Schellman, J. A. (1987) *Annu. Rev. Biophys. Chem. 16*, 115–137.
Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L., & Sigler, P. B. (1985) *Nature 317*, 782–786.
Schwarz, J., & Berget, P. B. (1989a) *Genetics 121*, 635–649.
Schwarz, J., & Berget, P. B. (1989b) *J. Biol. Chem. 264*, 20112–20119.
Seckler, R., Fuchs, A., King, J., & Jaenicke, R. (1989) *J. Biol. Chem. 264*, 11750–11753.
Smith, D. H., Berget, P. B., & King, J. (1980) *Genetics 96*, 331–352.
Sturtevant, J. M., Yu, M.-H., Haase-Pettingell, C., & King, J. (1989) *J. Biol. Chem. 264*, 10693–10698.
Tandon, S., & Horowitz, P. (1986) *J. Biol. Chem. 261*, 15615–15618.
Thomas, G. J., Becka, R., Sargent, D., Yu, M.-H., & King, J. (1990) *Biochemistry 29*, 4181–4187.
Villafane, R., & King, J. (1988) *J. Mol. Biol. 204*, 607–619.
Volkin, D. B., & Klibanov, A. M. (1989) in *Protein Function: a Practical Approach* (Creighton, T. E., Ed.) pp 1–24, IRL Press, Oxford.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697–701.
Wetzel, R., Perry, L. J., Mulkerrin, M. G., & Randall, L. M. (1990) in *Protein Design and the Development of New Therapeutics and Vaccines*, Smith Kline & French Laboratories Research Symposia Series (Hook, J. B., & Poste, G., Eds.) pp 79–115, Plenum Press, New York.
Wilson, I. A., Skehel, J. J., & Wiley, D. C. (1981) *Nature 289*, 366–373.
Winston, F. D., Botstein, D., & Miller, J. H. (1979) *J. Bacteriol. 137*, 433–439.
Yu, M.-H., & King, J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6584–6588.
Yu, M.-H., & King, J. (1988) *J. Biol. Chem. 263*, 1424–1431.

Thermodynamics of Antiparallel Hairpin–Double Helix Equilibria in DNA Oligonucleotides from Equilibrium Ultracentrifugation

Philip D. Ross,*,‡ Frank B. Howard,‡ and Marc S. Lewis§

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, and Biomedical Engineering and Instrumentation Program, National Center for Research Resources, National Institutes of Health, Bethesda, Maryland 20892

Received December 6, 1990; Revised Manuscript Received March 11, 1991

ABSTRACT: Five highly palindromic DNA dodecamers, four of which may form G-A or I-A purine–purine mispairs at either the 5,8 or 6,7 positions, have been studied at sedimentation equilibrium in the analytical ultracentrifuge. Each DNA oligonucleotide forms an equilibrium mixture of ordered antiparallel hairpin and double-stranded helical structures in solutions of 0.1 or 0.5 M NaCl between 5 and 40 °C. The dimeric duplex is favored by conditions of high salt and low temperature. The monomer–dimer equilibrium constants vary from 5×10^6 to 5×10^3 and are unique for each DNA dodecamer. Analysis of the temperature dependence of the equilibrium constants shows that the double helix to hairpin conversion is driven by a positive entropy change and is associated with an endothermic enthalpy change. The mispair substitutions at the 5,8 positions and the IA(6,7) mispair have the greatest tendency toward hairpin formation and exhibit significantly larger entropy changes than the nonmispaired dGGTACGCGTACC parent sequence and the thermodynamically similar GA(6,7) DNA. The consequences of such hairpin–double helix equilibria must be considered in the interpretation of other kinds of experiments carried out on oligonucleotides at different concentrations.

In recent years the production of DNA oligomers of defined nucleotide sequence has become commonplace due to advances in the methods of chemical synthesis and the availability of automated machines for the preparation of such molecules. The exact sequences of binding sites for biologically relevant proteins such as repressors and restriction endonucleases may be readily reproduced, and the effect of variation of sequence upon these and other interactions such as drug binding may be examined. Recent X-ray studies of crystals and two-dimensional nuclear magnetic resonance of oligonucleotides in solution have revealed a wide variety of structures that indicate considerable backbone flexibility and local sequence-dependent variation in DNA structure. Many of the properties of oligonucleotides such as their conformation, their mechanism of melting, and their interactions with ionic components in solution are highly chain-length dependent since oligonucleotides occupy the chain-length-sensitive transition region between small molecules and infinitely long-chain polyelectrolytes. It is clear that, on account of both their importance and their complexity, oligonucleotides should be examined from as many points of view as possible.

A spectroscopic and calorimetric study of five highly palindromic DNA dodecamers, four of which were designed with the potential of forming purine–purine mispairs in a duplex helical structure (F. B. Howard, C.-q. Chen, P. D. Ross, and H. T. Miles, manuscript in preparation) has recently been carried out. Based upon the conventional assumption of a two-state process, completely unreasonable values were derived for the thermodynamic parameters associated with the thermally induced melting transition. Since it has been known

* To whom correspondence should be addressed.
‡ National Institute of Diabetes and Digestive and Kidney Diseases.
§ National Center for Research Resources.

for over twenty years (Scheffler et al., 1968) that DNA oligonucleotides are capable of forming ordered helical hairpin structures in solution, we decided to examine these dodecamers in the analytical ultracentrifuge. This paper illustrates the capabilities of equilibrium ultracentrifugation in describing the nature and composition of these DNA dodecamers in solution and the thermodynamic parameters characterizing their associative equilibria.

