Oligonucleotides Containing G•A Pairs: Effect of Flanking Sequences on Structure and Stability

Ying Li and Sudhir Agrawal*

Hybridon Inc., One Innovation Drive, Worcester, Massachusetts 01605

Received March 8, 1995; Revised Manuscript Received May 26, 1995[⊗]

ABSTRACT: Sixteen oligodeoxyribonucleotides, 5′d(GXGAYC)3′, X and Y = G, A, C, or T, have been synthesized and studied by UV melting and ^{1}H and ^{31}P NMR methods. By varying X and Y, the sixteen resulting oligonucleotides can theoretically form 10 duplexes with all possible Watson−Crick base pairs flanking the center two G•A base pairs. Two-dimensional ^{1}H NMR data on 5′d(GCGAGC)3′ revealed that the center bases G and A pair through G amino hydrogen bonding and that the two consecutive G•A pairs form excellent purine−purine stacks. The concurrent appearance of one or more upfield-shifted imino proton peaks (∼10.5 ppm) and both upfield- and downfield- shifted ^{31}P signals (∼−2 and ∼−5.1 ppm) was a unique characteristic in imino ^{1}H and ^{31}P NMR spectra and was used as a conformational probe for this type of G•A pairs. Using this probe, seven out of 10 duplexes of 5′GXGAYC3′ were found to adopt the G•A base pairing with G amino proton bonding and G to G and A to A base stacking. Three were in the group comprising 5′pyrimidine−GA−purine3′, and four were in the group comprising 5′purine−GA−purine3′/5′pyrimidine−GA−pyrimidine3′. The G•A pairs in 5′purine−GA−pyrimidine3′ adopted a totally different conformation. Thermodynamic analysis indicated that duplexes in the 5′pyrimidine−GA−purine3′ group were more stable than the duplexes in the 5′purine−GA−pyrimidine3′ group. Overall, G•C base pairs were preferred as neighbors to this type of G•A pairs.

Antisense oligonucleotides have received significant attention recently because of their many potential uses in biomedical sciences (Zamecnik & Stephenson, 1978). Antisense oligonucleotides hybridize with complementary bases in RNA or DNA, inhibiting translation or transcription, and thereby controlling gene expression (Agrawal, 1992; Stein & Cheng, 1993; Agrawal & Iyer, 1995). For antisense oligonucleotides to effectively control gene expression, perfect Watson−Crick base pairing between oligo and the target RNA or DNA is absolutely required: mismatched base pairs severely decrease the bonding specificity and affinity. Single mismatched base pairs are less stable than Watson−Crick base pairs (Aboul-ela et al., 1985; Patel et al., 1984; Werntges et al., 1986): A and C will not form a base pair under normal conditions, and the most stable G•T (G•U) pairs are significantly less stable than A•T or G•C pairs (Kalnik et al., 1988). The bonding affinity is therefore better with a longer sequence length of Watson−Crick type antisense oligonucleotides (longer than 15 bases) (Walter et al., 1994).

Sequences with two or more consecutive mismatches have not been studied in detail. We presume that they either do not form a duplex or form a duplex that is too unstable to compete with complementary paired duplexes. In designing antisense oligonucleotides, however, we should pay particular attention to sequences with unexpected stable mismatched base pairs, especially consecutive ones since they are not very well documented and can be easily neglected. Adjacent G•A mismatches have been of great interest because of their unusual structural features (Figure 1) and specific stability (Chou et al., 1992; Ebel et al., 1992; Greene et al., 1994; Kan et al., 1983; Katahira et al., 1993; Lane et al., 1992; Li et al., 1991a,b; Maltseva et al., 1993; Maskos et al., 1993; Nikonowicz & Gorenstein, 1990; Orbons et al., 1987; Prive et al., 1987; SantaLucia et al., 1990; Webster et al., 1990; Wilson et al., 1988). The neighboring Watson−Crick base pair sequences have considerable influence on the thermostability of duplexes with adjacent G•A pairs (Cheng et al., 1992; Ebel et al., 1992; Li et al., 1991a). For the same base

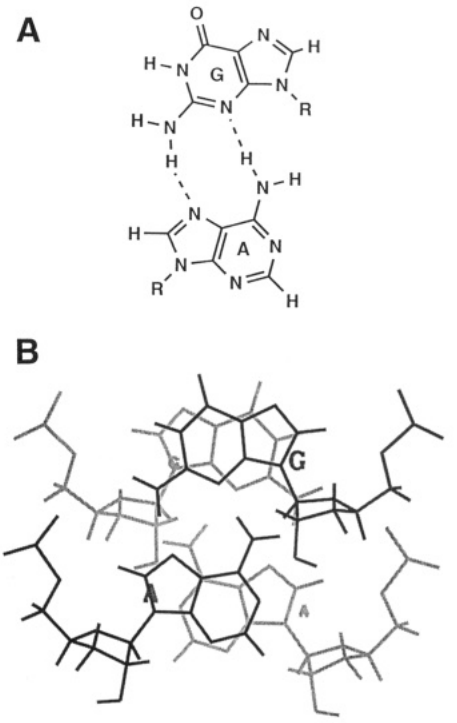

FIGURE 1: (A) G•A base pairing present predominantly with different neighboring sequences as shown in this study. (B) Extended purine−purine stacking with adjacent 5′GA3′/5′GA3′ pairs, G on top of G from the other strand and A on top of the other A, from our molecular modeling.

* To whom correspondence should be addressed.

[⊗] Abstract published in *Advance ACS Abstracts,* July 15, 1995.

