Analysis of DNA curvature distribution in mycobacterial promoters using theoretical models

Rupali N. Kalate[a], B.D. Kulkarni[a,*], V. Nagaraja[b]

[a]Chemical Engineering Division, National Chemical Laboratory, Pune 411 008, India
[b]Microbiology and Cell Biology Department, Indian Institute of Science, Bangalore 012, India

Abstract

In this paper, 125 different mycobacterial promoters are analyzed for their DNA curvature distribution using several di- and tri-nucleotide dependent models of DNA curvature. Different models give similar behavior and therefore qualitative validation of the results. Mycobacterial promoters resembling the *E. coli* σ^{70} type have almost 81% (85%) sequences having medium and high curvature profiles using dinucleotide-dependent models. Non-*E. coli* σ^{70} type mycobacterial promoters have comparatively higher percent of low curvature profiles. Very few extended -10 promoters have low curvature profiles. Mycobacterial promoters having $A_n T_m$ $(n+m \geq 3)$ tract in the upstream region of -35 box and repeated in phase with each other have high curvature profiles. *M. smegmatis* promoters have high curvature profiles compared to *M. tuberculosis* promoters.

Keywords: Mycobacterial promoters; Transcription; Curvature models

1. Introduction

Transcription process in *Mycobacteria* may differ from *E. coli* and many other bacteria as mycobacterial genome has high G+C content which affects codon usage and promoter recognition sites in an organism. Mycobacterial promoters like *M. tuberculosis* 65 kDa [1], *M. bovis BCG* 64 kDa [2], and *M. leprae* 65 kDa [3] are known to function in *E. coli*. However, mycobacterial promoters like *M. tuberculosis* 85A [4], recA [5] are known to be non-functional in *E. coli*. Thus, depending on the choice of expression host, myco-

bacterial promoters are classified as *E. coli* type and Non-*E. coli* type promoters. *M. smegmatis* and *M. tuberculosis* promoter analysis by Bashyam et al. [6] showed that occurrence of a TG motif near the -10 region is functionally significant for those having a non-functional -35 region. These promoters form a different class of promoters known as 'Extended -10 promoters'. The type of expression host, and the variation of the nucleotide sequence composition at the -35 and -10 regions of mycobacterial promoters [7] indicate that there exists immense variation in transcription initiation mechanism of mycobacterial promoters.

Transcription initiation is a multi-step, sequential process involving: (a) binding of RNA polymerase to the promoter leading to formation of a

*Corresponding author. Tel.: +91-20-589-3095; fax: +91-20-589-3041.
   E-mail address: bdk@ems.ncl.res.in (B.D. Kulkarni).

relatively weak closed initiation complex; (b) its isomerization to the more stable open complex that is accompanied by the separation of the DNA strands upstream and around the start site of the transcription; and (c) RNA polymerase escapes from the promoter after cycles of abortive initiation forming the stable elongation complex [8]. Promoter DNA undergoes drastic conformational changes during initiation of transcription. The necessary condition for open complex formation is that RNA polymerase must bind and bend the promoter DNA. This bending and subsequent torquing is responsible for melting the DNA and the formation of open complex [9,10].

The role of DNA curvature has been studied extensively in *E. coli* [11–17]. The conformation of DNA is a function of its nucleotide sequence [18,19]. The three-dimensional structure of DNA is the effect caused largely by interactions between neighboring base-pairs [20–29]. Generally, periodic repetitions of curved DNA in phase with the helical pitch cause the DNA to assume a macroscopically curved structure. Several theoretical models for estimating DNA curvature from di- or trinucleotides have been devised, and require various types of experimental data [23,25,28–31]. It is to be noted, however, that these models are being debated for their generality [32]. The importance of DNA conformation in transcription initiation is, however, clear and it would be interesting to study the DNA curvature distribution within the mycobacterial promoters especially in view of the large variation in their transcription mechanism. The objective of this paper is to use six different di- and trinucleotide-dependent models of curvature prediction for analysis of mycobacterial promoters.

2. System and methods

2.1. Data

The data for curvature analysis was taken from compilation of mycobacterial promoters (see Appendix A, the original reference details of each entry can be obtained from authors upon request). This data set contain 125 different mycobacterial promoters, out of which 80 promoters have their transcription start site (TSS) mapped while the other 45 are the putative promoters. In the listed compilation, we have considered the sequence stretches between −50 and +10 bp with respect to the TSS for the promoters whose TSS is mapped. For the putative promoters, we have documented the sequence stretch between the 15-bp upstream region of −35 box and 20 bp downstream of the −10 box. The promoter sequence length varies from 34 to 71 nucleotides based on the availability of the nucleotide sequence upstream of the −35 region and downstream of the −10 region. In a few cases, for the same gene two or more different sequence frames are considered based on the alternate consensus probability. Thus, total 135 mycobacterial promoter sequences are used in this study.

2.2. Curvature analysis

For the purpose of analyzing curvature distribution within mycobacterial promoter sequences, we have used the following dinucleotide models based on: (i) experimentally determined wedge angles [25]; (ii) energy minimized values of roll and tilt angles [31,33]; (iii) X-ray crystallography of DNA oligomers [30]; and (iv) Calladine–Dickerson rules [34,35]. The trinucleotide models used include: (i) the model based on tabulation of preferred sequence locations on nucleosomes [23,28]; and (ii) DNase I cutting frequencies [29].

1. CURVATURE [36]: To obtain curvature map of each mycobacterial promoter, a window size of a 21 bp nucleotide sequence is given as an input to the program and the curvature is obtained as an output. The results of this study are listed in Table 1 for each mycobacterial promoter. Various sub-groups of mycobacterial promoters are analyzed for nature of curvature profile and results are listed in Table 2.
2. De Santis et al. [33]: The curvature vector **C** (n,v) representing, in the complex plane (in modulus and phase), the directional change of the double helix axis between sequence number n and $n+v$ is calculated for each mycobacterial promoter sequence in the compilation. For this calculation, roll and tilt angle values (in

Table 1
Nature of curvature profile for mycobacterial promoters using dinucleotide models based on (i) experimentally determined wedge angles, and (ii) energy minimized values of roll and tilt angles

	Shipgelman et al. [36][a]	Santis et al. [33][b]
Low	• *M. tuberculosis*: T26, T119, T129, T130, gyrA, cpn60, KatG P_A (sp[c] = 19), KatG P_A (sp = 15), KatG P_C (sp = 22), KatG P_C (sp = 14), purC • *M. leprae* gro E1 • *M. smegmatis*: alrA, S5, S119, gyrB, ask, rrnA P3 • *M. paratuberculosis*: pAJB303, pAJB304, P_{AN}, pAJB73 • *M. fortuitum* repA	• *M. tuberculosis*: T3, T26, T119, recA rrnA P1, KatG P_A (sp[c] = 19), KatG P_A (sp = 15), KatG P_C (sp = 22), KatG P_C (sp = 14), purC • *M. bovis* BCG hsp60 P2 • *M. leprae*: gro E1, 36K, SOD • *M. smegmatis*: alrA, S33, S69, S119, gyr B, ask, rrnA P2 • *M. paratuberculosis*: pAJB303, pAJB300, pAJB73, pAJB301 • *Mycobacteriophage I3* ORF1 • *M. avium* pLR7
Medium	• *M. tuberculosis*: T3, T6, T80, T125, recA, rrnA P1, gyrB P3, 85A (sp[•] = 17), 85A (sp = 22), gyrB P2, glnA (sp = 18), glnA (sp = 10), KatG P_B (sp = 20), KatG P_B (sp = 22), purL, groE (sp = 19), groE (sp = 11), ahpC, 32 Kda, 10 Kda (sp = 17), 10 KDa (sp = 15), 10 Kda (sp = 8), 65 Kda, mpt64, metA, 38 Kda, ppgK • *M. bovis BCG*: hsp60 P2, ahpC, 23K, mpb 64,64 K • *M. leprae*: 16S rRNA, 28 Kda, 36 K, SOD • *M. smegmatis*: S4, S18, S21, S33, S65, S69, recA, acetamidase, rrnA P1, rrnA P2, rrnA PCL1, rpsL (sp = 18), rpsL (sp = 17), ahpC • *M. paratuberculosis*: pAJB86, pAJB125, pAJB300, pAJB305, pAJB301 • *M. fortuitum*: rrnA PCL1, rrnA P1, rrnA P2b, rrnA P3 • *M. phlei*: rrnA PCL1, rrnA P1, rrnA P3 • *Mycobacteriophage I3*: pKGR25, pKGR9, pKGR38, ORF1, ORF2, pKGR1 • *Mycobacteriophage L5*: 71 P2, 71 P_{left} • *M. avium*: avi-3, pLR7 • *M. neoaurum*: rrnA PCL1, rrnA P1, rrnA P3 • *M. absessus* rrnA P1 • *M. chelonae* rrnA P1	• *M. tuberculosis*: T80, T101, T125, T129, T130, gyrA, cpn60, gyrB P3, 85A (sp[•] = 17), 85A (sp = 22), gyrB P2, rrnA PCL1, 16S rRNA, glnA (sp = 18), glnA (sp = 10), KatG P_B (sp = 20), KatG P_B (sp = 22), purL, groE (sp = 19), groE (sp = 11), ahpC, 32 Kda, 10 Kda (sp = 17), 10 KDa (sp = 15), 10 Kda (sp = 8), 65 Kda, mpt64, rpsL, 38 Kda, ppgK • *M. bovis BCG*: rRNA, ahpC, 23K, mpb 64, 18 K, rpsL, mpb70, alpha • *M. leprae*: 16S rRNA, 18 Kda (sp = 17), 18 kda (sp = 18), 28 Kda, rpsL • *M. smegmatis*: S4, S5, S16, S18, S19, S21, S65, recA, acetamidase, rrnA P1, RrnA P3, rpsL (sp = 18), rpsL (sp = 17), ahpC • *M. paratuberculosis*: pAJB86, pAJB125, PAJB305, pAJB304, P_{AN} • *M. fortuitum*: repA, rrnA PCL1, rrnA P1, RrnA P2a, rrnA P2b, rrnA P3 • *M. phlei*: rrnA PCL1, rrnA P1, rrnA P3 • *Mycobacteriophage I3*: pKGR25, PKGR9, pKGR38, pKGR1 • *Mycobacteriophage L5*: 71 P2, 71 P_{left}, 71 P1 • *M. avium* avi-3 • *M. neoaurum*: rrnA PCL1, rrnA P1, RrnA P3, rrnA P2 • *M. absessus*: rrnA P1, rrnA PCL1 • *M. chelonae*: rrnA P1, rrnA PCL1
High	• *M. tuberculosis*: T101, T150, rrnA PCL1, 16S rRNA, rpsL • *M. bovis BCG*: rRNA, 18 K, rpsL, mpb70, alpha • *M. leprae*: 18 Kda (sp = 17), 18 Kda (sp = 18), rpsL • *M. smegmatis*: S6, S14, S16, S19, S30, S35, rrn B • *M. fortuitum* rrnA P2a • *Mycobacteriophage L5* 71 P1 • *M. neoaurum* rrnA P2 • *M. abscessus*: rrnA P4, rrnA PCL1, rrnA P2, rrnA P3	• *M. tuberculosis*: T6, metA • *M. bovis* BCG 64K • *M. leprae* 65 KD • *M. smegmatis*: S6, S12, S14, S30, S35, rrn B, rrnA PCL1 • *M. fortuitum* rrnA P2a • *M. phlei* rrnA P2 • *Mycobacteriophage I3* ORF2 • *Mycobacteriophage L5* 71 P1

Table 1 *(Continued)*

	Shipgelman et al. [36][a]	Santis et al. [33][b]
	•*M. chelonae*: rrnA P2, rrnA PCL1, rrnA P3, rrnA P4	
Very high	•*M. tuberculosis* gyrB P1	•*M. tuberculosis*: T150, gyrB P1
	•*M. leprae* 65 KD	•*M. smegmatis* S12
	•*M. smegmatis* S12	•*M. abscessus*: rrnA P4, rrnA P2, rrnA P3
	•*M. phlei* rrnA P2	•*M. chelonae*: rrnA P2, rrnA P3, rrnA P4

[a] Curvature maxima lying in the range [0.0–0.2], [0.2,0.4], [0.4,0.6]; and [0.6 and above] DNA curvature units are referred to as low, medium, high and very high curvature maps, respectively.

