von Hippel, P. H., Bear, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.

Winter, R. B., & von Hippel, P. H. (1981) *Biochemistry 20*, 6948–6960.

Winter, R. B., Berg, O. G., & von Hippel, P. H. (1981) *Biochemistry 20*, 6961–6977.

Yasuda, S., & Sekiguchi, M. (1976) *Biochim. Biophys. Acta 442*, 197–207.

Energetics of Internal GU Mismatches in Ribooligonucleotide Helixes[†]

Naoki Sugimoto,[‡] Ryszard Kierzek,[§] Susan M. Freier,[‡] and Douglas H. Turner*[,‡]

Department of Chemistry, University of Rochester, Rochester, New York 14627, and Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-704 Poznan, Noskowskiego 12/14, Poland

Received January 13, 1986; Revised Manuscript Received May 15, 1986

ABSTRACT: Thermodynamic parameters of helix formation were measured spectroscopically for 16 oligoribonucleotides containing either internal GU mismatches or the corresponding AU pairs. Internal GU mismatches stabilize each helix, but not as much as the corresponding AU pairs. The differences in the enthalpy and entropy changes of helix formation associated with replacing AU pairs with GU mismatches are less than previously realized. At both 25 and 37 °C, the decrease in helix stability associated with replacing an AU with a GU is also less than thought previously. Approximations are suggested for predicting the effects of GU mismatches on helix stability.

Internal GU mismatches in helical regions are known to occur in tRNA (Kim et al., 1974; Ladner et al., 1975; Sussman & Kim, 1976; Johnston & Redfield, 1981; Sprinzl et al., 1985) and are proposed to occur in the secondary structures of essentially all large RNA molecules (Fox & Woese, 1975; Noller, 1984; Woese et al., 1983; Steger et al., 1984; Cech et al., 1983). Romaniuk et al. (1979a,b) and Alkema et al. (1982) used NMR to demonstrate the existence of GU mismatches in RNA oligonucleotides. Corresponding GT mismatches in DNA oligomers have been studied by Patel et al. (1982), Aboul-ela et al. (1985), and Brown et al. (1985). Despite the prevalence of internal GU mismatches, there is only one experimental study of their thermodynamic properties (Uhlenbeck et al., 1971). This study was limited to A_4GUU_4, A_4UGU_4, and A_5UGU_5 because of the lack of methods for making defined-sequence oligonucleotides. Recent advances have made it possible to make oligoribonucleotides with more varied sequences (Uhlenbeck & Gumport, 1982; Kierzek et al., 1987; Markiewicz et al., 1984; Matteucci & Caruthers, 1980). This paper reports thermodynamic measurements on 16 oligomers containing either internal GU mismatches or the corresponding AU pairs. The results are important for improving predictions of RNA structure and stability (Tinoco et al., 1973; Borer et al., 1974; Gralla & Crothers, 1973), and the temperature dependence of these properties (Steger et al., 1984). The results indicate internal GU pairs are more stable and more temperature dependent than previously realized.

MATERIALS AND METHODS

Oligonucleotide Synthesis. Oligomers starting with AUG were synthesized by using T4 RNA ligase to make successive additions of nucleoside 5′,3′-bisphosphates to AUG (Sigma) as described previously (Hickey & Turner, 1985; Freier et al., 1985a). Ligase reactions differed widely in rates, requiring 8 h to 7 days at 37 or 4 °C for 100% yield. Reactions and product purity were monitored by HPLC using a methanol gradient (Petersheim & Turner, 1983). These oligomers have a 3′-terminal phosphate.

AUACGUAU, GCUGGC, GGCGUC, CUGCGG, CGGCUG, and GCUAGC were synthesized chemically on solid support by using phosphoramidite procedures, and purified by anion-exchange chromatography on DEAE-Sephadex (A-25) using NaCl gradients in 7 M urea, 10 mM Tris, pH 8.2. These oligomers do not have a 3′-terminal phosphate.

Melting Curves. Extinction coefficients were calculated with the nearest-neighbor approximation (Richards, 1975). For oligomers studied at 260 nm, the calculated extinction coefficients ($\times 10^{-4}$ cm^{-1} M^{-1}) at 260 nm are as follows: AUGCGUp, 4.89; AUGCAUp, 5.09; AUGCGCGUp, 6.14; AUGCGCAUp, 6.24; AUGCGUAUp, 6.72; AUACGUAU, 7.09; AUGUGCAUp, 6.89; AUGUACAUp, 7.58; GCUGGC, 4.38; GCUAGC, 4.81. For oligomers studied at 280 nm, the calculated extinction coefficients ($\times 10^{-4}$ cm^{-1} M^{-1}) at 280 nm are as follows: GGCGUC, 3.15; CUGCGG and CGGCUG, 3.18. Total strand concentrations, C_T, were determined from the absorbance measured at 90 °C.

