Davis, J. H., Jeffrey, K. R., Bloom, M., Valic, M. I., & Higgs, T. P. (1976) *Chem. Phys. Lett. 42*, 390-394.
Davis, J. H., Jeffrey, K. R., & Bloom, M. (1978) *J. Magn. Reson. 29*, 191-199.
Dufourc, E. J., & Smith, I. C. P. (1986) *Chem. Phys. Lipids 41*, 123-135.
Emsley, J. W., & Lindon, J. C. (1965) *NMR Spectroscopy Using Liquid Crystal Solvents*, Pergamon, Oxford.
Forrest, B. J., & Reeves, L. W. (1981a) *J. Am. Chem. Soc. 103*, 1641-1647.
Forrest, B. J., & Reeves, L. W. (1981b) *Chem. Rev. 81*, 1-14.
Hladky, S. B., & Haydon, D. A. (1972) *Biochim. Biophys. Acta 274*, 294-312.
Hunt, M. J., & MacKay, A. L. (1974) *J. Magn. Reson. 15*, 402-414.
Hunt, M. J., & MacKay, A. L. (1976) *J. Magn. Reson. 22*, 295-301.
Kar, L., Ney-Igner, E., & Freed, J. (1985) *Biophys. J. 48*, 569-595.
Killian, J. A., & de Kruijff, B. (1986) *Chem. Phys. Lipids 40*, 259-284.
Killian, J. A., Borle, F., de Kruijff, B., & Seelig, J. (1985) *Biochim. Biophys. Acta 854*, 133-142.
Koeppe, R. E., & Schoenborn, B. P. (1984) *Biophys. J. 45*, 503-507.
Koeppe, R. E., Hodgson, K. O., & Stryer, L. (1978) *J. Mol. Biol. 121*, 44-54.
Koeppe, R. E., Berg, J. M., Hodgson, K. O., & Stryer, L. (1979) *Nature (London) 279*, 723-725.
Moll, F., Nicholson, L. K., LoGrasso, P. V., Lay, J. C., Guy, C. A., Petefish, J., Fields, B. G., Van Wart, H. E., & Cross, T. A. (1987) *Biophys. J. 51*, 73a.
Momany, F. A., McGuire, R. F., Van, J. F., & Scheraga, H. A. (1971) *J. Phys. Chem. 75*, 2286-2297.
Momany, F. A., McGuire, R. F., Burgess, A. W., & Scheraga, H. A. (1975) *J. Phys. Chem. 79*, 2361-2381.
Morris, G. A., & Freeman, R. (1978) *J. Magn. Reson. 29*, 433-462.
Morrow, M. R., & Davis, J. H. (1988) *Biochemistry* (in press).
Pauling, L., Corey, R. B., & Branson, H. R. (1951) *Proc. Natl. Acad. Sci. U.S.A. 37*, 205-211.
Pauls, K. P., MacKay, A. L., Soderman, O., Bloom, M., Tanjea, A. K., & Hodges, R. S. (1985) *Eur. Biophys. J. 12*, 1-11.
Tanaka, H., & Freed, J. (1984) *J. Phys. Chem. 88*, 6633-6644.
Tanaka, H., & Freed, J. (1985) *J. Phys. Chem. 89*, 350-360.
Taylor, M. G., Akiyama, T., & Smith, I. C. P. (1981) *Chem. Phys. Lipids 29*, 327-339.
Urry, D. W. (1985) *The Enzymes of Biological Membranes* (Martonosi, A. N., Ed.) 2nd ed., Vol. 1, pp 229-255, Plenum, New York.
Urry, D. W., Trapane, T. L., & Prasad, K. U. (1983) *Science (Washington, D.C.) 221*, 1064-1067.
Venkatachalam, C. M., & Urry, D. W. (1983) *J. Comput. Chem. 4*, 461-469.
Venkatachalam, C. M., & Urry, D. W. (1984) *J. Comput. Chem. 5*, 64-71.
Wallace, B. A. (1983) *Biopolymers 22*, 397-402.
Weinstein, S., Wallace, B. A., Morrow, J. S., & Veatch, W. R. (1980) *J. Mol. Biol. 143*, 1-19.

Preferential Location of Bulged Guanosine Internal to a G·C Tract by ^{1}H NMR†

Sarah A. Woodson and Donald M. Crothers*
Department of Chemistry, Yale University, New Haven, Connecticut 06511
Received June 25, 1987; Revised Manuscript Received September 10, 1987

ABSTRACT: A series of double-helical oligodeoxyribonucleotides of sequence corresponding to a frame-shift mutational hot spot in the λ C_I gene, 5′-dGATGGGGCAG, are compared by proton magnetic resonance spectroscopy at 500 MHz of the exchangeable protons. Duplexes containing an extra guanine in a run of two, three, and four G·C base pairs are compared to regular helices of the same sequence and to another sequence containing an isolated bulged G, 5′-dGATGGGCAG·dCTGCGCCATC. The imino proton resonances are assigned by one-dimensional nuclear Overhauser effect spectroscopy. Resonances assigned to the G tract in bulge-containing duplexes are shifted anomalously upfield and are very broad. Imino proton lifetimes are determined by T_1 inversion–recovery experiments. The exchange rates of G-tract imino protons in bulged duplexes are rapid compared to those in regular helices and are discussed in terms of the apparent rate of solvent exchange for the isolated G bulge. Delocalization of a bulged guanosine in homopolymeric sequences can explain the observed changes in chemical shift and relaxation times across the entire G·C run, and the chemical shifts can be fit by a simple model of fast exchange between base-paired and unpaired states for the imino protons. This allows us to calculate the relative occupancies of each bulge site. In these sequences, we find the extra base prefers positions internal to the G tract over those at the edge.

Frame-shift mutations are generally known to occur in homopolymeric runs of bases in DNA, presumably due to the ability of the duplex to slip without loss of base pairing (Lerman, 1963; Ames et al., 1973). Streisinger proposed such a "bulged-base" model in which one or more looped-out bases are stabilized by a mutagenic intercalator, causing plus or minus frame shifts following repair and replication (Streisinger et al., 1966; Streisinger & Owen, 1985). In addition to a dependence on the length of the homopolymeric run, flanking sequences can affect mutation frequencies as much as 50-fold (Miller, 1985; Calos & Miller, 1981; Skopek & Hutchinson,

†Supported by Grant GM21966 from the National Institutes of Health.

0006-2960/88/0427-0436$01.50/0

1984). For example, runs of G's flanked on the 3′ side by a C have higher frequencies than those with other flanking bases (Skopek & Hutchinson, 1984; Calos & Miller, 1981). Binding of mutagenic drugs to DNA increases frame-shift frequencies, and models for the stacking of intercalators with extrahelical bases have been proposed on the basis of physical experiments (such as NMR[1]) on small drug–DNA complexes (Lee & Tinoco, 1978; Young & Kallenbach, 1981). Furthermore, ethidium bromide has been shown to bind bulge-containing oligonucleotides more strongly than regular helices (Nelson & Tinoco, 1985).

The large variations in overall mutation frequencies as well as lesions at specific sequences indicate that many factors are at work, making it more difficult to sort out their respective contributions. One might expect that the structure and flexibility of the DNA helix, sequence specificity of drug binding and drug-induced changes in the helical conformation, and recognition of drug–DNA complexes by polymerase and repair enzymes would all play a role in mutagenesis, but the mechanisms by which these processes interact are not yet understood. The availability of large quantities of synthetic DNA oligomers as well as recent advances in magnetic resonance techniques has made it possible for a number of oligonucleotides containing mismatched, extrahelical, or modified bases to be characterized by proton NMR (Morden et al., 1983; Hare et al., 1986a,b; Patel et al., 1982). Small oligomers can serve as convenient models for premutagenic lesions, allowing one to address some aspects of a complex biological problem from the viewpoint of physical chemistry.

We have chosen to examine the role of DNA sequence by comparing the physical properties of small oligonucleotides on the basis of a mutational hot spot sequence in the λ C_I gene (Skopek & Hutchinson, 1984), 5′-GATGGGGCAG. We previously examined bulge-containing duplexes from this same sequence by proton NMR and found that the destabilization due to an extra cytosine in a run of C's is delocalized over the entire C tract, whereas the effect of an extra adenine is localized as one might expect (Woodson & Crothers, 1987). These data lead to the conclusion that the position of the bulged C migrates along the C·G run in the fast-exchange limit on the NMR time scale. The duplex seems to sample all of the base-pairing conformations available, in a process that has a low apparent energy of activation. Furthermore, degeneracy of the bulge position lowers the free energy of duplex formation.