[b] Curvature maxima lying in the range [0–5], [5–10], [10–15]; and [15 and above] units are referred to as low, medium, high and very high curvature profiles, respectively.

[c] sp denotes spacer length in bp.

degrees) for the 16 different dinucleotide steps in DNA are taken from Anselmi et al. [31]. In our analysis, we have used integration step value as 31 (approx. three turns of B-DNA) in order to minimize the signal-to-noise ratio. The results of this study are also listed in Table 1 for each mycobacterial promoter. Various sub-groups of mycobacterial promoters are analyzed for the nature of curvature profile and these results are presented in Table 2. Curvature dispersion σ^2 quantifies the central dispersion of the local helical axes with respect to the average direction of the double helix. The σ^2 plot of cyclically permuted DNA sequence allows an easy alternative to experimental permutation assay for DNA tracts up to 700 bp long. Hence, σ^2 plots

of cyclically permuted mycobacterial promoters are prepared to see exact position of molecular bend locus. For simplicity of analysis mycobacterial promoter region is divided into the following five sub-regions: (i) region above -35 box; (ii) -35 region; (iii) spacer region; (iv) -10 region; and (v) region below -10 box. The position of molecular bend locus, for each mycobacterial promoter, with respect to the sub-regions specified above, is mentioned in Table 3.

3. Calladine–Dickerson Rule [34,35]: Calladine proposed four rules to understand the sequence-dependent departures from classical B-DNA due to simple steric hindrance of nearest neighbor purines on opposite strands. He suggested that

Table 2

Percentage of low, medium and high curvature profiles for various sub-groups of mycobacterial promoters using: (i) experimentally determined wedge angles [36]; and (ii) energy minimized values of roll and tilt angles [33]

Mycobacterial promoters	Low		Medium		High	
	Shipgelman et al. [36]	De Santis et al. [33]	Shipgelman et al. [36]	De Santis et al. [33]	Shipgelman et al. [36]	De Santis et al. [33]
Class I: *E. coli* σ^{70} type (sample size=69)	15	19	60	67	25	14
Class II: non-*E. coli* σ^{70} type (sample size=36)	22	27	56	54	22	19
Class II: extended -10 type (sample size=24)	17	4	25	58	58	38
Having optimum (17±1 bp) spacer length (sample size=79)	9	11	61	72	30	17
With high (≥50%) AT content (sample size=26)	12	15	54	58	35	27
Having A_nT_m $(n+m≥3)$ tract repeated in phase with each other and present at the upstream of -35 box (sample size=12)	17	17	25	33	58	50
M. tuberculosis (sample size=44)	25	23	61	68	14	9
M. smegmatis (sample size=28)	21	25	50	50	29	25
Entire compilation (sample size=135)	17	20	57	64	26	16

Table 3
Location of molecular bend locus with reference to the following sub-regions in the mycobacterial promoter (such as MT, *M. tuberculosis*; MB, *M. Bovis BCG*; ML, *M. leprae*; MS, *M. smegmatis*; MP, *M. paratuberculosis*; MF, *M. fortuitum*; MH, *M. phlei*; MI, *Mycobacteriophage I3*; MY, *Mycobacteriophage L5*; MV, *M. avium*; MN, *M. neoaurum*; MA, *M. abscessus*; MC, *M. chelonae*) sequence: (i) region above −35 box; (ii) −35 region; (iii) spacer region; (iv) −10 region; and (v) region below −10 box

Region above −35 box	−35 region	Spacer region	−10 region	Region below −10 box
Promoters whose transcription start site is determined				
MT T180	MT T119	MT T130	MT T101	MT T3
MT recA	MT T125	MT cpn60	MS S14	MT T6
MT 85A (sp^a=17)	MT T129	MT gyrB P1	MS rpsL (sp=17)	MT T26
MT KatG P$_C$ (sp=22)	MT 85A (sp=22)	MT gyrB P2	MP pAJB86	MT T150
MT purC	MT purL	MT katG P$_A$ (sp=19)	MY 71P2	MT rrnA P1
ML 16S rRNA	MS S4	MT katG P$_A$ (sp=15)	–	MT gyrA
MS S69	MS S5	MT katG P$_B$ (sp=20)	–	MT gyrB P3
MS gyrB	MS S19	MT katG P$_B$ (sp=22)	–	MT rrnA PCL1
MS ask	MS S21	MB hsp60 P2	–	MT 16S rRNA
MS rrnA P1	MS S119	MS S6	–	MT glnA (sp=18)
MS rrnA P2	MS rrnB	MS S12	–	MT glnA (sp=10)
MP pAJB300	MA rrnA P4	MS S16	–	MT KatG P$_C$ (sp=14)
MF rrnA PCL1	MC rrnA P2	MS S18	–	ML 18 kDa (sp=17)
MH rrnAPCL1	MC rrnA P3	MS S30	–	ML 18 kDa (sp=18)
–	–	MS S33	–	MS alrA
–	–	MS S35	–	MS S65
–	–	MP pAJB303	–	MS recA
–	–	MP P$_{AN}$	–	MS acetamidase
–	–	MF repA	–	MS rrnA P3
–	–	MY 71P1	–	MS rrnA PCL1
–	–	MA rrnA P1	–	MS rpsL (sp=18)
–	–	MA rrnAPCL1	–	MP pAJB125
–	–	MA rrnA P2	–	MP pAJB305
–	–	MA rrnA P3	–	MP pAJB304
–	–	MC rrnA P1	–	MP pAJB73
–	–	MC rrnA PCL1	–	MY 71P$_{left}$
–	–	–	–	MN rrnAPCL1
–	–	–	–	MC rrnA P4
16%	16%	30%	6%	32%
Putative promoters				
MT 32 kDa	MT ahpC	MT 10 kDa	ML SOD	MT groE
ML 28 kDa	MT metA	MT 38 kDa	MI pKGR25	MT groE
MF rrnA P1	MT rpsL	MT ppgK	MN rrnA P2	MT 10kDa
MN rrnA P1	MB ahpC	MB alpha	–	MT 10kDa
–	MB rpsL	MI pKGR38	–	MT 65kDa
–	ML 65 kDa	MI ORF2	–	MT mpt64
–	ML 36K	MV pLR7	–	MB rRNA
–	ML rpsL	–	–	MB 23K
–	MS ahpC	–	–	MB mpb64
–	MH rrnA P2	–	–	MB 18K
–	MI pKGR9	–	–	MB 64K
–	–	–	–	MB mpb70
–	–	–	–	ML groE1
–	–	–	–	MP pAJB301
–	–	–	–	MF rrnA P2a
–	–	–	–	MF rrnA P2b
–	–	–	–	MF rrnA P3
–	–	–	–	MH rrnA P1

Table 3 (*Continued*)

Region above −35 box	−35 region	Spacer region	−10 region	Region below −10 box
–	–	–	–	MH rrnA P3
–	–	–	–	MI ORF1
–	–	–	–	MI pKGR1
–	–	–	–	MV Avi−3
–	–	–	–	MN rrnA P3
8%	23%	15%	6%	48%

[a] sp denotes spacer length in bp.

the DNA chains may overcome these steric clashes in four possible ways: (i) the helix twist angle may be reduced; (ii) the base pairs can rotate along their long axes; (iii) the DNA backbone can shift sideways towards the pyrimidines; and (iv) the propeller twist can be suppressed. Dickerson quantified this by constructing four sum functions (Σ_1–Σ_4), by means of which the base sequence can be used to calculate the expected local variation in helix twist (Σ_1), base plane roll (Σ_2), torsion angle difference at the two ends of the base pair (Σ_3), and flattening of propeller twist (Σ_4). DNA helical structure variation at the molecular bend locus is studied here for mycobacterial promoters using Calladine–Dickerson rules. For this analysis, we have taken 11-bp-long sequence stretch obtained by taking five nucleotides on either side of the molecular bend locus of each mycobacterial promoter. For brevity, only Σ_1 function plots for the promoters whose TSS is mapped are shown in Fig. 1.

4. Propeller twist [30]: it is known that different types of dinucleotide step have different levels of conformational flexibility, which is very closely related to the Propeller-twist. Propeller twist values are obtained from X-ray crystallography of DNA oligomers. Dinucleotides with a large propeller-twist have a tendency to be more rigid than dinucleotides with low propeller twist. Higher (less negative) values correspond to higher flexibility. Flexibility profile was plotted using the propeller twist values from X-ray crystallography of DNA oligomers for overlapping dinucleotides.

5. DNase I derived bendability parameters [29]: the productive binding of bovine pancreatic deoxyribonuclease I (DNase I) requires DNA
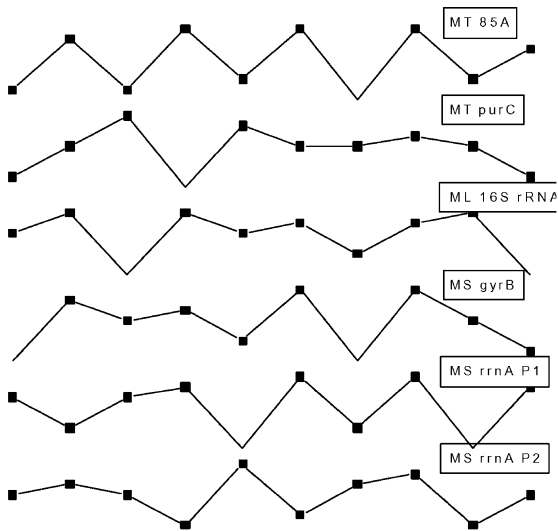
to be bent toward the major groove (positive roll). Base sequences that are flexible or inherently bent towards the major groove should therefore be more accessible to DNase I cleavage. DNase I cutting frequencies on naked DNA can be used as a quantitative measure of anisotropic bendability (major groove compressibility). Bendability profile was calculated using DNase I derived bendability parameters for overlapping trinucleotides of each mycobacterial promoter sequence.

6. Location preference [23]: from experimental investigations of the positioning of DNA in nucleosomes, it has been found that certain trinucleotides have strong preference for having minor grooves facing either towards or away from the nucleosome core. Based on the premise that flexible sequences can occupy any rotational position on nucleosomal DNA, while rigid sequences will be restricted in rotational location. We have calculated DNA flexibility profile using these location preference values for mycobacterial promoters at each position considering overlapping trinucleotides.
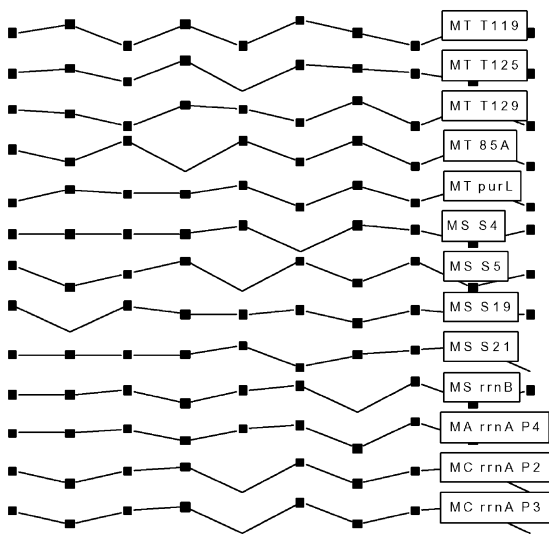
3. Results and discussion

The curvature distribution for various mycobacterial promoters as calculated using different models show similar trends. In order to aid the analysis the results obtained using: (i) experimentally determined wedge angles; and (ii) energy minimized values of roll and tilt angles, have been compared. The extent of curvature obtained using these models has been classified in terms of low, medium or high curvature and the results of the two models corroborate each other for most of the promoters barring a few promoter entries (e.g. *M. tubercu-*
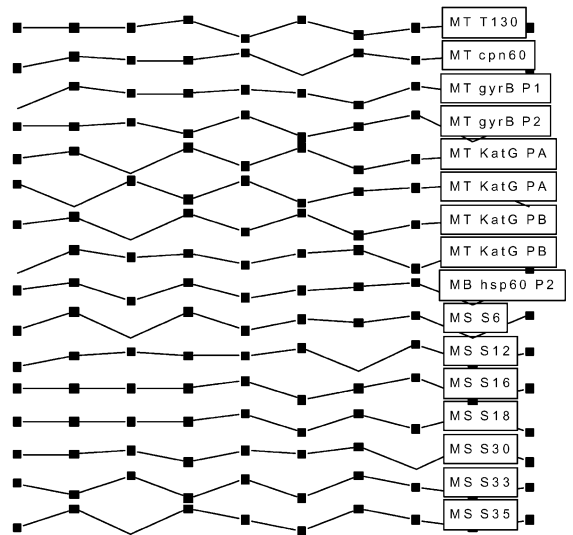
Region 1:

