17:47:36 OCA PAD AMENDMENT - PROJECT HEADER INFORMATION 10/06/93

Active

Project #: G-32-631 Cost share #: Rev #: 2
Center # : 10/24-6-R7099-3A0 Center shr #: OCA file #:
Work type : RES
Contract#: 5 R01 GM38045-03 Mod #: EXTENSION Document : GRANT
Prime #: Contract entity: GTRC

Subprojects ? : N CFDA: 93.821
Main project #: PE #:

Project unit: BIOLOGY Unit code: 02.010.134
Project director(s):
WARTELL R M BIOLOGY (404)894-5247

Sponsor/division names: DHHS/PHS/NIH / NATL INSTITUTES OF HEALTH
Sponsor/division codes: 108 / 001

Award period: 930101 to 941231 (performance) 950331 (reports)

Sponsor amount	New this change	Total to date
Contract value	0.00	83,660.86
Funded	0.00	83,660.86
Cost sharing amount		0.00

Does subcontracting plan apply ?: N

Title: THERMODYNAMICS OF BASE PAIR OPENING IN DNA

PROJECT ADMINISTRATION DATA

OCA contact: E. Faith Gleason 894-4820

Sponsor technical contact

DR. MICHELLE BROIDO, PROGRAM MGR.
(301)496-7585

NATIONAL INSTITUTES OF HEALTH
NAT INSTITUTE OF GEN MEDICAL SCIENCE
5333 WESTBARD AVENUE
BETHESDA, MD. 20892

Sponsor issuing office

ANN CALURE, LUCY CLARKE, I. GRISSON
(301)496-7275

NATIONAL INSTITUTES OF HEALTH
NAT INSTITUTE OF GEN MEDICAL SCIENCE
5333 WESTBARD AVENUE
BETHESDA, MD. 20892

Security class (U,C,S,TS) : U ONR resident rep. is ACO (Y/N): N
Defense priority rating : N/A NIH supplemental sheet
Equipment title vests with: Sponsor GIT X

Administrative comments -
* BUDGET PERIOD EXTENDED 12 MONTHS IN ACCORDANCE WITH DR. WARTELL'S LETTER OF SEPT. 9, 1993 AND "EXPANDED AUTHORITIES."

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

NOTICE OF PROJECT CLOSEOUT

Closeout Notice Date 03/07/95

Project No. G-32-631 Center No. 10/24-6-R7099-3A0

Project Director WARTELL R M School/Lab BIOLOGY

Sponsor DHHS/PHS/NIH/NATL INSTITUTES OF HEALTH

Contract/Grant No. 5 R01 GM38045-03 Contract Entity GTRC

Prime Contract No.

Title THERMODYNAMICS OF BASE PAIR OPENING IN DNA

Effective Completion Date 941231 (Performance) 950331 (Reports)

Closeout Actions Required:	Y/N	Date Submitted
Final Invoice or Copy of Final Invoice	Y	
Final Report of Inventions and/or Subcontracts	Y	950308
Government Property Inventory & Related Certificate	N	
Classified Material Certificate	N	
Release and Assignment	N	
Other	N	

Comments

Subproject Under Main Project No.

Continues Project No. G-32-628

Distribution Required:

Project Director	Y
Administrative Network Representative	Y
GTRI Accounting/Grants and Contracts	Y
Procurement/Supply Services	Y
Research Property Managment	Y
Research Security Services	N
Reports Coordinator (OCA)	Y
GTRC	Y
Project File	Y
Other	N
	N

NOTE: Final Patent Questionnaire sent to PDPI.

G-32-631
2

TO

NATIONAL INSTITUTE OF GENERAL MEDICAL SCIENCES
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MARYLAND

FINAL PROGRESS REPORT ON

GRANT NUMBER: GM38045

PROJECT TITLE: THERMODYNAMICS OF BASE PAIR OPENING IN DNA

PROJECT PERIOD: 1/1/91 TO 12/31/94

FROM

PRINCIPAL INVESTIGATOR:

ROGER M. WARTELL
SCHOOL OF BIOLOGY
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA 30332

PHONE NO. 404-894-3700
E-MAIL ADDRESS: ROGER.WARTELL@BIOLOGY.GATECH.EDU

2/22/94

Summary Progress Report:

Mutation Detection by Temperature Gradient Gel Electrophoresis. Several objectives were achieved in the development of temperature gradient gel electrophoresis (TGGE) as a method to detect mutations in DNA. In this method DNA fragments migrate in a vertical polyacrylamide gel with a temperature gradient superimposed parallel to the direction of mobility. A duplex DNA migrates until it reaches a temperature at which it partially unwinds. DNAs differing by a single base pair in an early melting domain unwind at different depths and can be readily distinguished.

DNA melting theory has been employed to predict the melting behavior of DNA sequences and thus determine the DNA fragments which enable mutations to be detected. While generally successful, denaturing gel methods have not always revealed base changes prredicted by calculations. In the study by Ke et al. (Electrophoresis, 1993, ref.8), new criteria were established to predict which DNA fragments within a long sequence will place designated base pairs in early melting domains. The criteria were tested in four "hot spot" regions of the human p53 tumor suppressor gene. 98% of approximately 200 known point mutations are clustered in four regions 30 to 70 bp long. Calculations with the new criteria indicated that three DNA fragments amplified from the gene's cDNA can detect base substitutions in the four hot spot regions. TGGE experiments were conducted with DNAs containing known single base substitutions. All mutations tested behaved as predicted and were detected by TGGE.

A computer program, MELTSCAN™ , was developed to automate the selection of DNA fragments for detecting mutations by any denaturing gel electrophoretic method (Brossette & Wartell Nucleic Acids Res.,1994, ref. 5). The program was first applied to the prediction of fragments to detect mutations in the cDNA and genomic DNA of the human p53 gene. TGGE experiments were carried out on DNA fragments amplified from DNA and RNA isolated from 19 tumor cell lines and patient cells. Point mutations in the p53 gene were detected in 9 of the 19 cell lines (S. Hosseini, Findley and Wartell in preparation, ref 3.).

TGGE was also employed to develop a method to rapidly screen for drug resistance in bacteria. The study employed Mycobacterium smegmatis, a relative of M. tuberculosis. Streptomycin and rifampicin resistance in M. smegmatis were examined. Streptomycin resistance in this organism has been shown to be due to point mutations in the s12 ribosomal gene. Rifampicin resistance is due to poitn mutations localized to the rpsL gene. MELTSCAN and the TGGE method have demonstrated that strains resistant to these drugs can be detected using several DNA fragments amplified from M. smegmatis genomic DNA (Churchward et al., Keystone Symposium, 1995, ref. 1).

DNA Nearest Neighbor Stacking Interactions. One objective of the project was to determine the thermodynamic parameters describing nearest neighbor stacking interactions in DNA. Although several sets of parameters have been published, a set of parameters that accurately describes the melting behavior of both short and long DNAs has not been obtained. Experimental UV melting curves were obtained in 0.1 M NaCl from four 15 bp DNA oligomer duplexes, four other DNA oligomers 13 to 25 bp in length, as well as 13 DNA fragments , 80-600 bp. A least squares fitting approach was employed to determine the 10 stacking interactions. Although the parameter

set gave a good fit, the comparison did not make predictions within expectations of experimental error. Data from some DNA oligomers indicated that interactions extending beyond nearest neighbor base pairs are significant. Oligomers with a sequence difference but with the same nearest neighbor stacking had different melting temperatures. TGGE has provided a set of relational constraints on the stacking free energies (see below). These are being used to re-evaluate the DNA stacking interactions.

Influence of Nearest Neighbor Sequence on the Stability of Base Pairs, Mismatches and Bulges in Long DNA. The temperature gradient gel electrophoresis method was used to determine the relative thermal stabilities of 48 DNA fragments differing by single base mismatches (Ke & Wartell Nucleic Acids Res., 1993, ref. 7). Results showed that both the bases at the mismatched site and neighboring stacking interactions influence destabilization. G·T, G·G and G·A mismatches were always among the most stable mismatches. Purine-purine mismatches were generally more stable than pyrimidine-pyrimidine mismatches.

TGGE was also employed to examine the effect of neighboring base pairs on single base bulges and base pairs in DNA (Ke & Wartell, Biochemistry, 1995 ref 4.). The relative stability of 32 DNA fragments differing by single unpaired bases and 17 DNA fragments differing by a single base pair were determined. When a DNA had an unpaired base which was identical to one its adjacent bases, it caused less destabilization than an unpaired base with an identity different from its neighbors. This general phenomena may explain the persistence of homogenous base tracts over evolutionary time.

The relative stability of base pair stacking interactions was determined from the stabilities of DNA fragments differing by a single base pair. The hierarchy of stacking interactions was the same as the hierarchy determined from some UV melting studies. This lends support to both approaches. It provides constraints for any least squares determination of the stacking interactions.

The relative stability of two adjacent base-base mismatches was also examined using TGGE. 28 DNA fragments that differ by two adjacent mismatches from completely paired DNAs were examined (Ke & Wartell, unpublished, ref 2.). We observed that the (dGpdA)·(dGpdA) mismatch in a long DNA creates a more stable duplex than a Watson-Crick homoduplex with dA·dT base pairs at the same site. The biological significance of this is unclear. In addition other adjacent mismatched pairs, e.g., (dApdA)·(dGpdA), cause relatively little destabilization of the duplex. Analysis of the DNA's melting curves indicate that some adjacent mismatches cause local base pair opening during the melting process.

Publications

1. "Identification of Streptomycin- and Rifampicin Resistant Mutants of M. smegmatis and identification of mutations by Temperature Gradient Gel Electrophoresis."
G. Churchward, T. Kenney, Y.Yu, S. Powell, and R.M. Wartell, poster at Keystone Symposium on Molecular Mechanisms in Tuberculosis (1995) (anticipated publication).

2. "The Thermal Stabilities of Adjacent Base Pair Mismatches in the d(CpXpYpG)·d(GpY'pX'pC) environment." S-H. Ke and R.M. Wartell, in preparation.

3. " Optimizing Mutation Detection by Temperature Gradient Gel Electrophoresis: Application to p53 Gene Mutations in Acute Lymphoblastic Leukemia Cells." S. Hosseini, H. Findley and R.M. Wartell, in preparation.

4. "The Influence of Neighboring Base Pairs on the Stability of Single Base Bulges and Base Pairs in a DNA Fragment." S-H Ke and R.M. Wartell Biochemistry in press (1995).

5. "A program for selecting DNA fragments to detect mutations by denaturing gel electrophoresis methods." S. Brossette and R.M. Wartell Nucleic Acids Research 22 4321-4325 (1994).

6. " Localization of the intrinsically bent DNA region upstream of the E. coli rrnB P1 promoter." T.Gaal, L. Rao, S.T. Estrem, J. Yang, R.M. Wartell, and R.L.Gourse Nucleic Acids Research 22 2344-2350 (1994).

7. " Dependence of Base Pair Mismatch Stabilities on Nearest Neighbor Sequence: Determination by Temperature Gradient Gel electrophoresis." S. Ke and R. M. Wartell Nucleic Acids Research 21 5137-5143 (1993).

8. " Selecting DNA fragments for mutation detection by temperature gradient gel electrophoresis: Application to the p53 gene cDNA." S. Ke, P. Kelly, R.M. Wartell and V. Varma Electrophoresis 14 561-565 (1993).

Song-hua Ke[1]
Patrick J. Kelly[1]
Roger M. Wartell[1]
Stephen Hunter[2]
Vijay A. Varma[2]

[1]School of Biology, Georgia Institute of Technology, Atlanta, GA
[2]Veterans Administration Hospital, Research Services, Department of Pathology, Emory University, Clairmont Road, Decatur, GA

Selecting DNA fragments for mutation detection by temperature gradient gel electrophoresis: Application to the p53 gene cDNA

Temperature gradient gel electrophoresis (TGGE) and related methods are widely employed to detect mutations in DNA fragments. DNA melting map calculations and GC clamps have been used to enhance the detection of mutations. While generally successful, these methods have not always revealed base changes within a DNA fragment. Previous work suggested that mutations are detected if they are in a DNA's first melting domain, and the melting domain is well separated from final strand dissociation. Two criteria from the DNA melting theory were established to determine when both of these conditions are met. The criteria involve calculating the derivative melting curve as well as the melting map of a DNA sequence. The approach was applied to the cDNA sequence of the human p53 gene. Mutations in the p53 gene are common in human cancers and are generally located in four 'hot spot' regions. Calculations indicated that three DNA fragments are needed to detect base substitutions in the four hot-spot regions. Predicted melting behavior was experimentally tested with eight single base substitutions distributed among the four hot-spot regions. All mutations tested behaved as predicted and were detected by vertical TGGE. Heteroduplex DNAs formed by melting and reannealing various ratios of wild type and mutant DNA fragments were also examined. Results indicated that point mutations can be detected by ethidium bromide staining from samples containing 10% mutant and 90% wild-type sequences.

1 Introduction

Denaturing gradient gel electrophoresis (DGGE) [1, 2] and the related methods of temperature gradient gel electrophoresis (TGGE) [3, 4] and constant denaturing gel electrophoresis (CDGE) [5, 6] have demonstrated their ability to detect single base substitutions. DGGE employs a gradient of denaturing solvent (urea and formamide) in a polyacrylamide gel. A duplex DNA migrates until it reaches a level of denaturant which induces the least stable domain to unwind. The mobility of partially melted DNA decreases dramatically. DNAs which differ in sequence in an early melting domain unwind at different depths in the gel. The temperature gradient approach, TGGE, simplifies several features of DGGE. Heating plates provide an easily controlled linear temperature gradient. The need for a chemical gradient gel is eliminated. A DNA migrates as a duplex until it reaches a destabilizing temperature and decreases in mobility. The vertical format of TGGE is similar technically to conventional polyacrylamide gel electrophoresis. CDGE employs a polyacrylamide gel with a uniform concentration of denaturant at a constant temperature. A perpendicular denaturing gradient gel is first used to determine the concentrations of the chemical denaturant at which a DNA fragment's melting domains unwind. Constant denaturant gels are then employed to screen for mutations in any observed melting domains.

The ability of all of the above methods to separate DNA fragments differing by a single base pair (bp) depends on the location of the base pair and the DNA unwinding process [7, 8]. DNA molecules tend to unwind in discrete domains. Mutations can be detected if they are in an early melting domain, providing the domain unwinds at a temperature (or denaturant level) well below the final duplex-to-strands dissociation step. GC-rich segments have been added to one end of a DNA to enhance this condition [9]. For DNAs under 500 bp however, the cooperative duplex-to-strands dissociation step may dominate a DNA's tendency to unwind in discrete domains [10]. Mutations may not be detectable when this situation occurs. Theoretical melting maps have frequently been employed to predict the melting behavior of a DNA and the regions where mutations may be detected [8]. Although highly useful, predicted DNA separations were not always observed [5, 11]. A possible explanation is that melting map calculations do not consider the duplex-to-single-strands dissociation step [8, 10]. Domains predicted by a melting map may not actually have a real experimental existence due to the cooperativity of the duplex-to-strands dissociation. We sought a procedure that would predict which DNA fragments within a long sequence will place designated regions in early melting domains well separated from the strands dissociation step. The procedure was tested in regions of the cDNA sequence of the human p53 gene. CDGE was previously used to screen for mutations in regions of the genomic DNA sequence of the p53 gene [6].

Mutations in the p53 tumor suppressor gene are among the most common genetic abnormalities reported in human cancer. Base substitutions in this gene have been found in a variety of common human cancers including neoplasms of the lung, colon, breast, brain and soft tissues [12–17]. The vast majority of known mutations cluster in four regions within exons 5 through 8 [17]. These mutational 'hot spot' regions correspond almost exactly to phylogenetically conserved domains of the p53 protein [17].

Correspondence: Dr. R. M. Wartell, School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA

Abbreviations: bp, base pair; **DMC**, derivative melting curve; **PCR**, polymerase chain reaction; **TGGE**, temperature gradient gel electrophoresis

0173-0835/93/0707-0561 $5.00+.25/0

2 Materials and methods

2.1 Gel electrophoresis

TGGE was carried out using a vertical gel electrophoresis apparatus previously described [4]. A 6.5% acrylamide gel was employed (acrylamide to *N,N'*-methylenebisacrylamide ratio of 37.5:1) in a solvent of 0.5 TBE (0.045 M sodium borate, 0.045 M Tris, 1 mM EDTA, pH 8.1) with 60% denaturant (100% denaturant = 7.0 M urea and 40% formamide). Ten μL of DNA samples (50–100 ng) were mixed with 3 μL of loading buffer (30% Ficol and Bromophenol Blue) prior to loading in each gel lane. Electrophoresis was generally carried out at voltages of 80–90 V and run for 8–14 h (overnight). Voltages up to 150 V were used for 5 h runs. Gels were stained with ethidium bromide and photographed with UV illumination.

2.2 DNA melting calculations

Theoretical melting analysis was employed to determine polymerase chain reaction (PCR) primers that would generate DNA fragments optimal for detecting base substitutions in the mutational hot-spot regions of the p53 gene. The melting behavior of DNA sequences was calculated using a previously described algorithm [10]. This algorithm is based on a statistical mechanical approach [18, 19]. The stability of each base pair depends on its base pair type (AT or GC) and the stacking interactions with neighboring base pairs. The model also includes the entropy difference between unwinding a DNA segment from an end *vs.* opening an internal loop of similar size, and the dissociation equilibrium of partial duplex to single strands. The calculations used parameters empirically determined from DNA melting studies in 0.1 M NaCl [10, 20]. Although the environment of the urea-formamide gel clearly differs from an aqueous salt solution, a correlation was observed between transition midpoints, T_m, of DNA fragments in 0.1 M Na^+, and experimental T_m values in the gels (P. J. Kelly and R. M. Wartell; unpublished). 45 °C was subtracted from the predicted T_m to estimate the midpoint temperatures in the temperature gradient gels. Preliminary comparisons suggest that differences in the stacking parameters between our algorithm and those used by others [8] do not make a significant difference in predicting the general melting behavior of short DNA fragment. Investigators may thus use other algorithms with equivalent results, providing strand dissociation equilibrium is included. Readers interested in obtaining the program used in the calculations should write to the corresponding author.

2.3 PCR

The polymerase chain reaction was used to amplify portions of the cDNA sequence of the p53 gene. Templates were DNA plasmids with cDNA inserts of the p53 gene (gift of B. Vogelstein, The Johns Hopkins University, Baltimore). Template DNA (0.1–0.5 ng) was employed in the 100 μL reaction mixtures. The solvent contained 50 mM Tris-HCl (pH 8.6), 10 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each of the dNTP, and 50 pmole of each primer, and 2.5 units of Taq polymerase (AmpliTaq, Cetus). The reactions were incubated in an MJ Research Inc. thermocycler (Watertown, MA) and initiated with a one-minute incubation at 94 °C, followed by 25–30 cycles of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 50 s. The primers were synthesized by Operon Inc. (Alameda, CA) or on an Applied Biosystems 380A DNA synthesizer. Primers were desalted, dried, and suspended in 10 mM Tris + 0.1 mM EDTA. DNA primer lengths were 40 bases or less. The 40-base primers contained 20-base-long'GC-clamp' segments on their 5' ends. Primer pairs employed in the synthesis of the DNA fragments were as follows: 280 bp A-fragment, anti-sense primer (5'-GCCCGCCGCG GCCGCCCGCCG AGCTCCCA-GA ATGCCAGAGG-3'), sense primer (5'-AATCAACC-CA CAGCTGCACA-3'); 140 bp B-fragment, anti-sense primer (5'-GCGCGTACCGC AGCGGCGCGC CCAC-

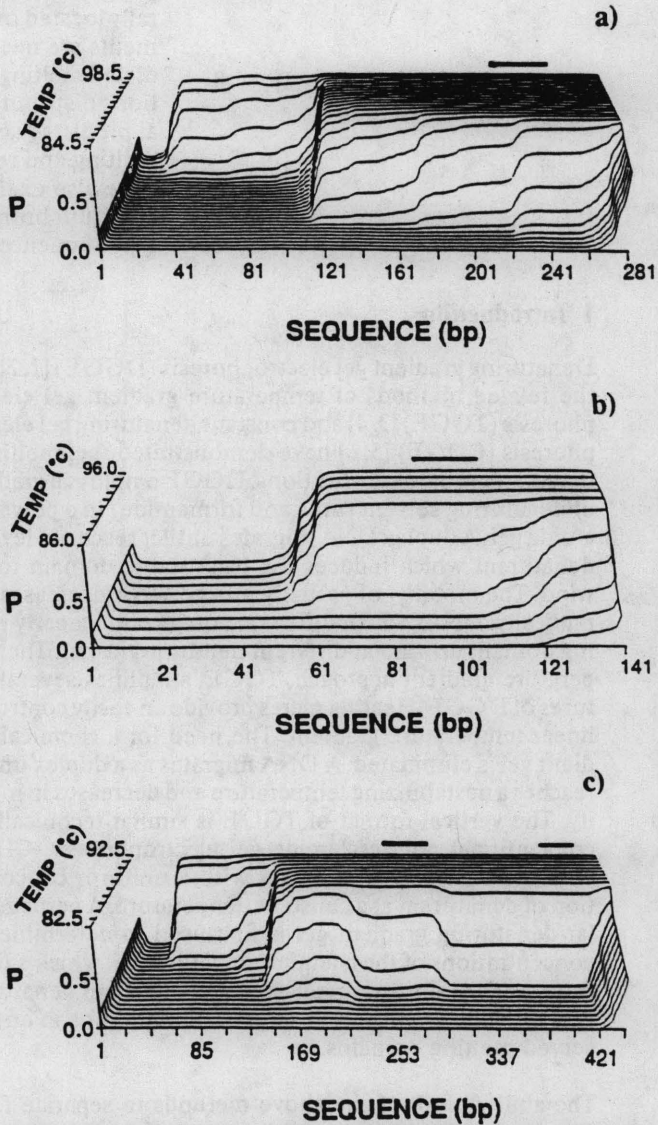


Figure 1. (a) Calculated melting profile of the 280-bp A-fragment. The fragment includes the sequence from base position 321 to 580 in the 1317-bp cDNA of the p53 gene, and a 20-bp GC clamp 5' to base 321. Probability of base pair broken, P, plotted for each base pair at various temperatures. (b) Melting profile of the 140-bp B-fragment. Sequence includes the sequence from base position 581 to 700, and a 20-bp GC clamp 5' to base 581. (c) Melting profile of 440-bp C/D-fragment. Sequence includes the sequence from base position 581 to 1000, and a 20-bp GC clamp 5' to base 581. The heavy lines above the melting profiles indicate the locations of the four hot-spot regions.

ACCCCC GCCCGGCACC-3′), sense primer (5′-CCAG-ACCATC GCTATCTGAG-3′); 440 bp C/D-fragment, anti-sense primer (same as anti-sense primer of 140 bp B-fragment), sense primer (5′-GATTCTCTTC CTCTGTGCGC-3′). To detect a mutation in codon 143, made available late in the study, a new sense primer was made for the A hot-spot region. This primer (5′-TACTATATGC CGGGC-GGGGG TGT-3′) creates a 304 bp DNA. Primers 25 or 34 bases long with single base mismatches to their complementary strands were used to produce mutations not available on plasmids. For example, to create a mutation in codon 139 in the 280 bp A-fragment, a 34-base oligomer was employed with the same 5′ end as the normal sense primer, but with a mismatch five bases from the 3′ end of the oligomer.

3 Results and discussion

Theoretical melting calculations were made to determine which DNA fragments could be used to detect mutations in the four hot-spot regions of the p53 gene. Each region was placed near one end of a potential DNA fragment and a 20 GC base pair segment was theoretically added to the other end. Overall DNA lengths were varied from 100 to 440 bp. The cDNA sequence on both sides of each hot-spot region were examined. The 20 bp GC-clamp length was selected since it was the shortest length that had a stabilizing effect on the end to which it was attached. Two criteria were used to select the DNA fragments. The first criteria required that a mutational hot-spot region was in the first melting domain of a DNA fragment. The melting domains of a DNA sequence were determined from its melting profile (*e.g.*, Fig. 1a). Each contour line in a melting profile plots the probability of being melted ($0 \leq P \leq 1$) for each base pair in the sequence at a fixed temperature. The projection into the page of a series of contour lines at increasing temperatures shows the locations of melting domains and their T_m.

The second criteria for selecting a DNA fragment required that its derivative melting curve (DMC) has at least two peaks separated by a minimum of 1 °C (*e.g.*, Fig. 2a). This criteria attempts to ensure the existence of two or more distinct melting domains during a DNA melting transition. Unlike a melting profile or melting map, which ignore strand dissociation, the calculation of a DMC includes the duplex- to single-strands dissociation equilibrium. The analysis predicted that mutations in the four hot-spot regions of the p53 cDNA sequence could be detected using three DNA fragments. These fragments were designated the 280-bp (or 304-bp) A-fragment, 140-bp B-fragment and 440 C/D-fragment.