Experimental Procedures

Materials

The DNA oligonucleotides used in this study were synthesized manually by Dr. C.-q. Chen using the cyanoethyl phosphoramidite method and were purified by DEAE-cellulose chromatography as described elsewhere (Howard et al., 1991). Purity was established by the observation of the appearance of a single band upon electrophoresis in 8 M urea on a 20% acrylamide gel. The parent Watson–Crick oligomer had the sequence 5′-GGTACGCGTACC-3′ and was designated WC. The other oligomers, designated by abbreviation according to their mispairs (in parentheses), had the sequences GGTACGAGTACC (GA(6,7)), GGTACIAGTACC (IA-(6,7)), GGTAAGCITACC (AI(5,8)), and GGTA-AGCGTACC (AG(5,8)). The DNA concentrations were determined with use of ultraviolet extinction coefficients (Howard et al., 1991) based upon phosphorus analysis (Muraoka et al., 1980). The oligonucleotide solutions used in the ultracentrifugation had an optical density at 280 nm of approximately 0.1 and 0.2 and were prepared in either 0.1 or 0.5 M NaCl containing 0.01 M sodium cacodylate buffer at pH 7.0.

Analytical Ultracentrifugation

Analytical ultracentrifugation was performed with use of a Beckman Model E analytical ultracentrifuge equipped with a scanning absorption optical system. Data were acquired from the scanner output with use of a Metrabyte DAS-8 12-bit analog to digital converter in a 10-MHz AST Premium/286 computer. The rapid scan rate was used and 90000 data points were acquired in the 18 s required to scan from the outer reference hole to the inner reference hole of the counterbalance. Each recorded point was the average of 100 acquired points; the actual data density was 42.5 points/mm of radial distance in the cell. Rapid data acquisition minimized the effects of rotor precession; individual scans were reproducible within the noise of the system (approximately 0.004 OD units at 280 nm). The inner and outer reference points and the region of interest were selected in the editing of the scan with the acquisition software, and the data were saved in the form of millivolts as a function of radial position.

At the time of analysis, the data were converted to the form of optical density at 280 nm as a function of radial position with use of a conversion factor obtained by calibration with solutions of cytidine monophosphate in 0.01 M HCl + 0.10 M NaCl. The low molecular weight of the cytidine monophosphate and immediate data collection precluded any significant solute redistribution. A linear response of millivolts output as a function of optical density at 280 nm was obtained over the range of 0–1.0 optical density units at the rotor speeds used in these experiments. The monochromator slit width was set at 0.4 mm in order to reduce stray light since 280 nm, the wavelength at which most of the experiments were performed, is away from the maximum in the absorption spectrum. This was the smallest monochromator slit width that could be used without significant degradation in the signal to noise ratio for the photomultiplier slit width set at 0.10 mm. Further decreasing the monochromator slit width would have required increasing the photomultiplier slit width with a corresponding loss of spatial resolution in the radial direction. The identical results obtained at 265 nm, described in Results, support the validity of this approach.

With one exception, run at 36000 rpm, all experiments were run at a rotor speed of 30000 rpm and with loading concentrations of approximately 0.1 and 0.2 OD units at 280 nm for each oligonucleotide. It has been demonstrated that the simultaneous fitting of data from experiments involving more than one rotor speed or more than one cell loaded with different concentrations or both, requiring the simultaneously fit models to have equilibrium constants as global fitting parameters and cell reference concentrations and base-line error terms as local fitting parameters, represents the most stringent criterion possible for establishing the validity of what appears to be a reversibly associating system (Roarke, 1975). The time required for reequilibration following speed changes significantly increases the time required for the experiment. When the values of the equilibrium constants are determined as a function of temperature, the time required for the complete experiment is approximately halved if a single rotor speed and more than one initial loading concentration of the desired solution are used. Two different initial loading concentrations proved sufficient for the solutions examined here.

All of the experiments were performed with use of a six-hole rotor with five cells and a counterbalance. Quartz windows and carbon-filled epoxy double-sector centerpieces with a 12-mm optical path length were used in the cells. The solution column lengths were approximately 7 mm. Two concentrations each of two oligonucleotides and one concentration of a third oligonucleotide were run in each experiment. The temperature-dependence series of experiments was begun at a temperature of 5 °C. Since the concentration gradients appeared to be invariant after 24 h of centrifugation, the final scans at this temperature were taken after 40–48 h. The temperature was then increased by 5 °C, and scans were taken after approximately 24 h, which was a more than sufficient period of time for reequilibration since only small perturbations in the gradients were introduced by the 5 °C temperature change. The temperature was sequentially raised, and scans were taken until deposition of oil vapor on optical surfaces within the vacuum chamber created sufficient noise in the data that the experiment had to be terminated. Terminal temperatures varied from 30 to 40 °C.

Results

Following data acquisition, further data manipulation was performed with use of software written specifically for that purpose and data analysis was performed with use of PC MLAB operating on the data acquisition computer. The ease of examining the suitability of different fitting models and the ability to fit more than one data set simultaneously and to generate appropriate weights for fitting each data set make this program well suited for the analysis of ultracentrifugal data. A value of 1×10^{-6} was selected as the convergence limit for terminating iteration during fitting after finding that smaller values did not result in any difference in the parameter values obtained. A minimum value of the root mean square (rms) error, the absence of systematic deviation in a plot of the distribution of the residuals about the fitting line, and the requirement that all parameter values be physically meaningful were the criteria used in selecting the most appropriate model when fitting the data.