Here we present data on duplexes of similar sequence that contain an extra G in runs of three, four, and five guanines and compare them to the appropriate regular duplexes. The sequences

5′-dGATGGGCAG CTAC-CGTC	5′-dGATGGGGCAG CTACC-CGTC
5′-dGATGGGGGCAG CTACC-CCGTC	5′-dGATGG-GCAG CTACCGCGTC
5′-dGATGGCAG CTACCGTC	5′-dGATGGGCAG CTACCCGTC
5′-dGATGGGGCAG CTACCCCGTC	5′-dGATGGGGGCAG CTACCCCCGTC

were chemically synthesized and compared by proton NMR. The imino proton resonances were assigned by one-dimensional NOE difference spectroscopy and their lifetimes measured by T_1 inversion–recovery experiments. Not only is the position of the bulged base averaged as in the case of the C tract, but the presence of a non-hydrogen-bonded imino proton on the unpaired guanine allows an estimate of the occupation of the bulge sites. Sequences containing a fixed bulged guanine and thymine serve as references for the unpaired state in the migrating bulge sequences.

Materials and Methods

Oligonucleotides were synthesized on a 10-μmol scale in an Applied Biosystems DNA synthesizer. Purification was by reverse-phase chromatography on a Vydac C_4 peptide column. Purity was checked by analytical chromatography and denaturing 20% polyacrylamide gel electrophoresis. Samples containing 1.0–1.5 μmol of duplex were dialyzed extensively against 1 M NaCl, 10 mM phosphate, pH 7.0, and 0.1 mM EDTA, followed by 83.9 mM NaCl, 10 mM sodium phosphate, and 0.1 mM EDTA, and dissolved in 350 μL of the same, with 5% D_2O. Sequences containing runs of four or more guanosines required heating at 90 °C for 10 min followed by cooling over several hours to room temperature to break up G·G aggregates.

^{1}H NMR spectra were acquired on the Yale 490 and Bruker WM-500 spectrometers, with twin-pulse observation (Kime & Moore, 1983) and ADA data acquisition (Roth et al., 1980) to suppress the water resonance. Chemical shifts were referenced to internal dioxane set at 3.471 ppm, relative to TSP. One-dimensional NOE difference spectra were acquired in a similar manner, with a 0.6–0.8-s irradiation time and 50% saturation of the desired resonance. A Lorentzian-to-Gaussian multiplication was applied to difference FIDs to enhance resolution. Inversion–recovery experiments used a $90°-D_1-90°-\tau-45°-D_2-45°$ pulse sequence with the carrier frequency in the center of the imino region and ADA acquisition with a 1.2-s relaxation delay. T_1 values were determined by a nonlinear least-squares program. Error in the relaxation rates was generally within 5%, except where chemical shift resolution was poor.

Optical melting curves were recorded on a Cary 219 UV–vis spectrometer. Samples were in the same buffer as for proton NMR experiments, with total strand concentrations from 5 to 25 μM. Data were fit to a two-state model by a nonlinear least-squares program made available by J. G. Nadeau. Extinction coefficients were calculated from values given by Janik (1971).

Results

Imino Proton Spectra of Regular 10-mer. The imino region of the proton spectrum of the regular 10-mer "C4G4" at 490 MHz is shown in Figure 1, and its sequence is diagrammed in Chart I. The imino protons of each neighboring base pair in a regular DNA helix are near enough to cross-relax with each other, and the imino resonances can be readily assigned in a sequential fashion by using one-dimensional or two-dimensional NOE spectroscopy (Roy & Redfield, 1981). The imino resonances of this molecule and of all the duplexes shown in Chart I were assigned by one-dimensional NOE difference spectroscopy, and their chemical shifts are listed in Table I.

The imino spectrum of the regular helix corresponds to what is usually observed for short DNA double-stranded helices (Pardi et al., 1981), as previously described for the perfect 9-mer from this sequence (Woodson & Crothers, 1987). All the resonances are resolved, and the broadening of the imino resonances due to solvent exchange proceeds in an orderly fashion beginning with terminal resonances (G1 and G10) at 283 K and proceeding with increasing temperature to resonances assigned to T2 and T8, followed by those of the internal

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; FID, free induction decay; EDTA, ethylenediaminetetraacetic acid; TSP, sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d*$_4$; HPLC, high-performance liquid chromatography.

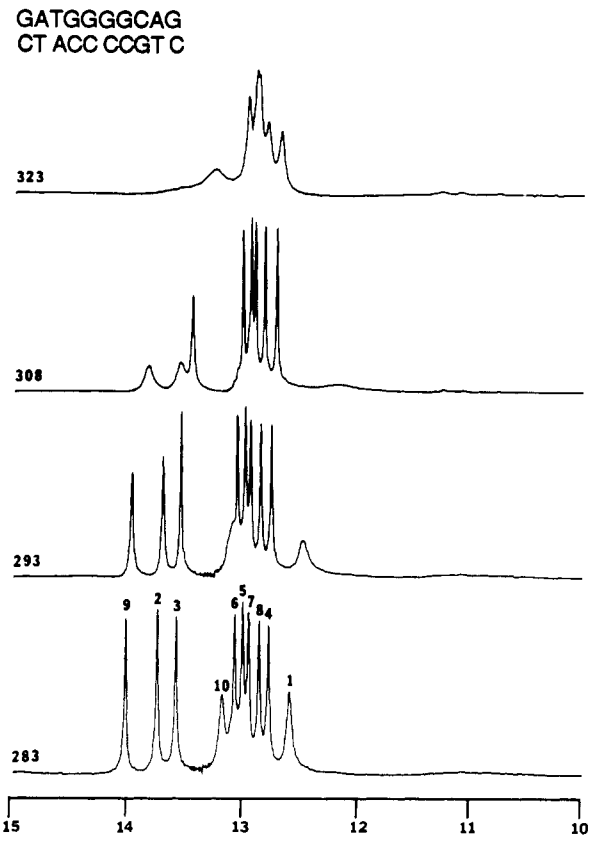

FIGURE 1: Proton magnetic resonance spectrum of 5′-dGATGGGGCAG·dCTGCCCCATC (C4G4) at 490 MHz, in 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0, with 5% D_2O. Chemical shifts are relative to TSP.

Chart I

C2G2	1 2 3 4 5 6 7 8 5′-d G A T G G C A G C T A C C G T C
C2G3	1 2 3 4 5 6 7 8 9 5′-d G A T G G G C A G C T A C - C G T C
C3G3	1 2 3 4 5 6 7 8 9 5′-d G A T G G G C A G C T A C C C G T C
C3G4	1 2 3 4 5 6 7 8 9 10 5′-d G A T G G G G C A G C T A C C - C G T C
C4G4	1 2 3 4 5 6 7 8 9 10 5′-d G A T G G G G C A G C T A C C C C G T C
C4G5	1 2 3 4 5 6 7 8 9 10 11 5′-d G A T G G G G G C A G C T A C C - C C G T C
C5G5	1 2 3 4 5 6 7 8 9 10 11 5′-d G A T G G G G G C A G C T A C C C C C G T C
CG4G3	1 2 3 4 5 6 7 8 9 10 5′-d G A T G G - G C A G C T A C C G C G T C
CT4G3	1 2 3 4 5 6 7 8 9 10 5′-d G A T G G - G C A C C T A C C T C G T C

G·C base pairs, which broaden above 323 K. The spectra of the regular 8-mer, C2G2, 9-mer, C3G3, and 11-mer, C5G5 (data not shown) are very similar, except the imino resonances of the central base pairs broaden at higher or lower temperatures, depending on the length of the helix.