Figure 1a shows the melting profile of the 280 bp A-fragment. The low temperature melting domain includes about 160 base pairs. Excluding the sense primer sequence, this includes codons 95–141. The derivative melting curve of dicts that unwinding of the first melting domain is separated by 5 °C from the strand dissociation step. A theoretical ated by 5 °C from the strand dissociation step. A theoretical examination of base substitutions from codons 117 to 141 indicated that mutations in this region alter the melting behavior of the first melting domain and should be detectable by TGGE. The derivative melting curve and melting profile of the 304-bp A-fragment were similar to the 280-bp A-fragment. The 304-bp A-fragment extends mutation

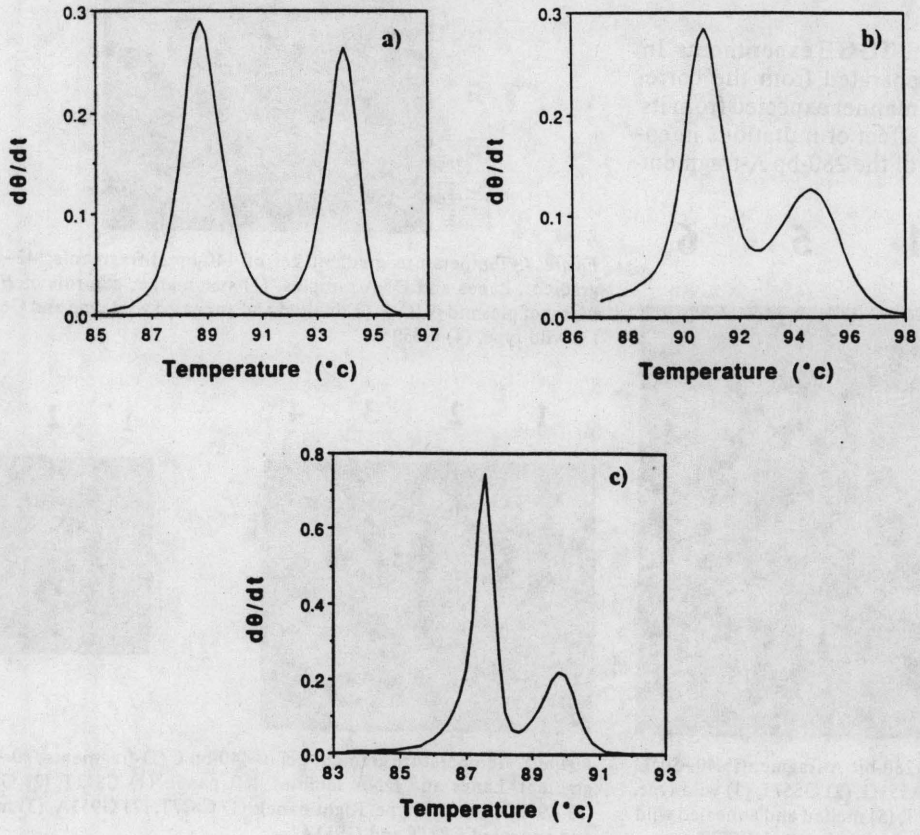


Figure 2. Calculated DMC of three DNA fragments described in Fig. 1. The temperature derivative of the fraction of broken base pairs, $d\theta/dT$, is plotted *vs.* temperature. DNA concentration was assumed to be 25 μg/mL. (a) 280-bp A-fragment, (b) 140-bp B-fragment, (c) 440-bp C/D-fragment.

detection in hot-spot region A to codon 149. Melting profiles of the 140-bp B-fragment and the 440-bp C/D-fragment are indicated in Figs. 1b and 1c. Hot-spot region B is within the first melting domain of the B-fragment, and hot-spot regions C and D are within the first melting domain of the C/D-fragment. The first melting domain of the C/D-fragment begins to unwind from the middle of the fragment, but it includes all 280 pase pairs from the right end. The DMCs of these two fragments show that their first melting domains unwind several degrees below the strand dissociation step (Figs. 2b and 2c). The above predictions were tested with eight single base substitutions. These sequence changes were previously observed in DNA derived from human tumors [17]. Table 1 lists the characteristics of the mutations. PCR products were amplified from wild-type and mutant cDNA sequences of the p53 gene inserted in plasmids. Lengths and purity of product DNAs were verified on standard 6.5% polyacrylamide gels (not shown).

Table 1. Characteristics of p53 gene mutations used in this study

Mutant[a)]	Region[b)]	codon	Nucleotide change
A551G[c)]	A	139	AAG→AGG
G557T[c)]	A	141	TGC→TTC
T563C	A	143	GTG→GCG
G659A	B	175	CGC→CAC
C670T	B	179	CAT→TAT
C877T	C	248	CGG→TGG
G953A	D	273	CGT→CAT
A977G[c)]	D	281	GAC→GGC

a) Mutant designation indicates base number of cDNA sequence. The first and last letters correspond to the wild-type and mutant base, respectively.
b) Region refers to mutational hot-spot regions.
c) Mutation created by using primer with mismatched base in PCR amplification of wild-type sequence.

Figures 3–5 show the results of the TGGE experiments. In all cases the mutant fragment separated from the corresponding wild-type fragment in a manner expected from its predicted melting behavior. The effect of mutations in codons 139 and 141 on the mobility of the 280-bp A-fragment is shown in Fig. 3. DNA with the mutation A551G in lane 1 migrated below the wild-type DNA sequence in lane 3. The fragment with mutation G557T in lane 2 migrated above the wild-type DNA. Lanes 4–6 in Fig. 3 show samples formed by heating equal amounts of wild-type and mutant PCR products or the two mutant PCR products for 3 min at 95 °C and reannealing for 10 min. The two uppermost bands in each lane correspond to heteroduplex DNAs with mismatched base pairs. Separation of the 304-bp A-fragment with mutation T563C from its homologous wild-type fragment was similar to lanes 1 and 3 (not shown). Figure 4 shows a temperature gradient gel comparing the G659A mutant with the wild-type sequence of the 140-bp B-frag-

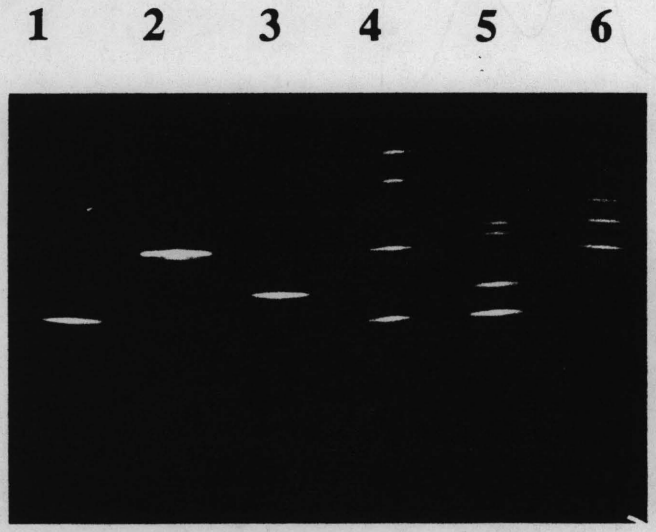


Figure 3. Temperature gradient gel of 280-bp A-fragments; 40–50 °C gradient. Lanes and DNA samples: (1) A551G, (2) G557T, (3) wild type, (4) melted and annealed A661 and G557T, (5) melted and annealed wild type and A551G, (6) melted and annealed wild type and G557T.

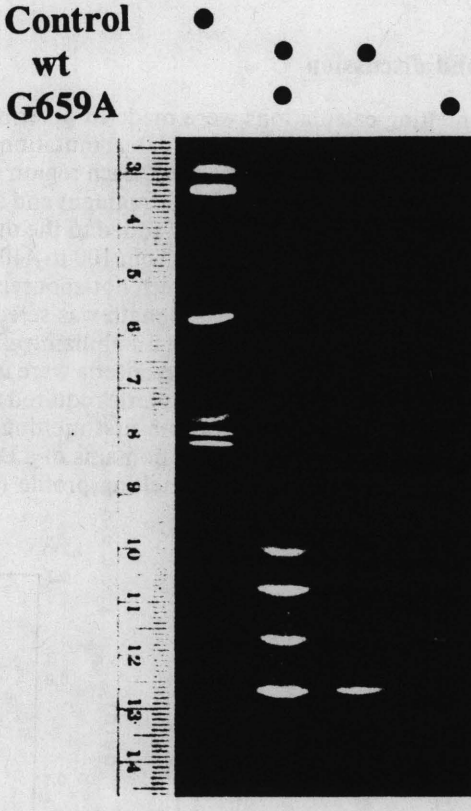


Figure 4. Temperature gradient gel of 140-bp B-fragments; 42–55 °C gradient. Lanes and DNA samples: (1) size marker controls of *Hae*III digest of plasmid pUC19, (2) melted and annealed wild type and G659 A, (3) wild type, (4) G659 A.

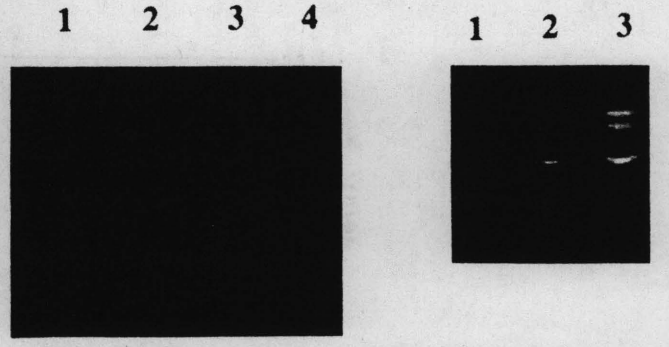


Figure 5. Temperature gradient gel of 440-bp C/D-fragments; 40–48 °C gradient. Lanes and DNA samples, left panel: (1) C877T, (2) G953A, (3) A977G, (4) wild type. Right panel: (1) C877T, (2) G953A, (3) melted and annealed C877T and G953A.

ment. The GC to AT base pair substitution results in a one cm separation. The two heteroduplex DNAs formed by melting and reannealing the wild-type and mutant samples melted higher in the gel. A similar result was observed with the other GC to AT mutation (G670A) in the B-fragment region (not shown). Heteroduplex formation exemplified in Figs. 3 and 4 provides a sensitive means for distinguishing DNAs with different base changes that melt at the same depth in a temperature gradient gel [7]. Figure 5 illustrates this with the 440-bp C/D-fragment. The left panel shows that DNAs with the G953A and C877T mutations melt prior to the wild-type sequence and have the same mobility. The presence of different mutations was shown by melting and reannealing an equimolar mixture of the mutant DNAs (Fig. 5, right panel). The lowest band corresponds to the two mutant DNAs that reannealed to their original homoduplex forms. The upper bands correspond to heteroduplex DNAs.

The ability of TGGE to detect a small amount of mutant DNA in the presence of a larger amount of wild-type DNA was examined. The 280-bp A-fragments with the wild-type sequence and the G557T mutation were mixed in molar ratios from 1:1 to 1:20 (mutant: wild type) and then melted and reannealed. The total DNA concentration was kept constant (100 ng). Figure 6 shows that the heteroduplex bands can be detected by ethidium staining when the mutant DNA is 10% of the total DNA. Although heteroduplexes were deliberately created, they might also be formed when a DNA region is amplified from tissue samples containing neoplastic and nonneoplastic cells. The results may be of relevance in mutational analysis of clinical samples in which the majority of the tissue may be non-neoplastic. They provide a best-case situation for ethidium staining since PCR amplifications from tissue samples are likely to produce less distinct bands.

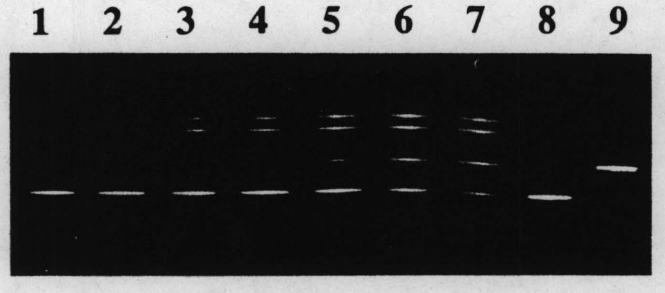


Figure 6. Temperature gradient gel of wild-type and G557T mutant 280-bp A-fragments; 40–50 °C gradient. Total DNA: 100 ng in each lane. Percentage of DNA that is G557T-mutant from left to right (lanes 1, 2, 3 ...): 0, 5, 10, 20, 30, 40, 50, 0, 100.

4 Discussion

The above results indicates that the DNA melting theory can be employed to select fragments for detecting mutations by TGGE. Eight test mutations distributed among three fragments from the p53 gene were successfully detected. In all cases the gel mobility behavior of the mutant DNAs was in qualitative agreement with theoretical expectations. Similar success with regard to predicted gel mobility behavior was previously obtained by TGGE with seven other point mutations in a bacterial DNA fragment [4]. Thus a total of fifteen mutations on four fragments were detected using the criteria described in this work. While the results are promising, a larger number of cases need to be examined to test the generality of the two melting criteria. DNA melting maps or profiles provide information on the regions of a DNA sequence that melt out first. The derivative melting curve criteria serves to ensure that early melting domains are well separated from the strand dissociation step. The latter criteria is also of practical importance since it estimates the minimum-length GC clamp needed to separate an early melting domain from strand dissociation. We estimate that a mutation-induced shift of 0.1 °C can be detected by TGGE based on this and related studies. A DNA melting theory developed specifically for gel conditions may be helpful in predicting which mutations can and cannot be detected.

We thank Dr. B. Vogelstein for providing plasmid DNAs containing p53 sequences. We also thank K. Abbott for expert technical assistance, and S. Hosseini for technical assistance with the 143 codon mutation. This work was supported by a grant to RMW from N.I.H. (GM 38045) and a Veterans Administration grant to VAV.

Received November 3, 1992

5 References

[1] Fischer, S. G. and Lerman, L. S., *Proc. Natl. Acad. Sci. USA* 1983, *80*, 1579–1583.
[2] Myers, R. M., Fischer, S. G., Maniatis, T. and Lerman, L. S., *Nucleic Acids Res.* 1985, *13*, 3111–3129.
[3] Rosenbaum, V. and Riesner, D., *Biophys. Chem.* 1987, *26*, 235–246.
[4] Wartell, R. M., Hosseini, S. H. and Moran, C. P., *Nucleic Acids Res.* 1990, *18*, 2699–2705.
[5] Hovig, B., Smith-Sorrensen, H., Brogger, A. J. and Borresen, A. L., *Mutation Research Let.* 1991, *262*, 63–71.
[6] Borresen, A.-L., Hovig, E., Smith-Sorensen, B., Malkin, D., Lystrad, S., Andersen, T. I., Nesland, J. M., Isselbacher, K. J. and Friend, S. H., *Proc. Natl. Acad. Sci. USA* 1991, *88*, 8405–8409.
[7] Lerman, L. S., Silverstein, K. and Grinfeld, N., *Cold Spring Harbor Symp. Quant. Biol.* 1986, *51*, 285–289.
[8] Lerman, L. S. and Silverstein, K., *Methods Enzymol.* 1987, *155*, 482–501.
[9] Sheffield, V. C., Cox, D. R., Lerman, L. S. and Myers, R. M., *Proc. Natl. Acad. Sci. USA*, 1989, *86*, 232–236.
[10] Wartell, R. M. and Benight, A. S., *Physics Reports* 1985, *126*, 67–107.
[11] Weber, C. K., Shaffer, D. J. and Sidman, C. L., *Nucleic Acids Res.* 1991, *19*, 3331–3335.
[12] Ahuja, H., Bar-eli, M., Advani, S., Benchimol, S. and Cline, M. J., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 6783–6787.
[13] Nigro, J. M., Baker, S. J., Preisinger, A. C., Jessup, J. M., Hostetter, R., Cleary, K., Bigner, S. H., Davidson, N., Baylin, S., Devilee, P., Glover, T., Collins, F. S., Weston, A., Modali, R., Harris, C. C. and Vogelstein, B., *Nature* 1989, *342*, 705–708.
[14] Bressac, B., Galvin, K. M., Liang, T. J., Isselbacher, K. J., Wands, J. R. and Ozturk, M., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 1973–1977.
[15] Bartek, J., Iggo, R., Gannon, J. and Lane, D. P., *Oncogene* 1990, *5*, 893–899.
[16] Mulligan, M., Matlashewski, G. J., Scrable, H. J. and Cavenee, W. K., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 5863–5867.
[17] Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C. C., *Science* 1991, *253*, 49–53.
[18] Poland, D., *Biopolymers* 1974, *13*, 1859–1871.
[19] Fixman, M. and Freire, J., *Biopolymers* 1977, *16*, 2693–2704.
[20] McCampbell, C. R., Wartell, R. M. and Plaskon, R. R., *Biopolymer* 1989, *28*, 1745–1758.

Nucleic Acids Research, 1993, Vol. 21, No. 22 ***5137–5143***

Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA: determination by temperature-gradient gel electrophoresis

Song-Hua Ke and Roger M.Wartell*
School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received July 27, 1993; Revised and Accepted October 4, 1993

ABSTRACT

Temperature-gradient gel electrophoresis (TGGE) was employed to determine the thermal stabilities of 48 DNA fragments that differ by single base pair mismatches. The approach provides a rapid way for studying how specific base mismatches effect the stability of a long DNA fragment. Homologous 373 bp DNA fragments differing by single base pair substitutions in their first melting domain were employed. Heteroduplexes were formed by melting and reannealing pairs of DNAs, one of which was ^{32}P-labeled on its 5′-end. Product DNAs were separated based on their thermal stability by parallel and perpendicular temperature-gradient gel electrophoresis. The order of stability was determined for all common base pairs and mismatched bases in four different nearest neighbor environments; d(GXT)·d(AYC), d(GXG)·d(CYC), d(CXA)·d(TYG), and d(TXT)·d(AYA) with X,Y = A,T,C, or G. DNA fragments containing a single mismatch were destabilized by 1 to 5°C with respect to homologous DNAs with complete Watson–Crick base pairing. Both the bases at the mismatch site and neighboring stacking interactions influence the destabilization caused by a mismatch. G·T, G·G and G·A mismatches were always among the most stable mismatches for all nearest neighbor environments examined. Purine·purine mismatches were generally more stable than pyrimidine·pyrimidine mispairs. Our results are in very good agreement with data where available from solution studies of short DNA oligomers.

INTRODUCTION

Non Watson–Crick or 'mismatched' base pairs occur during DNA replication, genetic recombination and from chemical reactions in cells (1,2,3). The frequency at which a base pair becomes a mutation depends on the frequency of mismatch formation, and the efficiency of mismatch removal by proofreading or repair. Statistical analysis of extant genes and pseudogene sequences indicate that spontaneous mutations do not occur with equal rates for all base pairs (4). The type of base pair substitution and the local sequence environment influence mutation rates. How neighboring base pairs effect the structure and/or stability of a base pair mismatch may be important to understanding the mechanisms that lead to spontaneous mutations.

In recent years, a number of investigations have examined the stability and structure of mismatched base pairs in short DNA duplexes (5–12). Nuclear magnetic resonance spectroscopy, X-ray crystallography, and UV absorbance melting studies have been employed. The stability of all mismatched bases were examined in two sequence environments (6,13,14). Results indicated that the stability and structural properties of a mismatch are influenced by its neighboring base pairs (1). No systematic study has yet to be reported on the effects of different single mismatches on the stability of long DNAs. The influence of end effects on the properties of short DNA duplexes make such a study desirable.

In addition to its interest with regard to mechanisms of spontaneous mutation, the influence of a mismatch on DNA stability is also relevant for methods that rely on thermal stability differences to separate DNAs with similar sequences. Knowledge of how a mismatch alters the stability of a DNA sequence can help optimize conditions in the selective binding of an oligonucleotide to a DNA site (15). TGGE and denaturant gradient gel electrophoresis (DGGE) are becoming widely used to detect single base mutations (16–19). The large effect of mismatched base pairs on DNA stability provides a sensitive means of detecting base pair changes (20). Understanding how a mismatch and its neighbors effect DNA stability can help identify the nature of a mutation.

In this work we have employed a vertical TGGE format (21) to determine the relative stabilities of all possible base pairs and base/base mismatches at four different positions within a 373 bp DNA. DNAs differing in thermal stability in their first melting domain unwind and decrease in mobility at different depths in a polyacrylamide gel with a superimposed temeprature gradient. In combination with site-directed mutagenesis by PCR, the temperature gradient approach provides a rapid method for examining the relative stabilities of mismatches at specific sites within a long DNA. DNAs differing in stability by 0.05 to 0.1°C were separated. Temperature gradients parallel to the direction

* To whom correspondence should be addressed

of electrophoresis were used to determine the relative stabilities of mismatches at a given site in the DNA. They provide the greatest resolution in detecting changes in DNA thermal stability. Experiments in which the temperature gradient was perpendicular to the direction of electrophoresis were used to obtain mobility transition curves.

MATERIALS AND METHODS

Materials

Taq DNA polymerase were obtained from Perkin Elmer and Promega. pUC8-31 and pUC8-36 plasmids were a gift from Dr C.Moran, Emory Univ. The plasmids contain a 130 bp segment of the *ctc* promoter region from *Bacillus subtilis* inserted between the HindIII and EcoRI sites of pUC8 (22). pUC8-31 has the wild type *ctc* sequence, and pUC8-36 has a GC to AT substitution (figure 2). DNA oligonucleotides were from Operon Inc., Alameda, CA. They were used as primers for polymerase chain reaction (PCR) amplification of the 373 bp region containing the *ctc* promoter. Sequences for the twelve upstream primers and one downstream primer are shown in figure 2. All upstream primers except UP14 were used to create a point mutation. Base positions of the mutations are underlined in the primer sequences (figure 2).

The downstream primer, designated as DP15, was end-labeled for some PCR amplifications with ^{32}P. 3 μl of γ-labeled ATP (3000 Ci/mmol, Amersham) was mixed with 1 μl (10 units) polynucleotide kinase (Promega), 1 μl 10× kinase buffer (400 mM Tris−HCl, pH 7.5, 100 mM $MgCl_2$, 50 mM DTT) and 1 μl of 10 μM primer and 4 μl of water. The mixture was incubated at 37°C for 30 mins, heated at 65°C for 5 mins and purified with a Nensorb™-20 cartridge (Du Pont).

PCR amplifications

The PCR conditions were similar to the protocol recommended by Perkin Elmer Cetus Inc. 100 μl reaction mixtures contained 50 pg of plasmid DNA, 0.6 μM of each primer, and 200 μM of each dNTP in a buffer of 10 mM Tris−HCl pH 8.3, 50 mM KCl and 2.5 mM $MgCl_2$. Reaction mixtures were overlaid with 100 μl mineral oil and thirty cycles of amplification carried out. The temperature cycles were 94°C for 1 min (except for a 4 minute first cycle), 44°C for 2 min and 72°C for 1 minute. 2−4 μl of each reaction was checked for size and purity on a 1% agarose or 7.5 % polyacrylamide gel. The PCR amplification was carried out with a 5′-end-labeled downstream primer when a labeled DNA was required. All 373 bp DNAs with or without mismatches ran with the same mobility in non-denaturing polyacrylamide gels.

TGGE

The apparatus for running the vertical temperature-gradient gel was described previously (21). A 6.5% polyacrylamide gel (37.5:1 , acrylamide:bisacrylamide) was used. The gel contained 4.2 M urea and 24% vol/vol formamide in 0.5× TBE (0.045 M sodium borate + 0.045 M Tris + 1 mM EDTA, pH 8.2). Formamide was deionized with mixed resin AG501-X8D (Bio-Rad). The gel running buffer was 0.5× TBE. Two aluminum heating blocks sandwiched the glass plates and established the temperature gradient either parallel or perpendicular to the electric field. Temperatures were measured at various positions in several test gels with a thermocouple probe (see below). The temperature gradient was linear and uniform within the region covered by the heating blocks.

The relative stability of DNAs were determined with the temperature gradient parallel to the direction of electrophoresis. DNA samples migrated from low temperature (top) to higher temperature (bottom) from 1 cm wells. For perpendicular temperature gradient gels, the DNA samples were loaded into a long well along the top of the gel. The electrophoretic direction was perpendicular to the temperature gradient. Mobility transition curves of duplex DNAs to their partially denatured states were detected as a decrease in mobility with increasing temperature (see figure 7).

Temperatures were evaluated in the gels with a needle-like thermocouple probe (TMTSS-020-6, Omega Inc.) connected to a digital thermometer (MDSD-465, Omega Inc, accuracy estimated as ± 0.1°C). Measurements were made at two position at the end of each transition run. The gels were stained with ethidium bromide and photographed. The positions where the temperature probe had been inserted were observed in the photograph and provided a temperature scale.

Photographs of the mobility transition curves were digitized and scaled using a digitizer tablet (SummaSketch II), and transmitted to a microcomputer. Transition curves were smoothed by the 'smoothlowess' function in the Axum graphics analysis package (Trimetrix Inc., Seattle, Wash.). This is based on a locally weighted regression analysis (23). The mobility transition temperature, T_u, was defined as the temperature at the peak of the derivative curve. The standard deviation of T_u measurements based on repeated experiments was 0.6°C. Differences in T_u values between two DNA transitions in a gel were reproducable within ± 0.2°C. This error estimate is based on three or four repeated measurements.

Electrophoresis conditions for several runs are described in the figure legends. In general, run times were 14 to 18 hours (overnight) at 4.5 to 6 volts/cm for the 20 cm long gels. Perpendicular temperature gradient gel results show that the 373 bp DNA melts out in several transition steps (not shown). The mobility transition temperature of the first melting domain, with Watson−Crick base pairs, is between 30 and 33.5°C under the gel conditions employed. Temperature gradients from 28.5 to 31.5°C or 28.5 to 32°C were used to optimize separation of DNAs with single base pair substitutions. A gradient from 26 to 29°C was used to optimize separation of DNAs with mismatched bases.