MT 85A
MT purC
ML 16S rRNA
MS gyrB
MS rrnA P1
MS rrnA P2

Region 3:

MT T130
MT cpn60
MT gyrB P1
MT gyrB P2
MT KatG PA
MT KatG PA
MT KatG PB
MT KatG PB
MB hsp60 P2
MS S6
MS S12
MS S16
MS S18
MS S30
MS S33
MS S35

Region 2:

MT T119
MT T125
MT T129
MT 85A
MT purL
MS S4
MS S5
MS S19
MS S21
MS rrnB
MA rrnA P4
MC rrnA P2
MC rrnA P3

Region 3:

MP pAJB303
MP PAN
MF repA
MY 71P1
MA rrnA P1
MA rrnA PCL1
MA rrnA P2
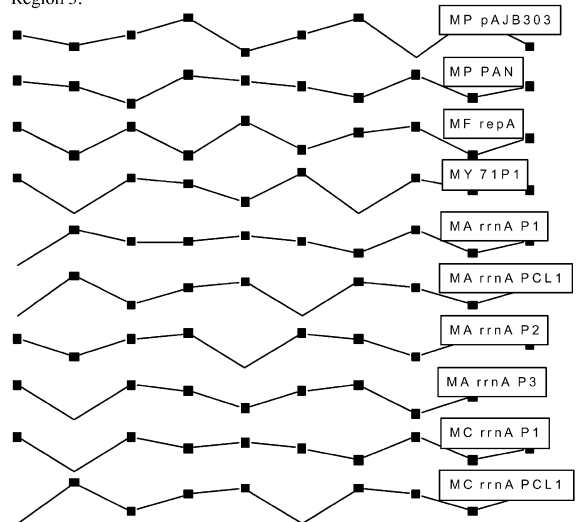MA rrnA P3
MC rrnA P1
MC rrnA PCL1

Fig. 1. Σ_1 function plots for the true mycobacterial promoters, sub-grouped depending upon the location of the molecular bend locus.

losis T3, T6, T101, T129, T130, recA, rrnA P1, gyrA, cpn60, rrnA PCL1, 16S rRNA, metA, rpsL, etc.) where the prediction of the two models differ.

In order to obtain a better insight for the results obtained by these two models, mycobacterial pro-

moters are sub-divided into various groups. These groups are as follows: (i) Class I, mycobacterial promoters resembling to *E. coli* σ^{70} type promoters; (ii) Class II, mycobacterial promoters which are different from *E. coli* σ^{70} type promoters, and
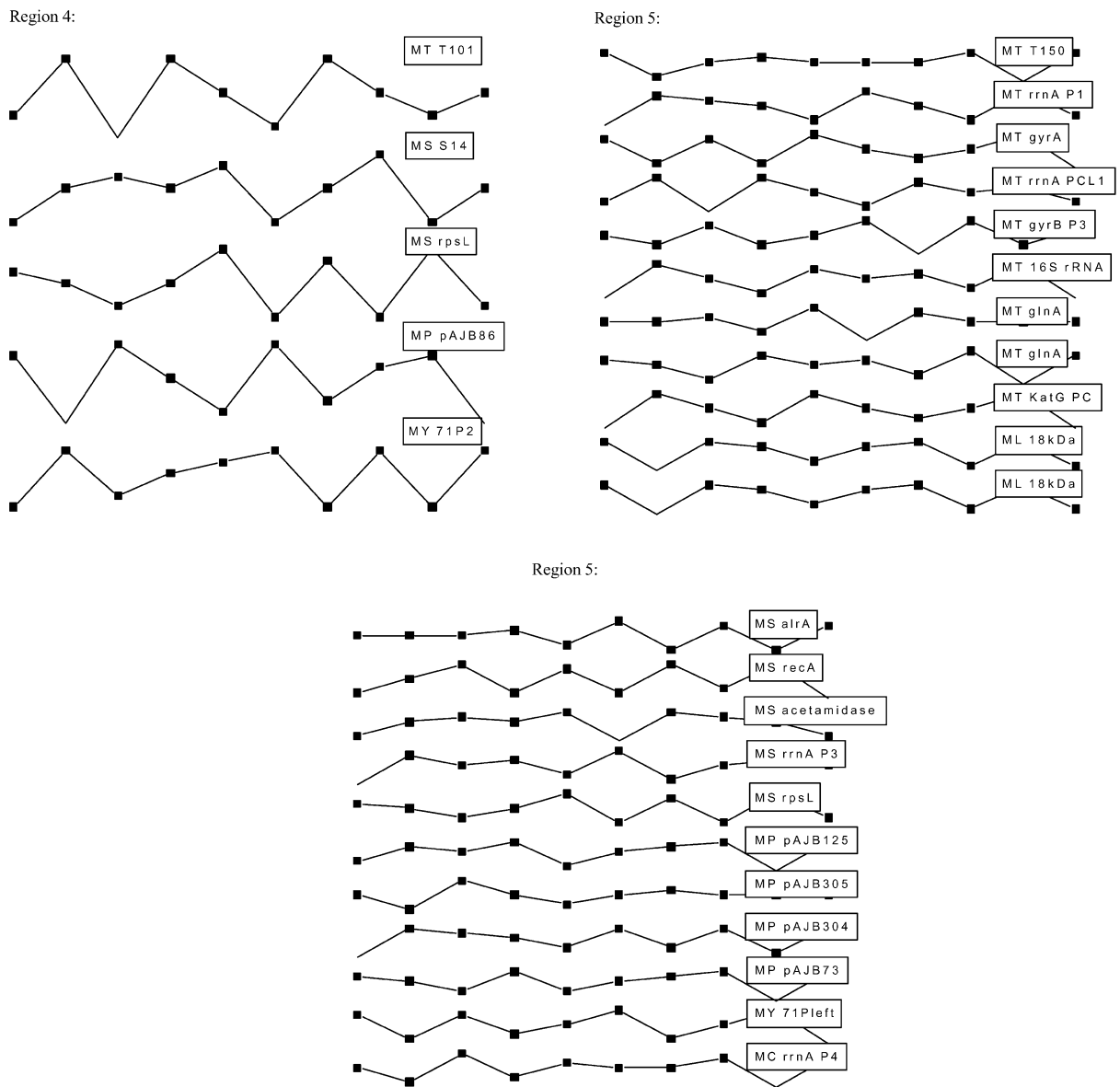
Region 4:

MT T101

MS S14

MS rpsL

MP pAJB86

MY 71P2

Region 5:

MT T150

MT rrnA P1

MT gyrA

MT rrnA PCL1

MT gyrB P3

MT 16S rRNA

MT glnA

MT glnA

MT KatG PC

ML 18kDa

ML 18kDa

Region 5:

MS alrA

MS recA

MS acetamidase

MS rrnA P3

MS rpsL

MP pAJB125

MP pAJB305

MP pAJB304

MP pAJB73

MY 71Pleft

MC rrnA P4

Fig. 1 (Continued.

constituting a class known as typical mycobacterial promoters; (iii) Class III, extended -10 promoters; (iv) mycobacterial promoters having optimum (17 ± 1 bp) spacer length; (v) mycobacterial promoters having high ($\geq 50\%$) AT content; (vi) mycobacterial promoters having $A_n T_m$ ($n+m \geq 3$) tract repeated in phase with each other and present at the upstream of the -35 box; (vii) *M. tuberculosis* promoters; (viii) *M. smegmatis* promoters; and (ix) entire mycobacterial promoter compilation. The curvature analysis of promoters classified in these groups are listed in Table 2. From Table 2, it can be seen that *E. coli* σ^{70} type mycobacterial promoters have 15% (19%), 60% (67%) and 25%

(14%) of low, medium, and high curvature profiles using curvature models of Shpigelman et al. [36] and De Santis et al. [33]. This distribution indicates that mycobacterial promoters resembling to *E. coli* σ^{70} type (Class I) have nearly 81% (85%) sequences having medium and high curvature profiles. Very few, i.e. 19% (15%) promoter sequences are having low curvature profiles. Considering percent distribution of curvature existing among *E. coli* σ^{70} type mycobacterial promoters, we can say that these promoters might be having good promoter activity. The analysis also indicates that the Non-*E. coli* σ^{70} type mycobacterial (Class II) promoters have 22% (27%), 56% (54%) and 22% (19%) of low, medium, and high curvature profiles (using both curvature models). This group of mycobacterial promoters has comparatively higher percent of low curvature profiles indicating that Non-*E. coli* σ^{70} type mycobacterial promoters might be poorly expressed compared to *E. coli* σ^{70} type mycobacterial promoters. The curvature models applied to the extended -10 (Class III) promoters show 17% (4%), 25% (58%) and 58% (38%) of low, medium, and high curvature profiles. The percent distribution of these promoters indicates that very few of these promoters have low curvature profiles. Extended -10 promoters might therefore have reasonably high promoter activity. *M. tuberculosis* T101, *M. smegmatis* S6, S16, and S19 promoters are extended -10 promoters, which are strongly curved. For such mycobacterial promoters sequence of the -35 region seems to be less important due to the presence of an extended TG motif in the immediate neighborhood of the -10 box along with the high curvature existing within it. Mycobacterial promoters lacking a consensus sequence at -35 and are curved are *M. tuberculosis* T150, *M. smegmatis* S12, S14, S30 and S35. Here, the curvature along with the -10 region might be useful for promoter activity although they do not possess a TG motif in the immediate neighborhood of the -10 box. The mycobacterial promoters having optimum (17 ± 1 bp) spacer length have 9% (11%) of sequences having low curvature profiles by both the models. The majority of sequences from this class has a curved structure. The favorable flexibility and/or curvature of DNA may compensate somewhat for a sub-optimal spacing of 16 or 18 bp between -35 and -10 regions during transcription initiation. The mycobacterial promoters with high percentage of AT have 12% (15%), 54% (58%) and 35% (27%) of sequences possess low, medium and high curvature profiles, respectively. The occurrence of curvature is obvious for the majority of sequences from this class due to their high percentage of AT content. Among mycobacterial promoters with $A_n T_m$ ($n+m \geq 3$) tract repeated in phase with each other and present at the upstream of the -35 box, 58% (50%) of sequences have high curvature trends. These promoters having upstream sequences, which can be expected to produce curvature in the DNA helical axis might be transcriptionally active promoters. *M. tuberculosis* promoters have 14% (9%), and *M. smegmatis* promoters have 29% (25%) of high curvature profiles. Such percentage distribution may be one of the causative factors for *M. smegmatis* to express better than *M. tuberculosis*. For the analysis performed in Table 2, it is important to realize that the percentage value of curvature predictions by both models sometimes differ significantly due to different conditions defined for low, medium, and high curvature profiles; and in a few cases predictions by the two models lie on the boundary conditions of low and medium, or medium and high curvature profiles. The sample size considered in this analysis is also small, and can cause large differences in the predictions by these two models. Results listed in Table 2 should only be used to see qualitative and semi-quantitative trends.

According to CURVATURE software, curvature maxima for *M. tuberculosis* gyrB P1, *M. bovis BCG* alpha, *M. fortuitum* rrnA P1, *Mycobacteriophage L5* 71P1, *M. neoaurum* rrnA PCL1, and rrnA P3 promoters lies above 0.3 DNA curvature units and it is present between the -35 and -10 regions. It will be interesting to study the transcription initiation mechanism in these promoters because in *E. coli* it is shown that the curvature between -35 and -10 regions seems to correlate significantly with promoter activity. In such cases the curved structure of promoter DNA enhances the binding of *E. coli* RNA polymerase to the promoter, when the curve is oriented correctly relative to the potential -10 and -35 regions,

and it also facilitates unwinding of the -10 region by thermal motion, as the DNA vibrates back and forth in solution between the twisted and curved forms [11].

