Nucleic Acids Research, 1994, Vol. 22, No. 25 5753–5760

# Sequence determinants of DNA bending in the *ilvlH* promoter and regulatory region of *Escherichia coli*

Qing Wang+, Fred G.Albert[2], Daniel J.Fitzgerald[2], Joseph M.Calvo[1] and John N.Anderson[2,*]
Sections of Genetics and Development and [1]Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 and [2]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA

Received April 19, 1994; Revised and Accepted September 29, 1994

## ABSTRACT

**Previous studies have shown that the promoter/regulatory region of the *ilvlH* operon displays intrinsic curvature, with the bend center located at position −120 relative to the transcription start site. In this report, a 57 bp sequence spanning the bend center was mutagenized *in vitro* in order to study the relationship between nucleotide sequence and curvature measured by electrophoresis. The strategy used for analyzing the results consisted of determining the strengths of the relationships between electrophoretic anomaly and predicted curvature calculated by computer programs that differ in wedge angle composition. The results revealed that programs which assume that bending occurs only at AA/TT display good predictive value, with correlation coefficients between electrophoretic anomaly and predicted curvature as high as 0.93. In contrast, a program which assumes that bending occurs at all 16 dinucleotide steps exhibited lower predictive value, while there were no significant relationships between the experimental data and curvature calculated by a program that was based on all non-AA/TT wedge values. These results show that the complete wedge model which incorporates values for all dinucleotide steps does not adequately describe the electrophoretic data in this report.**

## INTRODUCTION

The folding of DNA into protein-containing particles is a universal process within cells. Examples include the assembly of nucleosomes in chromatin and the formation of multi-protein complexes at sites of initiation of transcription, replication and site-specific recombination. It is logical to assume that the DNA which wraps around protein complexes would contain sequence elements that facilitate the folding process. Intrinsic curvature of DNA is thought to be one of these sequence features (1–5). Intrinsically curved DNAs are most often identified by their anomalously slow electrophoretic mobility through polycrylamide gels. This technique has been used to show that bent helices are characteristic of some promoter regions (6–11), DNA integration sites (12) and origins of replication (13–17). The importance of bending to function is supported by deletion analysis and by the observations that synthetic bent DNA can substitute functionally for natural bending elements in control regions for replication (18), transcription (19,20) and integration (21,22).

The *ilvlH* operon of *Escherichia coli* encodes acetohydroxyacid synthase III, one of the three isoenzymes catalyzing the first step common to the biosynthesis of isoleucine, leucine and valine (23). Transcription of the *ilvlH* operon is activated by the leucine-responsive regulatory protein (Lrp) and this activation is inhibited by leucine (24,25). The promoter region of the *ilvlH* operon has been defined by nuclease mapping and by mutational analysis (25). The 331 bp region upstream of the transcription start site is sufficient for both activation of transcription and leucine-mediated repression (25). Lrp binds to six sites within this region and the binding is required for transcriptional activation of the *ilvlH* operon *in vivo* and *in vitro* (26,27).

Recent studies have shown that the region upstream of the *ilvlH* promoter displays sequence-dependent curvature with the bend center located at position −130 relative to the transcription start site (11). The cooperative binding of Lrp to the region induces further bending and may facilitate the generation of a higher-order nucleoprotein structure (11). In this report, the nucleotide sequence determinants of curvature were examined by comparing electrophoretic mobilities of 57 randomly mutagenized sequences from this region to the curvature predicted by computer programs which differ in dinucleotide wedge contributions. The results revealed that programs which assume bending occurs only at AA/TT exhibit higher predictive value than those which incorporate all 16 dinucleotide wedge angles.

## MATERIALS AND METHODS

### Bacterial strains

*E.coli* strain CJ236 [(*dut*-1 *ung*-1 *thi*-1 relA1/pCJ105 ($Cm^r$)] (28) was used in site-directed mutagenisis experiments to make

*To whom correspondence should be addressed

+Present address: Department of Human Genetics, University of Utah, Health Sciences Center, Salt Lake City, UT 84112, USA

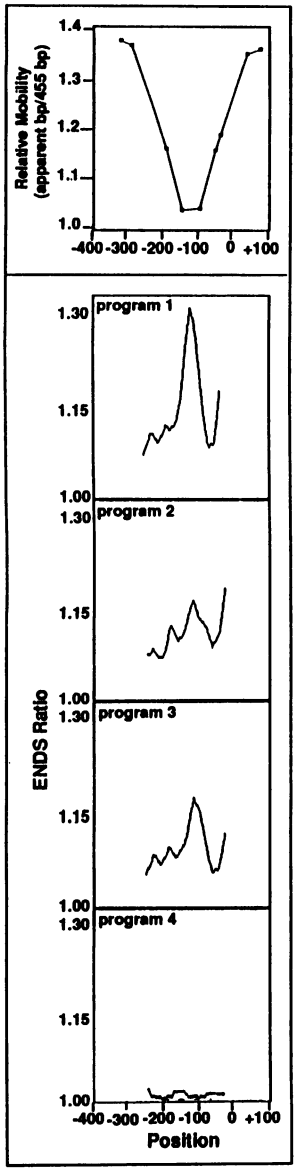



**Figure 1.** Experimental and computer mapping of the center of bending. (**Top** panel) A dimer containing the *ilvlH* promoter/regulatory region (from −345 to +110) was digested by restriction enzymes which cleave only once within the monomer. The products were analyzed by electrophoresis through a 6% polyacrylamide gel and the mobilities relative to the actual size of the monomer are plotted against position relative to the transcription start site. Further details are provided in Wang and Calvo (11). (**Bottom** panels) ENDS ratios as a function of sequence position were computed by the 4 programs at a window width of 120 bp.

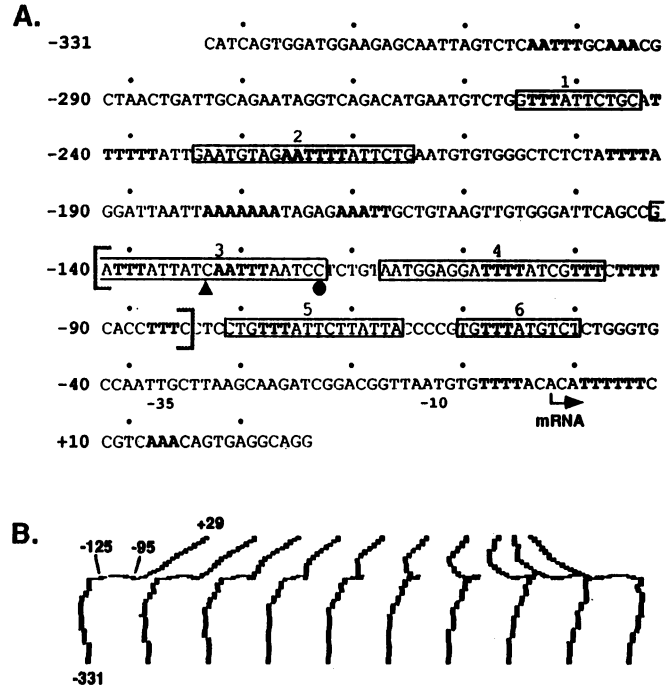



**Figure 2.** (**A**) Sequence of the *ilvlH* promoter/regulatory region. Indicated along the sequence are 6 Lrp binding sites (27) denoted by boxes, the position of the bend centers from Figure 1 determined by electrophoresis (●) or by programs 1 and 3 (▲), and bending elements of the form A ≥ 3, T ≥ 3 and $A_nT_m$ where n ≥ 2, m ≥ 2 and n + m ≥ 5. The brackets indicate the 57 bp mutagenized region described in Figure 4. Note the phase shift in the bending sequences that occurs between positions −125 and −95. These positions are indicated on the helix projections of the sequence (**B**) that were generated by program 1. Each successive projection is rotated by 18°.

uracil-containing single-stranded DNA of phage mp18QW156 (27). *E.coli* JM101 [*supE thi* Δ(*lac pro*/F′ *traD*36, *proAB lacZ* M15] was used as the host for phage M13mp18 and its derivatives.