Absorbance vs. temperature melting curves were measured at 260 or 280 nm on a Gilford 250 spectrometer (Freier et al., 1983a). The heating rate was 1 °C/min, and the buffer for measurements at 260 nm was 1 M NaCl, 10 mM Na_2HPO_4, 0.1 mM Na_2EDTA, pH 7. For measurements at 280 nm, the buffer was 1 M NaCl, 5 mM Na_2HPO_4, 0.5 mM Na_2EDTA, pH 7. For each oligonucleotide, 13–21 absorbance vs. temperature profiles were measured over a 100-fold range in strand concentration. Melting curves were analyzed with a two-state model as previously described (Petersheim & Turner, 1983; Freier et al., 1983a,b).

RESULTS

Temperature-Independent Thermodynamic Parameters. For a two-state transition, the reciprocal melting temperature, T_m^{-1}, depends linearly on the log of the total strand concentration, C_T (Borer et al., 1974):

$$T_m^{-1} = \frac{2.303R}{\Delta H^\circ}(\log C_T) + \frac{\Delta S^\circ}{\Delta H^\circ} \qquad (1)$$

Plots suggested by eq 1 are shown in Figure 1. Good straight

[†] This work was supported by National Institutes of Health Grant GM 22939.

[‡] University of Rochester.

[§] Polish Academy of Sciences.

0006-2960/86/0425-5755$01.50/0

Table I: Thermodynamic Parameters of Helix Formation[a,b]

	log C_t parameters			"temperature-independent" parameters		
oligomer	$-\Delta H^{\circ c}$ (kcal/mol)	$-\Delta S^{\circ c}$ (eu)	$T_M{}^{c,d}$ (°C)	$-\Delta H^{\circ e}$ (kcal/mol)	$-\Delta S^{\circ e}$ (eu)	$T_M{}^{d,e}$ (°C)
GU mismatches						
AUGCGUp	24.5	65.4	19.5	24.3	64.5	20.3
AUGCGCGUp	54.9	147.0	58.6	54.4	145.8	58.7
AUGCGUAUp	46.8	133.9	34.4	44.7	127.1	34.6
AUGUGCAUp	57.1	164.2	39.5	54.2	155.0	39.7
GCUGGC	59.1	169.7	41.5	59.1	169.5	41.7
GGCGUC	38.1	107.8	29.0	37.7	106.1	29.6
CUGCGG	41.4	119.6	26.8	38.7	110.5	27.1
CGGCUG	43.2	121.4	35.7	44.3	125.0	35.8
reference helixes						
AUGCAUp	41.7	119.2	30.0	38.4	108.4	30.2
AUGCGCAUp	64.4	174.8	60.3	60.7	163.6	60.5
AUACGUAU	54.4	154.2	41.9	55.6	158.1	42.0
AUGUACAUp	55.9	159.3	41.6	53.7	152.1	41.8
GCUAGC	59.1	165.1	49.3	63.0	177.1	49.3
GACGUC	58.1	164.0	46.2	57.3	161.0	46.3
CUGCAGp[i]	55.4	156.0	45.3	54.5	153.0	45.4
CAGCUGp[i]	51.6	145.0	43.2	50.2	140.0	43.6
ACGCGU[f]	(50.8)	(137.5)	(52.7)			
ACGCGCGU[f]	(73.0)	(191.7)	(74.4)			
GCCGGCp[g]	62.7	166.0	67.2			
AGCU[f]	(29.5)	(84.8)	(12.7)			
AGCGCU[h]	50.1	135.7	52.1			
AUCGAU[f]	(46.0)	(132.7)	(31.6)			
GCGC[h]	30.5	83.4	26.5			
GGCGCCp[j]	67.8	182.0	65.2			
CCGCGG[f]	(54.6)	(144.5)	(62.1)			
CGGCCG$_p$[g]	54.1	143.0	63.3			
CGCG[f]	(30.3)	(85.0)	(19.7)			