σ^2 plots of cyclically permuted mycobacterial promoters should allow an alternative to the experimental permutation assay for determining molecular bend locus of a mycobacterial promoter sequence. The model has been successful in predicting the experimental results for other systems [33,38,39], while promoters analyzed here have not been subjected to any such experimental investigations and hence, the theoretical predictions could not be tested. In Table 3, we have evaluated the percent occurrence of position of molecular bend locus in the (i) region above -35 box, (ii) -35 region, (iii) spacer region, (iv) -10 region, and (v) region below the -10 box. For this analysis, we have separated the entire promoter compilation into two groups: (i) promoters whose TSS is mapped (true promoters); and (ii) putative promoters. According to the percentage distribution for true promoters, the molecular bend locus lies predominantly in the spacer region and the region below the -10 box. Sixteen, 16, 30, 6 and 32% of true mycobacterial promoter sequences show that their molecular bend locus lies in the region above the -35 box, -35 region, spacer region, -10 region, and the region below the -10 box, respectively. For putative promoters, 8, 23, 15, 6 and 48% of sequences show their molecular bend locus in a region above the -35 box, -35 region, spacer region, -10 region, and a region below the -10 box, respectively. Thus, for true as well as putative mycobacterial promoters, the spacer region and region below the -10 box seems to be of frequent occurrence for the location of molecular bend locus. Similar studies by Nair and Kulkarni [40] on *E. coli* promoter sequences showed that 60% of these promoters have their minima (molecular bend locus) lying in the spacer region. However, for mycobacterial promoters, the position of the molecular bend locus can occur with varying percent distribution at a region above the -35 box, -35 region, spacer region and a region below the -10 box. Thus, mycobacterial promoters show variation in the position of the molecular bend locus compared to *E. coli* promoters.

The Calladine and Dickerson rule $(\Sigma_1 - \Sigma_4)$ gives a way of revealing possible structural homology between regions of DNA, when the similarity is not obvious by direct comparison of sequence alone. The helical structure variation at the molecular bend locus for the true mycobacterial promoters is sub-grouped according to the position of molecular bend locus. Thus, Fig. 1 is subdivided into five plots. The helical structure variation obtained using the Σ_1 function at the molecular bend locus lying in the (i) region above the -35 box, (ii) -35 region, (iii) spacer region, (iv) -10 box, and (v) region below the -10 box shows that each sub-group has structural similarity within that particular sub-group. The other sum functions also uphold the structural similarities (results not shown). The analysis of the sequence at the minima reveals that there exists homology among these sequences irrespective of the exact position of minima. The regions that are localized for mycobacterial promoters show significant commonality in structure, which is evident from the Σ_1 function plot. There seems to exist some structural commonalties among the each sub-group of mycobacterial promoters. We can therefore group the promoters based on the common structural features and advocate the notion of 'consensus structure' suggesting their common biological significance. The variation from these consensus structures can account for varying strength of the promoters. Such an analysis might help us in designing experiments to define the exact location and function of a promoter.

Although the entire mycobacterial promoter compilation has been analyzed using other curvature models [23,29,30], the results obtained using only three models are presented.

Mycobacterial promoters that are strongly curved are *M. tuberculosis* T150, and gyrB P1; *M. Leprae* 65KD; *M. smegmatis* S6, S12, S14, S30, S35, and rrnB; *M. Phlei* rrnA P2; *M. abscessus* rrnA P4, rrnA P2, and rrnA P3; *M. chelonae* rrnA P2, rrnA P3, and rrnA P4. Fig. 2 shows the curvature map expressed in DNA curvature units of these promoters using CURVATURE software. The curvature maxima of these curvature maps
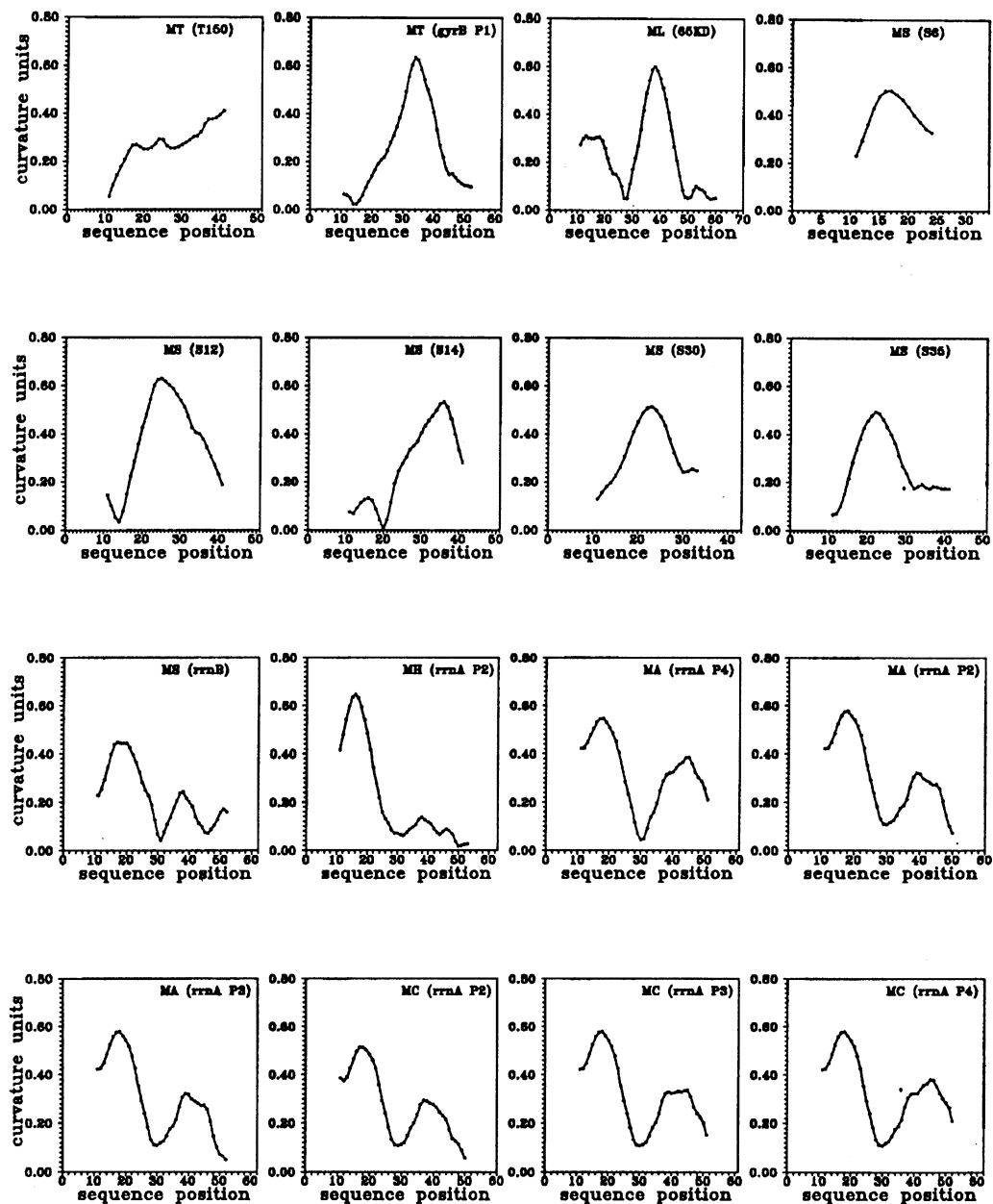
Fig. 2. Curvature map obtained using experimentally determined wedge angles for mycobacterial promoters. Curvature is expressed in DNA curvature units [37] where one curvature unit corresponds to the mean DNA curvature in the crystalline nucleosome (1/42.8 Å).

correspond to region having more curved structure. Fig. 3a,b presents the curvature analysis using energy-minimized values of roll and tilt angles.

The curvature vector is a complex function of the sequence with the modulus representing the deviation and the phase indicating the relative direc-
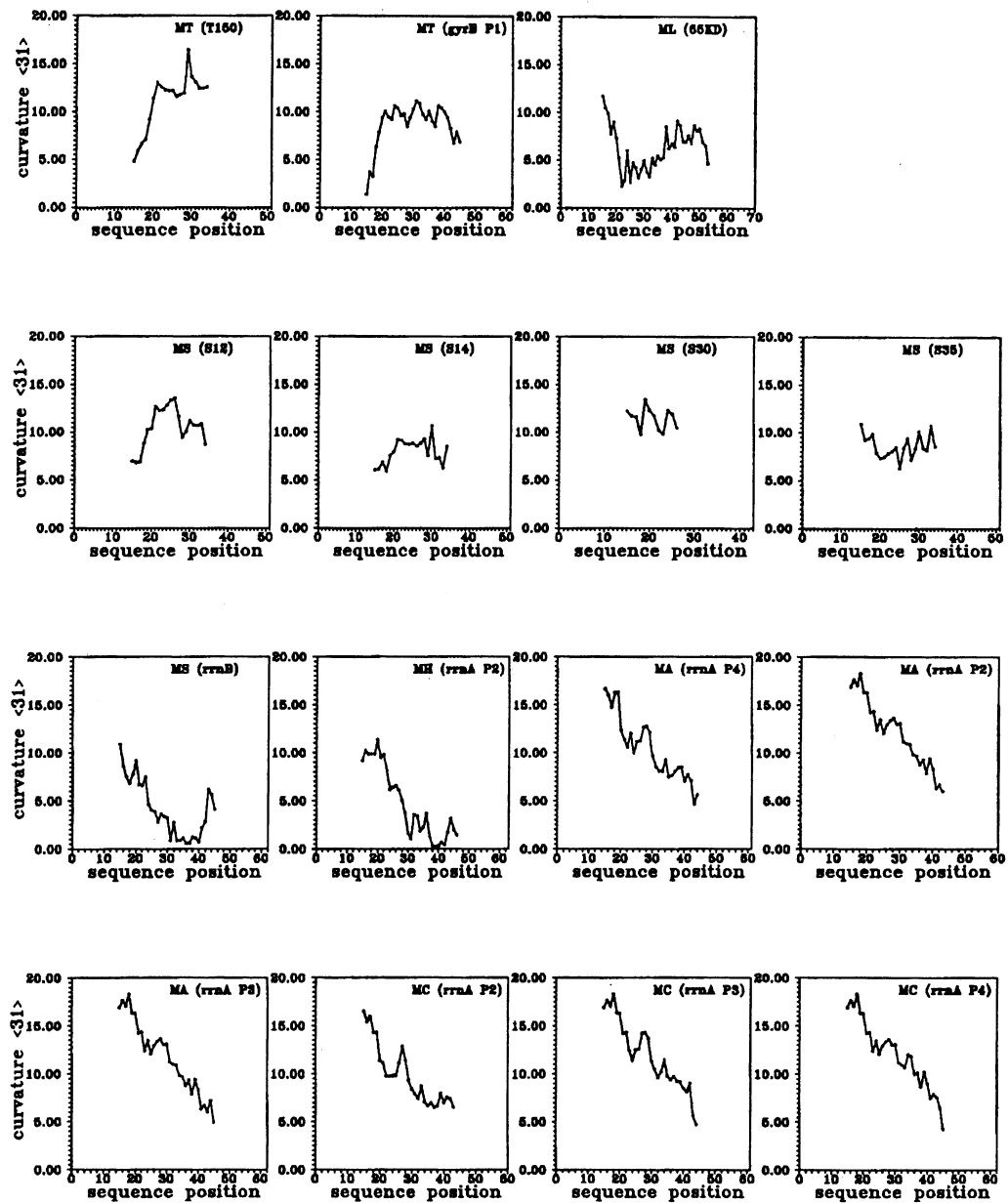
Fig. 3. (a) Curvature profiles obtained using energy minimized values of roll and tilt angles for mycobacterial promoters. The curvature is reported as $|C|$, the curvature modulus averaged over 31 bp (*M. smegmatis* S6 is excluded from this plot as grid size used for it is 21 bp). (b) Relative phase profiles of the mycobacterial promoters.

tion. The curvature diagrams for these mycobacterial promoters clearly show a DNA tract characterized by both a high curvature modulus (see Fig. 3a) and a constant phase (Fig. 3b). Fig.