The WC oligomer was first centrifuged at 20 °C in 0.1 M NaCl, and the resultant data was analyzed as a single-com-

ponent system. The molecular weight obtained from this analysis lay between the expected value of 3888 for monomer and 7776 for dimer; this and the fact that the distribution of the residuals from the fitting was poor and showed systematic deviations strongly indicated that we were observing a mixture of monomers and higher polymers with the possibility that this was a reversible equilibrium. Similar results yielding molecular weights intermediate between those of monomer and dimer were obtained with all of the other oligomers with the one-component model. The data was then fit with the model for a reversible monomer–dimer equilibrium

$$c_r = c_{b,1} \exp(AM_1(r^2 - r_b^2)) + k_{12}c_{b,1}^2 \exp(2AM_1(r^2 - r_b^2)) + \epsilon \quad (1)$$

where c_r is the total concentration at the radial position r in the cell, $c_{b,1}$ is the concentration of monomer at the radial position of the cell bottom, M_1 is the molecular weight of Na monomer (3888), k_{12} is the equilibrium constant for dimer formation based on the observed optical density concentration scale ($k_{12} = c_2/c_1^2$), and ϵ is a small base-line error correction term. The term A is given by

$$A = (\partial\rho/\partial c)_\mu \omega^2/2RT \quad (2)$$

where $\partial\rho/\partial c$ is the increment of density with concentration at constant chemical potential of solvent components and has the values of 0.472 and 0.446 for 0.1 and 0.5 M NaCl, respectively, at 25 °C (Eisenberg, 1990), ω is the rotor angular velocity, R is the gas constant (8.314×10^7 ergs deg^{-1} mol^{-1}), and T is the absolute temperature. Since the values of A and M_1 were known, only k_{12}, $c_{b,1}$ and ϵ were fitting parameters. The quality of fit obtained with this model was quite good. Additionally, measuring at a wavelength of 265 nm, near the absorption maximum, gave k_{12} values identical with those obtained at 280 nm where the absorption spectrum was steeper. The wavelength of 280 nm was used because this analytical ultracentrifuge gave an rms error of 0.004 OD units for a uniform initial distribution with an OD of 0.2 at that wavelength while double that rms error as well as approximately double the initial OD were observed at 265 nm.

When all of the oligonucleotides were examined at 10 and 20 °C and at 30 000 and 36 000 rpm, it was found that for each oligonucleotide at a given temperature the value of the dimerization constant, k_{12}, obtained by the simultaneous fitting of data from two rotor speeds was identical with the value obtained in other experiments from the simultaneous fitting of two initial loading concentrations at a single rotor speed at that temperature. These results demonstrate the equivalence of multiple rotor speed and multiple loading concentration experiments and establish that these monomer–dimer reactions are reversible. Thermal reversibility was also established; i.e., the values of k_{12} were independent of the order of the temperatures at which the measurements were made.

We next examined the possibility of thermodynamic nonideality and attempted fitting with a model that was a modification of eq 1, defining $M_{app,c} = M/(1 - Bc)$, which expresses the nonideality in terms of a single virial coefficient, B. We uniformly obtained zero values for B as a fitting parameter, thus indicating that no apparent nonideality was observed. Attempts made to fit with models involving higher aggregates in the presence and absence of dimer resulted in a worsening of the fit in every case.

Having demonstrated reversibility, the absence of virial coefficient effects, and poorer fits by assuming the presence of higher aggregates, we adopted an ideal reversible monomer–dimer association as the most appropriate model for describing the behavior of these oligonucleotides.

Since the values of A and M_1 were known, a simple transformation of the independent variable that introduces no error by distorting the weighting of the points enabled us to enhance the rate of computation 4-fold. Defining a function, $u(r)$, as

$$u(r) = \exp(AM_1(r^2 - r_b^2)) \quad (3)$$

eq 1 could then be written as a quadratic polynomial

$$c_u = c_{b,1}u + k_{12}c_{b,1}^2u^2 + \epsilon \quad (4)$$

that could then be used as a fitting function after transforming the data from the form of concentration as a function of radius to concentration as a function of the value of u. In order to ensure obtaining nonnegative values for the equilibrium constants and to optimize the shape of the minimum in the surface of the sum of squares as a function of the parameter values, it is desirable to use ln k_{12} as a fitting parameter. Equation 4 then has the form

$$c_u = c_{b,1}u + \exp(\ln k_{12})c_{b,1}^2u^2 + \epsilon \quad (5)$$

The saving of computational time that the use of this model effected was particularly significant when simultaneous fits of two data sets were performed, where a fitting function was written for each data set, where the values of $c_{b,1}$ and ϵ were fitting parameters specific to each set and the value of ln k_{12} was a global fitting parameter. In all cases an estimated weighting function was generated in MLAB by multiple smoothing of a data set, estimating the variance of each point from its deviation from the smoothed value, and using the reciprocal of the variance as a weight. Using weighting is particularly important when performing simultaneous fits where the range of values within each data set differ significantly since failure to do so results in the data set with values of greater magnitude having an undue influence upon the parameter values.

