assumed to make the loop more stable by -2.2, -2.0, and -1.0 kcal/mol, respectively. The free energy increments for the initial mismatches differ considerably from those used for hairpin loops. They are

[16] J. A. Jaeger, D. H. Turner, and M. Zuker, *Proc. Natl. Acad. Sci. U.S.A.* **86,** 7706 (1989).
[17] K. Weeks and D. M. Crothers, *Science* **261,** 1574 (1993).
[18] A. E. Peritz, R. Kierzek, N. Sugimoto, and D. H. Turner, *Biochemistry* **30,** 6429 (1991).
[19] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker, *Proc. Natl. Acad. Sci. U.S.A.* **91,** 9218 (1994).

primarily based on measurements of internal loops of four to six nucleotides.[18,20,21] It would not be surprising to find that larger internal loops behave more like hairpins. Results suggest the above-described rules may not even reasonably mimic the sequence dependence of symmetric internal loops of four nucleotides because it appears that $\begin{matrix}\text{GGAC}\\\text{CAGG}\end{matrix}$ motifs are about 2 kcal/mol more stable than $\begin{matrix}\text{CGAG}\\\text{GAGC}\end{matrix}$ motifs.[22] Examples for predicting the stability of duplexes containing internal loops are presented in Fig. 4. Prediction of ΔH° and ΔS° cannot be made with the currently available data. Clearly, additional experiments are required so that a more complete model can be developed for predicting internal loop stability.

Multibranch Loops

Multibranch loops are formed by the intersection of more than two helices, and usually also contain unpaired nucleotides. The factors determining their stability have not been studied systematically. A crude model for ΔG°_{37} that works reasonably well for predicting secondary structures is the following[16,19]:

$$\begin{aligned}\Delta G^\circ_{37}\,(\text{kcal/mol}) = {} & 4.6 + 0.4\,[\text{no. of unpaired nucleotides (nt)}]\\ & + 0.1\,(\text{no. of helices})\\ & + \Delta G^\circ_{37}\,(\text{coaxial stacking of helices})\\ & + \Delta G^\circ_{37}\,(\text{mismatches and unpaired nucleotides})\end{aligned} \quad (5)$$

For loops with more than six unpaired nucleotides, 0.4 (no. of unpaired nt) can be replaced with $2.4 + 1.987(.31015)(1.75)\ln(n/6)$. Coaxial stacking is possible when two helices are directly adjacent or separated by a GA mismatch. For directly adjacent helices, ΔG°_{37} (coaxial stacking of helices) has been approximated as the ΔG°_{37} for the equivalent nearest-neighbor base pair combination in an intact helix as listed in Table I.[19] For helices separated by a GA mismatch, ΔG°_{37} coaxial stacking has been approximated as -1 kcal/mol greater stability than the nearest-neighbor mismatch value. An example for predicting multibranch loop stability is presented in Fig. 5.

Limitations

In the above discussions, various "rules" have been presented for predicting the thermodynamic properties of RNA sequences. It should be kept

[20] J. Gralla and D. M. Crothers, *J. Mol. Biol.* **78,** 301 (1973).
[21] J. SantaLucia, Jr., R. Kierzek, and D. H. Turner, *Biochemistry* **30,** 8342 (1991).
[22] A. E. Walter, M. Wu, and D. H. Turner, *Biochemistry* **33,** 11349 (1994).

Bulge Loop of 1:

```
5'  GCGAGCG
    ••• •••
3'  CGC CGC
```

$$\Delta H^\circ = \Delta H^\circ{}^{GC}_{CG} + \Delta H^\circ{}^{CG}_{GC} + \Delta H^\circ{}^{GG}_{CC} + \Delta H^\circ{}^{GC}_{CG} + \Delta H^\circ{}^{CG}_{GC}$$

$$\Delta H^\circ = (-14.2)+(-8.0)+(-12.2)+(-14.2)+(-8.0) = -56.6 \text{ kcal/mol}$$

$$\Delta S^\circ = \Delta S^\circ{}^{GC}_{CG} + \Delta S^\circ{}^{CG}_{GC} + \Delta S^\circ{}^{GG}_{CC} + \Delta S^\circ{}^{GC}_{CG} + \Delta S^\circ{}^{CG}_{GC} + \Delta S^\circ_{int} + \Delta S^\circ_{bulge}$$

$$\Delta S^\circ = (-34.9)+(-19.4)+(-29.7)+(-34.9)+(-19.4)+(-10.8)+(-12.6)$$

$$= -161.7 \text{ eu}$$

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -56.6 - 310.15(-161.7)/1000 = -6.4 \text{ kcal/mol}$$

$$\Delta G^\circ_{37} = \Delta G^\circ{}^{GC}_{37CG} + \Delta G^\circ{}^{CG}_{37GC} + \Delta G^\circ{}^{GG}_{37CC} + \Delta G^\circ{}^{GC}_{37CG} + \Delta G^\circ{}^{CG}_{37GC} + \Delta G^\circ_{37int} + \Delta G^\circ_{37}\text{bulge loop of 1}$$

$$\Delta G^\circ_{37} = (-3.4)+(-2.0)+(-2.9)+(-3.4)+(-2.0)+3.4+3.9 = -6.4 \text{ kcal/mol}$$

$$T_M(\text{predicted } (10^{-4}\text{M})) = \frac{-56600}{-161.7 + R \ln (10^{-4}/4)} - 273.15 = 36.6^\circ\text{C}$$