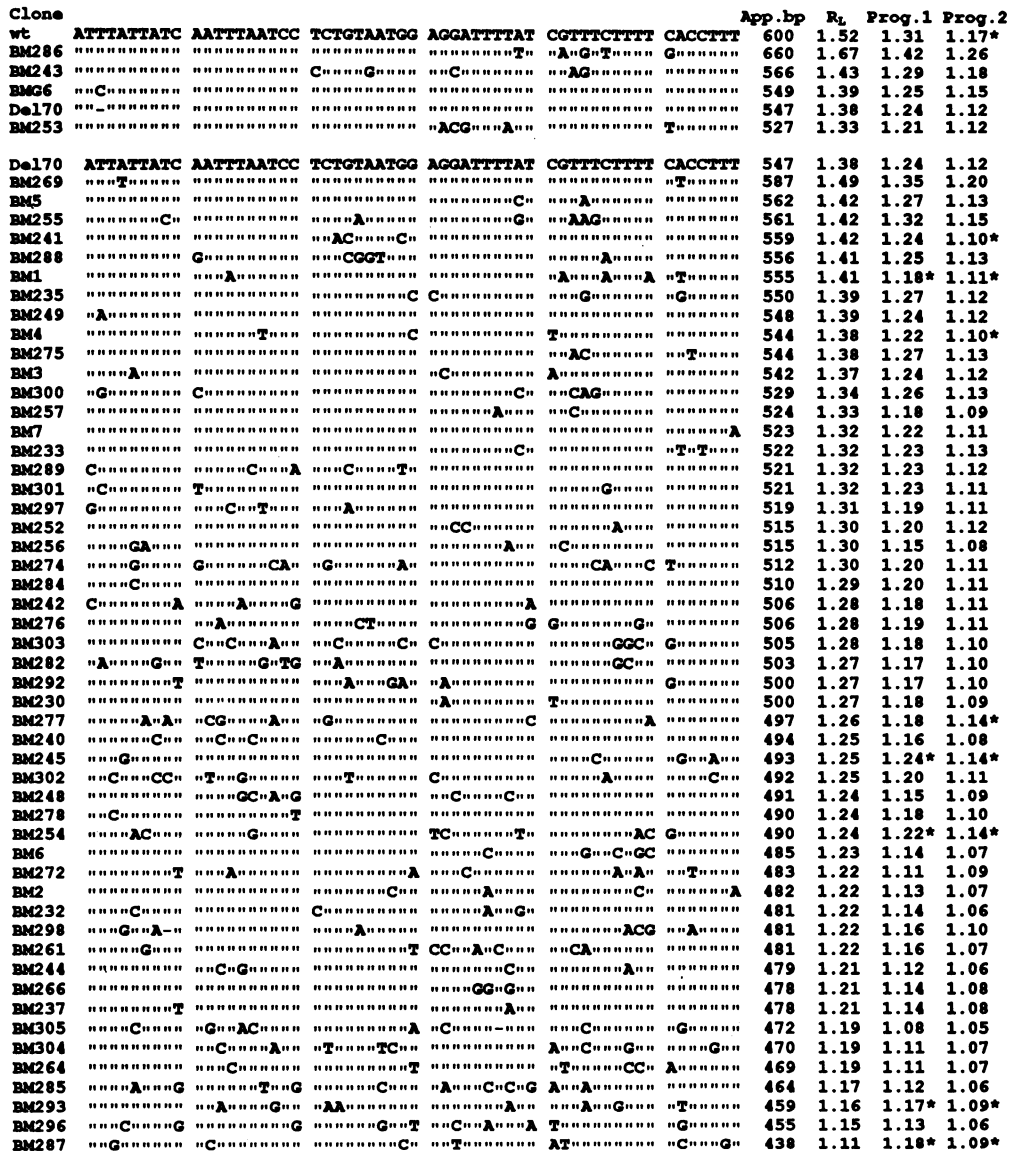
## General DNA techniques and electrophoresis

Standard molecular biological techniques were used for DNA manipulations (29). Both single-stranded and double-stranded DNA sequencing were performed by the dideoxy chain termination method (30) using the Sequenase kit from United States Biochemicals Inc. Oligonucleotides were synthesized by the Cornell University Biotechnology Program Oligonucleotide Synthesis Facility.

DNA fragments containing the *ilvlH* promoter/regulatory region were excised with appropriate restriction enzymes and analyzed by both native polyacrylamide gel (29:1 acrylamide/bisacrylamide) and agarose gel electrophoresis. Electrophoresis was carried out at 8−10 V/cm in 45 mM Tris−borate, 1.25 mM EDTA (pH 8.6) at 4°C. The apparent size of fragments was estimated by comparing their mobilities to a set of size standards prepared by cutting plasmid pBR322 with *Msp*I. Fragments were detected either by ethidium bromide staining (Figure 1) or by autoradiography following labeling with [$^{32}$P]dATP and reverse transcriptase (Figure 3). Relative length ($R_L$) is defined as the ratio of the apparent length to the actual length . Mapping of the center of bending was performed by using the circular permutaion gel analysis method of Wu and Crothers (31) as described previously (11).

