Thermodynamics of DNA Duplexes with Adjacent G·A Mismatches†

Ying Li, Gerald Zon,‡ and W. David Wilson*

Department of Chemistry, Georgia State University, Atlanta, Georgia 30303, and Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404

Received February 25, 1991; Revised Manuscript Received May 17, 1991

ABSTRACT: The sequence 5′-d(ATGAGCGAAT) forms a very stable self-complementary duplex with four G·A mismatch base pairs (underlined) out of ten total base pairs [Li et al. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30]. The conformation is in the general B-family and is stabilized by base-pair hydrogen bonding of an unusual type, by favorable base dipole orientations, and by extensive purine–purine stacking at the mismatched sites. We have synthesized 13 decamers with systematic variations in the sequence above to determine how the flanking sequences, the number of G·A mismatches, and the mismatch sequence order (5′-GA-3′ or 5′-AG-3′) affect the duplex stability. Changing A·T to G·C base pairs in sequences flanking the mismatches stabilizes the duplexes, but only to the extent observed with B-form DNA. The sequence 5′-pyrimidine-GA-purine-3′, however, is considerably more stable than 5′-purine-GA-pyrimidine-3′. The most stable sequences with two pairs of adjacent G·A mismatches have thermodynamic parameters for duplex formation that are comparable to those for fully Watson–Crick base-paired duplexes. Similar sequences with single G·A pairs are much less stable than sequences with adjacent G·A mismatches. Reversing the mismatch order from 5′-GA-3′ to 5′-AG-3′ results in an oligomer that does not form a duplex. These results agree with predictions from the model derived from NMR and molecular mechanics and indicate that the sequence 5′-pyrimidine-GA-purine-3′ forms a stable conformational unit that fits quite well into a B-form double helix. In order to minimize binding at secondary nucleic acid target sites, it is obviously of great importance to consider the possible formation of such stable conformational units when designing DNA probes or antisense drug molecules.

Base-pair mismatches can occur through errors in replication (Modrich, 1987) or recombination (Akiyama et al., 1989) and through base reaction chemistry (e.g., a G·C to G·U deamination) (Lewin, 1990). If not corrected by the cellular enzymatic repair system, such mismatches lead to irreversible mutations (Lu & Chang, 1988). The more stable purine–purine mismatches are also common structural components in the telomeres of chromosomes (Panyutin et al., 1990) and in duplex regions of RNA (Gutell & Fox, 1988). The use of oligonucleotides as chemotherapeutic agents (Zon, 1989) and as specific probes for infectious agents (Symons, 1989) requires high specificity in base pairing. The specificity of such interactions is decreased when stable mismatches occur since a significant fraction of the oligomer can be bound to incorrect sequences. Studies to understand the structure and stability of base-pair mismatches in more detail are, thus, necessary if incorrect pairing is to be minimized in the design and use of oligonucleotides in antisense therapy and as gene probes.

G·A mismatch base pairs are of particular interest since they can lead to mutations that are repaired with quite different efficiencies (Fazakerley et al., 1986), they are particularly stable components of RNA duplexes in some sequences (Gutell & Fox, 1988; SantaLucia et al., 1990), they are among the more stable DNA mismatches (Aboul-ela et al., 1985; Dodgson & Well, 1977), they can cause significant incorrect sequence recognition by gene probes and antisense oligonucleotides (Chang et al., 1989), and they are the most thoroughly studied type of mismatch base pair by high-resolution X-ray and NMR techniques (Brown et al., 1986, 1989; Gao & Patel, 1989; Hunter et al., 1986; Kan et al., 1983; Leonard et al., 1990; Li et al., 1991; Nikonowicz & Gorenstein, 1990; Patel et al., 1984; Prive et al., 1987; Webster et al., 1990, and references therein). It is also possible that adjacent mismatches of sequence 5′-GA-3′ are important components of ribozyme active sites (Li et al., 1991).

Single G·A mismatches have been observed in a variety of structural forms in DNA (Figure 1). A protonated G·A base pair (Figure 1, **III**) in sequences 5′-d(CGAGAATTCGCG), 5′-d(CGGGAATTCACG) (Gao & Patel, 1989), and 5′-d(CGCAAATTGGCG) (Brown et al., 1989) has been observed at acidic pH (mismatched base pairs are underlined). An A(anti)·G(anti) species (Figure 1, **I**) in 5′-d(CGAGAATTCGCG) (Patel et al., 1984) and in 5′-d(CCAAGATTGG) (Kan et al., 1983; Nikonowicz & Gorenstein, 1990; Prive et al., 1987), and an A(syn)·G(anti) base pair (Figure 1, **II**) in 5′-d(CGCGAATTAGCG) (Hunter et al., 1986) and 5′-d(CGCAAGCTGCGC) (Webster et al., 1990) with normal base tautomers have been observed at neutral pH and above. Although such G·A base pairs destabilize DNA duplexes, relative to Watson–Crick base pairs, in all sequences investigated to this time, some adjacent G·A mismatches have demonstrated significantly greater than the expected DNA duplex stability. For example, the sequence 5′-d(CCAAGATTGG) forms a quite stable duplex with central adjacent G·A mismatched base pairs in an A(anti)·G(anti) conformation with hydrogen bonding as shown in Figure 1, **I** (Kan et al., 1983; Nikonowicz & Gorenstein, 1990; Prive et al., 1987). In this duplex the adjacent G·A base pairs are highly propeller twisted and, in addition to the hydrogen bonds shown in Figure 1, are stabilized by additional three-center hydrogen bonds between the mismatch G-NH_2 groups and cross-strand thymine O_2 oxygens at the adjacent A-T base pairs (Prive et al.,

† This work was supported by Grant NIH-NIAID AI-27196 and by an NSF equipment grant for the Varian VXR-400.

