Sequence Effects on RNA Bulge-Induced Helix Bending and a Conserved Five-Nucleotide Bulge from the Group I Introns†

Kevin J. Luebke* and Ignacio Tinoco, Jr.

Department of Chemistry, University of California at Berkeley, and Structural Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720-1460

Received April 15, 1996; Revised Manuscript Received June 21, 1996⊗

ABSTRACT: Bulge loops introduce bends in RNA double helices. Thus, a role for bulge loops in the tertiary folding of RNA is to orient helical elements. The location, size, and sequence of a five-nucleotide bulge are conserved in many of the self-splicing group I introns. We have used gel electrophoretic analysis of helix bending to test the hypothesis that this bulge loop is conserved to control the angle between the flanking helices. Interruption of an RNA duplex by the five-nucleotide bulge of the group I intron from *Tetrahymena thermophila* results in an electrophoretically retarded species, indicative of bending by the bulge. However, mutation of conserved bases in the bulge has a small effect on the retardation, suggesting that the average induced bend angle is not strongly dependent on the conserved sequence. Electrophoretic analysis of a mixture of bulged duplexes containing all five-nucleotide bulges reveals that *most* five-nucleotide bulge sequences induce bends that are similar to the bend induced by the conserved bulge. We have calibrated relative electrophoretic mobilities with bends of known magnitude, and characterized the distribution of bulge sequences among bend angles. Though the entire range of bend angles induced by different five-nucleotide bulges is from approximately 45° to 75°, most (>85%) five-nucleotide bulge loops induce bends between 65° and 75°. We have identified several of the anomalous five-nucleotide bulge sequences that induce bends of magnitude smaller than 65°. They are generally, though not universally, pyrimidine-rich.

The global three-dimensional structure of a folded RNA molecule is characterized by the spatial orientation of its double helices. These universal elements of RNA structure are oriented with respect to each other at multibranch junctions and non-Watson−Crick structural elements. However, little is known about how the local features of structure determine the orientation of helices around them. The simplest example of the relationship between local structure and global folding is the helix bending induced by bulge loops.

A bulge loop is defined as one or more nucleotides that interrupt one strand of an otherwise continuous Watson−Crick base-paired double helix (Lilley, 1995). Gel electrophoresis (Bhattacharyya & Lilley, 1989; Bhattacharyya *et al.*, 1990; Hsieh & Griffith, 1989; Rice & Crothers, 1989; Riordan *et al.*, 1992; Tang & Draper, 1990, 1994; Wang & Griffith, 1991), electron microscopy (Wang *et al.*, 1992), fluorescence resonance energy transfer (Gohlke *et al.*, 1994), and transient electric birefringence experiments (Zacharias & Hagerman, 1995) have shown that bulge loops induce local, directional bends between the helical segments that flank them. The extent of bending depends on the number of nucleotides in the bulge, the base sequence of the bulge, the base pairs flanking the bulge, and the presence of divalent cations.

Studies of helix bending by RNA bulge loops have focused primarily on bulges composed entirely of uridines or adenosines [see Riordan *et al.* (1992) for an exception]. Within a given context of flanking duplex sequences, a bulge of uridines induces a smaller bend than an equal number of bulged adenosines (Bhattacharyya *et al.*, 1990; Zacharias & Hagerman, 1995). The difference is accentuated in the presence of millimolar concentrations of $MgCl_2$ (Zacharias & Hagerman, 1995). The structural basis of this sequence dependence is unknown, though sequence-dependent differences in base stacking within the bulge and between the bulge nucleotides and the flanking helices are likely to be important. Helix bending induced by mixed-sequence bulges has not been studied systematically.

A five-nucleotide bulge loop is highly conserved in the P5 extension of the group IB and IC introns (Figure 1) (Collins, 1988; Michel & Westhof, 1990; Michel & Cummings, 1985). The position of the bulge is conserved at 23 nucleotides from the catalytic core, and the sequence of the bulge is always adenosine-rich. Within those intron sequences containing the P5 extension, adenosines at positions labeled 1 and 3 in Figure 1 are absolutely conserved, and the adenosine at position 4 is highly conserved.

Murphy and Cech (1993, 1994) have identified an independently folding domain of tertiary structure within the group I intron from *Tetrahymena thermophila* and found that the bulge in P5a is required for the formation of that tertiary structure. One postulated role for this bulge in promoting the folding of the domain is to correctly orient the helical elements flanking it (Murphy & Cech, 1993). We initially

† This research was supported in part by National Institutes of Health Grant GM 10840 and by U.S. Department of Energy Grant DE-FG03-86ER60406. K.J.L. was supported in part by a postdoctoral fellowship from the American Cancer Society.

* Author to whom correspondence should be addressed at Hewlett-Packard Co., 3500 Deer Creek Rd., Building 25U-5, Palo Alto, CA 94304.

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1996.

S0006-2960(96)00914-2 CCC: $12.00

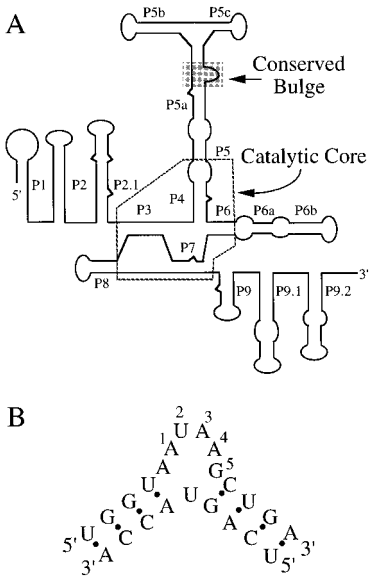

FIGURE 1: (A) Schematic representation of the secondary structure of the group I intron from *Tetrahymena thermophila*. The conserved five-nucleotide bulge is shaded, and the catalytic core is outlined. (B) Sequence of the five-nucleotide bulge loop and flanking double helices in the P5 extension of the group I intron from *Tetrahymena thermophila*. Nucleotides that have been varied in the work reported here are numbered 1−5.

hypothesized that the bulge in the P5 extension of the group I introns is conserved to conserve the angle between the flanking helices. In testing this hypothesis, we have undertaken a study of the sequence dependence of bending induced by five-nucleotide bulges of mixed sequence.