RESULTS

Figure 1 describes the method employed to determine the relative stability of DNAs differing by single base pairs or mismatches. Two DNAs differing by a base pair substitution were produced by PCR. One DNA was ^{32}P-labeled on the 5′-end of its downstream primer strand. The DNAs were heated for three minutes at 97°C, reannealed at 54°C for at least 10 minutes, and allowed to slowly cool to room temperature. The four resulting DNAs were analysed by parallel TGGE. Following electrophoresis, the gel was stained with ethidium bromide. DNA bands were located on a UV-transilluminator, excised, and their radioactivity measured by scintillation counting. Identities of the DNA bands were established from the radioactively labeled bands. Confirmation of band identities was made by switching the DNA that contained the labeled strand and/or by running one of the homoduplex DNAs in an adjacent lane.

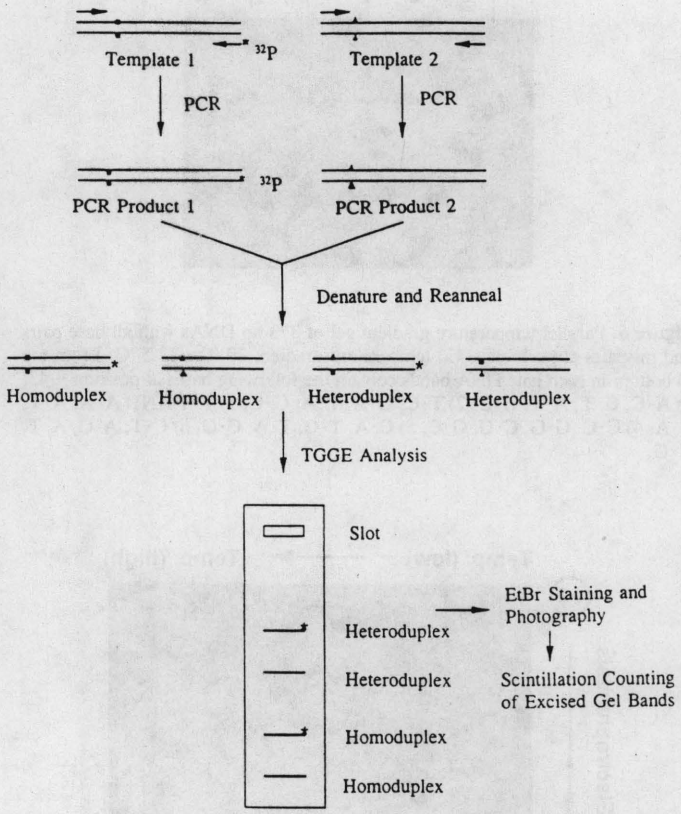


Figure 1. Protocol for determination of mismatch by PCR and TGGE. Heteroduplexes containing a single base-base mismatch were formed by melting and reannealing equal amounts of two PCR fragments, one with its downstream strand 5′-end labeled with ^{32}P. TGGE was used to separate the homoduplex and heteroduplex DNAs for analysis.

```
Upstream Primers:
      5'                                    3'
UP36C:    AATTCCATTTTTCGAGCTTTA
UP36T:    AATTCCATTTTTCGAGTTTTA
UP38T:    AATTCCATTTTTCGTGGTTTA
UP38G:    AATTCCATTTTTCGGGGTTTA
UP38C:    AATTCCATTTTTCGCGGTTTA
UP39T:    AATTCCATTTTTCTAGGT
UP39C:    AATTCCATTTTTCCAGGT
UP39A:    AATTCCATTTTTCAAGGT
UP43G:    AATTCCATTGTTCGAGGT
UP43C:    AATTCCATTCTTCGAGGT
UP43A:    AATTCCATTATTCGAGGT
UP14:     AATTCCATTTTTCG
                  ↕   ↕↕ ↕                                    ↓
      5'  AATTCCATTTTTCGAGGTTTAAATCCTTATCGTTATGGGTATTGTTTGTAATAGGACAA
                              A (pUC8-36)
                  -43    -36                                  +1

          CTAAAACGACAAGAGGATGGTGCTGAATATGGCAACTTTAACGGCAAAAGAAAGAACGG

          ACTTTACTCGGTCGACCTGCAGCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTG

          ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCC

          AGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCT

          GAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCAC

          ACCGCATATGGTGCACTCT  3'

           3' GTATACCACGTGAGA  5' DP15 (Downstream Primer)
```

Figure 2. The 373 bp DNA sequence between the EcoRI and RsaI sites from pUC8-31 plasmid is shown. Positions −43 and −36 are indicated. The DNA fragment from the plasmid pUC8-36 has the same sequence except for a G to A substitution at position −36. The upstream primers, and the downstream primer, DP15, employed in PCR are indicated. Upstream primers created base pair changes at the positions underlined.

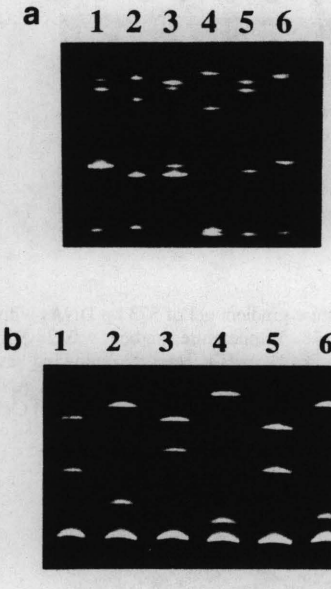


Figure 3. Parallel temperature gradient gel of 373 bp DNAs with all base pairs and mispairs at position −36. (**a**). Temperature gradient was from 28.5 to 33°C. Samples were run for 17 hrs at 90 volts. From the top to bottom in each lane DNA bands contain the following bases at position −36: 1) A·C, G·T, A·T, G·C, 2) T·C, G·A, T·A, G·C, 3) T·T, A·A, A·T, T·A, 4) C·C, G·G, G·C, C·G. 5) C·A, T·G, T·A, C·G and 6) C·T, A·G, A·T, C·G. (**b**). Temperature gradient was from 26 to 29°C. Run time and voltage were the same as above. The top two DNA bands in each each lane are the same as in (a). The lowest band in each lane contains both homoduplex DNAs, e.g., lane 1) A·C, G·T, A·T AND G·C.

The 373 bp DNAs used in the study are indicated in figure 2. PCR were used to generate thirteen DNAs differing from each other by a single base pair. The base pair changes occurred at four sites designated −36, −38, −39, and −43. This numbering scheme refers to base pair positions relative to the startpoint of transcription for the *ctc* promoter in the 373 bp DNA. Each site is located in the first melting domain of the DNA (21). Pairs of DNAs were melted and reannealed to produce 48 DNAs that contained all possible base mismatches at the four different sites each with a different base pair stacking environment.

Figures 3a and 3b show parallel TGGE experiments of 373 bp DNAs with all possible paired and mismatched bases at position −36. The nearest neighbor pairs surrounding this position are d(GXT)·d(AYC). The identity of the DNA bands in figures 3a and 3b are given in the figure caption and were based on the procedures described earlier. Figure 3a used a temperature gradient from 28.5 to 33°C to optimize the separation of base paired and mismatched DNAs in one gel. Figure 3b used a gradient from 26 to 29°C to optimize separation of the DNAs with mismatched bases. The lower temperatures of figure 3b sacrificed the ability to separate the base paired DNAs in order to determine the order of stability for all mismatched DNAs. In figure 3a, for example, it is difficult to order the relative stability of DNA bands containing A·A and T·T mismatches in lane 3 with the C·A and T·G bands in lane 5. Lane 3 of figure 3b shows a much greater separation between the DNAs with A·A and T·T mismatches. It is possible to rank their stability relative to the DNAs with C·A and T·G mismatches in lane 5.

Figures 4, 5 and 6 show parallel TGGE experiments of the 373 bp DNA with all possible paired and mismatched bases at

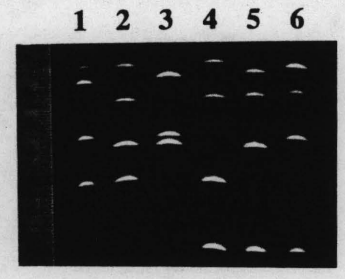


Figure 4. Parallel temperature gradient gel of 373 bp DNAs with all base pairs and mispairs at position −38. Temperature gradient was 28.5 to 33°C. From top to bottom in each lane DNA bands contain the following bases at −38: 1) A·C, G·T, A·T, G·C, 2) T·C, G·A, T·A, G·C, 3) A·A AND T·T, A·T, T·A, 4) C·C, G·G, G·C, C·G, 5) C·A, T·G, T·A, C·G, 6) C·T, A·G, A·T, C·G.

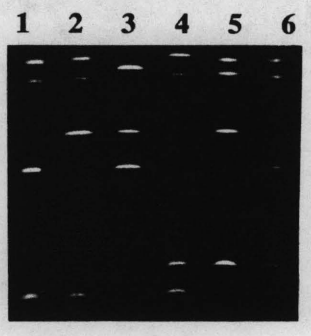


Figure 5. Parallel temperature gradient gel of 373 bp DNAs with all base pairs and mispairs at position −39, temperature gradient 29.0 to 32.5°C. From top to bottom in each lane DNA bands contain the following bases at position −39: 1) A·C, G·T, A·T, G·C, 2) T·C, G·A, T·A, G·C, 3) T·T AND A·A, T·A, A·T, 4) C·C, G·G, C·G, G·C, 5) C·A, T·G, T·A, C·G, 6) C·T, A·G, A·T, C·G.

positions −38, −39 and −43 respectively. The base pairs surrounding these positions are d(GXG)·d(CYC) for −38, d(CXA)·d(TYG) for −39, and d(TXT)·d(AYA) for −43. Four bands are observed in each lane of these figures except for the third lane. The top band in the third lane of each figure contains the heteroduplex DNAs with A·A and T·T bases. A temperature gradient from 26 to 29°C was able to separate this band and show that the A·A mismatch is slightly more stable than the T·T mismatch at position −43 (not shown). This temperature gradient was unable to separate the band containing these mismatches at positions −38 and −39.

Table 1 summarizes the results from the parallel TGGE experiments. All homoduplex DNAs were more stable than their corresponding heteroduplex DNAs. This table and figures 3−6 show that both the bases at a mismatched site and the neighboring stacking interactions influence the destabilization caused by a mismatch. The least destabilizing base in a mismatch is G and the most destabilizing base to have in a mismatch is C. A similar observation was made in a systematic study of mismatches in the oligomer d(CT_3XT_3G)·d(CA_3YA_3G) (6). In general our results show that purine·purine (both homo- and hetero-) mismatches are more stable than pyrimidine·pyrimidine mismatches.

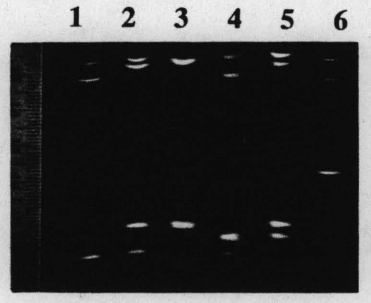


Figure 6. Parallel temperature gradient gel of 373 bp DNAs with all base pairs and mispairs at position −43, temperature gradient 29.0 to 32.5°C. From top to bottom in each lane DNA bands contain the following bases at position −43: 1) A·C, G·T, A·T, G·C, 2) T·C, G·A, T·A, G·C, 3) T·T AND A·A, A·T, T·A, 4) C·C, G·G, C·G, G·C, 5) C·A, T·G, T·A, C·G, 6) C·T, A·G, A·T, C·G.

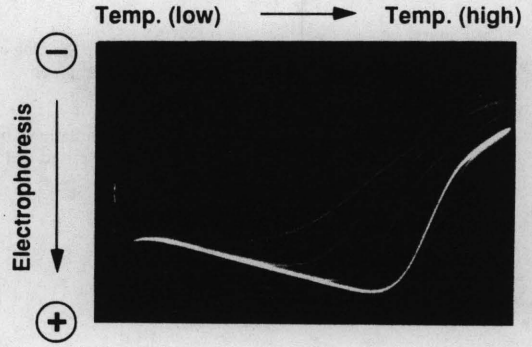


Figure 7. A typical perpendicular temperature gradient gel of 373 bp PCR fragments. Electrophoresis was conducted for 14 hrs at 90 volts. The temperature gradient was 17 to 35.5°C from left to right. A total of 3−5 μg of DNA was added across the top of the gel. The sample contained melted and reannealed DNAs with C·G and G·C at position −43, and the native DNA with T·A at the same position. The transitions from left to right correspond to DNAs with the following base pairs at position −39: C·C, G·G, T·A, C·G, G·C.

The most stable mismatch is different for different nearest-neighbor environments. G·T is the most stable mismatch for positions −39 and −43. G·A and G·G are the most stable mismatches for positions −38 and −36 respectively. NMR and X-ray crystallographic studies indicate that both G·T and G·A pairs can form two hydrogen bonds and stack within a DNA B-conformation duplex with relatively little distortion (1). Studies on DNA duplex oligomers with G·A mismatches indicate that intrahelical base pairing occurs but that the nature and extent of helix distortion is strongly sequence dependent (7,24). Our results confirm a sequence dependence on the properties of the G·A mismatch.

Results from the TGGE studies are in good agreement with available data on mismatch stabilities in short DNA oligomers. NMR studies (24) on d(CGXGAATTCYCG) where X·Y formed the pairs T·G, A·G, C·A, or C·T indicated an order of stability of T·G > A·G > C·A > C·T. We observed the same heirarchy in the equivalent nearest neighbor environment at position −38 (Table 1). UV absorbance melting curves were obtained by Aboul-ela *et al.* (6) from the oligomers d(C-T_3XT_3G)·d(CA_3YA_3G) and by Gaffney and Jones (14) from the

Table 1. Comparison of Watson–Crick and mismatched base pair stabilities

Base Pair Position	5' Flanking Base Pair	3' Flanking Base Pair	The Ranking of Stability of Normal and Mismatched Base Pair
-36	G·C	T·A	C·G > G·C > T·A > A·T > G·G > A·G > G·A > G·T > T·G > A·A > C·A > T·T ≥ A·C > T·C ≥ C·T > C·C
-38	G·C	G·C	C·G > G·C > T·A > A·T > G·A > G·G > T·G > A·G > G·T > A·A = T·T > C·A > C·T, A·C > T·C > C·C
-39	C·G	A·T	G·C > C·G > A·T > T·A > G·T > G·A > G·G > A·G ≥ T·G > A·A = T·T > A·C > C·A, T·C, C·T > C·C
-43	T·A	T·A	G·C > C·G > T·A > A·T > G·T > G·G, A·G > G·A, T·G > A·A > T·T > A·C > T·C > C·C, C·A, C·T

oligomers $d(G_2T_2XT_2G_2)\cdot d(C_2A_2YA_2C_2)$ with X and Y substituted by all four DNA bases. Both studies had a similar heirarchy of stability for mismatched DNAs. The Gaffney and Jones work yielded the following order of stabilities based on T_{max}, the peak of the derivative melting curves: G·T > G·G = A·G > T·G > G·A = T·T > T·C > A·C > C·T > A·A > C·A > C·C. Table 1 shows a similar but not identical rank order of stability at position −43. The most significant difference is the relative ranking of A·A which is more stable in the DNA fragment than in the oligomers. The differences are not due to mobility differences of mismatched DNAs. In the absence of temperatures sufficient for melting, all DNA fragments have the same mobility. This is best illustrated in the perpendicular gels such as figure 7. Prior to the onset of melting the DNAs had the same mobility. Potential causes for the differences are discussed below.

Table 1 also provides direct information on the relative stabilities of Watson–Crick base pairs in four stacking environments inside a long DNA. DNAs with G·C or C·G base pairs were more stable than those with A·T or T·A pairs in all cases examined. The relative stability of several base pair stacking interactions was consistent with observations from DNA polymer melting studies. Solution studies have shown that poly (dA)·poly (dT) is more stable than poly (dAT)·poly (dAT), and poly (dGC)·poly (dGC) is more stable than poly (dG)·poly (dC) (25). These observations are consistent with the greater stability of the d(TTT)·d(AAA) sequence at position −43 relative to d(TAT)·d(ATA), and the enhanced stability of d(GCG)·d(CGC) at position −38 relative to d(GGG)·d(CCC). Similarly, the greater stability of the d(GTG)·d(CAC) sequence at position −38 when compared to d(GAG)·d(CTC) is consistent with the higher T_m of poly (dGT)·poly (dAC) when compared to poly (dGA)·poly (dTC) (25). The above agreement can not be automatically expected since the solvent employed in TGGE differs from salt solutions commonly used in UV absorbance melting studies.

Perpendicular TGGE experiments display the mobility transitions of intact double-stranded DNA to the denatured state. Figure 7 shows transition curves of five of the DNAs examined. The initial increase in mobility with increasing temperature prior to the main transition is due to the effect of temperature on the gel. The sigmoidal decrease in mobility is due to the unwinding of DNA strands. The temperature range was selected to include the first melting domain of the 373 bp DNAs. The two leftmost transitions in figure 7 correspond to DNAs contain C·C or G·G mismatches at position −39. The middle transition corresponds to the DNA with the T·A base pair at position −39. The two closely spaced transitions on the right correspond to DNAs with C·G and G·C base pairs at position −39. The curve of the C·C mismatch DNA is much broader then the other transitions. This observation was common for the transitions involving DNAs with mismatched base pairs of low stability. An analogous observation was made by UV absorbance melting studies of DNA oligomers with mismatches (13), and may indicate that internal melting plays a significant role in the melting process.

Table 2. Perpendicular TGGE [a]T_u measurements of 373 bps DNA first melting domain

-36 set (°C)		-38 set (°C)		-39 set (°C)		-43 set (°C)	
C·G:	31.8	C·G:	33.4	[b]G·C:	31.8	G·C:	32.4
[b]G·C:	31.8	G·C:	32.6	C·G:	31.6	C·G:	31.9
T·A:	31.0	T·A:	32.1	A·T:	30.6	[b]T·A:	31.8
A·T:	30.9	[b]A·T:	31.8	T·A:	30.1	A·T:	31.2
G·G:	29.6	G·A:	30.6	G·T:	28.8	G·T:	29.8
A·G:	29.5	G·G:	30.4	G·A:	28.8	G·G:	29.7
G·A:	29.4	T·G:	30.3	G·G:	28.7	A·G:	29.7
G·T:	28.8	A·G:	30.1	A·G:	28.6	G·A:	29.5
T·G:	28.5	G·T:	29.8	T·G:	28.6	T·G:	29.5
A·A:	28.3	A·A:	29.5	A·A:	28.5	A·A:	29.3
C·A:	28.1	T·T:	29.5	T·T:	28.5	T·T:	29.3
T·T:	28.0	C·A:	29.4	A·C:	28.4	A·C:	29.2
A·C:	28.0	C·T:	29.0	C·A:	28.2	T·C:	29.2
T·C:	27.9	A·C:	29.0	T·C:	28.2	C·C:	29.1
C·T:	27.9	T·C:	28.7	C·T:	28.2	C·A:	29.1
C·C:	27.9	C·C:	28.5	C·C:	28.1	C·T:	29.1

[a] Tu was defined as the temperature at the peak of the derivative curve calculated from the smoothed DNA mobility transition profile. Estimated precision in T_u relative to the standard DNA fragment is ± 0.2°C. Each experiment contained the pUC8-31 DNA fragment as an internal standard. The mean Tu of this DNA was 31.8°C (± 0.6°C) based on 19 repeated experiments.
[b] This is the same pUC8-31 DNA fragment.

Table 3. Comparison of DNA destabilization by single-base mismatch in different nearest-neighbor environments

Base Pair Position	DNA SEQUENCE	*ΔT_u (°C)	
		T_u(G·C)-T_u(C·C)	T_u(G·C)-T_u(G·G)
-36	d(GXT)·d(AYC)	3.9	2.2
-38	d(GXG)·d(CYC)	4.1	2.2
-39	d(CXA)·d(TYG)	3.7	3.1
-43	d(TXT)·d(AYA)	3.3	2.7

* T_u differences are obtained by subtracting T_u values of DNA duplexes with C·C or G·G mismatch from T_u values of DNA duplexes with G·C base pair at the same position.

Transition temperatures, T_u, for transitions of all of the base paired DNAs and DNAs with mismatches are given in Table 2. T_u values of DNA molecules with a single mismatch are lower than the corresponding values for homoduplex DNAs by 1 to 5°C. The order of stability for each position is based on results from Table 1 since the resolution of the parallel gradient gel was better than perpendicular temperature gradient gels. All of the transition curve data were consistent with the results from the parallel TGGE. Table 3 compares the destabilization caused by two different base changes in the four nearest neighbor environments. It illustrates that the destabilization caused by a mismatch depends on its nearest neighbor base pairs. Converting a G·C base pair to a C·C has its largest effect in the d(GXG)·d(CYC) environment. Changing a G·C to a G·G has its largest effect in the d(CXA)·d(TYG) environment.

DISCUSSION

TGGE provides a rapid method to characterize the relative stability of different base pairs and mismatched bases at specific sites within a DNA molecule. Agreement is observed between the heirarchy of Watson–Crick base pair stacking interactions from TGGE and DNA polymer melting studies. The presence of urea and formamide does not appear to cause a major alteration in the heirarchy of base pair stacking interactions. This is consistent with previous melting studies that indicate urea and formamide lower the thermal stability of GC and AT base pairs by approximately equivalent amounts (16,26,27,28,29).

The heirarchy of mismatch stabilities in Table 1 is also in relatively good agreement with data available from DNA oligomer melting studies in aqueous solutions. For the d(GXG) environment the heirarchy of stability is the same as that observed for four mismatches studied in a DNA oligomer (24). The most stable and least stable mismatches we observe for the d(TXT)·d(AYA) environment are also among the most and least stable mismatches observed in DNA oligomer melting studies (6,14). However, as described earlier, the stability ranking of several mismatches in this environment differ between the fragment and oligomer data. The discrepancies do not appear to be from measurement uncertainties. TGGE indicate that the G·T mismatch is more stable than the G·G mismatch. This agrees with the 2.3°C separation in melting curves of related DNA oligomers (14). Yet TGGE can not separate DNA fragments with G·A and T·G mismatches, although DNA oligomer melting studies show the T·G mismatch to be more stable than G·A by 2.8°C (6,14).

Several possible causes exist for the descrepancies in the mismatch rank order between the oligomer and fragment data for d(TXT)·d(AYA). An obvious possibility is the difference in solutions employed. The urea-formamide solution may alter the relative stability of some mismatches when compared to NaCl solutions used for the oligomer studies. Another consideration is the oligomer vs. polymer contexts of the studies. Different mismatches may differentially influence the duplex-strands dissociation step which is assumed to be a two state process for all oligomers. This step dominates DNA oligomer melting. It is absent in the unwinding of a fragment's first melting domain.

The influence of the twelve possible mismatched bases on DNA stability varies with nearest neighbor environment. For the four sites examined G·T, G·G, and G·A pairs are always among the most stable mismatches, and the pyrimidine·pyrimidine mismatches are among the least stable. However the specific mismatch that creates the most or least instability depends on the neighboring sequence (Table 3). Results from this and related work should be of value for methods that utilize DNA duplex formation for sequence-specific recognition. Competitive oligonucleotide priming (15) and related methods require conditions that maximize the difference in stability between a completely complimentary DNA duplex and a duplex with one mismatch. Results from this study may also be of value in characterizing a base pair substitution detected by TGGE or DGGE. The pattern of bands produced by melting and reannealing two DNA molecules differing by a single base pair substitution depends on the base pair change and its nearest neighbor pairs. One may be able to characterize the base pairs neighboring the mutant site as well as the type of base pair change that has occurred.

This study illustrates the ressolving power of vertical TGGE and its sensitivity to the temperature gradient employed. The most effective gradient in our parallel TGGE experiments was 3 to 4°C. This gradient was spread over the 15 cm height of the heating blocks. One can readily distinguish DNA bands separated by 2 mm. Thus the parallel TGGE experiment can separate DNAs differing in stability by 0.05°C. Figures 3a and 3b demonstrate how changes in the midpoint temperature of the gradient allow for the separation of DNA bands differing by a base pair or by a mismatch. In our experiments samples of melted and reannealed DNAs differing by a single base pair substitution produced three to four bands for gradients of 8°C or less. The temperature range selected was an important parameter in determining which bands separated.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the technical assistance of Thomas Maier in the physics department of Georgia Tech. This work was supported by N.I.H. grant GM38045.