* To whom correspondence should be addressed.

‡ Applied Biosystems.

0006-2960/91/0430-7566$02.50/0

Table I: Thermodynamic Parameters for Duplex Formation[a]

name	sequence	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/mol deg)	$\Delta G°$ (kcal/mol)[b]	T_m (°C)[c]
IAA	AAGAGCGATT TTAGCGAGAA	−62.0	−178	−9.0	34.0
IAT	ATGAGCGAAT TAAGCGAGTA	−70.5	−198	−11.5	45.4
IAG	AGGAGCGACT TCAGCGAGGA	−89.0	−254	−13.3	52.2
IAC	ACGAGCGAGT TGAGCGAGCA	−97.0	−277	−14.5	54.2
ITA	TAGAGCGATA ATAGCGAGAT	−57.8	−162	−9.5	37.3
I-T	TGAGCGAA AAGCGAGT	−51.9	−144	−9.0	36.3
IIAT	ATGAATGAAT TAAGTAAGTA	−43.5	−124	−6.5	22.1
IITA	ATGATAGAAT TAAGATAGTA				8[d]
IIGC	ATGAGCGAAT TAAGCGAGTA	−70.5	−198	−11.5	45.4
IICG	ATGACGGAAT TAAGGCAGTA	−59.0	−170	−8.3	31.7
III-0	ATGAGCTCAT TACTCGAGTA	−68.0	−187	−12.3	50.0
IIIC2	ATGAGCGCAT TACGCGAGTA	−46.7	−129	−8.3	31.7
IIIT2	ATGAGCTAAT TAATCGAGTA	−36.0	−105	−4.7	6.8
III-4	ATGAGCGAAT TAAGCGAGTA	−70.5	−198	−11.5	45.4
IVGA	ATGACGGAAT TAAGCGAGTA	−70.5	−198	−11.5	45.4
IVAG	ATAGCAGAT TAGAAGGATA	ND[e]	ND	ND	ND

[a] Thermodynamic data were obtained from van't Hoff plots. The samples were in PIPES buffer with 1 M NaCl. G·A pairs are underlined. [b] $\Delta G°$ values were calculated at 25 °C. [c] T_m values were taken at 6.1×10^{-6} M in strands in PIPES buffer with 1 M NaCl. [d] The T_m value was estimated at 3×10^{-5} M in strands in PIPES buffer with 1 M NaCl. [e] No T_m was detectable above 0 °C.

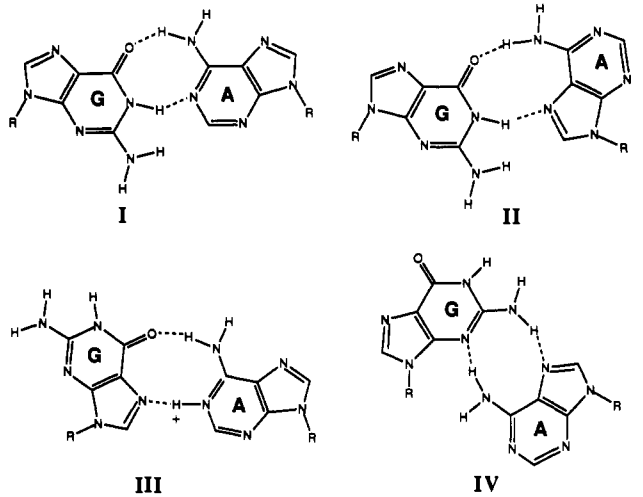

FIGURE 1: Four experimentally observed G·A mismatched base pairs: **I**, G(anti)·A(anti); **II**, G(anti)·A(syn); **III**, protonated G(syn)·A(anti); **IV**, another G(anti)·A(anti) but with completely different hydrogen bonding (see the text for details).

1987). The four central base pairs of this sequence thus form a stable structural unit.

We have found that adjacent G·A mismatches of sequence 5′-GA-3′ can form an entirely different type of structural unit in nucleic acid duplexes (Figure 1, **IV**) (Li et al., 1991). The oligomer 5′-d(ATGAGCGAAT), for example, forms a self-complementary duplex, with two pairs of G·A mismatches, that has stability similar to the non-self-complementary Watson–Crick duplex with one strand of the same sequence (Wilson et al., 1988). Tinoco and co-workers (Aboul-ela et al., 1985) have shown that, although G·A pairs are among the more stable mismatches, they are still generally destabilizing relative to Watson–Crick base pairs. The unusual stability of the sequence shown above with four G·A mismatches out of ten total base pairs is, thus, quite surprising. NMR and molecular modeling studies (Li et al., 1991) have shown that the adjacent G·A base pairs form a structural unit that is stabilized by extensive purine–purine stacking interactions (more detail under Discussion). The molecular modeling studies suggested that this structural unit should be quite sensitive to the mismatch sequence order and to the base pairs flanking the mismatches. To test this model and provide additional information on the thermodynamics and stability of these unusual G·A base pairs, we have synthesized the 13 oligomers shown in Table I. These sequences allow us to determine how flanking sequences, the number of G·A mismatches, and the mismatch sequence order (5′-GA-3′ or 5′-AG-3′) affect the duplex stability. The oligomers were investigated by thermal melting experiments at different salt concentrations and at different total oligomer strand concentrations.

Experimental Procedures

Materials. Oligodeoxyribonucleotides were synthesized and purified as described previously (Wilson et al., 1988).