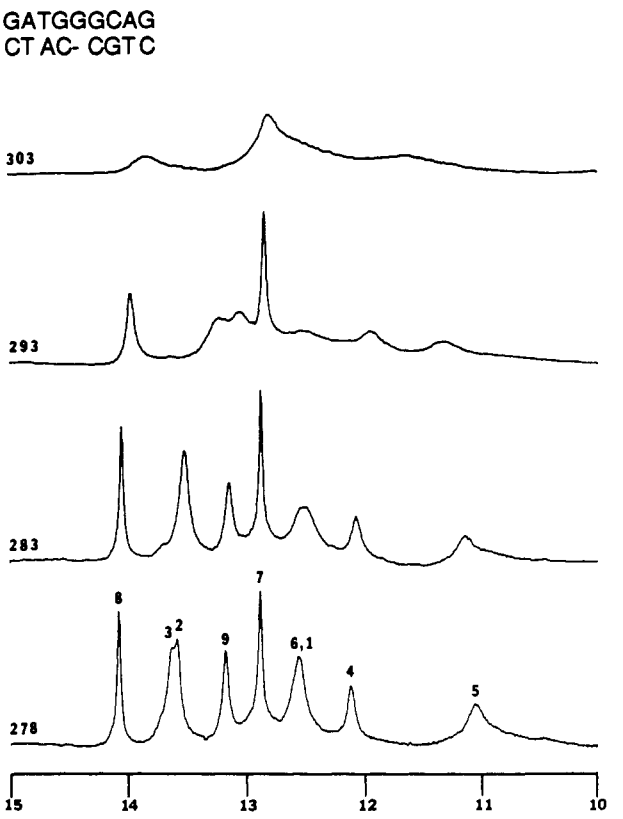

FIGURE 2: Proton NMR spectrum of 5′-dGATGGGCAG·dCTGCCATC (C2G3) at 490 MHz, as described in the legend to Figure 1.

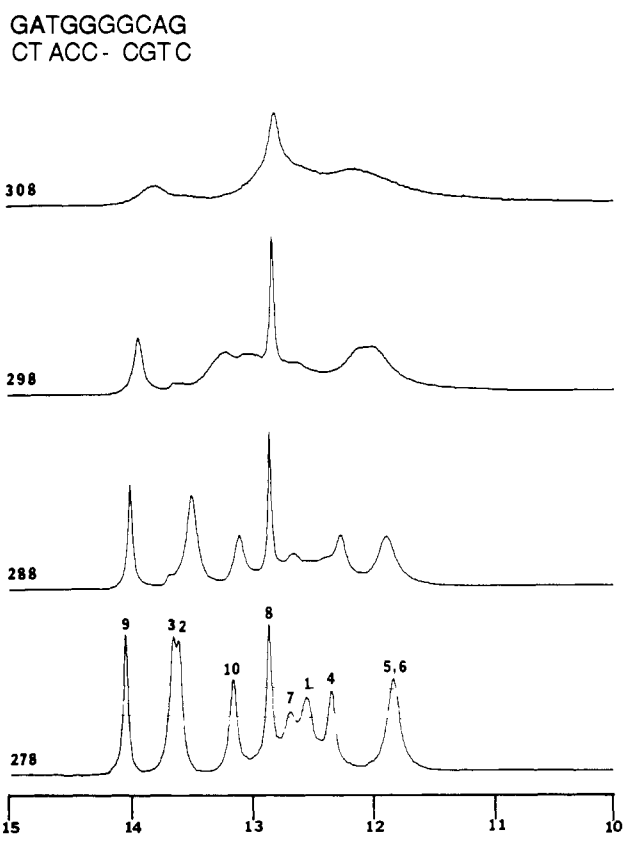

FIGURE 3: Proton NMR spectrum of 5′-dGATGGGGCAG·dCTGCCCATC (C3G4) at 490 MHz, as described in the legend to Figure 1.

Imino Proton Spectra of Bulge-Containing Duplexes. The imino regions of the proton spectra of the bulge-containing

Table I: Chemical Shifts of Imino Resonances (ppm)[a]

base	C2G2	C2G3	C3G3	C3G4	C4G4	C4G5	C5G5	CG4G3	CT4G3
1	12.60	12.56	12.60	12.56	12.58	12.58	12.63	12.63	12.63
2	13.72	13.65	13.75	13.63	13.74	13.67	13.78	13.73	13.74
3	13.57	13.59	13.59	13.65	13.57	13.67	13.60	13.73	13.66
4	12.80	12.10	12.76	12.36	12.76	12.47	12.78	12.82	12.74
5	12.90	11.03	13.04	11.83	12.99	12.05	13.00	12.92	13.12
6	12.82	12.56	12.94	11.83	13.06	12.14	13.04	10.47	11.20
7	14.01	12.87	12.84	12.68	12.94	12.32	13.08	12.92	12.74
8	13.17	14.09	14.03	12.86	12.85	12.81	12.96	12.92	12.91
9		13.18	13.18	14.06	14.02	12.88	12.86	14.14	14.09
10				13.17	13.17	14.07	14.05	13.19	13.17
11						13.19	13.20		

[a] Relative to TSP, at 278 K, in 10 mM phosphate and 100 mM sodium, pH 7.0.

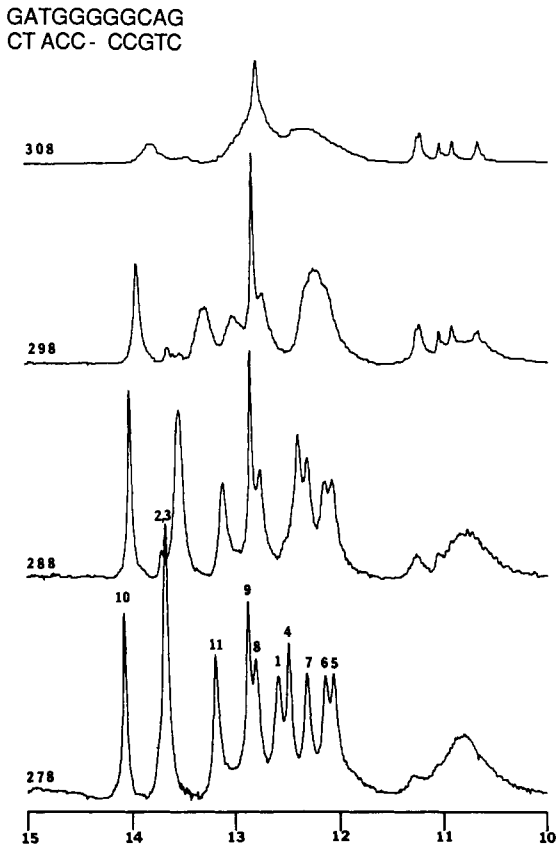

FIGURE 4: Proton NMR spectrum of 5′-dGATGGGGGCAG·dCTGCCCCATC (C4G5) at 490 MHz, as described in the legend to Figure 1. Peaks of smaller intensity between 10 and 11 ppm correspond to an extremely stable G·G aggregate which could not be entirely eliminated from this sample.

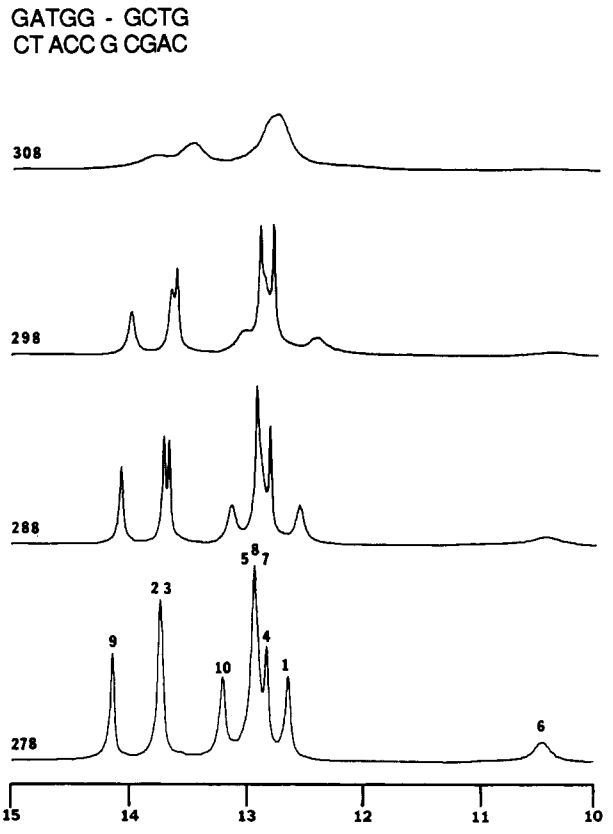

FIGURE 5: Proton NMR spectrum of 5′-dGATGGGCAG·dCTGCGCCATC (CG4G3) at 490 MHz, as described in the legend to Figure 1.

duplexes are shown in Figures 2–5. In the sequences C2G3, C3G4, and C4G5 (see Chart I), the bulge base is an extra guanine placed in the G tract, in such a way that one guanine is not able to pair with a cytosine on the opposite strand. The sequences differ only in the length of the C·G run. In the proton spectra of these molecules, shown in Figures 2–4, several striking differences from the spectra shown in Figure 1 are immediately apparent. Most of the resonances are extremely broad, even at 278 K, and several of the resonances are shifted unusually far upfield for imino resonances in double-stranded DNA oligomers. These resonances are also very temperature sensitive and completely broaden at much lower temperatures than the few sharp resonances in each series.