## Mutagenesis

Most of the mutations were made by the so-called 'dirty bottle' method within a 57 bp region which is located at −83 to −140 relative to the transcription start site of the *ilvlH* operon (28,32). A 70mer having the following sequence was chemically synthesized: 5′-CAGGAGG[AAA GGTGAAAAGA AACG-ATAAAA TCCTCCATTA CAGAGGATTA AATTGATAAT AAAT]CGGCTG-3′. The synthesis was designed such that, for



| Clone | | | | | | | App.bp | $R_L$ | Prog.1 | Prog.2 |
|---|---|---|---|---|---|---|---|---|---|---|
| wt | ATTTATTATC | AATTTAATCC | TCTGTAATGG | AGGATTTTAT | CGTTTCTTTT | CACCTTT | 600 | 1.52 | 1.31 | 1.17* |
| BM286 | ·····T· | ····· | ····· | ·····T· | ·A·G·T····· | G····· | 660 | 1.67 | 1.42 | 1.26 |
| BM243 | ····· | ····· | C····G····· | ···C····· | ···AG····· | ····· | 566 | 1.43 | 1.29 | 1.18 |
| BMG6 | ···C····· | ····· | ····· | ····· | ····· | ····· | 549 | 1.39 | 1.25 | 1.15 |
| Del70 | ···-····· | ····· | ····· | ····· | ····· | ····· | 547 | 1.38 | 1.24 | 1.12 |
| BM253 | ····· | ····· | ····· | ·ACG···A·· | ····· | T····· | 527 | 1.33 | 1.21 | 1.12 |
| Del70 | ATTATTATC | AATTTAATCC | TCTGTAATGG | AGGATTTTAT | CGTTTCTTTT | CACCTTT | 547 | 1.38 | 1.24 | 1.12 |
| BM269 | ···T····· | ····· | ····· | ····· | ····· | ·T····· | 587 | 1.49 | 1.35 | 1.20 |
| BM5 | ····· | ····· | ····· | ·····C· | ···A····· | ····· | 562 | 1.42 | 1.27 | 1.13 |
| BM255 | ·····C· | ····· | ····A····· | ·····G· | ···AAG····· | ····· | 561 | 1.42 | 1.32 | 1.15 |
| BM241 | ····· | ····· | ···AC····C· | ····· | ····· | ····· | 559 | 1.42 | 1.24 | 1.10* |
| BM288 | ····· | G····· | ···CGGT··· | ····· | ·····A····· | ····· | 556 | 1.41 | 1.25 | 1.13 |
| BM1 | ····· | ···A····· | ····· | ····· | ·A···A···A | ·T····· | 555 | 1.41 | 1.18* | 1.11* |
| BM235 | ····· | ····· | ·····C | C····· | ···G····· | ·G····· | 550 | 1.39 | 1.27 | 1.12 |
| BM249 | ·A····· | ····· | ····· | ····· | ····· | ····· | 548 | 1.39 | 1.24 | 1.12 |
| BM4 | ····· | ·····T··· | ·····C | ····· | T····· | ····· | 544 | 1.38 | 1.22 | 1.10* |
| BM275 | ····· | ····· | ····· | ····· | ···AC····· | ···T····· | 544 | 1.38 | 1.27 | 1.13 |
| BM3 | ····A····· | ····· | ····· | ·C····· | A····· | ····· | 542 | 1.37 | 1.24 | 1.12 |
| BM300 | ·G····· | C····· | ····· | ·····C· | ···CAG····· | ····· | 529 | 1.34 | 1.26 | 1.13 |
| BM257 | ····· | ····· | ····· | ·····A··· | ···C····· | ····· | 524 | 1.33 | 1.18 | 1.09 |
| BM7 | ····· | ····· | ····· | ····· | ····· | ·····A | 523 | 1.32 | 1.22 | 1.11 |
| BM233 | ····· | ····· | ····· | ·····C· | ····· | ·T·T··· | 522 | 1.32 | 1.23 | 1.13 |
| BM289 | C····· | ·····C···A | ···C····T· | ····· | ····· | ····· | 521 | 1.32 | 1.23 | 1.12 |
| BM301 | ·C····· | T····· | ····· | ····· | ·····G···· | ····· | 521 | 1.32 | 1.23 | 1.11 |
| BM297 | G····· | ···C··T··· | ···A····· | ····· | ····· | ····· | 519 | 1.31 | 1.19 | 1.11 |
| BM252 | ····· | ····· | ····· | ···CC····· | ·····A···· | ····· | 515 | 1.30 | 1.20 | 1.12 |
| BM256 | ····GA···· | ····· | ····· | ·····A·· | ·C····· | ····· | 515 | 1.30 | 1.15 | 1.08 |
| BM274 | ····G····· | G·····CA· | ·G·····A· | ····· | ····CA···C | T····· | 512 | 1.30 | 1.20 | 1.11 |
| BM284 | ····C····· | ····· | ····· | ····· | ····· | ····· | 510 | 1.29 | 1.20 | 1.11 |
| BM242 | C·····A | ····A·····G | ····· | ·····A | ····· | ····· | 506 | 1.28 | 1.18 | 1.11 |
| BM276 | ····· | ···A····· | ····CT····· | ·····G | G·····G· | ····· | 506 | 1.28 | 1.19 | 1.11 |
| BM303 | ····· | C··C···A·· | ···C·····C· | C····· | ·····GGC· | G····· | 505 | 1.28 | 1.18 | 1.10 |
| BM282 | ·A····G·· | T·····G·TG | ···A····· | ····· | ·····GC·· | ····· | 503 | 1.27 | 1.17 | 1.10 |
| BM292 | ·····T | ····· | ···A···GA· | ·A····· | ····· | G····· | 500 | 1.27 | 1.17 | 1.10 |
| BM230 | ····· | ····· | ····· | ·A····· | T····· | ····· | 500 | 1.27 | 1.18 | 1.09 |
| BM277 | ·····A·A· | ·CG····A·· | ·G····· | ·····C | ·····A | ····· | 497 | 1.26 | 1.18 | 1.14* |
| BM240 | ·····C·· | ···C··C···· | ·····C···· | ····· | ····· | ····· | 494 | 1.25 | 1.16 | 1.08 |
| BM245 | ···G····· | ····· | ····· | ····· | ····C····· | ·G··A·· | 493 | 1.25 | 1.24* | 1.14* |
| BM302 | ···C···CC· | ·T··G····· | ···T····· | C····· | ·····A···· | ·····C·· | 492 | 1.25 | 1.20 | 1.11 |
| BM248 | ····· | ····GC·A·G | ····· | ···C····C·· | ····· | ····· | 491 | 1.24 | 1.15 | 1.09 |
| BM278 | ···C····· | ·····T | ····· | ····· | ····· | ····· | 490 | 1.24 | 1.18 | 1.10 |
| BM254 | ····AC···· | ·····G···· | ····· | TC·····T· | ·····AC | G····· | 490 | 1.24 | 1.22* | 1.14* |
| BM6 | ····· | ····· | ····· | ·····C···· | ····G···C·GC | ····· | 485 | 1.23 | 1.14 | 1.07 |
| BM272 | ·····T | ···A····· | ·····A | ···C····· | ·····A·A· | ···T···· | 483 | 1.22 | 1.11 | 1.09 |
| BM2 | ····· | ····· | ·····C·· | ·····A···· | ·····C· | ·····A | 482 | 1.22 | 1.13 | 1.07 |
| BM232 | ·····C···· | ····· | C····· | ·····A···G· | ····· | ····· | 481 | 1.22 | 1.14 | 1.06 |
| BM298 | ····G···A·· | ····· | ····A····· | ····· | ·····ACG | ···A···· | 481 | 1.22 | 1.16 | 1.10 |
| BM261 | ·····G···· | ····· | ·····T | CC···A·C···· | ···CA····· | ····· | 481 | 1.22 | 1.16 | 1.07 |
| BM244 | ····· | ···C·G····· | ····· | ·····C·· | ·····A·· | ····· | 479 | 1.21 | 1.12 | 1.06 |
| BM266 | ····· | ····· | ····· | ·····GG·G·· | ····· | ····· | 478 | 1.21 | 1.14 | 1.08 |
| BM237 | ·····T | ····· | ····· | ·····A·· | ····· | ····· | 478 | 1.21 | 1.14 | 1.08 |
| BM305 | ····C····· | ·G···AC···· | ·····A | ·C·····-···· | ···C····· | ·G····· | 472 | 1.19 | 1.08 | 1.05 |
| BM304 | ····· | ···C····A·· | ·T·····TC·· | ····· | A···C···G·· | ·····G·· | 470 | 1.19 | 1.11 | 1.07 |
| BM264 | ····· | ···C····· | ·····T | ····· | ·T·····CC· | A····· | 469 | 1.19 | 1.11 | 1.07 |
| BM285 | ·····A···G | ·····T···G | ·····C···· | ·A···C·C·G | A···A····· | ····· | 464 | 1.17 | 1.12 | 1.06 |
| BM293 | ····· | ···A····G·· | ·AA····· | ·····A·· | ···A···G···· | ·T····· | 459 | 1.16 | 1.17* | 1.09* |
| BM296 | ···C·····G | ·····G | ·····G···T | ···C···A···A | T····· | ·G····· | 455 | 1.15 | 1.13 | 1.06 |
| BM287 | ···G····· | ·C····· | ·····C· | ···T····· | AT····· | ·C····G· | 438 | 1.11 | 1.18* | 1.09* |

