Solution Structure of a Trinucleotide A-T-A Bulge Loop within a DNA Duplex†

Mark A. Rosen, Lawrence Shapiro, and Dinshaw J. Patel*

Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032

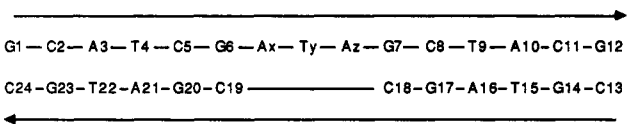
Received December 16, 1991

ABSTRACT: We have synthesized an oligodeoxynucleotide duplex, d(G-C-A-T-C-G-A-T-A-G-C-T-A-C-G)·d(C-G-T-A-G-C-C-G-A-T-C-G), with a three-base bulge loop (A-T-A) at a central site in the first strand. Nuclear Overhauser experiments (NOESY) in H_2O indicate that the GC base pairs flanking the bulge loop are intact between 0 and 25 °C. Nuclear Overhauser effects in both H_2O and D_2O indicate that all bases within the bulge loop are stacked into the helix. These unpaired bases retain an anti conformation about their glycosidic bonds as they stack within the duplex. The absence of normal sequential connectivities between the two cytosine residues flanking the bulge site on the opposite strand indicates a disruption in the geometry of this base step upon insertion of the bulged bases into the helix. This conformational perturbation is more akin to a shearing apart of the bases, which laterally separates the two halves of the molecule, rather than the "wedge" model often invoked for single-base bulges. Using molecular dynamics calculations, with both NOE-derived proton–proton distances and relaxation matrix-calculated NOESY cross peak volumes as restraints, we have determined the solution structure of an A-T-A bulge loop within a DNA duplex. The bulged bases are stacked among themselves and with the guanine bases on either side of the loop. All three of the bulged bases are displaced by 2–3 Å into the major groove, increasing the solvent accessibility of these residues. The ATA-bulge duplex is significantly kinked at the site of the lesion, in agreement with previously reported electron microscopy and gel retardation studies on bulge-containing duplexes [Hsieh, C.-H., & Griffith, J. D. (1989) *Proc. Natl. Acad. Sci. U.S.A 86*, 4833–4837; Bhattacharyya, A., & Lilley, D. M. J. (1989) *Nucleic Acids Res. 17*, 6821–6840]. Bending occurs in a direction away from the bulge-containing strand, and we find a significant twist difference of 84° between the two base pairs flanking the bulge loop site. This value represents 58% of the twist difference for base pairs four steps apart in B-DNA. These results suggest a structural mechanism for the bending of DNA induced by unpaired bases, as well as accounting for the effect bulge loops may have on the secondary and tertiary structures of nucleic acids.

In the preceding paper of this issue (Rosen et al., 1992), we have presented solution NMR data for a series of DNA dodecamer duplexes containing a variable number of unpaired adenine bases at a central site. We have shown that for either an A, A-A, or A-A-A bulge loop, all of the bulged adenine residues adopt intrahelical conformations, stacking within the duplex. Furthermore, the insertion of two or three residues into the helix causes severe disruptions in the stacking interaction between the bases flanking the bulge site on the opposite strand. This base step exhibits an unusual geometry in which some of the sequential proton–proton distance connectivities are weakened while others are strengthened. We also find that the backbone conformation at the 3′ end of multiple-base bulge loops is perturbed, as is evident by the downfield shift of a lone phosphorus resonance at the 3′ end of the bulge site.

The chemical shifts of certain proton and phosphorus resonances within the A_n-bulge duplex series exhibit a roughly linear dependence on the number of bases within the bulge loop. This pattern is reminiscent of the results of gel mobility studies on duplex DNA containing bulge loops (Hsieh & Griffith, 1989; Bhattacharyya & Lilley, 1989), in which an increase in the number of bulged bases led to a parallel increase in the amount of anomalous gel mobility retardation. We were therefore interested in obtaining a more quantitative description of the geometry of a bulge loop within DNA. Unfortunately, the proton resonances within the purine-rich bulge loops studied were often poorly resolved in the NOESY spectra. Such overlap severely hampered our efforts to measure the proton–proton distances needed for structural determination via NOE-restrained molecular dynamics simulations. We therefore looked for alternative loop sequences which would exhibit both a single conformation in solution and ample dispersion of the proton resonances for the residues located within or near the bulge loop site. In this paper, we report on the two-dimensional proton NMR spectroscopy and the structural determination of an A-T-A bulge loop within a dodecamer DNA duplex. We refer to this molecule from here on as the ATA-bulge duplex and present the sequence and numbering scheme in Chart I.

Chart I



Experimental Procedures

Synthesis, Purification, and Preparation of NMR Samples. DNA oligomers were prepared as described in the preceding paper (Rosen et al., 1992) using standard phosphoramidite chemistry on solid supports. Oligomers were purified by reversed-phase HPLC. Purity was judged both by analytical HPLC and the appearance of the one-dimensional NMR spectra. Stoichiometric complexes were obtained via titration of the two individual strands at high temperature, followed by one-dimensional NMR. DNA-containing solutions were

† This research was supported by NIH Grant GM34504 to D.P. M.R. was supported by NIH MSTP Training Grant 5-T32-GM07376. The NMR spectrometers were purchased from funds donated by the Robert Woods Johnson Trust toward setting up an NMR center in the Basic Medical Sciences at Columbia University.

* Author to whom correspondence should be addressed.

0006-2960/92/0431-4015$03.00/0

allowed to cool slowly in buffered solution to facilitate duplex formation. Final buffer conditions were 10 mM sodium phosphate, pH 6.5, containing 100 mM NaCl and 0.2 mM EDTA. Observation of sharp imino proton resonances between 12.0 and 14.0 ppm indicated that the two strands had properly annealed.

Deuteration of Purine Residues at the C8 Position. The DNA oligomer was dissolved in 10 mM sodium phosphate with 0.1 mM EDTA. The pH of the sample was adjusted to 9.0 with NaOH and was then lyophilized and redissolved in 400 μL of D_2O. The uncorrected pH reading was readjusted to 9.0 with NaOD, after which the sample was incubated at 75 °C for 96 h. The reaction was followed periodically by one-dimensional NMR in D_2O. At the end of 96 h, deuteration was judged to be >98% complete for guanine residues and >95% complete for adenine residues.

NMR Spectroscopy. One- and two-dimensional proton NMR spectra were recorded on Bruker AM-400 and AM-500 spectrometers. Spectra in H_2O were acquired with a jump-and-return pulse sequence (Plateau & Gueron, 1982) to suppress the solvent signal. Those in D_2O were recorded with a low-power preirradiation of the residual HDO signal during the recovery period. All spectra were referenced to an external 3-trimethylsilyl(*2,2,3,3,-d*$_4$)propionate (TSP) standard. Phase-sensitive NOESY experiments were acquired in the States mode (States et al., 1982). Two-dimensional NOESY and COSY spectra were acquired as described previously (Rosen et al., 1992). All two-dimensional data sets were processed on a Vax 11/780 computer using the FTNMR software package (D. Hare, Hare Research, Inc.).

NOE Quantitation. NOESY data sets for NOE cross peak volume measurements were collected in succession with mixing times of 50, 90, 140, 200, and 300 ms. The two-dimensional data sets were processed as described above, except for the substitution of 90° shifted sine bell window functions in order not to deemphasize cross peaks from protons with shorter relaxation times. The initial NOE buildup rates were calculated by plotting the cross peak volume versus mixing time and measuring the slope of the linear portion of the graph. Distances were calculated using the isolated two-spin approximation, with the C(H5)–C(H6) distance of 2.5 Å used as a reference. For weak or severely overlapped cross peaks, only minimum distance estimates were used.

Molecular Dynamics Calculations. Restrained molecular dynamics simulations were performed using the program XPLOR (A. Brunger, Yale University) on a Convex C2 computer. Molecular structures were visualized with the program Insight II on a Silicon Graphics IRIS workstation. Molecular dynamics simulations were done in vacuo using an all-atom potential force field derived from CHARMM (Nilsson & Karplus, 1986), with reduced phosphate charges and a distance-dependent dielectric, $\epsilon = r$. Explicit hydrogen-bonding terms were included in order to maintain Watson–Crick base pairing. Distance constraints from experimental data were input using a biharmonic potential with flat-well width of 0.4 Å (0.8 Å for cross peaks that were resolved only after partial deuteration of the duplex). Key distances involving selected imino protons were estimated and included with a 2.0-Å well width. Watson–Crick hydrogen bonds were reinforced by NOE restraints as well. The dihedral angles of the residues within the four terminal base pairs on either end (16 bases in all) were restrained to the following equilbrium values: $\alpha = -70°$, $\beta = 180°$, $\epsilon = 160°$, and $\zeta = 100°$, with a flat-well width of 60° in the potential function. Energy-minimized starting models were subjected to 6 ps of restrained molecular dynamics (1.0-fs time step) at 400 K, during which time the NOE energy scale factor was slowly increased from 1.0 to 8.0 kcal/mol. The molecules then underwent an additional 8 ps of restrained-molecular dynamics at 300 K with full-scale NOE restraints. The SHAKE algorithm (Ryckaert et al., 1977) was used to maintain proper hydrogen-bond distances throughout the simulation. The coordinates of the last 4 ps were averaged and subjected to a final 2000 steps of conjugate gradient energy minimization.