[a] Measurements were in 1 M NaCl, 10 mM Na_2HPO_4, and 0.1 mM Na_2EDTA, pH 7, except for GGCGUC, CUGCGG, CGGCUG, GACGUC, CUGCAGp, and CAGCUGp. The latter oligomers were measured in 1 M NaCl, 5 mM Na_2HPO_4, and 0.5 mM Na_2EDTA. [b] Although estimated errors in ΔH° and ΔS° are ±5%, additional significant figures are given to allow accurate calculation of T_M. [c] From plots of reciprocal melting temperature vs. log C_T. [d] Calculated for 10^{-4} M strand concentration. [e] Temperature-independent thermodynamic parameters are the average of those from plots of $T_M{}^{-1}$ vs. log C_T and those from averaging fits of individual melting curves to a two-state model with sloping base lines. [f] Parameters predicted by using nearest neighbor parameters of Freier et al. (unpublished results). [g] Freier et al. (1985a). [h] Freier et al. (1986b). [i] D. Groebe, V. Cameron, S. M. Freier, D. H. Turner, and O. C. Uhlenbeck, unpublished results. [j] Freier et al. (1986a).

lines indicate the two-state model is reasonable. Thermodynamic parameters derived from these plots are listed in Table I. Also listed in Table I are "temperature-independent" parameters obtained by averaging parameters from eq 1 with parameters derived from fitting the shapes of the melting curves with the two-state model. Table I also contains thermodynamic parameters for additional reference oligomers. These parameters were either predicted from nearest-neighbor parameters (Freier et al., unpublished results) or measured previously.

Temperature-Dependent Thermodynamic Parameters. Enthalpy and entropy changes derived from fitting individual melting curves are slightly dependent on T_m. Plots of ΔH° vs. T_m and ΔS° vs. ln (T_m) provide heat capacity changes, $\Delta C_p{}^\circ$, for the coil to helix transition. These are listed in Table II along with the values of ΔH° and ΔS° at 37 °C derived from these plots. Within experimental error, substitution of a GU mismatch for an AU pair does not substantially affect $\Delta C_p{}^\circ$.

DISCUSSION

The purpose of this work is to improve predictions for thermodynamic properties of oligonucleotides containing internal GU mismatches. Table III lists increments, $\Delta\Delta H^\circ$, $\Delta\Delta S^\circ$, and $\Delta\Delta G^\circ$, associated with inserting GU mismatches into a helix and with substituting GU mismatches for AU or GC base pairs. For example, the free energy increment for inserting each GU mismatch in AUGCGU is $\Delta\Delta G^\circ(\text{GU}) = 0.5[\Delta G^\circ(\text{AUGCGU}) - \Delta G^\circ(\text{AGCU})]$. Values in parentheses were derived by calculating the reference helixes from nearest-neighbor parameters (Freier et al., unpublished results).

Table II: Temperature-Dependent Thermodynamic Parameters of Helix Formation[a,b]

oligomer	$-\Delta H^\circ{}_{37}{}^c$ (kcal/mol)	$-\Delta S^\circ{}_{37}{}^d$ (eu)	$-\Delta C^\circ{}_p$ (cal mol^{-1} K^{-1})
GU mismatches			
AUGCGUp	31.4	88.2	443
AUGCGCGUp	51.7	138.0	126
AUGCGUAUp	43.5	123.7	397
AUGUGCAUp	51.8	147.8	199
GCUGGC	56.8	164.0	991
GGCGUC	45.2	131.2	937
CUGCGG	39.3	112.9	236
CGGCUG	46.8	132.9	464
reference helixes			
AUGCAUp	37.8	107.1	379
AUGCGCAUp	53.5	140.5	178
AUACGUAU	55.0	156.2	368
AUGUACAUp	51.0	145.5	290
GCUAGC	61.1	171.1	416
GACGUC	54.2	151.3	255
CUGCAG$_p$[g]	52.2	144.1	260
CAGCUG$_p$[g]	48.0	134.1	730
GCCGGC$_p$[e]	61.5	162	730
GGCGCC$_p$[f]	58.4	154	185
CGGCCG$_p$[e]	53.8	142	111
AGCGCU	44.9	119	406
GCGC[f]	38.7	110	640

[a] Measurements were in 1 M NaCl, 10 mM Na_2HPO_4, and 0.1 mM Na_2EDTA, pH 7, except for GGCGUC, CUGCGG, CGGCUG, GACGUC, CUGCAGp, and CAGCUGp. The latter oligomers were measured in 1 M NaCl, 5 mM Na_2HPO_4, and 0.5 mM Na_2EDTA. [b] Although estimated errors in ΔH° and ΔS° at the T_M are ±10%, additional significant figures are given to allow accurate calculation of T_M. [c] From plots of ΔH° vs. T_M. [d] From plots of ΔS° vs. ln T_M. [e] Freier et al. (1985a). [f] Freier et al. (1986). [g] D. Groebe, V. Cameron, S. M. Freier, D. H. Turner, and O. C. Uhlenbeck, unpublished results.