MATERIALS AND METHODS

Preparation of RNA Duplexes. RNA strands were synthesized by transcription of DNA templates with T7 RNA polymerase. DNA templates were chemically synthesized by standard methodology and amplified with the polymerase chain reaction. Approximately 0.2 pmol of template was amplified per 100 μL reaction, using 100 pmol of each primer. Primers for the strands complementary to the bulged strands were 5′TAATACGACTCACTATAGGGACAAAGCTTCTTGCAGT3′ and 5′GGATCCGAAGACCTCAGTCTTGTCCA3′. Primers for the strands that contained the bulge loops were 5′TAATACGACTCAGTATAGGATCCGAAGACCTCAGTCTT3′ and 5′GGGACAAAGCTTCTTGCAGTACGGTT3′ (or 5′GGGACAAAGCTTCTTGCAGTACGGTA3′ for the strands used to calibrate electrophoretic mobilities with bend angles).

After amplification with 25 cycles of the polymerase chain reaction, the reaction mixtures were precipitated with ethanol. The precipitates were used directly in transcription reactions. The product of two 100 μL amplification reactions was used in each 100 μL transcription reaction. In addition to DNA template, the transcription reactions contained 2.4 mM adenosine triphosphate, 4 mM guanosine triphosphate, 4 mM cytidine triphosphate, 4 mM uridine triphosphate, α-^{32}P labeled adenosine triphosphate (Amersham, 10 mCi/mL), 5 mM dithiothreitol, 15 mM $MgCl_2$, 40 mM Tris·HCl, pH 8.1, 0.01% Triton X-100, and T7 RNA polymerase. T7 RNA polymerase was prepared as described previously (Davanloo *et al.*, 1984; Wyatt *et al.*, 1991).

Transcription reactions were allowed to proceed for 2−3 h at 37 °C, after which the precipitated pyrophosphate was removed from the reaction mixtures by centrifugation. The RNA was precipitated from the supernatants with ethanol, and the precipitates were purified by electrophoresis in a 6% polyacrylamide gel containing 8 M urea. The product bands were visualized by UV-shadowing, excised, crushed, and soaked in 0.3 M sodium acetate, pH 5.2, overnight at room temperature. The gel slurries were filtered, and product RNA was precipitated from the filtrates with ethanol.

To form bulged duplexes, the precipitates were dissolved in water, and roughly equimolar amounts (as judged by radioactivity) of complementary strands were mixed. The resulting mixtures were heated to 95 °C, and then allowed to cool to room temperature over the course of about 2 h. The samples were chilled overnight at 4 °C prior to electrophoretic analysis. Samples were equilibrated in the electrophoresis buffer at the temperature of the gel for at least 15 min before addition of a glycerol loading buffer for electrophoresis.

Nondenaturing Gel Electrophoresis. All nondenaturing gel electrophoresis was in 10% polyacrylamide, cross-linked 19:1. Gels were run at room temperature in 100 mM Tris−HEPES buffer, pH 7.8, containing 5 mM $MgCl_2$. In experiments to calibrate relative mobility with bend angle, gels were run at 4 °C in 45 mM Tris−borate buffer, pH 8.0, containing 5 mM $MgCl_2$. In all cases, buffer was circulated between cathodic and anodic reservoirs at a rate of approximately 4 mL/min.

Gels were visualized by autoradiography and storage phosphor analysis (Molecular Dynamics PhosphorImager). Absolute electrophoretic mobilities were measured between the point of maximum radioactivity of each band and the bottom of each initial well. The relative mobility of a bulged duplex is defined as the ratio of its absolute mobility in a given electrophoretic experiment to the absolute mobility in the same experiment of the unbulged control duplex having the same number and sequence of base pairs.

Electrophoretic Selection of Bulged Duplexes. Rare bulge sequences that induce anomalously small bends were selected by electrophoresis from 150 pmol of a mixture of duplexes containing all possible five-nucleotide bulges. The mixture of RNA bulges was prepared by transcription of a mixture of DNA templates synthesized using an equimolar mixture of four phosphoramidites in the coupling steps corresponding to the bulge nucleotides. After electrophoresis of the bulged RNA duplexes, a 4 mm long slice of the gel was excised at the position of the duplex containing a bulge of five uridines (approximately 300 mm from the well) which was run in a lane beside the mixture on the gel. The RNA was extracted from the gel slice by crushing and soaking, and then precipitated with ethanol and glycogen.

The bulged strand of the precipitated RNA was reverse-transcribed at 70 °C using 5 units of Tth (*Thermus thermophilus*) polymerase in the presence of 1 mM $MnCl_2$ (Myers & Gelfand, 1991). The Mn^{2+} was then chelated with 1 mM EGTA, $MgCl_2$ was added to the solution to a final concentration of 1.2 mM, and the DNA-directed DNA polymerase activity of Tth polymerase in the presence of Mg^{2+} was employed to amplify the cDNA with the polymerase chain reaction (Myers & Gelfand, 1991). The T7 promoter was added during amplification, and the resulting DNA duplex was used as a template for the transcription of the selected RNA. Another portion of the DNA duplex was cloned for analysis of individual selected sequences.