Closer examination of each series of spectra in turn reveals a pattern. In Figure 2, the resonance at 11.00 ppm is assigned to the central guanosine in the G tract of C2G3, G5. This is very far upfield for an imino proton in a DNA duplex; comparison with the regular 9-mer shows that the chemical shift would normally be around 13 ppm. With increasing temperature, the resonances of G5, and also of G4 and G6, begin to broaden, while the resonances of G7, T8 and even T2 and T3 remain much sharper. CG7 appears to be the most stable base pair in this duplex, in contrast to the regular duplex where the most central base pairs (C·G 4, 5, and 6) are the most stable.

In the spectrum of C3G4 shown in Figure 3, the most upfield resonance is now at 11.83 ppm and is again assigned to the most central C·G pairs, G5 and G6. Again, the resonances of all the guanines in the G tract broaden at much lower temperatures than the other resonances, and the C·G base pair flanking the G tract (G8) is the most stable. While the chemical shifts of the A·T base pairs and the terminal G·C base pairs are similar in each G-bulge duplex, the chemical shifts of the central G·C pairs have moved downfield.

These trends become even more apparent with consideration of C4G5, shown in Figure 4. The most central G imino resonances are still the furthest upfield, but less so than in the previous duplexes, at 12.05 ppm for G5 and 12.14 ppm for

G6. Again, the resonances corresponding to the G tract broaden at lower temperatures than the other resonances, and G9 at 12.88 ppm is the sharpest resonance. The small peaks around 11 ppm are due to residual G·G aggregate which is extremely stable and could not be entirely removed in these samples.

Comparison to Isolated G Bulge. Our previous work showed that the effect of an extra cytosine in a run of C's is delocalized over the entire C tract (Woodson & Crothers, 1987), and so it seemed reasonable that the effect of an extra G should be similarly spread over a run of G's. This is in fact what we observe. In all three duplexes with G tracts of varying lengths, all of the resonances assigned to the guanines in the G run are broadened, even in C4G5 where there are four G·C pairs. The imino resonances of the C-bulge-containing molecules, however, were not substantially shifted from those of the regular 9-mer and were narrow at low temperatures, whereas those of the G-bulge duplexes are very broadened and shifted significantly upfield relative to the regular helices.

Since we are observing imino proton resonances which are associated with the guanine in G·C base pairs, the G-bulge duplexes will contain at least one unpaired imino proton. Assuming delocalization of the bulged G over the G tract, isomerization of a non-hydrogen-bonded imino proton could explain the apparent differences in the proton spectra. In order to test this idea, we synthesized a molecule with an extra guanine in a fixed position, "CG4G3", whose imino proton spectra are shown in Figure 5. As expected, the resonance corresponding to the unpaired guanine imino proton is visible at low temperatures, although it is extemely broad, and falls at 10.5 ppm. Furthermore, the effects of the fixed bulge are localized, as seen in the A-bulge duplex examined previously (Woodson & Crothers, 1987). Although the resonances are not well resolved in this sequence, careful inspection of the spectra shows that the imino resonances flanking the bulge, G5 and G7, broaden at lower temperatures than do G4 and G8. This confirms that an unpaired G imino is observable in these sequences and that its chemical shift would fall between 11.0 and 10.5 ppm.

The imino proton spectrum of a duplex containing an extra thymine inserted into the helix at the same position ("CT4G3") is shown in Figure 6. The chemical shift of the unpaired thymine imino proton falls at 11.2 ppm, which agrees with chemical shift values of thymine imino protons in hairpin loops (Ikuta et al., 1986; Haasnoot et al., 1983). We observe the difference in chemical shift for unbonded imino protons on a thymine compared to a guanine base to be 0.7 ppm, with identical flanking sequence.

Comparison of Chemical Shifts. The chemical shifts of the imino protons in each duplex are diagrammed in Figure 7. The lines drawn in Figure 7A connect analogous protons in each bulge-containing duplex. In Figure 7B, the correspondence between analogous protons in bulged (solid circles) and regular duplexes (open circles) is shown. The shortest sequence, C2G3, is in the lower part of the diagram, and the longest sequence, C5G5, is at the top. The diagram shows that the chemical shifts of the guanines in a bulge-containing G tract are intermediate between those in the regular helices and the isolated G bulge. As the length of the G run becomes longer, the chemical shifts of the central resonances in the bulge-containing sequences move downfield, more like that of a normal G·C pair, and less like that of the bulged G in CG4G3. Furthermore, the proton resonances of the G's on the edges of the run are less affected, and the chemical shift of the C·G pair 3′ to the G run is nearly the same in each bulged duplex and barely changed from the regular duplexes. While the chemical shifts of all the guanine imino protons in the G tract are affected by the presence of a bulged nucleotide, those in the center of the tract are affected more than those at the edge, and the change becomes less apparent as the run of guanines becomes longer.

Table II: T_1 Relaxation Times of C2G3 (ms)[a]

temp (K)	1	2	3	4	5	6	7	8	9
278	22	95	83	40	21		123	136	48
283	15	49	44	25	15		147	98	30
288	19	21	21	15	7		118	52	18
293	3	9	9	8	6		67	27	12
298				4	4		32	13	6
303							7	5	

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0.

Table III: T_1 Relaxation Times of C2G2 (ms)[a]

temp (K)	1	2	3	4	5	6	7	8
278	38	163	253	203	195	200	124	33
283	22	121	289	270	227	264	86	21
288	13	70	258	306	276	292	50	13
293	5	33	174	329	277	298	24	6
298	4	19	97	321	277	272	13	4
303		9	36	189	176	154	6	
308			12	58	57	49		
313				10	10	9		

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0.

Table IV: T_1 Relaxation Times of C3G4 (ms)[a]

temp (K)	1	2	3	4	5	6	7	8	9	10
278	18	70	70	29	19	19	19	156	118	33
283	11	34	34	22	13	13	15	159	75	18
288	11	19	19	23	12	12	22	161	50	16
293		12	12	13	9	9	11	104	26	11
298		6	6	7	7	7	10	57	16	9
303				7	7	7	11	22	8	8
308								9	5	

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0.

Table V: T_1 Relaxation Times of C3G3 (ms)[a]

temp (K)	1	2	3	4	5	6	7	8	9
278	37	202	217	262	240	247	252	177	32
283	40	230	388	407	377	362	380	182	29
288	21	116	314	318	322	302	317	96	11
293	11	66	235	335	326	299	315	54	16
298	6	37	164	338	328	292	319	29	14
303		17	75	324	349	277	270	15	
308		7	36	264	284	188	176	7	
313			14	82	88	78	68		
318				22	25	24	21		

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0.

Table VI: T_1 Relaxation Times of C4G5 (ms)[a]

temp (K)	1	2	3	4	5	6	7	8	9	10	11
278	35	76	76	22	22	24	26	27	128	114	33
283	28	44	44	16	19	18	21	20	157	80	22
288	18	23	23		13	14	18	15	155	52	13
293	10	14	14		8	7	10	10	105	29	10
298	7	7	7		7	7	6	7	77	17	8
303					4	4	6		41	9	2
308									20	5	
313									8	2	

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0.