Table III: Thermodynamic Increments for GU Mismatches[a]

	$\Delta\Delta H°$ (kcal/mol)			$\Delta\Delta S°$ (eu)			$\Delta\Delta G°_{37}$ (kcal/mol)		
oligomer	(GU-AU)[b]	(GU-GC)[b]	+GU[c]	(GU-AU)	(GU-GC)	+GU	(GU-AU)	(GU-GC)	+GU
AU̇GCĠUp	8.6	(13.2)	(2.5)	26.9	(36.1)	(9.7)	0.26	(1.97)	(−0.51)
AU̇GCGCĠUp	4.8	(9.1)	−2.4	13.9	(22.4)	−5.7	0.44	(2.12)	−0.65
AUĠCGU̇AUp	3.8	8.8	(−0.4)	10.2	20.5	(−0.6)	0.65	2.46	(−0.21)
AUGU̇ĠCAUp	0.6	3.7	−7.7	−2.5	5.3	−22.5	0.16	2.01	−0.72
GCU̇ĠGC	0.0	1.8	−14.3	−2.3	−1.9	−43.2	0.71	2.38	−0.92
GĠCGU̇C	10.0	14.9	−3.8	28.1	37.1	−12.2	1.29	3.34	−0.02
CU̇GCĠG	7.0	(6.6)	(−5.6)	18.2	(12.5)	(−17.3)	1.36	(2.74)	(−0.18)
CĠGCU̇G	4.2	5.5	(−6.5)	11.8	10.8	(−18.2)	0.54	2.10	(−0.80)

[a] Increments were calculated from parameters derived from plots of T_M^{-1} vs. log C_T. Values in parentheses were calculated by using the nearest-neighbor parameters of Freier et al. (unpublished results) for the reference helix. [b] Increments are defined as in the following example: for AUGCGUp, $\Delta\Delta H°[(\text{GU-AU})] = 0.5[\Delta H°(\text{AUGCGUp}) - \Delta H°(\text{AUGCAUp})]$. [c] Increments are defined as in the following example: for AUGCGUp, $\Delta\Delta H°(+\text{GU}) = 0.5[\Delta H°(\text{AUGCGUp}) - \Delta H°(\text{AGCU})]$.

Table IV: Thermodynamic Parameters for RNA Helix Propagation by Internal GU Mismatches in 1 M NaCl[a]

sequence	$\Delta G°_{37}$ (kcal/mol)	sequence	$\Delta G°_{37}$ (kcal/mol)
$\genfrac{}{}{0pt}{}{5'\mathrm{CG}3'}{3'\mathrm{GU}5'}$	−1.5	$\genfrac{}{}{0pt}{}{5'\mathrm{UG}3'}{3'\mathrm{GU}5'}$	−0.6
$\genfrac{}{}{0pt}{}{5'\mathrm{GU}3'}{3'\mathrm{UA}5'} = \genfrac{}{}{0pt}{}{5'\mathrm{UG}3'}{3'\mathrm{AU}5'}$	−0.7	$\genfrac{}{}{0pt}{}{5'\mathrm{GG}3'}{3'\mathrm{UC}5'}$	−1.5
$\genfrac{}{}{0pt}{}{5'\mathrm{GC}3'}{3'\mathrm{UG}5'}$	−1.9	$\genfrac{}{}{0pt}{}{5'\mathrm{GG}3'}{3'\mathrm{CU}5'}$	−1.3

[a] Data from plots of T_M^{-1} vs. log C_T were analyzed by assuming that all thermodynamic parameters for $\genfrac{}{}{0pt}{}{5'\mathrm{GU}3'}{3'\mathrm{UA}5'}$ and $\genfrac{}{}{0pt}{}{5'\mathrm{UG}3'}{3'\mathrm{AU}5'}$ are equal and that thermodynamic parameters for nearest neighbors involving Watson–Crick base pairs are identical with those determined from 45 oligomers by Freier et al. (unpublished results.)

The comparisons for several oligomers neglect effects of 3′-terminal phosphates. These effects are small, however (Freier et al., 1985a).