0006-2960/95/0434-10056$09.00/0

composition, with adjacent G·A pairs in 5′GA3′ series, switching the neighboring bases from purines to pyrimidines or vice versa, e.g., from 5′A-GA-T3′ to 5′T-GA-A3′, causes the center two G·A pairs to adopt a totally different conformation (Greene et al., 1994). We are interested in a series of questions that arises: How common are G·A pairs? What conformation is most often adopted? What is the duplex stability? We address these questions by systematically altering the neighboring sequences of the center 5′GA3′ in a series of hexaoligonucleotides, 5′GXGAYC3′, X and Y = A, T, G, or C, and investigating their structural and thermal properties using NMR and temperature melting techniques. Our results indicate that (a) all of the oligomer DNA sequences with 5′GA3′ form duplexes; (b) altering the flanking sequences of the GAs changes the duplexes' stability substantially; and (c) G·A pairs exist in every duplex, and the dominant conformation for the base pairs is as shown in Figure 1.

MATERIAL AND METHODS

Oligodeoxyribonucleotides. Oligodeoxyribonucleotides were synthesized on an automated synthesizer (Millipore 8700) using β-cyanoethyl phosphoramidite chemistry. After deprotection, the oligodeoxynucleotides with a DMT group were first purified using reverse-phase HPLC on a C_{18} column. After detritylation with 80% acetic acid for 1 h, the samples were repurified using reverse-phase HPLC and then desalted by dialysis. The final sample purity was checked by a denaturing electrophoresis gel assay.

Twenty oligodeoxyribonucleotides were synthesized. The first 16 oligomers of 5′GXGAYC3′ (X and Y = A, T, G, or C) include all the possibilities of the four deoxyribonucleotides at both the 5′ and 3′ sides of 5′GA3′ and form 10 duplexes. The 10 duplexes are divided into three groups, shown in Table 1, on the basis of the flanking base type: there are three duplexes in the 5′py−GA−pu3′ group, four in the 5′pu−GA−pu3′/5′py−GA−py3′ group, and three in the 5′pu−GA−py3′ group. The duplexes in the first and last groups have the opposite base types at the 5′ and 3′ sides of the 5′GA3′. For the purpose of comparison, four duplexes with perfect Watson−Crick base pairs are listed in the reference group.

Melting Experiments. Thermal melting (T_m) data were collected using a GBC 920 spectrophotometer, which has six 10-mm cuvettes mounted in a dual carousel. The temperature is controlled by software directing a peltier-effect temperature controller. A VWR 1166 refrigerated bath is connected to the peltier-effect temperature controller to absorb the heat. A nitrogen purge is used in the sample compartment at low temperature. T_m experiments were directed and controlled by a computer with DNA-Melting software, which is supplied by GBC. T_m data were analyzed by both the first-derivative method and the midpoint method in the software. The error of T_m values was within 0.5 °C on repeated experiments and curve fittings. The buffer contained 10 mM PIPES, 1 mM EDTA, and 1 M NaCl at pH 7.0. Oligomer DNA strand concentration was determined by using absorbance values at 260 nm with their extinction coefficients (Fasman, 1975). Melting temperatures were used to construct plots of $1/T_m$ versus ln C_t according to the van't Hoff equation:

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

$N = 1$ if the duplex is self-complementary, and $N = 4$ for non-self-complementary duplexes.

NMR Experiments. Exchangeable proton NMR experiments were performed on a Varian UNITY 300. The samples were contained in 5-mm tubes in a buffer consisting of 10 mM PIPES, 0.1 mM EDTA, 0.2 M NaCl, pH 7.0, and H_2O/D_2O = 90/10. A solvent suppression pulse sequence (Hore, 1983) of either 1331 or 11 was used with the transmitter frequency set on water resonance. Typical experimental parameters were as follows: spectral width, 7000 Hz; acquisition time, 1 s; recycle delay, 1 s; 1K to 7K scans; and temperature, 0 °C. The data were Fourier-transformed with line broadening at 2 Hz. Water resonance was used as the internal reference and was set at 5 ppm. One-dimensional NOE spectra for the imino proton assignments were obtained by alternately placing the decouplor on and off the resonances of interest. The decoupler power was adjusted to saturate ~80% of peaks in 1−1.5 s. The off-resonance FID was subtracted from the on-resonance FID and then Fourier-transformed to generate the difference spectra.

The same samples were used for ^{31}P NMR experiments on the UNITY 300 set at 121 MHz. Typical experimental parameters were as follows: spectral width, 2420 Hz; acquisition time, 1; s recycle delay, 0.7 s; 5K to 10K scans; temperature, 0 °C. Trimethyl phosphate (TMP) was used as the external reference.

2D ^1H NMR experiments were performed on a UNITY 500 (supporting information).

RESULTS

Table 1 shows the thermodynamic parameters and T_m values for all the G·A-containing duplexes. All the double strands shown in Table 1 (except 5′GGGATC3′/5′GAGACC3′) formed duplexes, as indicated by their concentration-dependent T_m values (Figure S1, supporting information). As can be seen, the thermodynamic parameters and T_m values varied significantly when the sequences flanking the center two G·A pairs were changed. Among all the duplexes, 5′GCGAGC3′/5′GCGAGC3′ in the 5′py-GA-pu3′ group showed the highest stability and was chosen for the structural study by 2D NMR (supporting information, Figure S2). The T_m values of 5′GGGATC3′/5′GAGACC3′ probably reflect the association of only 5′GGGATC3′ (see below).