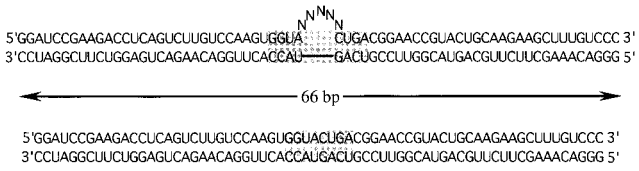

FIGURE 2: RNA duplexes for electrophoretic analysis. Shaded base pairs are the same as those flanking the five-nucleotide bulge in the group I intron from *Tetrahymena thermophila*. Electrophoretic mobilities of 66 base pair duplexes containing central five-nucleotide bulge loops (indicated by NNNNN) were compared with the continuous 66 base pair control duplex.

Sequence Analysis of Bulges. The ends of the amplified cDNA were rendered cohesive by digestion with *Bam*HI and *Hin*dIII, and the resulting mixture of duplexes was ligated into the large *Bam*HI/*Hin*dIII fragment of pUC19. The resulting plasmids were propagated in *Escherichia coli* of the strain DH5α. Individual colonies were cultured from the resulting mixture. Plasmid was isolated from 5 mL cultures of the individual colonies, and the insert region was sequenced by the dideoxy chain termination method of Sanger (Sanger *et al.*, 1977).

The composition of the mixture was analyzed by reverse-transcribing some of the RNA, amplifying the resulting cDNA by the polymerase chain reaction, cloning the DNA into *E. coli*, and sequencing 16 individual clones. Of the 80 (5 × 16) variable nucleotides sequenced, 21 were adenosine, 22 were cytidine, 17 were guanosine, and 20 were uridine. The four bases were distributed evenly between the five variable positions, and all possible dinucleotides were represented.

RESULTS

Effect of Mutations in the P5 Bulge on Electrophoretic Mobility. A bend in an RNA or DNA double helix results in electrophoretic retardation of that helix compared to a linear duplex of the same number of base pairs. The amount of retardation increases with the magnitude of the bend (Zacharias & Hagerman, 1995; Crothers *et al.*, 1990). We initially hypothesized that the sequence of the five-nucleotide bulge in the P5 extension of the group I introns is conserved to ensure a particular orientation of the flanking helices. To test this hypothesis, we used gel electrophoresis to investigate the effect of mutations in this bulge sequence on the extent of bending it induces. We prepared a series of 66 base pair RNA duplexes with centrally located five-nucleotide bulge loops (Figure 2). The sequence of four base pairs flanking each bulge was the same as the sequence context of the bulge in the group I intron. The sequence of one of these bulge loops was identical to the sequence of the corresponding bulge from the group I intron of *Tetrahymena thermophila*. One of three conserved adenosines was changed to a uridine in each of three other bulged duplexes. Duplexes with bulges of five adenosines and five uridines were also prepared. The electrophoretic mobilities of these bulged duplexes were compared with the electrophoretic mobility of a continuous duplex of the same sequence (Figure 3).

All of the bulged duplexes were retarded on the gel with respect to the continuously base-paired duplex, indicating that each of the bulges creates a bend in the double helix. A bulge of five adenosines bends the duplex to a greater extent than a bulge of five uridines (relative mobilities of 0.889 and 0.951, respectively), consistent with observations by

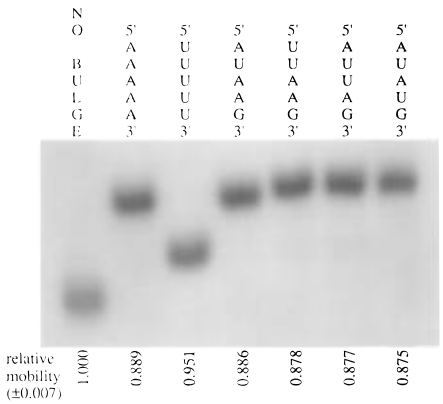

FIGURE 3: Electrophoretic mobility of RNA duplexes containing five-nucleotide bulges. The gel was run at room temperature in 100 mM Tris−HEPES buffer, pH 7.8, containing 5 mM $MgCl_2$. The bulge sequence is shown above each lane. Relative mobilities were determined by comparison with the mobility of the continuously base paired control duplex in the lane labeled "NO BULGE".

others (Bhattacharyya *et al.*, 1990; Zacharias & Hagerman, 1995). The duplex containing the bulge sequence from the intron (lane labeled AUAAG) was retarded essentially the same as the duplex containing the bulge of five adenosines (relative mobilities of 0.886 and 0.889, respectively). The three duplexes in which adenosines from the intron bulge were substituted with uridines were retarded with respect to the control duplex by approximately equal amounts, slightly more than the bulge of five adenosines and the bulge from the intron sequence.

Distribution of Bend Angles among Five-Nucleotide Bulges. A mixture of bulged duplexes in which all possible five-nucleotide bulge sequences were represented was prepared. The mixture of bulged duplexes was analyzed electrophoretically, and the result is shown in Figure 4A. The band on the gel is broader for the mixed population of bulged duplexes than for any of the discrete duplexes, bulged or unbulged, indicating a range of bend angles represented within the population. However, the mixture of bulged duplexes is not uniformly distributed among mobilities in the range between the bulge of five uridines and the bulge of five adenosines (Figure 4B,C,D).

The shape of the distribution reflects the distribution of sequences in the mixture, the distribution of bulge sequences among different bend angles, and the nonlinear relationship between bend angle and electrophoretic mobility (Zacharias & Hagerman, 1995; Crothers *et al.*, 1990). Sequence analysis of the bulge mixture indicates that the bulge sequences are uniformly distributed among all possibilities (see Materials and Methods). If there were a corresponding uniform distribution of bend angles, the range of mobilities would favor the highest mobilities (Zacharias & Hagerman, 1995). Instead, the observed mobility distribution favors the lowest mobilities—the largest bend angles.

