Biophysical Chemistry, 48 (1994) 383–393
Elsevier Science B.V., Amsterdam

BIOCHE 01802

On the consensus structure within the *E. coli* promoters

T. Murlidharan Nair and B.D. Kulkarni *

Chemical Engineering Division, National Chemical Laboratory [1], Pune 411 008 (India)

(Received 12 February 1993; accepted in revised form 26 July 1993)

Abstract

Using the theoretical model of DNA curvature, we have studied about 112 different *E. coli* promoters with a view to obtain some common super structures associated with them. Out of the 112 promoters analyzed by theoretical gel electrophoresis permutation about 66 of them have their minima lying between the −10 and the −35 region. The analysis of the bases at the minima reveals strong structural similarities. The differences can account for the varying strengths of the promoters as well as for different degree with which the RNA polymerase binds to these regions. The effects of mutation in each of these 112 promoters and their changes in curvature dispersion have also been evaluated.

Keywords: DNA bending; Consensus sequence; Consensus structure; Promoters

1. Introduction

A promoter is a start signal at the beginning of a gene or a gene cluster that directs the RNA polymerase to initiate RNA synthesis [1]. Since the enzyme *E. coli* RNA poylmerase binds to precisely the promoter region, this region is expected to share some characteristic feature of protein-DNA recognition which may be nucleotide sequence and/or DNA structure. Earlier analysis of the *E. coli* promoter sequences have revealed that there are three regions about 35, 16 and 10 base pairs (bp) upstream from the transcription start site [1–4] and have extended the concept of "consensus" promoter sequence [5] namely a −35 (TTGACA) and −10 (TATAAT) region separated by 17 bp where the transcription initiation is known to occur at a purine about 7 bp downstream from the 3′ end of the −10 region. Weak nucleotide homology was detected at the −16 region. None of the promoters contain all 12 nucleotides but some of these bases occur more frequently than the others. For example, the last T of the −10 region occurs invariably. The first TA of the −10 region and the TTG of the −35 site are also very frequent. Thus we see only a partial sequence homology.

The promoter regions have also been the sites of nearly all mutations which were known to affect the transcriptional strength, while the other bases flanking these regions in addition to the start point sometimes affect promoter activity [5,6]. The variation in spacing of region between the −10 and the −35 also play a role in the promoter strength [7–10]. Both genetic and biochemical studies indicate that these two regions and the distance between them are of prime

* To whom correspondence should be addressed.

[1] NCL communication number: 5665.

SSDI 0301-4622(93)E0056-B

importance in defining the overall strength or transcription initiation rate of a promoter.

There have been numerous attempts to predict the location of the promoter sequences on the basis of homology either by searching for consensus sequence or to a reference list of promoters. Numerous algorithms have been also proposed and used in searching for new sequences [11,12]. In the studies carried out by Bossi and Smith [13], and Gourse et al. [14] the DNA region responsible for the enhancement of transcription exhibited lower than expected electrophoretic mobility on polyacrylamide gels. Physiochemical evidence indicates that the abnormally low electrophoretic mobilities of DNA fragments can be due to a curved or bent DNA structure [15–25]. Bending may directly facilitate the separation of strands required in open complex formation. Molecular modelling studies on the bends indicate that the bends lower the activation energy for the opening of the double helix [26]. Sequences likely to bend or those that are predicted to assume unusual helical twists are frequently found upstream from [27–29] and within promoter sequences [13,14]. Promoter mutations have been correlated with altered gel electrophoretic mobility and is taken as evidence of bending of the DNA alone [13,14]. Genetic and biochemical studies can be used to identify promoter elements which can pinpoint bases which interact with RNA polymerase but such data is not available for most genes.

Recently, De Santis et al. have proposed a theoretical model for DNA curvature [30–32]. The model predicts the intrinsic DNA superstructures that have been experimentally investigated with a very high degree of reliability [33]. The predictive power of such a model has led us to use it as a tool in understanding the superstructures associated with the promoters which has been a subject of intensive theoretical investigation. In the present study we have not biased ourselves to the three consensus region, the results reported here demonstrate structural homologies with respect to the four basic rules for predicting the spatial behavior of a DNA segment put forth by Calladine [34] and Dickerson [35] at the region of minima obtained by using the theoretical model.