**Figure 3.** Characteristics of mutations that affect bending. Sequences shown represent the 57 bp regions from −140 to −84. These sequences are part of a longer 395 bp fragment that was analyzed by electrophoresis. For the two groups of mutants, comparisons are made to either the wild-type (wt) sequence or to Del70, which has a single T deletion. The apparent size (App. bp) of the 395 bp fragments and their relative lengths ($R_L$) are compared to ENDS ratio values computed by programs 1 and 2. ENDS ratios were calculated at a window width of 120 bp centered at position −130, which is the approximate center of bending. The ENDS ratios indicated by the stars were the most deviant values expected from the relationships given in the top two panels of Figure 4.

each nucleotide position within the 57 base pair region denoted by brackets, 95% of the molecules would have the indicated base and 5% of the molecules would have one of the three other bases. With this design, an average of 3 mutations were expected per molecule. The actual average turned out to be about 6. The mutated sequences were annealed to the uracil-containing single-stranded mp18QW156 and the product was made uniformly double-stranded by treatment with T4 DNA polymerase. Transformation into *E.coli* JM101 eliminated the wild-type, uracil-containing strand. The efficiency of the mutagenesis was about 75%, which made screening via sequencing possible. For each mutant analyzed, double-stranded DNA was isolated and a 395 bp region containing the *ilvIH* promoter/regulatory region was excised by cutting with *EcoRI* and *Hin*dIII. All but one (BMG6) of the mutations analyzed in Figure 3 were isolated by the procedure described above.

For reasons that we do not understand, all but 3 of the mutant constructs that were produced by this approach had a deletion of an A base near the 3′ end of the oligomer at position −138. To create a suitable parental derivative for comparative studies, we used site-directed mutagenesis to create Del70, which has the −138 T deleted and no other mutations. The primer used for this latter construction was 5′-AATTGATAATAAT-CGGCTGAAT-3′.

Mutant BMG6 was isolated after the chemical mutagenesis procedure as described by Myers *et al.* (33). Phage mp18QW156

contains region −331 to +30 from the *ilvIH* operon cloned into the *Sma*I site within the polylinker region of phage M13mp18. Single-stranded DNA of mp18QW156 was treated with nitrous acid, hydrazine and formic acid and the region from the *ilvIH* operon was made double stranded by primer extension from the universal M13 sequencing primer (New England Biolab). A fragment containing the *ilvIH* promoter/regulatory region was excised using *Eco*RI and *Hin*dIII and fractionated through a 6% polyacrylamide gel. Fragments containing mutations that affected the extent of bending migrated faster or slower than the wild-type fragment. DNA fragments within regions of the gel above and below the wild-type fragment were eluted and subcloned into phage M13mp18. Double-stranded DNA was prepared from infected cells isolated after transformation. Following digestion with *Eco*RI and *Hin*dIII, samples were analyzed at 4°C on 6% polyacrylamide gels.

### Computer analysis

The program for calculating DNA bending from nucleotide sequence has been described previously (15). The program incorporates the dinucleotide twist angles reported by Kabsch *et al.* (34) using the parameters of the wedge model for bent DNA (35). Four derivative programs are compared in this work. Program 1 assumes bending occurs only at AA/TT and uses a wedge angle of 8.7°, as estimated by Ulanovsky *et al.* (36). Program 2 uses the 16 wedge angles estimated by Bolshoy *et al.* (37). Program 3 assumes bending occurs only at AA/TT and uses a wedge angle of 7.2°, as given in Bolshoy, while program 4 incorporates all non-AA/TT wedge angles from Bolshoy *et al.* (37). The values of 8.7° and 7.2° for the AA/TT step were determined from cyclization and electrophoretic mobility experiments carried out at 12°C (36) and room temperature (37).

The ENDS ratio is used as the index of predicted bending (15). The ENDS ratio is defined as the ratio of the contour length of a nucleotide segment to the shortest distance between the ends of the segment. ENDS ratios were computed at the window widths indicated in the text and at a window step of 10 nucleotides. Regression plots of $R_L$ vs. ENDS ratios for the data mentioned in the text and in Figure 5 are available upon request.

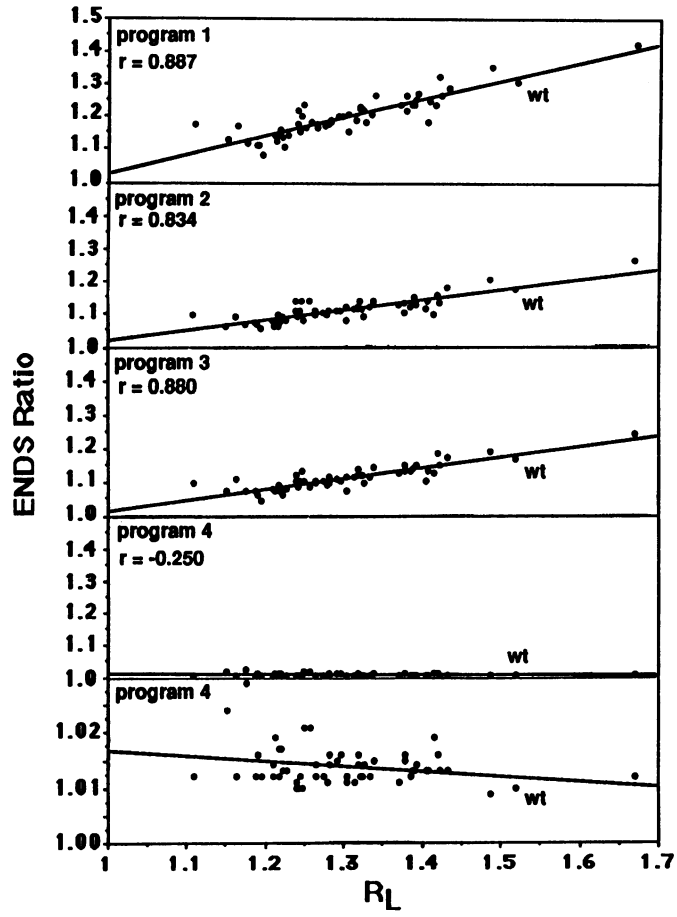



**Figure 4.** Relationship between electrophoretic anomaly and predicted curvature at the bend center. $R_L$ values from Figure 3 are plotted against ENDS ratios computed by the 4 programs. Correlation coefficients (*r*) of the relationships are given. An expanded ENDS ratio scale is used in the bottom panel for program 4. ENDS ratios were calculated at a window width of 120 bp centered at position −130. The values for the wild-type (wt) sequence are indicated. Equations for the regression lines are given in Figure 5.