T_m Determinations. Thermal melting curves from a Cary 219 spectrophotometer were collected by computer and analyzed as previously described (Kibler-Herzog et al., 1990). T_m experiments were conducted in a buffer at pH 7.0 containing 10 mM PIPES, 1 mM EDTA, and Na^+ concentration up to 1.0 M (added NaCl). The extinction coefficients for the oligomers were calculated from mononucleotide and dinucleotide data by using the nearest-neighbor approximation (Fasman, 1975). In units of 10^4 M^{-1} cm^{-1}, the calculated single-strand extinction coefficients at 260 nm are as follows: IAA, 10.6; IAT, 10.6; IAG, 10.2; IAC, 10.4; ITA, 10.9; I-T, 8.48; IIAT, 11.0; IITA, 11.2; IICG, 10.6; III-0, 9.9; IIIC2, 10.0; IIIT2, 10.4; IVAG, 10.8. Oligomer concentrations were determined with these extinction coefficients by using absorbance values obtained from linear extrapolation of the high-temperature baseline in UV melting curves to 25 °C.

T_m, $\Delta H°$, and $\Delta S°$ values for individual curves are determined with a Macintosh computer by using a nonlinear fitting program that uses sloping baselines in both the low- and high-temperature regions (Aboul-ela et al., 1985; Bower et al., 1987; Petersheim & Turner, 1983). Melting temperatures are used to construct plots of $1/T_m$ versus ln C_T according to the van't Hoff equation

$$1/T_m = (R/\Delta H°) \ln (C_T/N) + \Delta S°/\Delta H° \quad (1)$$

where $1/T_m$ is the reciprocal of the melting temperature in degrees K, C_T is the total oligomer concentration in single strands, N is 1 for self-complementary duplexes and 4 for non-self-complementary duplexes, and $\Delta H°$ and $\Delta S°$ are the enthalpy and entropy for duplex dissociation. $\Delta H°$ and $\Delta S°$ values can be determined from the van't Hoff plots with the assumptions of a two-state transition and a temperature-independent transition enthalpy (Marky & Breslauer, 1987; Martin et al., 1971).

Results

Oligomer Selection. A total of 13 oligomers have been synthesized (grouped by related sequences in Table I). The oligomer sequences are designed to compare duplexes with none, two single, and two adjacent G·A mismatches, to probe the effects of neighboring sequences on duplexes with adjacent G·A mismatches, and to determine the effect of switching the adjacent mismatch sequence from 5′-GA-3′ to 5′-AG-3′. The oligomers are divided into four groups for comparison purposes (Table I).

Group I: These oligomers have the same central hexanucleotide core, 5′-GAGCGA-3′, with four G·A mismatches out of six base pairs. They have variations in the outer base pairs in the sequence such that base pairs on the 5′ side of the first set of G·A mismatches and on the 3′ side of the last set of G·A mismatches are varied. The concerted changes preserve the self-complementary base pairing of the oligomers. The first four oligomers in the group, 5′-d(AX-core-YT) (IAX), have the same terminal 5′-A and 3′-T and have variations in only the bases adjacent to the mismatches. The next oligomer, ITA, has the reverse sequence at the 5′ and 3′ ends when compared to IAT. The last oligomer I-T, is an octamer with the same central sequence as IAT.

Group II: These four oligomers have the same sequence of their outer eight base pairs, 5′-d(ATGAXYGAAT) (IIXY), and the central base pairs are varied over all possibilities with self-complementary base pairing. These changes affect the base pairs at the 3′ side of the first set of G·A mismatches and at the 5′ side of the second set of mismatches, the opposite side of the mismatches relative to the changes of Group I.

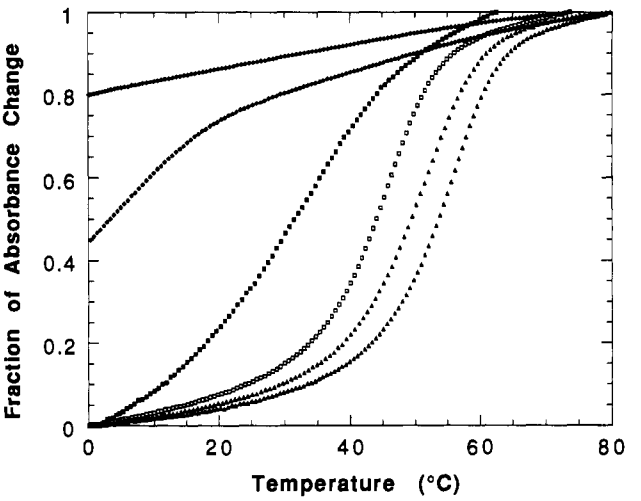

FIGURE 2: Absorbance–temperature profiles for DNA oligomers: right to left, IAC, the most stable sequence with adjacent mismatches; III-0, Watson–Crick sequence; IVGA, a mismatched duplex investigated by NMR (Li et al., 1991); IIIC2, a sequence with single mismatches; IITA, the least stable sequence with adjacent mismatches; and IVAG, a sequence with reversed mismatch order, 5′-AG-3′. These oligomers were in a buffer of 10 mM PIPES, 1 mM EDTA, 1 M NaCl, pH 7, and at 1×10^{-5} M in strands. IITA has a T_m of approximately 0 °C.

Group III: These four oligomers have the same first six and last two bases, 5′-d(ATGAGCXYAT). They compare this sequence with zero (III-0), two (IIIC2 and IIIT2), and four (III-4) G·A mismatch base pairs while maintaining the self-complementary base pairing in the duplexes. Bases X and Y change to generate a sequence with two adjacent G·A mismatches (III-4, four G·A mismatches in total), two sequences with two single G·A mismatches, one with adjacent G·C pairs (IIIC2), another with adjacent A·T pairs (IIIT2), and one Watson–Crick self-complementary sequence with no G·A mismatches (III-0). These are all the possible variations at these two positions with G·A mismatches and self-complementary pairing.