2. Model

The model [30] predicts the DNA superstructures based on the local deformations of the dinucleotide double helix fragments. The structural deviations from the canonical B-DNA structure is obtained in terms of the curvature vector $C(n, v)$ which represents in the complex plane (in modulus and phase) the directional change of the double helix axis between sequence number n and $n+v$ is given per turn of the B-DNA by

$$C(n, v) = v^0/v \sum_{j=n}^{n+v} d_j \, e^{2\pi i(j-1)/v^0} \qquad (1)$$

where $v^0 = 10.4$ is the average periodicity of the DNA and $d_j = (\rho_j - i\tau_j)$ represents the orientational deviations of the jth base pair average plane from the canonical B-DNA. ρ_j and τ_j are the roll and tilt angles respectively. The energy minimized values of roll and tilt as obtained by De Santis et al. is given in the following matrices.

ρ (rad)	A	T	G	C
T	−0.16	0.09	−0.12	−0.04
A	0.09	0.12	0.02	0.04
C	−0.12	0.02	−0.10	−0.01
G	−0.04	0.04	−0.01	0.07

τ (rad)	A	T	G	C
T	0.00	0.00	0.01	−0.03
A	0.00	0.00	0.03	0.03
C	−0.01	−0.03	0.00	−0.02
G	0.03	−0.03	0.02	0.00

The curvature dispersion, σ^2, can be evaluated as follows

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} \left\{ \sum_{j=1}^{n} d_j \, e^{2\pi i(j-1)/v^0} \right\}^2 - \left\{ \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{n} d_j \, e^{2\pi i(j-1)/v^0} \right\}^2 \qquad (2)$$

De Santis et al. [30] have shown that this complex plane σ^2 reproduces the trend of the

permutation gel electrophoresis of DNA tracts that have been investigated by different authors for their biological relevance [36–39]. A direct comparison between the experimental permutation gel electrophoresis and the pertinent σ^2 diagram was also reported. The full agreement between the theoretical and experimental permutation assay for DNA tracts up to 700 bp, indicates a long range conservation of the DNA writhing (or at leasts of its second moment) and offer an easy theoretical alternative to the experimental technique [40].

3. The Calladine–Dickerson rules

Calladine [34] formulated a set of four rules and explained the departure of the DNA dodecamer CGCGAATTCGCG, from the ideal regular helical structure of B-DNA. The DNA dodecamer was crystalized and experimentally analyzed by Dickerson and Drew earlier [41–45]. The departures from the classical B-DNA is attributed to simple steric hindrance of the nearest neighbor purine on the opposite strand which stems from (i) the larger size of the purine, and (ii) the propeller twist of the base pairs in DNA. Let us consider the latter effect in more detail. The two bases which constitute a base pair in a B-DNA do not lie in the same plane, instead they are rotated relative to each other on their sugar–sugar virtual bond, i.e one is rotated clockwise and its complement anti-clockwise. This rotation or propeller twist, improves the stacking of the bases on the same strand. The double-ring purine extends past the center of the base pair and the hydrogen bonded gap does not lie midway between the two bases. Thus the helix axis passes through the purine.

It is easy to comprehend from the above discussion that the problem of steric hindrance will arise if the bases of unequal size on the two strands are twisted in opposite direction. If the nearest neighbors are of the type Pu–Py/Pu–Py or Py–Pu/Py–Pu, no strain will occur. But in the case of the base pairs Pu–Py/Py–Pu or Py–Pu/Pu–Py, the distance between the purines as pointed out by Calladine would drop from the van der Waals' spacing of 3.4 Å to about 2.8 Å, leading to a strong steric clash. In the case of Py → Pu on each strand, a severe clash would occur in the minor groove of the DNA while Pu → Py would lead to a weaker clash in the major groove.

Calladine used principles of elastic beam mechanics to analyze this clash and ways of relieving it. He proposed that the DNA chain may ameliorate these van der Waals clashes in four ways:

(i) Flatten the propeller twist in one or both pairs.