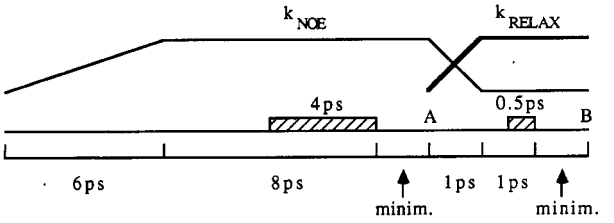


FIGURE 1: Molecular dynamics protocol. The time line shows the evolution of molecular dynamics simulations. The scaling factors, k_{NOE} and k_{RELAX}, refer to the force constants used with the harmonic potentials defined for NOE-derived proton–proton distances and for relaxation-matrix calculated NOESY cross peak volumes, respectively. These constants were gradually scaled between their minimum and maximum values (1.0–8.0 kcal/mol for k_{NOE} and 10–120 kcal/mol for k_{RELAX}). Cross-hatched areas represent periods where averaging of the molecular coordinates occurred before the subsequent step. "Minim." refers to 2000 steps of conjugate gradient minimization. Most simulations were carried through to point A only.

Relaxation Matrix Refinement. For one of the converged, distance-refined structures, we ran an additional 2 ps of molecular dynamics during which the NOE (proton–proton distance) energy term was reduced while a RELAX energy function based on the full relaxation matrix calculation of NOESY cross peak volumes (Yip & Case, 1989; Nilges et al., 1991) was introduced. Volume inputs for relaxation matrix refinement were used with a uniform 10% error range. Only volumes greater than two times the random noise in a given spectral region were included. The coordinates of the last 0.5 ps of this final 2 ps of molecular dynamics were averaged and minimized. *R* factors were calculated as the weighted average of the absolute value of the difference between observed and calculated intensities. Figure 1 shows a schematic of the molecular dynamics protocol used.

Helical Analysis. Global helical parameters for all structures generated were determined using the CURVES algorithm (Lavery & Sklenar, 1988, 1989). Helical parameters were calculated for the individual duplex stems (six base pairs each) as well as for the entire molecule.

RESULTS

Exchangeable Protons. The downfield portion of the one-dimensional proton NMR spectrum of the ATA-bulge duplex in H_2O buffer at 10 °C is shown in Figure 2A. The imino proton assignments in the figure are discussed below. Hydrogen-bonded imino protons resonate between 12 and 14 ppm, as opposed to imino protons of unpaired bases which resonate farther upfield and have broader line shapes due to frequent exchange with solvent. The lone resonance at 10.4 ppm is characteristic of imino protons of unpaired thymines and is therefore assigned to the Ty residue within the bulge loop. Due to the overlap of the remaining imino resonances in the one-dimensional spectrum, NOEs between imino protons of neighboring base pairs were unhelpful toward making sequential assignments. We therefore assigned imino protons on the basis of NOEs to known nonexchangeable protons. Thymine imino protons in AT base pairs were assigned on the basis of the strong NOE to the H2 proton of the partner

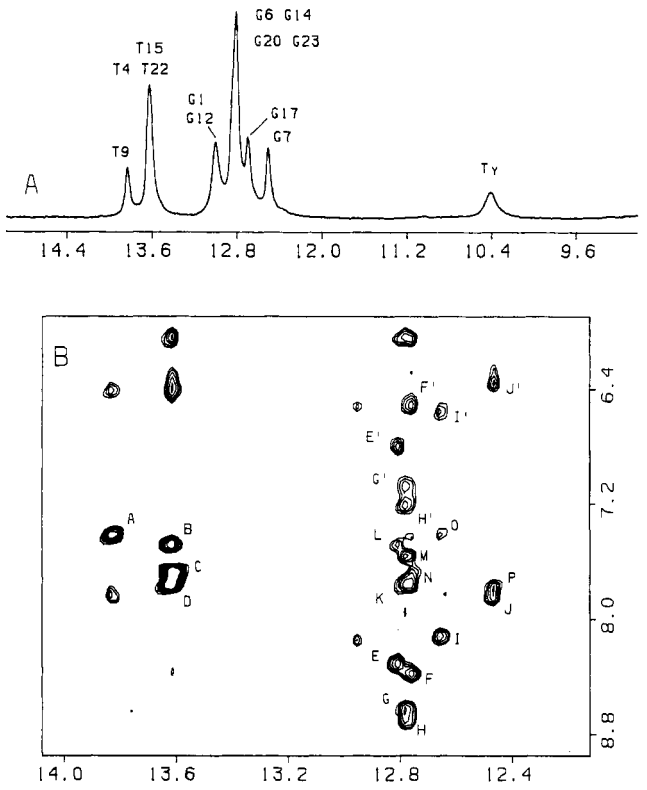


FIGURE 2: (A) Expanded downfield region of the proton NMR spectrum (9.0–15.0 ppm) of the ATA-bulge duplex in 0.1 M NaCl, 10 mM sodium phosphate, and 0.2 mM EDTA, pH 6.5, in H_2O at 10 °C. The imino proton assignments shown are discussed in the text. (B) Expanded contour plot of a NOESY experiment (150-ms mixing time) of the ATA-bulge duplex in H_2O buffer at 0 °C. Assignments for cross peaks labeled are as follows: (A) T9(imino)–A16(H2); (B) T15(imino)–A10(H2); (C) T22(imino)–A3(H2); (D) T4(imino)–A21(H2); (E,E′) G14(imino)–C11(amino, b/e); (F,F′) G23(imino)–C2(amino, b/e); (G,G′) G20(imino)–C5(amino, b/e); (H,H′) G6(imino)–C19(amino, b/e); (I,I′) G17(imino)–C8(amino, b/e); (J,J′) G7(imino)–C18(amino, b/e); (K) G20(imino)–A21(H2); (L) G14-(imino)–A10(H2); (M) G6(imino)–Ax(H2); (N) G23(imino)–A3-(H2); (O) G17(imino)–A16(H2); (P) G7(imino)–Az(H2). The symbols "b" and "e" refer to the hydrogen-bonded and exposed cytosine amino protons, respectively, involved in Watson–Crick base pairing. Cross peaks H,H′ and J,J′ demonstrate intact base pairing for residues G6·C19 and G7·C18, respectively. Cross peaks M and P localize the two bulge adenines to intrahelical positions stacked with their respective flanking GC base pair.

adenine. These latter protons were assigned on the basis of characteristic intra- and interstrand NOEs to nearby H1′ protons seen in the NOESY spectra in D_2O buffer. Guanine imino protons in GC base pairs were assigned on the basis of observed NOEs to the hydrogen-bonded and exposed amino protons of the cytosine residue with which it is paired. These two geminal protons show a strong NOE to each other and also to the nonexchangeable H5 proton of the same residue. We first assigned the amino protons of cytosine on the basis of NOEs to the C(H5) protons. The amino pairs from all eight cytosines could be assigned in this manner. Guanine iminos were then assigned on the basis of NOEs to the cytosine amino pair.

An expanded region of the 150-ms mixing time NOESY spectrum of the ATA-bulge duplex in H_2O buffer at 0 °C is shown in Figure 2B. Imino proton chemical shifts are plotted along the horizontal axis, while those of amino and base protons are along the vertical axis. On the basis of the strong NOEs seen to adenine H2 protons (peaks A–D, Figure 2B), the four resonances between 13.5 and 14.0 ppm are assigned to the imino protons of the four Watson–Crick thymines: T4, T9, T15, and T22. Specific assignments of these cross peaks

Table I: Exchangeable and A(H2) Proton Chemical Shifts of the ATA-Bulge Duplex in H_2O Buffer, pH 6.5, at 0 °C

	chemical shifts (ppm)				
	H3	H1	NH_2b^a	NH_2e^b	H2
G1·C24			8.17	6.51	
C2·G23		12.78	8.39	6.53	
A3·T22	13.63				7.70
T4·A21	13.64				7.78
C5·G20		12.80	8.64	7.09	
G6·C19		12.79	8.72	7.23	
Ax					7.59
Ty	10.44				
Az					7.80
G7·C18		12.48	7.86	6.37	
C8·G17		12.68	8.14	6.58	
T9·A16	13.85				7.43
A10·T15	13.63				7.50
C11·G14		12.83	8.33	6.82	
G12·C13			8.14	7.03	

[a] Hydrogen-bonded cytosine amino protons. [b] Exposed cytosine amino protons.

are listed in the caption to the figure. The remaining resonances, located between 12.4 and 13.0 ppm, belong to guanine imino protons. For each of the six nonterminal guanine residues, a pair of cross peaks is observed from the imino proton to the two amino protons of the corresponding cytosine on the partner strand. These pairs are labeled E,E′–J,J′ in Figure 2B. It should be noted that both G6 and G7, which flank the bulge site in the ATA-bulge duplex, show observable NOEs to the C19 (peaks H,H′) and C18 (peaks J,J′) cytosine amino pairs, respectively. The downfield chemical shift of these imino protons, as well as the presence of NOEs to cytosines on the opposite strand, confirm that these two guanines remain base paired with their partner cytosines even though they flank a three-base bulge loop.

Two additional NOE cross peaks, one each from the G6 and G7 imino protons, are observed in Figure 2B. These are distance connectivities from each guanine imino proton flanking the bulge site to the H2 proton of the neighboring unpaired adenine at the bulge site. Specifically, we see an NOE between G6(imino) and Ax(H2) (peak M, Figure 2B) and between G7(imino) and Az(H2) (peak P, Figure 2B). Identification of NOEs from the H2 protons of each unpaired adenine to its neighboring guanine imino proton allows us to place both of these bulged bases within the duplex in a "stacked-in" conformation. Note that the G6(imino)–Ax(H2) NOE in Figure 2B (peak M) is better resolved for this ATA-bulge duplex than had been the case for a duplex with an A-A-A bulge loop (Rosen et al., 1992). Other than this qualitative difference, the NOESY spectrum of the ATA-bulge duplex in H_2O buffer resembles quite well that of the A_3-bulge duplex. No NOEs were observed to or from the unpaired Ty imino proton in the ATA-bulge duplex under these conditions. However, one-dimensional NOE experiments revealed the presence of weak NOEs from this Ty imino proton to the H2 protons of each of the bulged adenines, Ax and Az, in the ATA-bulge duplex. These data suggest that Ty also stacks into the helix under these conditions. The chemical shifts of the exchangeable and A(H2) protons assigned for the ATA-bulge duplex at 0 °C are listed in Table I.