The range of behavior observed in this work is shown in Figures 1A and 2A. Figure 1A illustrates the normalized concentration distributions of the Watson–Crick oligomer, GGTACGCGTACC, in 0.5 M NaCl at 10 °C, a temperature at which most of the DNA was in the dimeric form and profound quadratic behavior of the distributions is observed. In contrast, Figure 2A illustrates the normalized concentration distributions of the AG(5,8) mispair, GGTAAGCGTACC, in 0.1 M NaCl at 25 °C, where all of the DNA appears to be in the monomeric form and the concentration distributions are linear. The graphs illustrating the distribution of the residuals (Figures 1B and 2B) demonstrate the absence of any significant systematic deviation and represent rms errors lying between 0.005 and 0.007 OD units. These values are noteworthy since they are less than twice the intrinsic rms error associated with the data acquisition system of the ultracentrifuge at 280 nm.

The values of ln k_{12} obtained in the fitting procedure using eq 5 are based on an optical density concentration scale and must be transformed to the molar concentration scale for thermodynamic calculations. This is accomplished with use of the relationship

$$K'_{12} = C_2/C_1^2 = k_{12}E_1/2 \quad (6)$$

where K'_{12} is the equilibrium constant on the molar concentration scale (C) and E_1 is the molar extinction coefficient of the oligonucleotide (Howard et al., 1991). This formulation assumes that the value of the molar extinction coefficient of the dimer is exactly twice that of the monomer; the fact that

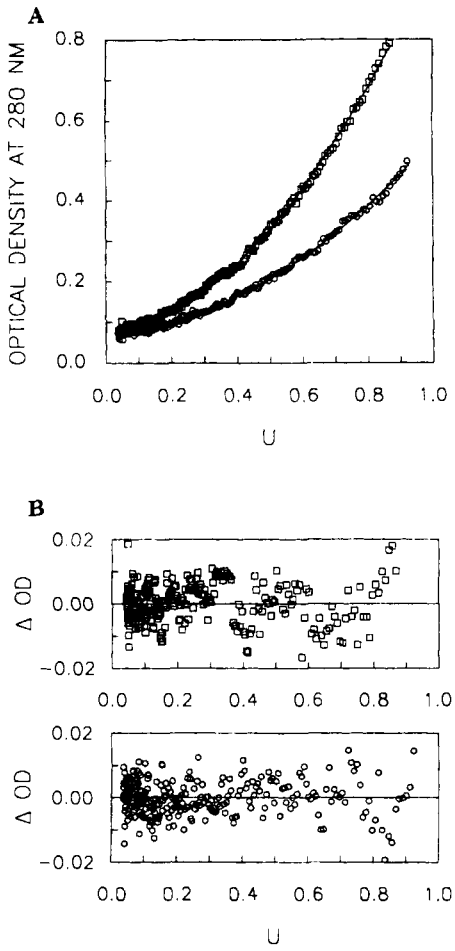

FIGURE 1: (A) Concentration distributions of the Watson–Crick oligomer, GGTACGCGTACC, as a function of normalized distance in the cell (see text) in 0.5 M NaCl at 10 °C and 30 000 rpm. The fitting lines are given by eqs 5 and 6 with a common best fit value of $\ln K'_{12} = 15.108$. The upper curve is for an initial loading of 0.186 OD and the lower curve for an initial loading of 0.090 OD at 280 nm. (B) Distribution of the residuals of the fits to the data shown in (A). The upper panel is for the 0.186 OD loading; the lower panel is for the 0.090 OD loading. These distributions correspond to rms errors lying between 0.005 and 0.006 OD units at 280 nm, which should be compared to the intrinsic noise of the system of 0.004 OD units at this wavelength.

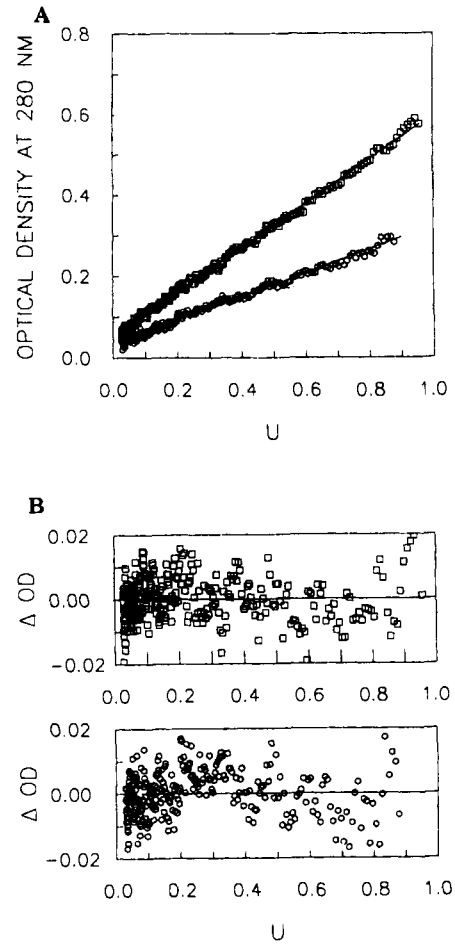

FIGURE 2: (A) Concentration distributions of the AG(5,8) mispaired oligomer, GGTAAGCGTACC, as a function of normalized distance in the cell in 0.1 M NaCl at 25 °C and 30 000 rpm. The fitting lines are given by eqs 5 and 6 with a common best fit value of an indeterminately large negative value of $\ln K'_{12}$ ($K'_{12} = 0$). The upper curve is for an initial loading of 0.199 OD and the lower curve for an initial loading of 0.102 OD at 280 nm. (B) Distribution of the residuals of the fits to the data shown in (A). The upper panel is for the 0.199 OD loading; the lower panel is for the 0.102 OD loading. These distributions correspond to rms errors lying between 0.006 and 0.007 OD units at 280 nm, which should be compared to the intrinsic noise of the system of 0.004 OD units at this wavelength.