All free energy increments for inserting GU mismatches into helixes are negative so that adding GU mismatches stabilizes these duplexes. The increments in $\Delta G°_{37}$ range from −0.02 to −0.9 kcal/mol GU mismatch, so sequence affects the stability of GU mismatches. The effect of sequence is also seen in the increments associated with substituting GU mismatches for AU or GC pairs (see Table III). For example, $\Delta\Delta G°_{37}$ (GU-AU) ranges from 0.2 to 1.4 kcal/mol of GU mismatch. The sequence dependence appears to extend beyond a nearest-neighbor dependence. For example AUGCGU and AUGCGCGU have different thermodynamic increments relative to AUGCAU and AUGCGCAU, respectively. The non-nearest-neighbor dependence is most striking when CUGCGG and CGGCUG are compared. The corresponding oligomers with AU base pairs, CUGCAG and CAGCUG, have the same nearest neighbors, and at 10^{-4} M have similar melting temperatures of 45 and 43 °C, respectively. When the AU base pairs are replaced by GU mismatches, the order of melting reverses and the difference increases to 9 °C: 27 vs. 36 °C, respectively, for CUGCGG and CGGCUG. The free energy increments, $\Delta\Delta G°_{37}$ (GU-AU) are 1.4 and 0.5 kcal/mol, respectively, even though both pairs of oligomers have identical nearest neighbors. This may be a special case. A similar destabilization is observed for UCAUGA vs. UGAUCA (Kierzek et al., 1987). UCAUGA and UGAUCA also have identical nearest neighbors, so there may be an unusual destabilization associated with YYRYGR sequences where Y is a pyrimidine and R a purine.

The results in Table III suggest the detailed effects of internal GU mismatches on helix stability may be complex. Even a nearest-neighbor approximation requires determination of 11 parameters for each thermodynamic property. Thus the oligomers in Table I are not sufficient for a complete nearest-neighbor analysis. They are sufficient, however, to improve on parameters currently in use for prediction of RNA structure (Tinoco et al., 1973; Gralla & Crothers, 1973). These parameters were derived from studies of A_4GUU_4, A_4UGU_4, and A_5UGU_5 (Uhlenbeck et al., 1971). Due to the limited sequences available, it was necessary to assume that all GU/AU nearest neighbor interactions are identical and that a GU/GC stack is more stable than GU/AU by the same amount that GC/AU is more stable than AU/AU (Gralla & Crothers, 1973). Omitting CUGCGG, which may be anomalous, there are seven oligomers in Table I that have GU mismatches. They contain seven different nearest-neighbor interactions involving GU. The data are only sufficient, however, to solve for six parameters involving GU. Thus two stacking interactions must be arbitrarily set equal. We have chosen to set $\genfrac{}{}{0pt}{}{5'\mathrm{GU}3'}{3'\mathrm{UA}5'} = \genfrac{}{}{0pt}{}{5'\mathrm{UG}3'}{3'\mathrm{AU}5'}$. With this assumption and with all non-GU interactions set equal to nearest-neighbor values determined from results on 45 oligomers (Freier et al., unpublished results), the data in Table I were analyzed by multiple linear regression (Bevington, 1969) upon the nearest-neighbor model to give the parameters in Table IV. Values derived by assuming $\genfrac{}{}{0pt}{}{5'\mathrm{GG}3'}{3'\mathrm{UC}5'} = \genfrac{}{}{0pt}{}{5'\mathrm{GG}3'}{3'\mathrm{CU}5'}$ and $\genfrac{}{}{0pt}{}{5'\mathrm{CG}3'}{3'\mathrm{GU}5'} = \genfrac{}{}{0pt}{}{5'\mathrm{GC}3'}{3'\mathrm{UG}5'}$ were similar for $\Delta G°_{37}$. Since the data set used to derive Table IV is small, we expect values to change as more data becomes available.

Four nearest-neighbor interactions involving GU mismatches are not represented in the oligomers in Table I: $\genfrac{}{}{0pt}{}{5'\mathrm{GG}3'}{3'\mathrm{UU}5'}$, $\genfrac{}{}{0pt}{}{5'\mathrm{GU}3'}{3'\mathrm{UG}5'}$, $\genfrac{}{}{0pt}{}{5'\mathrm{AG}3'}{3'\mathrm{UU}5'}$, $\genfrac{}{}{0pt}{}{5'\mathrm{GA}3'}{3'\mathrm{UU}5'}$. Until these parameters are measured, we recommend a value of −0.5 kcal/mol for $\Delta G°_{37}$ for each of these sequences. This is based on the average difference between the GU/AU interactions in Table IV, and the corresponding AU/AU interactions determined by Freier et al. (unpublished results).