Nonexchangeable Protons. Sequential assignments of nonexchangeable protons can be made in D_2O by observation of through-space connectivities between base (H8 for purines, H6 for pyrimidines) and sugar protons. In a right-handed helix, a base proton will show an NOE to its own sugar H1′ proton as well to that from the preceding nucleotide in the 5′ direction (Hare et al., 1983; Wüthrich, 1986). Due to the lack

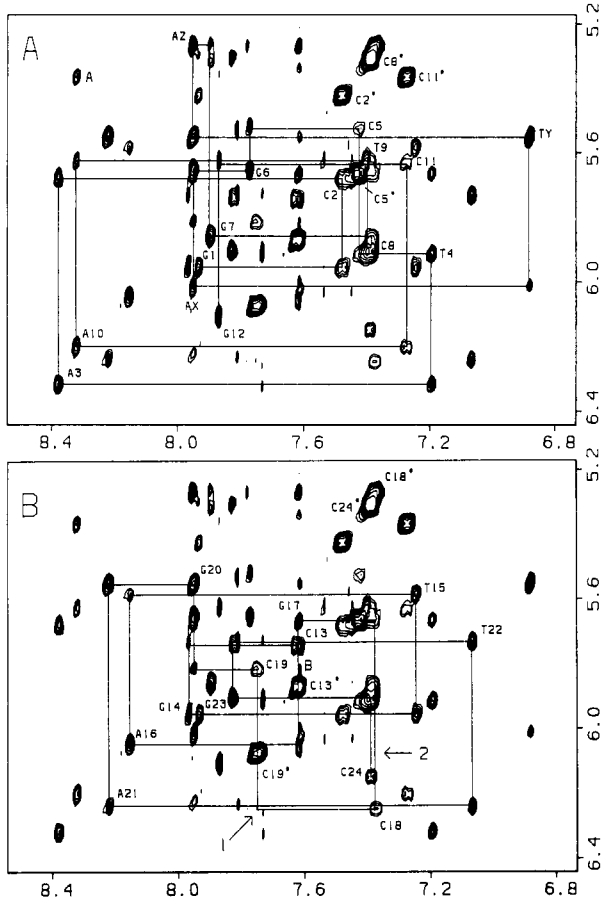


FIGURE 3: Expanded contour plots (base to H1′ region) of the NOESY (250-ms mixing time) spectrum of the ATA-bulge duplex in 0.1 M NaCl, 10 mM phosphate, and D_2O, pH 6.8, at 25 °C. Sequential assignments for individual strands are traced separately in panels A and B. Intraresidue base to H1′ proton cross peaks are labeled. Asterisks indicate cytidine H6–H5 cross peaks. Arrow 1 in panel B denotes the position of the absent C18(H1′)–C19(H6) cross peak. Arrow 2 in panel B represents the position of a weak C18(H6)–C19(H5) cross peak (apparent at lower contour levels only). Cross peak A in panel A represents the A10(H8)–C11(H5) cross peak for comparison. Cross peak B in panel B represents an interstrand NOE from Ax(H2) to C19(H1′).

of absolute resolution of proton resonances in the ATA-bulge duplex, and the resulting overlap of NOE cross peaks, we analyzed the entire NOESY spectrum, as well as the COSY spectrum, to complete our sequential assignments. Correlated cross peaks in the COSY spectrum enabled us to identify the individual spin systems within each sugar moiety. Through-space connectivities in the NOESY spectrum were then analyzed to establish sequential assignments. These assignments were extended to the H3′ and H4′ sugar protons by identifying COSY cross peaks in conjunction with H1′–H4′ and base to H3′ and H4′ NOESY cross peaks. The latter are commonly seen in experiments with long mixing times, despite the large distances involved, due to the efficiency of spin diffusion among sugar protons. No attempt was made to assign the H5′/H5″ proton pairs due to the poor resolution in this region of the spectrum.

The duplex exhibits a B-like conformation, as judged by the large intensity of the intraresidue base–H2′ cross peaks as well as the presence of strong H1′–H2′ couplings in the COSY spectrum. Expanded contour plots of the base to H1′ region of the 250-ms NOESY experiment on the ATA-bulge duplex in D_2O buffer at 25 °C are shown in duplicate in Figure 3. For clarity, sequential assignments for each strand are indicated separately in panels A and B. From the diagram, it is

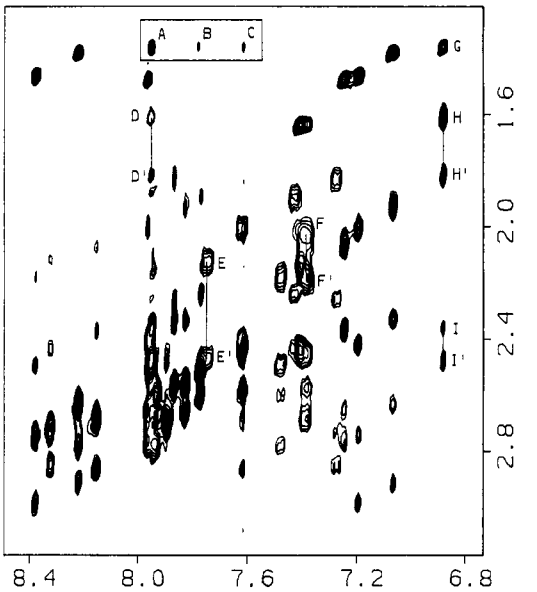


FIGURE 4: Expanded NOESY contour plot of the base to H2′,H2″ and methyl region of the 250-ms mixing time NOESY spectrum of the ATA-bulge duplex in D_2O buffer. Selected cross peaks detailing sequential NOEs within the bulge loop region are labeled. Assignments are as follows: (A) Ax(H8)–Ty(Me); (B) Az(H2)–Ty(Me); (C) Ax(H2)–Ty(Me); (D,D′) Ty(H2′,2″)–Az(H8); (E,E′) C19-(H2′,2″)–C19(H6); (F,F′) C18(H2′,2″)–C18(H6) (G) Ty(CH_3)–Ty(H6); (H,H′) Ty(H2′,2″)–Ty(H6); (I,I′) Ax(H2′,2″)–Ty(H6).

evident that the sequential distance connectivities throughout the three-base bulge remain intact. That is, we can trace NOE cross peaks through the sequence G6(H1′)-Ax(H8)-Ax-(H1′)-Ty(H6)-Ty(H1′)-Az(H8)-Az(H1′)-G7(H8) without interruption, as is shown in Figure 3A. On the opposite strand (Figure 3B), there is the conspicuous absence of a cross peak between the H1′ of C18 and the H6 of C19 (arrow 1, Figure 3B). This occurs at the dinucleotide step flanking the bulge site on the opposite strand. Additionally, we see a very weak cross peak (not apparent at this contour level) between C18-(H6) and C19(H5) (arrow 2, Figure 3B). For right-handed helices, one generally observes a relatively strong NOE between a base proton and the H5 (or CH_3) proton(s) of the subsequent base in the 3′ direction. The absence or weakening of NOE connectivities between C18 and C19 indicates an increase in the separation between these two bases in the ATA-bulge duplex. A special cross-strand connectivity between the H2 proton of Ax and the H1′ proton of C19 is also seen here (peak B, Figure 3B). This NOE confirms the intrahelical location of Ax, as was deduced from the NOESY data in H_2O buffer.

We also analyzed a short mixing time (50 ms) NOESY experiment on the ATA-bulge duplex in D_2O buffer. When compared with the intensity of the C(H6)–C(H5) cross peaks, all of the base to H1′ NOEs are quite weak. This indicated that all of the bases within the duplex, including those within the bulge loop, retain an anti configuration about the glycosidic bond (Patel et al., 1982b). In the X-ray crystal structure of a single unpaired adenine within a DNA tridecamer (Joshua-Tor et al., 1992), one of the two symmetrically related bulged adenines residues adopts the syn conformation. No such phenomenon is observed in the ATA-bulge duplex in solution.

Figure 4 shows the base to H2′,H2″, and methyl region of the same NOESY experiment on the ATA-bulge duplex. Sequential base to sugar connectivities for the Ax–Ty step (peaks I and I′, Figure 4) and the Ty–Az step (peaks D and D′) are indicated. Further assignments of connectivities in-

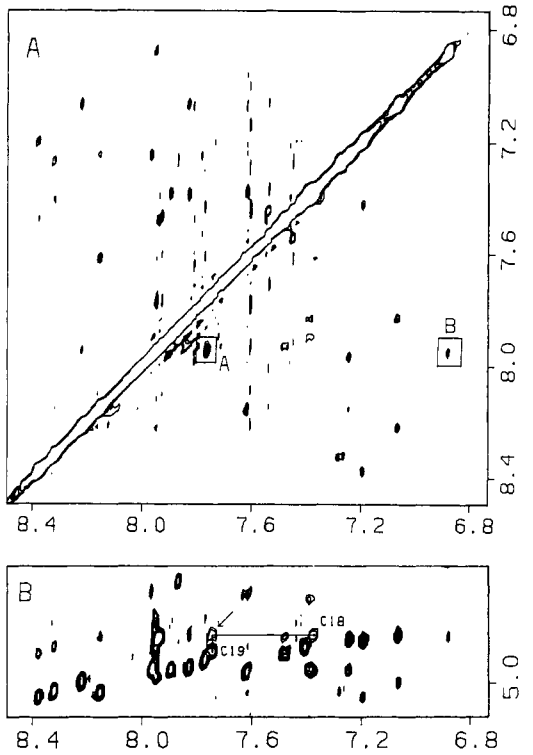


FIGURE 5: (A) Expanded contour plot of the base diagonal region of the 250-ms mixing time NOESY of the ATA-bulge duplex in D_2O buffer. The two labeled cross peaks establish distance connectivities between base protons within the bulge-loop region. Peak A represents an NOE between G6(H8) and Ax(H8). Peak B represents an NOE between Ax(H8) and Ty(H6). (B) Expanded contour plot of the base to H3′ region of the same experiment. The intraresidue cross peaks for C18 and C19 are labeled. The arrow points to an intact sequential connectivity for this base step via the C18(H3′) proton.

volving H2′ and H2″ protons within the bulge site were hampered by the extreme density of cross peaks in this area. On the opposite strand, the intraresidue NOEs from base to H2′ and H2″ for C18 (peaks F and F′) and C19 (peaks E and E′) are indicated. As was the case in the base to H1′ region, there is still a disruption of the sequential connectivities between these two bases.