Bulge Loop Greater than 1:

```
5'  GCGAAGCG
    ••• •••
3'  CGC CGC
```

$$\Delta H^\circ = \Delta H^\circ{}^{GC}_{CG} + \Delta H^\circ{}^{CG}_{GC} + \Delta H^\circ{}^{GC}_{CG} + \Delta H^\circ{}^{CG}_{GC}$$

$$\Delta H^\circ = (-14.2)+(-8.0)+(-14.2)+(-8.0) = -44.4 \text{ kcal/mol}$$

$$\Delta S^\circ = \Delta S^\circ{}^{GC}_{CG} + \Delta S^\circ{}^{CG}_{GC} + \Delta S^\circ{}^{GC}_{CG} + \Delta S^\circ{}^{CG}_{GC} + \Delta S^\circ_{int} + \Delta S^\circ_{bulge}$$

$$\Delta S^\circ = (-34.9)+(-19.4)+(-34.9)+(-19.4)+(-10.8)+(-10.0) = -129.4 \text{ eu}$$

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -44.4 - 310.15(-129.4)/1000 = -4.3 \text{ kcal/mol}$$

$$\Delta G^\circ_{37} = \Delta G^\circ{}^{GC}_{37CG} + \Delta G^\circ{}^{CG}_{37GC} + \Delta G^\circ{}^{GC}_{37CG} + \Delta G^\circ{}^{CG}_{37GC} + \Delta G^\circ_{37int} + \Delta G^\circ_{37}\text{bulge loop of 2}$$

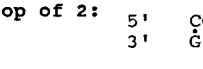
FIG. 4. Prediction of thermodynamic properties for duplexes with bulge and internal loops using nearest-neighbor model.

in mind, however, that these rules are based on a crude nearest-neighbor model, and should be considered only gross approximations. This is true even for helices containing only Watson–Crick base pairs. For example, the duplexes $\begin{matrix}\text{UGAUCA}\\ \text{ACUAGU}\end{matrix}$ and $\begin{matrix}\text{UCAUGA}\\ \text{AGUACU}\end{matrix}$ both contain the same nearest neighbors and are predicted to have equal stability, but their ΔG°_{37} values

$\Delta G^\circ_{37} = (-3.4)+(-2.0)+(-3.4)+(-2.0)+3.4+3.1 = -4.3$ kcal/mol

$$T_M(10^{-4}M)) = \frac{-44400}{-129.4 + R \ln (10^{-4}/4)} - 273.15 = 22.0^\circ C$$

Internal loop of 2:



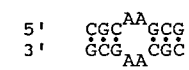
ΔG°_{37}loop of 2 = 0.8 (sequence independent)

$$\Delta G^\circ_{37} = \Delta G^\circ_{37}{}^{CG}_{GC} + \Delta G^\circ_{37}{}^{GC}_{CG} + \Delta G^\circ_{37}\text{loop of 2} + \Delta G^\circ_{37}{}^{GC}_{CG} + \Delta G^\circ_{37}{}^{CG}_{GC} + \Delta G^\circ_{37}\text{int}$$

$\Delta G^\circ_{37} = (-2.0)+(-3.4)+0.8+(-3.4)+(-2.0)+3.4 = -6.6$ kcal/mol

Note: Data set is not sufficient to predict ΔH° and ΔS° for internal loops.

Internal loop of 4 (symmetric):



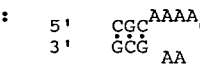
$$\Delta G^\circ_{37}\text{loop} = \Delta G^\circ_{37}\text{i,loop 4} + \Delta G^\circ_{37}\text{MM1} + \Delta G^\circ_{37}\text{MM2} + \Delta G^\circ_{37}\text{asymmetry penalty}$$

ΔG°_{37}loop = +4.9 + (-1.5) + (-1.5) + 0 = 1.9 kcal/mol

$$\Delta G^\circ_{37}\text{duplex} = \Delta G^\circ_{37}{}^{CG}_{GC} + \Delta G^\circ_{37}{}^{GC}_{CG} + \Delta G^\circ_{37}\text{loop} + \Delta G^\circ_{37}{}^{GC}_{CG} + \Delta G^\circ_{37}{}^{CG}_{GC} + \Delta G^\circ_{37}\text{int}$$

ΔG°_{37}duplex = (-2.0)+(-3.4)+(1.9)+(-3.4)+(-2.0)+3.4 = -5.5 kcal/mol

Internal loop of 6 (asymmetric):



$$\Delta G^\circ_{37}\text{loop} = \Delta G^\circ_{37}\text{i,loop of 6} + \Delta G^\circ_{37}\text{MM1} + \Delta G^\circ_{37}\text{MM2} + \Delta G^\circ_{37}\text{asymmetry penalty}$$