The majority of the bulged duplexes in the mixture comigrate with the bulge of five adenosines and the bulge from the *Tetrahymena* intron (compare Figure 4B and Figure 4C). Most five-nucleotide bulge sequences induce a bend of nearly the same angle as the conserved bulge sequence. Electrophoretic comparison of the mixture of bulged duplexes with the duplex containing the bulge of five uridines reveals that a bulge of five uridines induces an anomalously small bend (Figure 4D). Less than 5% of the RNA in the mixture,

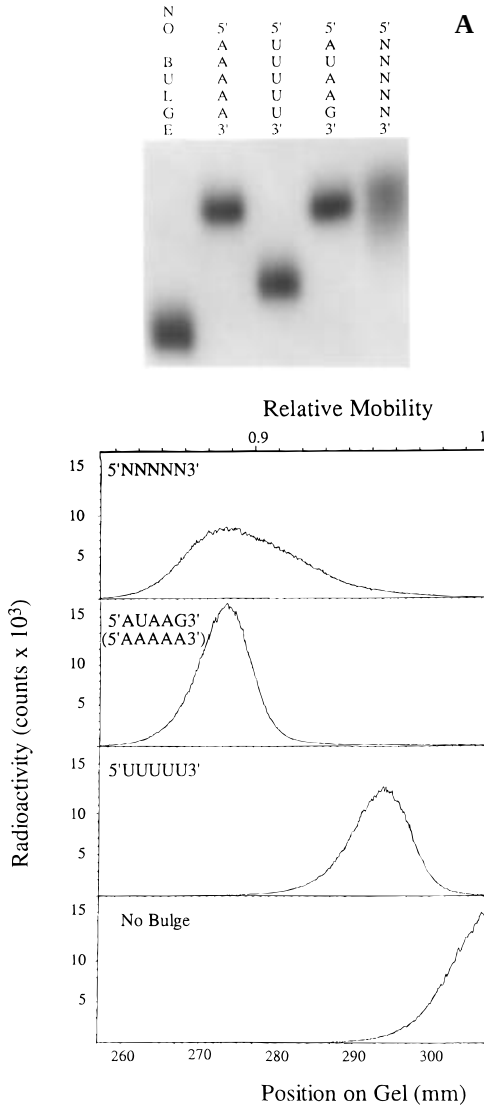

FIGURE 4: Electrophoretic analysis of a mixture of bulged duplexes containing all five-nucleotide bulge sequences and comparison with discrete bulged duplexes. (A) Autoradiogram of a polyacrylamide gel comparing electrophoretic mobilities of bulged and control duplexes with the mixture of bulged duplexes in lane labeled 5′NNNNN3′. Electrophoresis conditions were as in Figure 3. (B−E) Plots of radioactivity *vs* position on the gel for the polyacrylamide gel in (A). The trace in (C) is for the duplex containing the bulge sequence 5′AUAAG3′. However, the position of the peak maximum for the duplex containing the bulge sequence 5′AAAAA3′ is indistinguishable from this peak maximum, and the peak shapes are similar.

as quantitated by phosphor storage analysis, was as mobile or more mobile than the duplex containing a bulge of five uridines.

Selection of Five-Nucleotide Bulges with Anomalously Small Bends. To obtain a clearer picture of the sequence effects on bulge-induced helix bending, other bulge sequences were identified which induce anomalously small bends. These sequences were isolated from the mixture of five-nucleotide bulges based on their electrophoretic mobilities. The mixture was separated by gel electrophoresis, and a slice of gel was excised at the same mobility as the duplex with a bulge of five uridines. The bulged strand of the RNA from this gel slice was reverse-transcribed, and the cDNA was amplified. The amplified DNA was transcribed, and the resulting RNA was annealed to the complementary 66-nucleotide strand to reconstitute the selected bulged duplexes.

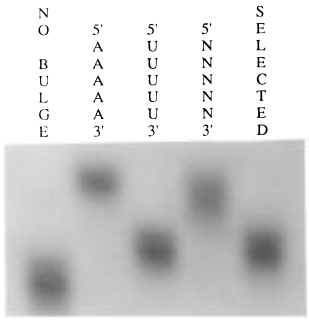

FIGURE 5: Electrophoretic analysis of bulged duplexes selected for anomalously high electrophoretic mobility from a mixture of all five-nucleotide bulges. Electrophoresis conditions were as in Figure 3.

Table 1: Five-Nucleotide Bulge Sequences Identified in Populations of Bulged Duplexes Selected for Anomalously High Electrophoretic Mobility

five pyrimidines	5′UCUCU3′
	5′UUCUU3′
	5′CUUUC3′
	5′UUCUC3′
four pyrimidines	5′UCCAC3′
	5′UUGUU3′
three pyrimidines	5′GGUUU3′ [a]
	5′GAUUC3′
	5′UUGGC3′
	5′UAAUC3′
	5′CCCAA3′
two pyrimidines	5′UAAAU3′

[a] Bulge sequence found in 2 of 13 clones sequenced.

The resulting mixture of bulged duplexes was analyzed by electrophoresis, and the result is shown in Figure 5. The selected mixture of bulge sequences is enriched in high-mobility species, compared with the mixture of all five-nucleotide bulges. There is a high degree of enrichment of bulged duplexes having the electrophoretic mobility of the duplex containing a central bulge of five uridines.

The amplified cDNA from the selection was cloned, and 13 individual clones were sequenced. The bulge sequences identified are listed in Table 1. They are generally pyrimidine-rich, containing three, four, or five pyrimidines. Of the sequences, one (5′GGUUU3′) was represented twice in the limited sample sequenced. One purine-rich bulge sequence (5′UAAAU3′) was identified.

The electrophoretic behaviors of duplexes interrupted by specific selected bulge sequences were examined (Figure 6). Each of the five selected bulge sequences analyzed induces a smaller bend in the duplex than the $5'A_5 3'$ bulge. Of these, the lowest mobility duplex is interrupted by 5′GGUUU3′. The relative mobilities of the three duplexes interrupted by all-pyrimidine bulges are similar but not identical to each other and to the relative mobility of the duplex interrupted by the $5'U_5 3'$ bulge.