Group IV: The two oligomers in this group have the same sequence adjacent to the mismatches, 5′-d(ATXYGCX-YAT), and have the mismatches in both possible directions, 5′-GA-3′ and 5′-AG-3′.

The sequence 5′-d(ATGAGCGAAT) appears in each group under different names, IAT, IIGC, III-4, and IVGA, for direct comparison.

T_m Comparisons and Effects of Salt Concentration. All of the oligomers, except two sequences, IVAG and IITA, at low salt concentration, exhibit cooperative transitions as the temperature is increased. Figure 2 shows some typical T_m curves of the oligomers: III-0, a fully complementary Watson–Crick base-paired duplex; IAC and IAT, duplexes with two sets of adjacent G·A mismatches; IIIC2, a similar duplex with two single G·A mismatches; IITA, the least stable duplex with two sets of adjacent G·A mismatches; IVAG, a sequence with two adjacent G·A mismatches in the reverse order, 5′-AG-3′. As can be seen, IAC with two adjacent G·A mismatches has a remarkably high T_m, even higher than that of the Watson–Crick sequence III-0. IITA, which has the same mismatch sequence, however, showed no melting at low salt concentrationd, and even at high salt concentration only a partial transition curve was observed. At salt concentrations up to 1 M, the profiles of absorbance vs temperature for IVAG did not show an observable transition above 0 °C. The other oligomers have cooperative melting transitions, and T_m values, under a fixed set of conditions, are collected in Table I. The

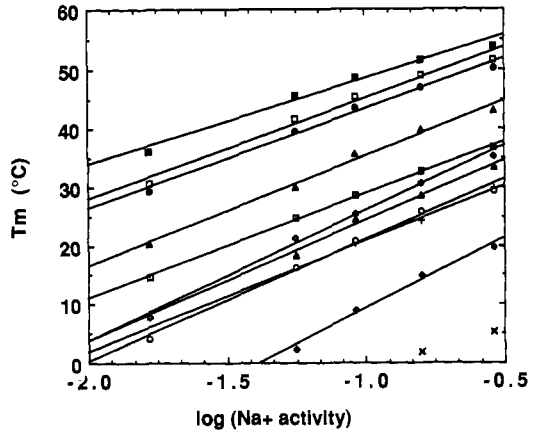

FIGURE 3: T_m vs log (Na^+ activity) plots for Δ, IAA; ▲, IAT; □, IAG; ■, IAC; ⊡, I–T; ◇, ITA; ◆, IIAT; ○, IICG; ●, III-0; +, IIIC2; ×, IIIT2. Solution conditions are the same as for Figure 2.

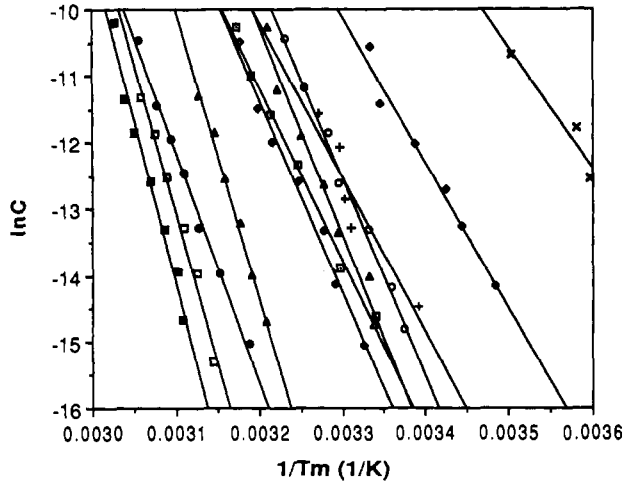

FIGURE 4: $1/T_m$ vs log (oligomer strand concentration) plots for the sequences shown in Figure 3 (symbols are as in Figure 3). Solution conditions are the same as described in the legend to Figure 2.

T_m values increase as salt concentration increases and Figure 3 shows linear plots of T_m vs log (Na^+ activity). The slopes of these plots are in the general range (17 ± 2) expected for DNA transitions (Cantor & Schimmel, 1980).

Thermodynamic Comparisons. The T_m values of the oligomers also depend on the oligomer concentration, and Figure 4 shows linear plots of $1/T_m$ vs ln C_T. All of the thermodynamic parameters derived from van't Hoff plots agree with those from individual curve fittings within 10% for most of the oligomers. The free energy values calculated by the two methods agree within 7% for all of the duplexes. The agreement indicates that there is only one two-state transition for all the sequences (Marky & Breslauer, 1987). Table I lists the sequences with their thermodynamic results, and, as can be seen, there are strikingly large variations in the thermodynamic parameters in all four of the oligomer groups. Enthalpy values vary by 60 kcal/mol, entropy values by 170 eu, and free energy values by almost 10 kcal/mol. Correlation of enthalpy and entropy values results in the smaller variation range observed for the free energy values.

When comparing oligomers of the same AT/GC composition, the least stable sequence is IVAG with adjacent mismatches of sequence 5′-AG-3′, the next level of stability is exhibited by oligomers with two separated single mismatches (IIIC2 and IIIT2), and the most stable duplexes have adjacent 5′-GA-3′ mismatches. As can be seen in Table I, very slight changes in the sequences flanking the G·A mismatches can have very large effects on the thermodynamics of duplex formation (compare, for example, IIAT and IITA or IIGC and IICG).