4 shows flexibility profiles based on propeller twist values from X-ray crystallography of DNA oligomers. Dinucleotides with a large propeller-twist have a tendency to be more rigid than
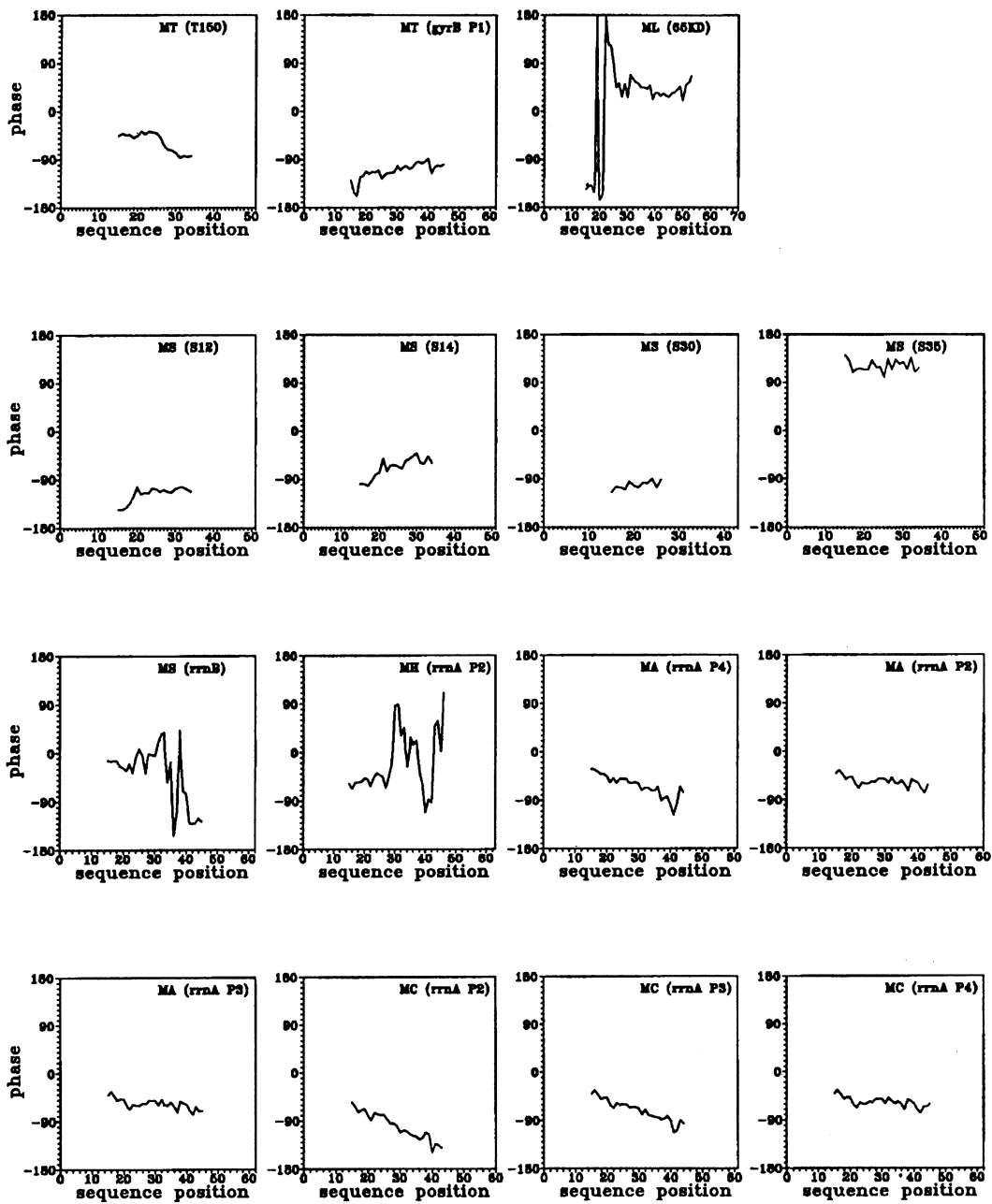
Fig. 3 (Continued).

dinucleotides with low propeller-twist. Thus, sequence positions corresponding to higher (less negative) values represent regions of higher flexibility for mycobacterial promoter. Fig. 5 presents flexibility profile calculated using a trinucleotide model based on preferred sequence location on nucleosomes. Sequence positions corresponding to lower values of location preference represent a
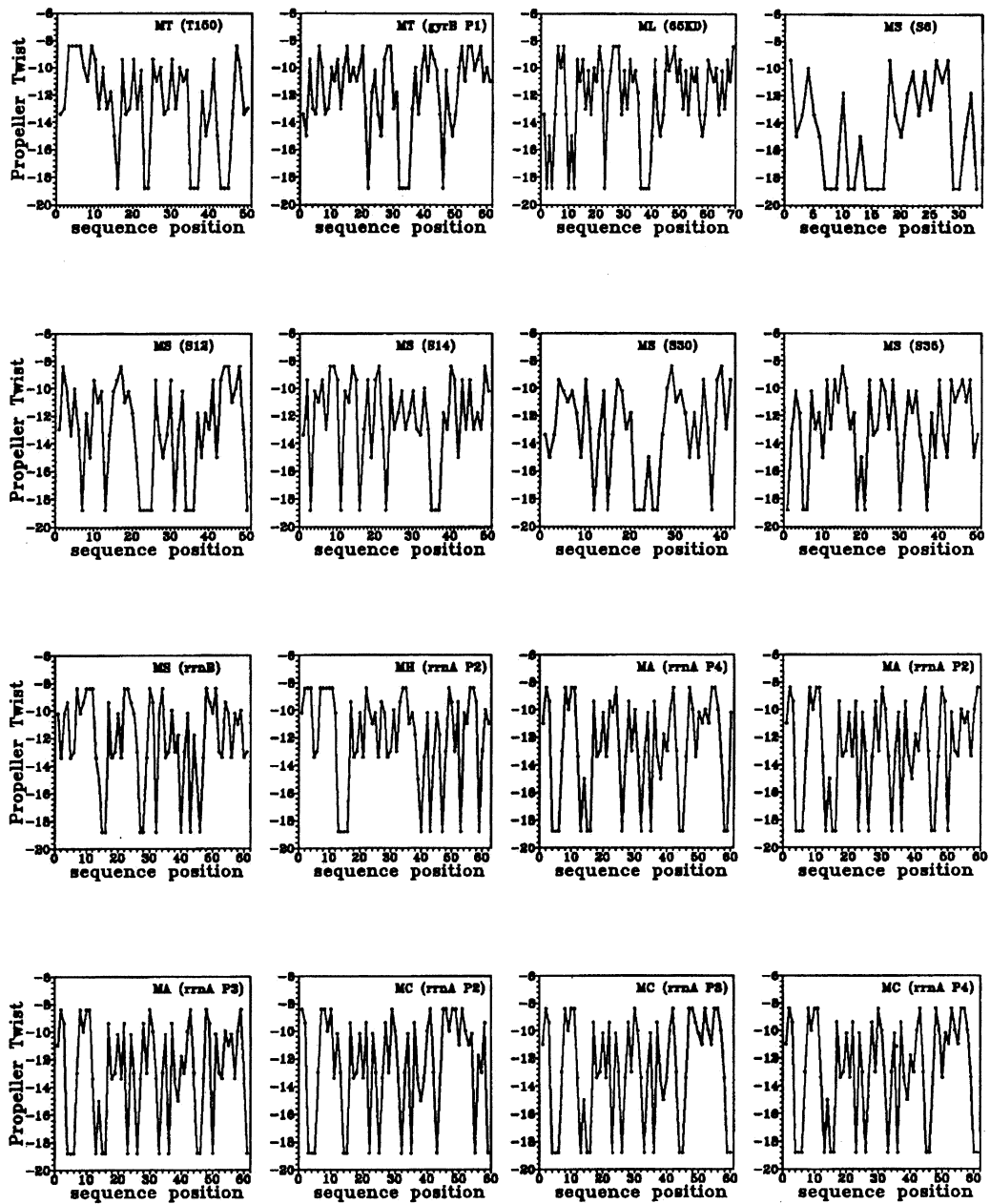
Fig. 4. Flexibility profile calculated using propeller-twist values obtained from X-ray crystallography of DNA oligomers for myco-bacterial promoters.

more flexible region of the mycobacterial promot-er, which have less preference for being positioned specifically. Fig. 6 shows a bendability profile in

mycobacterial promoters calculated using DNase I derived bendability parameters. A sequence posi-tion corresponding to higher bendability parame-
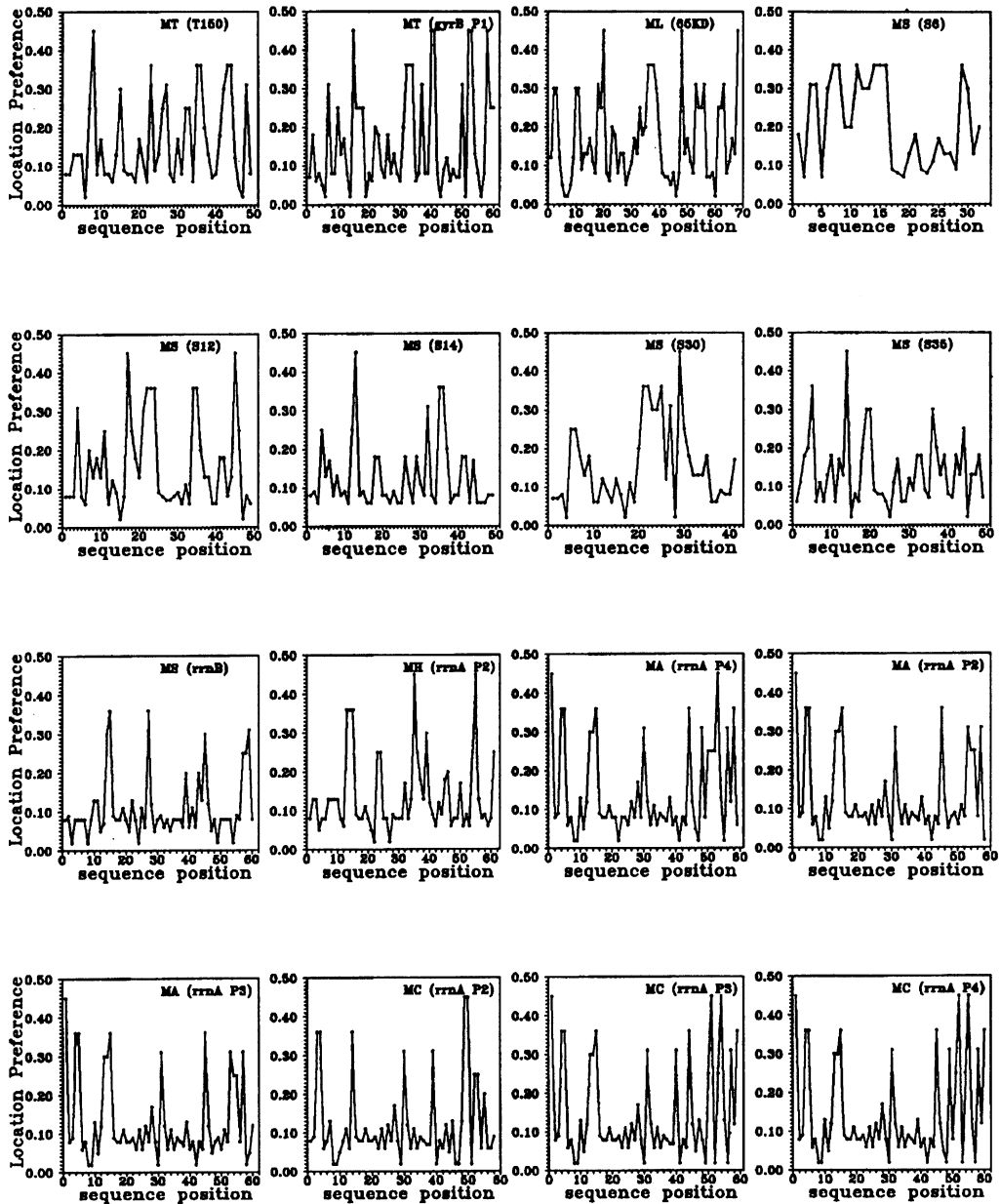
Fig. 5. Flexibility profile calculated using trinucleotide model based on preferred sequence location on nucleosomes for mycobacterial promoters.