ΔG°_{37}loop = +5.7+(-1.5)+(-1.5)+(0.3(4-2)) = 3.3 kcal/mol

$$\Delta G^\circ_{37}\text{duplex} = \Delta G^\circ_{37}{}^{CG}_{GC} + \Delta G^\circ_{37}{}^{GC}_{CG} + \Delta G^\circ_{37}\text{loop} + \Delta G^\circ_{37}{}^{GC}_{CG} + \Delta G^\circ_{37}{}^{CG}_{GC} + \Delta G^\circ_{37}\text{int}$$

ΔG°_{37}duplex = (-2.0)+(-3.4)+(3.3)+(-3.4)+(-2.0)+3.4 = -4.1 kcal/mol

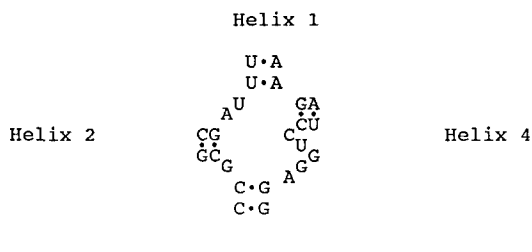
FIG. 4. (*continued*)

for duplex formation are −5.1 and −3.9 kcal/mol, respectively.[23] Nonnearest-neighbor effects are likely to be more important for loops, and have been observed for helices containing bulges,[24] and tandem GU mismatches.[6] It should also be kept in mind that the available database is still

[23] R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, and S. M. Freier, *Biochemistry* **25,** 7840 (1986).

[24] C. E. Longfellow, R. Kierzek, and D. H. Turner, *Biochemistry* **29,** 278 (1990).

Multi Branch Loop from unmodified Phenylalanyl t-RNA



ΔG°_{37} coaxial stack of helix 1 and 4

$\Delta G^\circ_{37} = \Delta G^\circ_{37}\ {}^{GA}_{CU} = -2.3$ kcal/mol

ΔG°_{37} AG mismatch on helix 2 and 3' unpaired A on helix 3

$\Delta G^\circ_{37} = \Delta G^\circ_{37}{}^{CG}_{GA} + \Delta G^\circ_{37}{}^{GA}_{C} = (-1.4)+(-1.1) = -2.5$ kcal/mol

Number of unpaired nucleotides = 8

Number of helixes = 4

ΔG°_{37} multi branch loop = 4.6 + ΔG°_{37} (unpaired nucleotides)

+ 0.1 (# of helixes) + ΔG°_{37} coaxial stacking

+ ΔG° (mismatches and unpaired nucleotides)

ΔG°_{37} multi branch loop =

4.6 + (2.4+[1.987(310.15)(1.75)/1000)ln(8/6)]

+0.1(4)+(-2.3)+(-2.5) = +2.9 kcal/mol

Note: The thermodynamic approach does not predict the correct local geometry between helixes 2 and 3.

FIG. 5. Prediction of free energy increment of multibranch loop using nearest-neighbor model.

small. Thus most of the parameters listed above are based on a limited number (often one) of model systems. Presumably, as the database gets larger, predictions will improve both because parameters will improve and also because more sophisticated models will be developed. Even with these limitations, however, the current state of knowledge is sufficient for aiding design and interpretation of many experiments.

Prediction of RNA Secondary Structure

Another thermodynamic property of RNA that can be predicted in principle is the equilibrium folding. This may or may not be the physiologically important folding depending on the kinetics of folding. Attempts to predict RNA secondary structures on the basis of thermodynamic data described above with slight modifications have been about 70% successful[19] when compared with those determined by phylogenetic analysis and/or chemical mapping. In phylogenetic analysis, sequences for RNA molecules with similar functions are compared to find common folding.[25] The cloverleaf secondary structure for tRNA is a typical example. In chemical mapping, the RNA is allowed to react with reagents selective for single- or double-stranded regions.[26] The reactivity of each nucleotide then provides constraints on possible secondary structures. Presumably, both phylogenetic analysis and chemical mapping reflect physiologically important secondary structures. The similarity to structures predicted from thermodynamic considerations suggests thermodynamics is at least a major determinant of secondary structure folding. In practice, the most powerful way to deduce the secondary structure from sequence is to combine all of these methods.

Acknowledgments

This work is supported by NIH Grant GM22939 (D. H. T.), NIH Grant GM49429 (M. J. S.), the Research Corporation (M. J. S.), and the DANA Foundation. D. H. T. is a Guggenheim Fellow and an American Cancer Society Scholar.

[25] C. R. Woese and N. R. Pace, *in* "The RNA World" (R. F. Gesteland and J. F. Atkins, eds.), Chapter 4. Cold Spring Harbor Lab. Press, Plainview, NY, 1993.
[26] R. Parker, this series, Vol. 180, p. 510.

[12] Thermodynamics and Mutations in RNA–Protein Interactions

By KATHLEEN B. HALL and JAMES K. KRANZ

To describe the association between an RNA and a protein, it is necessary to define the local interactions between nucleotides and amino acids and also to determine the energetics of the association. The local interactions will show how the specificity of the association is conferred; the energetics will provide the assembly parameters that encompass both the individual interactions and their interdependence.