Discussion

The duplex structure of sequence 5′-d(ATGAGCGAAT) (named in different groups as IAT, IIGC, III-4, and IVGA) has been investigated by NMR methods, and a molecular model has been built with the NMR constraints and energy minimized by molecular mechanics (Li et al., 1991). With four G·A mismatched base pairs, this sequence has a very unusual duplex conformation with characteristics briefly described as follows: the sequence forms a stable self-complementary duplex with four G·A mismatches, each of which is paired as shown in Figure 1, IV. The adjacent mismatched G·A pairs stack in an unusual way such that one G is over the G on the opposite strand and one A is over the cross-strand A (Figure 5, center). As can be seen, these cross-strand interactions give excellent base stacking and alignment of base dipoles at the adjacent mismatch steps. Good stacking is also observed between the mismatched bases and particular bases of their Watson–Crick neighboring base pairs. The stacking of the first mismatched G in the sequence with its 5′ neighbor T (Figure 5, top), and the stacking of the second mismatched G with its 5′ neighbor C (Figure 5, bottom) is quite good. The stacking between the mismatched A bases and their 3′ neighboring purines in Watson–Crick base pairs, however, is less extensive (Figure 5).

In spite of the four G·A mismatches out of ten base pairs in total, the conformation described above is in the general right-handed B-family of duplexes with very few of the backbone torsional angles outside of the normal range for a B-type helical conformation (Figure 6). The oligomer conformation presented in Figures 5 and 6 is still under refinement by use of an expanded set of NMR structural constraints, and the specific features of the conformation should be viewed as preliminary. The global conformation and hydrogen-bonding scheme are, however, clearly defined by the NMR results (Li et al., 1991). The conformational model can be used to explain the effects of neighboring sequence, the number of mismatches, and the adjacent base order on the thermodynamic results of this paper. An additional important point in the structural analysis of G·A mismatches that emerges from the results presented here is that both the glycosidic torsional angles and the type of base pairing must be specified to define the conformation. Both the type I and type IV base pairs in Figure 1, for example, are in the anti–anti glycosidic conformation, but they have different hydrogen-bonding arrangements and completely different global conformations.

Nearest-Neighbor Effects: Group I and Group II. Nearest-neighbor base sequence effects are very important for the stability of the mismatched sequences in Table I. The sequences in Group I have the same core, 5′-GAGCGA-3′, with the adjacent mismatched G·As at each end of the core. Changing the 5′ nearest neighbor of the first mismatch results in a change of the 3′ nearest neighbor of the second mismatch in the first four sequences (Table I). Two interesting observations are clear from the comparison of $\Delta H°$ and $\Delta G°$ values for the transitions of these oligomers. First, G·C base pairs are more preferred (higher stability) than A·T at either side of the core (compare IAG and IAC to IAA and IAT and the four sequences of Group II in Table I). Second, the combination 5′-pyrimidine(core)purine-3′ is more stable than 5′-purine(core)pyrimidine-3′ (compare IAC to IAG, IAT to IAA, IIAT to IITA, and IIGC to IICG). The results of all possible nearest neighbors in Group I and Group II, thus, show that

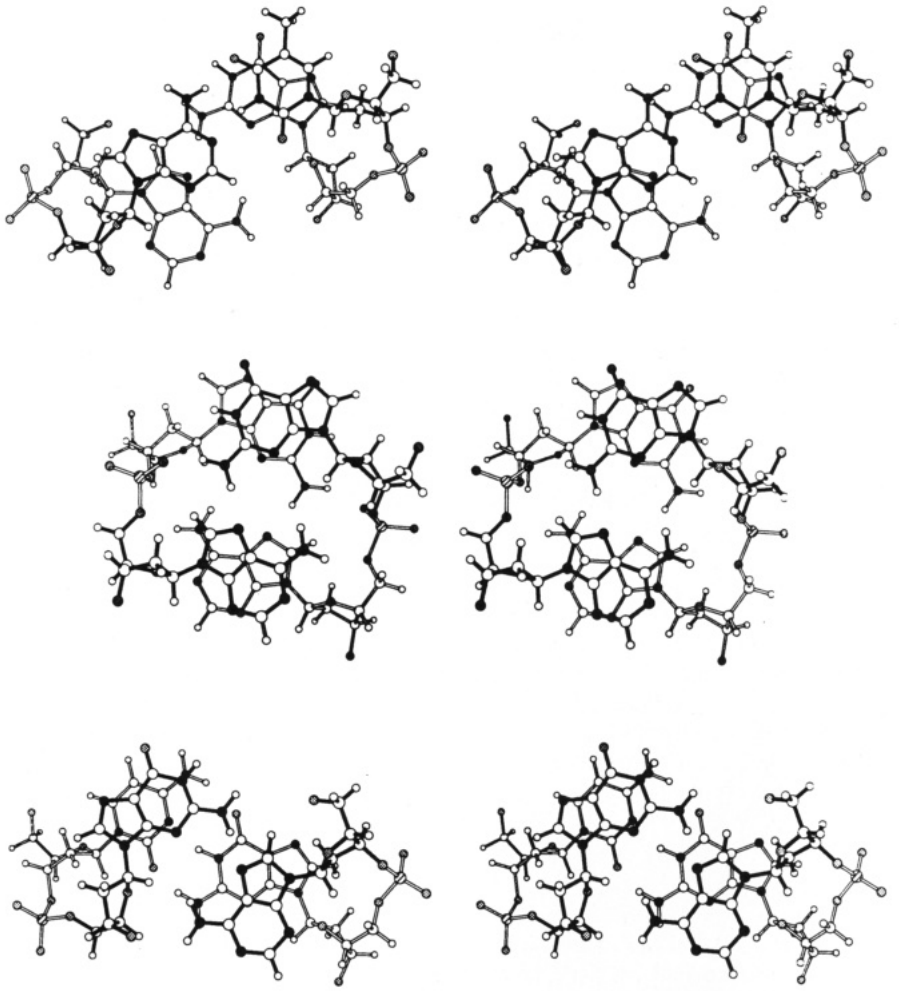

FIGURE 5: Stereoview from the top of the stacking of bases at the G·A mismatch sites from the NMR and molecular modeling studies on 5′-d(ATGAGC) (Li et al., 1991): (top) (5′-TG-3′)·(5′-AT-3′); (center) (5′-GA-3′)·(5′-GA-3′); (bottom) (5′-AG-3′)·(5′-CG-3′). The bonds in the upper base pairs are darkened in each diagram for visualization. Atom labeling is as follows: carbon, white; hydrogen, smaller white; nitrogen, black; oxygen, stippled; and phosphorous, diagonal lines.