A few additional NOEs of interest are seen in this region. They are labeled A–C within the box in the upper part of Figure 4. Peak A represents an NOE between the H8 proton of Ax and the methyl protons of Ty. Peaks B and C represent NOEs from the H2 protons of Az and Ax, respectively, to the same Ty methyl protons. Together, these NOEs indicate the presence of base stacking interactions within the bulge-loop segment of the ATA-bulge duplex. This information confirms the intrahelical nature of all three bases within the bulge loop, as was deduced from the data in H_2O buffer.

Additional evidence of base stacking within the bulge loop region is seen in the base diagonal region of the NOESY spectrum of the ATA-bulge duplex in D_2O buffer, shown in Figure 5A. Two cross peaks demonstrating the stacking of G6 on Ax (peak A, Figure 5A) and of Ax on Ty (peak B, Figure 5A) are labeled in this diagram. Other NOEs of this type in the ATA-bulge duplex are obscured by overlapping cross peaks or by proximity to the diagonal. Figure 5B shows another region of the NOESY experiment on the ATA-bulge duplex. In this panel, NOE connectivities between base and H3′ protons are shown. The diagram traces a sequential connectivity between C18(H3′) and C19(H6). The presence of this cross peak is unusual, given the absence of any sequential connectivities between C18 and C19 involving either

Table II: Nonexchangeable Proton Chemical Shifts of the ATA-Bulge Duplex in D_2O Buffer at 25 °C

	chemical shifts (ppm)							
	H8/H6	H5/CH_3	H2	H1′	H2′	H2″	H3′	H4′
G1	7.93			5.96	2.61	2.78	4.84	4.23
C2	7.48	5.43		5.68	2.18	2.50	4.90	4.22
A3	8.38		7.73	6.32	2.75	2.99	5.06	4.45
T4	7.20	1.48		5.92	2.02	2.43	4.86	4.17
C5	7.43	5.67		5.53	1.90	2.25	4.76	4.07
G6	7.77			5.66	2.49	2.60	4.93	4.28
Ax	7.95		7.62	6.02	2.38	2.48	4.85	4.23
Ty	6.89	1.37		5.58	1.62	1.83	4.56	3.68
Az	7.96		7.78	5.27	2.48	2.48	4.76	4.18
G7	7.90			5.87	2.71	2.71	4.96	4.36
C8	7.39	5.32		5.91	2.04	2.47	4.71	4.23
T9	7.41	1.65		5.63	2.14	2.45	4.88	4.17
A10	8.32		7.54	6.21	2.73	2.86	5.04	4.43
C11	7.28	5.37		5.64	1.84	2.26	4.76	4.16
G12	7.87			6.11	2.57	2.35	4.64	4.15
C13	7.62	5.87		5.74	2.02	2.41	4.67	4.06
G14	7.97			5.95	2.66	2.77	4.96	4.34
T15	7.25	1.48		5.59	2.08	2.38	4.85	4.16
A16	8.16		7.45	6.05	2.70	2.87	5.04	4.39
G17	7.62			5.67	2.44	2.59	4.97	4.34
C18	7.38	5.28		6.25	2.06	2.22	4.84	4.28
C19	7.75	6.08		5.82	2.14	2.48	4.92	4.22
G20	7.96			5.55	2.71	2.79	4.99	4.30
A21	8.22		7.81	6.25	2.63	2.91	5.00	4.44
T22	7.07	1.39		5.72	1.94	2.34	4.84	4.12
G23	7.83			5.91	2.57	2.67	4.95	4.34
C24	7.39	5.30		6.15	2.18	2.18	4.47	4.04

the H1′ or H2′,2″ protons. Such a result, similar to that seen for the A_2-bulge and A_3-bulge duplexes in the preceding paper, implies an unusual conformation for this base step within the ATA-bulge duplex. The presence of a C18–C19 sequential connectivity via the C18(H3′) proton indicates that the separation of these two bases, induced by the insertion of three unpaired bases on the opposite strand, is more complex than simply a wedging apart of these two cytosine residues.

To aid in the assignment of cross peaks within crowded regions of the NOESY spectrum of the ATA-bulge duplex in D_2O buffer, a NOESY experiment was performed with an ATA-bulge duplex sample in which the purine H8 protons of the partner (bulgeless) strand had been exchanged with deuterium. Figures S1 and S2 (Supplementary Material) show expansions of the regions shown in Figures 3 and 4 for the NOESY experiment run with either fully protonated (panel A) or hemideuterated (panel B) sample. The resulting spectral simplification allowed us to confirm assignments of distance connectivities involving key protons within the bulge loop region of the ATA-bulge duplex. Comparison of cross peak volumes from NOESY experiments using either fully protonated or hemideuterated samples enabled us to apportion NOE intensity arising from cross peaks that were overlapped in the original spectrum.

The chemical shifts of all nonexchangeable protons, except for the H5′/H5″ pairs, in the ATA-bulge duplex are listed in Table II. Two observations based on chemical shifts should be noted here. We observe that the resonances of base protons (H8 and H6) from residues in the bulge loop are found approximately 0.2–0.4 ppm upfield from those of similar protons on residues within the double-helical stems. Specifically, the H8 protons of both Ax and Az resonate at 7.95 ppm (compared with 8.1–8.4 ppm for the four Watson–Crick paired adenines) while the H6 proton of Ty is found at 6.89 ppm (versus 7.1–7.3 ppm for Watson–Crick paired thymines). Also, the base protons of residue C19 and the H1′ proton of C18 are all located significantly downfield from those of other cytosine residues. The C19 base protons resonate to lower field

than even those of the 5′-terminal C13 base. This chemical shift information is consistent with the positioning of bulged residues within the helix. Looped-out bases would be deshielded relative to fully stacked bases, and their protons would be expected to resonate farther downfield. In addition, loss of the stacking interaction between C18 and C19 might account for the downfield position of protons from these two residues.

Comparison with Control Duplex. NOESY and COSY experiments in both D_2O and H_2O buffer on a DNA duplex of the same sequence, but without any bulged bases, have been reported in the preceding paper. The control duplex also exhibits the general properties of B-DNA in solution. All bases participate in standard Watson–Crick pairing between 0 and 25 °C. All expected NOE cross peaks are observed, and there are no unusual NOEs that would indicate a nonstandard DNA conformation.

The chemical shift differences between equivalent protons within the ATA-bulge duplex and the control duplex are listed in Tables S-I and S-II (Supplementary Material), for the exchangeable and nonexchangeable protons, respectively. Little change in chemical shift is seen for protons located more than two residues away from the bulge loop. Chemical shift changes versus the control are greatest for the base and amino protons of C19. As stated before, we believe that the relative chemical shift values reflect proximity of nearby aromatic rings and the presence (or absence) of base stacking within the bulge loop region. The specific deshielding of C19 is presumably due to wedging apart of the C18–C19 base step upon insertion of three unpaired bases into the helix. This feature is examined further under Discussion.

Generation of Starting Models. Two different starting models of the bulge duplex, referred to as "high-twist" and "low-twist", were generated. For each, we began with a bulgeless dodecamer duplex of the appropriate sequence in standard B-form geometry. The backbone between the sixth and seventh base pairs was broken, and the two halves of the molecule were separated by 10.2 Å (three base pair steps) along the helical axis. Next, a right-handed, stacked trinucleotide fragment, A-T-A, was manually inserted into the central gap. For the high-twist model one helical segment was rotated by three base pair turns (108°) about the helical axis. This rotation created a contiguous sugar–phosphate backbone on the bulge-containing strand, with only a single break (between C18 and C19) on the partner strand. No such rotation was made in the low-twist structure, and so there existed three discontinuities in the backbone of this model: at the G6-Ax, the Az-G7, and the C18–C19 steps. In each case, the backbone was joined by conjugate gradient energy minimization using XPLOR, during which a gentle attractive force was maintained between the phosphorus and oxygen atoms to be ligated. When the distance between these atoms reached 2.5 Å, bonds were created and the minimization was continued for another 2000 steps. Creation of backbone bonds spanning larger discontinuities led to distorted structures. In the two minimized starting models, the GC base pairs flanking the bulge site are buckled, and the backbone connecting C18 and C19 is quite extended. Otherwise, the structures are relatively undistorted, with no significant kinking in any direction. The two models differed most significantly in the twist angle between the two helical segments, 136° and 39°, for the high-twist and low-twist structures, respectively.

Generation of Distance Restraints. Distance restraints between nonexchangeable proton pairs were generated based on observed NOE buildup rates as described under Experimental Procedures. Strong NOEs were converted into absolute distance ranges, while weak or overlapped NOEs were converted into minimum distance estimates only. Selected key restraints involving exchangeable protons within the bulge region were also included with looser restraints. A total of 508 distance restraints were used in all. Figure S3 (Supplementary Material) shows representative NOE buildup curves for proton pairs spanning either the Ax–Ty or the A21–T22 base steps. The second column of Table III lists some of the key distance restraints we obtained for protons located within the bulge loop region.