Figure 2 shows the imino proton NMR spectra of all 10 5′GA3′-containing duplexes. The peaks were identified by NOE, temperature-dependent spectra, or the knowledge that the imino protons of A·T base pairs resonate more downfield than that of a G·C pair. Both sequences, 5′GCGAGC3′ and 5′GTGAAC3′, form self-complementary duplexes; therefore, only three imino proton signals are observed (Figure 2A,C).

Starting with 5′GCGAGC3′ (Figure 2A), the non-hydrogen-bonded G imino proton resonates at 10.5 ppm, and the imino proton resonance of the end G·C pairs migrates and broadens as the temperature increases (data not shown). 5′GTGAAC3′ has a G3 imino proton that resonates at 10.5 ppm and a T2 imino proton at 14.0 ppm, typical of an A·T pair (Figure 2C). Duplex 5′GTGAGC3′/5′GCGAAC3′ is not self-

Table 1: Thermodynamic Parameters of Duplex Formation

group	sequence	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/mol K)	$\Delta G°$ [a] (kcal/mol)	T_m [b] (°C)
5′py-GA-pu3′	5′GCGAGC 3′CGAGCG	−48.4 ± 7	−134 ± 23	−8.58 ± 0.17	39.5
	5′GCGAAC 3′CGAGTG	−45.2 ± 3	−127 ± 13	−7.48 ± 0.04	30.0
	5′GTGAAC 3′CAAGTG	−45.9 ± 5	−136 ± 15	−5.30 ± 0.45	18.8
5′pu-GA-pu3′/3′py-GA-py5′	5′GGGAGC 3′CCAGCG	−43.2 ± 5	−119 ± 19	−7.76 ± 0.03	28.1
	5′GGGAAC 3′CCAGTG	−52.7 ± 3	−158 ± 10	−553 ± 0.11	14
	5′GAGAGC 3′CTAGCG	−58.7[d]	−181	~−4.6	13
	5′GAGAAG 3′CTAGTC	NA	NA	NA	~6
5′pu-GA-py3′	5′GGGACC 3′CCAGGG	−33.2 ± 6	−91.8 ± 20	−5.83 ± 0.17	18.0
	5′GAGATC 3′CTAGAG	−28.0 ± 6	−73.1 ± 28[d]	~−4.7 ± 0.38	~8
	5′GGGATC 3′CCAGAG	NA[c]	NA	NA	~8[e]
reference 5′GA3′/5′TC3′	5′GCGAGC 3′CGCTCG	−51.4 ± 5	−141 ± 20	−9.53 ± 0.17	40.2
	5′GGGACC 3′CCCTGG	−32.7 ± 3	−84.5 ± 10	−7.52 ± 0.38	32.3
	5′GTGAAC 3′CACTTG	−43.6 ± 2	−124 ± 8	−6.57 ± 0.10	20.3
	5′GAGATC 3′CTCTAG	−53.5 ± 7	−162 ± 22	−5.21 ± 0.62	13.4

[a] $\Delta G°$ values were derived at 25 °C. [b] T_m values were obtained in a buffer of 0.01 M PIPES, 1 M NaCl, and 1 mM EDTA, pH 7.0, and the strand concentration was ~1.2×10^{-5} M. [c] Data are not available. [d] Data for this duplex are only qualitatively taken because of the low T_m values and, therefore, larger errors (≥20%). [e] This T_m value is more likely for the self-associated 5′GGGATC3′ (Figure S4).

complementary; consequently, six imino proton signals are observed (Figure 2B): two of them are from the center G•A pairs (~10.5 ppm), and two are overlapped at 13.4 ppm from the first and the last G•C pairs (Figure S3).

In the 5′pu-GA-pu3′/5′py-G-py3′ group, all four duplexes show the imino proton signals about 10.5 ppm. 5′GGGAAC3′/GTGACC3′ shows four well-separated resonances at the typical Watson−Crick paired region other than at 10.5 ppm, corresponding to the four normal base pairs. NOE experiments assigned the peaks as in Figure 2E. With 5′GGGAGC3′/5′GCGACC3′, four imino proton signals are located at 13.2−13.7 ppm, the regular G•C pair region (Figure 2D). We were unable to get NOE difference spectra because of a high noise level (the sample strand concentration is less than 0.025 mM; at higher concentrations the purine-rich strands tend to be self-associated). The last two duplexes have very low T_m values and high dynamic properties, indicated by the broad and out-of-proportion area. The peaks at ~10.5 ppm, however, are clearly observable and, therefore, are tentatively assigned as the imino protons from the center G•A pairs (Figure 2F,G).

The 5′pu-GA-py3′ group gives relatively regular imino proton spectra; that is, all the imino proton signals appear in the hydrogen-bonded region (~12−14 ppm). This spectral pattern is obviously different from that of the previous two groups. We did not attempt to identify these peaks.

We observed a very similar imino proton spectral pattern for both 5′GGGATC3′/5′GAGACC3′ and 5′GGGATC3′ alone (Figure S4), confirming that oligomer 5′GGGATC3′ tends to self-associate in solution even at very low concentrations. Similar phenomena are also observed for 5′GGGAGC3′ and 5′GGGAAC3′ presumably due to the three consecutive G's (data not shown) at higher concentration. In order to obtain conclusive structural information, we designed two dodecamers, 5′GTACGGATGGCC3′ and 5′GGCCAGACGTAC3′, which are complementary to each other and have the same sequences flanking the center 5′GA3′ as in 5′GGGATC3′, i.e., 5′G-GA-T3/5′A-GA-C3′, but with longer Watson−Crick type base pairs. The formation of a duplex was confirmed by their downfield-shifted imino proton signals (Figure 2J). Integration of this area indicates that there are about 12 imino proton signals in the 12−14 ppm region, suggesting that only one species exists in the solution. No signal is observed in the region upfield of about 10.5 ppm.