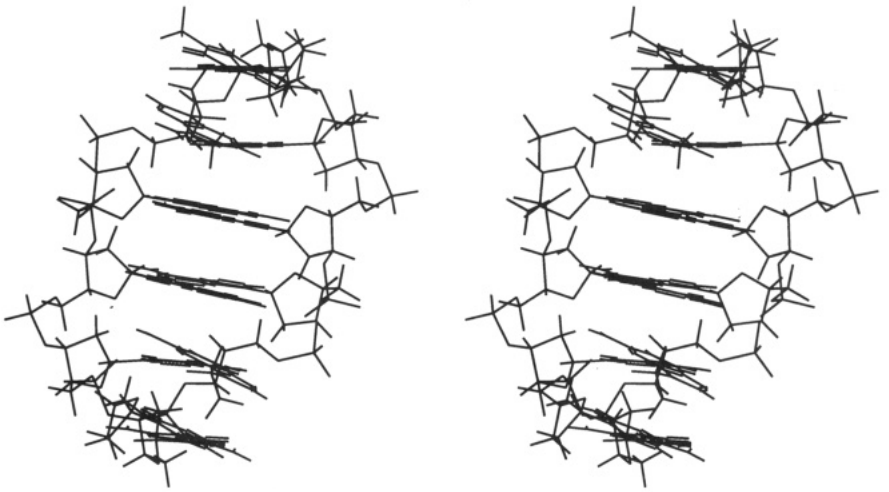

FIGURE 6: Stereoview of the mismatched duplex of (5′-ATGAGC-3′)·(5′-GCGAAT-3′) (Li et al., 1991). The center two base pairs are the mismatched G·A pairs. As can be seen, the duplex has little backbone distortion and basically is in the B-form family.

G·C base pairs enhance stability more than A·T at both sides of an adjacent G·A mismatch and that the sequence combination 5′-pyrimidine-GA-purine-3′ is more stable than 5′-purine-GA-pyrimidines-3′.

The nearest-neighbor stabilization of the mismatched duplexes by G·C base pairs can be explained by the fact that they are intrinsically more stable than A·T base pairs. The free energy differences, $\Delta\Delta G°$, between **I**AA and **I**AG, between **I**AT and **I**AC, between **II**GC and **II**AT, and between **II**CG and **II**TA come primarily from the extra hydrogen bonds of the G·C base pairs (one for each G·C, two G·Cs in total in the sequence). The maximum magnitude of $\Delta\Delta G°$ for one hydrogen bond in oligomers is about 2 kcal/mol (Turner et al., 1987). The $\Delta\Delta G°$ values are 2.1 kcal/mol between **I**AG and **I**AA, 1.5 kcal/mol between **I**AC and **I**AT in group **I**, and 2.5 kcal/mol between **II**GC and **II**AT in group **I**. The $\Delta\Delta G°$

Table II: Thermodynamic Parameters for Duplex Formation

name	sequence	ΔH_{cal} (kcal/mol)[a]	$\Delta\Delta H_{cal}$ (kcal/mol)	ΔH_{obs} (kcal/mol)[b]	$\Delta\Delta H_{obs}$ (kcal/mol)
IAT	ATGAGCGAAT	17.2 + T-core-A		−70.6	
I-T	TGAGCGAA	T-core-A	17.2	−51.9	18.6
IAA	AAGAGCGATT	18.2 + A-core-T		−62.0	
ITA	TAGAGCGATA	12.0 + A-core-T	6.2	−57.8	4.2
III-0	ATGAGCTCAT	−66.7		−68.0	

[a] Calculated by using the database of Breslauer et al. (1986). [b] From Table I.

values for these sequences, thus, indicate that the primary contribution to the enhanced duplex stability due to G·C base pairs is essentially the same as that predicted for Watson–Crick duplexes of similar sequence. The van't Hoff plot for sequence IITA in group II can not be obtained because of the low T_m values, and, therefore, a quantitative comparison between IITA and IICG is not possible.

In the molecular mechanics energy-minimized model of sequence IAT, the unique and excellent interactions between the mismatched purines (Figure 5, center) provide a major contribution to stabilize the duplex (Figure 6). The stacking between the mismatched purines and their Watson–Crick neighbor base pairs (Figure 5, top and bottom) is also as good as that in a regular B-form DNA (Wing et al., 1980). At this point we do not have an experimental or molecular mechanics derived structure to explain precisely why switching 5′-pyrimidine-GA-purine-3′ to 5′-purine-GA-pyrimidine-3′ decreases the duplex stability, but it is clear that in the structural unit 5′-pyrimidine-GA-purine-3′, for example, (5′-TGAG-3′)·(5′-CGAA-3′) in IAT, all of the interactions are highly optimized, and changes in the neighbor sequence decrease the stacking and the duplex stability.