## RESULTS

The location of sequence-directed bent DNA in the *ilvIH* promoter/regulator region has been identified previously by circular permutation studies (11). An example of the analysis, which places the center of bending at a position about 120 bp upstream from the transcription start site, is presented in the top panel of Figure 1. The profile of the electrophoretic anomaly is compared to the bending predicted by four computer programs which differ in dinucleotide wedge angles (Figure 1). Predicted bending is expressed as the ENDS ratio, which is defined as the ratio of the contour length of a segment to the shortest distance between the ends of the segment. Programs 1 and 3 assume bending occurs only at AA and TT sequences, with the value for the AA/TT wedge angle being 8.7° and 7.2°, respectively (36,37). In both cases, the positions of the ENDS ratio peaks coincide with the center of bending determined by the electrophoretic analysis. Program 2, which incorporates all 16 dinucleotide wedge angles (37), appears to display less predictive capacities. The relatively high ENDS ratio peak at position −200 is not expected from the electrophoretic data. In addition, the ratio of the value of the ENDS ratio peak at the bend center (−120) to the average ENDS ratio signal is lower for program 2 than for programs 1 and 3. The ENDS ratio values generated by program 4 are low and relatively uniform throughout the sequence. This program uses wedge values for all non-AA/TT dinucleotides.

The DNA sequence in the vicinity of the bend center at position −120 to −130 contains several $A_nT_n$ tracts which are in approximate phase with the DNA helical repeat (Figure 2A). Eight of the tracts are located within the 6 Lrp binding sites (27). Binding of Lrp to these sites has been shown to enhance curvature of the region (11). The tracts at positions −180 to −125 bp are out of phase with those at positions −94 to −53 by an average of about 3 bp. This phase shift should change the direction of bending in the vicinity of the bend center. Projections of the DNA axis of the sequence, produced by program 1, are shown in Figure 2B. The change in direction of curvature in the vicinity of the bend center reflects the phase shift in $A_nT_n$ tracts.

In order to study the sequence determinants of bending, mutations were made within a 57 bp region of the *ilvIH*

POSITION

WINDOW SIZE

**90 bp**

| Program | -160 r | -160 slope | -160 Y int | -130 r | -130 slope | -130 Y int | -115 r | -115 slope | -115 Y int | -90 r | -90 slope | -90 Y int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.507 | 0.043 | 1.087 | 0.797 | 0.237 | 1.038 | 0.805 | 0.363 | 1.011 | 0.801 | 0.264 | 1.017 |
| 2 | 0.121 | 0.010 | 1.088* | 0.767 | 0.170 | 1.008 | 0.752 | 0.237 | 1.002 | 0.607 | 0.136 | 1.027 |
| 3 | 0.416 | 0.022 | 1.063 | 0.796 | 0.144 | 1.022 | 0.831 | 0.227 | 1.004 | 0.762 | 0.150 | 1.015 |
| 4 | -.341 | -.008 | 1.019* | -.356 | -.009 | 1.014* | -.202 | -.004 | 1.010* | 0.014 | 0.000 | 1.007* |

**120 bp**

| Program | -160 r | -160 slope | -160 Y int | -130 r | -130 slope | -130 Y int | -115 r | -115 slope | -115 Y int | -90 r | -90 slope | -90 Y int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.632 | 0.119 | 1.083 | 0.887 | 0.558 | 1.032 | 0.886 | 0.565 | 1.003 | 0.820 | 0.311 | 1.028 |
| 2 | 0.254 | 0.038 | 1.091* | 0.834 | 0.301 | 1.018 | 0.805 | 0.307 | 0.997 | 0.606 | 0.176 | 1.045 |
| 3 | 0.543 | 0.058 | 1.061 | 0.880 | 0.313 | 1.020 | 0.879 | 0.328 | 1.004 | 0.787 | 0.184 | 1.021 |
| 4 | -.370 | -.012 | 1.024* | -.201 | -.014 | 1.019* | 0.010 | 0.000 | 1.011* | 0.211 | -.006 | 1.012* |

**200 bp**

| Program | -160 r | -160 slope | -160 Y int | -130 r | -130 slope | -130 Y int | -115 r | -115 slope | -115 Y int | -90 r | -90 slope | -90 Y int |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.865 | 0.669 | 1.053 | 0.887 | 0.784 | 0.977 | 0.893 | 0.727 | 0.976 | 0.898 | 0.532 | 1.024 |
| 2 | 0.669 | 0.280 | 1.067 | 0.754 | 0.265 | 1.036 | 0.721 | 0.225 | 1.048 | 0.626 | 0.166 | 1.078 |
| 3 | 0.847 | 0.349 | 1.048 | 0.877 | 0.397 | 0.999 | 0.880 | 0.372 | 0.997 | 0.877 | 0.286 | 1.025 |
| 4 | -.288 | -.015 | 1.027* | -.309 | -.024 | 1.034* | -.306 | -.026 | 1.036* | -.273 | -.020 | 1.029* |

**290 bp**

| Program | -160 r | -160 slope | -160 Y int | -130 r | -130 slope | -130 Y int | -115 r | -115 slope | -115 Y int |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.801 | 0.950 | 1.032 | 0.822 | 0.755 | 1.022 | 0.836 | 0.614 | 1.050 |
| 2 | 0.593 | 0.281 | 1.105 | 0.404 | 0.132 | 1.132 | 0.132 | 0.036 | 1.171* |
| 3 | 0.780 | 0.450 | 1.061 | 0.787 | 0.341 | 1.048 | 0.790 | 0.266 | 1.063 |
| 4 | -.337 | -.029 | 1.039* | -.337 | -.036 | 1.046* | -.223 | -.103 | 1.075* |

**375 bp**

| Program | -160 r | -160 slope | -160 Y int |
|---|---|---|---|
| 1 | 0.709 | 0.622 | 1.143 |
| 2 | 0.052 | -.240 | 1.064* |
| 3 | 0.574 | 0.241 | 1.149 |
| 4 | -.341 | -.042 | 1.145* |