The duplex containing the 5′UAAAU3′ bulge, the one purine-rich bulge identified in the selection, has an electrophoretic mobility similar to the duplexes containing all-pyrimidine bulges. To account for this exception, we considered the possibility that the secondary structure of this bulge is exceptional. For example, the 5′UA3′ in the bulge can be imagined to displace the the U and A of the flanking Watson−Crick base pairs (see sequence in Figure 2),

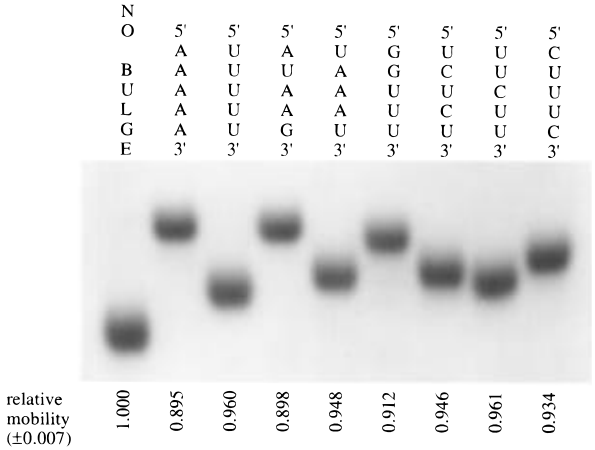

FIGURE 6: Electrophoretic mobility of RNA duplexes containing discrete bulge sequences identified in the mixture selected for high mobility. Electrophoresis conditions were as in Figure 3.

resulting in two bulge loops of two and three nucleotides separated by two Watson−Crick base pairs. We tested that hypothesis by preparing the bulged duplex with the U·A base pair two nucleotides from the bulge changed to an A·U base pair, leaving all other nucleotides the same. Though this change was expected to eliminate the divided bulge structure suggested, the relative electrophoretic mobility of the resulting bulged duplex was within experimental error of the electrophoretic mobility of the bulged duplex with the original flanking sequence (data not shown).

Quantitation of Bend Angles. Quantitative determination of bend angles cannot be made from electrophoretic mobility data alone. However, Zacharias and Hagerman (1995) have shown that electrophoretic mobility data can be calibrated with bend angles determined by the method of transient electric birefringence. We have estimated the bend angles of bulged duplexes by calibrating our electrophoretic experiment with bend angles determined by Zacharias and Hagerman. We prepared a series of 66 base pair calibration duplexes in which the 8 central base pairs were the same as the 8 central base pairs of the duplexes studied by Zacharias and Hagerman. This substitution was made to control for the effect of flanking sequence on the bend angle induced by a bulge. The remaining base pairs were identical to the duplexes for which we intended to estimate bend angles. The calibration duplexes were interrupted by adenosine and uridine bulge sequences for which bend angles have been determined. The electrophoretic mobilities were measured for these duplexes at 4 °C in a gel containing Tris−borate buffer and 5 mM $MgCl_2$ as were the electrophoretic mobilities of the bulged duplexes for which bend angles were to be estimated. Relative electrophoretic mobilities measured at 4 °C in Tris−borate magnesium buffer were within experimental error of relative electrophoretic mobilities measured at room temperature in Tris−HEPES magnesium buffer.

The electrophoretic mobilities and reported bend angles for the calibration duplexes are shown in Table 2, and the relative electrophoretic mobilities, μ_{rel}, are plotted *versus* bend angle, θ, in Figure 7. The relative mobilities can be fit to bend angles by a quadratic or quartic equation (Zacharias & Hagerman, 1995), but a function of the form $\mu_{rel} = \cos b\theta$, where b is an adjustable parameter, fits the data well with a least-squares fit to a value of 0.4 for b.

Table 2: Relative Electrophoretic Mobilities and Bend Angles of Duplexes for Electrophoretic Calibration

bulge sequence[a]	relative mobility (±0.007)[b]	bend angle (deg)[c]
A_2	0.987	31 ± 4
A_3	0.954	58 ± 3
A_5	0.870	70 ± 3
A_6	0.828	83 ± 4
U_6	0.931	43 ± 3
no bulge (control)	1.000	0

[a] Bulges interrupted a duplex in which the bulged strand had the sequence 5′-GGATCC GAAGACCTCAGTCTTGTCCAGCTGAGC-X_n-GCTCAGCTACCGTACTGCAAGAA GCTTTGTCCC-3′, where underlined bases are the same as those flanking the bulges in the study of Zacharias and Hagerman (1995), X_n represents the bulged nucleotides, and the other strand was the Watson−Crick complement to all but the bulged nucleotides. [b] Relative electrophoretic mobilities were determined as described under Materials and Methods using 10% nondenaturing polyacrylamide gels, 19:1 cross-linking, with 45 mM Tris−borate, pH 8.0, 5 mM $MgCl_2$ run at 4 °C. Reported mobilities are from one experiment. The uncertainty of ±0.007 was estimated from the range of mobilities obtained for bulged duplexes in multiple experiments. [c] Bend angles were as determined by Zacharias and Hagerman (1995). They are defined as the supplements of the small interhelix angles. Thus, a bend angle of 0° describes a linear molecule.