The thermodynamic parameters for some Watson–Crick base pairs in the oligomers of Table I can be calculated from the database of Breslauer et al. (1986) and compared to experimental results. For example, in IAA $\Delta H°$ (AAGAGCGATT) = $\Delta H°$ (AA) + $\Delta H°$ (AcoreT) + $\Delta H°$ (TT), where the core is the sequence underlined, and in ITA, $\Delta H°$ (TAGAGCGATA) = $\Delta H°$ (TA) + $\Delta H°$ (AcoreT) + $\Delta H°$ (TA). Assuming that $\Delta H°$ (AcoreT) remains essentially the same in all the sequences, then the difference between IAA and ITA will be $\Delta\Delta H° = [\Delta H°\text{ (AA)} + \Delta H°\text{ (TT)}] - [\Delta H°\text{ (TA)} + \Delta H°\text{ (TA)}]$, which can be calculated. IAT and I-T can be compared in a similar manner, and the results of such comparisons are listed in Table II. G·A mismatches are involved in neighbor interactions at other positions, and, since no empirical database is available for them, other comparisons can not be made at this point. The calculated enthalpy differences ($\Delta\Delta H°_{cal}$) are very close to the observed enthalpy differences ($\Delta\Delta H°_{obs}$) for the change from 5′-AA step to 5′-TA step and from 5′-AT step to 5′-T step (Table II). The good agreement between the calculated and observed values indicates that in these sequences base pairs adjacent to the core behave as they do in standard Watson–Crick duplexes. The free energy differences ($\Delta\Delta G°$) are also very close between calculated and observed values for IAT and I-T ($\Delta\Delta G$ is 3.0 and 2.5 kcal/mol for calculated and experimental values, respectively), again suggesting that the next nearest-neighbor Watson–Crick base pairs of the core behave as they do in standard B-form DNA. These observations support the NMR and molecular mechanics results that indicate an overall B-conformation for the mismatched duplexes as shown in Figure 6.

For IAA and ITA, however, the difference between calculated $\Delta\Delta G°$ (2.0 kcal/mol) and observed $\Delta\Delta G°$ (−0.5 kcal/mol) values is larger than expected. The purine A at the 5′ side of the core is the least favorable nearest neighbor for the adjacent G·A mismatches (see Table I, group I). This may be due to initiation of an AA track type conformation (Nelson et al., 1987), which could distort the mismatches, or due to purine–purine steric clash at the adjacent mismatch conformation. As indicated above, either an A or G as the 5′ neighbor of the 5′-GA-3′ sequence gives duplexes of lower stability than those with 5′ pyrimidines.

Effect of the Number of G·A Mismatches on Duplex Stability: Group III. DNA duplexes with mismatched base pairs are generally less stable than the corresponding duplex with only Watson–Crick base pairs (Aboul-ela et al., 1985). a T_m reduction of about 17 °C, for example, has been seen on replacing two G·C by two G·A pairs in the sequence 5′-d(CGAGAATTCGCG) (Patel et al., 1984). G·A mismatches in the sequences 5′-d(CGCAATTGGAG) cause a decrease in T_m of about 25 or 22 °C when the two G·A pairs replace two G·C pairs or two A·T pairs, respectively (Leonard et al., 1990). A reasonably high T_m value has been observed from NMR analysis of the sequence 5′-d(CCAAGATTGG) with adjacent G·A mismatches (Nikonowicz & Gorenstein, 1990). The stability of this sequence is enhanced by the two extra hydrogen bonds at the mismatched sites as shown by X-ray studies (Prive et al., 1987). With two sets of adjacent mismatched G·A pairs, the sequence 5′-d(ATGAGCGAATA) also has exceptional stability with T_m values and thermodynamic parameters comparable to those for the corresponding Watson–Crick duplex (Wilson et al., 1988).

With no mismatched base pairs, oligomer III-0 in group III is a self-complementary sequence, and the experimental and calculated $\Delta H°$ values for melting of this oligomer agree very well (Table II). Sequences IIIC2 and IIIT2, which have two single G·A mismatches, are much less stable than III-0. These results are consistent with the general observation of duplex destabilization by single G·A mismatches. The destabilization by single G·A mismatches occurs because the purine–purine base pair is too large to fit into the standard diameter of a Watson–Crick B-form duplex. The mismatches, thus, cause a bulge in the duplex backbone or undergo an anti–syn conversion of one of the mismatch purine bases to create a hydrogen-bonded base pair that is closer to the width of standard base pairs. Either of these effects contributes a positive free energy term to the duplex conformational energy.

Our results for oligomers with two pairs of adjacent G·A mismatches are reversed relative to the sequences with two single G·A mismatches. III-4, for example, has T_m, $\Delta H°$, and $\Delta G°$ values comparable to the corresponding Watson–Crick sequence III-0. Instead of a bulge as expected for single G·A mismatches, a cross-chain purine–purine stack occurs in III-4 (Li et al., 1991). This structural unit lets the two mismatches fit into a B-form helix, and its extra stability comes primarily from the excellent purine–purine interactions (Figure 5). It is the *unique local structure* of the mismatched G·A base pairs and not the number of mismatches that determines the duplex stability for this sequence. In the sequence 5′-d(CCAAGATTGG) a bulge is created in the B-form duplex, but the

positive free energy from this distortion is largely offset by the extra hydrogen bonds formed by the adjacent mismatch base pairs and their neighboring bases. Thus, the two different oligomers with adjacent G·A mismatches have two completely different structural units that markedly stabilize G·A mismatches in B-form duplexes.