Values of $\Delta G°_{37}$ for nearest neighbors containing GU mismatches next to AU and GC pairs range from −0.5 to −0.7 and −1.3 to −1.9 kcal/mol, respectively. The $\Delta G°_{37}$ values from parameters used most often for prediction of RNA structure are −0.3 and −1.5 kcal/mol, respectively, for GU/AU and GU/GC nearest neighbors (Gralla & Crothers, 1973; Tinoco et al., 1973). Thus, in general, GU mismatches appear to be more stable than previously realized. Values of $\Delta H°$ for nearest neighbors containing GU mismatches are very dependent on which nearest neighbors are set equal. The average $\Delta H°$ value is −6 kcal/mol when it is assumed that 5′GU3′ = 5′UG3′. Below 54 °C, previous values for $\Delta H°$ were −0.3 and +2.0 kcal/mol for GU/AU and GU/GC neighbors, respectively (Gralla & Crothers, 1973). Thus the temperature dependence of GU mismatches is larger than previously realized.

The results in Table III can be compared with results of Patel et al. (1982) on d(CGTGAATTCGCG) and d-(CGCGAATTCGCG), and of Aboul-ela et al. (1985) on the series $dCA_3XA_3G + dCT_3YT_3G$ where X and Y are G, C, or T. Both groups found that substitution of GT mismatches

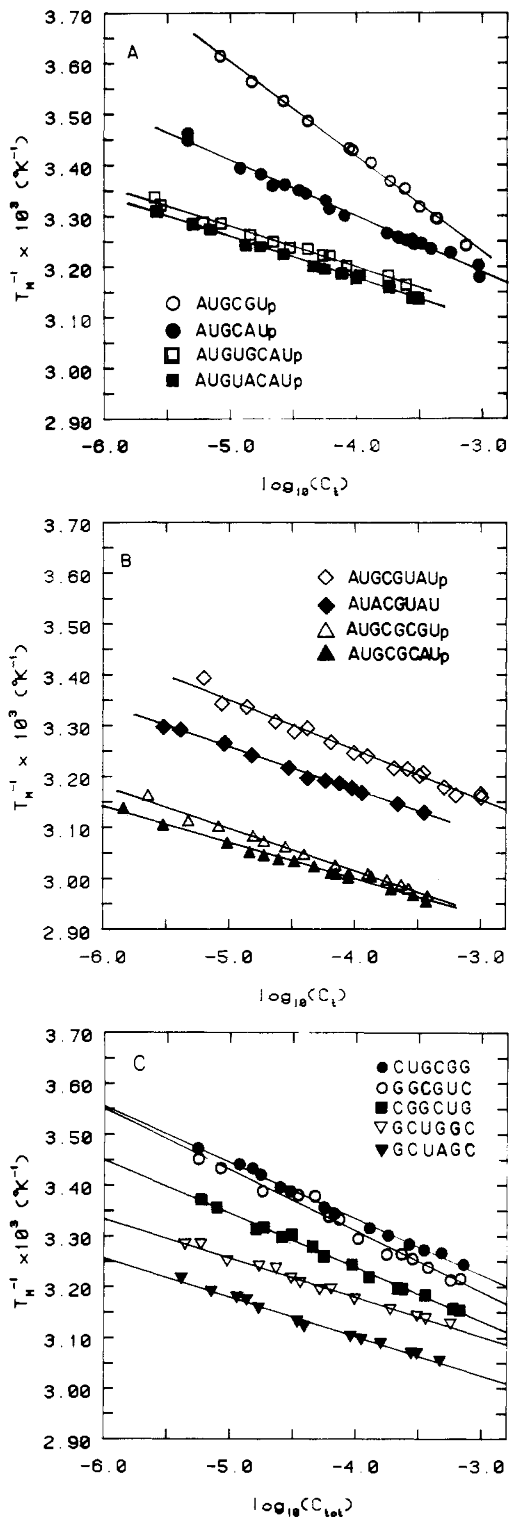

FIGURE 1: Plots of T_M^{-1} vs. log C_T: (A) AUGCGUp (○), AUGCAUp (●), AUGUGCAUp (□), AUGUACAUp (■); (B) AUGCGUAUp (◇), AUACGUAU (◆), AUGCGCGUp (△), AUGCGCAUp (▲); (C) GCUGGC (▽), GCUAGC (▼), GGCGUC (○), CUGCGG (●), CGGCUG (■).

for GC base pairs lowered the melting temperature about 20 °C. This destabilization is associated with a more unfavorable $\Delta S°$ of helix formation for the GT oligomer with a slightly more favorable $\Delta H°$ in the work of Patel et al. (1982), and with opposite trends in the work of Aboul-ela et al. (1985). For the RNA oligomers in Table III, substitution of GU for GC is always associated with a less favorable $\Delta H°$ of helix formation. Oligomers with isolated GU mismatches have a more favorable $\Delta S°$ of helix formation. For the adjacent GU mismatches in AUGUGCAU and GCUGGC, however, the data in Table II indicate $\Delta S°_{37}$ becomes less favorable relative to the equivalent GC pairs, suggesting adjacent GU mismatches may be conformationally more rigid than isolated GU mismatches.