Figure 3 shows the ^{31}P NMR spectra of the 10 duplexes. The assignments in Figure 3 are based on previous work (Greene et al., 1994; Li and Wilson, 1992). The unique split pattern of downfield (~-2 ppm) and upfield (~-5 ppm) phosphorus signals in the sequence 5′GCGAGC3′ is different from that of a regular Watson−Crick paired duplex (~-3.8 to -4.5 ppm) and is representative of the distorted backbone caused by two consecutive stacked G•A pairs (Figure 1) (see

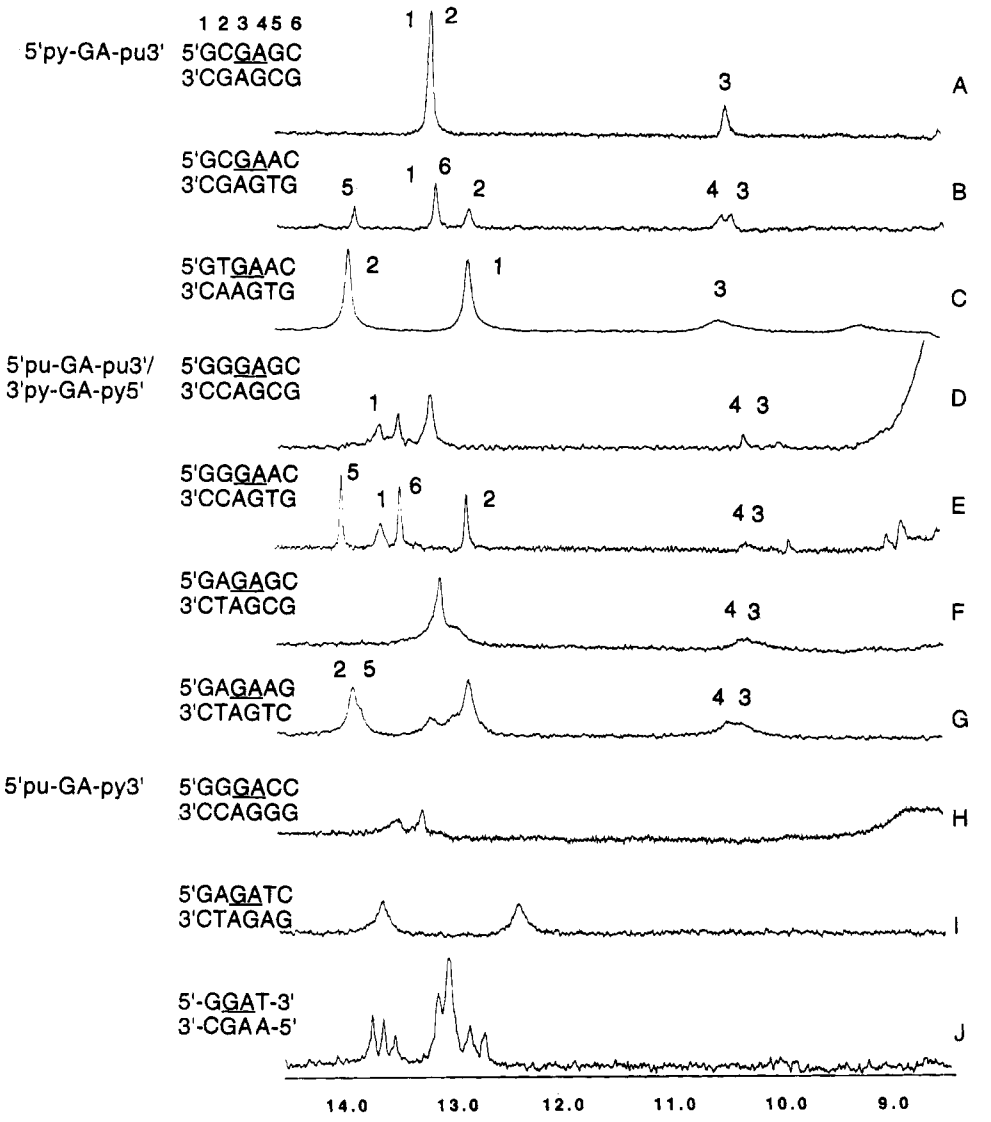

FIGURE 2: Imino proton NMR spectra of 5′GXGAYC3′, X and Y = A, C, G, or T, at 0 °C. Sequences are listed with each counterpart spectrum with the peak assignments. The last spectrum, 5′-GGAT-3′/5′-AGAC-3′, was obtained with a dodecamer duplex. 5′GTACG-GATGGCC3′/5′GGCCAGACGTAC3′, which has the same flanking sequences as shown in the figure but extended Watson–Crick base pairs to ensure the formation of a duplex.

Discussion). Of the 10 duplexes, seven, which cover all the sequences in the 5′py-GA-pu3′ group and the 5′pu-GA-pu3′/5′py-GA-py3′ group, demonstrate the same spectral shifting, indicating that they all have the same type of backbone distortion. The 5′pu-GA-py3′ group does not show such shifting in the spectra. The explicit difference in the ^{31}P NMR spectra between the first two groups and the third group parallels that of the imino proton NMR spectra. The dodecamer duplex, 5′GTACGGATGGCC3′/5′GGCCAGACG-TAC3′, which maintains the neighboring sequences in a duplex form, also shows parallel patterns of the ^{31}P and imino proton spectra.