the DNA solutions exhibit minimal change in optical density over the temperature range of the study while the ratio of monomer to dimer changes radically validates this assumption. K'_{12} is converted to the dimensionless thermodynamic equilibrium constant K_{12} by choice of standard states. The values of the thermodynamic monomer–dimer equilibrium constants, K_{12}, and their associated standard errors as obtained by equilibrium ultracentrifugation are reported in Table I. The average uncertainty in K_{12} was about 6% for the oligonucleotides in 0.1 M NaCl and about 9% for the oligonucleotides in 0.5 M NaCl.

The temperature dependence of the equilibrium constant was analyzed by the method of Clarke and Glew (1966) using the Taylor's series expansion about the reference temperature, θ

$$R \ln K_{12} = \frac{-\Delta G^\circ_\theta}{\theta} + \Delta H^\circ_\theta\left[\frac{1}{\theta} - \frac{1}{T}\right] + \Delta C_p^\circ{}_\theta\left[\frac{\theta}{T} - 1 + \ln\left(\frac{T}{\theta}\right)\right] \quad (7)$$

This form of the expression describing the temperature dependence of the equilibrium constant has the desirable property that the fitted parameters ΔG°_θ, ΔH°_θ, and $\Delta C_p^\circ{}_\theta$ are independent of one another. The reference temperature θ was taken to be 293.15 K, approximately in the middle of the temperature range of the experiments, in order to minimize the standard errors of the derived thermodynamic parameters.

When eq 7 was used for fitting the values of $R \ln K_{12}$ as a function of temperature, the data were weighted by the normalized reciprocals of the variances of $\ln K_{12}$ obtained when fitting the ultracentrifuge concentration data. Although no statistical improvement in fit was obtained for the oligonucleotide IA(6,7) using a nonzero value for $\Delta C_p^\circ{}_\theta$, in most of the other cases a statistically significant improvement in fit was found with a nonzero value for this parameter. Our ultracentrifuge data indicates nonzero values for $\Delta C_p^\circ{}_\theta$ since their magnitudes are several times greater than those that would be expected to arise from propagation of error in the values of the equilibrium constants (King, 1965). It is known that finite but small heat capacity changes accompany structural transitions in nucleic acids (Filimonov, 1986). The values of all of the thermodynamic parameters and their standard errors are presented in Table II, and the three-parameter fits to the weighted data are shown in Figure 3, panels

Table I: Monomer–Dimer Association Constants[a] of DNA Dodecamers from Equilibrium Ultracentrifugation[b]

T(°C)	WC	GA(6,7)	IA(6,7)	AI(5,8)	AG(5,8)
			$10^{-4}K_{12}$ in 0.1 M NaCl[c]		
5	66.8 ± 4.1	49.4 ± 3.2	5.28 ± 0.28	59.3 ± 2.1	24.9 ± 1.8
10	80.5 ± 5.4	45.7 ± 2.4	4.30 ± 0.22	44.9 ± 2.8	12.7 ± 0.6
15	70.1 ± 4.5	31.1 ± 1.4	1.63 ± 0.17	21.4 ± 1.4	5.74 ± 0.30
20	46.4 ± 2.2	25.8 ± 1.1	0.90 ± 0.11	13.6 ± 0.8	2.38 ± 0.16
25	23.0 ± 1.1	14.1 ± 0.6	0.56 ± 0.25	5.84 ± 0.46	all monomer
30	13.2 ± 0.9	9.20 ± 0.46	all monomer	2.82 ± 0.24	all monomer
			$10^{-4}K_{12}$ in 0.5 M NaCl[c]		
5	426 ± 61	292 ± 76	50.1 ± 4.1	68.2 ± 5.5	77.5 ± 5.9
10	364 ± 58	274 ± 36	38.7 ± 3.6	63.5 ± 5.9	51.6 ± 2.7
15	252 ± 34	217 ± 21	19.2 ± 1.3	39.1 ± 3.4	22.7 ± 1.0
20	183 ± 19	175 ± 18	11.3 ± 0.8	27.0 ± 2.1	11.9 ± 0.4
25	141 ± 15	161 ± 17	7.09 ± 0.36	18.9 ± 0.9	6.14 ± 0.24
30	86.7 ± 8.0	106 ± 10	3.33 ± 0.25	10.7 ± 0.8	1.53 ± 0.10
35	49.5 ± 6.3	58.7 ± 7.5			
40	36.1 ± 5.2	40.7 ± 6.0			

[a] Reaction: 2 M ⇌ D. [b] The standard state was chosen to be 1 mol L^{-1} of oligonucleotide monomer in each solvent. [b] All solutions also contained 0.01 M sodium cacodylate and were at pH 7.0. Indicated uncertainties are the estimated standard deviations from the combined weighted fits of the concentration distributions.