The wobble hypothesis (Crick, 1966) attributes special stability to terminal GU mismatches. Freier et al. (1986b) have shown that terminal GU mismatches next to GC pairs are essentially equivalent to terminal AU pairs. The results in Table III indicate internal GU mismatches are an average of 0.7 kcal/mol of mismatch less stable than AU pairs. This is consistent with the wobble hypothesis.

The results presented here indicate that internal GU mismatches are more stable and more temperature dependent than previously believed. Previous predictions of properties at low temperature are particularly prone to error. Approximations are proposed that should improve predictions of the structures, stability, and thermodynamic parameters of RNA molecules. In addition, the results suggest that adjacent GU mismatches may be more constrained conformationally than isolated GU mismatches.

Registry No. G, 73-40-5; U, 66-22-8; A, 73-24-5; AUGCGUp, 103793-80-2; AUGCGCGUp, 103793-81-3; AUGCGUAUp, 103793-82-4; AUGUGCAUp, 103793-83-5; GCUGGC, 103793-84-6; GGCGUC, 103793-85-7; CUGCGG, 103793-86-8; CGGCUG, 103793-87-9; AUGCAUp, 103793-88-0; AUGCGCAUp, 103793-89-1; AUACGUAU, 103793-90-4; AUGUACAUp, 103793-91-5; GCUAGC, 103793-92-6; GACGUC, 103793-93-7; CUGCAGp, 103793-94-8; CAGCUGp, 103793-95-9; ACGCGU, 103816-27-9; ACGCGCGU, 103816-28-0; AGCU, 55048-61-8; AUCGAU, 103793-96-0; CCGCGG, 103816-29-1; CGCG, 89435-89-2; UG, 3474-04-2; 5′CG3′/3′GU5′, 103793-97-1; 5′GU3′/3′UA5′, 103793-98-2; 5′GC3′/3′UG5′, 103793-99-3; 5′GG3′/3′UC5′, 103794-00-9; 5′GG3′/3′CU5′, 103794-01-0.

REFERENCES

Aboul-ela, F., Koh, D., Tinoco, I., Jr., & Martin, F. H. (1985) *Nucleic Acids Res. 13*, 4811–4824.

Alkema, D., Hader, P. A., Bell, R. A., & Neilson, T. (1982) *Biochemistry 21*, 2109–2117.

Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.

Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843–853.

Brown, T., Kennard, O., Kneale, G., & Rabinovich, D. (1985) *Nature* (*London*) *315*, 604–606.

Cech, T. R., Tanner, N. K., Tinoco, I., Jr., Weir, B. R., Zuker, M., & Perlman, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3903–3907.

Crick, F. H. C. (1966) *J. Mol. Biol. 29*, 548–555.

Fox, G. E., & Woese, C. R. (1975) *Nature* (*London*) *256*, 505–506.

Freier, S. M., Burger, B. J., Alkema, D., Neilson, T., & Turner, D. H. (1983a) *Biochemistry 22*, 6198–6206.

Freier, S. M., Albergo, D. D., & Turner, D. H. (1983b) *Biopolymers 22*, 1107–1131.

Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., & Turner, D. H. (1985a) *Biochemistry 24*, 4533–4539.

Freier, S. M., Sinclair, A., Neilson, T., & Turner, D. H. (1985b) *J. Mol. Biol. 185*, 645–647.

Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., & Turner, D. H. (1986a) *Biochemistry 25*, 3214–3219.

Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986b) *Biochemistry 25*, 3209–3213.

Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 73*, 497–511.

Biochemistry **1986**, *25*, 5759–5765

Hickey, D. R., & Turner, D. H. (1985) *Biochemistry 24*, 2086–2094.

Johnston, P. D., & Redfield, A. G. (1981) *Biochemistry 20*, 1147–1156.

Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., & Freier, S. M. (1987) *Biochemistry* (in press).

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974) *Science* (*Washington, D.C.*) *185*, 435–440.

Ladner, J. E., Jack, A., Robertus, J. D., Brown, J. S., Rhodes, D., Clarke, B. F. C., & Klug, A. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4414–4418.