We have also done 2D NMR experiments on 5′GC-GAGC3′. With the analysis of NMR assignments and molecular modeling, it is clear that the center G•A base pair adopt the pairing and stacking conformation shown in Figure 1 (supporting information).

DISCUSSION

In these studies, we have identified a probe for the structural conformation of G•A pairs and then correlated the G•A conformation and thermostability to the sequences that flank the G•A. The model sequence we used, 5′GCGAGC3′, has 5′pyrimidine–GA–purine3′ neighbors. The structural conformation and detailed NMR characteristics of our results (Table S1) agree well with several recent reports. For example, the sequence 5′GGACGAGTCC3′, reported by Katahira et al. (1993), has the same flanking sequences as in our 5′GCGAGC3′, and the detailed NMR characteristic and structural results at the 5′GA3′ sites are very similar. Greene et al. (1994) have reported the same G•A pairing and stacking in the sequence 5′CCATGAATGG3′. Another sequence that shares the 5′T-GA-A3′ flanking sequence, 5′GTGAACTT3′ (Lane et al., 1992), has very similar conformation and NMR features as well. The sequence 5′ATGAGCGAAT3′ (Li et al., 1991b) is the first duplex reported with two sets of such adjacent G•A pairs separated by two Watson–Crick base pairs. In this sequence, the adjacent G•A pairs are flanking as 5′T-GA-G3′ and 5′C-GA-A3′. The same type of G•A pairs, with the same flanking sequences, were observed in the sequence 5′GCGA-

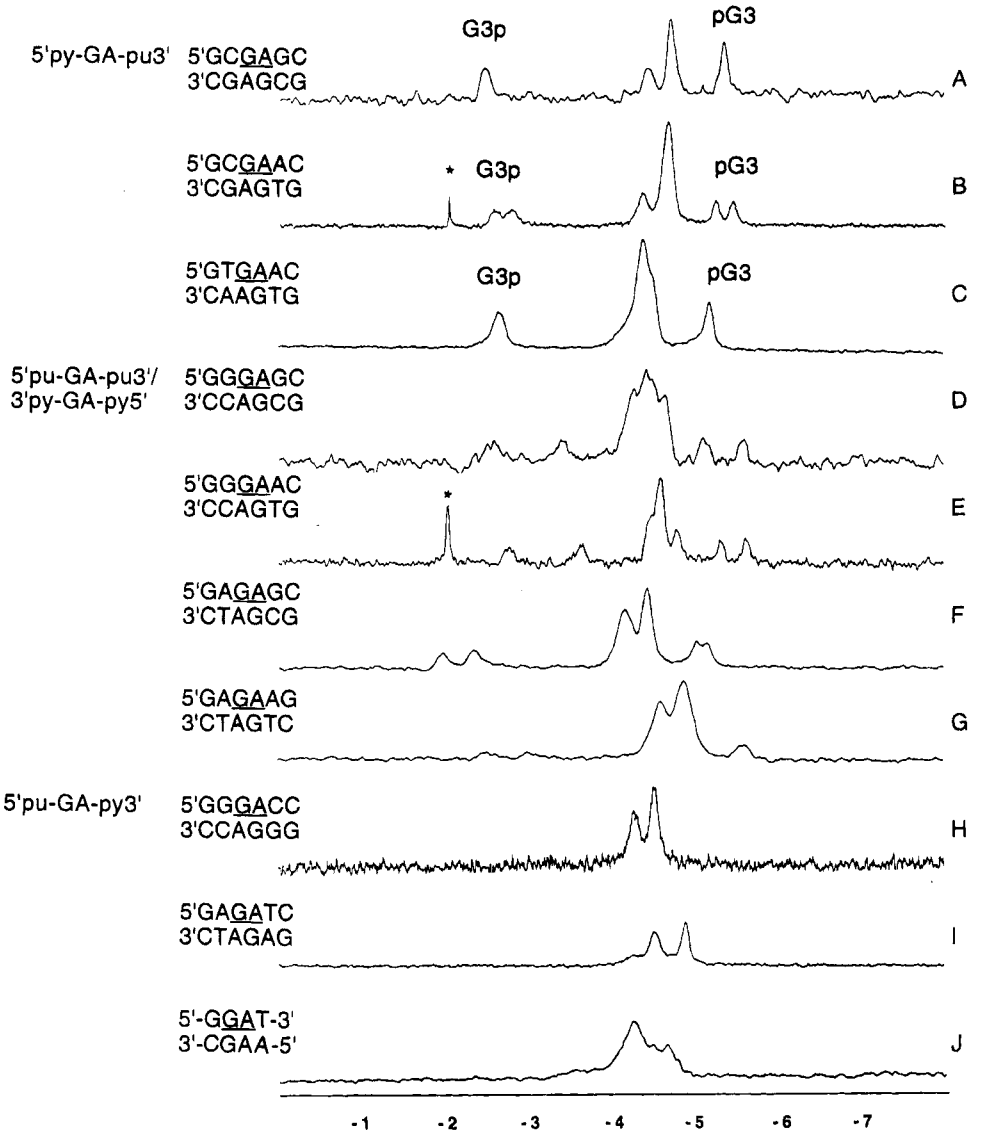

FIGURE 3: Phosphorus NMR spectra of 5′GXGAYC3′, X and Y = A, C, G, or T, at 0 °C. The order of the sequence layout is parallel to that in Figure 2. Impurities from the buffer are marked with asterisks (*).