(ii) Open up the roll angle between base pairs on the side where clash is found.

(iii) The DNA backbone can shift sideways towards the pyrimidines.

(iv) Decrease the local helix twist angle at the step at which clash occurs.

Calladine showed that the above four rules were employed by the DNA dodecamer, CGCGAATTCGCG. Dickerson [35] extended this analysis and quantified it 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 δ at the two ends of the base pair (Σ_3), and flattening of propeller twist (Σ_4) and tested it by calculating structures of sequences for which the crystal structures were already known. His results are in agreement with the crystal structures. The algorithm for constructing these sum functions is as follows:

(1) The twist angle perturbation in the helix is measured by the sum function (Σ_1). Relieving the purine steric hindrance means decreasing the twist angle at that base pair step and increasing it at the two neighboring steps. In the case of the major groove clash X–Pu–Py–X, the values +1,

	R	-	R	-	Y	-	Y	-	Y	-	Y	-	R	-	R
Σ_1(twist)			+1		-2		+1				+2		-4		+2
Σ_2(roll)			+1		-2		+1				-2		+4		-2
$\Sigma_3(\delta)$					+1		-1						-2		+2
Σ_4(pr.tw)					-1		-1						-2		-2

Fig. 1. The algorithm for calculating the sum functions.

−2, +1 are assigned to the corresponding base pair steps. (X = any base). For the more severe minor groove clash X–Py–Pu–X the values are doubled +2, −4, +2.

(2) The variations of the roll angle θ_R is measured using the sumfunction (Σ_2). The roll angle is positive if it opens towards the minor groove and negative if it opens towards the major groove. Relieving a clash in the minor groove leads to more positive θ_R and for alleviating a clash in the major groove yields more negative θ_R As the minor groove clash is about twice as severe,