Markiewicz, W. T., Biala, E., & Kierzek, R. (1984) *Bull. Pol. Acad. Sci., Chem. 32*, 433–451.

Matteucci, M. D., & Caruthers, M. H. (1980) *Tetrahedron Lett. 21*, 719–722.

Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119–162.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Dallas, J., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 437–444.

Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 256–263.

Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology*: *Nucleic Acids* (Fasman, C. D., Ed.) 3rd ed., Vol. I, p 197, CRC, Cleveland, OH.

Romaniuk, P. J., Hughes, D. W., Gregoire, R. J., Neilson, T., & Bell, R. A. (1979a) *J. Chem. Soc., Chem. Commun.*, 559–560.

Romaniuk, P. J., Hughes, D. W., Gregoire, R. J., Bell, R. A., & Neilson, T. (1979b) *Biochemistry 18*, 5109–5116.

Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol. 42*, 985–1002.

Sprinzl, M., Möll, J., Meissner, F., & Hartman, T. (1985) *Nucleic Acids Res. 13*, r1–r49.

Steger, G., Hofmann, H., Förtsch, J., Gross, H. J., Randles, J. W., Sänger, H. L., & Riesner, D. (1984) *J. Biomol. Struct. Dyn. 2*, 542–571.

Sussman, J. L., & Kim, S.-H. (1976) *Biochem. Biophys. Res. Commun. 68*, 89–96.

Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973) *Nature* (*London*), *New Biol. 246*, 40–41.

Tsong, T. Y., Hearn, R. P., Wrathall, D. P., & Sturtevant, J. M. (1970) *Biochemistry 9*, 2666–2677.

Uhlenbeck, O. C., & Gumport, R. I. (1982) *Enzymes* (*3rd Ed.*) *15*, 31–58.

Uhlenbeck, O. C., Martin, F. H., & Doty, P. (1971) *J. Mol. Biol. 57*, 217–229.

Woese, C. R., Gutell, R., Gupta, R., & Noller, H. F. (1983) *Microbiol. Rev. 47*, 621–669.

Compact Units in Proteins[†]

Micheal H. Zehfus and George D. Rose*

Department of Biological Chemistry, The Milton S. Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania 17033

Received November 5, 1985; Revised Manuscript Received April 23, 1986

ABSTRACT: An explicit measure of geometric compactness called the coefficient of compactness is introduced. This single value figure of merit identifies those continuous segments of the polypeptide chain having the smallest solvent-accessible surface area for their volume. These segments are the most compact units of the protein, and the larger ones correspond to conventional protein domains. To demonstrate the plausibility of this approach as a method of identifying protein domains, the measure is applied to lysozyme and ribonuclease to discover their constituent compact units. These units are then compared with domains, subdomains, and modules found by other methods. To show the sensitivity of the method, the measure is used to successfully differentiate between native and deliberately misfolded proteins [Novotný, J., Bruccoleri, R., & Karplus, M. (1984) *J. Mol. Biol. 177*, 787–818]. Methods that utilize only backbone atoms to define domains cannot distinguish between authentic and misfolded molecules because their backbone conformations are virtually superimposable. Compact units identified by this method exhibit a hierarchic organization. Such an organization suggests possible folding pathways that can be tested experimentally.

The exposure of nonpolar residues to solvent is thermodynamically disfavored in both model compounds and proteins (Kauzmann, 1959; Nozaki & Tanford, 1971). As a result, the nonpolar residues of proteins tend to aggregate, minimizing the total hydrophobic surface area they expose to solvent. This phenomenon, known as the hydrophobic effect, is thought to be a major force driving the protein-folding process (Kauzmann, 1959; Tanford, 1980). Consistent with this idea, the average surface area that residues lose upon folding is found to scale linearly with hydrophobicity (Rose et al., 1985). For protein molecules then, a strong correlation exists between hydrophobicity and surface area. This correlation suggests, in turn, an underlying relationship between conformational free energy and molecular surface area.

In this paper, a normalized measure of molecular surface area is introduced. The parameter, called the coefficient of compactness, is applied to X-ray-elucidated proteins to identify those continuous-chain regions having the smallest accessible surface area for their volume. Geometrically, these units are compact, while, energetically, they have efficiently minimized their conformational free energy through the reduction of exposed surface area.

[†] This research was supported by grants from the National Institutes of Health [GM10824 (M.H.Z.) and GM29458 (G.D.R.)].

* Author to whom correspondence should be addressed.

0006-2960/86/0425-5759$01.50/0