ters represent higher propensity for major groove compressibility of the mycobacterial promoter. Essentially, all the models predict similar behavior for these promoters. Thus, a nucleotide sequence position corresponding to high (low) curvature shows a high (low) curvature trend with all the other models. Mycobacterial promoters of *M. abscessus* rrnA P4, rrnA P2, and rrnA P3; *M.*
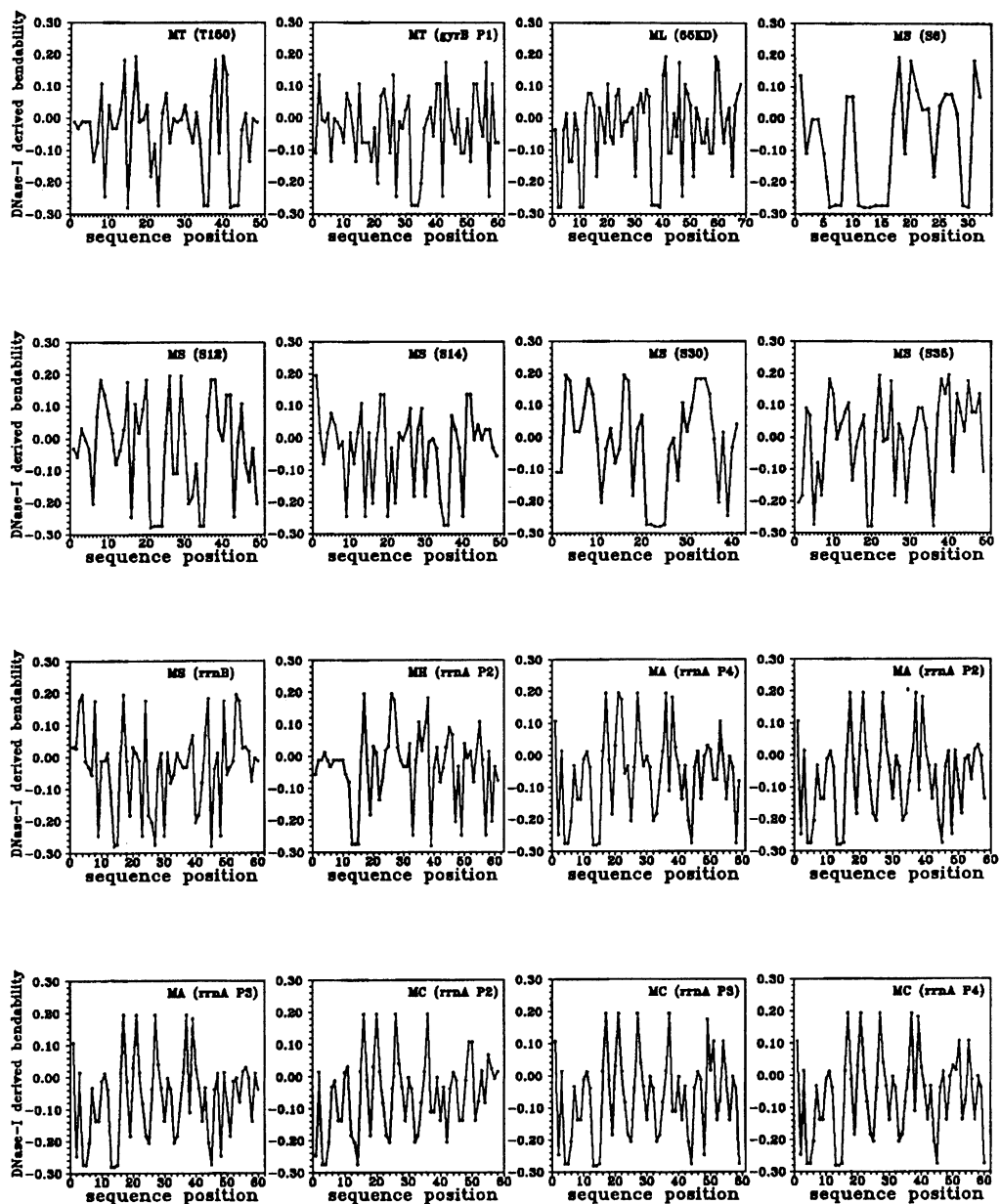
Fig. 6. Bendability profile calculated using DNase I derived bendability parameters for the mycobacterial promoters.

chelonae rrnA P2, rrnA P3, and rrnA P4 have similar curvature trends as their nucleotide sequence shows maximum homology with each other. The similar curvature trends suggest a common mechanism for transcription initiation.

Regions with high DNA curvature would be expected to exhibit anomalous mobility by the gel electrophoresis assay. It will be of interest to examine fragments containing these regions for the structural feature of DNA curvature, and the

corresponding functional feature of transcriptional activation. Plasmids containing stiff, flexible or curved DNA structure near the cleavage site of commonly used restriction enzymes can be helpful for studying the role of DNA structure in transcription mechanism of mycobacterial promoters.

Thus, analysis of DNA curvature distribution for mycobacterial promoters reveals the following important features. (i) The curvature distribution for various mycobacterial promoters calculated using different models show similar trends. (ii) Mycobacterial promoters resembling to *E. coli* σ^{70} type have nearly 81% (85%) sequences having medium and high curvature profiles. (iii) Non-*E. coli* σ^{70} type mycobacterial promoters have com

paratively higher percent of low curvature profiles. (iv) Very few of the extended -10 promoters have low curvature profiles. (v) Mycobacterial promoters having $A_n T_m$ $(n+m \geq 3)$ tract in the upstream region of the -35 box and repeated in phase with each other have high curvature profiles. (vi) *M. smegmatis* promoters have high curvature profiles compared to *M. tuberculosis* promoters.

Experimental studies based on curvature distribution and its role in transcription mechanism for particular mycobacterial promoter(s) or representative examples from various groups of mycobacterial promoters showing some distinct features, will throw light on our understanding of transcription mechanism of *Mycobacteria*.

Appendix A: Compilation of Mycobacterial promoters[1]

M. tuberculosis	
T3	ATCGACGGCCACGGCTGGTCTAGGACGAGGTACCCGG(**TAACAT**)GCTGGGC[G]
T6	CCGTCCAGTCTGGCAGGCCGGAAACATCGGTCAGCAGA(**TAGGCT**)TTACCA[G]
T26	CTGCGAGCATCATATGCCGCGTGCGTGGTGATGCGGCAG(**GATGTT**)GGACC[A]
T180	GATCACTCCGAGCATGCGCCCATTGTTGTGCATAGGG(**CAGGAT**)GCCCTG[G]
T101	AGCGATCGCAGCCGACGTGATACCTGACCGTTGT*TGA*(**TAGTGT**)CGGCGGC[A]
T119	CCCCGTGCTCGTAGTAGGCGTCCAGCCGACCCGCC*CGC*(**TACCAT**)GCACAAG[T]
T125	CCGAGGTAAGGACTGAGCATGGGCCCGATAAAGTGAC(**TATTAT**)GGATTTC[T]
T129	ACTCGCGGCAGATTACGCCGACGGTTCCTGGCG*TGG*(**TTCAAT**)ATTCGCCG[A]
T130	ACTCCAACAGGTCGATAACCTCCTGCGCCTGCTC̲GTC(**TATGCT**)GCGATCC[G]
T150	GACCCCCGCCACGTATTGACACTTTGCGACACGCTTT(**TATCAT**)TTTCCGA[C]
RecA	TTCGGAGCAGCCGAC(**TTGTCA**)GTGGCTGTC(**TCTAGT**)GTCACGGCC[A]ACCGACCGAT
RrnA P1	GAGAACCTGGTGAGT(**CTCGGT**)GCCGAGATCGAACGGG(**TATGCT**)GTTAGGC[G]ACGGTCACCT
GyrA	GATGGGCGAGGACGT(**CGACGC**)GCGGCGCAGCTTTATCA(**CCCGCA**)ACGCCAA[G]GATGTTCGGT
Cpn60	CCCCGGCGATCCCCG(**TGCTCA**)CCACGGGTGATTTCCGG(**GGCGGC**)ATGCGTT[A]GCGGACTAGC
GyrB P1	GATGTCCGACGCACG(**GCGCGG**)TTAGATGGGTAAAAACG(**AGGCCA**)GAAGATC[G]GCCCTGGCGC
GyrB P3	CAAGGGGCCTCGCCA(**TATTGC**)CGGTAGGGGTCCGCGCG(**ACACCT**)ACGGATA[A]CACGTCGATC
85A	GAAGTTGTGGTTGAC(**TACACG**)AGCACTGCCGGGCCCAG(**CGCCTG**)CAGTCTG[A]CCTAATTCAG
85A	CGCCCGAAGTTGTGG(**TTGACT**)ACACGAGCACTGCCGGGCCCAG(**CGCCTG**)CAGTCTG[A]CCTAATTCAG
GyrB P2	AGCGGTTGGCAACGA(**TGTGGT**)GCGATCGCTAAAGATCAC(**CGGGCC**)GGCACC[A]TCGTGGCGCA
RrnA PCL1	TGACCGAACCTGGTC(**TTGACT**)CCATTGCCGGATTTGTAT(**TAGACT**)GGCAGG[G]TTGCCCGAAA
16S rRNA	TGACCGAACCTGGTC(**TTGACT**)CCATTGCCGGATTTGTAT(**TAGACT**)GGCAGG[G]TTGCCCCGAA
GlnA	TCGGCATGCCACCGG(**TTACGA**)TCTTGCCGACCATGGCCC(**CACAAT**)AGGGCCGGGG[A]GACCCGGCGT
GlnA	CCACCGGTTACGATC(**TTGCCG**)ACCATGGCCC(**CACAAT**)AGGGCCGGGG[A]GACCCGGCGT
KatG P_A	GGTCATCTACTGGGG(**TCTATG**)TCCTGATTGTTCGATATCC(**GACACT**)TCGCGATC[A]CATCCGTGAT
KatG P_A	ATCTACTGGGGTCTA(**TGTCCT**)GATTGTTCGATATCC(**GACACT**)TCGCGATC[A]CATCCGTGAT
KatG P_B	GAGGCGGAGGTCATC(**TACTGG**)GGTCTATGTCCTGATTGTTC(**GATATC**)CGACAC[T]TCGCGATCAC
KatG P_B	ACGAGGCGGAGGTCA(**TCTACT**)GGGGTCTATGTCCTGATTGTTC(**GATATC**)CGACAC[T]TCGCGATCAC
KatG P_C	CCTGATTGTTCGATA(**TCCGAC**)ACTTCGCGATCACATCCGTGAT(**CACAGC**)CCGATAA[C]ACCAACTCCT
KatG P_C	TTCGATATCCGACAC(**TTCGCG**)ATCACATCCGTGAT(**CACAGC**)CCGATAA[C]ACCAACTCCT
PurL	CGGCTTGTCCGTTTC(**CACGCG**)GCCGCAGCGCGATGGGGCCTAGC(**TAGACT**)GCCTCC[G]TGATGTCTCC

[1] Consensus regions are shown in bold letters and transcription start sites are shown in a square bracket.