ATAAGCG3′ by Maskos et al. (1993), although there is an A•A pair adjacent to a G•A pair.

The sequences mentioned above are all 5′pyrimidine−GA−purine3′ neighboring sequences. Given these results, we can conclude that G•A pairs in a duplex flanked by a 5′pyrimidine−GA−purine3′ sequence will adopt the conformation shown in Figure 1. Our goal here is to establish a quick and simple way to screen the G•A conformation in a given duplex using NMR. The work described above details unique and similar NMR characteristics (Table S1), such as strong NOEs between cross-stranded AH2 and AH1′ at the G•A sites; strong NOEs between cross-stranded GNH1 and GH8 or AH8; upfield-shifted proton signals of GNH1 (10−10.5 ppm), AH8 (~7.4 ppm), AH2′ (~1.1 ppm), and 5′pyrimidine H2′ (~1.5 ppm); and upfield- and downfield-shifted ^{31}P signals of 5′pG (~−5 ppm) and Gp3′ (~−2.5 ppm) at the G•A pair sites (Greene et al., 1994; Li and Wilson, 1992).

The most logical representative of this type of G•A pairs (Figure 1) appear to be the upfield-shifted G imino proton signals in the proton NMR spectra and the well-separated upfield- and downfield-shifted phosphorus pG and Gp signals in the ^{31}P NMR spectra. The reasons for this choice are as follows: (1) The spectral characteristics in both ^{1}H and ^{31}P NMR spectra are clear and unique: the non-hydrogen-bonded G imino proton signals in G•A pairs as in Figure 1 (~10.5 ppm) are about 2 ppm away from the hydrogen-bonded imino proton signals (~12.5 ppm) and are extremely well separated from the main band of resonances. The split ^{31}P NMR pattern, with 5′pG upfield ~0.5 ppm and Gp3′ downfield ~2 ppm from the main bands (Li and Wilson, 1992; Greene et al., 1994; Nikonowicz & Gorenstein, 1990, could also represent backbone distortion caused by the two adjacent G•A pairs' cross-strand stacking. (2) Imino protons reside inside a duplex, and phosphates residue outside a duplex. The combination of NMR spectra from both nuclei provides direct information on base pair structure and on backbone conformation. (3) The concurrent appearance of the upfield chemical shifts in imino proton pattern and the large upfield- and downfield-shifted phosphorus NMR spectra are highly consistent and can be called typical of this type of G•A pairing and stacking in a 5′GA3′ series. Duplexes in the 5′pyrimidine−GA−purine3′ group all show the coexisting phenomena of upfield-shifted imino proton signals and largely split phosphorus signals (Figures 2 and 3), and they

all adopt the base pairing and stacking pattern shown in Figure 1.

Flanking Sequence Requirement for G·A Pairs. We also observed the concurrent imino proton and phosphorus NMR features with all of the sequences in the 5′purine−GA−purine3′/5′pyrimidine−GA−pyrimidine3′ group. The remarkable spectral similarity between this group and the 5′pyrimidine−GA−purine3′ group strongly suggests that in all of the 5′purine−GA−purine3′/5′pyrimidine−GA−pyrimidine3′ duplexes the center 5′GA3′ pairs and stacks as shown in Figure 1. We were unable to perform 2D NMR on 5′GGGAGC3′/5′GCGACC3′ because of severe aggregation of the oligomers. Researhers have recently reported the results of extensive 2D NMR study on the duplex 5′GGACGACATC3′/5′GATGGAGTCC3′ (Katahira et al., 1993). They found that when the center 5′GA3′ was flanked as 5′GGA-G3′/5′C-GA-C3′, the 5′GA3′ pairs and stacks as in Figure 1. Although they did not show ^{31}P NMR results, their imino proton spectral data are similar to ours. More significantly, our probe is validated by their detailed structural data. Integrating both our and their results, our knowledge of the sequence requirement for this type of adjacent G·A pair can be extended from 5′pyrimidine−GA−purine3′ sequences to all 5′purine−GA−purine3′/5′pyrimidine−GA−pyrimidine3′ sequences. Indeed, Turner and colleagues have used upfield imino proton signals as an indicator for this type of G·A pair in their RNA studies (Walter et al., 1994). At this point, we have limited knowledge about concurrent splitting on both ^{1}H and ^{31}P NMR spectra in RNA sequences.

The 5′purine−GA−pyrimidine3′ group displays no upfield-shifted imino proton signals and no scattered ^{31}P signals. In this group, 5′A-GA-T3′ has been well-studied by NMR and X-ray crystallography methods for the sequence 5′CCAAGATTGG3′ (Kan et al., 1983; Nikonowicz & Gorenstein, 1990; Prive et al., 1987). The G·A pairs in this neighboring series take a conventional G(anti)·A(anti) conformation through G imino proton hydrogen bonding and do not exhibit cross-strand stacking. The imino proton of G therefore resonates in the normal downfield region, and the ^{31}P signals of phosphates at the adjacent G·A pair sites are very close to the main bands in ^{31}P NMR spectra.