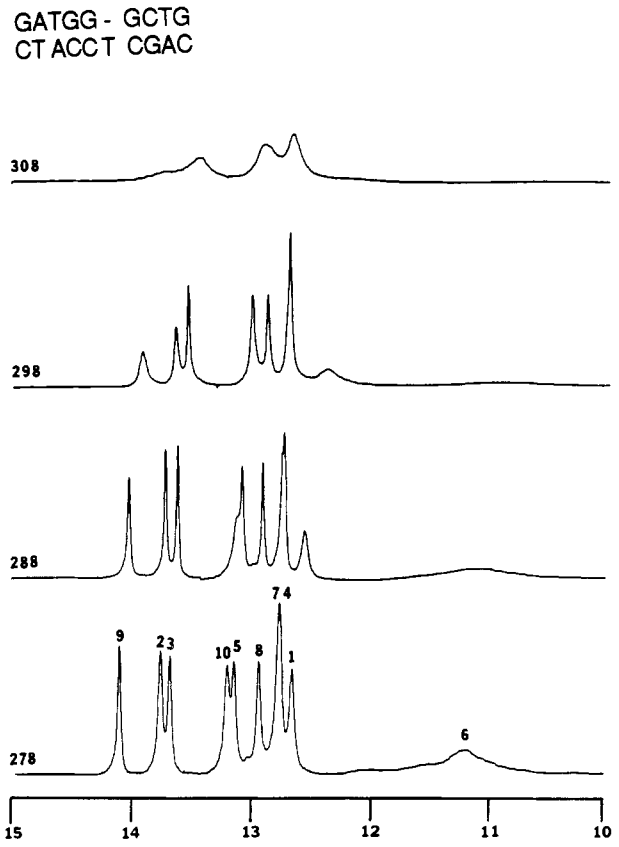

FIGURE 6: Proton NMR spectrum of 5'-dGATGGGCAG·dCTGCTCCATC (CT4G3) at 490 MHz, as described in the legend to Figure 1.

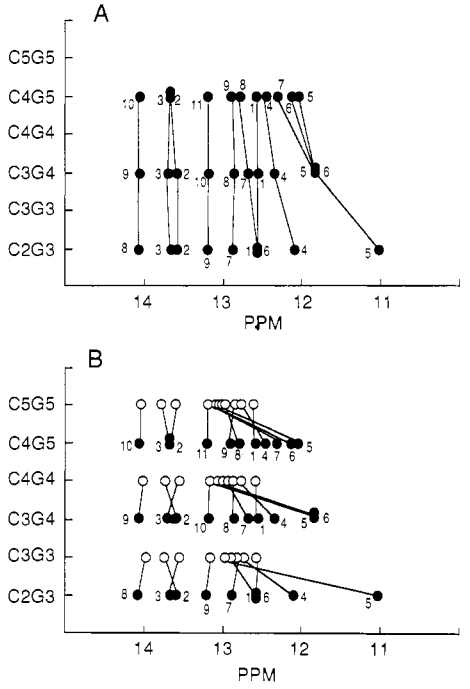

FIGURE 7: Comparison of chemical shifts of bulge-containing and regular duplexes. (A) Chemical shifts of imino proton resonances at 278 K of the G-bulge-containing duplexes are plotted along the horizontal axis, with the sequences arranged on the vertical axis from shortest (bottom) to longest (top). Lines connect resonances assigned to an analogous base pair in each duplex. The numbers indicate the assignment of the imino resonance (refer to Chart I). (B) Chemical shifts of imino proton resonances at 278 K are plotted as in (A). Open circles indicate regular duplexes, and solid circles indicate bulge-containing duplexes. The lines connect resonances assigned to analogous base pairs in each duplex, and the assignments are indicated as above.

Table VII: T_1 Relaxation Times of C4G4 (ms)[a]

temp (K)	1	2	3	4	5	6	7	8	9	10
278	43	198	277	191	234	239	199	220	146	36
283	18	151	296	231	304	260	252	252	121	25
288	12	101	280	291	321	265	291	288	86	15
293	10	64	230	298	303	307	326	320	48	
298	6	39	159	339	368	372	314	335	30	
303		20	85	296	328	331	246	292	16	
308		13	41	230	302	362	235	211	9	
313			19	145	219	242	164	125		
318			8	56	81	91	72	51		
323				18	24	25	24	16		

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0.

Imino Proton Exchange Rates. The T_1 relaxation times for each duplex are listed in Tables II–VII. Twenty τ values at each temperature, in 10 mM phosphate and 100 mM sodium, pH 7.0, were fit to the equation

$$I_z = I_0 + e^{-k\tau+b}$$

where I_z is the observed intensity at time τ, I_0 is the equilibrium intensity at a long value of τ, and k is the observed relaxation rate, or $1/T_{1,\text{obsd}}$. Both magnetic and exchange components contribute to the observed relaxation of exchangeable protons, where

$$1/T_{1,\text{obsd}} = 1/T_{1,\text{magn}} + 1/\tau_{\text{exchange}}$$

At higher temperatures, exchange predominates over dipolar relaxation, and the observed relaxation rates can be used to estimate the rate of solvent exchange.

The theory of imino proton exchange of oligonucleotides has been described in detail (Pardi & Tinoco, 1982; Crothers et al., 1974), and this method has been used to characterize a number of duplex oligomers (Morden et al., 1983; Patel et al., 1982; Cheung et al., 1984; Quignard et al., 1985). The observed rate of exchange depends on the rates of helix opening and closing, and on the rate of base-catalyzed proton transfer (Crothers et al., 1974). Extrapolation to high buffer concentration is necessary to directly determine helix opening rates from rates of proton exchange (Leroy et al., 1985). Since we are mainly interested in perturbations of a given sequence due to the insertion of extrahelical nucleotides, we rely on comparison of proton exchange rates at analogous sites in a set of molecules.

The T_1 lifetimes of the bulge-containing duplexes C2G3, C3G4, and C4G5 are listed in Tables II, IV, and VI, and the lifetimes of the analogous regular duplexes are listed in Tables III, V, and VII for comparison. The relaxation rates show the same destabilization of the G-tract base pairs as described for the imino spectra. Comparison of C2G3 and C2G2 (both duplexes contain eight base pairs, but C2G3 contains nine imino protons) in Tables II and III shows that all the exchange rates in the bulge-containing molecules are much higher. The optical melting temperature of C2G3 is 40 °C, extrapolated to 2 mM strand concentration, while that of G2G2 is 50 °C at 2 mM strand. The lifetime of G5 imino, however, is extremely short for an internal base pair, only 21 ms at 278 K and 7 ms at 288 K. The lifetime of G4 is somewhat longer, 40 ms at 278 K, but is still very short compared to that of G7 (123 ms at 278K) or that of the regular 8-mer, where the lifetime of G4 is 203 ms at 278 K in these buffer conditions. Resonance overlap between G6 and G1 made it impossible to reliably determine the exchange rates of G6.

Comparison of the T_1 values for C3G4 and C3G3 listed in Tables IV and V gives a similar picture. As expected, the

Table VIII: T_1 Relaxation Times of CG4G3 (ms)[a]

temp (K)	1	2	3	4	5	6[b]	7	8	9	10
278	46	146	146	125		12		100	107	28
283	27	141	141	162	117	9	117	133	88	20
288	15	107	122	184	106	7	100	119	61	11
293	9	52	83	161	44	5	44	100	33	8
298	7	28	42	85		4		74	19	5
303			13	26				22	7	
308			4	6				6	3	

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0. [b] Lifetimes for these resonances are T_2 relaxation times taken from the natural line widths and are somewhat approximate.

Table IX: T_1 Relaxation Times of CT4G3 (ms)[a]

temp (K)	1	2	3	4	5	6[b]	7	8	9	10
278	37	192	242	177	174	5	177	227	128	31
283	22	143	266	220	172	4	213	271	88	18
288	15	93	222	206	174	3	173	266	54	14
293	9	45	131	180	137		152	213	28	10
298	6	24	71	129	89		163	110	16	
303		9	21	36	33		24	25	6	
308		5	5	9				8	3	

[a] In 10 mM phosphate, 100 mM sodium, and 0.1 mM EDTA, pH 7.0. [b] Lifetimes for these resonances are T_2 relaxation times taken from the natural line widths and are somewhat approximate.

bulge lowers the overall stability of the helix, which is reflected in the imino exchange rates and the optical melting temperatures, 46 °C for C3G4 compared to 60 °C for C4G4 at 2 mM strand. The central G·C pairs in C3G4, however, have extremely short lifetimes, as in C2G3. G5 and G6 have the same chemical shift, and only one common lifetime was determined for both resonances. G5, G6, and G7 all have lifetimes of 19 ms at 278 K, and G4 has a lifetime of 29 ms at 278K, compared to 250 ms in C3G3. The central resonances of the G tract are the most affected by the bulge, but the lifetimes of all the protons in the G tract are shortened, as was true for the C-bulge 9-mer reported earlier (Woodson & Crothers, 1987).