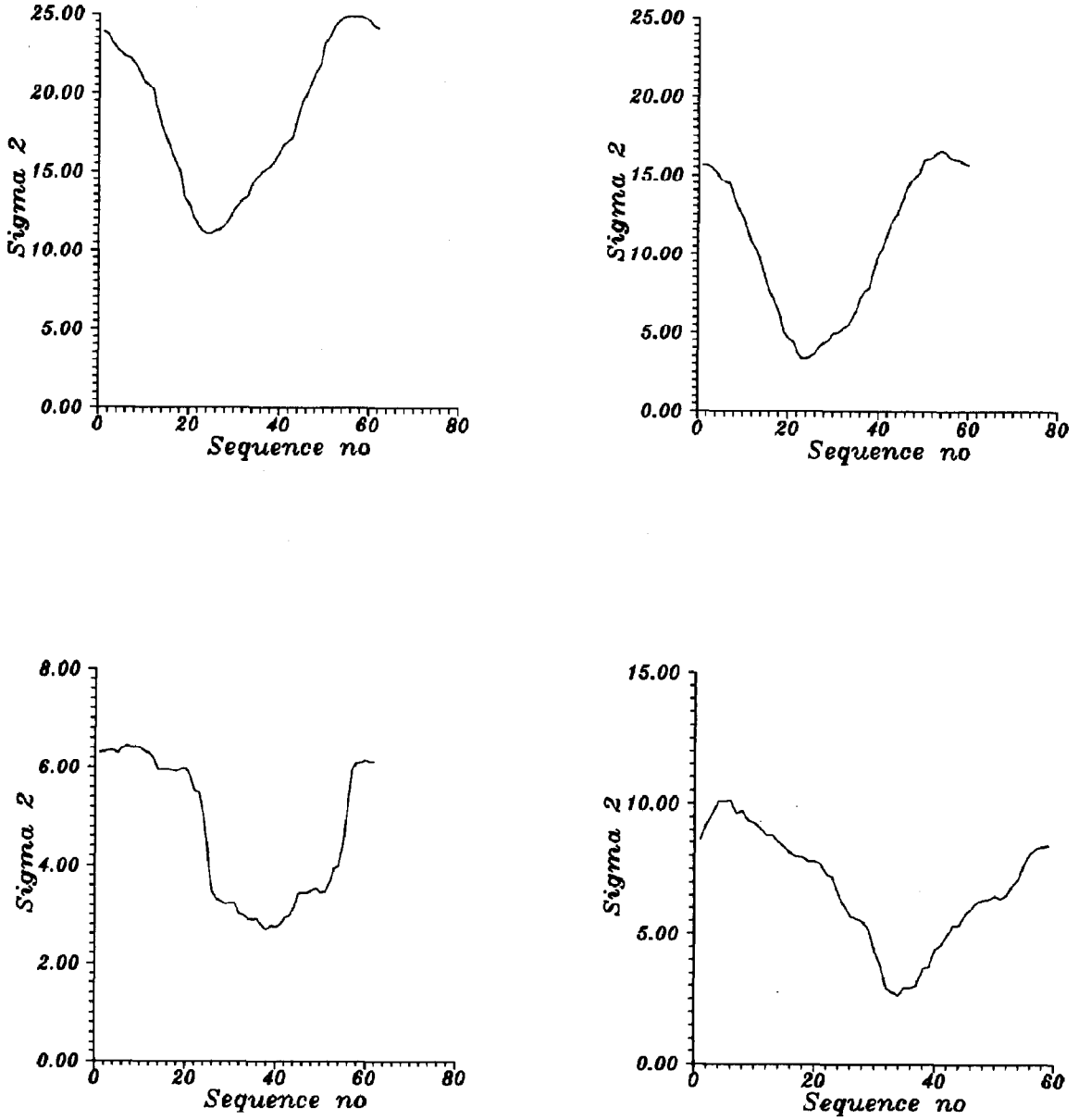

Fig. 2. Calculated variance σ^2 by a cyclic permutation of the sequence for the promoters *araBAD*, *galP2*, ϕ*XD*, λ*cI7*. The minima corresponds to the region where the bend is localized.

Table 1

The promoters for *E. coli* RNA polymerase (see Hawley and McClure [5]). The sequences at the minima are underlined