REFERENCES

1. Modrich,P. (1987) *Annu. Rev. Biochem.*, 56, 435–466.
2. Akiyama,M., Maki,H., Sekiguchi,M. and Horiuchi,T. (1989) *Proc. Natl. Acad. Sci. USA.*, 86, 3949–3952.
3. Singer,B., Chavez,F., Goodman.M.F., Essigmann,J.M., Dosanjh,M.K. (1989) *Proc.Natl.Acad. Sci. USA*, 86, 8271–8274.
4. Blake,R.D., Hess,S.T. and Nicholson-Tuell,J. (1992) *J. Mol. Evol.* 34, 189–200.
5. Brown,T., Leonard,G.A., Booth,E.D. and Kennard,O. (1990) *J. Mol. Biol.*, 212, 437–440.
6. Aboul-ela,F., Koh,D., Tinoco,I.Jr., and Martin,F.H. (1985) *Nucleic Acids Res.*, 13, 4811–4824.
7. Brown,T., Hunter,W.N., Kneale,G.G. and Kennard,O. (1986) *Proc. Nat. Acad. Sci. USA*, 83, 2402–2406.
8. Gao,X. and Patel,D.J. (1988) *J. Am. Chem. Soc.*, 110, 5178–5182.
9. Patel,D.J., Kozlowski,S.A., Marky,L.A., Rice,J.A., Broka,C., Dallas,J., Itakura,K. and Breslauer,K.J. (1982) *Biochemistry* 21, 437–444.
10. Roongta,V.A., Jones,C.R. and Gorenstein,D.G. (1990) *Biochemistry*, 29, 5245–5258.
11. Kneale,G., Brown,T., Kennard,O. and Rabinovich,D. (1985) *J. Mol. Biol.*, 186, 805–814.
12. Arnold,F.H., Wolk,S., Cruz,P. and Tinoco,I.J. (1987) *Biochemistry*, 26, 4068–4075.
13. Werntges,H., Steger,G., Riesner,D. and Fritz,H-J. (1986) *Nucleic Acids Res.*, 14, 3773–3790.
14. Gaffney,B.L. and Jones,R.A. (1989) *Biochemistry*, 26, 5881–5889.
15. Gibbs,R.A., Nquyen,P-N., Caskey,C.T. (1989) *Nucleic Acids Res.*, 17, 2437–2448.
16. Lerman,L.S., Fischer,S.G., Hurley,I., Silverstein,K. and Lumelsky,N. (1984) *Annu. Rev. Biophys. Bioeng.*, 13, 399–423.

17. Guldberg,P. and Guttler,F. (1993) *Nucleic Acids Res.*, 21, 2261–2262.
18. Ke,S-H., Kelly,P.J., Wartell,R.M., Hunter,S.H. and Varma,V.A. (1993) *Electrophoresis*, **14** 561–565.
19. Riesner,D., Henco,K., and Steger,G. (1991) In Chrambach,A., Dunn,M.J., and Radola,B.J. (eds.), Advances in Electrophoresis. VCH Pub., New York, Vol. 4,pp.171–250.
20. Abrams,E.S., Murdaugh,S.E., and Lerman,L.S., (1990) *Genomics* 7 463–475.
21. Wartell,R.M., Hosseini,S.H., and Moran,J.D. (1990) *Nucleic Acids Res.*, 18, 2699–2705.
22. Tatti,K.M. and Moran,C.P.Jr. (1985) *Nature*, 314, 190–192.
23. Cleveland,W.S. and Devlin,S.J. (1988) *J. Amer. Stat. Assoc.*, 83, 596–610.
24. Patel,D.J., Kozlowski,S.A., Ikuta,S. and Itakura,K. (1984) *Federation Proc.*, 43, 2663–2670.
25. Wells,R.D., Wartell, R.M. (1974) In Burton, K.(ed.), Biochemistry of Nucleic Acids. Butterworth & Co., London, Vol. 6, pp. 41–64.
26. Nishigaki,K., Husimi,Y., Masuda,M., Kaneko,K. and Tanaka,T. (1984) *J. Biochem.*,95, 627–635.
27. Hutton, J. R. (1977) *Nucleic Acids Res.*, 4, 3537–3555.
28. Klump,H., and Burkart,W. (1977) *Biochim. Biophys. Acta*, 475, 601–604.
29. Casey,J. and Davidson,N. (1977) *Nucleic Acids Res.*, 4, 1539–1552.

Nucleic Acids Research, 1994, Vol. 22, No. 20 4321–4325

A program for selecting DNA fragments to detect mutations by denaturing gel electrophoresis methods

Stephen Brossette and Roger M.Wartell*
School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received April 25, 1994; Revised and Accepted August 22, 1994

ABSTRACT

A computer program was developed to automate the selection of DNA fragments for detecting mutations within a long DNA sequence by denaturing gel electrophoresis methods. The program, MELTSCAN, scans through a user specified DNA sequence calculating the melting behavior of overlapping DNA fragments covering the sequence. Melting characteristics of the fragments are analyzed to determine the best fragment for detecting mutations at each base pair position in the sequence. The calculation also determines the optimal fragment for detecting mutations within a user specified mutational hot spot region. The program is built around the statistical mechanical model of the DNA melting transition. The optimal fragment for a given position is selected using the criteria that its melting curve has at least two steps, the base pair position is in the fragment's lowest melting domain, and the melting domain has the smallest number of base pairs among fragments that meet the first two criteria. The program predicted fragments for detecting mutations in the cDNA and genomic DNA of the human p53 gene.

INTRODUCTION

Denaturing gel electrophoresis methods are widely used to detect point mutations or polymorphisms in DNA fragments. The methods include denaturing gradient gel electrophoresis (DGGE; 1,2), temperature gradient gel electrophoresis (TGGE) in both horizontal (3,4) and vertical formats (5,6), constant gradient gel electrophoresis (CDGE; 7) and temperature sweep gel electrophoresis (TSGE; 8).

The above methods utilize DNA melting properties to separate fragments differing by as little as 1 bp substitution or mismatch. A duplex DNA migrates in a polyacrylamide gel until it reaches a temperature or denaturant concentration which induces the least stable domain to unwind. The branched structure of a partially denatured DNA results in a large decrease in gel mobility. Homologous DNAs with different stabilities in their first melting domain unwind at different depths in the gel. Base pair changes are detectable if they are in the first melting domain and the domain unwinds at a temperature or denaturant concentration well below the last stage of strand unwinding. Although mutations may also be detected in all but the last temperature domain (9), we have focused attention on first domain melting (6).

Computer simulations of DNA melting behavior are frequently employed to predict if a DNA fragment can be used to detect mutations. DNA melting maps (1) or melting profiles (6) can determine if a designated base pair or base pair region is in an early melting domain, and derivative melting curves determine if a melting domain is separated from the final melting step. A G+C rich segment (a GC clamp) is often added to one end of a DNA to stabilize it and promote a two-step melting process (10,11). Computer simulations provide a way to select a fragment for detecting mutations with a minimum GC clamp length. Simulations will also be useful if psoralen chemi-clamps are employed (12). In the latter case, modifications to the theory are required to accommodate an effective T_m for the chemi-clamp, and the elimination of strand dissociation.

The current approach for selecting a DNA fragment involves calculating the melting behavior of trial fragments surrounding the region of interest. This procedure is straightforward when applied to one or a few regions. It can be laborious if one wishes to determine a large set of DNA fragments to detect mutations throughout a gene. The program described in this paper scans through a long DNA sequence and selects optimum fragments for detecting mutations at each base pair position. The program also selects the optimum fragment for detecting mutations within a user specified region in the sequence. The program, called MELTSCAN™, should be useful to workers applying DGGE, TGGE and related methods to detect mutations in DNA.

METHODS AND RESULTS

MELTSCAN is based on the statistical mechanical model of the DNA helix–coil transition (1,6,13,21). In this model the stability of each base pair depends on its base pair type (AT or GC) and stacking interactions with neighboring base pairs. The model includes a loop entropy term for opening an internal segment of base pairs, and the dissociation equilibrium of partial duplex to single strands. Thermodynamic parameters employed were from recent DNA melting studies; stacking interactions from Delcourt and Blake (14), dissociation parameters and loop entropy from McCampbell *et al.* (15). DNA melting transitions were

*To whom correspondence should be addressed

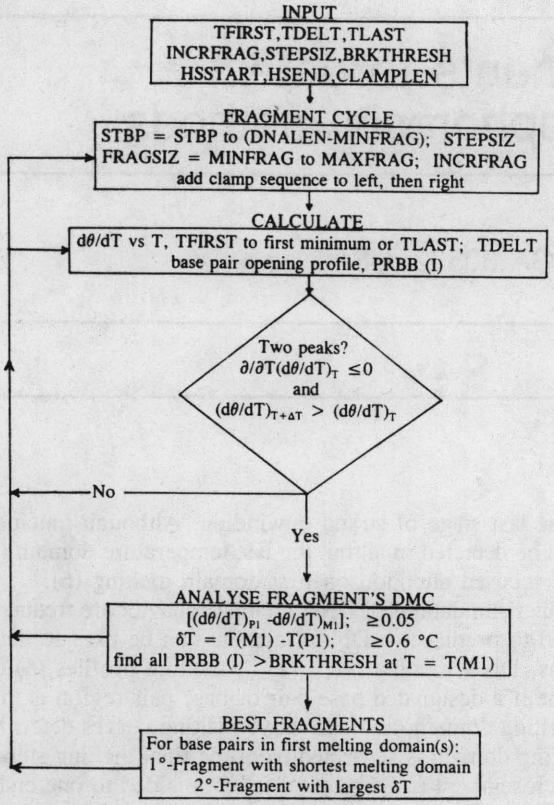


Figure 1. Flow chart outlining the parameters and procedures used in the computer program MELTSCAN. See text for definition of parameters and description of program.

calculated for a solution of 0.1 M Na^+. To estimate the mid-temperature in an experimental temperature gradient gel with 60 % denaturant (4.2 M urea and 24% formamide), 46°C was subtracted from the predicted T_m of the first melting domain (6). This empirical shift is consistent with observations that urea and formamide lower DNA melting temperatures by an amount that is, to a first approximation, independent of base composition (1,9).

Computer algorithm

Figure 1 shows the flow diagram of MELTSCAN. The program was written in FORTRAN-77 and compiled using the Microsoft Corp. Powerstation software. The hardware required is an IBM-PC or IBM-PC compatible personal computer with a 80386 or higher processor running an MS-DOS version 3.3 or later operating system. The Microsoft Corp. Powerstation run-time file DOSXMSF.EXE is also required. Program execution initiates from a batch file with the command 'MELTSCAN SEQNAME'. SEQNAME is the user-defined DNA sequence file. The first line of this file is for descriptive information of the sequence. The remaining lines list the DNA sequence to be analyzed. Blank spaces, numbers, or characters other than A, T, C, and G in the second and subsequent lines of SEQNAME are ignored. The following paragraphs describe how the program works. A user-oriented description of an interaction with the program with sample input parameters and acceptable parameter limits is shown in Table 1.

Table 1. Typical user application of MELTSCAN. User input underlined. Comments within (...)

```
C:>MELTSCAN p53cdna <CR>   (The sequence file here is called p53cdna)
                           (Sequence file must not have an extension in name)

MELTSCAN DNA SEQUENCE ANALYSIS PROGRAM
          .          .
          .          .
TYPE IN VALUES FOR
TFIRST,TDELT,TLAST
78,0.5,94 <CR>             (Values for TDELT; 0.1 to 0.5)
                           (Values for TFIRST/TLAST; 75-80 °C/93-99°C)
                           (TFIRST & TLAST may be adjusted to suit %GC of fragments)
                           (78 to 94 °C works for 46 < %GC < 68)
                           (Extremes of TFIRST AND TLAST may cause program to crash)
TYPE IN VALUES FOR
STEPSIZ,INCRFRAG,BRKTHRESH
20,20,0.5<CR>              (Values for STEPSIZ or INCRFRAG; 2 to 40)
                           (Values for BRKTHRESH; .3 to .7)

TYPE SEQ NUMBER OF FIRST BP IN HOTSPOT (HSSTART)
654<CR>                    (Range for HSSTART; 1 to DNALEN-1)
                           (If no hotspot examined, use 100)

TYPE SEQ NUMBER OF LAST BP IN HOTSPOT (HSEND)
669<CR>                    (Range for HSEND; {HSSTART +1} to DNALEN)
                           (If no hotspot examined, use 101)
TYPE GC CLAMP LENGTH
20<CR>                     (Range for GC clamp; 0 to 60)

 -----ANALYSIS PARAMETERS------
(The program prints the input parameters to the screen. It then prints)

   ---SCREEN PROGRESS INDICATOR---
          1          90
(The first number indicates the program is at bp position 1 and is starting to calculate the
properties of the shortest fragment, 90 bp. As the run proceeds, new lines are printed. They
indicate the program has reached the next position to calculate properties of a 90 bp fragment.
When the run ends, 'stop terminated'is printed. The output file listing the results is called
SEQNAME.OF where the user's sequence name replaces SEQNAME.)
```

The program prompts the user for the input parameters listed in the top box of Figure 1. The algorithm analyzes the DNA sequence file from its first base pair to its last base pair, DNALEN. DNALEN is evaluated when the sequence file is read in. TFIRST, TDELT, and TLAST are the first temperature, the temperature increment, and the last temperature used in calculating the derivative melting curves of the fragments. Values of 78, 0.5, and 94°C were used. The temperature range (78−94°C) adequately evaluated the melting behavior of DNA fragments while minimizing calculation time. Similarly, a value of 0.5°C for TDELT gave sufficient resolution for the derivative melting curves while minimizing calculation time. INCRFRAG is the increment in the DNA fragment length used in the fragment cycle, and STEPSIZ the step size between new starting base pairs in the DNA fragment cycle (described below). Values of 10 or 20 bp were used for the latter two parameters. BRKTHRESH is a threshhold value that is compared with the calculated probability of being open for each base pair in a fragment. It defines when a base pair is considered broken, i.e. non-hydrogen bonded and unstacked. A value of 0.5 was used. The apparent melting temperature for a base pair, T'_m, is the temperature at which the probability of being open exceeds BRKTHRESH. HSSTART and HSEND are the base pair positions in the DNA sequence that start and end the user designated 'hot spot' region. A hot spot region is where mutations/polymorphisms are anticipated. CLAMPLEN is the length of the GC clamp which may be added to the fragment ends. A GC clamp segment may range from 0 to 60 bp and is selected from the 60 bp sequence used by Abrams *et al.* (2). MELTSCAN seeks DNA fragments that contain the entire hot spot region in the first melting domain.

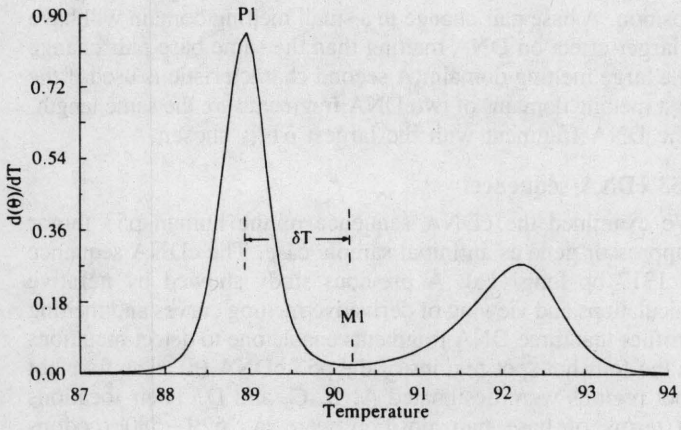


Figure 2. The derivative melting curve, DMC, of the 310 bp DNA sequence from position 311 to 600 in the p53 cDNA (16) with a 20 bp GC clamp left of 311. P1 is the peak of the first melting domain and δT is the temperature interval from P1 to the first minimum M1.

Table 2. Example output; analysis of cDNA p53 sequence and hot spot B. Using 20 bp GC lamp

```
>>>>>>>>>> HOT SPOT ANALYSIS<<<<<<<<<<<<<<<
HOT SPOT:
    STARTING BP:          654
    ENDING BP   :         669

OPTIMAL FRAGMENT FOR HOT SPOT :
          581 --->          690
    LENGTH (W/OUT G-C CLAMP) =          110
    G-C CLAMP ON LEFT
    TEMP AT DTHETA MAXIMUM =          91.000000
    DELT T =                           2.000000
    NUMBER OF OPEN BASE
    PAIRS AT DTHETA MIN    =           78
================================================
      BASE PAIR      --- OPTIMAL FRAGMENT ----     G-C     DELTA     NO. BP
       NO.'S        START BP   END BP   LENGTH   CLAMP ON    T     OPEN AT TMIN
        1 -->  115         0        0        1                .0        600
    116 -->  135       101      490      390    RIGHT      2.5        232
    136 -->  155       121      430      310    RIGHT      2.5        212
    156 -->  174       141      390      250    RIGHT      3.0        192
    175 -->  175         1      250      250    LEFT       3.0         76
    176 -->  177         1      230      230    LEFT       5.0         55
    178 -->  178       101      230      130    LEFT       4.0         53
    179 -->  215       121      230      110    LEFT       3.5         52
    216 -->  235        61      250      190    LEFT       2.5         75
    236 -->  255         1      270      270    LEFT       3.0         95
    256 -->  275       241      390      150    RIGHT      4.0         92
      .        .         .        .        .        .        .          .
      .        .         .        .        .        .        .          .
      .        .         .        .        .        .        .          .
   1236 --> 1255       941     1270      330    LEFT        .5        202
   1256 --> 1275       941     1290      350    LEFT        .5        222
   1276 --> 1295       941     1310      370    LEFT        .5        242
   1296 --> 1317         0        0        1                .0        600

Stop
```

We have applied MELTSCAN to sequences assuming hot spot regions 50 bp or less.

The second box of the flow diagram of Figure 1, FRAGMENT CYCLE, outlines how the program cycles through DNA fragments. The first DNA fragment contains the sequence from the starting base pair, STBP = 1, to (MINFRAG-CLAMPLEN) base pairs to the right. MINFRAG is the minimum fragment length examined. It was set at 110 bp. The GC clamp segment is added to the left end of the first fragment and a melting curve is calculated. The calculation is then repeated after moving the GC clamp to the right end of the fragment. The fragment length is then increased by adding to the right end the next INCRFRAG base pairs of the parent sequence. Melting calculations are again made with the GC clamp on each end. The length of the fragment is continuously incremented by INCRFRAG base pairs and two melting curves are calculated for each length until the fragment reaches or exceeds (MAXFRAG-CLAMPLEN). MAXFRAG was set at 520 bp. The starting base pair, STBP, is then moved to the left by STEPSIZ base pairs and the next series of fragments from 110 to 520 bp long are examined.

The core of the program is described from the CALCULATE box of Figure 1 to the end of the flow diagram. Derivative melting curves are calculated for all DNA fragments generated by the fragment cycle. The melting characteristics of the fragments are analyzed to find the best fragment for detecting mutations at each base pair position in the sequence. The program also determines the optimal fragment for the user specified segment of contiguous base pairs, i.e. the hot spot region.

The initial screening of DNA fragments looks for fragments with derivative melting curves (DMC) that have two peaks. A DMC is a plot of the temperature derivative of the fraction of intact base pairs, $d\theta/dT$, vs temperature, T. If the slope of $d\theta/dT$ at temperature T is negative or zero and then becomes positive at a higher temperature, there are at least two peaks in a DMC (Figure 2).

If the above analysis indicates two peaks in a DMC, the next step is to look for a clear separation between the first and second melting peak. This selection is described in the box ANALYSE FRAGMENT's DMC. A DNA must have a DMC with a difference between $d\theta/dT$ at the first peak (P1) and the first minimum (M1) of at least 0.05, and a temperature interval, δT, between the first peak and first minimum of at least 0.5°C (Figure 2). The values of 0.05 and 0.5°C are arbitrary, but were chosen based on observations of experimental DMCs of DNAs with two or more steps (13,15). We note that the above conditions do not actually determine the temperature difference between the first two peaks in a DMC. The selection conditions are, however, strongly correlated with well separated peaks. A test of these values was made by plotting DMCs of 33 DNA fragments selected by the algorithm (from Table 2). All show two (or more) peaks separated by at least 0.8°C. Since the selection decision is made at the first minimum of a DMC rather than the second peak, execution time is reduced.

Base pairs contributing to the first melting peak are determined from the calculated probability of each base pair being melted, PRBB. The file PRBB(I) contains the probability that each base pair (I = 1, 2, ...*n*) is melted at the calculated temperatures. A base pair is considered to be within the first melting peak if PRBB(I) is greater than BRKTHRESH at temperature M1. DNA fragments with two or more melting peaks are stored and indexed by the base pairs in their first melting domains. Figure 3a shows a melting map for the 310 bp DNA sequence used in the example of Figure 2. The apparent melting temperature, T'_m, for each base pair is shown. For the 310 bp DNA, the temperature M1 is 90.2°C. Base pairs considered to be in the first melting domain are, from Figure 3a, between positions 120 and 310. It may be worth noting that the calculation used for Figure 3a does not consider the duplex-to-single-strands dissociation (1,6). One cannot rely on melting maps alone to accurately predict the melting behavior of late melting domains (6). This is indicated by the observation that base pairs unwinding under the second melting peak in Figure 2 melt at ≈92.2°C, while T'_m values for these base pairs as indicated by Figure 3a are ≈94°C.

After the melting characteristics of all DNA fragments are calculated, the program scans through each base pair position and compares the characteristics of DNA fragments that have that base pair in their first melting domain. The 'best' fragment for a given position is selected using one of two characteristics. The first characteristic sought is the DNA fragment with the smallest number of base pairs in its first melting domain. This

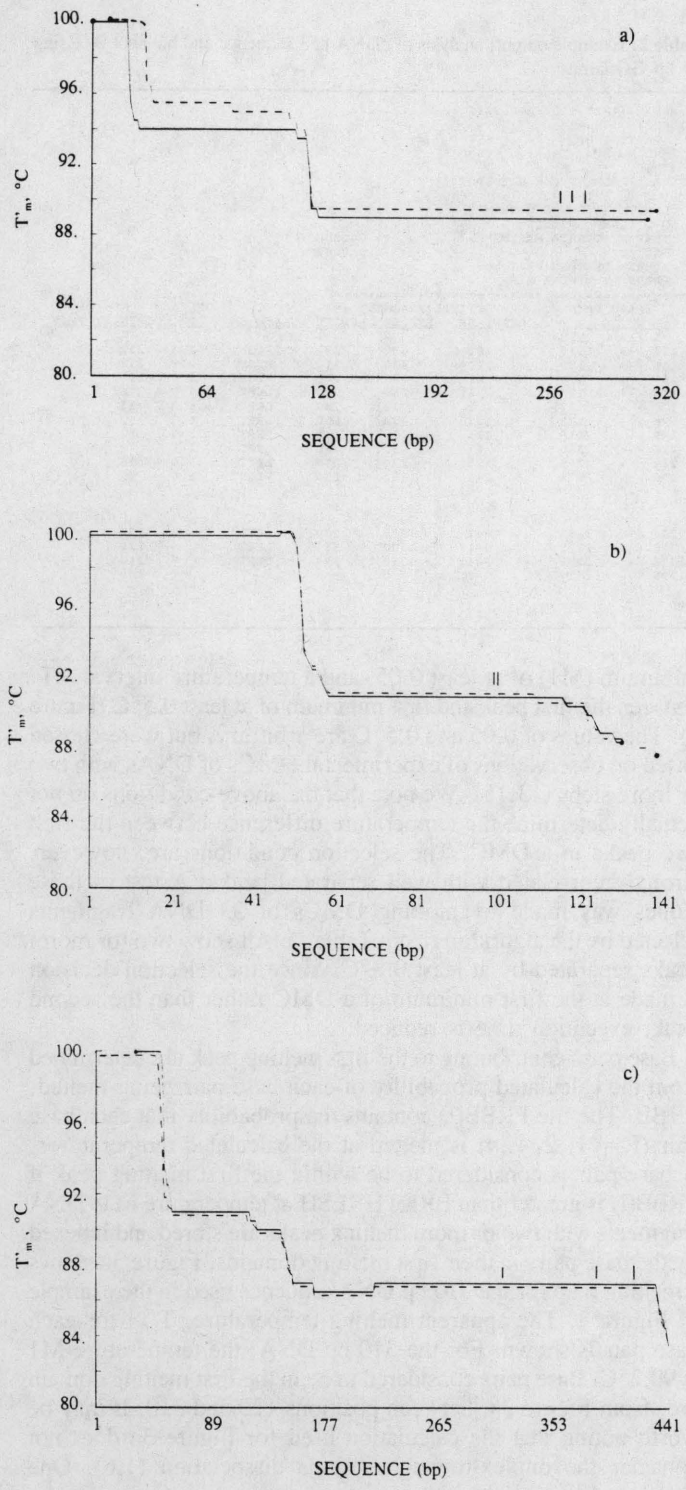


Figure 3. Melting maps of three DNA fragments determined by MELTSCAN for p53 cDNA hot spots and related DNA fragments used in ref. 6. The apparent melting temperature, T'_m, is plotted for each base pair. T'_m values in excess of 100°C are represented by 100°C. Short vertical lines indicate location of detected point mutations. (**a**) A-hot spot fragments: 310 bp DNA, 20 GC's and sequence 311–600 (—); 304 bp DNA, 20 GC's and sequence 321–604 (- - -). (**b**) B-hot spot fragments: 130 bp DNA, 20 GC's and sequence 581–690 (—); 140 bp DNA, 20 GC's and sequence 581–700 (- - -); (**c**) C/D-hot spot fragments: 410 bp DNA, 20 GC's and sequence 611–1000 (—), 440 bp DNA, 20 GC's and sequence 581–1000 (- - -).

enhances the ability to detect a base pair change at a given position. A base pair change in a small melting domain will have a larger effect on DNA melting than the same base pair change in a large melting domain. A second characteristic is used if the first melting domains of two DNA fragments are the same length. The DNA fragment with the largest δT is chosen.

p53 cDNA sequence

We examined the cDNA sequence of the human p53 tumor suppressor gene as an initial sample case. The cDNA sequence is 1317 bp long (16). A previous study showed by iterative calculations and viewing of derivative melting curves and melting profiles that three DNA fragments enable one to detect mutations in the four hot spot regions of the p53 cDNA (6). The four hot spot regions were designated A, B, C, and D. Their locations in terms of base pair position were A, 529–580 (codons 132–149); B, 654–669 (codons 174–179); C, 841–877 (codons 236–248); and D, 949–976 (codons 272–281) (17). The cDNA sequence region 841–976, C/D, was considered to be one hot spot region. Position numbers are based on the sequence from reference (16).