Table III: Distances between Various Proton Pairs within the Bulge Loop Region of Refined Structures

		distance-refined structure	
proton pair	bounds (Å)	low-twist starting model	high-twist starting model
bulge strand			
Ax(H1′)–Ty(H6)	4.4–4.8	4.8	4.7
Ax(H2′)–Ty(H6)	3.5–3.9	3.9	4.0
Ax(H2″)–Ty(H6)	3.1–3.5	2.8	2.7
Ax(H3′)–Ty(H6)	4.7–5.1	5.0	5.0
Ax(H2)–Ty(CH_3)	>5.0	6.4	7.4
Ty(H1′)–Ax(H8)	>4.5	6.1	6.2
Ty(H2″)–Az(H8)	3.6–4.0	3.9	3.9
Ty(H2′)–Az(H8)	>4.0	4.8	4.6
Ty(CH_3)–Az(H2)	5.2–5.6	5.6	5.7
partner strand			
C18(H1′)–C19(H6)	>3.9	6.8	6.4
C18(H2′)–C19(H6)	>5.0	5.1	4.9
C18(H2″)–C19(H6)	>2.6	5.5	4.6
C18(H3′)–C19(H6)	3.3–3.7	3.2	3.4
C18(H6)–C19(H5)	>3.9	7.6	7.2
cross-strand			
Ax(H2)–C19(H1′)	3.9–4.5	4.1	4.4
exchangeable protons			
G6(NH1)–Ax(H2)	1.8–3.8	3.9	3.7
G7(NH1)–Az(H2)	1.8–3.8	3.4	3.4
Ty(NH3)–Ax(H2)	3.0–5.0	2.9	2.8
Ty(NH3)–Az(H2)	3.0–5.0	3.2	2.9

Molecular Dynamics Simulations. We subjected our starting models to 14-ps NOE-restrained molecular dynamics, as described under Experimental Procedures. Our first simulations resulted in compact, folded structures with van der Waals contacts between extreme ends of the molecule. The two helical stems were bent at certain points where few distance restraints had been included due to overlapped cross peaks in the NOESY spectrum. To eliminate these "weak links" in the molecule, we constrained the dihedral angles of the eight terminal bases (four at either end) of each strand. We also included artificial, NOE-like restraints between internally located protons for the terminal base pair steps. These latter restraints prevented base pair sliding motions that were seen to occur otherwise. Our resulting structures contained stiff, B-form helices at the ends, enabling us to focus on structural features of the central bulge loop segment.

Results from subsequent MD simulations indicated that the resulting structures exhibited conformational heterogeneity at the bulge site. In many of these structures, we noted the existence of hydrogen bonds between functional groups of the bulged bases and phosphate oxygens on the opposite strand. Two such structures are plotted in Figure S4A (starting from the high-twist model) and Figure S4B (starting from the low-twist model) of the Supplementary Material. In order to eliminate the possible electrostatic bias represented by simulations in vacuo, we removed the partial charges and hydrogen-bonding potential of the functional groups of the bulged bases and repeated our simulations. This modification led to much more consistent and reproducible results. The resulting

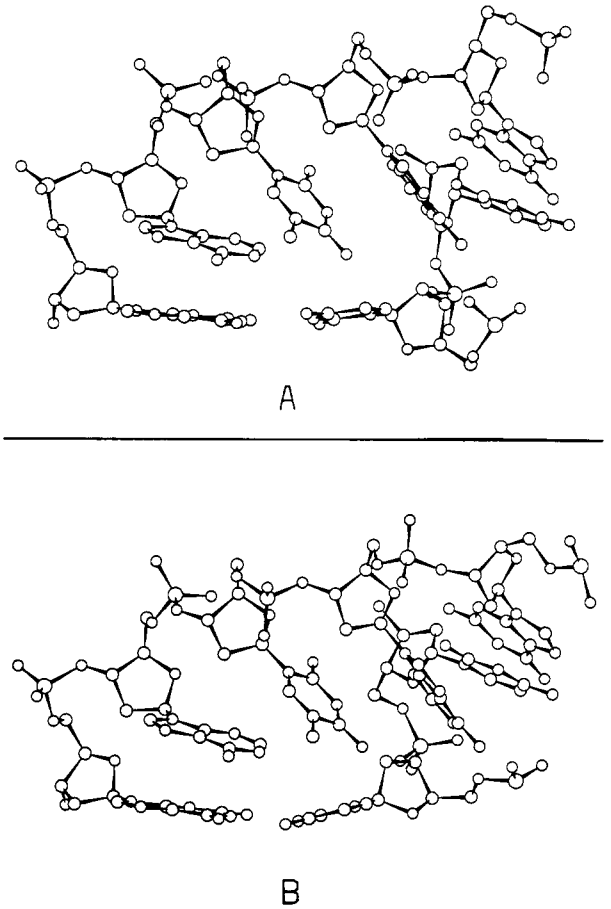


FIGURE 6: Ball-and-stick diagrams of the central portion of distance-refined structures of the ATA-bulge duplex. Only bulge loop and flanking GC pairs are shown. Hydrogens are omitted for clarity. The bottom base pair is G7·C18 in each figure. (A) Refinement starting from the "high-twist" model. (B) Refinement starting from the "low-twist" model. Each model represents the resulting structure from a 14-ps molecular dynamics simulation (point A in Figure 1) using NOE-distance derived restraints, with coordinates averaged over the last 4 ps and subjected to energy minimization. Electrostatic and hydrogen-bonding potentials were omitted for the bulged bases.

distance-refined structures containing the bulge loop and flanking GC pairs are shown in Figure 6A (based on the high-twist starting model) and in Figure 6B (based on the low-twist starting model). Within this central bulge region, the root-mean-squared difference (RMSD) of the atomic coordinates for the starting "high-twist" and "low-twist" models is 3.9 Å. For the two distance-refined structures, this difference is reduced to 1.2 Å. The distance-refined structures for the entire ATA-bulge duplex molecule starting from the high-twist and low-twist models are shown in Figure S5, panels A and B, respectively (Supplementary Material).

Relaxation Matrix Refinement. We performed an additional 2-ps MD simulation with relaxation matrix based refinement on one of the distance-refined structures (starting from the low-twist model), so as to adjust the structure to fit the experimental data more precisely. We used an overall correlation time of 7 ns for the molecule, as this value produced the closest initial fit to the experimental NOESY spectra. During the course of the matrix refinement, the calculated R factor decreased from 0.21 to 0.09. The most notable difference between the structures before and after matrix refinement is the reduction in the magnitude of base pair buckling within the duplex stems of the matrix-refined structure. A stereoview of the relaxation matrix-refined structure, based on the low-twist starting model, is shown in Figure 7. In Figure 8, stacked plots of the base to H1′ region of the experimental NOESY spectrum are compared with those back-calculated from the distance-refined and relaxation matrix-refined structures. A similar comparison of NOESY stacked plots for the base to H2′,H2″ and CH_3 region is shown in Figure 9. The agreement between experimental and calculated spectra is good in both cases, but subtle improvements are seen for the spectrum derived from the matrix-refined structure. Helical bending parameters for the starting models, distance-refined structures, and relaxation matrix-refined structure are listed in Table IV.

Table IV: Helical Bending Parameters for the ATA-Bulge Structures

structure	angle[a]	direction[b]	twist[c]
starting			
high-twist	18	−115	136
low-twist	4.4	−135	39
distance-refined			
high-twist	49	−100	95
low-twist	59	−85	109
R matrix refined			
low-twist	55	−84	84

[a] Angle between the best linear axes of individual stems. [b] Direction of the bend relative to the dyad axis (major groove), averaged over G6·C19 and G7·C18 segments. [c] Twist angle between G6·C19 and G7·C18 helical axis segments.

Geometry of the Bulged Duplex. The ATA-bulge duplex features a partially stacked trinucleotide located between two DNA duplex segments. All three bulged bases are displaced toward the major groove by 2–3 Å. The major groove is flattened by the protrusion of the three bulged bases. The two cytosines opposite the bulge site are separated by about 7 Å. As was indicated by the NOESY data, there is a significant lateral displacement and twisting of one base relative to another. This is reflected in the large helical twist between the G6·C19 and G7·C18 base pairs (see Discussion). Figure S6 (Supplementary Section) shows an expanded view of this base step alone in the relaxation matrix-refined structure. A short C18(H3′)–C19(H6) distance (3.2 Å) is observed, as was predicted by the NMR data, and is indicated in the figure.

Discussion

General Features of the ATA-Bulge Duplex. We have shown in the preceding paper that the incorporation of an A-A-A bulge loop into a DNA dodecamer has little perturbing effect outside the immediate vicinity of the lesion. This finding has now been extended to a mixed sequence A-T-A bulge loop. In general, the two-dimensional NMR results for these two duplexes are quite similar. Many of the spectral and structural characteristics described here are shared in common by these two molecules.

In the ATA-bulge duplex, all nonterminal Watson–Crick base pairs, including G6·C19 and G7·C18 flanking the bulge loop, are intact between 0 and 25 °C. All through-space connectivities usually present for right-handed B-DNA duplexes are seen, with the exception of those at the C18–C19 base step, opposite the bulge loop. Disruption of sequential NOEs at this step is consistent with a large separation between these two bases. All sequential base to sugar NOEs are present throughout the bulge loop, as are certain base–base NOEs in H_2O and D_2O that allow us to place the three bulged bases within the duplex.