In our study, the sequence 5′GAGATC3′ shows all the imino ^{1}H and ^{31}P signals in the typical regions, presumably taking the same conformation as mentioned above. The severe aggregation of sequence 5′GGGACC3′ at millimolar concentrations makes 2D experiments futile, and no detailed structure information can be given at this point. At low concentrations, however, the imino proton NMR signals at 13−14 ppm (Figure 2H) demonstrate that virtually all the imino protons are involved in hydrogen bonding. Our results on the formation of the duplex 5′GGGACC3′ differ from the results reported by Cheng et al. (1992), who found no duplex formation for the sequence 5′CTGGACAG3′. The completely different imino ^{1}H and ^{31}P NMR speectra of the sequences suggest that the G·A pairs in the 5′purine−GA−pyrimidine3′ group take a different conformation, presumably the conventional G(anti)·A(anti) form.

Stability. Duplex stability depends on the base sequence (Breslauer et al., 1986), with Watson−Crick base-paired sequences being the most stable, as is clearly demonstrated in the reference group, in which all the duplexes are Watson−Crick base-paired. For example, changing the flanking sequences of the center 5′GA3′/5′TC3′ base pairs results in a notable free energy and T_m change (Table 1). Duplex 5′GCGAGC3′/5′GCTCGC3′ ($\Delta G°$ = −9.53 kcal/mol) is more stable than duplex 5′GGGACC3′/5′GGTCCC3′ ($\Delta G°$ = −7.49 kcal/mol) by 2 kcal/mol even though they contain the same base composition. With two consecutive G·A pairs, the flanking sequence effect is even more pronounced. The sequence 5′GCGAGC3′ has $\Delta G°$ = −8.58 kcal/mol, and $\Delta G°$ is only −5.83 kcal/mol for the sequence 5′GGGACC3′, a difference of about 2.8 kcal/mol. The larger free energy difference as compared with the first two fully Watson−Crick base paired sequences in the reference group can be attributed to the structural change of the G·A pairs. For the sequences with A·T or T·A pairs flanking center G·A pairs, the same rationale, in principal should stand; however, the very low T_m values of these sequences make free energy calculation less accurate.

It has been proposed that extended purine−purine stacking contributes substantially to duplex stability (Li et al., 1991a). A comparison of $\Delta G°$ values of 5′GCGAGC3′ and 5′GGGACC3′ with those of their corresponding Watson−Crick paired duplexes, 5′GCGAGC3′/5′GCTCGC3′ and 5′GGGACC3′/5′GGTCCC3′, supports this proposal. Take the perfect Watson−Crick paired duplexes as a reference: $\Delta\Delta G°$(CGAG) = $\Delta G°$(5′GCGAGC3′/5′GCTCGC3′) − $\Delta G°$(5′GCGAGC3′) = −0.95 kcal/mol, and $\Delta\Delta G°$(GGAC) = $\Delta G°$(5′GGGACC3′/5′GGTCCC3′) − $\Delta G°$(5′GGGACC3′) = −1.69 kcal/mol. The value of $\Delta\Delta G°$(CGAG), −0.95 kcal/mol, being lower than that of $\Delta\Delta G°$(GGAC), −1.69 kcal/mol, indicates that 5′GCGAGC3′ is closer to its analogous Watson−Crick type duplex in stability than is 5′GGGACC3′. The extra stability in 5′GCGAGC3′ is furnished by the extensive G·A pairs' stacking. The conventional G(anti)·A(anti) pairs in the 5′purine−GA−pyrimidine3′ group do not have such excellent base stacking; as a result, the $\Delta\Delta G°$ values are consistently lower (Table 1).

Although the type of G·A pairing and stacking shown in Figure 1 exists in all the 5′purine−GA−purine3′/5′pyrimidine−GA−pyrimidine3′ duplexes, the $\Delta G°$ and T_m values vary greatly depending on the flanking sequences, presumably because more restricted neighboring sequence is demanded for the base stacking. Walter et al. (1994) have suggested that a 5′G stabilizes this type of G·A pair in RNA. In our DNA studies, the $\Delta G°$ comparison of 5′GGGAAC3′/5′GTGACC3′ (−5.53 kcal/mol with 5′G and 3′A) and 5′GAGAGC3′/5′GCGATC3′ (~-4.61 kcal/mol with 5′A and 3′G) in the 5′purine−GA−purine3′/5′pyrimidine−GA−pyrimidine3′ group suggests that 5′G is superior to 3′G in stabilizing duplexes with this type of G·A base pair. The closeness of the $\Delta G°$ values of 5′GGGAGC3′/5′GCGACC3′ (−7.76 kcal/mol with 5′G and 3′G) and 5′GTGAGC3′/5′GCGAAC3′ (−7.48 kcal/mol with 5′T and 3′G) suggests that 5′pyrimidine and 3′purine both predominantly contribute to the higher duplex stability with this type of G·A base pair. This conclusion takes into account that A·T pairs are less stable than G·C pairs because former lack one hydrogen bond.

In each of the three G·A-containing groups, duplexes with G·C pairs flanking the center G·A pairs have the highest stability, supporting previous work using RNA (Walter et al., 1994). This may be attributed to some structural consideration, such as base stacking and base propelling, other than the fact that G·C pairs have one more hydrogen

bond than A·T pairs. The oligomer DNAs are six bases long in this study, and we keep the bases at both ends constant in order to see only the nearest neighbor effect. Although we assume the second nearest neighbor effect is the same on each sequence, we cannot discount that the severe distortion of the nearest base pairs in some duplexes may actually affect the conformation of the second nearest neighbors, and this effect (if any) may not be observed because of the end effect.