Table 1 shows an example of a user's application of the program with typical input parameter values. The DNA sequence file contained the 1317 bp p53 gene cDNA sequence. Acceptable parameter values and comments are in italics and between parentheses. Table 2 shows the output of this application of MELTSCAN. The hot spot region B described above was used in this example. Table 2 lists the optimal fragment for the specified hot spot region and the optimal fragment for each base pair or contiguous group of base pairs. Except for the first 105 bp, last 21 bp, and one 21 bp internal segment, the algorithm found fragments for mutation detection at all base pair positions. Sixty-six percent of the DNA fragments had δT values of 1.0°C or larger. The run time for the program employing a PC-compatible computer with a 80486 processor running at 66 MHz was 25 min.

The optimal fragments predicted for the three p53 hot spot regions A, B, and C/D were from base pair positions 311–600, 581–690 and 611–1000 respectively. A 20 bp GC clamp was on the left end in each case. These DNAs are essentially identical to the three DNA fragments previously shown to detect eight point mutations (6). The latter DNAs encompassed positions 321–604, 581–700, and 581–1000 with the 20 bp GC clamp also on the left ends. The agreement between the two sets of DNA fragments is not surprising since the criteria used to select DNA fragments in reference 6 is the same as that employed by MELTSCAN. The difference is that the previous method required the user to arbitrarily select DNA endpoints and iteratively calculate DNA fragment melting properties until a fragment with appropriate properties was found. Figure 3 shows the melting maps of the three DNA fragments predicted by MELTSCAN and the fragments employed in reference 6. The similarity of the melting maps indicates that one may move the endpoints of a predicted DNA fragment by up to ±10 bp without compromising the fragment's usefulness for detecting mutations. This can be helpful in optimizing primers to produce DNA fragments by the polymerase chain reaction (PCR).

p53 genomic sequence

Borrensen *et al.* (7) employed constant denaturant gel electrophoresis to detect p53 mutations in PCR-amplified fragments from human genomic DNA. Theoretical melting map calculations were

made to predict which DNA fragments would detect mutations in the four hot spot regions. The fragments employed were amplified using primers with 40 nucleotide long GC clamps attached to 20 nucleotide complementary sequences.

MELTSCAN was applied to the genomic p53 sequence to determine if shorter GC clamp primers would also allow mutation detection. The 1000 bp genomic sequence surrounding each hot spot was analyzed. For this calculation we assumed hot spot regions: A, codons 128–153; B, codons 161–185; C, codons 237–253; and D, codons 265–301 (7). Fragments were obtained for three of the four hot spot regions using 20 or 25 bp GC clamps. The fragment for hot spot A required a 40 bp GC clamp. The predicted fragments for the hot spot regions were: A, 13061–13140 with a 40 GC clamp on the left (120 bp); B, 13131–13245 with a 25 bp GC clamp on the left (140 bp); C, 13926–14110 with a 25 bp GC clamp on the left (210 bp); and D, 14021–14520 with a 20 GC clamp on the left (520 bp). Position numbers are based on the sequence presented in reference 18.

The DNA fragments predicted for hot spots B and C were tested by TGGE using genomic DNAs (19). Earlier work has shown that GC to AT mutations occur in codons 175 (hot spot B) and codon 248 (hot spot C) in the genomic DNA of cell line CEM (20). A different mutation in codon 248, CG to TA, was also characterized in cell line Namalwa (20). The 140 bp B-fragment and 210 bp C-fragment were amplified by PCR from genomic DNA isolated from the CEM and Namalwa cell lines, 11 other tumor cell samples, and peripheral blood cells with no p53 mutations (19). TGGE experiments verified the above mutations and found five mutations in hot spot regions B and C from the other tumor cells (19).

The utility of DNA melting theory for selecting DNA fragments for denaturing gel electrophoresis methods has been previously demonstrated (1,2,4,6,7,11). The new feature of MELTSCAN is its ability to automatically scan through a DNA sequence selecting an appropriate fragment for a hot spot region and for base pairs in the sequence. The results described above, although limited in scope, verify that MELTSCAN selects DNA fragments for detecting mutations. The program should simplify and broaden applications of denaturing gel methods. MELTSCAN is available at no cost to academic investigators. For further information readers may contact the corresponding author (internet e-mail address, roger.wartell@biology.gatech.edu.). MELTSCAN is proprietary to the Georgia Tech Research Corporation. MELTSCAN is a service mark of the Georgia Tech Research Corporation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support provided by grant GM38045 from the National Institutes of Health. We also thank S.H.Hosseini for providing information on TGGE results of PCR-amplified fragments from genomic p53 DNA samples, and T. Maier for programming advise.

REFERENCES

1. Lerman,L.S., Silverstein,K. (1987) *Methods Enzymol.*, **155**, 482–501.
2. Abrams,E.S., Murdaugh,S.E., and Lerman,L.S., (1990) *Genomics*, **7**, 463–475.
3. Rosenbaum,V. and Riesner,D. (1987) *Biophys. Chem.*, **26**, 235–246.
4. Riesner,D., Henco,K., and Steger,G. (1991) In Chrambach,A., Dunn,M.J., and Radola,B.J. (eds), *Advances in Electrophoresis*. VCH Pub., New York, Vol. 4, pp.171–250.
5. Wartell,R.M., Hosseini,S.H., and Moran,J.D. (1990) *Nucleic Acids Res.*, **18**, 2699–2705.
6. Ke,S-H., Kelly,P.J., Wartell,R.M., Hunter,S.H., and Varma,V.A. (1993) *Electrophoresis*, **14**, 561–565.
7. Borresen, A-L., Hovig,E., Smith-Sorensen,B., Malkin,D., Lystrad,S., Andersen T.I., Nesland,J.M., Isselbacher, K.J., and Friend, S.H. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 8405–8409.
8. Yoshino,K., Nishigaki,K., and Husimi,Y. (1991) *Nucleic Acids Res.*, **19**, 3153.
9. Lerman,L.S., Fischer,S.G., Hurley,I., Silverstein,K., and Lumelsky,N. (1984) *Annu. Rev. Biophys. Bioeng.*, **13**, 399–423.
10. Myers,R.M., Lerman,L.S., and Maniatis,T. (1985) *Science*, **229**, 242–246.
11. Sheffield,V. C., Cox,D.R., Lerman,L.S., and Myers,R.M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 232–236.
12. Costes,B., Girodon,E., Ghanem,N., Chassignol,M., Thong,N.T., Dupret,D., and Goossens,M. (1993) *Hum. Mol. Genet.*, **2**, 393–397.
13. Wartell,R.M. and Benight,A.S. (1985) *Physics Rep.*, **126**, 67–107.
14. Delcourt,S.G. and Blake,R.D. (1991) *J. Biol. Chem.*, **266**, 15160–15169.
15. McCampbell,C.R., Wartell, R.M., and Plaskon, R.R. (1989) *Biopolymer*, **28**, 1745–1758.
16. Zakout-Houri,R. Bienz-Tadmor,B., Givol,D., and Oren,M. (1985) *EMBO J.*, **4**, 1251–1255.
17. Nigro,J.M., Baker,S.J., Preisinger,A.C., Jessup,J.M., Hostetter,R., Cleary,K., Bigner,S.H., Davidson,N., Baylin,S., Devilee,P., Glover,T., Collins,F.S., Weston,A., Modali,R., Harris,C.C., and Vogelstein,B. (1989) *Nature*, **342**, 705–708.
18. Chumakov,P.M., Almazov,V.P. And Jenkins,J.R. (unpublished). GenBank X54156.
19. Hosseini,S.H. (1994) Ph.D. thesis, Georgia Institute of Technology.
20. Cheng,J. and Haas,M. (1990) *Mol. Cell. Biol.*, **10**, 5502–5509.
21. Fixman,M. and Friere,J. (1977) *Biopolymers*, **16**, 2693–2704.

Biochemistry, in press

The Influence of Neighboring Base Pairs on the Stability of Single Base Bulges and Base Pairs in a DNA Fragment†

Song-Hua Ke and Roger M. Wartell*

School of Biology
Georgia Institute of Technology
Atlanta, GA 30332

† This work was supported by NIH Grant GM38045.

* Author to whom correspondence should be addressed.

Running Title: Single Base Bulges in DNA

Key Words: Temperature-Gradient, Electrophoresis, Melting, Mutations

ABSTRACT

Temperature-Gradient Gel Electrophoresis (TGGE) was used to determined the relative thermal stabilities of 32 DNA fragments that differ by a single unpaired base (base bulge) and 17 DNAs differing by a base pair. Homologous 373 bp and 372 bp DNA fragments differing by a single base pair substitution or deletion were employed. Heteroduplexes containing a single base bulge were formed by melting and reannealing pairs of 372 bp and 373 bp DNAs. Product DNAs were separated based on their thermal stability by parallel and perpendicular TGGE. The order of stability was determined for all single unpaired bases in four different nearest neighbor environments; (GXT)·(AYC), (GXG)·(CYC), (CXA)·(TYG), and (TXT)·(AYA) with X = A, T, G or C, and Y = no base, or *visa versa*. DNA fragments containing a base bulge were destabilized by 2°C to 3.6°C (± 0.2 °C) with respect to homologous DNAs with complete Watson-Crick base pairing. Both the identity of the unpaired base and the sequence of the flanking base pairs influenced the degree of destabilization. The range of temperature shift correspond to estimated unfavorable free energies from 2.5 to 4.6 kcal/mole. Purine base bulges were generally not as destabilizing as pyrimidine base bulges. An unpaired base which was identical to one of its adjacent bases generally caused less destabilization than an unpaired base with an identity differing from its nearest neighbors. This implies that positional degeneracy of an unpaired base within a run of two or more identical bases is an important factor effecting stability. The ability of TGGE to order the stabilities of DNA fragments differing by a single base pair was used to determine the relative stabilities of base pair stacking interactions. The results determined by TGGE were consistent with the relative stabilities determined from UV melting transitions.

Extra unpaired bases, or "bulges" in double-helical DNA are created when one or more consecutive bases are unpaired to bases on the opposite strand. They can arise from recombination between sequences that are imperfectly homologous, or from errors during replication of DNA. A duplex with a bulged base is the proposed intermediate in the process of frame-shift mutagenesis (Streisinger et al., 1966; Ionov et al., 1993). Extra bases in duplex segments of nucleic acids are also implicated in site specific recognition by nucleic acid binding proteins (Christiansen et al., 1985; Wu & Uhlenbeck, 1987). Understanding how different bulged bases and their sequence environment influence DNA structure and stability will help explain their role in biological processes. Information on the effect of bulged bases on DNA stability will also help optimize methods that rely on thermal stability differences to hybridize probes to specific sites.

Previous studies of unpaired bases have focused primarily on short DNA duplexes. Investigators have examined the stability and structure of single unpaired bases by NMR (Woodson & Crothers, 1987), X-ray crystallography (Miller et al., 1988; Joshua-Tor et al., 1992), UV absorbance melting studies and gel retardation assays (Rice & Crothers, 1989; Wang & Griffith, 1991; LeBlanc & Morden, 1991). Results indicate that both the identity of the bulged base and its sequence environment influence the amount by which a duplex oligomer is destabilized. No systematic study has yet to be reported on the effects of different single-base bulges on the stability of longer DNAs. The influence of end effects on the properties of short DNA duplexes (Olmsted et al., 1991) and the desire for a more complete survey of context effects makes such a study desirable.

A vertical TGGE format (Wartell et al.,1990) was employed to determine the relative stabilities of the four possible single-base bulges on each strand in four different nearest neighbor environments within 372/373 bp DNA duplexes. This investigation utilized an approach employed in an earlier study on the influence of single base pair mismatches on DNA stability (Ke & Wartell, 1993). In TGGE DNA fragments migrate through a polyacrylamide gel with a superimposed temperature gradient. A DNA moves at a constant mobility until the least stable melting domain denatures. The branched structure of the partially melted DNA results in a large decrease in mobility. DNAs of identical lengths but differing in the stability of their first melting domain migrate to different depths before they denature. A temperature gradient parallel to the direction of electrophoresis provided the greatest resolution for detecting changes in DNA thermal stability. Experiments in which the temperature gradient was perpendicular to the direction of electrophoresis were carried out to obtain mobility transition curves.

MATERIALS AND METHODS

Materials: Plasmids pUC8-31 and pUC8-36 plasmids were previously described (Tatti & Moran, 1985). The plasmids contain a 130 bp segment of the *ctc* promoter region from *Bacillus subtilis* inserted between the HindIII and EcoRI sites of pUC8. pUC8-31 has the wild type *ctc* sequence, and pUC8-36 has a G·C to A·T transition at position -36. The designation "-36 " denotes the base pair position located 36 base pairs upstream of the *ctc* transcription startpoint. The polymerase chain reaction (PCR) was used to amplify a 373 bp region of the plasmids containing the *ctc* promoter (figure 1). Four sets of specific point mutations or deletions were created by PCR through the application of upstream primers with single base substitutions

or deletions. The DNA oligonucleotides employed (Operon Inc., Alameda, CA.) for the sixteen upstream primers and one downstream primer are shown in figure 1. Base positions of the substitutions are underlined and single-base deletions in the upstream primers are indicated with a "-" at the relevant positions. The downstream primer DP15 was end-labeled for some PCR amplifications with ^{32}P as described previously (Ke & Wartell, 1993). Thirteen 373 bp DNAs differing from each other by a single base pair were produced by PCR. Four 372 bp DNAs were created by PCR using upstream primers with one base deleted (figure 1). The base pair changes and deletions occurred at the four positions designated -36, -38, -39, and -43.

Taq DNA polymerase was obtained from Perkin Elmer Cetus Inc. or Promega Inc. The PCR protocol recommended by Perkin Elmer Cetus Inc. was employed. The 100 ul reactions contained 50 pg of plasmid DNA, 0.6 uM of each primer in a buffer of 10 mM Tris-HCl, pH 8.3, 50 mM KCl, and 2.5 mM $MgCl_2$. 200 uM of each dNTP was employed. Further details are described in Ke & Wartell (1993). 2 to 4 ul of each reaction was checked for size and purity on a 6.5 or 7.5% polyacrylamide gel.

Although bulged bases phased with the helical repeat of DNA and RNA induce kinking and reduce electrophoretic mobilities in gels (Tang & Draper, 1990, Wang & Griffin, 1991), no single unpaired base caused a significant perturbation of gel mobility for the DNA molecules employed. The bending induced by a single bulge near the end of the DNA is apparently not large enough to affect the mobility of the DNA. In the absence of a temperature sufficient for unwinding, all DNA fragments had the same mobility. This is best illustrated in the perpendicular gel such as figure 6. Prior to the onset of melting the DNAs had the same mobility.

TGGE: TGGE was carried out using a vertical gel electrophoresis apparatus previously described (Wartell et al. 1990; Ke & Wartell, 1993). Two aluminum heating plates sandwiching the glass plates were used to establish a temperature gradient either parallel or perpendicular to the electric field. Temperature measurements in the gel (described below) confirmed that the temperature gradient was linear and uniform within the region covered by the heating plates. A 6.5% polyacrylamide gel (37.5:1, acrylamide:bisacrylamide) was used. The gel contained 3.36 M urea and 19.2% vol/vol formamide in 0.5X TBE (0.045 M sodium borate + 0.045 M Tris + 1 mM EDTA, pH 8.1). This corresponds to 48% of what is often referred to as 100% denaturant solution (7M urea + 40% formamide). Formamide was deionized with mixed resin AG501-X8D (Bio-Rad). The gel running buffer was 0.5X TBE. Run times were generally from 14 to 16 hours at 4.5 to 6 volts/cm. The gels were stained with ethidium bromide and photographed.

The relative stabilities of DNA fragments were determined from experiments in which the temperature gradient was parallel to the direction of electrophoresis (e.g., figure 2). The low and high temperatures of the gradients are described in the text. DNA mobility transitions were obtained from perpendicular temperature gradient gels (e.g., figure 6). Approximately 1 ug of each DNA fragment was loaded into a long well along the top of the gel and electrophoresed through a gradient with the high temperature sufficient to unwind at least the first melting domain of the DNAs.

Temperatures in the gels were determined with a thermocouple probe (TMTSS-020-6, Omega Inc.) connected to a digital thermometer (MDSD-465, Omega Inc., accuracy estimated as $\pm$ 0.1 °C). Measurements at different depths in test gels verified the linearity of the gradient.

Temperature measurements were made for the perpendicular temperature gradient gels at two or more positions at the end of each transition run. The positions where the probe had been inserted were observed as dark lines in the photographs and provided a temperature scale.

Photographs of the mobility transition curves were digitized and scaled using a digitizer tablet (SummaSketch II), and input to a microcomputer. Transition curves were smoothed prior to obtaining derivative curves by the "smoothlowess" function in the Axum graphics analysis package (Trimetrix Inc., Seattle, Wash.). This is based on a locally weighted regression analysis (Cleveland & Devlin, 1988). The mobility transition temperature, T_u, was defined as the temperature at the peak of the transition's derivative curve.

RESULTS

DNA fragments containing single base bulges were made by melting and annealing two PCR generated DNAs differing in length by a base pair. The DNAs were heated for three minutes at 97 °C, reannealed at 54 °C for at least 10 minutes, and allowed to slowly cool to room temperature. Two heteroduplex DNAs with a single-base bulge each and the two homoduplexes were created. The four resulting DNAs were separated by parallel TGGE. In some experiments one of the DNAs was ^{32}P-labeled on the 5'-end of its downstream primer strand to identify separated heterodupex DNAs. The gel was stained with ethidium bromide and photographed. When ^{32}P-labeled DNA was used the bands were located on a UV-transilluminator, excised, and their radioactivity measured by scintillation counting. DNA band identities were established from the radioactively labeled bands. Confirmation of band identities was made by switching the DNA that contained the labeled strand and/or by running one of the

homoduplex DNAs at a larger concentration.

Parallel TGGE. Figure 2a and 2b show parallel TGGE experiments of 372/373 bp DNAs with all possible paired, deleted and unpaired bases at position -36. The nearest-neighbor base pairs surrounding this position are (GXT)·(AYC). The identity of the bands in figure 2a and 2b are given in the figure caption and were based on the procedures described above. DNA fragments are designated by specifying the base (A,T,G,C) at position X or Y, or a dash (-) for the absence of a base. Thus (GCT)·(A-C) implies the DNA with a C bulge on the top strand (position X), and (G-T)·(A-C) denotes the 372 bp homoduplex. A temperature gradient from 35.5 to 42.5 °C was used for the gel shown in figure 2a to optimize the separation of homoduplex and heteroduplex DNAs in one gel. The uppermost band in lanes 1, 2, and 3 actually corresponds to two heteroduplex DNAs with single-base bulges. A larger separation for two heteroduplex DNAs is seen in lane 4. Figure 2b shows a gel that employed a temperature gradient from 32 to 35 °C. This shallower gradient optimized the separation of the DNAs with single-base bulges, sacrificing the ability to separate the homoduplex DNAs. The brighter bottom band in all lanes of figure 2b contains two homoduplex DNAs. The most stable single-base bulge DNA in this series is (G-T)·(AAC) in lane 4 of figure 2b. This band is separated by 1.6 cm from the least stable single-base bulge DNA, (GAT)·(A-C), in lane 3. We note that the A bulge in the DNA represented by (G-T)·(AAC) could also shift to the adjacent position, i.e. (GT-)·(AAC). For simplicity only one representation is used.

Figures 3, 4 and 5 show parallel TGGE experiments of DNAs with all possible base pairs and deleted or unpaired bases at positions -38, -39, and -43 respectively. The base pairs surrounding these positions are (GXG)·(CYC) for -38, (CXA)·(TYG) for -39, and

(TXT)·(AYA) for -43. Figures 3 and 4 used similar temperature gradients (~36 °C to ~42 °C) to optimize the separation of homoduplex and heteroduplex DNAs in one gel. The largest separation of two homoduplex DNAs with transversion of the same base pair is observed in lanes 3 and 4 of figure 3. The DNA fragment with the (GCG)·(CGC) sequence centered at position -38 moved 1.8 cm further into the gel than the DNA with the (GGG)·(CCC) sequence. Figure 5 used a temperature gradient from 33 to 36 °C to optimize the separation of the DNAs with single-base bulges at position -43. Gels with a temperature gradient from 33 to 36 °C were also used to analyse the relative stabilities of DNA fragments with single-base bulges at positions -38 and -39 (not shown).

Table 1 summarizes the results from the parallel TGGE experiments. Both the identity of the unpaired base and its neighboring base pairs influence the destabilization caused by the unpaired base. A single base bulge with the same identity as a nearest neighbor base generally induces less destabilization than a base-bulge with an identity different from its nearest neighbors. The data at position -36 in Table 1 illustrates this point. For each of the four unpaired bases the most stable situation occurs when the bulged base is identical to a neighboring base, e.g. (GGT)·(A-C) > (G-T)·(AGC). A similar situation occurs for the other positions as well.

At position -43, except for the two most stable situations,(T-T)·(AAA) and (TTT)·(A-A), the base bulge located within the stretch of A's is more stable than the base bulge located in the strand with the stretch of T's. A similar observation was made in a study of single-base bulges in a DNA oligomer (LeBlanc & Morden, 1991). Table 1 also shows that at each position examined the most stable single-base bulge is a purine and the least stable base bulge is a pyrimidine. The greater stability of purine bulges is consistent with NMR evidence indicating

that unpaired purines are generally stacked into the helix (Woodson & Crothers, 1988, 1989) while unpaired pyrimidines may be extrahelical or intrahelical depending on the sequence context and temperature (Morden et al., 1990; van der Hoogen et al., 1988; Kalnik et al. 1989).

Mobility Transition Curves. Perpendicular TGGE experiments display the temperature induced mobility transition of intact double-stranded DNA to the denatured state. Figure 6 shows transition curves of the first melting domain of five of the DNAs examined. The upper temperature limit employed in figure 6 was selected to induce unwinding of only the first melting domain. Several steps are observed in this DNA's complete mobility transition (Wartell et al. 1990, figure 7). The initial quasi-linear increase in mobility with increasing temperature appears to be due to the effect of temperature on gel viscosity and/or pore size. The larger sigmoidal decrease in mobility is due to the unwinding of DNA strands.

The two leftmost transitions in figure 6 correspond to DNAs altered at position -36 to have bulges (GTT)·(A-C) and (G-T)·(AAC). The middle transition corresponds to the DNA with the T·A base pair at position -36. The two transitions on the right correspond to DNAs with the base pair deletion and the G·C base pair at position -36. The transition curves of the DNAs with bulged bases are noticably broader than the other DNAs. This may indicate that internal melting from the bulged base plays a role in the melting process. It may also reflect an influence of the extra base on the configuration of partially melted DNA. Broader transitions were also observed for the 373 bp DNAs containing certain mismatched base pairs (Ke & Wartell, 1993).

The transition temperatures of the DNA fragments' first melting domain, T_u, are given in Table 2. The T_u data were consistent with the results from the parallel TGGE (Table 1).

When T_u values appear to be the same for two DNAs the order of stability listed in Table 2 was determined from Table 1 since resolution of the parallel gradient gels was better than the perpendicular temperature gradient gels. T_u values of the DNAs with a single-base bulge were lower than the values for corresponding 372 bp homoduplex DNAs by 2 to 3.6 °C (± 0.2). Mobility transitions of the first domains of homoduplex DNAs in gels with 48% and 58% denaturant differed by 5.3 °C but gave essentially identical slopes and the same order of homoduplex stabilities (not shown). This is consistent with previous melting studies indicating that urea and formamide lower the thermal stability of G·C and A·T base pairs by roughly equivalent amounts (Hutton, 1977; Klump & Burkart, 1977; Casey & Davidson, 1977).

Nearest Neighbor Base Pair Stacking. The T_u data of the homoduplex DNAs in Table 2 provides information on the relative stability of nearest neighbor stacking for each pair of DNAs with an A·T to T·A or G·C to C·G transversion at a specific position. Although T_u is not defined as the temperature where 50% of the domain's base pairs are melted - the T_m - we consider $\Delta T_u = \Delta T_m$ for the homoduplex DNAs. This assumption is supported by an analysis of the full DNA mobility curves which indicate $T_u \approx T_m$ for the first melting domain (see below). Since the slopes of the mobility transitions of the first domains of the homoduplex DNAs are the same (e.g. figure 6), and the temperature differences are small, the ΔT_u values may be related to stacking free energy differences.