Whereas singly-bulged purine residues in DNA have been shown to stack within the DNA duplex (Patel et al., 1982a; Hare et al., 1986; Roy et al., 1987; Woodson & Crothers, 1988; Kalnik et al., 1989a; Nikonowicz et al., 1989, 1990),

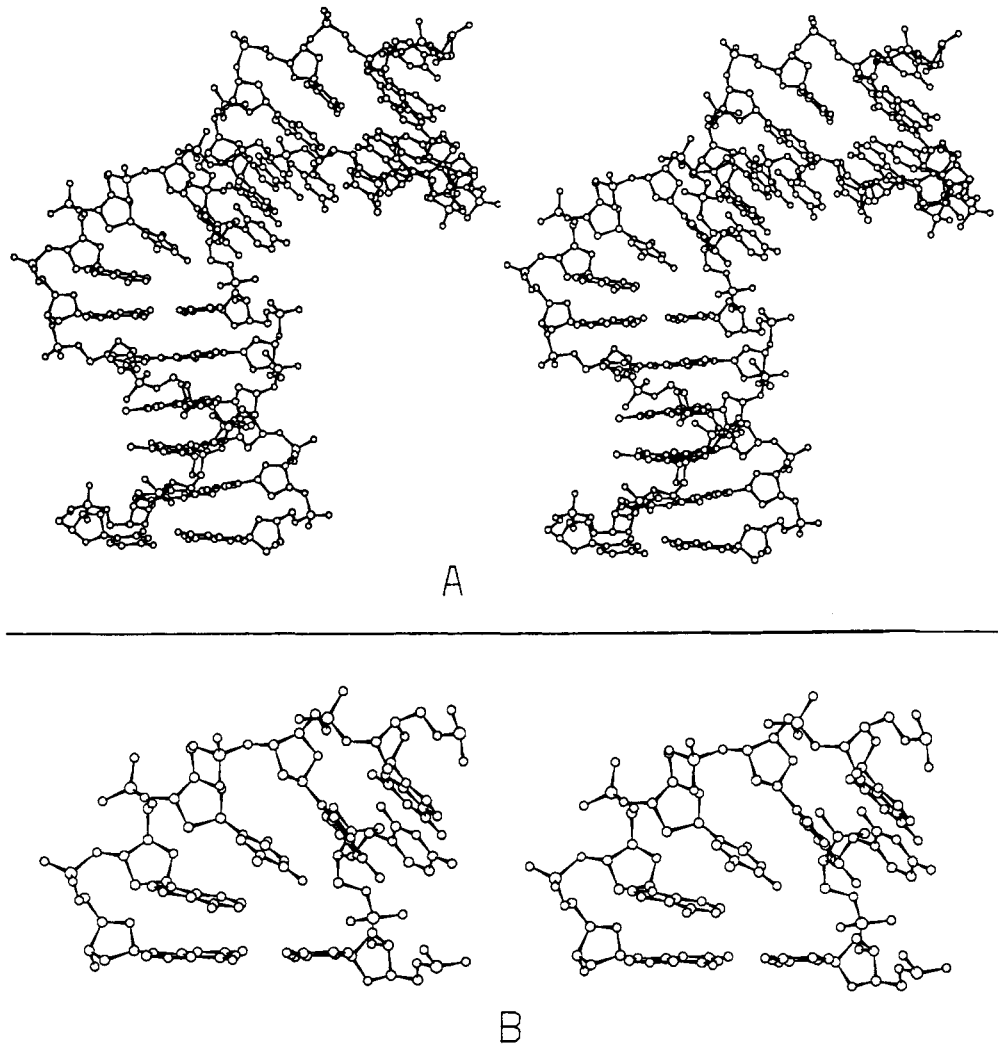


FIGURE 7: Stereoviews of the ball-and-stick structures of the ATA-bulge duplex following relaxation matrix refinement of the distance-refined structure ("low-twist" starting model). (A) Full molecule, with the G12·C13 base pair on the bottom. (B) Close-up of the bulge region, with the G7·C18 base pair on the bottom. Hydrogen atoms are omitted for clarity.

pyrimidine residues may either stack into or loop out from the double helix (Morden et al., 1983, 1990; van den Hoogen et al., 1988; Kalnik et al., 1989b, 1990). In the ATA-bulge duplex, all bulged bases are stacked within the helix, with the Ty residue located between the two bulged adenine residues, Ax and Az. Thus the ATA-bulge duplex behaves much like the A_3-bulge duplex in respect to the intrahelical location of the unpaired bases. It is unclear at present what conformation a multiple pyrimidine bulge would adopt. Such information would be of interest since nucleic acids containing either purine or pyrimidine bulges of equal length exhibit different degrees of gel mobility retardation (Bhattacharyya & Lilley, 1989; Tang & Draper, 1990).

Ax–Ty–Az Step at the Bulge Site. NOE connectivities between protons of the bulged bases in the ATA-bulge duplex demonstrate the existence of base stacking within the bulge loop. Furthermore, the observation of NOEs from the H2 protons of bulged adenines to the flanking guanine imino protons places these bulge loop bases in an intrahelical position within the duplex. The absence of strong base to H1′ NOEs at short mixing times indicates that the unpaired bases remain in the anti configuration about the glycosidic bond as they insert into the duplex. In the refined structures of the ATA-bulge duplex, the bases within the bulge loop stack with one another and with neighboring bases in the duplex. This is consistent with the upfield shift seen for base protons located within the bulge loop. Protons from looped out residues would be magnetically deshielded and would therefore be expected to lie further downfield in the one-dimensional NMR spectrum.

The sequential NOEs between Ty and Az are the weakest of all those within the bulge loop region. For example, the Ty(H1′)–Az(H8) cross peak in Figure 3A, which is overlapped with the C19(H1′)–G20(H8) cross peak (Figure 3B), is revealed to be quite weak in the NOESY experiment on the hemideuterated ATA-bulge duplex sample (Figure S1B, Supplementary Material).

The base stacking at the G6–Ax and the Az–G7 base steps is demonstrated largely through the existence of NOEs between the guanine imino protons and the adenine H2 protons at each of these steps. NOEs involving solvent-exchangeable protons are difficult to quantify in DNA. Therefore, we have used broader bounds in our distance restraints for these protons pairs. This limited our ability to define rigorously the stacking interactions at the G6–Ax and Az–G7 base steps. Each of these steps are sites of conformational heterogeneity among our refined structures. This is most true for the G6–Ax base step, where changes in the degree of stacking between these two bases lead to variations in the overall angle of bending in the molecule. The third bulged adenine, Az, stacks over G7, but is significantly displaced toward the major groove. We do note that this displacement places the H1′ proton of Az

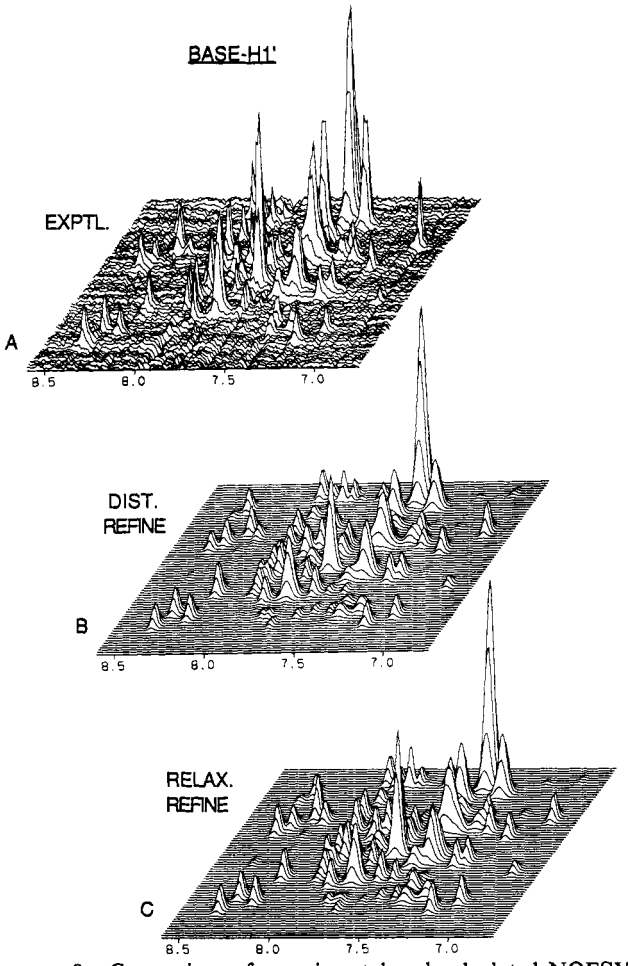


FIGURE 8: Comparison of experimental and calculated NOESY (300-ms mixing time) stacked plots of the base to H1′ proton region for the ATA-bulge duplex in D_2O buffer at 25 °C. (A) Experimental data. (B) Calculated spectrum using the distance-refined structure ("low-twist" starting model). (C) Calculated spectrum using the relaxation matrix-refined structure ("low-twist" starting model).

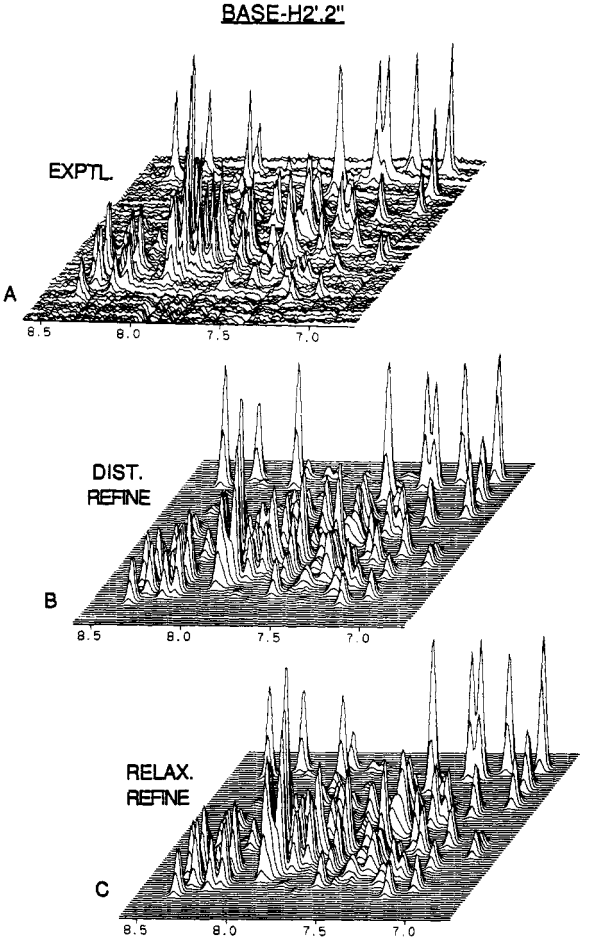


FIGURE 9: Comparison of experimental and calculated NOESY (300-ms mixing time) stacked plots of the base to H2′,2″ and CH_3 proton region for the ATA-bulge duplex in D_2O buffer at 25 °C. (A) Experimental data. (B) Calculated spectrum using the distance-refined structure ("low-twist" starting model). (C) Calculated spectrum using the relaxation matrix-refined structure ("low-twist" starting model).

directly over the G7 ring. This would explain the extreme upfield chemical shift of the Az(H1′) proton (5.27 ppm) relative to those of the H1′ protons on other adenine residues (see Figure 3A and Table II).