The Effect of Sequence Order on Stability of the Adjacent Mismatches: Group IV. The small absorbance changes with temperature for IVAG (Figure 1) indicate a typical strand unstacking process with no duplex to single-strand transition. Although the stability of IVAG is not expected to be the same as IVGA, it was initially surprising than no duplex occurs for IVAG. As indicated above, however, the IVGA duplex is strongly stabilized by purine–purine stacking, by base–dipole alignment, and by hydrogen bonding. During the molecular model building for IVGA, we found that the order 5′-GA-3′ is crucial for this type of base pairing and base stacking. The opposite sequence 5′-AG-3′ can form type IV hydrogen bonds, but the purine–purine stacking no longer exists and the duplex is severely distorted. SantaLucia et al. (1990) have studied RNA oligomers with adjacent G·A mismatches, 5′-GA-3′ and 5′-AG-3′, in sequences 5′-(GCGAGCU) and 5′-(GCAGGCG) and also find that the 5′-GA-3′ sequence is more stable. Both the RNA sequences form duplexes with base-paired mismatches, but, on the basis of imino proton NMR results, the hydrogen bonding for 5′-AG-3′ is different from that for 5′-GA-3′. The mismatched G·A pairs are in an anti conformation (Figure 1, I) in the 5′-AG-3′ RNA sequences (SantaLucia et al., 1990), while we feel that in the RNA 5′-GA-3′ sequence the mismatches pair as in IVGA on the basis of their similar imino proton chemical shifts (Li et al., 1991).

Conclusion

It is, thus, clear from all of the thermodynamic results presented here that adjacent 5′-GA-3′ mismatched can form a quite stable structural unit (Figure 5), especially in the sequence 5′-pyrimidine-GA-purine-3′. The unit with reversed purine and pyrimidine flanking bases is less stable, and the reversed mismatch sequence 5′-AG-3′ has extremely low stability. Duplexes with isolated G·A mismatch base pairs can not form the type of structural unit shown in Figure 5, and they have lower stability than similar sequences with adjacent G·A mismatches. In order to minimize binding at secondary nucleic acid target sites, it is obviously of great importance to consider the possible formation of stable adjacent G·A conformational units when designing DNA probes or antisense drug molecules. As discussed above, the appearance of similar conformational units in RNA also appears probable and must be considered in RNA structure prediction algorithms. Finally, as one simply ponders the aesthetics of the base pairing and conformation in Figure 5, the question of how many such stable conformational units exist in nucleic acids naturally arises.

References

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

Akiyama, M., Maki, H., Sekiguchi, M., & Horiuchi, T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3949–3952.

Bower, M., Summers, M. F., Kell, B., Hoskins, J., Zon, G., & Wilson, W. D. (1987) *Nucleic Acids Res. 15*, 3531–3547.

Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.

Brown, T., Hunter, W. N., Kneale, G., & Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2402–2406.

Brown, T., Loenard, G. A., Booth, E. D., & Chambers, J., (1989) *J. Mol. Biol. 207*, 455–457.

Cantor, C. R., & Schimmel, R. R. (1980) *Biophysical Chemistry*, W. H. Freeman and Co., San Francisco, CA.

Chang, E. H., Yu, Z., Shinozuki, K., Zon, G., Wilson, W. D., & Strekowska, A. (1989) *Anti-Cancer Drug Des. 4*, 221.

Dodgson, J. B., & Well, R. D. (1977) *Biochemistry 16*, 2367–2374, 2374–2379.

Fasman, G. D., Ed. (1975) *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, Vol. I, 3rd ed., CRC Press, Cleveland, OH.

Fazakerley, G. V., Quinard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H., Jones, M., & Radman, M. (1986) *EMBO J. 5*, 3697–3703.

Gao, X., & Patel, D. J. (1989) *J. Am. Chem. Res. 110*, 5178–5182.

Gutell, R. R., & Fox, G. E. (1988) *Nucleic Acids Res. 16* (supplement), r175–r269.

Hunter, W. N., Brown, T., & Kennard, O. (1986) *J. Biomol. Struct. Dyn. 4*, 173–191.

Kan, L. S., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.

Kibler-Herzog, L., Kell, B., Zon, G., Shinozuka, K., Mizan, S., & Wilson, W. D., (1990) *Nucleic Acids Res. 18*, 3545–3555.

Leonard, G. A., Booth, E. D., & Brown, T. (1990) *Nucleic Acids. Res. 18*, 5617–5623.

Lewin, B. (1990) in *Genes IV*, Oxford University Press, New York.

Li, Y., Zon, G., & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.

Lu, A.-L., & Chang, D.-Y. (1988) *Cell 54*, 805–812.

Marky, L. A., & Breslauer, K. J. (1987) *Biopolymers 26*, 1601–1620.

Martin, F. H., Uhlenbeck, O. C., & Doty, P., (1971) *J. Mol. Biol. 57*, 201–215.

Modrich, P. (1987) *Annu. Rev. Biochem. 56*, 435–466.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.

Panyutin, I. G., Kovalsky, O. I., Budowsky, E. I., Dickerson, R. E., Rikhirev, M. E., & Lipanov, A. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 867–870.

Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Biochemistry 23*, 3207–3217.

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

Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.

SantaLucia, J. J., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813–8819.

Symons, R. H. (1989) in *Nucleic Acid Probes* (Symons, R. H., Ed.) CRC Press, Inc., Boca Raton, FL.

Turner, D. H., Sugimoto, N., Kierzek, R., & Dreoker, S. D. (1987) *J. Am. Chem. Soc. 109*, 3783–3785.

Webster, G. D., Sanderson, M. R., Skelly, J. V., Neidle, S., Swann, P. F., Li, B. F., & Tickle, I. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6693–6697.

Wilson, W. D., Dotrong, M., Zuo, E. T. & Zon, G. (1988) *Nucleic Acids Res. 16*, 5137–5151.

Zon, G. (1989) in *Oligonucleotides: Antisense Inhibitors of Gene Expression* (Cohen, J. S., Ed.) pp 233–249, CRC Press, Boca Raton, FL.