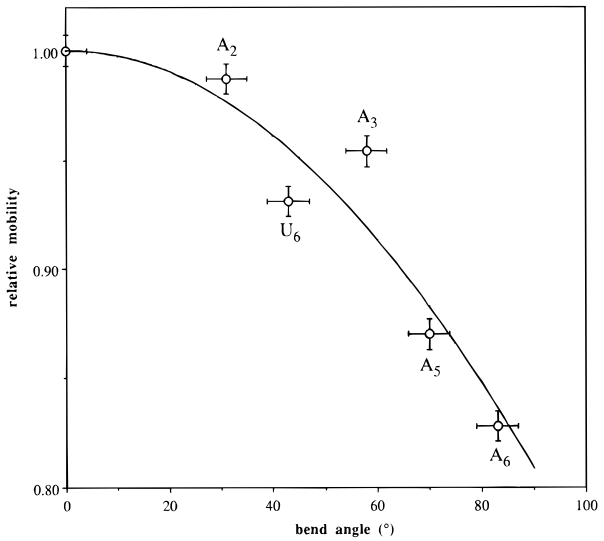

FIGURE 7: Relative electrophoretic mobility *vs* bend angle. Electrophoretic mobilities were determined at 4 °C in 45 mM Tris−borate buffer, pH 8.0, containing 5 mM $MgCl_2$ for the bulged and control duplexes described in Table 2. The curve drawn through the data is a least-squares fit of the function $\mu_{rel} = \cos b\theta$, where $b = 0.4$.

Using this function, bend angles were estimated for duplexes interrupted by various five-nucleotide bulges (Table 3). The angles estimated for the 5′AAAAA3′ and 5′UUUUU3′ bulges, 69° and 43°, respectively, are within experimental error of the bend angles determined by Zacharias and Hagerman for those bulge sequences, despite the difference in flanking sequences. The bulge from the self-splicing intron induces a bend of approximately 69°.

The largest fraction of the mixture of all possible five-nucleotide bulges also migrated at a mobility corresponding to a bend angle of approximately 69°. Eighty-five percent of the bulge mixture migrated with mobilities that correspond to bend angles of 60−80°. This estimate of the range of angles induced by 85% of the five-nucleotide bulge sequences is an upper limit, because the width of the band due to a bulged duplex of a single sequence is significant

Table 3: Relative Electrophoretic Mobilities and Estimated Bend Angles of Duplexes Interrupted by Various Five-Nucleotide Bulges

bulge sequence[a]	relative mobility (±0.007)[b]	bend angle (±5°)[c]
A_5	0.886	69
U_5	0.956	43
5′AUAAG3′	0.886	69
5′GGUUU3′	0.891	67
5′UAAAU3′	0.939	51
5′NNNNN3′ [d]	0.885	69
no bulge (control)	1.000	0

[a] Bulges interrupted the duplex shown in Figure 2. [b] Relative electrophoretic mobilities were determined as described under Materials and Methods using 10% nondenaturing polyacrylamide gels, 19:1 cross-linking, with 45 mM Tris−borate, pH 8.0, 5 mM $MgCl_2$ run at 4 °C. [c] Bend angles were estimated from the function $\theta = (\cos^{-1} \mu_{rel})/0.4$, where μ_{rel} is the relative mobility and θ is the bend angle defined in Table 2. [d] Peak maximum for the mixture of all five-nucleotide bulges.

with respect to the range of electrophoretic mobilities observed. Accounting for this band width, 65−75° is a more realistic estimate of the range of bend angles induced by the majority (>85%) of five-nucleotide bulge loop sequences. It is noteworthy that a purine-rich bulge, 5′UAAAU3′, induces a bend smaller than this range (51°), whereas the bend induced by a pyrimidine-rich bulge, 5′GGUUU3′ (67°), lies within this range. Although the duplex containing the bulge 5′GGUUU3′ is reproducibly of higher electrophoretic mobility than the duplexes containing the bulges 5′AUAAG3′ and 5′AAAAA3′, the bend angles estimated for those bulges are within experimental error of each other (Table 3). This result is consistent with the observed relationship between electrophoretic mobility and bend angle, in which electrophoretic mobility becomes increasingly sensitive to small changes in bend angle as bending increases from 30° to 85°.

DISCUSSION

Helix Bending by the Five-Nucleotide Bulge from the Group I Introns. We have shown that the five-nucleotide bulge from the group I intron of *T. thermophila* induces a bend between the flanking double helices. Interruption of a double helix (with the same flanking sequence context as the intron) by the five-nucleotide bulge results in electrophoretic retardation of the helix, consistent with a directional bend induced by the bulge. However, our results fail to support the hypothesis that the sequence of the bulge is conserved to control the extent of that bend.

Electrophoretic retardation alone does not prove the existence of a directional bend, because isotropic flexibility can also contribute to electrophoretic retardation (Mills *et al.*, 1994; Kahn *et al.*,1994). However, other five-nucleotide bulges have been shown by phasing analysis and permutation analysis to induce directional bends (Bhattacharyya & Lilley, 1989; Bhattacharyya *et al.*, 1990; Hsieh & Griffith, 1989; Rice & Crothers, 1989; Riordan *et al.*, 1992; Tang & Draper, 1994). Furthermore, among these bulges is the bulge of five adenosines which results in the same amount of electrophoretic retardation as the bulge from the intron. Thus, a directional bend at the site of the bulge from the intron is implicated.

The extent of bending induced by the bulge from the intron is not strongly dependent on the conserved sequence. Mutation of each of three highly conserved adenosines in the bulge to uridines slightly decreased the electrophoretic mobility of the duplex. Based on the relationship that we have observed between average bend angle and electrophoretic mobility (Figure 7), the difference in electrophoretic mobility that results from mutation of the conserved residues corresponds to a very small increase in bending, on the order of a few degrees. Furthermore, electrophoretic analysis of a mixture of duplexes in which all five-nucleotide bulge sequences are represented reveals that *most* five-nucleotide bulge sequences induce a bend of nearly the same average angle as the bulge sequence that is conserved. Many bulge sequences could effectively replace the conserved bulge if its only structural role were to fix the orientation of the flanking helices, and phylogenetic conservation would not be expected.