S no.	Promoter	minima	TTGACA	TATAAT	+1
1	araBAD	−28	TTAGCGGATCCTACCTGACGCTTTT	TATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTT	
2	araC	−12	GCAAATAATCAATGTGGACTTTTCT	GCCGTGATTATAGACACTTTTGTTACGCGTTTTTGT	
3	galP1	−30	CTAATTTATTCCATGTCACACTTTTCGCA	TCTTTGTTATGCTATGGTTAT	TTCATACCATAAG
4	galP2	−26	CACTAATTTATTCCATGTCACACTT	TTCGCATCTTTGTTATGCTATG	GTTATTTCATACC
5	lacP1	+1	TAGGCACCCCAGGCTTTACACTTTA	TGCTTCCGGCTCGTATGTTGTG	TGGAATTGTGAGC
6	lacI	−35	GACACCATCGAATGGCGCAAAACCT	TTCGCGGTATGGCATGATAGC	GCCCGGAAGAGAGT
7	malEFG	−30	AGGGGCAAGGAGGATGGAAAGAGGT	TGCCGTATAAAGAAACTAGA	GTCCGTTTAGGTGT
8	malK	−33	CAGGGGGTGGATGATTTAAGCCATC	TCCTGATGACGCATAGTCAG	CCCATCATGAATG
9	tnaA	−27	AAACAATTTCAGAATAGACAAAAAC	TCTGAGTGTAATAATGTAGC	CTCGTGTCTTGCG
10	doeP1	−15	CAGAAACGTTTTATTCGAACATCGA	TCTCGTCTTGTGTTAGAATTCT	AACATACGGTTGC
11	trp	−4	TCTGAAATGAGCTGTTGACAATTAA	TCATCGAACTAGTTAACTAGT	ACGCAAGTTCACGT
12	trpR	−12	TGGGGACGTCGTTACTGATCCGCAC	GTTTATGATATGCTATCGTACT	CTTTAGCGAGTACA
13	aroH	−26	GTACTAGAGAACTAGTGCATTAGCT	TATTTTTTTGTTATCATGCT	AACCACCCGGCGAG
14	trpP2	−30	ACCGGAAGAAAACCGTGACATTTTA	ACACGTTTGTTACAAGGTAAA	GGCGACGCCGCCC
15	leu	−18	GTTGACATCCGT	TTTTGTATCCAGTAACTCTAA	AAGCATATCGCATT
16	ilvGEDA	−21	GGCCAAAAAATATCTTGTACTATTT	ACAAAACCTATGGTAACTCTTT	AGGCATTCCTTCG
17	argCBH	−8	TTTGTTTTTCATTGTTGACACACCT	CTGGTCATGATAGTATCAATATTCATGCAGTATT	
18	bioA	−30	GCCTTCTCCAAAACGTGTTTTTTGT	TGTTAATTCGGTGTAGACTTGT	AAACCTAAAMTCT
19	fol	−1	CATCCTCGCACCAGTCGACGACGGT	TTACGCTTTACGTATAGTGGC	GACAATTTTTTTT
20	uvrB P1	−9	TCCAGTATAATTTGTTGGCATAATT	AAGTACGACGAGTAAAATTAC	ATACCTGCCCGC
21	uvrB P2	−44	TCAGAAATATTATGGTGATGAACTG	TTTTTTTATCCAGTATAATTTG	TTGGCATAATTAA
22	uvrB P3	−15	ACAGTTATCCACTATTCCTGTGGAT	AACCATGTGTATTAGAGTTAG	AAAACACGAGGCA
23	lexA	−26	TGTGCAGTTTATGGTTCCAAAATCG	CCTTTTGCTGTATATACTCAC	AGCATAACTGTAT
24	ampC	−30	TGCTATCCTGACAGTTGTCACGCTG	ATTGGTGTCGTTACAATCTA	ACGCATCGCCAATG
25	lpp	−31	CCATCAAAAAAATATTCTCAACATA	AAAAACTTTGTGTAATACTTGT	AACGCTACATGGA