Table II: Thermodynamic Parameters of Monomer–Dimer Association[a]

oligonucleotide	$\Delta G°_{20°}$ (kcal mol^{-1})	$\Delta H°_{20°}$ (kcal mol^{-1})	$\Delta C_p°_{20°}$ (cal mol^{-1} deg^{-1})	$\Delta S°_{20°}$ (cal mol^{-1} deg^{-1})	$\Delta G°_{7.5°}$(calcd) (kcal mol^{-1})
			0.1 M NaCl[b]		
WC	−7.59 ± 0.03	−16.3 ± 1.2	−1530 ± 240	−30 ± 4	−7.53
GA(6,7)	−7.21 ± 0.03	−13.6 ± 1.0	−700 ± 220	−22 ± 3	−7.29
IA(6,7)	−5.41 ± 0.12	−26.9 ± 6.8	−1200 ± 910	−73 ± 30	−6.01
AI(5,8)	−6.85 ± 0.04	−23.2 ± 1.2	−800 ± 910	−56 ± 4	−7.34
AG(5,8)	−5.87 ± 0.01	−31.9 ± 0.4	−850 ± 50	−89 ± 1	−6.74
			0.5 M NaCl[b]		
WC	−8.42 ± 0.02	−11.9 ± 0.5	−390 ± 80	−12 ± 2	−8.47
GA(6,7)	−8.42 ± 0.03	−8.5 ± 0.6	−560 ± 110	−0.3 ± 2	−8.29
IA(6,7)	−6.81 ± 0.03	−19.3 ± 1.1	−470 ± 240	−43 ± 4	−7.25
AI(5,8)	−7.32 ± 0.03	−14.1 ± 0.9	−580 ± 190	−23 ± 3	−7.47
AG(5,8)	−6.83 ± 0.05	−27.4 ± 2.2	−1060 ± 480	−70 ± 8	−7.43

[a] The standard state was chosen to be 1 mol L^{-1} of oligonucleotide monomer in each solvent. Unitary changes in free energy and entropy, in which the entropy of mixing contribution arising from the choice of concentration units is eliminated, can be obtained by adding 7.98 cal mol^{-1} deg^{-1} to each standard entropy change, $\Delta S°_{20°}$, reported above and making each standard free energy change, $\Delta G°_{20°}$, 2.34 kcal mol^{-1} more negative. [b] All solutions also contained 0.01 M sodium cacodylate and were at pH 7.0. Uncertainties are the estimated standard deviations from the weighted three-parameter fits using eq 7.

A–E. The lines drawn in Figure 3, panels A–E, show the excellent fits of eq 7 to the equilibrium constant data for both salt concentrations studied.

Discussion

The primary finding of this study is that *all* of these highly palindromic DNA dodecamers form *equilibrium mixtures* of hairpin and double-stranded helical structures in solution at temperatures below their respective main order–disorder transitions. It is of considerable interest that in the first paper to describe oligonucleotide hairpin helices, Scheffler et al. (1968), employing equilibrium ultracentrifugation, reported a continuous decrease in relative molecular weight of nearly a factor of 2 for $d(TA)_{19}$ in the region below the main optical transition. This behavior is very similar to what we are reporting for these dodecamers.

The position and temperature dependence of the monomer–dimer helix equilibrium are unique for each dodecamer studied. It is seen in Table I that the values of the monomer–dimer equilibrium constant, K_{12}, are different for the various oligonucleotides and decrease to different extents with increasing temperature. The equilibrium ultracentrifugation measurements reported in this paper were mostly carried out at temperatures below the onset of the cooperative order–disorder transition as determined by ultraviolet absorption spectroscopy of oligonucleotide solutions at the same concentration (Howard et al., 1991). In this pretransition region, IR studies clearly indicate base-pairing (F. B. Howard, C.-q. Chen, P. D. Ross, and H. T. Miles, manuscript in preparation), while there is little or no change in the UV absorbance. Therefore, in this region, we identify the species of molecular weight 7776 as a dimeric double helix and the species of half this molecular weight as a monomeric intramolecular hairpin. Thus, the technique of equilibrium ultracentrifugation reveals extensive information about the composition of these oligonucleotide solutions in the pretransition region. The tendency for double-helix formation is greatest at higher salt concentration and lower temperatures, whereas the monomeric species are favored at lower salt concentration and higher temperatures. These results are consistent with earlier studies (Scheffler et al., 1968; Marky et al., 1983; Xodo et al., 1988) that indicated that an antiparallel intramolecular hairpin is the more stable ordered form at higher temperatures.

The signs of the thermodynamic parameters characterizing these DNA dodecamers reported in Table II show that hairpin formation (i.e., reversing signs given in Table II) is a wholly entropically driven process that is opposed by an unfavorable change in enthalpy. Recent computer simulations at low ionic strength reported by Olmsted et al. (1989) show a marked decrease in counterion concentration at the surface of an oligonucleotide upon halving the chain length, N, for the values of N relevant to this study. Assuming that the charge spacing