An obvious role for the P5 bulge is formation of direct tertiary interactions. Tertiary contacts have been proposed for the bulge loop based on chemical probing and mutational studies (Flor *et al.*, 1989; Murphy & Cech, 1993, 1994), and conservation of base-specific or conformation-specific tertiary interactions could account for conservation of bulge sequence. The global tertiary structure would nonetheless need to accommodate helix bending induced by the bulge.

The electrophoretic assay we have used provides information about the average bend angle induced by the bulge in the double helix. It does not provide information about the energetic penalty for deforming the bulge to a conformation that induces a different average bend angle. Within the context of the folded structural domain, favorable tertiary interactions could compensate for such a penalty. The thermodynamics of deforming a bulge-induced bend are likely to have a sequence dependence that is separate from the sequence dependence of the average bend induced in the isolated duplex. In that context, it is noteworthy that the secondary structure of the isolated bulge loop determined by nuclear magnetic resonance spectroscopy is different from the secondary structure of the bulge loop in the intron predicted by comparative sequence analysis (Landry, 1995; Luebke *et al.*, manuscript in preparation; Michel & Westhof, 1990).

Distribution of Bulge Sequences among Bend Angles. Five-nucleotide uridine and adenosine bulges are near the extremes of bending for bulges of this size. No five-nucleotide bulge sequences that result in electrophoretic mobilities greater than that of the duplex containing the 5′$U_5$3′ bulge are evident in the mixture of bulges. Of the bulges selected based on high electrophoretic mobility that were analyzed individually, all five induced a greater gel retardation than the 5′$U_5$3′ bulge. At the other end of the range of electrophoretic mobilities, the mixture of bulged duplexes extends to slightly lower mobilities than the 5′$A_5$3′ bulged duplex. The single adenosine to uridine mutants of the bulge from the group I intron exemplify bulge sequences that induce a slightly greater electrophoretic retardation than the 5′$A_5$3′ bulge. These lowest electrophoretic mobilities are the most sensitive to small changes in bend angle, and we estimate that the lowest electrophoretic mobility represented in the mixture of bulge sequences corresponds to a bend angle of approximately 75°.

Thus, for five-nucleotide bulges in the presence of millimolar concentrations of magnesium ion, a range of induced bend angles of approximately 32° (from 43° to 75°) can be obtained. Although the induced bend angle for a given bulge sequence is somewhat dependent on the identities

of flanking base pairs, the close agreement between the bend angles induced by 5′$U_5$3′ and 5′$A_5$3′ bulges in the different sequence contexts of our study and the study of Zacharias and Hagerman (1995) suggests that the observed range of angles is general for bulges of this size. We cannot rule out the possibility of very rare bulge sequences that induce bends outside of the range detected in the bulge mixture. One bulge sequence out of 1024 (4^5, the total number of possible five-nucleotide bulges) would not be detectable within the mixture if present at an isolated position on the gel. However, we were unable to reverse-transcribe and amplify any RNA that formed bulged duplexes migrating outside of the apparent range of electrophoretic mobilities.

Although the full range of bend angles induced by all five-nucleotide bulge loops extends between 43° and 75°, most five-nucleotide bulge sequences induce bends that fall within a much narrower range of angles, estimated to be between 65° and 75° for bulges in the context of these experiments. We observe an even narrower range of electrophoretic mobilities for five-nucleotide bulges in the absence of magnesium ion than in its presence. This observation is supported by the previously reported similarity of bend angles induced by adenosine and uridine bulges in the absence of magnesium ion (Zacharias & Hagerman, 1995).

Because the range of bend angles induced by most five-nucleotide bulges is much narrower than the full range of bend angles, the anomalous bulge sequences that induce bends smaller than the most populated range are of particular interest. In studies of single-nucleotide DNA bulges, it was found that purine nucleotides induce a sharper bend than pyrimidine nucleotides (Wang & Griffith, 1991). Similarly, in RNA bulge loops, a bulge of from one to seven adenosines induces a sharper bend than the same number of uridines (Bhattacharyya *et al.*, 1990; Zacharias & Hagerman, 1995). Based on these observations, the anomalous five-nucleotide bulge sequences might be expected to be pyrimidine-rich. In fact, bulges that are entirely or predominantly composed of pyrimidines are prevalent among the sequences that we identified. Of 13 sequences, all but 1 were pyrimidine-rich. The noteworthy exception is the purine-rich bulge sequence 5′UAAAU3′, which induces a bend that is below the range of most of the bend angles. This bulge induces a smaller bend than the pyrimidine-rich sequence 5′GGUUU3′. In general, the relative bend angles induced by bulge loops cannot be predicted based on *a priori* inspection of sequence composition. Information about the internal structure of the bulge and the basis of bending in that structure is necessary.

Structural Basis of Sequence Dependence of Bulge-Induced Helix Bending. The relationship between the bend angle induced by an RNA bulge loop and the conformation of that loop is unknown. The experiments described here have demonstrated two consequences of that relationship: (1) a relatively narrow distribution of bend angles among bulge sequences, and (2) the existence of a small number of exceptional bulge sequences which induce anomalously small bends. The exceptional sequences are generally, but not universally, pyrimidine-rich.

The narrow distribution of bend angles among bulge sequences might reflect insensitivity of the bend angle to conformational variation for a given bulge size or a limited conformational diversity among five-nucleotide bulges. Though little experimental information is available, the conformational diversity of bulge loops might be expected to increase with an increasing number of loop nucleotides. Crystallographic and spectroscopic data for one-nucleotide bulges of RNA (van den Hoogen *et al*., 1988*b*) and DNA (Hare *et al.*, 1986; Joshua-Tor *et al.*, 1988; Kalnik *et al.*, 1989, 1990; Miller *et al.*, 1988; Morden *et al.*, 1983, 1990; Nikonowicz *et al.*, 1989; van den Hoogen *et al.*, 1988a; Woodson & Crothers, 1988a,b; Woodson & Crothers, 1989) have indicated that the extra base can be intercalated into the helical stack or extruded from the helix. The equilibrium occupancy of the two conformations depends upon the identities of the bulged nucleotide and the flanking base pairs, the temperature, and the buffer composition. In larger bulge loops, a greater structural diversity is expected because stacking and hydrogen bonding interactions are possible within the bulge as well as between bulge nucleotides and the flanking double helices.