There is no direct experimental evidence which would predict the lateral displacement of the bulged bases toward the major groove. There are no specific NOE restraints in this region that would tend to pull the bulged bases in this direction. Therefore, indirect forces such as steric repulsion and nonbonded attraction must account for this structural result in our calculations. However, in the A_3-bulge duplex, we have observed a strong NOE between the H2 proton of the middle adenine, Ay, and the H5 proton of C18 [see Figure 6B, Rosen et al. (1992)]. Displacement of the bulged bases toward the major groove would bring the adenine H2 protons, normally found in the minor groove, closer to the H5 proton of C18. In other words, the structural details we see in the ATA-bulge duplex help to explain the appearance of an unusual cross-strand NOE in the related A_3-bulge duplex molecule. There are likely to be structural differences between an A-T-A bulge loop and an A-A-A bulge loop. However, our chemical shift and NOE data for these two sequences are quite similar, leading us to believe that such differences are minor.

C18–C19 Base Step Opposite the Bulge Loop. Loss of base stacking across from the bulge site causes prominent downfield movement in chemical shift. The protons of C19 are most greatly affected. When compared with the chemical shifts seen in the control duplex, the H6 and H5 protons of this residue move 0.39 and 0.59 ppm downfield, respectively (Table S-II). All sugar protons of this residue move downfield as well. We believe this is due to the unstacked nature of C19 in the ATA-bulge duplex. In our structure, C19 has lost not only its 5′ stacking partner, C18, but also its stacking base on the 3′ side, G20, due to a displacement of the G6·C19 base pair (relative to C5·G20) in the direction of its long axis. By way of contrast, the stacking of G17 over C18 is not disrupted in a similar manner, and, consequently, the base protons of C18 show much smaller chemical shift differences between the ATA-bulge and control duplexes.

The absence of the base to H1′ or H2′,2″ sequential NOEs between C18 and C19 supports the existence of a spatial separation between these two bases in the ATA-bulge duplex. However, the presence of a sequential connectivity within the H3′ region implies that the disruption is not a simple wedging apart of these two bases, as it is often pictured in schematic drawings of bulge loops within duplexes. In our structure, we find that the disruption is primarily due to a lateral displacement and twisting apart of these two bases, more in the nature of shearing than wedging (see Figure S6). This conformational feature has an important influence on the overall geometry of helical bending at the bulge loop site.

The backbone torsion angles of all bases at or near the bulge site were left unconstrained throughout the molecular dynamics simulations. As a result, we find that the same general

structure can be achieved through a wide variety of torsional angle combinations in this segment. The C18–C19 base step is a case in point. The γ torsion of C19 is trans in the distance-refined structure derived from the high-twist starting model but has the more usual gauche$^+$ conformation in the structure derived from the low-twist starting model. As there is no direct information regarding these torsion angles, we cannot choose one configuration over the others. Nevertheless, despite the conformational heterogeneity within the DNA backbone of our structures, we can reproducibly define the relative orientation of the bulged bases within the loop solely on the basis of the interproton distances derived from our NMR data in solution.

Relaxation Matrix versus Distance-Refined Structures. In general, relaxation-matrix refinement will improve the accuracy of structures determined by NMR, since spin diffusion is fully accounted for in the relaxation matrix approach (Keepers & James, 1984). Proton–proton distances longer than the reference will most likely be underestimated in the isolated two-spin approximation, leading to distortion of the resulting three-dimensional structure. Correction of these errors through matrix-based refinement can relieve this stress. This is evident, for example, in the reduction of base pair buckling in our structure after relaxation matrix refinement (compare Figures 7 and S5).

There are some apparent discrepancies between the original distance restraints listed in Table III and the proton–proton distances in the matrix-refined structure. Whereas certain distance restraints were obtained from NOESY experiments using a partially deuterated duplex, our volume inputs for matrix-based refinement were derived from the NOESY experiment on the fully protonated sample only. We therefore find that, in certain cases, the large separation between proton pairs that was maintained during distance refinement has decreased when the structure is refined against overlapped NOESY cross peak volumes.

We also note that the critical distances between the CH_3 protons of Ty and the H2 protons of both Ax and Az are quite large in both the distance-refined and the relaxation matrix-refined structures. These NOEs, which were present but quite weak in the experimental data, are well accounted for in the back-calculated NOESY data, due to the existence of spin-diffusion pathways involving the H1′ protons of residues Ax and Ty. The use of shorter distance restraints for the Ax(H2)–Ty(CH_3) and the Az(H2)–Ty(CH_3) proton pairs, based on the observed NOEs (box, Figure 4), would have led to structural inaccuracies. We note that our conservative use of minimum distance estimates only for weak cross peaks helped to minimize this potential error in our distance-refined structures.

Comparison of Structural and Footprinting Results at Bulge Sites. Our structure for the DNA trinucleotide bulge differs from an earlier proposed model (Bhattacharyya & Lilley, 1989), in which the bulged base at the 5′ end of the loop stacks within the duplex whereas all subsequent bases are looped out and exposed to solvent. This model was based on results of modification studies using two chemical probes, diethyl pyrocarbonate (DEP) and osmium tetroxide, specific for solvent-exposed adenine and thymine residues, respectively. For A_n and T_n bulge loops (n = 3 or 5), the authors found that the 5′ base of the loop was partially protected from modification while remaining bases were hyperreactive. In our structure, all bulged bases are at least partially stacked within the duplex. However, our results do not necessarily disagree with the chemical modification data. The protrusion of the bulged bases toward the major groove, as seen in our structure, leads to increased solvent exposure of the Ax and Az N7 atoms, and of the Ty C5–C6 bond, the targets of DEP and osmium tetroxide, respectively. Therefore, it is conceivable that partially stacked bases would still be accessible to small molecule attack from the major groove. In our model, Az protrudes farthest from the helical center, with concomitantly greater solvent exposure of its N7 atom. This is consistent with the observation that bases toward the 3′ end of a bulge loop are more susceptible to chemical modification than are those located at the 5′ end.

Kink and Twist Angles at the Bulge Site. The angle of bending at the bulge loop site in our refined structures ranges between 50° and 60°. However, all of our distance constraints are short-range, and propagation of small errors could yield large deviations in global structural features, such as the total angle of helical bending. Therefore, it is conceivable that structures with larger or smaller kink angles would also be compatible with our NMR data. We did observe that additional molecular dynamics simulations occasionally resulted in structures with greater or lesser degrees of kinking. In general, this variation was due to partial unstacking of the G6–Ax or Az–G7 base steps. We previously noted that these base steps were susceptible to conformational heterogeneity due to the more qualitative nature of the A(H2) to G(imino) distance constraints. On the basis of the chemical shift data indicative of a well-stacked geometry within the bulge loop region, we have presented those structures with the greatest amount of stacking between the adenines of the bulge loop and the flanking guanines on either side.

Other aspects of the geometry of bending in the ATA-bulge duplex structure, such as the direction of the bend and the relative twist between the two helical segments, are better defined. Convergence of the "helical advance" parameter in the two distance-refined structures is especially pleasing, since we generated starting models with radically different twists angles about the bulge site. In our relaxation matrix-refined structure, the helical twist between the two base pairs, flanking the bulge is 84°, or about 58% of the value expected between base pairs separated by four steps in B-DNA. The value of this geometric parameter must be considered in the interpretation of gel mobility results for DNA molecules containing multiple-base bulge loops separated by a fixed number of base pairs [see Discussion in Tang & Draper (1990)].

The large twist angle between G6·C19 and G7·C18 requires that one must carefully define the terms major and minor groove when discussing the geometry of the bulge site. These features, which are characteristic of a linear double helix, are less easily described in bent or otherwise deformed structures. We have observed the strict definition of major and minor grooves relative to the dyad axis of a given base pair within the duplex. By such a definition, the position of the grooves changes rather abruptly as one traverses the bulge site. This transition is eased somewhat by the spatial separation of the two helical segments. Any description of the direction of DNA bending is dependent on the frame of reference used. The values listed in the second column of Table IV for the direction of DNA bending are expressed as an average of the values relative to the dyad axes of the G6·C19 and G7·C18 base pairs, as defined by the best global helical axis of each six base pair stem according to the CURVES algorithm (Lavery & Sklenar, 1989, 1990). A value of 0° for this parameter represents bending toward the major groove, 180° toward the minor groove, +90° toward the first (bulge-containing) strand, and −90° toward the second (partner) strand. In our structures,

bending occurs away from the bulge-containing strand (between −80° and −100°), with little deviation toward either groove. This is in agreement with an analysis based on gel electrophoresis (Rice & Crothers, 1989) for a single-base bulge in DNA.

Electrostatic Effects of Unpaired Bases during in Vacuo Calculations. We have found that our simulations in vacuo are more reproducible if the electrostatic and hydrogen-bonding potentials of the bases within the loop are removed. The hydrogen bonds that otherwise form are, we believe, an artifact of the force field we have used, one in which hydrogen bonding is explicitly included in order to maintain the Watson–Crick alignment of the base pairs. Furthermore, a distance-dependent dielectric, which has been shown to be useful for DNA modeling in vacuo (Gelin & Karplus, 1975), gives greater emphasis to short-range electrostatic interactions. We believe that our conformational searches were hindered by the formation of favorable electrostatic pairs between unpaired bases in the loop and phosphate backbone oxygens. These interactions "froze" the molecule in conformations not necessarily representative of that in solution. Simulations that include explicit water molecules and counterions could alleviate this problem in future computational studies involving DNA or RNA molecules with unpaired bases.