**Highest ENDS ratio peak in the 57 bp mutagenized region**

**120 bp**

| Program | r | slope | Y int |
|---|---|---|---|
| 1 | 0.875 | 0.524 | 1.010 |
| 2 | 0.710 | 0.226 | 1.037 |
| 3 | 0.873 | 0.296 | 1.007 |
| 4 | -.348 | -.012 | 1.020* |

**Highest ENDS ratio peak in entire 395 bp fragment**

**120 bp**

| Program | r | slope | Y int |
|---|---|---|---|
| 1 | 0.930 | 0.605 | 1.015 |
| 2 | 0.475 | 0.042 | 1.183 |
| 3 | 0.814 | 0.189 | 1.015 |
| 4 | 0.355 | -.009 | 1.025* |

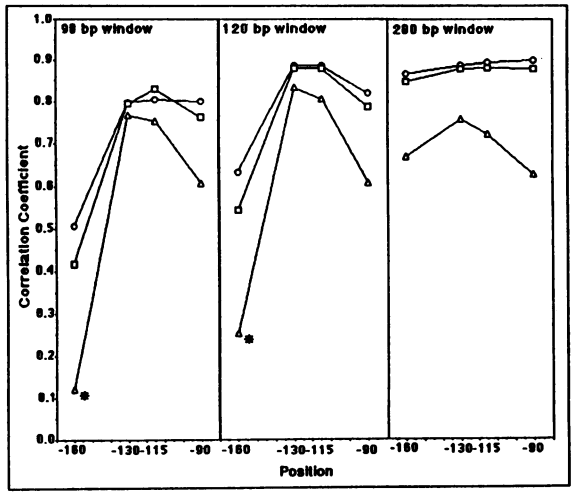



**Figure 5.** Relationship between electrophoretic anomaly and predicted curvature at various positions along the sequence. ENDS ratios centered at the indicated positions were computed at the indicated window widths by programs 1–4. The results are given in the form of characteristics of the regression lines. The Y intercept (Y int) values correspond to X = 1.0 (see Figure 4). Relationships indicated by the stars were not significant ($P > 0.005$). The insert shows the relationships between correlation coefficients and ENDS ratios computed by programs 1 (○), 2 (Δ) and 3 (□) at the indicated positions and window widths. The relationships in the two panels adjacent to the insert were calculated using the highest ENDS ratio peaks in the mutagenized region and in the entire fragment regardless of peak position.

promoter/regulatory region that contains the bend center. The mutagenized region is located from −83 to −140 relative to the transcription start site and is enclosed by the brackets in Figure 2. The identity of the introduced mutations are given in Figure 3. On average, 6 changes per molecule were made. Four of the mutations (BM286, BM243, BMG6 and BM 253) have the same number of bp as the wild-type sequence and these are shown at the top of Figure 3. However, most of the mutations have a T deletion at position −138. For this reason, comparisons of these fragments are made with Del70, which has this deletion but no other change. Wild-type and mutated sequences were analyzed by electrophoresis as 393–395 bp fragments containing the entire *ilvIH* promoter/regulatory region shown in Figure 2. As expected, the 56 mutated DNA fragments had the same mobility as the 395 bp wild-type sequence when analyzed on agarose gels (data not shown). However, significant mobility differences were observed upon electrophoresis through 6% polyacrylamide gels run at 4°C.

The apparent lengths (App. bp) and relative lengths ($R_L$ = apparent length/actual length) of the fragments are given in Figure 3. A few general observations can be made. In some mutants, bending as measured by polyacrylamide gel electrophoresis was increased, whereas in others bending was unaffected. In most of them, however, the extent of DNA bending was reduced. In the mutant having the greatest change (BM287), the apparent length was reduced to 438 bp, which is close to the size of the 394 bp fragment. This illustrates the importance of the 57 bp mutagenized region in the bending of the entire *ilvIH* promoter/regulatory element. The importance of the first T tract is clearly illustrated from inspection of the data. For example, mutations in constructs BMG6 (TTT→TCT) and Del70 (TTT→TT) reduced the apparent length from 600 bp to 549 and 547 bp, respectively. Likewise, a mutation in this region of construct BM269 (TTATT→TTTTT) is probably responsible for the increase in apparent mobility from 547 to 587 bp. For the most part, however, the sequences were difficult to analyze in this manner, since most of them contained multiple nucleotide changes.

The strategy used for analyzing the experimental results in Figure 3 consisted of determining the strengths of the relationships between the electrophoretic data and the predicted curvature

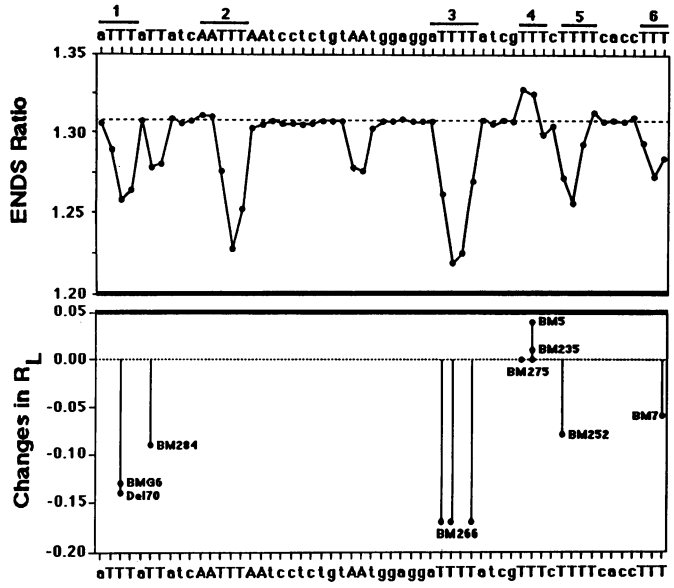



**Figure 6.** Predicted and observed changes in bending upon disruption of single tracts. **Top** panel: The 6 tracts in the 57bp mutagenized region are indicated at the top of the figure. Each nucleotide in the region was converted to G and the ENDS ratios of the 57 derivative sequences were determined by Program 1 as described in Figure 3. The ENDS ratio of the wild-type sequence is indicated by the dotted line. **Bottom** panel: Deviations in $R_L$ values from control displayed by the 9 mutagenized sequences from Figure 3 which disrupt a single AA/TT dinucleotide or single tract. Control $R_L$ values (wt or Del 70) are indicated by the dotted line.

calculated by the 4 computer programs that differ in wedge angle composition. Figure 4 shows $R_L$ measurements from Figure 3 plotted against the ENDS ratio values generated by the 4 programs. Linear relationships are assumed, since these regression lines yielded correlation coefficients which were at least as high as those seen when a variety of higher-order curves were fitted to the data (not shown). ENDS ratios were calculated at a window width of 120 bp centered at position −130, which is the approximate bend center of the *ilvIH* regulatory/promoter region (Figure 1). High correlation coefficients were observed with programs 1−3 while no significant relationship ($P > 0.05$) was detected for program 4, which is based on the non-AA/TT wedge values. Similarly, strong relationships between predicted bending and $R_L$ values were observed when the analysis was resstricted to sequences with single base mutations using programs 1−3 ($r = 0.993$, 0.941, and 0.984 respectively) but not program 4. In addition, the average ENDS ratio magnitude produced by program 4 was no more than 10% of the average values produced by the other three programs. This is in agreement with the results in Figure 1.