26	Pori-r	−22	GATCGCACGATCTGTATACTTATTT	GAGTAAATTAACCCACGATCCC	AGCCATTCTTCTGC
27	spot42	−11	ATTACAAAAAGTGCTTTCTGAACTG	AACAAAAAAGAGTAAAGTTAG	TCGCGTAGGGTACA
28	MI RNA	−21	ATGCGCAACGCGGGGTGACAAGGGC	GCGCAAACCCTCTATACTGCG	CGCCGAAGCTGACC
29	trpS	−14	CTACGGCGAGGCTATCGATCTCAGC	CAGCCTGATGTAATTTATCAGTCTATAAATGACC	
30	tyrT	−20	TCTCAACGTAACACTTTACAGCGGC	GCGTCATTTGATATGATGCG	CCCCGCTTCCCGAT
31	leul tRNA	−17	TCGATAATTAACTATTGACGAAAAG	CTGAAAACCACTAGAATGCGCCTCCGTGGTAGCAA	
32	supB - E	−17	CCTTGAAAAAGAGGTTGACGCTGCA	AGGCTCTATACGCATAATGCG	CCCCGCAACGCCGA
33	rrnAB P1	−11	TTTTAAATTTCCTCTTGTCAGGCCG	GAATAACTCCCTATAATGCGCCACCACTGACACGG	
34	rrnD P1	−15	GATCAAAAAAATACTTGTGCAAAAA	ATTGGGATCCCTATAATGCGCCTCCGTTGAGACGA	
35	rrnE P1	−44	CTGCAATTTTTCTATTGCGGCCTGC	GGAGAACTCCCTATAATGCGCCTCCATCGACACGG	
36	rrnX P1	−44	ATGCATTTTTCCGCTTGTCTTCCTG	AGCCGACTCCCTATAATGCGCCTCCATCGACACGG	
37	RRNG P2	−13	AAGCAAAGAAATGCTTGACTCTGTA	GCGGGAAGGCGTATTATGCA	CACCGCCGCGCCG
38	rrnDEx	−34	CCTGAAATTCAGGGTTGACTCTGAA	AGAGGAAAGCGTAATATACG	CCACCTCGCGACAG
39	str	−28	TCGTTGTATATTTCTTGACACCTTT	TCGGCATCGCCCTAAAATTCG	GCGTCCTCATAT
40	spc	−17	CCGTTTATTTTTTCTACCCATATCC	TTGAAGCGGTGTTATAATGCC	GCGCCCTCGATA
41	s10	−18	TACTAGCAATACGCTTGCGTTCGGT	GGTTAAGTATGTATAATGCG	CGGGCTTGTCGT
42	rpoA	−39	TTCGCATATTTTTCTTGCAAAGTTG	GGTTGAGCTGGCTAGATTAGC	CAGCCAATCTTT
43	rplJ	−12	TGTAAACTAATGCCTTTACGTGGGC	GGTGATTTTGTCTACAATCTT	ACCCCCACGTATA
44	rpoB	−20	CGACTTAATATACTGCGACAGGACG	TCCGTTCTGTGTAAATCGCA	ATGAAATGGTTTAA
45	T7 A3	−16	GTGAAACAAAACGGTTGACAACATG	AAGTAAACACGGTACGATGTA	CCACATGAAACGAC
46	T7 C	−32	CATTGATAAGCAACTTGACGCAATG	TTAATGGGCTGATAGTCTTAT	CTTACAGGTCATC
47	T7 D	−10	CTTTAAGATAGGCGTTGACTTGATG	GGTCTTTAGGTGTAGGCTTTA	GGTGTTGGCTTTA
48	λPR	−30	TAACACCGTGCGTGTTGACTATTTT	ACCTCTGGCGGTGATAATGGT	TGCATGTACTAAG
49	λPRM	−30	AACACGCACGGTGTTAGATATTTAT	CCCTTGCGGTGATAGATTTAA	CGTATGAGCACAA
50	λPL	−30	TATCTCTGGCGGTGTTGACATAAAT	ACCACTGGCGGTGATACTGAG	CACATCAGCAGGA