PurC	ATCTCATACCAGAGA(**TACCAG**)CACAGGGCGCCGTCGTGCGGCGGA(**TAGGCT**)GGCGTG[A]TGCGCCCCGC
GroE	CAGGAAGCAAGGGGGCG(**CCCTTG**)AGTGCTAGCACTCTCA*TGT*(**ATAGAG**)TGCTAGATGGCAATCGGCTA
GroE	CAGGAAGCAAGGGGG(**CGCCCTTG**)AGTGCTAGCAC(**TCTCATGTATAGAG**)TGCTAGATGGCAATCGGCTA
AhpC	TGTGATATATCACCT(**TTGCCT**)GACAGCGACTTCACGG(**TACGAT**)GGAATGTCGTAACCAAATGC
32 kDa	ACATGCATGGATGCG(**TTGAGA**)TGAGGATGAGGGAAGC(**AAGAAT**)GCAGCTTGTTGACAGGGTTC
10 kDa	AAGCAAGGGGCGCCC(**TTGAGT**)GTCAGCACTCTCATGTA(**TAGAGT**)GCTAGATGGCAATCGGCTAA
10 kDa	AAGCAAGGGGCGCCC(**TTGAGT**)GTCAGCACTCTCATG(**TATAGA**)GTGCTAGATGGCAATCGGCT
10 kDa	AAGCAAGGGGCGCCC(**TTGAGT**)GTCAGCAC(**TCTCAT**)GTATAGAGTGCTAGATGGCA
65 kDa	GCGTAAGTAGCGGGG(**TTGCCG**)TCACCCGGTGACCCCCG(**TTTCAT**)CCCCGATCCGGAGGAATCAC
Mpt64	GAGTCTGGTCAGGCA(**TCGTCG**)TCAGCAGCGCGATGCCC(**TATGTT**)TGTCGTCGACTCAGATATCG
MetA	TCCGGCCCCCGCGAT(**TTGGCG**)AGCTTCGTGCGTGTTCGG(**TAGCCT**)GGCATTTACCGACGCGGGGT
RpsL	GCCGCAACGCCCGCT(**TTGACC**)TGCCAGACTGGCGGCGGG(**TATTGT**)GGTTGCTCGTGCCTGGCGGC
38 kDa	CGTCGCCGGACTGTCGGGGGACGTCAAGGACGCCAAGCGCG(**GAAATT**)GAAGAGCACAGAAAGGTATG
Ppgk	CGGGCCGCAGTTTAAGGTGAGGGTCATCCACGTCTCGCCGAGGAGATTCGATGACCAGCAC

M. bovis BCG

Hsp60 P2	CGGTGCGGGGCTTCTTGCACTCGGCATAGGCGAG*TGC*(**TAAGAA**)TAACGTT[G]
RRNA	TGACCGAACCTGGTC(**TTGACT**)CCATTGCCGGATTTG(**TATTAG**)ACTGGCAGGGTTGCCCCGAA
AhpC	TGTGATATATCACCT(**TTGCCT**)GACAGCGACTTCACGG(**TACGAT**)GGAATGTCGCAACCAAATGC
23K	GAGTCTGGTCAGGCA(**TCGTCG**)TCAGCAGCGCGATGCCC(**TATGTT**)TGTCGTCGACTCAGATATCG
mpb64	GAGTCTGGTCAGGCA(**TCGTCG**)TCAGCAGCGCGATGCCC(**TATGTT**)TGTCGTCGACTCAGATATCG
18K	TGGCGTCCGAAACAC(**TTGAGG**)TGCGGCCCAGCAAGGGGC(**TACAGG**)TTTTTTCCTTCACCTACGGA
64K	GCGTAAGTAGCGGGG(**TTGCCG**)TCACCCGGTGACCCCCGG(**TTTCAT**)CCCCGATCCGGAGGAATCAC
rpsL	GCCGCAACGCCCGCT(**TTGACC**)TGCCAGACTGGCGGCGGG(**TATTGT**)GGTTGCTCGTGCCTGGCGGC
mpb70	TGGCGTCCGAAACAC(**TTGAGG**)TGCGGCCCAGCAAGGGGC(**TACAGG**)TTTTTTCCTTCACCTACGGA
alpha	CGACTTTCGCCCGAA(**TCGACA**)TTTGGCCTCCACACACGG(**TATGTT**)CTGGCCCGAGCACACGACGA

M. leprae

16S rRNA	TAGTCAACCCGGGAC(**TTGACT**)CCTCTGCTGGATC*TGT*(**ATTAAT**)CTGGCTG[G]GTTGCCGAAG
18 Kda	CTTGTCTATCACAAC(**TTGCAT**)CAATATATCGACCAGTG(**CTATAT**)CAAATCTA[T]GTAGTCAGGA
18 Kda	CTTGTCTATCACAAC(**TTGCAT**)CAATATATCGACCAG*TGC*(**TATATC**)AAATCTA[T]GTAGTCAGGA
28-kDa	TCAATATAACCACTC(**TGGTCA**)CACTAACCATACTCG(**TACCAT**)CAACCGTGTGGGGCTAATCC
groE1	AGCAGCGGGCCGGCC(**TTGAGT**)GCTAGCACTCGCGTGTA(**TAGAGT**)GCTAGATGGCAGTCGGCCAG
65 kd	GAATTCCGGAA(**TTGCAC**)TCGCCTTAGGGGAG*TGC*(**TAAAAA**)TGATCCTGGCACTCGCGATC
36k	GTTGGG(**TTTCCT**)CTCGGAGGGCGCACCGC(**TACGTT**)AGCGGGATG
SOD	GG(**TGGGCG**)CGATCATGGCGCAGCGTT(**GATTAT**)GCTAGTCG
rpsL	CGCCGTTGGGTCGCT(**TTGACC**)TGCCCGAGCAGGGACGGG(**TATTGT**)GTTTCTCGTTCCTGACGGCT

M. smegmatis

alrA	GTCTGCGGCCTCTGG(**GACAAT**)GGGCGCC[G]GAGATTATGA
S4	AAGCCGAATCGAGACCTTTTGGGTTCGTACACACTTGCTT(**TATAAG**)CCTC[G]
S5	AACAAGATTCCGTTAATCGTGTCTGGTGGAGCTGG*TGG*(**TAAGCT**)TGATCC[G]
S6	CATCGATTTTAAATTTT*TGA*(**TAGAG**T)GCAAATA[A]
S12	ACCTCGTTATGCTTCTGGCTATTTTTGATCAACTTT(**TATACA**)TGGGCGGT[T]
S14	TCAAGCACCCAAGCCAACATGGTTGTAGTAGTCGTTT(**TACCAT**)GTGTACC[T]
S16	TCCACGCGAACCGCTTCGGCGTGCCCCGTTTTCCC*TGT*(**TATAAT**)ATCGGC[G]
S18	GATCATTGTCTTCTGTTGTCTTTCGTA(**TAAAGT**)TGTTACT[G]
S19	TTTGATGTAGCCAAAGGCTCTCACCACCTGAGCCA*TGA*(**TAGTAT**)CCATCC[C]
S21	ACATGGCATTTTTCATTTAAAACAGGACTCAGGTGG(**TATGGT**)TGACATCG[A]
S30	GATCAGCTATGTTCTTCAGTAAAATTTCGGC(**TATATG**)TTGGT[G]
S33	GATCCGCTCTTCTTATGATGCCAGTTATGGTATC(**TATGGT**)TATC[G]
S35	AACTAAAGTATGTGCCGTAATTGACAGTGTTCTAGAT(**TATGAT**)GCTGCAT[C]
S65	GGCACAGCTCGAAGTTCTACTACATGGCTTGCTGAA(**TCCAGT**)CACATTAC[T]
S69	ATCACGATGTCTTCATGCTTGGCTTCAATGCTCCGGTC(**TACAAT**)CAGTTC[A]
S119	GATCAAGAAGCCAATGATT*TGT*(**TAAACG**)CAATTAAT[G]
gyrB	CAGAATCGGTGCTGT(**CGCTAT**)CTCGCGG(**TAGACT**)GGACGAC[G]GATCTCAGGC
recA	AGAGTTCGACCGGAC(**TTGTCG**)GTGGTC*TGC*(**TCTAAC**)GTCACGGCC[A]ACCGATCGGA
ask	GT(**TTGCCC**)GCCGCGGCGCCC(**CACGAT**)GAACCGC[A]CGGGCTGACG
acetamidase	GGCCGGCGTTCACCC(**TTGACT**)TTTATTTTCATCTGGA(**TATATT**)TCGGGT[G]AATGGAAAGG

rrnB	CTCTGACCTGGGGAT(**TTGACT**)CCCAGTTTCCAAGGACG(**TAACTT**)ATTCCAG[G]TCAGAGCGAC
rrnA P1	GAAAACCTGGTCAGC(**CTCGGA**)GCCGAGATCGAGAGAG(**TAAGCT**)CGTAG[G]AAGCAAGACC
rrnA P2	CTCTGACCAGGCGAT(**TTGCAA**)TCGCGACGAACCTCGTAT(**TATCTT**)TATGAA[G]TCGCCGCGGA
rrnA P3	CCGGGCCAGAGCGAC(**TTGACA**)AGCCAGCCGAGATCGTAC(**TAAGCT**)GGCGAG[G]TTGCCTCAGA
rrnA PCL1	CCGGTCCAGAGCGAC(**TTGACA**)AGCCAGACAAAGCAGTAT(**TAAGCT**)GGCAGG[G]TTGCCCCAAA
rpsL	CCGCCGTGCACGAGT(**TTGTTT**)CGTCGCGGTCGCCCC*TGG*(**TATTGT**)GGTGGATC[G]TGCCTGGCCC
rpsL	CGTGCACGAGTTTGT(**TTCGTC**)GCGGTCGCCCCTGGTAT(**TGTGGT**)GGATC[G]TGCCTGGCCCGAAA
ahpC	TGTGATATATCACCT(**TTGCCT**)GACAGCGACTTCACGG(**CACGAT**)GGAATGTCGCAACCAAATGC

M. paratuberculosis

pAJB303	GACGACGAGGGCGG(**TGGCGT**)CGCCGGTGTAGCCGAA(**CGGCAC**)GTGCGCG[T]AGGCCCAGAT
pAJB86	CCACCTTACTCCCGA(**TGACGT**)TGCACGGCTGGGATTAA(**CGGTCC**)GCGTGC[T]CCAGGAGACA
pAJB125	GCAACGAGCGCATCA(**TTAAAG**)ATCGANGGCGCCGGGNT(**CATGTC**)CCTTCAC[C]CCGCCCAGCT
pAJB300	TCGAGTTCAAGACCC(**TGACGC**)TGGCCGACCTCGGCGCG(**CAGCCG**)ACCGCGC[A]GCGGTGCACG
pJB305	ATCCGGACGGGCAGT(**TGTTGG**)AGTTTCTGTCGGACGGT(**TGGTTG**)GCGGCAT[T]TCCGGCGAGG
pAJB304	CACCAGGTACACGCC(**AAGGAC**)AACGGCCGTATCCGGTA(**CCAACG**)GGTGTGC[G]AGCTGGACGG
P_{AN}	CTGGTGAAGGGTGAA(**TCGACA**)GGTACACACAGCCGCCA(**TACACT**)TCGCTTC[A]TGCCCTTACG
pAJB73	GATCGGTG(**TGCCGC**)TTGAACCGGCCCAGCTCCCG(**CTCCAG**)GGTGACG[T]GCTCGAGCTC
pAJB301	GATCTGGCGGGCGG(**TCCAGT**)ACACCGCGAGTTCGCGCACG(**CTGGCC**)GGCAGCGTCTTGGACGCCCG

M. fortuitum

repA	GAGCTCGTGTCGGACCATACACCGGTGATTAATCG*TGG*(**TCTACT**)ACCAAG[C]
rrnA PCL1	CCAGGATGATGCAAC(**TTGACT**)TGCCGGCAAGATTCGAAT(**TAAGCT**)GGCGGG[G]TTGCCCCAAA
rrnA P1	GAAAACCTGTTGAGC(**CTCGGA**)GCCGAGATCGAAAGAG(**TAGGGT**)CGTAAACAGCAGTCCGGGCC
rrnA P2a	CGCTGACCAGCCGAT(**TTGACC**)TTGTAGGCAGGCCCGCGC(**TAATCT**)TTTGAAGTCGCGCGGAGCGG
rrnA P2b	CCGGGCCAGAGCGAC(**TTGACA**)AGCCAGCCGAGATCGTAC(**TAAGCT**)GGCGAGGTTGCCTCAGACCG
rrnA P3	CAGGATGATGCAACT(**TGACTT**)GCCGGCAAGATTCGAATT(**AAGCTG**)GCGGGGTTGCCCCAAAACAG

M. phlei

rrnA PCL1	ACTGGGGACGAGGTC(**TTGACG**)CCCCTGATCAGATCGGTA(**TAGACT**)GGCAGG[G]TTGCCCGAAA
rrnA P1	GAGAACCTCCGCAGT(**CTCGGC**)GCCGAGATCGAGAGGG(**TCGCCT**)GAAACATGCCGTTACCTGC
rrnA P2	AGGGGACCCCCCTTT(**TTGACT**)CCGCTCAGACGTGGGC(**TATTCT**)TCTAACCACAAGCCCAACGC
rrnA P3	CTGGGGACGAGGTCT(**TGACGC**)CCCTGATCAGATCGGTAT(**AGACTG**)GCAGGGTTGCCCGAAAGCAA