NMR studies of three-nucleotide bulges have revealed that the bulged nucleotides are intercalated into the helical stack in DNA bulges of 5′ATA3′ (Rosen *et al.*, 1992) and 5′$A_3$3′ (Aboul-ela *et al.*, 1993) and the RNA bulge of 5′UCU3′ from the TAR element of human immunodeficiency virus RNA (Puglisi *et al.*, 1992). However, the spectroscopically determined solution conformation of the five-nucleotide bulge from the group I intron from *T. thermophila* cannot be described simply as intercalation or extrusion of all of the bulged nucleotides (Landry, 1995; Luebke *et al.*, manuscript in preparation). Base stacking within the loop is discontinuous. The extent to which the points of stacking discontinuity depend on the sequence is not discernible from a single structure.

The existence of exceptional bulge sequences which induce anomalously small bends indicates some sensitivity of the induced bend angle to conformational variation. However, the nature of the conformational features resulting in an anomalously small induced bend is not obvious from the bulge sequences identified. Extrusion of extra bases from the helix, allowing the flanking base pairs (or other bulge nucleotides) to stack together, is expected to allow partial straightening of the helix. The preponderance of pyrimidine-rich sequences among the anomalous bulges is consistent with this expectation and the lower propensity for pyrimidine nucleotides to stack relative to purine nucleotides. On the other hand, the small bend induced by the purine-rich bulge identified indicates that the propensity for internal base stacking is not the sole determinant of the bulge-induced bend angle.

ACKNOWLEDGMENT

We thank Professor Paul Hagerman for many useful and interesting discussions. We also thank Mr. David Koh for chemical synthesis of DNA and Ms. Barbara Dengler for managing the laboratory.

REFERENCES

Aboul-ela, F., Murchie, A. I. H., Homans, S. W., & Lilley, D. M. J. (1993) *J. Mol. Biol. 229*, 173−188.

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

Bhattacharyya, A., Murchie, A. I. H., & Lilley, D. M. J. (1990) *Nature 343*, 484−487.

Collins, R. A. (1988) *Nucleic Acids Res. 16*, 2705−2714.

Crothers, D. M., Haran, T. E., & Nadeau, J. G. (1990) *J. Biol. Chem. 265*, 7093−7096.

Davanloo, P., Rosenberg, A. H., Dunn, J. J., & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035−2039.
Flor, P. J., Flanegan, J. B., & Cech, T. R. (1989) *EMBO J. 8*, 3391−3399.
Gohlke, C., Murchie, A. I. H., Lilley, D. M. J., & Clegg, R. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11660−11664.
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. 225*, 397−431.
Kahn, J. D., Yun, E., & Crothers, D. M. (1994) *Nature 368*, 163−166.
Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989) *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.
Landry, S. M. (1995) Ph.D. Thesis, University of California, Berkeley.
Lilley, D. M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7140−7142.
Michel, F., & Cummings, D. J. (1985) *Curr. Genet. 10*, 69−79.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 85−86.
Mills, J. B., Cooper, J. P., & Hagerman, P. J. (1994) *Biochemistry 33*, 1797−1803.
Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I. (1983) *Biochemistry 22*, 5557−5563.
Morden, K. M., Gunn, B. M., & Maskos, K. (1990) *Biochemistry 29*, 8835−8845.
Murphy, F. L., & Cech, T. R. (1993) *Biochemistry 32*, 5291−5300.
Murphy, F. L., & Cech, T. R. (1994) *J. Mol. Biol. 236*, 49−63.
Myers, T. W., & Gelfand, D. H. (1991) *Biochemistry 30*, 7661−7666.
Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714−8725.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76−80.
Rice, J. A., & Crothers, D. M. (1989) *Biochemistry 28*, 4512−4516.
Riordan, F. A., Bhattacharyya, A., McAteer, S., & Lilley, D. M. J. (1992) *J. Mol. Biol. 226*, 305−310.
Rosen, M. A., Shapiro, L., & Patel, D. J. (1992) *Biochemistry 31*, 4015−4026.
Sanger, F., Nicklen, S., Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Tang, R. S., & Draper, D. E. (1990) *Biochemistry 29*, 5232−5237.
Tang, R. S., & Draper, D. E. (1994) *Nucleic Acids Res. 22*, 835−841.
Van den Hoogen, Y. T., van Beuzekom, A. A., de Vroom, E., van den Elst, G. A., van Boom, J. H., & Altona, C. (1988a) *Nucleic Acids Res. 16*, 5013−5030.
Van den Hoogen, Y. T., van Beuzekom, A. A., de Vroom, E., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988*b*) *Nucleic Acids Res. 16*, 2971−2986.
Wang, Y.-H., & Griffith, J. (1991) *Biochemistry 30*, 1358−1363.
Wang, Y.-H., Barker, P., & Griffith, J. (1992) *J. Biol. Chem. 267*, 4911−4915.
Woodson, S. A., & Crothers, D. M. (1988a) *Biochemistry 27*, 436−445.
Woodson, S. A., & Crothers, D. M. (1988b) *Biochemistry 27*, 3130−3141.
Woodson, S. A., & Crothers, D. M. (1989) *Biopolymers 28*, 1149−1177.
Wyatt, J. R., Chastain, M., & Puglisi, J. D. (1991) *BioTechniques 11*, 764−769.
Zacharias, M., & Hagerman, P. J. (1995) *J. Mol. Biol. 247*, 486−500.

BI960914R