Table 3 lists the ΔT_u value for pairs of DNAs with A·T to T·A or G·C to C·G changes at specific positions. Within the framework of the nearest neighbor model of DNA stability, the DNA with the higher T_u value has the more stable pair of stacking interactions at the specified position. The adjacent column lists the difference in stacking energy terms for the pairs of DNAs

based on stacking parameters evaluated by Delcourt & Blake (1991). For all eight cases examined, the TGGE data and the parameters of Delcourt & Blake give the same relative ranking of stacking interactions, i.e. the DNA with the higher T_u has the lower net stacking energy. This was also observed using stacking parameters evaluated from other studies (Wartell & Benight, 1985; Breslauer et al. 1986; Doktycz et al. 1992). A quantitative correlation between ΔT_u and the net stacking energy is not observed for any of the stacking parameter sets. This lack of quantitative agreement is not unexpected given the difference in solvents between the TGGE experiments and that used to obtain the melting parameters.

A comparison was made between the ΔT_u values of DNAs differing by one base pair at a specific position with the ΔT_m values predicted from calculated melting curves. The calculations employed the Delcourt & Blake (1991) stacking parameters, and other parameters corresponding to a solvent of 0.1 M Na^+ (McCampbell et al., 1989). The ΔT_m's predict the same order of stability for the DNAs as the measured ΔT_u's although quantitative agreement is again not observed. One interesting observation correctly predicted by the theory is that deleting a base pair at each of the four positions always produced a DNA with a stability higher than the DNA with an A·T (or T·A) pair and lower than the DNA with a G·C (or C·G) pair.

The predictions of DNA melting theory were also compared with the full experimental mobility curve of the wt 373 bp DNA. Lerman et al. (1984) introduced an equation to describe the electrophoretic retardation of DNA accompanying partial melting in a polyacrylamide gel. The relative mobility of a long duplex DNA with a short melted domain may be written as

$$\mu(T) = \mu_o \exp[-p(T)/L_r] \qquad (1)$$

where μ_o is the mobility of the completely duplex DNA, L_r is the length of the flexible unit of

a DNA strand in number of bases, and p(T) is the length of the melted region at temperature T measured in the number of melted base pairs. p(T) is calculated from melting theory and is the sum of probabilities of each base pair being melted (Lerman et al. 1984). This is equivalent to the number of opened base pairs if strand dissociation is neglected. Since equation (1) is applicable only when the melted domain is a small fraction of the entire DNA, neglecting strand dissociation is a reasonable assumption. Equation (1) is not expected to accurately predict the mobility transition curve beyond the initial stages of melting.

The mobility curve of the wild type 373 bp DNA was obtained from a perpendicular temperature gradient gel. A 48% denaturant 6.5% polyacrylamide gel was used with a temperature gradient from 30 °C to 55 °C. The mobility curve was first normalized by dividing the distance to which DNA migrated at each temperature by the distance migrated at 30 °C. Then the small pre-transition linear increase in DNA mobility with temperature (see e.g. figure 6) was subtracted from the entire transition. This subtraction corrects for the effect of temperature on gel pore size and/or viscosity, and its influence on the mobility of duplex DNA. The solid line in figure 7a shows the adjusted experimental mobility transition. A theoretical mobility curve was calculated from eq. (1). The melting theory parameters employed were from solution studies in ~ 0.1 M Na^+ (McCampbell et al. 1989, Delcourt & Blake, 1991) except for the melting temperatures of the average AT and GC base pairs, T_{AT} and T_{GC}. A uniform decrease in these parameters was needed to compensate for the lower ionic strength of the gel buffer (0.045 M Tris borate) and the urea-formamide denaturant. The dash-dot line in figure 7a shows the predicted mobility transition with values of T_{AT} =24.1 °C and T_{GC} =68.0 °C and L_r = 100. The T_{AT} and T_{GC} values are about 40 °C lower than values appropriate for 0.1 M Na^+

Figure 7a shows that eq. (1) using DNA melting theory and its parameters provides an excellent fit to the first phase of the mobility melting curve. A comparison of the calculated number of opened base pairs for the first phase of the experimental mobility curve shows a very close correspondence (< 0.1 °C) between the calculated T_m and experimental T_u. The good agreement between the calculated and experimental curves supports the application of the theory for interpreting the mobility curve data. Figure 7b shows the calculated derivative melting curve for the 373 bp DNA. The first melting domain is well separated from unwinding of the rest of the DNA. The area under the first melting domain corresponds to the unwinding of 52 base pairs. Analysis of the DNA's melting map (not shown) shows this region is from the EcoRI end of the sequence where the single base bulges occur.

DISCUSSION

The degree of destabilization due to a single-base bulge was influenced by both the base type and its neighboring base sequence. The results from this study provide information that may be related to mutational events, and illustrate how TGGE can be applied to determine the stabilities of specific non-Watson-Crick structures in long nucleic acid duplexes.

Frame-shift mutations are known to occur predominantly in simple repeated sequences (Ionov et al., 1993; Calos & Miller, 1981). According to a model of frame-shift mutation proposed by Streisinger et al. (1966), slippage of one strand of DNA causes one or more base to loop out. If the stability of a base bulge is the dominant factor determining the liklihood of a frameshift mutation, one may expect the mutation frequency at a site to be correlated with the

stability of the base bulge. NMR studies (Woodson & Crothers, 1987) on $(GATG_3CAG) \cdot d(CTGC_4ATC)$, and $d(GATG_3CAG) \cdot d(CTGAC_3ATC)$ indicated a bulged C base, when identical to neighboring bases, is more stable than a bulged A base which differs from neighboring bases.

We have observed as a general phenomena that a single base bulge with the same identity as one of its adjacent bases causes less destabilization than the bulged base when it has different nearest-neighbors. This can be rationalized on the basis of the positional degeneracy of an unpaired base within a homogeneous tract which increases the entropy and therefore the stability of the DNA. If this phenomena minimizes error correction of random base insertions, the results suggest that homogenous tracts would tend to persist or perhaps enlarge over evolutionary time.

The analysis of the full mobility transition of the 373 bp DNA showed that the unwinding of the first melting domain is well separated from the remainder of the DNA molecule (figure 7). Since the single base bulges are in this domain and its unwinding can be treated as a separate transition, the shift in the domain's melting temperature due to a bulged base can be related to a free energy change. The destabilizing free energy of a bulge can be expressed as (Woodson & Crothers, 1987)

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

in which ΔS is the entropy change for melting the fully base paired domain, and δT_m is the shift in T_m due to a bulge. We assume $\delta T_u \approx \delta T_m$ and ΔS is independent of temperature. ΔS can be estimated by multiplying the number of base pairs in the first melting domain, N_1, times ΔS° ,the average entropy change/base pair. Analysis of the 373 bp DNA melting curve using the parameters employed for figure 7 indicates that $N_1 \approx 52$ bp. This value was determined from

the relative area under the first domain of the derivative melting curve, 14%, and confirmed by a melting map analysis of the DNA (e.g., Wartell et al. 1990). The latter analysis also shows that the first melting domain is an end-melting domain encompassing the bulge sites. An estimate of error for N_1 was made by calculating and analysing the melting curves of the 373 bp DNA using several different sets of stacking parameters in addition to those of Delcourt & Blake (1991) (McCampbell et al. 1989, Gotoh & Tagashira, 1981, Doktycz et al. 1992). We found $N_1 = 52.5 \pm 3.5$ bp.

Utilizing an average value per base pair of $\Delta S^\circ = -24.8$ cal K^{-1} (mole bp)$^{-1}$ (Delcourt & Blake, 1991), the destabilizing free energy for temperature shifts of 2 to 3.6 °C ranges from $2.5 < \delta(\Delta G) < 4.6$ kcal/mole. These values are in the same range determined for single-base bulges in DNA oligomers (Morden et al., 1983; Woodson & Crothers, 1987; LeBlanc & Morden, 1991). Given the errors in the temperature shifts, N_1, and ΔS° , the value of $\delta(\Delta G)$ for a specific bulge is approximately $\pm$ 10%.

The relative stabilities of the single base bulges determined by TGGE are in good agreement with data where available from short DNA oligomers. NMR studies were carried out on the four self-complimentary duplexes each with a single-base bulge (underlined); d(CCGGAATTCACGG), d(CCGAGAATTCCGG), d(CCGTGAATTCCGG), d(CCGGAATTCTCGG) (Kalnik et al., 1989b; Kalnik et al., 1990). The results demonstrated that the 13-mer duplex with (GAG)·(C-C) is more stable than the duplex with (G-G)·(CAC), and the 13-mer duplex with (GTG)·(C-C) is more stable than the duplex with (G-G)·(CTC). Comparing the T_m values of the bulged duplexes in these two studies, the following order of stabilities is obtained: (GAG)·(C-C) > (GTG)·(C-C) > (G-G)·(CAC) > (G-G)·(CTC). We

observed the same ranking in the equivalent nearest neighbor environment at position -38 (table 1). UV absorbance melting curves (LeBlanc & Morden, 1991) were obtained from oligomers $d(GCGA_2XA_2GCG) \cdot d(CGCT_4CGC)$ and $d(GCGA_4GCG) \cdot d(CGCT_2YT_2CGC)$ with X = C, T, or G in A-strand and Y = C, A, or G in T-strand. Comparison of DNA oligomer T_m values yielded the following relative order of stability of single-base bulges; bulges in A-strand > bulges in T-strand; T bulge > C bulge > G bulge in A-strand; and C bulge > G bulge > A bulge in T-strand. In our experiments, a similar ranking was observed in the -43 set where the nearest-neighbor sequence is the same as the above oligomers. The only differences were the relative order of duplexes which have very similar thermal stabilities. This may be due to the difference between the two experimental conditions, or the oligomer vs. polymer contexts.

The mobility transition curves provide a novel approach for directly evaluating the heirarchy of stacking interactions in a long DNA. It differs from the least squares analysis fitting the ten stacking parameters to melting data from a number of DNAs (e.g., Delcourt & Blake, 1991; Breslauer et al. 1986). The consistency between the results obtained by TGGE and the stacking interactions obtained by least squares analysis lends support to the determined heirarchies. A more meaningful and quantitative test of stacking interaction values will require a quantitative comparison using data obtained under similar solvent conditions.

ACKNOWLEDGEMENTS. The authors wish to acknowledge Thomas Maier for the technical assistance and equipment maintenance.

References

Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746-3750.

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

Casey, J., & Davidson, N. (1977) *Nucleic Acids Res. 4*, 1539-1552.

Christiansen, J., Douthwaite, S. R., Christensen, A., & Garrett, R. (1985) *EMBO J. 4,* 1019-1024.

Cleveland, W. S., & Devlin, S. J. (1988) *J. Amer. Stat. Assoc. 83*, 596-610.

Delcourt, S.G. & Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160-15169.

Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., & Benight, A. S. (1992) *Biopolymers 32*, 849-864.

Gotoh O. & Tagashira Y. (1981) *Biopolymers 20*, 1033-1042.

Hutton, J. R. (1977) *Nucleic Acids Res. 4*, 3537-3555.

Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., & Perucho, M. (1993) *Nature, 363,* 558-561.

Joshua-Tor, L., Frolow, F., Applella, E., Hope, H., Rabinovich, D., & Sussman, J. L. (1992) *J. Mol. Biol. 225*, 397-431.

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

Kalnik, M. W., Norman, D. G., Swann, P. F., & Patel. D. J. (1989b) *J. Biol. Chem. 264,* 3702-3715.

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

265, 636-647.

Ke, S-H., & Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137-5143.

Klump, H., & Burkart, W. (1977) *Biochim. Biophys. Acta 475*, 601-604.

LeBlanc, D. A., & Morden, K. M. (1991) *Biochemistry 30*, 4042-4047.

Lerman, L. S., Fischer, S. G., Hurley, I., Silverstein, K., & Lumelsky, N. (1984) *Ann. Rev. Biophysics & Bioengineering 13*, 399-424.

McCampbell, C. R.,Wartell, R. M., & Plaskon, R. R.(1989) *Biopolymers 28*, 1745-1758.

Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 85-86.

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

Olmsted, M. C., Anderson, C. F., & Record, M. T. Jr. (1991) *Biopolymers 31,* 1593-1604.

Rice, J. A., & Crothers, D. M. (1989) *Biochemistry 28*, 4512-4516.

Streisinger, G., Okada, Y., Emrich, J. N., Tsugita, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77-84.

Tang, R. S., & Draper, D. E. (1990) *Biochemistry 29*, 5232-5237.

Tatti, K. M., & Moran, C. P. Jr. (1985) *Nature 314*, 190-192.

van den Hoogen, Y. Th., van Beuzekom, A. A., van den Elst,H., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988) *Nucleic Acids Res. 16*, 2971-2986.

Wang, Y-H., & Griffith, J. (1991) *Biochemistry 30*, 1358-1363.

Wartell, R. M., Hosseini, S. H., & Moran, J. D. (1990) *Nucleic Acids Res. 18*, 2699-2750.

Wartell, R. M., & Benight, A. S. (1985) *Physics Reports 126*, 67-107.

Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 3130-3141.

Woodson, S. A., & Crothers, D. M. (1989) *Biopolymers 28*, 1149-1177.

Woodson, S. A., & Crothers, D. M. (1987) *Biochemistry 26*, 904-912.

Wu, H-N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221-8227.

Table 1. Comparison of Bulged Base Stabilities in 373 bp DNA Fragment

Bulge Position	5' Flanking Base Pair*	3' Flanking Base Pair*	The Ranking of Single-Base Bulge Stability
-36	G·C	T·A	G/- > -/A > T/- > -/C ≥ -/T ≥ -/G > A/- > C/-
-38	G·C	G·C	G/- > -/C ≥ A/- > T/- ≥ -/G > C/- ≥ -/A > -/T
-39	C·G	A·T	-/G > A/- ≥ G/- > C/- ≥ -/A > -/T > T/- > -/C
-43	T·A	T·A	-/A > T/- > -/G = -/T ≥ -/C > G/- = A/- ≥ C/-

* e.g. the single-base bulge designated A/- at position -36 represents the following sequence:

5'---- GAT ---- 3'

3'---- C-A ---- 5'

Table 2. Perpendicular TGGE [a]T_u Measurements of First Melting Domain of 372/373 bps DNA

-36 set (°C)	-38 set (°C)	-39 set (°C)	-43 set (°C)
C·G: 37.1	C·G: 38.8	[b]G·C: 37.1	G·C: 37.7
[b]G·C: 37.1	G·C: 38.0	C·G: 36.8	C·G: 37.3
[c]-/-: 36.5	[c]-/-: 37.6	[c]-/-: 36.2	[c]-/-: 37.3
T·A: 36.2	T·A: 37.4	A·T: 35.8	[b]T·A: 37.1
A·T: 36.1	[b]A·T: 37.1	T·A: 35.3	A·T: 36.5
G/-: 34.1	G/-: 35.6	-/G: 34.0	-/A: 35.1
-/A: 33.9	-/C: 34.7	A/-: 33.8	T/-: 34.9
T/-: 33.7	A/-: 34.7	G/-: 33.7	-/G: 34.6
-/C: 33.4	T/-: 34.5	C/-: 33.5	-/T: 34.6
-/T: 33.4	-/G: 34.3	-/A: 33.5	-/C: 34.6
-/G: 33.4	C/-: 34.3	-/T: 33.4	G/-: 34.4
A/-: 33.2	-/A: 34.2	T/-: 33.4	A/-: 34.4
C/-: 33.2	-/T: 34.0	-/C: 33.2	C/-: 34.4

a). T_u was defined as the temperature at the peak of the derivative curve calculated from the smoothed DNA mobility transition profile. Estimated precision in T_u relative to the standard DNA fragment is $\pm$ 0.2 °C averaged over all DNAs. Each experiment contained the pUC8-31 DNA fragment as an internal standard. The mean T_u of this DNA was 37.1 °C ($\pm$0.4°C) based on 15 repeated experiments.

b). This is the same pUC8-31 DNA fragment .

c). -/- corresponds to the 372 bp duplex DNAs where the base pair at the designated position has been deleted.

Table 3. Comparison of Base Pair Stacking Parameters.

ΔT_u and stacking energy difference for base pair transversion at specific positions in 373 bp DNA

DNA Fragments[a] (position)	ΔT_u (°C)	$\{[\delta G_{JK} + \delta G_{KL}] - [\delta G_{PQ} + \delta G_{QR}]\}$[b] (cal/mole)
[TTT - TAT] (-43)	0.6	-81
[TGT - TCT] (-43)	0.5	-186
[GCG - GGG] (-38)	0.8	-615
[GTG - GAG] (-38)	0.3	-186
[GTT - GAT] (-36)	0.1	-240
[GCT - GGT] (-36)	0.05	-179
[CAA - CTA] (-39)	0.5	-27
[CGA - CCA] (-39)	0.2	-250

(a) Each DNA is designated by the three top-strand bases that include the position of the base pair transversion in the center. Thus for the second row TGT is the DNA with a G·C base pair at position -43.

(b) δG_{JK} corresponds to the stacking energy parameter for the base pair stack (JpK)·(K'pJ'). Values from Delcourt & Blake (1991) were used.

FIGURE LEGENDS

FIGURE 1. The 373 bp DNA sequence between the EcoRI and RsaI sites from pUC8-31 plasmid is shown. Positions -43 and -36 are indicated. The DNA fragment from the plasmid pUC8-36 has the same sequence except for a G to A substitution at position -36. The upstream primers, and the downstream primer, DP15, employed in PCR are indicated. Upstream primers created base pair changes at the positions underlined. Single-base deletions in the upstream primers are indicated with a "-" at the positions.

FIGURE 2. Parallel temperature gradient gel of 372/373 bp DNAs with all base pairs and single-base bulges at position -36. Samples were run for 17 to 20 hrs at 90 volts. (A) Temperature gradient was from 35.5 to 42.5 °C. From top to bottom in each lane DNA bands contain the following bases at position -36: 1)* -/C, G/-, *-/-, G·C; 2) C/-, -/G, -/-, C·G; 3) A/-, -/T, A·T, -/-; 4) T/-, -/A, T·A, -/-. (B). Temperature gradient was from 32 to 35 °C. The top two DNA bands in each lane are the same as in (A). The lowest band in each lane contains both homoduplex DNAs, e.g., lane 1) -/C, G/-, -/- and G·C. (* Note: -/C means no base on top strand and C on bottom strand, -/- means the base pair at the position has been deleted.)

FIGURE 3. Parallel temperature gradient gel of 372/373 bp DNAs with all base pairs and single-base bulges at position -38. Temperature gradient was 36 to 42 °C. Samples were run for 20 hrs at 90 volts. From top to bottom in each lane DNA bands contain the following bases at position -38: 1) -/A and T/-, T·A, -/-; 2) -/T, A/-, A·T, -/-; 3) C/- and -/G, -/-, C·G; 4) -/C, G/-, -/-, G·C.

FIGURE 4. Parallel temperature gradient gel of 372/373 bp DNAs with all paired bases and single unpaired bases at position -39. Samples was run for 20 hrs at 90 volts with a temperature gradient from 36 to 41 °C. From top to bottom in each lane DNA bands contain the following bases at position -39: 1) T/- and -/A, T·A, -/-; 2) -/T, A/-, A·T, -/-; 3) C/-, -/G, -/-, C·G; 4) -/C, G/-, -/-, G·C.

FIGURE 5. Parallel temperature gradient gel of 372/373 bp DNAs with all paired bases and single unpaired bases at position -43. Samples was run for 16.5 hrs at 90 volts with a temperature gradient from 33 to 36 °C. From top to bottom in each lane DNA bands contain the following bases at position -43: 1) T/-, -/A, T·A and -/-; 2) A/-, -/T, A·T and -/-; 3) C/-, -/G, -/- and C·G; 4) G/-, -/C, -/- and G·C.

FIGURE 6. A typical perpendicular temperature gradient gel of 373 bp PCR fragments. The transition curves only show the first melting domain of the DNA fragments. Electrophoresis was conducted for 14.5 hrs at 90 volts. The temperature gradient was 26 to 44 °C from left to right. The sample contained melted and reannealed DNAs with -/- and T·A at position -36, and the native DNA with G·C at the same position. The transition from left to right correspond to DNAs with the following bases at position -36: T/-, -/A, T·A, -/-, G·C.

FIGURE 7. **(a)** The normalized experimental mobility transition curve of the 373 bp DNA with wild type sequence (———) is compared to the predicted mobility curve (— · —) calculated using eq. (1). In the experiment the DNA was run for 14 hours at 90 volts in a perpendicular temperature gradient gel containing 48% denaturant and 6.5 % polyacrylamide. Normalization procedures for the experimental curve and the theoretical parameters for the calculated curve are described in the text. **(b)** The derivative melting curve of the 373 bp

DNA calculated using the parameters described in the text for the mobility curve of (a).

Upstream Primers:

```
       5'                                3'
UP36C:     AATTCCATTTTTCGAGCTTTA
UP36T:     AATTCCATTTTTCGAGTTTTA
UP36D:     AATTCCATTTTTCGAG-TTTAA
UP38T:     AATTCCATTTTTCGTGGTTTA
UP38G:     AATTCCATTTTTCGGGGTTTA
UP38C:     AATTCCATTTTTCGCGGTTTA
UP38D:     AATTCCATTTTTCG-GGTTTA
UP39T:     AATTCCATTTTTCTAGGT
UP39C:     AATTCCATTTTTCCAGGT
UP39A:     AATTCCATTTTTCAAGGT
UP39D:     AATTCCATTTTTC-AGGT
UP43G:     AATTCCATTGTTCGAGGT
UP43C:     AATTCCATTCTTCGAGGT
UP43A:     AATTCCATTATTCGAGGT
UP43D:     AATTCCATT-TTCGAGGTTTA
UP16:      AATTCCATTTTTCGAG
                   ↕   ↕↕ ↕                                          ↓
      5'   AATTCCATTTTTCGAGGTTTAAATCCTTATCGTTATGGGTATTGTTTGTAATAGGACAA
                              A (pUC8-36)
                      -43     -36                                    +1

           CTAAAACGACAAGAGGATGGTGCTGAATATGGCAACTTTAACGGCAAAAGAAAGAACGG

           ACTTTACTCGGTCGACCTGCAGCCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTG

           ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCC

           AGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCT

           GAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCAC

           ACCGCATATGGTGCACTCT  3'

            3' GTATACCACGTGAGA  5'
```

DP15 (Downstream Primer)

Figure 1.