Mycobacteriophage I3

pKGR25	CCTGTACACCCTCGC(**TGCACT**)CGCCGAGGACAAG(**CACTAT**)CGCCCCGACGTCCCGGCCTGG
pKGR9	ACCACGAGCACCCGG(**TCGTCA**)GGACTGCGACACTCGA(**TGTTGT**)AGACGCACTGGTGCAGCATG
pKGR38	ATCTGGTCGACCTGC(**TCGACG**)AGGTCGATCATCTTCT(**TCATCT**)CGCCGAACGGGATGCCCTGG
ORF1	ACCTCATGGAGCACT(**TCGAGG**)TCACTGAGCACGCCCA(**CGAACT**)ACGAGAGGCCGTGGGACTGG
ORF2	TACTTTTTGTACCGT(**TCGACA**)CCAGCGGTTTCCGCTTCCT*TGC*(**CAATCT**)CCTGCAAACAAACCACAATG
	ACACAGACCAGGAGC(**TCGACA**)TGACCGCCACCGCCCCC*TAC*AGCG(**TCATCT**)GGTTCGAAGGCACCCCGGAT
pKGR1	

Mycobacteriophage L5

71 P2	TACCTGTCACAAGGT(**TTGCTA**)CCGAGTGGGGCAGGCCGC(**TACATT**)TACGACC[G]CGTAACGCCA
71 P_{left}	TTTGCGATTAGGGC(**TTGACA**)GCCACCCGGCCAGTAGTG(**CATTCT**)TGTGTC[A]CCGCAGCAGC
71 P1	ACAACTGAATATGGT(**TCCGCA**)GACGCAACTAAATTAGGGG(**TATCCT**)TGACA[G]GCACCAACAT

M. avium

Avi-3	GCCGGCGATCGTGGG(**CTGATA**)AGTCTTATCGGGCATAC(**TATAAG**)TGTAGTGGGAAATATCACCT
pLR7	AGCCTTGTTGGCGGC(**CAACTG**)CCGGACGATCGCGGCGGC(**CATCGT**)CCTCGAGCTCGGCCCCGTGC

M. neoaurum

rrnA PCL1	GCGAGACAGAGAAGC(**TTGACT**)CGCCAGACAAGATAGTT(**TAAGCT**)GGCAGG[G]TTGCCCCGAA
rrnA P1	GAAAACCTGGTCAGC(**TTGGGC**)GCCGGGATCGAGCGAG(**TACACT**)CGTAAGAGACCGGTCGAGTG
rrnA P3	GCGAGACAGAGAAGC(**TTGACT**)CGCCAGACAAGATAGTT(**TAAGCT**)GGCAGGGTTGCCCGAAACG
rrnA P2	CTCTGACCAGCGGAT(**TTGACT**)TCCGAAGGCACAAAGTTC(**TAATCT**)TTTGAAGTCGCCGCGGGGAG

M. abscessus

rrnA P4	GCCAAAACCGGGAAT(**TTGACT**)CAGGTTCACGAACT*TGA*(**TACGGT**)TTCCGA[G]CGCCCGAAAG
rrnA P1	GGCGGGTCTAGTGGC(**GGACGG**)CGTCACAGAGGTAT*ACGA*(**TGTGTT**)TCATATCG[A]CCGCGGTTAC

rrnA PCL1	GCCCCCGACCCGAAG(**TTGACT**)CAAGTTCATTGGACT*TGG*(**TACAGT**)GGTCGG[G]TTGCCCTGAA
rrnA P2	GCCAAAACCGGGAAT(**TTGACT**)CAAGTTCACCGAACT*TGA*(**TACGGT**)TTCC[A]AGTCGCTCGG
rrnA P3	GCCAAAACCGGGAAT(**TTGACT**)CAAGTTCACCGAACT*TGA*(**TACGGT**)TTCCAA[G]TCGCTCGGAA

M. chelonae

rrnA P2	CCAAAACCCGGAGTT(**TGACTC**)AAGTTCACCGAACT*TGA*(**TCGGTT**)CCCGG[G]CCGCTTACAA
rrnA P1	GGCGGGGTTAGTGGC(**GGATGG**)CGTCACCGAGGTATACGA(**TGTGTT**)TCATATC[G]ACCGCGGTTA
rrnA PCL1	CCCCAGAACCCGAAG(**TTGACT**)CAAGTTCATTGGACT*TGG*(**TACAGT**)GGTCGG[G]TTGCCCTGAA
rrnA P3	GCCAAAACCGGGAAT(**TTGACT**)CAAGTTCACCGAACT*TGA*(**TCGGTT**)TCCCA[G]CCGCCCGAAA
rrnA P4	GCCAAAACCGGGAAT(**TTGACT**)CAAGTTCACCGAACT*TGA*(**TACGGT**)TTCCGA[G]CCGCCCGAAA

References

[1] T.M. Shinnick, The 65-kilodalton antigen of *Mycobacterium tuberculosis*, J. Bacteriol. 169 (1987) 1080–1088.

[2] J.E. Thole, H.G. Dauwerse, P.K. Das, D.G. Groothuis, L.M. Schouls, E.J.D. van, Cloning of *Mycobacterium bovis* BCG DNA and expression of antigens in *Escherichia coli*, Infect. Immun 50 (1985) 800–806.

[3] V. Mehra, D. Sweetser, R.A. Young, Efficient mapping of protein antigenic determinants, Proc. Natl Acad. Sci. USA 83 (1986) 7013–7017.

[4] L. Kremer, A. Baulard, J. Estaquier, J. Content, A. Capron, C. Locht, Analysis of the *Mycobacterium tuberculosis* 85A antigen promoter region, J. Bacteriol. 177 (1995) 642–653.

[5] F. Movahedzadeh, M.J. Colston, E.O. Davis, Determination of DNA sequences required for regulated *Mycobacterium tuberculosis* RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist, J. Bacteriol. 179 (1997) 3509–3518.

[6] M.D. Bashyam, D. Kaushal, S.K. Das Gupta, A.K. Tyagi, A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements, J. Bacteriol. 178 (1996) 4847–4853.

[7] M.A. Mulder, H. Zappe, L.M. Steyn, Mycobacterial promoters, Tubercle. Lung Dis 78 (1997) 211–223.

[8] C.B. Hoopes, W.R. McClure, Strategies in regulation of transcription initiation, in: F.C. Neidhart (Ed.), *Escherichia coli* and *Salmonella typhimurium*, American Society of Microbiology, Washington DC, 1987, pp. 1231–1239.

[9] J. Ramstein, R. Lavery, Energetic coupling between DNA bending and base pair opening, Proc. Natl. Acad. Sci. USA 85 (1988) 7231–7235.

[10] J.D. Parvin, R.J. McCormick, P.A. Sharp, D.E. Fisher, Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor, Nature 373 (1995) 724–727.

[11] C.M. Collis, P.L. Molloy, G.W. Both, H.R. Drew, Influence of the sequence-dependent flexure of DNA on transcription in *E. coli*, Nucl. Acids Res 17 (1989) 9447–9468.

[12] R.R. Plaskon, R.M. Wartell, Sequence distributions associated with DNA curvature are found upstream of strong *E. coli* promoters, Nucl. Acids Res 15 (1987) 785–796.

[13] L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann, H. Buc, Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*, EMBO J. 8 (1989) 4289–4296.

[14] T. Ohyama, M. Nagumo, Y. Hirota, S. Sakuma, Alteration of the curved helical structure located in the upstream region of the beta-lactamase promoter of plasmid pUC19 and its effect on transcription, Nucl. Acids Res 20 (1992) 1617–1622.

[15] T. Gaal, L. Rao, S. Estrem, J. Yang, R. Wartell, R. Gourse, Localization of the intrinsically bent DNA region upstream of the *E.coli* rrnB P1 promoter, Nucl. Acids Res 22 (1994) 2344–2350.

[16] M. Engelhorn, A. Geiselmann, Maximal transcriptional activation by the IHF protein of *Escherichia coli* depends on optimal DNA bending by the activator, Mol. Microbiol 30 (1998) 431–441.

[17] S.E. Aiyar, R.L. Gourse, W. Ross, Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit, Proc. Natl. Acad. Sci. USA 95 (1998) 14652–14657.

[18] E.N. Trifonov, Curved DNA, CRC Crit. Rev. Biochem. 19 (1985) 89–106.

[19] P.J. Hagerman, Sequence-directed curvature of DNA, Annu. Rev. Biochem 59 (1990) 755–781.

[20] A. Klug, A. Jack, M.A. Viswamitra, O. Kennard, A. Shakked, T.A. Steitz, A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the *lac*-repressor protein, J. Mol. Biol. 131 (1979) 669–680.

[21] R.E. Dickerson, H.R. Drew, Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure, J. Mol. Biol. 149 (1981) 761–786.

[22] P.J. Hagerman, Evidence for the existence of stable curvature of DNA in solution, Proc. Natl. Acad. Sci. USA 81 (1984) 4632–4636.

[23] S.C. Satchwell, H.R. Drew, A.A. Travers, Sequence periodicities in chicken nucleosome core DNA, J. Mol. Biol. 191 (1986) 659–675.

[24] C.R. Calladine, H.R. Drew, M.J. McCall, The intrinsic curvature of DNA in solution, J. Mol. Biol. 201 (1988) 127–137.

[25] A. Bolshoy, P. McNamara, R.E. Harrington, E.N. Trifonov, Curved DNA without A–A: experimental estimation of all 16 DNA wedge angles, Proc. Natl. Acad. Sci. USA 88 (1991) 2312–2316.

[26] C.A. Hunter, Sequence-dependent DNA structure. The role of base stacking interactions, J. Mol. Biol. 230 (1993) 1025–1054.

[27] C.A. Hunter, Sequence-dependent DNA structure, Bioessays 18 (1996) 157–162.

[28] D.S. Goodsell, R.E. Dickerson, Bending and curvature calculations in B-DNA, Nucl. Acids Res 22 (1994) 5497–5503.

[29] I. Brukner, R. Sanchez, D. Suck, S. Pongor, Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides, EMBO J. 14 (1995) 1812–1818.

[30] M.A. EI Hassan, C.R. Calladine, Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA, J. Mol. Biol. 259 (1996) 95–103.

[31] C. Anselmi, G. Bocchinfuso, P. De Santis, M. Savino, Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability, J. Mol. Biol. 286 (1999) 1293–1301.

[32] M. Dlakic, R.E. Harrington, The effects of sequence context on DNA curvature, Proc. Natl. Acad. Sci. USA 93 (1996) 3847–3852.

[33] P. De Santis, A. Palleschi, M. Savino, A. Scipioni, A theoretical model of DNA curvature, Biophys. Chem 32 (1988) 305–317.

[34] C.R. Calladine, Mechanics of sequence-dependent stacking of bases in B-DNA, J. Mol. Biol. 161 (1982) 343–352.

[35] R.E. Dickerson, Base sequence and helix structure variation in B and A DNA, J. Mol. Biol. 166 (1983) 419–441.

[36] E.S. Shpigelman, E.N. Trifonov, A. Bolshoy, curvature: software for the analysis of curved DNA, Comp. Appl. Biosci 9 (1993) 435–440.

[37] E.N. Trifonov, L.E. Ulanovsky, Inherently curved DNA and its structural elements, in: R.D. Wells, S.C. Harvey (Eds.), Unusual DNA Structures, Springer-Verlag, Berlin, 1987, pp. 173–187.

[38] P. De Santis, A. Palleschi, M. Savino, A. Scipioni, Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature, Biochemistry 29 (1990) 9269–9273.

[39] P. De Santis, A. Palleschi, M. Savino, A. Scipioni, Theoretical prediction of the gel electrophoretic retardation changes due to point mutations in a tract of SV40 DNA, Biophys. Chem 42 (1992) 147–152.

[40] T.M. Nair, B.D. Kulkarni, On the consensus structure within the *E. coli* promoters, Biophys. Chem 48 (1994) 383–393.