Conclusions

Our results demonstrate that a dodecamer duplex can easily accommodate a A-T-A bulge loop with little structural perturbation beyond the immediate vicinity of the loop itself. Despite the potentially destabilizing effect of the adjacent unpaired bases, the guanine and cytosine bases that flank the bulge loop site participate in normal Watson–Crick base pairing. Our NOESY data in H_2O and D_2O demonstrate that all three bases of an A-T-A bulge loop stack within the dodecamer DNA duplex, in an anti configuration about their glycosidic bonds. In the preceding paper in this issue, we have demonstrated that for A, A-A, and A-A-A bulged loops, all of the unpaired adenines are stacked within the DNA helix. The results of this work extend this observation to a mixed-sequence bulge loop containing one thymine and two adenine residues. Absence of certain sequential NOEs across from the bulge loop indicates a disruption of stacking at this base step.

Previous studies involving multiple-base bulge loops within DNA duplexes have shown that these defects cause significant bending of the helical axis. In this combined NMR and molecular dynamics study, we have endeavored to expand the structural description to a more quantitative level. Molecular dynamics simulations, using both proton–proton distance and NOESY cross peak volume restraints, yield a structural model for the bulge loop in DNA. The loop bases stack within the duplex and upon each other, while protruding into the major groove. On the opposite strand, the bases flanking the bulge loop are laterally displaced by approximately 7 Å. The angle of bending at the bulge site ranges between 50° and 60° in our refined structures, with bending in the direction away from the bulge-containing strand. We also find a significant "helical advance" as one moves across the bulge site, slightly more than 50% of the helical twist that would be produced by three fully paired bases.

The results of this study have implications for the tertiary structure of large RNA molecules, many of which are thought to contain bulge loop of varying number of bases. The geometry of bulge loops may be crucial in determining the relative positioning of helical segments in large RNA molecules. Our results demonstrate a structural mode for the bending of DNA caused by bulged loops. Differences between our structure (Figure 7) and an earlier model of bulge loops within DNA (Bhattacharyya & Lilley, 1989) point to the importance of NMR studies that allow for greater resolution of details in the secondary and tertiary structure of complex nucleic acid molecules in solution.

Acknowledgments

We thank A. Brunger and M. Nilges at Yale University for making available an early version of the RELAX protocol within XPLOR. We acknowledge the use of the Molecular Modeling Facility for Molecular Biology at Columbia University, supported in part by NSF Grant DIR-8720229.

Supplementary Material Available

Two tables listing exchangeable and nonexchangeable proton chemical shift differences between ATA-bulge and control duplexes and six figures showing expanded contour plots of the NOESY spectra of the ATA-bulge duplex in D_2O buffer, representative buildup curves for NOEs spanning the Ax–Ty or the A21–T22 base step, ball-and-stick diagrams of distance-refined structures of the ATA-bulge duplex, and a stick drawing showing the relative orientation of bases C18 and C19 in the relaxation matrix-refined structure of the ATA-bulge duplex (11 pages). Ordering information is given on any current masthead page.

References

Bhattacharyya, A., & Lilley, D. M. J. (1989) *Nucleic Acids Res. 17*, 6821–6840.

Gelin, G. R., & Karplus, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2002–2006.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

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

Hsieh, C.-H., & Griffith, J. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4833–4837.

Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D., & Sussman, J. L. (1992) *J. Mol. Biol.* (in press).

Kalnik, M. W., Norman, D. G., Swann, P. F., & Patel, D. J. (1989a) *J. Biol. Chem. 264*, 3702–3712.

Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989b) *Biochemistry 28*, 294–303.

Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., & Patel, D. J. (1990) *J. Biol. Chem. 265*, 636–647.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–426.

Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63–91.

Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655–667.

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

Morden, K. M., Gunn, B. M., & Maskos, K. (1990) *Biochemistry 29*, 8835–8845.

Nikonowicz, E., Roongta, V., Jones, J. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714–8725.

Nikonowicz, E. P., Meadows, R. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 4193–4204.

Nilges, M., Habazettl, J., Brunger, A. T., & Holak, T. A. (1991) *J. Mol. Biol. 219*, 499–510.

Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591–616.

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

Patel, D. J., Kozlowski, S. A., Nordheim, A., & Rich, A. (1982b) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1413-1417.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310-7311.
Rice, J. A., & Crothers, D. M. (1989) *Biochemistry 28*, 4512-4516.
Rosen, M. A., Live, D., & Patel, D. J. (1992) *Biochemistry* (preceding paper in this issue).
Roy, S., Sklenar, V., Appella, E., & Cohen, J. S. (1987) *Biopolymers 26*, 2041-2052.
Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327-341.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286-292.
Tang, R. S., & Draper, D. E. (1990) *Biochemistry 29*, 5232-5237.
van den Hoogen, Y. T., van Beuzekon, A. A., van den Elst, H., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988) *Nucleic Acids Res. 16*, 2971-2986.
Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 3130-3141.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Yip, P., & Case, D. (1989) *J. Magn. Reson. 83*, 643-648.

Synthesis of Fully Active Biotinylated Analogues of Parathyroid Hormone and Parathyroid Hormone-Related Protein as Tools for the Characterization of Parathyroid Hormone Receptors

Eliahu Roubini,[‡] Le T. Duong,[‡] Susan W. Gibbons,[‡] Chih T. Leu,[‡] Michael P. Caulfield,*[,‡] Michael Chorev,[§] and Michael Rosenblatt[‡]

Merck, Sharp and Dohme Research Laboratories, West Point, Pennsylvania, 19486, and Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, 91120, Israel

Received November 14, 1991; Revised Manuscript Received January 30, 1992

ABSTRACT: The synthesis, purification, and characterization of biotinylated analogues of parathyroid hormone (PTH) and PTH-related protein (PTHrP) are described. A novel methodology was developed which allowed the selective biotinylation during solid-phase synthesis of either the Lys^{13} or Lys^{26} residue in PTH/PTHrP sequences. Incorporation of orthogonally protected N^{α}-Boc-Lys(N^{ϵ}-Fmoc) at a selected position in the sequence, followed by selective side-chain deprotection and biotinylation of the ϵ-amino group, permitted modification of the specific lysine only. Biotinylated analogues of [$Nle^{8,18}$,Tyr^{34}]bPTH(1-34)NH_2 (analogue **1a**) were prepared by modification of Lys^{13} with a biotinyl group (analogue **1**) or a biotinyl-ϵ-aminohexanoyl group (analogue **2**) or at Lys^{26} with a biotinyl-ϵ-aminohexanoyl group (analogue **3**). A biotinylated PTHrP antagonist [Leu^{11},D-Trp^{12},Lys^{13}(N^{ϵ}-(biotinyl-β-Ala))]PTHrP(7-34)NH_2 (analogue **5**), was also prepared. In a different synthetic approach, selective modification of the thiol group of [Cys^{35}]PTHrP(1-35)NH_2, in solution, with *N*-biotinyl-*N'*-(6-maleimidohexanoyl)hydrazide, resulted in analogue **4**. The high affinities of the biotinylated analogues for PTH receptors present in human osteosarcoma B-10 cells or in porcine renal cortical membranes (PRCM), were comparable to those of the underivatized parent peptides. The analogues were also highly potent in stimulation of cAMP formation (analogues **1-4**) or inhibition of PTH-stimulated adenylyl cyclase (analogue **5**) in B-10 cells. The most potent analogue (analogue **1**) had potencies in B-10 cells (K_b = 1.5 nM, K_m = 0.35 nM) and in porcine renal membranes (K_b = 0.70 nM) identical or similar to those of its parent peptide, respectively. Furthermore, these high binding affinities were retained in the presence of either avidin, streptavidin, or anti-biotin (1 μM). Radioiodination of PTH analogues (analogues **1-2**) generated highly potent radioligands which bind to a single class (B_{max} = 390-450 fmol/mg of protein) of high affinity (K_d = 0.24-0.61 nM) PTH receptors on renal membranes and which also bind to immobilized streptavidin or anti-biotin. Cross-linking of biotinylated radioligands (analogues **1-3**) and analysis by SDS-PAGE and autoradiography showed specific labeling of renal PTH receptors (M_r = 75 000). The biotinylated analogues prepared in this study, and their radiolabeled derivatives, serve as useful tools for the identification and isolation of PTH receptors.

Biotinylated derivatives of peptide hormones have been useful as specific probes for the localization and visualization of membrane receptors in cells by means of affinity cytochemistry techniques (Childs et al., 1986; Yamasaki et al., 1988; Wilchek & Bayer, 1990) and in the purification of hormone receptors in combination with avidin affinity chromatography (Finn et al., 1984; Kohanski & Lane, 1985; Hazum et al., 1986; Lee et al., 1989). Parathyroid hormone, a linear peptide of 84 amino acids, acts primarily at bone and kidney tissue to regulate calcium levels in blood. Parathyroid hormone-related protein (PTHrP), a 141 amino acid peptide associated with the clinical syndrome of humoral hypercalcemia of malignancy, has been isolated and cloned (Suva et al., 1987; Strewler et al., 1987; Mangin et al., 1988). PTHrP shares considerable sequence homology with PTH within the N-terminal 1-13 region of the peptide: 8 positions contain identical amino acids in the two hormones. There is little homology beyond position 13, including the C-terminal 25-34

* Address correspondence to this author.
[‡] Merck, Sharp and Dohme Research Laboratories.
[§] The Hebrew University of Jerusalem.

0006-2960/92/0431-4026$03.00/0