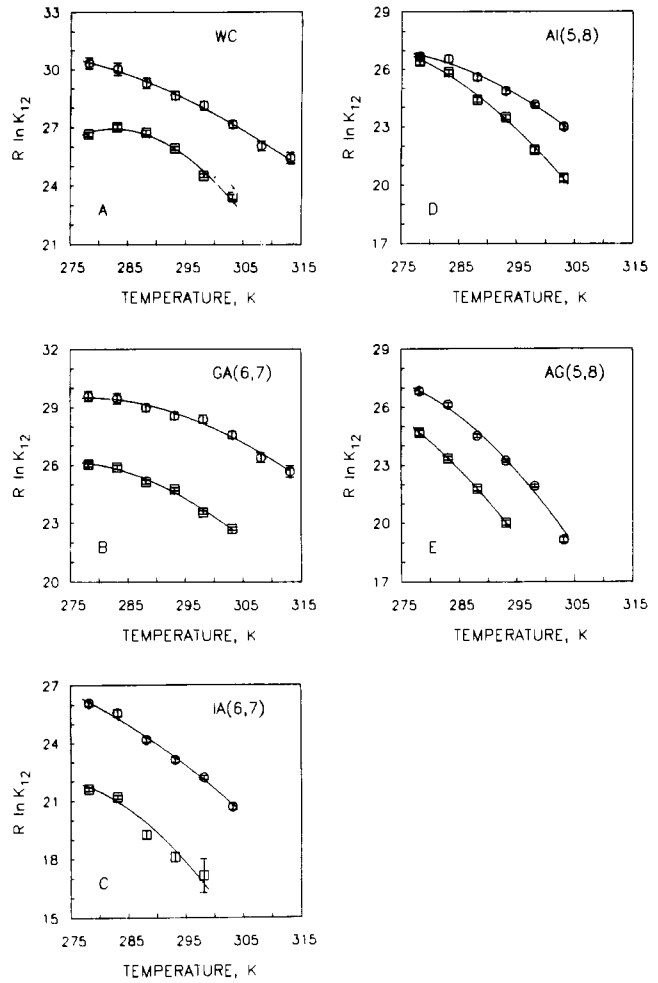

FIGURE 3: Fits of $R \ln K_{12}$ as a function of temperature in terms of the parameters $\Delta G°_{\theta}$, $\Delta H°_{\theta}$, and $\Delta C_p°_{\theta}$ in eq 7. The upper curves (O) refer to the oligonucleotides in 0.5 M NaCl; the lower curves (□) refer to the oligonucleotides in 0.1 M NaCl. The error bars indicate the estimated standard error for each value. Oligonucleotides: (A) GGTACGCGTACC, (B) GGTACGAGTACC, (C) GGTACIAGTACC, (D) GGTAAGCITACC, (E) GGTAAGCGTACC.

and cylindrical radii are the same for the double helix and the monomeric hairpin structures, these calculations would indicate that the formation of hairpin from double helix will be accompanied by a substantial release of counterions that could constitute a major source of the observed positive entropy change driving the conversion of double helix to ordered hairpin structure.

The magnitudes of the thermodynamic parameters shown in Table II fall roughly into two groups: the parent Watson–Crick base-paired dodecamer, WC, and the GA(6,7) mispair, which are the strongest double helix formers, are associated with smaller entropy and enthalpy changes while the three remaining dodecamers, which have a greater tendency for hairpin formation, are associated with considerably larger changes in enthalpy and entropy. Specifically, in 0.1 M NaCl, the parent Watson–Crick base-paired dodecamer, WC (Figure 3A), and the GA(6,7) mispair (Figure 3B) are mainly double helices up to 30 °C ($K_{12} > 10^5$), while in contrast the IA(6,7) and the AG(5,8) mispairs (Figure 3C,E) are essentially monomeric above 20 °C ($K_{12} < \approx 10^4$). The AI(5,8) dodecamer (Figure 3D) is mainly helical at low temperatures, resembling WC and GA(6,7), but above room temperature becomes primarily monomeric similar to IA(6,7) and AG(5,8). In 0.5 M NaCl, WC and GA(6,7) are again predominantly double helical at all temperatures studied (Figure 3A,B), and again both IA(6,7) and AG(5,8) are primarily monomeric at room temperature and above (Figure 3C,E). AI(5,8) most resembles IA(6,7) and AG(5,8) in 0.5 M NaCl but with somewhat reduced hairpin-forming tendencies (Figure 3D).

In order to examine the effects of the variation of chemical structure and the location of the mispairs upon the hairpin to double helix reaction, it is desirable to compare the dodecamers under conditions approximating a constant environment as far removed as possible from the main order–disorder transition so that any contributions from that phenomenon are minimized. The results in Table I suggest that the monomer–dimer equilibrium constant changes least at temperatures below 10 °C where the helix form is most stable. In order to obviate the use of uncertain $\Delta C_p°$ values, we have averaged the values of K_{12} determined at 5 and 10 °C and using $\Delta G° = -RT \ln K_{12}$ have obtained the standard-state Gibbs free energy change associated with the hairpin to double helix equilibrium at 7.5 °C. These Gibbs free energy changes presented in the last column of Table II are close to the values obtained at 20 °C from the global fit of all the data; this reflects the general insensitivity of the free energy function to temperature. In both 0.1 and 0.5 M NaCl, the free energy changes at 7.5 °C are very similar for all of the dodecamers and the values all lie within the narrow range of 1.5 kcal mol^{-1}, which corresponds to a maximum difference of only a factor of 12 in K_{12} between the most helical dodecamers, WC and GA(6,7), and the least helical dodecamer, IA(6,7). These extremely small differences in Gibbs free energy change ($\Delta\Delta G°$) render difficult any interpretation of these thermodynamic parameters in terms of possible structural origin.