Finally, examination of the imino proton lifetimes of C4G5 and C5G5 which are listed in Tables VI and VII confirms the trend seen in the previous two cases. The lifetimes of G4, G5, G6, and G7 are comparable and are extremely short, around 22–26 ms at 278 K. G9 certainly has a shorter lifetime in the bulged molecule than in the regular helix, 157 ms in C4G5 compared to 252 ms for G8 in C4G4 at 283 K, and so is destablilized by the presence of the bulge, but does not undergo extremely fast solvent exchange like the G imino protons in the G tract. It is interesting to note that even as the length of the G·C run becomes longer, there is no substantial difference in the lifetimes of the imino protons, except that those on the very edge of the run have slightly slower exchange rates.

Optical Melting Temperatures. The melting temperature of each duplex was determined by fitting optical melting curves recorded at 260 nm to a simple two-state model at several concentrations, ranging from 3 to 25 μmol of strand. The T_m and thermodynamic parameters are listed in Table X. Comparison of perfect and bulge-containing duplexes shows greater destabilization of the shorter duplexes by the presence of the bulge. At the highest concentrations, the T_m of the 8-mer is lowered roughly 14 °C, that of the 9-mer 12 °C, and that of the 10-mer about 10 °C.

Discussion

The broad and temperature-sensitive imino proton resonances as well as the T_1 data for the bulge-containing duplexes presented above indicate destabilization of the entire G tract

Table X: Apparent Thermodynamic Parameters from Optical Melting Data

	concn (μM)[a]	T_m (°C)	ΔH (kcal)	ΔS (cal/deg)	mean ΔH (kcal)	mean ΔS (cal/deg)
C2G2	7.28	32.52	−60.6	−171	−59.7 ± 0.9[b]	−169 ± 3[b]
	14.74	34.53	−59.9	−170		
	23.60	35.30	−58.8	−167		
C2G3	5.34	15.85	−45.0	−129	−44.9 ± 0.3	−128 ± 1
	11.78	20.06	−44.6	−127		
	26.10	21.66	−45.13	−129		
C3G3	4.16	38.05	−70.0	−197	−69.4 ± 1.2	−196 ± 4
	11.52	41.53	−68.0	−191		
	22.80	43.27	−70.3	−198		
C3G4	8.63	30.18	−63.9	−185	−59.8 ± 4.1	−172 ± 13
	12.11	30.20	−59.7	−172		
	24.28	31.08	−55.8	−159		
C4G4	3.53	44.60	−77.0	−215	−75.9 ± 1.0	−212 ± 3
	9.19	46.59	−75.3	−210		
	20.90	47.07	−75.4	−211		
C4G5	3.17	35.92	−71.9	−205	−73.1 ± 5.3	−209 ± 17
	7.79	37.30	−68.5	−195		
	12.35	39.58	−79.0	−227		
C5G5	3.80	48.92	−88.4	−247	−89.0 ± 2.2	−249 ± 7
	8.33	50.93	−87.2	−243		
	16.80	52.04	−91.5	−257		
CG4G3	6.37	25.80	−64.8	−190	−61.9 ± 3.4	−181 ± 11
	7.75	25.47	−62.8	−184		
	23.30	28.58	−58.2	−169		

[a] Determined from absorbance at 60 °C, in 10 mM phosphate and 100 mM sodium, pH 7.0. [b] Standard deviation.

in those sequences where an extra guanine has been inserted in the run of G's. We previously observed a similar delocalization of the effect of a bulged cytosine in a run of C's and concluded that this was due to migration of the bulged C along the C·G run. In that case, the resonances were not detectably broader than those of the regular 9-mer. The relaxation times were similar to each other but increased somewhat (lifetime of 255 ms at 283 K) over the rate of exchange of the internal resonances of the regular 9-mer (380 ms at 283 K). Although the chemical shifts and relaxation times in the G-bulge molecules are very different from those of the C-bulge duplex, we wished to see if this difference could be attributed to the presence of a non-hydrogen-bonded imino proton on the unpaired guanine, and whether the observed chemical shifts and relaxation rates could be accounted for by a bulge migration model.

Chemical Shift. For a molecule in or near the fast-exchange limit on the NMR time scale, the chemical shift of the observed resonances should be a weighted average of the chemical shifts of the resonances in each conformer. Each duplex can adopt all of the pairing conformations shown in Figure 8. For example, C2G3 can form eight base pairs in three different combinations. The observed chemical shift of any of the resonances can be written as

$$\omega_{obsd} = f_A\omega_A + f_B\omega_B + f_C\omega_C$$

where ω is the chemical shift and f is the relative mole fraction of each conformer. In order to simplify solution of the equations, we assume that the chemical shifts of each proton in base-paired states are nearly the same (since the flanking bases are always the same, this is not unreasonable). We also assume that, for all the protons, the difference in chemical shift between an unpaired state and a base-paired state is some constant $\Delta\omega$. The equations can be rewritten for each proton in terms of ω_{obsd}, $\omega_{H\text{-}bond}$, and $\Delta\omega$. The set of equations is solved for each molecule, where

$$\omega_{i,obsd} = \omega_{i,H\text{-}bond} - \Delta\omega f_i$$

$$\sum_i f_i = 1$$

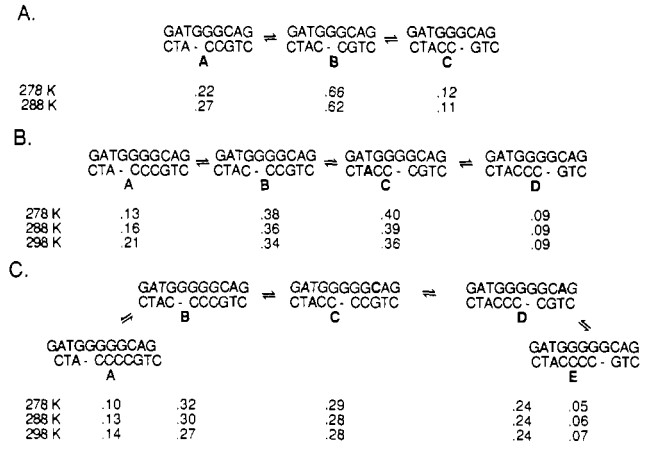

FIGURE 8: Isomerization of the bulge site in C2G3, C3G4, and C4G5. All the possible conformers of each bulge-containing duplex are diagrammed. The relative populations of each conformer were calculated as described in the text from the imino proton chemical shifts at 278, 288, and 298 K. The calculated relative population of a given conformer is indicated below the corresponding sequence diagram. (A) Bulge conformers of C2G3; (B) bulge conformers of C3G4; (C) bulge conformers of C4G5.

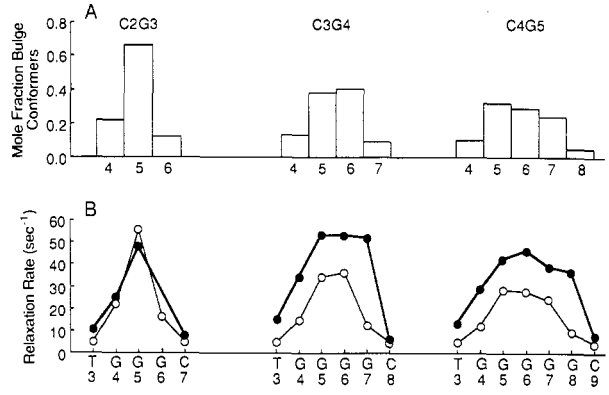

FIGURE 9: (A) Relative populations of bulge conformers at 278 K. The bars indicate the relative mole fraction of conformers where the guanine in the position indicated is unpaired, taken from Figure 8. The sequence is indicated above. (B) Imino proton T_1 relaxation rates (s^{-1}) are plotted against sequence position and correspond to the sequences and bulge conformers plotted above in (A). Solid circles and heavy lines are experimentally determined T_1 values, and open circles and light lines indicate values for these sequences calculated from relaxation rates taken from the isolated G-bulge duplex, CG4G3, and A-bulge 9-mer.

and $\omega_{i,\text{obsd}}$ = observed chemical shift of imino resonance i; $\omega_{i,\text{H-bond}}$ = chemical shift of resonance i when base paired, taken from the perfect helix; $\Delta\omega = \omega_{\text{H-bond}} - \omega_{\text{bulge}}$; and f_i = fraction of time imino i spends non base paired.