In Figure 3, the 57 bp mutagenized segments were analyzed by electrophoresis in the context of a 395 bp fragment containing the entire *ilvIH* promoter/regulatory region. It was important to compare calculated bending at various positions along the fragment to the electrophoretic data. In Figure 5, ENDS ratios were calculated by the 4 programs at the indicated window widths and positions along the *ilvIH* promoter/regulatory region. The figure gives a summary of the relationships between the calculated bending for each program and $R_L$ values. Four trends are evident from the 17 analyses. First, programs 1 and 3 gave the highest correlation ratios in all cases. The highest correlation coefficient ($r = 0.93$) was seen when using the highest ENDS ratio per sequence computed by program 1 at a 120 bp window width. Previous studies have shown that program 1 accurately predicted the electrophoretic mobility of synthetic DNA fragments and fragments from 107 kb of natural DNA using these parameters (15). Second, there were no significant relationships between $R_L$ values and bending calculated by program 4. Third, the highest correlation coefficients were seen at the bend center (−130) when bending was calculated at window widths of 90−200 bp by programs 1−3. Fourth, the magnitude of the decline in $r$ values as a function of distance from the bend center was greater with program 2 than programs 1 and 3. This feature is illustrated in the insert to the figure. Likewise, the strength of the relationships declined at window widths greater than 200 bp and the decrease was most pronounced with program 2. These four trends were also seen when the analysis was restricted to the sequences from Figure 3 with single base changes (data not shown).

The results in Figure 5 suggest that the curvature scores calculated by programs 1 and 3 more accurately reflect the electrophoretic data than the scores from program 2. However, one could argue that the $R_L$ values of a small subset of the 58 sequences is best described by program 2 but that the ENDS ratio signals are masked by the bulk of the correlated sequences. This does not appear to be the case, since $R_L$ values are either overestimated or underestimated to about the same extent by programs 1−3. This can be seen from the data in Figure 3, which shows that the 5 most deviant $R_L$ values seen with program 1 are a subset of the 8 most deviant $R_L$ values seen with program 2. The possibility cannot be excluded that both programs fail to detect a sequence characteristic that influences the electrophoretic mobilities of the fragments. However, there are no obvious sequence features that are common to the 5 deviant molecules detected by both programs used in Figure 3. Experimental error could also account for most if not all of the deviations seen when programs 1 and 3 were used to analyze the data.

The results in Figure 5 indicate that the sequence-directed curvature of the *ilvIH* promoter-regulatory region as measured by electrophoretic mobility can best be described by assuming that AA/TT stacks are the major determinants of bending. Thus, it was of interest to determine the relative contributions of these dinucleotides to the overall bending of the region. The computer simulation shown at the top of Figure 6 was performed to address this question. Each nucleotide in the wild-type sequence was converted to G and the ENDS ratio for each modified sequence was calculated by Program 1. The analysis predicts that the indicated tracts contribute diffentially to bending. Specifically, disruptions of tracts 1,2,3,5 and 6 should decrease curvature of the overall region while disruption of tract 4 should increase bending. This is because tract 4 is out of phase with tracts 3, 5, and 6, and consequently deflects the helical axis in the opposite direction. The results of this simulation are compared to the changes in $R_L$ values seen in the 9 mutagenized sequences indicated in the bottom of the figure which show a disruption of a single tract or AA/TT dinucleotide. The results illustrate that the direction and relative magnitude of the deviations in each $R_L$ value was predicted by the computer simulation. A correlation coefficient of 0.976 characterized this relationship.

## DISCUSSION

Since the discovery of phased A tracts in kinetoplast DNA, it has generally been assumed that short runs of As are the major determinant of DNA bending (2–5). The relative magnitudes of curvature predicted by the 4 programs used in this work are consistent with this view. The results in Figures 1, 4 and 5 show that, on average, about 90% of each ENDS ratio value generated with a program that includes all dinucleotide wedge values (program 2) is due to bending at AA/TT. The disparity in the magnitude of bending assigned to AA/TT vs. non-AA/TT steps has also been seen for other natural DNA sequences (38) and can be attributed to more effective wedge summations in oligo(A/T) tracts when compared to non-AA/TT sequences, as discussed in McNamara and Harrington (39). Our results do not distinguish between the major models of bending at oligo(A) tracts which assume that helical deflections occur either at the level of individual AA wedges within the tract (1,5,35,37) or at the junctions of the tract and flanking DNA (3,40). However, the results do suggest that a more complete wedge model proposed by Bolshoy *et al.* (37), which contains wedge values for all dinucleotide steps, does not adequately describe the electrophoretic data in this report. This is seen from the data in Figures 1, 4 and 5 which show that: (i) there were no significant relationships between electrophoretic anomaly and curvature values predicted by program 4; and (ii) the predictive power of electrophoretic mobility was consistently highest when sequences were analyzed with programs which assume that AA/TT is the sole source of bending (programs 1 and 3). The decline in the correlation ratios seen with program 2 compared with programs 1 and 3 as a function of increasing window widths and increasing distances from the bend center shown in Figure 5 can thus be attributed to the dilution of AA/TT bending signals with spurious signals from non-AA/TT elements.

Electrophoretic retardation of non-AA/TT synthetic sequences has been reported by several laboratories (14,37,39,41,42). However, the magnitudes of the anomalies are invariably much less than those seen with A tract-containing fragments. Unlike A tract sequences, the electrophoretic anomaly of most of these fragments actually decreases with increasing fragment length (14,37,39,41). In addition, the anomaly of the 7 non-AA/TT fragments analyzed by Bolshoy *et al.* (37) did not correlate with their predicted curvature values, which were calculated using a program that is comparable to program 2. Taken together, these observations imply that the features responsible for the electrophoretic anomaly of AA-containing vs. non-AA fragments are distinct. Whether the mild electrophoretic anomaly of non-AA/TT fragments is due to sequence-dependent differences in flexibility (14,42), kinking (37,41) or some other local variation in helical structure is not yet known.