A) 1 2 3 4

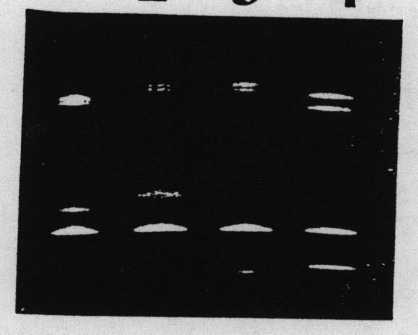


Figure 1 (a)

B) 1 2 3 4

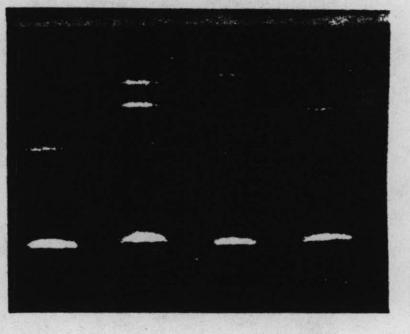


Figure 2 (b)

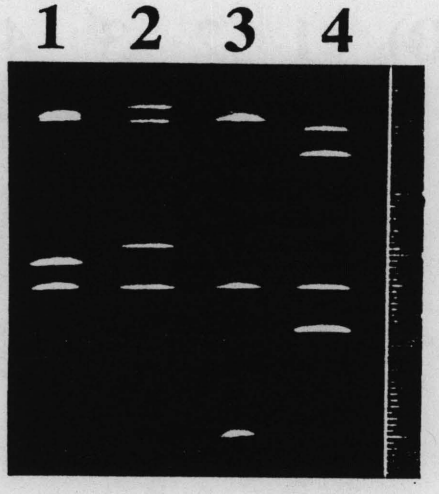


Figure 3

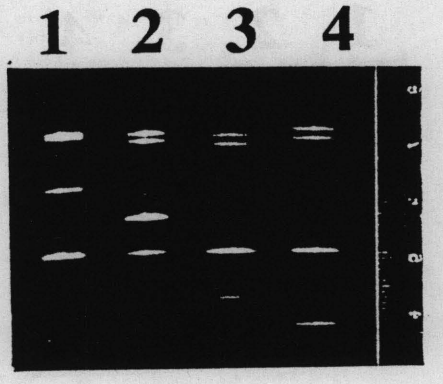


Figure 4

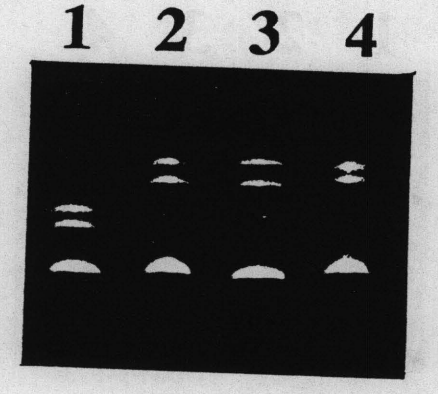
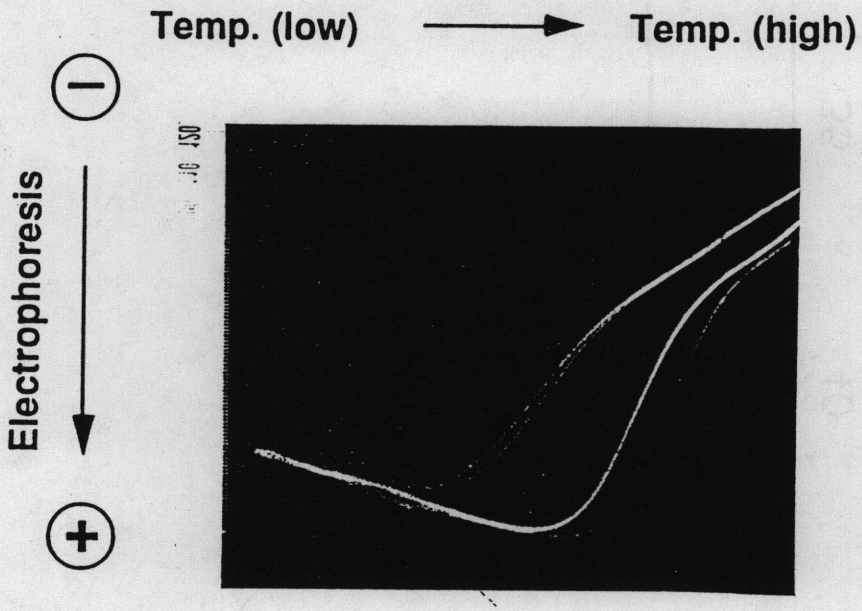


Figure 5


Temp. (low)
Temp. (high)
Electrophoresis
−
+

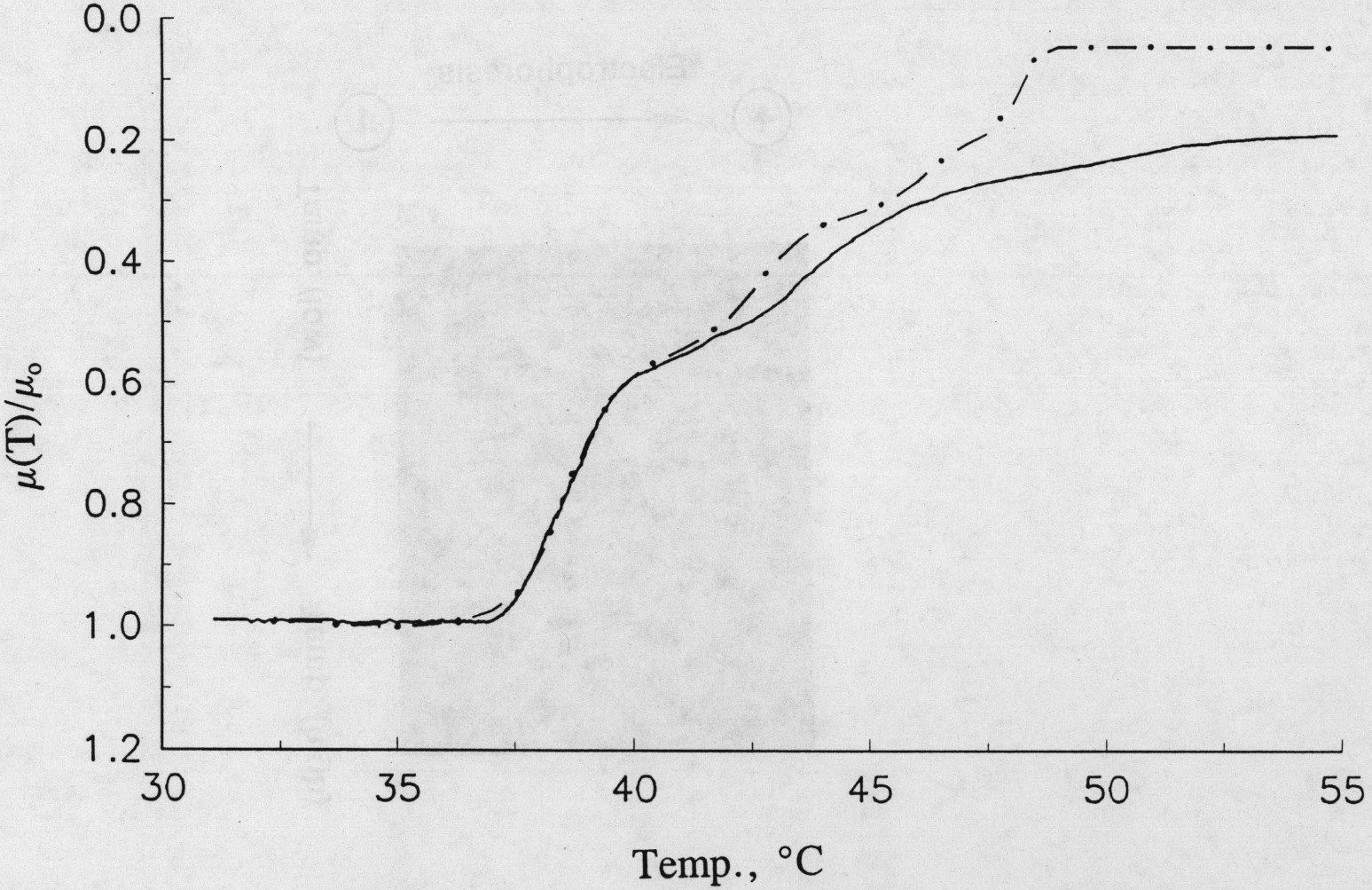


Figure 7a

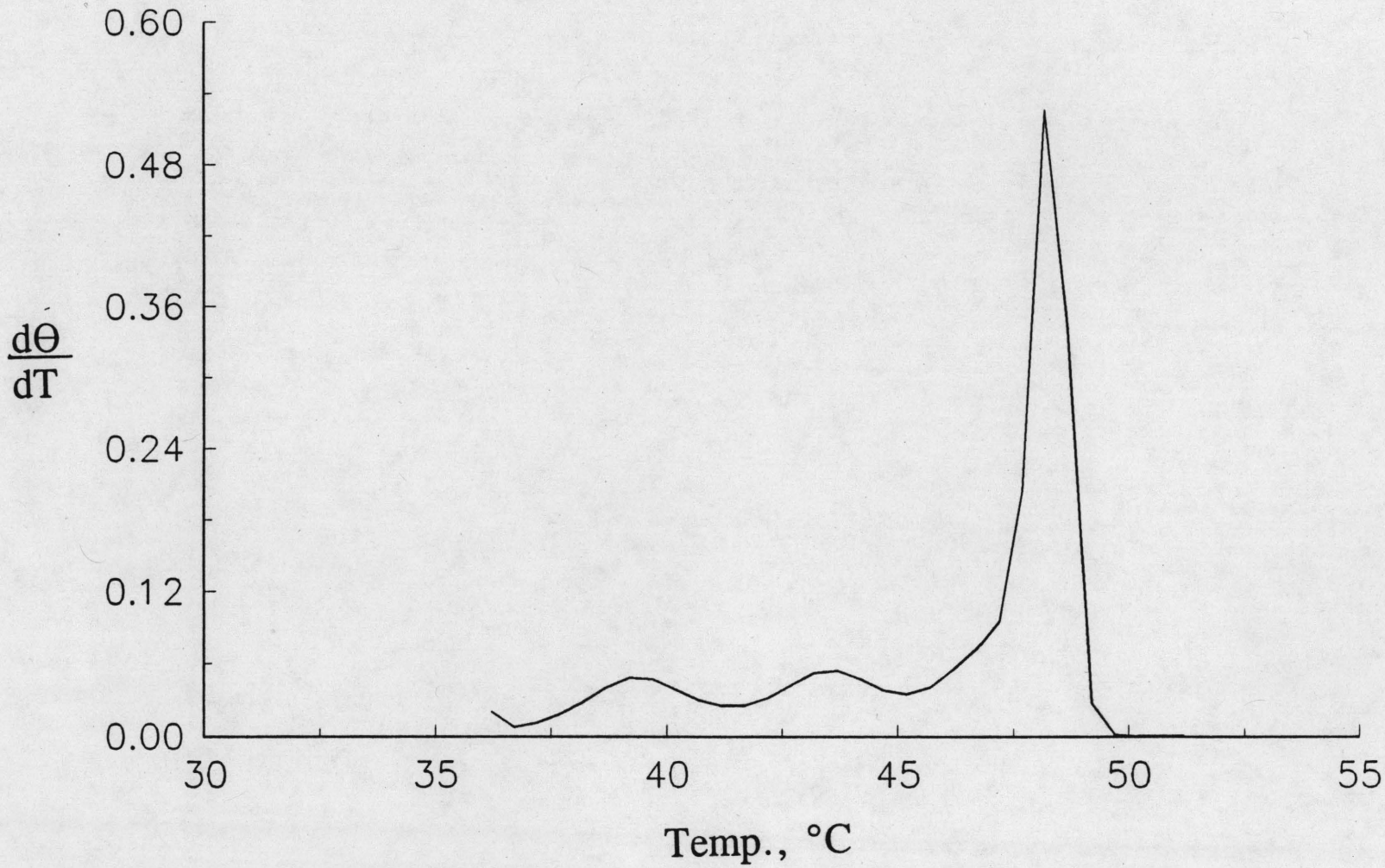


Figure 7b

IDENTIFICATION OF STREPTOMYCIN- AND RIFAMPICIN-RESISTANT MUTANTS OF *M. smegmatis* AND IDENTIFICATION OF MUTATIONS BY TEMPERATURE GRADIENT GEL ELECTROPHORESIS.

Gordon Churchward*, Teresa Kenney*, Yi-Yi Yu*, Sandra Powell**, and Roger Wartell**,
* Department of Microbiology and Immunology, Emory University, Atlanta, GA 30322 and **School of Biology, Georgia Institute of Technology, Atlanta, GA 30332.

To determine the molecular basis of drug resistance in mycobacteria, two collections of spontaneous mutants of *M. smegmatis* resistant to streptomycin or rifampicin were isolated. Using genetic complementation experiments, the streptomycin resistant mutants were shown to affect the *rpsL* gene of *M. smegmatis*. This gene encodes a ribosomal protein. DNA fragments containing either the *rpsL* gene (streptomycin-resistant mutants) or a segment of the *rpoB* gene (rifampicin-resistant mutants) were amplified from chromosomal DNA isolated from the mutants, and the nucleotide sequence of the amplified DNA fragments was determined. A majority of the streptomycin-resistant mutants were caused by one of four single nucleotide changes that affected two codons. A majority of the rifampicin-resistant mutations contained one of two single nucleotide changes affecting a single codon. In both cases similar mutations cause resistance in other organisms. To develop a system to detect drug-resistant mycobacteria, the melting behavior of fragments from the *rpsL* and *rpoB* genes was analyzed theoretically to select appropriate sequences so that wild type and mutant DNA fragments produced by PCR amplification would show different migration during temperature gradient gel electrophoresis. All mutant fragments tested behaved as expected during electrophoresis and could be distinguished from wild type fragments.

2344–2350 Nucleic Acids Research, 1994, Vol. 22, No. 12

Localization of the intrinsically bent DNA region upstream of the *E.coli rrnB* P1 promoter

Tamas Gaal, Lin Rao, Shawn T.Estrem, Jin Yang[1], Roger M.Wartell[1] and Richard L.Gourse*
Department of Bacteriology, University of Wisconsin–Madison, 1550 Linden Drive, Madison, WI 53706 and [1]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received March 3, 1994; Revised and Accepted May 8, 1994

ABSTRACT

DNA sequences upstream of the *rrnB* P1 core promoter (−10, −35 region) increase transcription more than 300-fold *in vivo* and *in vitro*. This stimulation results from a *cis*-acting DNA sequence, the UP element, which interacts directly with the alpha subunit of RNA polymerase, increasing transcription about 30-fold, and from a positively acting transcription factor, FIS, which increases expression another 10-fold. A DNA region exhibiting a high degree of intrinsic curvature has been observed upstream of the *rrnB* P1 core promoter and has thus been often cited as an example of the effect of bending on transcription. However, the precise position of the curvature has not been determined. We address here whether this bend is in fact related to activation of rRNA transcription. Electrophoretic analyses were used to localize the major bend in the *rrnB* P1 upstream region to position approximately −100 with respect to the transcription initiation site. Since most of the effect of upstream sequences on transcription results from DNA between the −35 hexamer and position −88, i.e. downstream of the bend center, these studies indicate that the curvature leading to the unusual electrophoretic behavior of the upstream region does not play a major role in activation of rRNA transcription. Minor deviations from normal electrophoretic behavior were associated with the region just upstream of the −35 hexamer and could conceivably influence interactions between the UP element and the alpha subunit of RNA polymerase.

INTRODUCTION

The P1 promoters of the seven *E.coli* ribosomal RNA operons are among the strongest in the cell with the capacity to produce as much RNA as all the other cellular promoters combined. This exceptional strength results from a near consensus core promoter (−10, −35 region) activated more than 300-fold by an upstream activator region (UAR) (1–3).

The UAR contains two functionally distinct elements (Fig. 1B). A promoter proximal region, the UP element, is a *cis*-acting DNA sequence located between −60 and −40 with respect to the transcription start site which interacts directly with the C-terminal portion of the α subunit of RNA polymerase and increases transcription about 30-fold (2–6 and W.Ross, E.Blatter, R.H.Ebright, and R.L.Gourse, unpublished results). Furthermore, a promoter distal region, located between −150 and −60, increases transcription about 10-fold by binding the FIS protein at three sites (7). FIS Site I, the site closest to the promoter, accounts for 70–80% of the effect of FIS on transcription (7). FIS most likely stimulates transcription by interacting directly with RNA polymerase (6, 8, and A.J.Bokal, W.Ross, and R.L.Gourse, unpublished results).

DNA fragments containing the *rrnB* P1 promoter region have retarded mobility on polyacrylamide gels (1, 9), a feature characteristic of intrinsic DNA curvature (10). DNA curvature is generally believed to be a function of oligo-dA tracts repeated in phase with the DNA helix (11–13), although curvature without A tracts has also been reported (14). It is possible to localize curvature within a fragment by electrophoresis, since DNA fragments in which the bend is at the center are substantially more retarded than fragments with the bend near an end (10, 15, 16). Furthermore, the ratio of the apparent length of a DNA fragment to its actual length can be correlated to an apparent DNA bend angle based on co-electrophoresis of fragments containing A_6 tracts (17, 18). Detection of DNA bending is facilitated by electrophoresis in the cold, by the presence of Mg^{2+} (19), and by high polyacrylamide concentrations (11). The relationship between bending and electrophoretic mobility is most likely an approximation which could also be affected by other structural distortions in DNA (13).

Predictions from computer analyses indicate that bent DNA sequences often occur in the 5′ flanking regions of protein coding sequences (20) and are preferentially located upstream of strong *E.coli* promoters (21). Direct cloning of fragments displaying unusual electrophoretic mobility also suggested that curvature is often associated with promoters (22). A+T-rich upstream DNA displaying curvature can sometimes replace a transcriptional activator (e.g. see 23–25). However, in some cases it has not been shown that the position of the bend and the position of the sequences which increase transcription are the same. Furthermore, when the sequences do overlap, it is difficult to distinguish effects of curvature on RNAP activity (e.g. by

*To whom correspondence should be addressed

facilitating RNAP binding or DNA strand opening) from effects of the A+T-rich sequences on transcription independent of the curvature.

In this report, we address the relationship between rRNA transcription activity and DNA bending by identification of the sequences required for DNA curvature and comparison of those regions with information now available about the regions required for stimulation of transcription of *rrnB* P1 (2, 3, 7). Our results suggest that the large intrinsic curvature in the vicinity of the *rrnB* P1 promoter does not play a major role in rRNA transcription.

MATERIALS AND METHODS

General methods

Plasmid DNA was purified using Qiagen columns (Qiagen Inc.), phenol extracted, and ethanol precipitated. Restriction enzymes and linkers were obtained primarily from New England Biolabs, and sequencing primers were provided by the NutraSweet Co. (Mt Prospect, Illinois). All plasmid constructions were confirmed by DNA sequencing using Sequenase (USB).

Plasmids

Restriction fragments are referred to by their *rrnB* P1 DNA endpoints (with respect to the transcription start site, +1). Eight bp *EcoRI* and 12 bp *BamHI* linkers were ligated to the upstream and downstream ends, respectively, of a fragment containing the *rrnB* P1 sequences from −154 (*AluI*) to −28 (*HaeIII*), and inserted into the *EcoRI* and *BamHI* sites of pUC19 and pRW4 to make pJNBend130 and pJY237, respectively (J.T.Newlands, R.L.Gourse, J.Yang, and R.M.Wartell, unpublished results). Fragments indicated in Table I were generated by digestion of these plasmids with *EcoRI* (−154), *BamHI* (−28), *DraI* (−46), and *TaqI* (−104 and −114). Other restriction fragments (illustrated in Figure 1B) were obtained as *EcoRI*−*HindIII* fragments from previously existing plasmids (3, and L.Rao and R.L.Gourse, unpublished results), and inserted between the unique *EcoRI* and *HindIII* sites of the permutation vector pSL6 (8), a derivative of pBEND2 (16) (Figure 1A).

Analysis of DNA bending by polyacrylamide gel electrophoresis

Fragments from pJY237 were examined at 25°C on 8% acrylamide:bisacrylamide (29:1) gels in 1×TBE buffer (26) for 8.5 hours at 5−7 V/cm as described (17). *HaeIII* and *HinfI* digested pBR322 served as size standards. pSL6 derivatives were digested separately with at least seven different restriction enzymes to generate a set of fragments of equal length containing the promoter DNA segment of interest at different positions within the restriction fragment. The fragments were electrophoresed on 220×200×1.5 mm 10% acrylamide: bisacrylamide (30:1) gels in 0.5×TBE buffer at 3V/cm for 48 hours at 6°C. A *HaeIII* digest of pUC19 was used as a molecular weight standard. As controls, identical digests were run under conditions where bent and non-bent fragments migrate similarly, e.g. in the presence of 0.5 μg/ml ethidium bromide or at 55°C (25 V/cm) (27). After electrophoresis, gels were stained with ethidium bromide and photographed with UV illumination. The apparent length of each fragment was determined from semi-logarithmic plots of molecular weight versus migration distance and expressed as a ratio of the fragment's actual size (K value).

The errors in determining a K value for a specific fragment between different experiments were usually quite small (≤ 4%). Furthermore, differences in electrophoretic mobility between fragments of the same size but with either different sequence composition or different endpoints could be detected and visualized reproducibly when the fragments were in adjacent gel lanes or within the same gel lane. Thus, differences in the mobilities of fragments in certain DNA regions which have K values very close to 1.0 (zero curvature) could still be detected reproducibly.

Construction of mutations in the UP element and evaluation of their effects on promoter strength

Mutations were constructed using the method of Kunkel (28). Primer 5′-ATTTAAAATAATATTCTGACCGCG-3′ was used to create a substitution of a T for an A at position −55 (A−55T). Primer 5′-GACAAGAGGAATATTAAAATAATT-3′ was used to create the double substitution T−43A, A−44T. Operon fusions with *lacZ* were constructed in strain NK5031, and β-galactosidase activities were measured as described previously (29).

RESULTS

Detection of bending by electrophoresis

Intrinsic DNA curvature can be identified by unusual electrophoretic behavior. The K value is the ratio of the apparent length of a fragment to its actual length: the greater the K value, the larger the deviation in apparent size from real size, which can be indicative of an increase in the bend angle. The K value is also a function of acrylamide concentration, electrophoresis conditions, the length of the DNA fragment, and the position of the bend within the fragment (10, 15−19). The closer the bend is to the center of the fragment, the slower is the electrophoretic mobility.

An rRNA promoter containing *rrnB* P1 endpoints −154 to +50 with respect to the transcription initiation site was shown previously to have maximum rRNA transcription activity and to migrate aberrantly on acrylamide gels (1). K values of the *rrnB* P1 sub-fragments listed in Table 1 were determined at 25°C and indicated that an anomalous structure, perhaps a bend, was located between −154 and −46. Assuming the anomalous structure represents DNA bending, these data were not sufficient to determine the distribution of the curvature within this region nor the location of the bend center. However, the −154 to −114 and −104 to −46 sub-fragments had normal mobilities, suggesting that the endpoints in the −100 region either disrupted the curved region or placed it too close to the end of the fragment

Table 1. K Values for restriction fragments containing UAR DNA

fragment endpoints[a]	K value[b]
−154 to −28	1.29
−154 to −46	1.20
−104 to −28	1.05
−104 to −46	1.00
−154 to −114	1.00

[a]*rrnB* P1 sequence coordinates. +1 is the transcription initiation site.
[b]The K value is the apparent electrophoretic mobility relative to the actual size of the fragment. Electrophoresis was performed at 25°C.

to affect electrophoretic migration substantially. The K value of 1.29 for the −154 to −28 fragment resulted in an estimate of a bend angle of approximately 108° when calculated using the empirical equations developed by Koo and Crothers (17) for DNAs with A_6 tracts, a value of 18° per A_6 tract (30, 31), and a value of 10.5 bp/helix turn (i.e. the retarded mobility of this region was equivalent to that which would be afforded by six A_6 tracts).

We next analyzed the rRNA promoter region using the permutation vector pSL6 (8, 16) illustrated in Figure 1A. Fragments of DNA associated with different functionally important sequences in the region of the *rrnB* P1 promoter (Figure 1B) were cloned between the two tandem repeats in the vector. Cleavage at different restriction sites in the tandem repeat generated fragments of the same absolute length but with the bend (if present) at different positions with respect to the ends. The apparent sizes of the fragments were then determined by electrophoresis. Since multiple bends within a fragment can limit the usefulness of this technique, we examined relatively short pieces of DNA.

A set of at least seven different restriction digestions was performed on plasmid constructs containing each *rrnB* insert. The set of digests containing a single *rrnB* insert was electrophoresed on a 10% polyacrylamide gel at 6°C alongside known length markers (data not shown). The digests generating the slowest and fastest running fragments from each permuted fragment set were then electrophoresed on the same gel to permit direct comparison of the K values of different *rrnB* inserts. The results are shown in Figure 2A, illustrated graphically in Figure 3A, and tabulated in Table 2. Only the fragment containing the −154 to −64 insert (and the larger fragments which contain this region, −154 to −46 and −154 to +50) had K values of 1.4 to 1.5. All other fragments tested, including several containing the UP element region with a variety of fragment endpoints, deviated in mobility from their actual sizes by less than or equal to about 10%. These results indicated that the major bend in the *rrnB* P1 promoter region is located upstream of the UP element.

Electrophoresis of the same digests in the presence of 0.5 μg/ml ethidium bromide or at 55°C, either of which reduces the effect of bending on electrophoretic mobility (32), results in migration

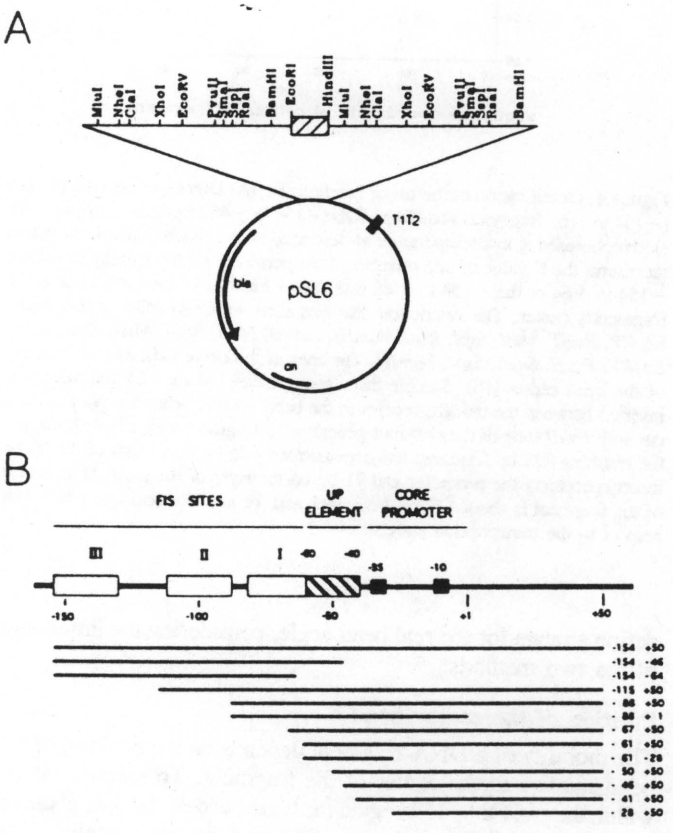


Figure 1. (A) Plasmid pSL6 (8). The plasmid has two identical regions of DNA containing multiple restriction sites flanking *EcoRI* and *HindIII* sites for promoter fragment insertion (hatched box). T1T2: transcription termination region from *rrnB*. bla: ampicillin resistance gene. ori: origin of DNA replication. (B) The *rrnB* P1 promoter region and DNA fragments used for permutation analysis. The core promoter region, UP element, and FIS binding sites are indicated (2, 3, 7). The endpoints of *rrnB* fragments inserted into pSL6 are shown on the right. The resulting plasmids are listed in Table 2. The fragments are displayed under the schematic in order to illustrate the functionally relevant regions contained within each fragment.