Solving these equations for the chemical shifts of C2G3, C3G4, and C4G5 (listed at 278 K in Table I; also see Figure 7) results in the relative populations at 278, 283, and 288 K for each conformer shown in Figure 8. In addition, one can calculate a value for $\Delta\omega$, which theoretically should be the same in each sequence. The agreement is remarkably good, with values ranging from 2.96 to 3.12 ppm at 278 K and from 2.75 to 3.07 ppm at 283 K. This results in a calculated chemical shift for the bulged G of 10.0–10.2 ppm, which, after differences in flanking sequence and rate of solvent exchange are taken into account, is in reasonably good agreement with the observed chemical shift of 10.47 ppm for the isolated bulged G in CG4G3.

Examination of the calculated distribution of the bulge across the G tract shows that the flanking positions are occupied much less than the internal positions. As the run becomes longer (C4G5), the internal positions are occupied nearly evenly, resulting in a square rather than a bell-shaped distribution (see Figure 9A). The slight preference for the 5′ end of the G tract is probably due to the differences in flanking sequence. As the temperature increases, the relative occupancies of the flanking positions increase slightly, which could be explained by an increase in the rate of end fraying. The qualitative observations about the large changes in chemical shift made in Figure 7 can be understood in terms of a simple model in which the time-averaged position of the bulge is distributed over a homopolymeric sequence, which in G-bulge duplexes implies fast averaging of hydrogen-bonded and nonbonded states of imino protons.

Imino Proton Relaxation Rates. Next, we wished to see whether this same model could predict the observed exchange rates. For spin–lattice relaxation, the observed rate of exchange should be a weighted average of the exchange rates of each conformer in the fast-exchange limit. We attempted to calculate relaxation rates for the internal imino protons of the migrating G-bulge duplexes by using the exchange rates of the bulged guanosine and flanking base pairs in CG4G3 and weighting them according to the distribution of the bulge calculated from the chemical shift data. For example, the relaxation rates of the internal imino resonances of C2G3 can be written as

$$k_{3,\text{obsd}} = f_A k_{3'} + (f_B + f_C)k$$

$$k_{4,\text{obsd}} = f_A k_b + f_B k_{3'} + f_C k$$

$$k_{5,\text{obsd}} = f_A k_{5'} + f_B k_b + f_C k_{3'}$$

$$k_{6,\text{obsd}} = f_A k + f_B k_{5'} + f_C k_b$$

$$k_{7,\text{obsd}} = (f_A + f_B)k + f_C k_{5'}$$

where $k = 4.469\ s^{-1}$, $k_{5'} = 8.095\ s^{-1}$, and $k_{3'} = 6.263\ s^{-1}$, taken from the exchange rates of G4, G6, and G7 in the A-bulge 9-mer (Woodson & Crothers, 1987), and $k_b = 80\ s^{-1}$, taken from the exchange rate of the extrahelical guanine in CG4G3 at 278 K. (The lifetimes for CG4G3 are listed in Table VIII, and those of CT4G3 in Table IX.)

The experimental and calculated exchange rates are plotted in Figure 9B with the relative populations of the conformers calculated from the chemical shift values shown above in Figure 9A. As in the chemical shifts, the iminos on the edge of the G run are less affected (or have lower exchange rates) than those internal to the G tract. While the general shape of the curves is the same in the calculated and experimental plots, the application of the exchange rate from the isolated bulged G results in exchange rates that are significantly lower than those actually observed in C3G4 and C4G5. This may be due to a difference in the rates of opening of flanking base pairs, accessibility of those base pairs to solvent, or other factors that are not addressed by this simple model. While the experimental exchange rates can be fit more closely by more complicated models, we feel the data do not support an analysis requiring additional parameters.

Free Energy of the Bulge Defect. The thermodynamic parameters obtained from the optical melting transitions enable us to estimate the increase in free energy for a bulged nucleotide. The free energy change for a given shift in T_m due to the addition of a bulged nucleotide to a duplex of given length is obtained from

$$\delta(\Delta G^\circ) = -\Delta S^\circ \delta T_m$$

in which ΔS° is set equal to the change in entropy calculated

for the melting transition of the analogous regular duplex (Table X). In the case of duplexes where the position of the bulge is isomerized, the total change in free energy due to the bulge is smaller than for isolated bulges, but this difference can be accounted for by the degeneracy of the bulge site. The destabilizing free energy of the bulge is now given by

$$\delta(\Delta G^\circ) = \delta(\Delta G^\circ)_{\text{isolated bulge}} - RT \ln \sum_i W_i$$

where W_i is the weighting of each state relative to the most probable state. This leads to values of 2.7 kcal/mol for C2G3, 2.8 kcal/mol for C3G4, and 2.7 kcal/mol for C4G5 for the change in free energy due to an isolated bulge in a location corresponding to the most probable state, which is internal to the G tract. These values are in good agreement with the value of 2.9 kcal/mol calculated for CG4G3, with those obtained previously for the A-bulge and C-bulge 9-mers (3.2 and 2.9 kcal/mol, respectively; Woodson & Crothers, 1987), and with early studies on RNA helices containing bulged nucleotides (Fink & Crothers, 1972).

Conclusions

Clearly, the way in which a given DNA sequence accommodates a bulge defect depends on many factors, such as local base-stacking interactions and backbone flexibility. Homopolymeric sequences are able to have the extra nucleotide in several sites and so have another way in which to lower the destabilization of the bulge defect. Our studies have shown that the extra base occupies all the available positions with some frequency, although at least for guanine, the flanking positions are occupied less often. Bulge migration appears to be on the millisecond time scale or faster, with an apparently low energy of activation.

It is unfortunately not possible to determine the precise mechanism of bulge migration from these experiments. The exchange rates in the C-bulge 9-mer implied protection of flanking base pairs from solvent during isomerization of the bulge site. Clearly, the rate of proton exchange of an unpaired guanine imino is much higher than for an imino proton involved in base pairing. The data from the isolated G-bulge duplex, and the broad distribution of fast relaxation rates in the migrating G-bulge duplexes, indicate increased opening of the base pairs that flank the bulge site. It is difficult to determine whether the broadening of the imino resonances in the G-bulge duplexes is due entirely to the presence of an imino proton on the extrahelical base, or whether there is additional line broadening from slower exchange between the bulge conformers. We cannot be certain that the rate of bulge migration is the same for oligopyrimidine and oligopurine tracts. Data from our laboratory and from Patel and coworkers indicate that extrahelical purines remain stacked in the helix, while pyrimidines are in exchange between stacked and unstacked conformations (Woodson and Crothers, unpublished results; Hare et al., 1986b). Futhermore, the line width of the bulged thymine imino resonance in CT4G3 is twice that of the guanine in CG4G3 at the same temperature (194 and 80 Hz, respectively), implying that it is more accessible to exchange with solvent. It is possible that a bulged cytosine unstacks with its neighbors, even during rapid migration.

We have begun to rationalize the observed tendency of frame shifts to occur in homopolymeric sequences by describing the delocalization of the bulge defect across those sequences and demonstrating that this results in less destabilization of the duplex, as expected. Furthermore, we have demonstrated that longer bulge-containing sequence tracts are stabilized relative to shorter ones and shown that, at least for guanines, the extra base is preferentially located internal to the run. Not only does bulge migration stabilize a defective duplex, but it is rapid enough that enzymes may "see" an average structure that more nearly approximates a normal helix. Further investigation of these sequences and their complexes with various mutagenic drugs is under way. Hopefully, we will be able to better understand the roles of flanking sequence, local DNA structure, and drug affinities in determining the relative rates of frame-shift mutagenesis.