Our preliminary molecular modeling results show that the extensive cross-strand purine stacking makes the G·A pairs very compact. Instead of bulging out as in conventional G(anti)·A(anti) conformation, the duplex at G·A sites is actually slimmer than a Watson–Crick paired duplex. This stacking requires a more defined conformation on the nearest flanking base pairs, and the combination of 5′pyrimidine and 3′purine is probably the best fit that brings the duplex back to its normal conformation immediately. While the conformation of the nearest flanking sequences cannot fully compensatefor the distortion caused by the purine stacking, the distortion may extend to the next nearest neighboring base pairs and destabilize the duplex. The 5′purine–GA–purine3′/5′pyrimidine–GA–pyrimidine3′ group may represent this case. With the least favorable flanking sequences, this type of purine stacking cannot balance the energy cost of the neighboring backbone distortion. The bulge at the G·A sites is more energy saving, and a totally different G·A conformation emerges, as in the 5′purine–GA–pyrimidine3′ group.

SUMMARY

We have systematically studied the effect of flanking sequences on the structure and stability of adjacent 5′GA3′/5′GA3′ base pairs using ^{1}H and ^{31}P NMR and T_m methods. Our results indicate that flanking sequences of the G·A pairs control the change in the G·A base pairing pattern, which is mainly responsible for the thermodynamic difference. The sequence effect on duplex stability is more significant for adjacent 5′GA3′/5′GA3′ base pairs than for their corresponding Watson–Crick base pairs. Our results show that, despite the drastically different stabilities, all sequences containing 5′GA3′ form duplexes with every combination of the Watson–Crick flanking base pairs, and the G·A pairs are present predominantly in G amino H-bonded form (Figure 1). This fortuitous outcome largely enriches our knowledge of basic nucleic acid base pairing.

As antisense oligonucleotides, mismatches in general lead to nonspecific bonding and low affinity between antisense oligonucleotides and the target sequences. With this type of G·A pairs, however, the flanking sequence requirement is so specific that G·A pairs can be advantages for antisense studies. We are trying to explore the possibility of using these G·A pairs as antisense agents. We have preliminary T_m results showing that the stabilities of several oligodeoxyribonucleotides (25-mers) with G·A base pairs duplexes with the target sense DNA sequence are comparable to those of the perfect Watson–Crick paired duplexes. The definition of "sense" and "antisense" may also extend to G·A pairs in addition to Watson–Crick base pairs, giving the proper flanking sequences.

SUPPORTING INFORMATION AVAILABLE

One table listing the ^{1}H NMR assignments of oligomer DNA 5′GCGAGC3′, one 2D spectrum of the same sequence, two 1D imino proton NOE spectra, and one figure showing $1/T_m$ versus ln C_t plots (9 pages). Ordering information is given on any current masthead page.

REFERENCES

Aboul-ela, F., Koh, D., & Tinoco, I., Jr. (1985) *Nucleic Acids Res. 13*, 4811–4824.
Agrawal, S. (1992) *Trends Biotechnol. 10*, 152–158.
Agrawal, S., & Iyer, R. P., (1995) *Curr. Opin. Biotechnol. 6*, 12–19.
Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.
Cheng, J.-W., Chou, S.-H., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 1037–1041.
Chou, S.-H., Cheng, J.-W., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 138–155.
Ebel, S., Lane, A., & Brown, T. (1992) *Biochemistry 31*, 12083–12086.
Fasman, G. (1975) *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, Vol. 1, CRC Press, Cleveland.
Greene, K. L., Jones, R. L., Li, Y., Robinson, H., Wang, A. H.-J., Zon, G., & Wilson, W. D. (1994) *Biochemistry 33*, 1053–1062.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.
Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988) *Biochemistry 27*, 108–115.
Kan, L.-S., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.
Katahira, M., Sato, H., Mishima, K., Uesugi, S., & Fujii, S. (1993) *Nucleic Acids Res. 21*, 5418–5424.
Lane, A., Martin, S. R., Ebel, S., & Brown, T. (1992) *Biochemistry 31*, 12087–12095.
Li, Y., & Wilson, W. D. (1992) Ph.D. Thesis, Georgia State University, Atlanta, GA (unpublished data).
Li, Y., Zon, G., & Wilson, W. D. (1991a) *Biochemistry 30*, 7566–7572.
Li, Y., Zon, G., & Wilson, W. D. (1991b) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.
Maltseva, T., Sandstrom, A., Ivanova, I. M., Sergeyev, D. S., Zarytova, V. F., & Chattopadhyaya, J. (1993) *J. Biochem. Biophys. Methods 26*, 172–236.
Maskos, K., Gunn, B. M., LeBlanc, D. A., & Morden, K. M. (1993) *Biochemistry 32*, 3583–3595.
Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.
Orbons, L. P. M., van der Marel, G. A., van Boom, J. H. & Altona, C. (1987) *Eur. J. Biochem. 170*, 225–239.
Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1994) *Fed. Proc. 43*, 2663–2670.
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.
Stein, C. A., & Cheng, Y. (1993) *Science 261*, 1004–1012.
Walter, A. E., Wu, M., & Turner, D. H. (1994) *Biochemistry 33*, 11349–11354.
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.
Werntges, H., Steger, G., Riesner, D., & Fritz, H.-J. (1986) *Nucleic Acids Res. 14*, 3773–3790.
Wilson, W. D., Dotrong, M.-H., Zou, E. T., & Zon, G. (1988) *Nucleic Acids Res. 16*, 5137–5151.
Zamecnik, P. C., & Stephenson, M. L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 280–284.

BI950521C