The introduction of two GA purine–purine mismatches at the 6 and 7 positions in the center of the dodecamer introduces virtually no thermodynamic perturbation with respect to the normally paired WC DNA and reduces the free energy of the helix–hairpin equilibrium by only 200 cal mol^{-1}. However, introduction of AG mispairs at the 5,8 positions results in changes of Gibbs free energy of about 1000 cal mol^{-1} in both 0.1 and 0.5 M NaCl, demonstrating the marked effect of the sequence position of the mispair for these compositional isomers. In the case of IA substitution, the hairpin is most favored by the 6,7 substituent, while the 5,8 AI substituent is most similar to the GA(6,7) dodecamer in 0.1 M NaCl and displays an intermediate tendency toward hairpin formation in relation to the other dodecamers in 0.5 M NaCl. These results demonstrate both specific base substituent effects and sequence-specific effects in determining the helix–hairpin propensities of these dodecamers, which are considered in detail elsewhere (Howard et al., 1991; F. B. Howard, C.-q. Chen, P. D. Ross, and H. T. Miles, manuscript in preparation).

The fact that for room temperature and above, the value of the hairpin–dimer equilibrium constant, K_{12}, is of the order of 10^5 has important implications for these and similar oligonucleotides. If we assume a value of $K_{12} \approx 10^5$, it is readily seen that at low concentrations, of 10^{-5} M, such as used in ultraviolet spectrophotometry, one has an approximately equimolar population of helical dimers and monomeric hairpins, whereas at the higher concentrations that are employed in calorimetric, NMR, or infrared studies (e.g., 10^{-4} M) one might have a mixture that is 90% helix and 10% hairpin. It appears to be a widespread practice to relate the values of apparent "thermodynamic" parameters derived either from the shape or concentration dependence of UV melting curves obtained at low oligonucleotide concentrations on the basis of an assumed model (usually a two-state helix to coil or hairpin

to coil melting process) to structural features deduced (from, e.g., two-dimensioanl NOE studies of the same oligonucleotides) at concentrations at least an order of magnitude higher where the composition of the system is markedly different. Thus, one may be attributing apparent thermodynamic parameters characteristic of monomeric hairpins or of hairpin–helix mixtures with structural features associated with a double helical conformation. Additionally, on account of the temperature dependence of the helix to hairpin equilibrium, the composition of the system itself changes continuously throughout the melting process and the observed thermal transition may not be correctly defined in terms of simple two-state or sequential helix to hairpin to coil melting models, thus undermining the significance of the values of the thermodynamic parameters deduced from such assumptions. Consequently, great caution is warranted in attributing small differences in the values of derived thermodynamic parameters as arising from structural perturbation induced by changes either of chemical composition or of base sequence. In order to overcome these difficulties, we suggest that oligonucleotide systems first be characterized by molecular weight distribution studies such as the ultracentrifuge investigations described in this paper in order to establish the composition of the oligonucleotide solutions that are subsequently examined by other techniques. In a future publication describing a detailed calorimetric and spectroscopic study of these dodecamers (F. B. Howard, C.-q. Chen, P. D. Ross, and H. T. Miles, manuscript in preparation), we will present a more complete thermodynamic analysis of these systems that is based in part upon the dimer–monomer equilibria reported here.

Acknowledgments

We thank Dr. W. A. Hagins, NIDDK, NIH, and Richard Shrager, DCRT, NIH, for helpful dicussions concerning various aspects of the fitting procedures. We also thank Drs. Barry Bunow and Gary D. Knott of Civilized Software, Bethesda, MD, for making available several versions of PC MLAB for the data analysis and for helpful discussions concerning its use.

Registry No. WC, 113341-04-1; (GA(6,7)), 130641-72-4; (IA(6,7)), 130641-68-8; (AI(5,8)), 130668-56-3; (AG(5,8)), 130641-71-3.

References

Clarke, E. C. W., & Glew, D. N. (1966) *Trans. Faraday Soc. 62*, 539–547.

Eisenberg, H. (1990) in *Landolt–Börnstein: Numerical Data and Functional Relationships in Science and Technology*, New Series, Group VII (Saenger, W., Ed.) Vol. 4, Springer Verlag, Berlin (in press).

Filimonov, V. V. (1986) in *Thermodynamic Data for Biochemistry and Biotechnology* (Hinz, H.-J., Ed.) pp 377–399, Springer Verlag, Berlin.

Howard, F. B., Chen, C.-q., Ross, P. D., & Miles, H. T. (1991a) *Biochemistry 30*, 779–782.

King, E. J. (1965) *Acid–Base Equilibria*, pp 184–217, Pergamon Press, New York.

Marky, L. A., Blumenfeld, K. S., Kozlowski, S., & Breslauer, K. J. (1983) *Biopolymers 22*, 1247–1257.

Muraoka, M., Miles, H. T., & Howard, F. B. (1980) *Biochemistry 19*, 2429–2439.

Olmsted, M. C., Anderson, C. F., & Record, M. T., Jr. (1989) *Proc. Nat. Acad. Sci. U.S.A. 86*, 7766–7770.

Roark, D. E. (1976) *Biophys. Chem. 5*, 185–196.

Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1968) *J. Mol. Biol. 36*, 291–304.

Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G., & van Boom, J. H. (1988) *Biochemistry 27*, 6321–6326.