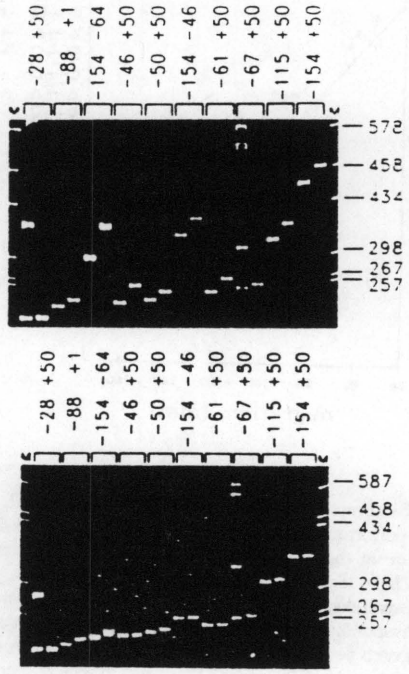


Figure 2. Gel electrophoresis of permuted fragments of *rrnB* P1. (A) electrophoresis performed at 6°C. (B) electrophoresis performed at room temperature in the presence of 0.5 μg/ml ethidium bromide. In each digest, the *rrnB* insertion is embedded within 137 bp of vector DNA. The two restriction digests resulting in the maximal and minimal electrophoretic mobilities of each permuted fragment set were run on the same gel. Plasmid DNAs used in each lane are identified by the endpoints of their *rrnB* DNA. The top of the gel is not pictured. Restriction enzymes used in each lane were: M: pUC19 digested with *HaeIII* (molecular weight standard; sizes of the fragments in base pairs are indicated at the right of each panel); −28 to +50: *SspI*, *MluI*. (*SspI* cleaves at multiple sites within the vector resulting in additional higher molecular weight bands); −88 to +1: *MluI*, *PvuII*; −154 to −64: *MluI*, *SmaI*; −46 to +50: *BamHI*, *XhoI*; −50 to +50: *BamHI*, *XhoI*; −154 to −46: *BamHI*, *EcoRV*; −61 to +50: *BamHI*, *MluI*; −67 to +50: *RsaI*, *MluI*. (*RsaI* cleaves at multiple siteswithin the vector resulting in additional bands. The *rrnB* sequences are contained within the fragment migrating near the position of the 257 and 267 bp standards.); −115 to +50: *BamHI*, *EcoRV*; −154 to +50: *BamHI*, *MluI*.

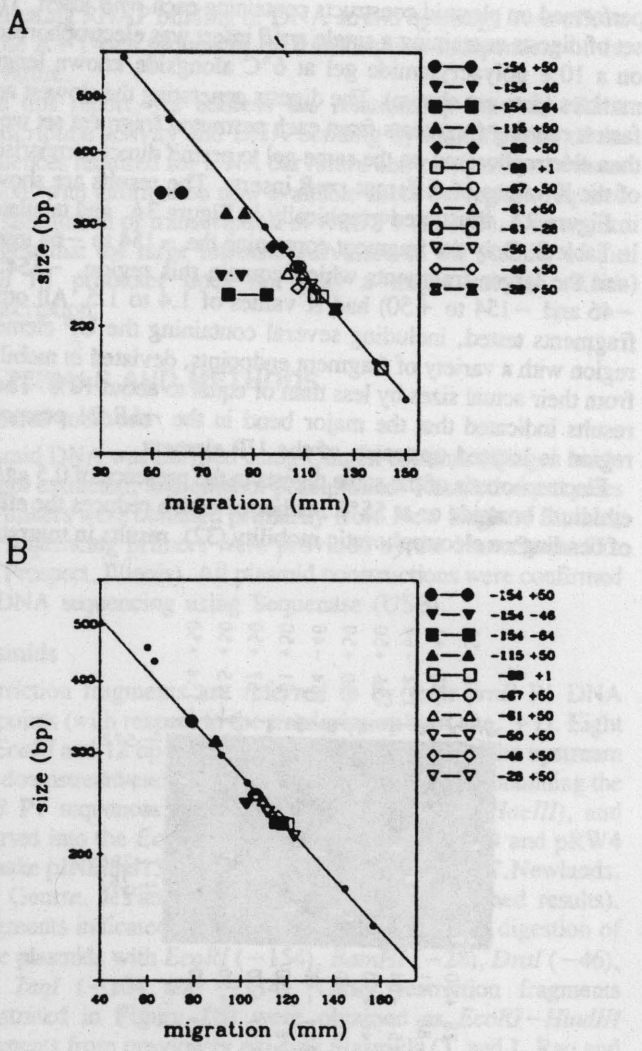


Figure 3. Graphical representation of data pictured in Figure 2. (A) Data from electrophoresis performed at 6°C. (B) Data from electrophoresis performed at room temperature in the presence of 0.5 μg/ml ethidium bromide. Apparent fragment length in bp is plotted versus distance migrated in mm. A line is drawn through points (small dots) derived from the molecular weight standard. The pair of symbols corresponding to the maximum and minimum apparent lengths of fragments from each permuted set are connected. The symbols for each set are defined on the right side of the graph. From this graphical representation, it is apparent that the fragments containing *rrnB* sequences −154 to +50, −154 to −46, and −154 to −64 migrate substantially slower at 6°C than expected from their lengths.

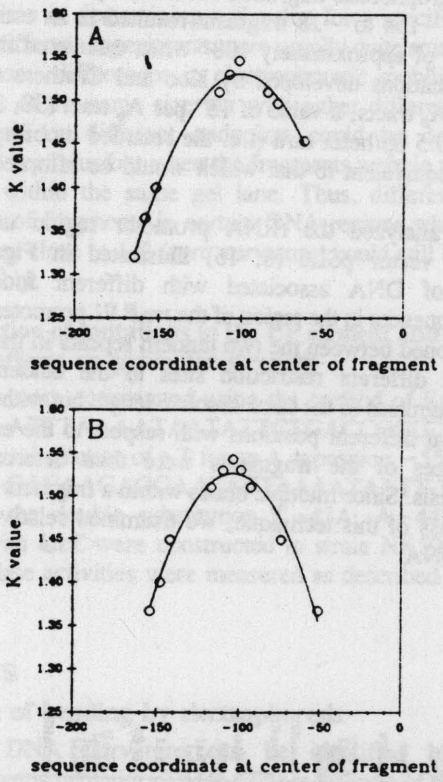


Figure 4. Identification of the major bend center. (A) Digests of plasmid pLR18 (−154 to −64 fragment) and (B) pLR20 (−154 to −46 fragment). Digests were electrophoresed at low temperature as described above. Each point on the graph represents the K value of one member of the permuted set (containing either the −154 to −64 or the −154 to −46 insert) as a function of the coordinate at the fragment's center. The restriction enzymes used were (A) *MluI*, *BglII*, *NheI*, *EcoRV*, *PvuII*, *SmaI*, *SspI*, *RsaI*, *BamHI*; and (B) *MluI*, *BglII*, *NheI*, *XhoI*, *DraI*, *EcoRV*, *PvuII*, *SmaI*, *SspI*, *BamHI*. The apex of the curve indicates the position of the bend center (16). Sample calculation: the −154 to −64 fragment was inserted between the tandem repeats in the bend vector. When the plasmid was cut with *PvuII* (one of the enzymes generating a fragment with a high K value), the resulting 227 bp fragment had approximately 66 bp on the left of the 90 bp insert containing the promoter and 71 bp on the right of the insert. The center of the fragment is about 113 bp from each end, or at about position −107 with respect to the transcription start site.

almost as predicted by the actual fragment sizes (Figure 2B, Figure 3B, Table 2, and data not shown). Thus, deviations from normal mobility are a function of the structure of the fragment and not from miscalculation of actual length.

The electrophoretic mobility of the −154 to −64 fragment was also compared directly with that of fragments containing different numbers of phased A tracts from the Thompson−Landy set (18) in order to obtain an estimate of the major bend angle occurring within the *rrnB* promoter region. The −154 to −64 fragment migrated with a K value between that of fragments with 4 and 5 A-tracts, providing an estimated bend angle of about 80° (18). This estimate of 80° and that noted above of 108° likely define a range for the real bend angle, considering the limitations of the two methods.

Position of the major bend

The mobility of a DNA fragment depends on the position of the bend relative to the center of the fragment. To identify where within the −64 to −154 region the bend resides, the data obtained from a set of 9 different digests of the vector containing the −154 to −64 fragment and a set of 10 different digests of that containing the −154 to −46 fragment were plotted as Figures 4A and 4B, respectively. In these figures, the K value for each of the fragments in the set was plotted against the sequence coordinate at the center of the fragment. The fragments with the highest K value are those where the center of the fragment is at approximately −100, thereby identifying this position as the bend center. This position is well upstream of both the UP element and FIS Site I, the major upstream determinants of promoter strengh at *rrnB* P1.

Table 2. K values from circular permutation experiments[a]

plasmid	rrnB P1 endpoints	actual size (bp)[b]	apparent size (bp)[c]		K values[d]	
			6°C	Etbr	6°C	Etbr
pSL6	−154+50	(341)	440 – 484	356 – 356	1.29 – 1.42	1.04 – 1.04
pLR20	−154−46	(245)	328 – 365	264 – 270	1.34 – 1.49	1.08 – 1.10
pLR18	−154−64	(227)	293 – 350	227 – 232	1.29 – 1.54	1.00 – 1.02
pSL8	−115+50	(302)	328 – 353	309 – 315	1.09 – 1.17	1.02 – 1.04
pSL9	−88+50	(275)	267 – 289	nd[e]	0.97 – 1.05	nd
pSL13	−88+1	(226)	224 – 233	221 – 226	0.99 – 1.03	0.98 – 1.00
pSL10	−67+50	(254)	249 – 254	249 – 254	0.98 – 1.00	0.98 – 1.00
pSL11	−61+50	(248)	246 – 265	245 – 248	0.99 – 1.07	0.99 – 1.00
pLR1	−61−28	(170)	170 – 177		nd	1.00 – 1.04
pSL12	−50+50	(237)	235 – 246	237 – 240	0.99 – 1.04	1.00 – 1.01
pLR21	−46+50	(233)	233 – 256	230 – 230	1.00 – 1.10	0.99 – 0.99
pLR9	−41+50	(228)	217 – 233		nd	0.95 – 1.02
pLR19	−28+50	(215)	219 – 219	210 – 215	1.02 – 1.02	0.98 – 1.00

[a]As described in Materials and Methods and in text.
[b]In each case, the *rrnB* insertion is embedded within 137 bp of vector DNA.
[c]Minimum and maximum apparent lengths of a set of permuted fragments of the same size with different endpoints, from electrophoresis at 6°C or in the presence of ethidium bromide (Etbr).
[d]Minimum and maximum K values (apparent size/actual size) of a set of permuted fragments of the same size with different endpoints, from electrophoresis at 6°C or in the presence of ethidium bromide (Etbr). The K values represent the range of mobilities for a given fragment set. Errors in K value determination are within 4% (see Methods).
[e]nd: not determined.

Minor deviations from linearity in fragments not containing the major bend

Examination of fragments downstream of the major bend locus (−67 to +50; −61 to +50; −50 to +50; −46 to +50) did not lead to a simple interpretation of curvature in this region. For example, the −67 to +50 fragment migrates more normally than the −61 to +50 or −46 to +50 fragments. It is possible that there are multiple small deviations from linearity caused by these sequences (or by the sequences created at the *rrnB*-vector junctions), some of which compensate for one another by bending in opposite directions.

We tested whether these minor electrophoretic abnormalities reflect curvature in the UP element (−40 to −60) region by comparing the mobility of a fragment with −88 to +1 *rrnB* endpoints to that of an identical fragment containing a substitution of the sequences from −41 to −59 [the SUB mutation, sequence provided in legend of Figure 5 (2, 3)]. The wild type fragment ('W'; Figure 5), which contains two A-tracts in phase, migrated only slightly slower than the fragment of the same length containing the SUB mutation ('S'), consistent with the results reported in Table 2 indicating that only minor curvature exists in this region. (The maximum K value of the wild type fragment was about 1.03, while the K value of the SUB fragment was about 0.98.) The small difference in the mobilities of the wild type and SUB fragments is likely to reflect bending, however, since the difference was greatest when the A-tract regions were near the center of the fragment (Figure 5, *PvuII* or *XhoI* digests).

Functional role of A-tracts within the UP element

The UP element consists of 90% adenine and thymine residues, contains phased A/T-tracts (Figure 6), and increases transcription of *rrnB* P1 at least 30-fold *in vivo* and *in vitro* (2, 3). Replacement of the −40 to −60 region in *rrnB* P1 with the SUB mutation completely eliminates activation by the UP element and partially reduces activation by FIS bound at FIS Site I (3), resulting in a reduction of transcription even greater than 30-fold.

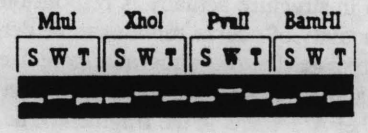


Figure 5. Electrophoretic mobility of fragments containing *rrnB* P1 derivatives with substitutions upstream of the −35 hexamer. Promoter fragments with *rrnB* P1 sequences from −88 to +1 were inserted into pSL6 and digested with either *MluI*, *XhoI*, *PvuII*, or *BamHI*. S, SUB mutant (2, 3); W, wild type; T, C−37T substitution (29, 34). Electrophoresis was performed at 6°C as described for the other figures. The sequence of the SUB mutation in the UP element of *rrnB* P1 is (−59) 5'-GACTGCAGTGGTACCTAGG-3' (−41) (3).

The functional role of the phased A·T-tracts in the −40 to −60 region was addressed by disrupting the A/T-tracts with substitution mutations without changing the A+T content. The T-tract at −41 to −43 and the A-tract at −44 to −46 (top strand sequences provided in Figure 6) were altered with a double substitution (T−43A, A−44T). In a separate experiment, the −54 to −57 A-tract was disrupted with a single substitution, A−55T. The mutant promoters (with *rrnB* endpoints of −88 to +1) were fused to *lacZ* and inserted into the bacterial chromosome as monolysogens of bacteriophage lambda. The activities of these fusions are shown in Table 3, compared to a fusion with the wild type promoter with the same endpoints. The A/T-tract mutations reduce transcription only 1.3 to 1.6 fold, implying that the continuous runs of A or T residues do not account for the 30-fold effect of the UP element.

Altered DNA structure from a mutation at −37

We have reported previously that a mutation at position −37 (C−37T) reduces *rrnB* P1 promoter activity 10−20 fold, but that substitution of an A or G residue at this position reduces activity only 2−3 fold (29, 34). We speculated that the large allele-specific effect of the T substitution might result not from

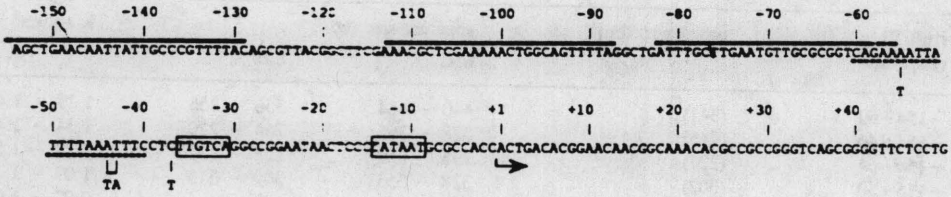


Figure 6. DNA sequence of the *rrnB* P1 region [from (33)]. The three FIS Sites are indicated above the sequence, the UP element is indicated below the sequence as a broken line, the −10 and −35 hexamers are boxed, and the transcription start site is indicated by an arrow. The substitutions described in the text at positions −37, −43, −44, and −55 are also indicated.

the importance of the wild type C residue to promoter function, but from the creation of a series of four T residues in phase with another T tract upstream, which might result in a bend that might be detrimental to promoter function *in vivo*.

In order to test this hypothesis, we compared the mobility of the promoter fragment containing the wild type −88 to +1 region with that of the same promoter fragment containing the C−37T mutation at position −37. As shown in Figure 5, the fragment containing the C−37T mutation ('T') migrates differently from that containing the wild type promoter region. Whether or not this alteration in structure actually is responsible for the defect in promoter activity of course remains to be determined. Since the C−37T promoter fragment migrates more like that containing the SUB mutation, the newly created A/T tract may be out of phase with other A/T tracts in the fragment. The resulting bends in opposing directions might then cancel overall effects on mobility.

DISCUSSION

Using permuted sets of restriction fragments carrying different portions of the *rrnB* P1 promoter and its upstream sequences we were able to localize a major structural anomaly, most likely a bend, centered around position −100. This conclusion agreed with the results of other electrophoretic analyses of the *rrnB* promoter region [Table 1 and (1)], and with conclusions based on the electrophoretic behavior of fragments containing linker scanning mutations in the upstream sequence (9). However, this bend is unlikely to play a prominent role in *rrnB* P1 transcription, since it is upstream of the sequences responsible for UP element function, growth rate dependent regulation, or stringent control of this promoter (1, 3, 35, 36 and C.A.Josaitis and R.L.Gourse, unpublished results). Likewise, the bend is upstream of FIS Site I, which accounts for the majority of FIS-mediated activation (7). It is conceivable, however, that the bend at about −100 could have a role in mediating the small effects on transcription activation attributable to FIS Sites II and III.

Two methods used to estimate the intrinsic bend angle resulted in a range of values from 80−108°. Both methods rely on certain problematical assumptions [see (17, 37)], but the presence of the upstream region gives DNA fragments a K value equivalent to that expected from the presence of four to six phased A_6 tracts. Figure 6 shows that there are five A (or T)-tracts 3bp or more in length (at −81, −92, −104, −113, and −134) in the vicinity of the bend center. Other sequences could add to the degree of curvature observed, and/or anomalies in structure in addition to bending could also contribute to the observed retarded electrophoretic mobility. Nevertheless, the position of the bend does not correlate with the region responsible for transcription activation, and as a result, we have not attempted to extend the structural analysis further in order to resolve these questions. The estimate of the bend angle, but not the bend position, should thus be regarded as tentative.

Table 3. Expression of fusions with A tract mutations

Promoter[a]	β-gal activity[b]
wild type	4970 ± 140
T−43A A−44T	3850 ± 177
A−55T	3030 ± 192

[a]Promoter−*lacZ* fusions are described in the text.
[b]Miller units. Under comparable conditions, the SUB mutation eliminating UP element function reduces β-galactosidase activity more than 30-fold (3).

The region upstream of −60 in each of the seven *rrn* operons contains multiple phased A/T tracts which might be expected to lead to unusual electrophoretic mobility (21) [see (38) for sequences]. However, the phased A/T tracts are not in exactly the same position as those that generate the bend centered at about −100 in *rrnB*, nor are they uniformly positioned relative to the core promoter. Therefore, their function, if any, remains unknown.

Fragments containing at least a portion of the *rrnB* UP element but without the −100 region exhibited a slight degree of aberrant electrophoretic mobility (R values less than or equal to 1.1, which is less than or equal to the deviation expected from the presence of two phased A/T tracts under our conditions). A mutation at position −37 which creates a T-tract increased the electrophoretic mobility of an *rrnB* P1 promoter fragment slightly, further indicating that a subtle, non-standard DNA structure probably exists in this region.

Three point mutations disrupting the A/T tracts in the wild type UP element had only small effects on transcription activity. Therefore, if the A/T tracts are responsible for the small degree of curvature associated with this region, either the mutations do not affect this curvature, or the curvature plays little role in the 30-fold increase in transcription resulting from UP element function. On the other hand, either displacement of the UP element by non-integral portions of a helical turn (6, 29, 34) or creation of an additional T tract with a T substitution at position −37 (29, 34) severely decreases promoter activity. Thus, correct positioning of the RNAP alpha subunit (which binds to the UP element) relative to the sigma subunit (which binds to the −10 and −35 hexamers) most likely plays a major role in promoter activity (2). The sequence and structural determinants responsible

for interaction of the RNA polymerase alpha subunit with the −40 to −60 region and thus for UP element function are under investigation. Although a high degree of intrinsic curvature appears not to play a role in this interaction, it is certainly possible that binding of alpha may be facilitated by subtle structural distortions determined by the DNA sequence.

In summary, although the obvious intrinsic curvature in the *rrnB* P1 upstream region is often cited as an example of the effects of DNA bending on transcription, apparently this bend plays little or no role in FIS-dependent activation, UP element function, or in core promoter strength, which together account for the high activity of rRNA promoters.

ACKNOWLEDGEMENTS

We thank Janet Newlands and the late Sigrid Leirmo for contributions in the initial stages of these studies, David Wheeler for discussions on computer modeling, and Wilma Ross for critical reading of the manuscript. This work was supported by research grants GM37048 (R.L.G.) and GM38045 (R.M.W.) from the National Institutes of Health and by an NIH predoctoral training grant (S.T.E.).

REFERENCES

1. Gourse, R.L., deBoer, H.A. and Nomura, M. (1986) *Cell* **44**, 197–205.
2. Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R.L. (1993) *Science* **262**, 1407–1413.
3. Rao, L. Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T. Jr. and Gourse, R.L. (1994) *J. Mol. Biol.* **235**, 1421–1435.
4. Leirmo, S. and Gourse, R.L. (1991) *J. Mol. Biol.* **220**, 555–568.
5. Newlands, J.T., Ross, W., Gosink, K.K. and Gourse, R.L. (1991) *J. Mol. Biol.* **220**, 569–583.
6. Newlands, J.T., Josaitis, C.A., Ross, W. and Gourse, R.L. (1992) *Nucleic Acids Res.* **20**, 719–726.
7. Ross, W., Thompson, J.F., Newlands, J.T. and Gourse, R.L. (1990) *EMBO J.* **9**, 3733–3742.
8. Gosink, K.K., Ross, W., Leirmo, S., Osuna, R., Finkel, S.E., Johnson, R.C. and Gourse, R.L. (1993) *J. Bacteriol.* **175**, 1580–1589.
9. Zacharias, M., Goringer, H.U. and Wagner, R. (1992) *Biochem.* **31**, 2621–2628.
10. Wu, H.-M. and Crothers, D.M. (1984) *Nature* **308**, 509–513.
11. Marini, J.C., Levene, S.D, Crothers, D.M and Englund, P.T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7664–7668.
12. Trifonov, E.N. (1985) *CRC Crit. Rev. Biochem.* **19**(2), 89–106.
13. Hagerman, P.J. (1990) *Annu. Rev. Biochem.* **59**, 755–781.
14. Bolshoy, A., McNamara, P., Harrington, R.E. and Trifonov, E.N. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2312–2316.
15. Prentki, P., Pham, M.H. and Galas, D.J. (1987) *Nucleic Acids Res.* **15**, 10060.
16. Kim, J., Zwieb, C., Wu, C. and Adhya, S. (1989) *Gene* **85**, 15–23.
17. Koo, H.-S. and Crothers, D.M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1763–1767.
18. Thompson, J.F. and Landy, A. (1988) *Nucleic Acids Res.* **16**, 9687- 9705.
19. Diekmann, S. and James,C.W. (1985) *J.Mol.Biol.* **186**, 1–11.
20. VanWye, J.D., Bronson, E.C. and Anderson, J.N. (1991) *Nucleic Acids Res.* **19**, 5253–5261.
21. Plaskon, R.R. and Wartell, R.M. (1987) *Nucleic Acids Res.* **15**, 785–796.
22. Tanaka, K., Muramatsu, S., Yamada, H. and Mizuno, T. (1991) *Mol. Gen. Genet.* **226**, 367–376.
23. Gartenberg, M.R. and Crothers, D.M. (1991) *J. Mol. Biol.* **219**, 217–230.
24. Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. and Buc, H. (1989) *EMBOJ.* **8**, 4289–4296.
25. Nachaliel, N., Melnick, J., Gafny, R. and Glaser, G. (1989) *Nucleic Acids Res.* **17**, 9811–9822.
26. Peacock, A.C. and Dingman, C.W. (1968) *Biochem.* **7**, 668–674.
27. Diekmann, S. (1992) *Meth. Enzymol.* **212**, 30–46.
28. Kunkel, T. A. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 488–492.
29. Gaal, T., Barkei, J., Dickson, R.R., deBoer, H.A., deHaseth, P.L., Alavi, H. and Gourse, R.L. (1989) *J. Bacteriol.* **171**, 4852–4861.
30. Levene, S.D., Wu, H.-M., and Crothers, D.M. (1986) *Biochem.* **25**, 3988–3995.
31. Griffith, J., Bleyman, M., Rauch, C.A., Kitchin, P.A., and Englund, P.T. (1986) *Cell* **46**, 717–724.
32. Diekmann, S. and Lilley, D.M. (1987) *Nucleic Acids Res.* **15**, 5765–5774.
33. Brosius, J., Dull, T.J., Sleeter, D.D., and Noller, H.F. (1981) *J. Mol. Biol.* **148**, 107–127.
34. Josaitis, C.A., Gaal, T., Ross, W. and Gourse, R.L. (1990) *Biochim. Biophys. Acta* **1050**, 307–311.
35. Dickson, R.R., Gaal, T., deBoer, H.A., deHaseth, P.L. and Gourse, R.L. (1989) *J. Bacteriol.* **171**, 4862–4870.
36. Zacharias, M., Goringer, H.U. and Wagner, R. (1989) *EMBO J.* **8**, 3357–3363.
37. Haran, T.E. and Crothers, D.M. (1989) *Biochem.* 28, 2763–2767.
38. Condon, C., Philips, J., Fu, Z.Y., Squires, C. and Squires, C.L. (1992) *EMBO J.* **11**, 4175–4185.