Bent DNA in the vicinity of promoters plays an important role in regulation of transcription in some systems. Evidence for this view has been provided by correlations between promoter strength and predicted bending (7), by mutation analysis (6,8,43,44) and by the observation that synthetic bent DNA can substitute functionally for naturally curved upstream activating sequences (19,20). The overall spatial organization of the bent segments, as well as their orientation relative to promoter elements, are also critical factors for stimulation of transcription by upstream curved DNA. For example, McAllister and Achberger (8) have shown that insertions of oligonucleotides (6–29 bp) between the 2 regions of curvature in the Alu156 promoter region from phage SP82 altered the electrophoretic mobility and promoter activity in a sinusoidal 10 bp pattern in a manner that suggested an S-shaped molecule. Computer modeling of these sequences using program 1 revealed that the ENDS ratio values also displayed a sinusoidal pattern that was highly correlated with their electrophoretic data ($r = 0.92$). These results are consistent with the interpretation of McAllister and Achberger and suggest that A/T tracts within the 2 bends that comprise the S-shaped structure are out of phase with each other in the wild-type molecule. The *ilvIH* promoter/regulatory region from *E.coli* (Figure 2) and from *Salmonella typhimurium* (data not shown) also appear S-shaped using program 1. This conservation in structure may be relevant to the common function of the regions, since there is considerable divergence (~40%) in the nucleotide sequence between the segments from the two organisms (45). The offset in the sequence implies that there is a change in the rotational orientation of the helix in the vicinity of the bend center (46,47) that is located in the approximate center of the S-shaped structure (Figure 2). This central site also contains a T tract which is out of phase with flanking T tracts. This arrangement inhibits local bending in the region (Figure 2B, 6A and 6B). DNA that preferentially positions nucleosomes is also offset in sequence which gives rise to computer generated S-shaped structures that resemble the predicted structure of the *ilvIH* promoter/regulatory region (38). These sequencese also contain irregularly spaced oligo A/T tracts at the central site (38,46). Although the precise function of these apparently common features is presently unknown, it seems likely that the S-shaped structures in naked DNA play some role in facilitating the winding of DNA into high-order particles following protein binding (38).

## REFERENCES


1. Trifonov, E. N. and Sussman, J. L. (1980) Proc. Natl Acad. Sci. USA **77**, 3816–3820.
2. Marini, J. C., Levene, S. D., Crothers, D. M. and Englund, P. T. (1982) Proc. Natl Acad. Sci. USA **79**, 7664–7668.
3. Crothers, D. M., Haran, T. E. and Nadeau, J. G. (1990) J. Biol. Chem. **265**, 7093–7096.
4. Hagerman, P. J. (1990) Annu. Rev. Biochem. **59**, 755–781.
5. Trifonov, E. N. (1991) TIBS **16**, 467–470.
6. Bossi, L. and Smith, D. M. (1984) Cell **39**, 643–652.
7. Plaskon, R. R. and Wartell, R. M. (1987) Nucleic Acids Res. **15**, 785–795.
8. McAllister, C. F. and Achberger, E. C. (1989) J. Biol. Chem. **265**, 10451–10456.
9. VanWye, J. D., Bronson, E. C. and Anderson, J. N. (1991) Nucleic Acids Res. **19**, 5253–5261.
10. Lavigne, M., Herbert, M. Kolb, A. and Buc, H. (1992) J. Mol. Biol. **224**, 293–306.
11. Wang, Q. and Calvo, J. M. (1993) EMBO J. **12**, 2495–2501.
12. Ross, W., Shulman, M. and Landy, A. (1982) J. Mol. Biol. **156**, 505–529.
13. Zahn, K. and Blattner, F. (1985) Nature **317**, 451–453.
14. Anderson, J. N. (1986) Nucleic Acids Res. **14**, 8513–8533.
15. Eckdahl, T. T. and Anderson, J. N. (1987) Nucleic Acids Res. **15**, 8531–8545.
16. Zahn, K. and Blattner, F. (1987) Science **236**, 416–422.
17. Eckdahl, T. T. and Anderson, J. N. (1990) Nucleic Acids Res. **18**, 1609–1612.
18. Williams, J. S., Eckdahl, T. T. and Anderson, J. N. (1988) Mol. Cell Biol. **8**, 2763–2769.
19. Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. and Buc, H. (1989) EMBO J. **8**, 4289–4296.
20. Gartenberg, M. R. and Crothers, D. M. (1991) J. Mol. Biol. **219**, 217–230.
21. Goodman, S. D. and Nash, H. (1989) Nature (London) **341**, 251–254.
22. Snyder, U. K., Thompson, J. F. and Landy, A. (1989) Nature (London) **341**, 255–257.
23. Squires, C. H., DeFelice, M., Wessler, S. R. and Calvo, J. M. (1981) J. Bacteriol. **147**, 797–804.

24. DeFelice, M., Lago, C. T., Squires, C. H. and Calvo, J. M. (1982) Ann. Microbiol. (Inst. Pasteur) **133A**, 251–256.
25. Haughn, G. W. Squires, C. H., DeFelice, M., Lago, C. T. and Calvo, J. M. (1985) J. Bacteriol. **163**, 186–198.
26. Willins, D. A. and Calvo, J. M. (1992) J. Bacteriol. **174**, 7648–7655.
27. Wang, Q. and Calvo, J. M. (1993) J. Mol. Biol. **229**, 306–318.
28. Kunkel, T. A. ( 1985) Proc. Natl Acad. Sci. USA **82**, 488–492.
29. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).
30. Sanger, F., Nicklen, S., Coulson, A. R. (1977) Proc. Natl Acad. Sci. USA **74**, 5463–5467.
31. Wu, H.-M. and Crothers, D. (1984) Nature (London) **308**, 509–513.
32. Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987) Meth. Enzymol. **154**, 367–382.
33. Myers, R., Lerman, L. S. and Maniatis, T. (1985) Science **229**, 242–247.
34. Kabsch, W., Sander, S. and Trifonov, E. N. (1982) Nucleic Acids Res. **10**, 1097–1104.
35. Ulanovsky, L. E. and Trifonov, E. N. (1987) Nature (London) **326**, 720–722.
36. Ulanovsky, L. E., Bodner, M., Trifonov, E. N. and Choder, M. (1985) Proc. Natl Acad. Sci. USA **83**, 862–866.
37. Bolshoy, A., McNamara, P., Harrington, R. E. and Trifonov, E. N. (1991) Proc. Natl Acad. Sci. USA **88**, 2312–2316.
38. Fitzgerald, D. J., Dryden, G. L., Bronson, E. C., Williams, J. S. and Anderson, J. N. (1994) J. Biol. Chem. 269, 21303–21314.
39. McNamara, P. T. and Harrington, R. E. (1991) J. Biol. Chem. **266**, 12548–12554.
40. Koo, H. S., Wu, H. M. and Crothers, D. M. (1986) Nature (London) **320**, 501–506.
41. McNamara, P. T., Bolshoy, A., Trifonov, E. N. and Harrington, R. E. (1990) J. Biomol. Structure Dynamics **8**, 529–538.
42. Lyubchenko, Y. L., Shlyakhtenko, L. S., Appella, E. and Harrington, R. E. (1993) Biochemistry **32**, 4121–4127.
43. Mizuno, T. (1987) Gene (Amst.) **54**, 57–64.
44. Bauer, B. F., Kar, E. G., Elford, R. M. and Holmes, W. M. (1988) Gene (Amst.) **63**, 123–134.
45. Wang, Q., Sacco, M., Ricca, E., Lago, C. T., DeFelice, M. and Calvo, J. M. (1993) Mol. Microbiol. **7**, 883–891.
46. Satchwell, S. C., Drew, H. R. and Travers, A. A. (1986) J. Mol. Biol. **191**, 659–675.
47. Travers, A. A. and Klug, A. (1987) Phil. Trans. R. Soc. Lond. **317**, 537–561.