Acknowledgments

Our thanks are due to Franklin Hutchinson for his continuing advice on this problem and to James Nadeau for use of his computer fitting programs.

Registry No. C2G2, 111689-23-7; C2G3, 111662-56-7; C3G3, 105784-72-3; C3G4, 111662-58-9; C4G4, 111662-59-0; C4G5, 111689-25-9; C5G5, 111689-27-1; CG4G3, 111662-61-4; CT4G3, 111662-63-6; guanosine, 118-00-3.

References

Ames, B., Lee, F. D., & Durston, W. E. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 782–786.

Calos, M. P., & Miller, J. H. (1981) *J. Mol. Biol. 153*, 39–66.

Cheung, S., Arndt, K., & Lu, P. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3665–3669.

Crothers, D. M., Cole, P. E., Hilbers, C. W., & Shulman, R. G. (1974) *J. Mol. Biol. 153*, 39–66.

Fink, T. R., & Crothers, D. M. (1972) *J. Mol. Biol. 66*, 1–12.

Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1983) *Biomol. Stereodyn., Proc. Symp., 1981 1*, 115–129.

Hare, D., Shapiro, L., & Patel, D. J. (1986a) *Biochemistry 25*, 7445–7456.

Hare, D., Shapiro, L., & Patel, D. J. (1986b) *Biochemistry 25*, 7456–7464.

Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., & Kearns, D. R. (1986) *Biochemistry 25*, 4840–4849.

Janik, B. (1971) *Physicochemical Characteristics of Oligonucleotides and Polynucleotides*, Plenum, New York.

Kime, M. J., & Moore, P. B. (1983) *Biochemistry 22*, 2615–2622.

Lee, C.-H., & Tinoco, I., Jr. (1978) *Nature (London) 274*, 609–610.

Lerman, L. S. (1963) *Proc. Natl. Acad Sci. U.S.A. 49*, 94–102.

Leroy, J.-L., Broseta, D., & Gueron, M. (1985) *J. Mol. Biol. 184*, 165–178.

Miller, J. H. (1985) *J. Mol. Biol. 182*, 45–64.

Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I. (1983) *Biochemistry 22*, 5557–5563.

Nelson, J. W., & Tinoco, I., Jr. (1985) *Biochemistry 24*, 6416–6421.

Pardi, A., & Tinoco, I., Jr. (1982) *Biochemistry 21*, 4686–4693.

Pardi, A., Martin, F. H., & Tinoco, I., Jr. (1981) *Biochemistry 20*, 3986–3996.

Patel, D., Kozlowski, S. A., Marky, L. A., Broka, C., Rice, J. A., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 445–451.

Quignard, E., Teoule, R., Guy, A., & Fazakerley, G. V. (1985) *Nucleic Acids Res. 13*, 7829–7836.

Roth, K., Kimber, B. J., & Feeney, J. (1980) *J. Magn. Reson. 41*, 302–309.

Skopek, T. R., & Hutchinson, F. (1984) *MGG, Mol. Gen. Genet. 195*, 418–423.

Streisinger, G., & Owen, J. (1985) *Genetics 109*, 633–659.
Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77–84.
Woodson, S. A., & Crothers, D. M. (1987) *Biochemistry 26*, 904–912.
Young, P. R., & Kallenbach, N. R. (1981) *J. Mol. Biol. 145*, 785–813.

Structural and Dynamic Aspects of Binding of a Prototype Lexitropsin to the Decadeoxyribonucleotide d(CGCAATTGCG)$_2$ Deduced from High-Resolution ^{1}H NMR Studies†

Moses Lee,‡ Ding-Kwo Chang,‡ John A. Hartley,‡ Richard T. Pon,§ Krzysztof Krowicki,‡ and J. William Lown*,‡

Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada, and Regional DNA Synthesis Laboratory, University of Calgary, Calgary, Alberta T2N 4N1, Canada

Received May 19, 1987; Revised Manuscript Received September 9, 1987

ABSTRACT: Structural and dynamic properties of the self-complementary decadeoxyribonucleotide d-(CGCAATTGCG)$_2$ and the interaction between a prototype lexitropsin, or information-reading oligopeptide, and the decadeoxyribonucleotide are deduced by using high-resolution ^{1}H NMR techniques. The nonexchangeable and imino proton resonances of d(CGCAATTGCG)$_2$ have been completely assigned by two-dimensional NMR studies. The decadeoxyribonucleotide exists as a right-handed B-DNA. In the ^{1}H NMR spectrum of the 1:1 complex, the selective chemical shifts and removal of degeneracy of AH2(4), AH2(5), T-CH$_3$(6), and T-CH$_3$(7) due to the anisotropy effects of the heterocyclic moieties of the ligand, and with lesser effects at the flanking base sites C(3) and G(8), locate the drug centrally in the decadeoxyribonucleotide. This conclusion is supported by plots of individual chemical shift changes across the decadeoxyribonucleotide. Similarly, imino protons IV and V experience larger shifts and II and III smaller shifts in accord with this conclusion while drug complexation permits the detection of imino proton I. Strong nuclear Overhauser effects (NOEs) between pyrrole H5 and AH2(5), and weaker NOEs to AH1′(5), TH3′(6), and AH2′(5), firmly locate the ligand in the minor groove. Intraligand NOEs between the adjacent heterocyclic moieties indicate that the lexitropsin is subject to propeller twisting about the N6–C9 bond in both the bound and free forms. Nuclear Overhauser effect spectroscopy (NOESY) and correlated spectroscopy (COSY) experiments also indicate that the removal of degeneracy of the C16 methylene protons upon complexation may arise from restricted rotation about the C15–N9, C15–C16, and C16–C17 bonds. Specific hydrogen bonds between amide NH groups on the concave face of the ligand (N4H, N6H, N9H) and adenine N3 or thymine O2 on the floor of the minor groove are in accord with displacement of the hydration shell by the drug. NOE measurements on the decadeoxyribonucleotide in the 1:1 complex confirm it exists as a right-handed helix and belongs to the B family. Exchange NMR effects permit an estimate of a rate of $\approx$44 s^{-1} for the two-site exchange of the lexitropsin between two equivalent sites on the decamer with $\Delta G^{\ddagger} \approx 70 \pm 5$ kJ mol^{-1} at 294 K. Alternative mechanisms for this exchange process are considered.

The sequence-specific molecular recognition of DNA by proteins is central to the regulation of many biological processes (Caruthers, 1980; Frederick et al., 1984; Gurskii et al., 1977; Kim et al., 1974; Takeda et al., 1983). The oligopeptide antibiotics netropsin and distamycin can serve as models of sequence-specific and groove-selective DNA-binding molecules (Hahn, 1975; Kolchinskii et al., 1975; Kopka et al., 1985; Lown et al., 1986; Patel, 1982; Wartell et al., 1974; Zimmer, 1975; Zimmer et al., 1986). Netropsin and distamycin bind within the minor groove of DNA (Zimmer, 1975) and demand binding sites consisting of (A·T)$_4$ and (A·T)$_5$, respectively. Physical studies including X-ray analysis of a complex of netropsin with the self-complementary dodecamer d-(CGCGAATTCGCG)$_2$ (Kopka et al., 1985), ^{1}H NMR investigations (Patel, 1982; Gupta et al., 1984), and circular dichroism (CD) studies (Zimmer, 1975) have provided some structural details on the nature of the interactions between drug and receptor that contribute to their marked specificity. An analysis of these structural requirements for the molecular recognition suggested that replacement of one or more pyrrole groups by hydrogen bond accepting heterocycles such as imidazole should alter this (A·T)$_n$ specificity in a predictable way. This prediction was recently confirmed experimentally, indicating that the imidazole moieties permit the recognition of GC sites by invoking new hydrogen bonds between G–C(2)-NH$_2$ and the imidazole N3 (Lown et al., 1986; Kissinger et al., 1987).

It was clear however that a number of other factors are important determinants of recognition. For example, electrostatic interactions contribute significantly since the marked negative potential wells in the minor groove of (A·T)$_n$ sequences confer a bias for the binding of doubly positively

†This investigation was supported by grants (to J.W.L.) from the National Cancer Institute of Canada and the Biotechnology Strategic Grants Programme of the National Sciences and Engineering Research Council of Canada.
‡University of Alberta.
§University of Calgary.

0006-2960/88/0427-0445$01.50/0