Table 1 (continued)

S no.	Promoter	minima	TTGACA	TATAAT	+1
51	λPO	−26	TACCTCTGCCGAAGTTGAGTATTTT	TGCTGTATTTGTCATAATGAC	TCCTGTTGATAGAT
52	λPR′	−11	TTAACGGCATGATATTGACTTATTG	AATAAAATTGGGTAAATTTGA	CTCAACGATGGGTT
53	λPRE	−32	TAGAGCCTCGTTGCGTTTGTTTGCA	CGAACCATATGTAAGTATTTC	CTTAGATAACAAT
54	λP1	−14	CGGTTTTTTCTTGCGTGTAATTGCG	GAGACTTTGCGATGTACTTGA	CACTTCAGGAGTG
55	434 PRM	−13	ACAATGTATCTTGTTTGTCAAATAC	AGTTTTTCTTGTGAAGATTGG	GGGTAAATAACAGA
56	P22 PRM	−9	CATCTTAAATAAACTTGACTAAAGA	TTCCTTTAGTAGATAATTTA	AGTGTTCTTTAAT
57	φXD	−12	TAGAGATTCTCTTGTTGACATTTTA	AAAGAGCGTGGATTACTATCTGAGTCCGATGCTGTTC	
58	φXB	−31	GCCAGTTAAATAGCTTGCAAAATAC	GTGGCCTTATGGTTACAGTATG	CCCATCGCAGTT
59	fD X	−38	TCCTCTTAATCTTTTTGATGCAATTCGCTTTGCTTCTGACTATAATAGA		CAGGGTAAAGACCT
60	pBR322 bla	−25	TTTTTCTAAATACATTCAAATATGT	ATCCGCTCATGAGACAATAAC	CCTGATAAATGCT
61	pBR P1	−9	TTCATACACGGAGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACC		GCATTAAAGCTTA
62	pBR RNAI	−6	GTGCTACAGAGTTCTTGAAGTGGTG	GCCTAACTACGGCTACACTAGA	AGGACAGTATTTG
63	pBR primer	−26	ATCAAAGGATCTTCTTGAGATCCTT	TTTTTCTGCGCGTAATCTGCT	GCTTGCAAACAAAA
64	pBR322 P4	−37	CATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTG		CTCTGATGCCGCAT
65	ColE1 P1	−26	GGAAGTCCACAGTCTTGACAGGGAA	AATGCAGCGGCGTAGCTTTTA	TGCTGTATATAAAA
66	ColE1 P2	−11	TTATTTTTAACTTATTGTTTTAAAA	GTCAAAGAGGATTTTATAATGGA	AACCGCGGTAGCGT
67	RSF primer	−3	GGAATAGCTGTTCGTTGACTTGATA	GACCGATTGATTCATCATCTC	ATAAATAAAGAA
68	100 RNAII	−6	ATGGGCTTACATTCTTGAGTGTTCA	GAAGATTAGTGCTAGATTACT	GATCGTTTAAGGAA
69	RI RNAII	−10	ACTAAAGTAAAGACTTTACTTTGTG	GCGTAGCATGCTAGATTACT	GATCGTTTAAGGAA
70	R100RNAIII	−24	GTACCGGCTTACGCCGGGCTTCGGC	GGTTTTACTCCTGTATCATATG	AAACAACAGAG
71	Cat	−19	ACGTTGATCGGCACGTAAGAGGTTC	CAACTTTCACCATAATGAA	ATAAGATCACTACC
72	Tn10Pout	−30	AGTGTAATTCGGGGCAGAATTGGTA	AAGAGAGTCGTGTAAAATATC	GAGTTCGCACATC
73	Tn10 tetA	−3	ATTCCTAATTTTTGTTGACACTCTA	TCATTGATAGAGTTATTTTACC	ACTCCCTATCAGT
74	GLTNPR	−19	ATTCATTAACAATTTTGCAACCGTC	CGAAATATTATAAATTATC	GCACACATAAAAAC
75	Tn5 IR	−15	TCCAGGATCTGATCTTCCATGTGAC	CTCCTAACATGGTAACGTTCA	TGATAACTTCTGCT
76	λcI7	−14	GGTGTATGCATTTATTTGCATACAT	TCAATCAATTGTTATAATTGT	TATCTAAGGAAAT
77	λL57	−11	TTGATAAGCAATGCTTTTTTATAAT	GCCAACTTAGTATAAAATAGC	CAACCTGTTCGACA
78	IS2 I-II	−22	ATGTCTGGAAATATAG	GGGCAAATCCACTAGTATTAA	GACTATCACTTATT

X–Py–Pu–X is given the value −2, +4, −2 and X–Pu–Py–X is assigned +1, −2, +1.

(3) The torsion angle δ is defined by the sum function (Σ_3) which is the property of the base pairs themselves, and not base pair steps. Since the shift is towards the pyrimidine, δ increases at the purine side and decreases at the pyrimidine side. The Pu–Py major groove clash produces the value +1, −1 and the Py–Pu clash produces a clash in the minor groove which is twice as severe and is assigned −2, +2.

(4) The flattening of the propeller twist is given by the sum function (Σ_4) which is also a property of the base pair. The value of −1, −1 is assigned for Pu–Py clash and −2, −2 for Py, Pu hindrance.

The definitions of (Σ_1) (helical twist), (Σ_2) (base roll), (Σ_3) (torsion angle difference), and (Σ_4) (propeller twist) are summarized in Fig. 1.

4. Method

The method employed here is same as the one used by De Santis et al. [30] for the theoretical gel electrophoresis permutation which predicts the experimental permutation gel electrophoresis for localizing the DNA curvature, as introduced by Wu and Crothers [36] with a very high degree of reliability. We have calculated the variance σ^2 of all the 112 promoter sequences compiled by Hawley and McClure [5] by cyclic permutation of the sequence. The curvatures were calculated by

taking an integration grid of 31 bp which corresponds to three complete turns of the helix.

The minima obtained by the cyclic permutation of the sequence corresponds to the region where the bend is localized (see Fig. 2). About five bases on either side of the minima were taken and aligned with the sequences of the promoter.

We then analyzed these localized region by applying Calladines rules and calculated the four sum functions.

These promoters were then analyzed for the changes in the curvature dispersion due to a single base change using the same model. The model has been very successfully used to predict the changes in the electrophoretic retardation as

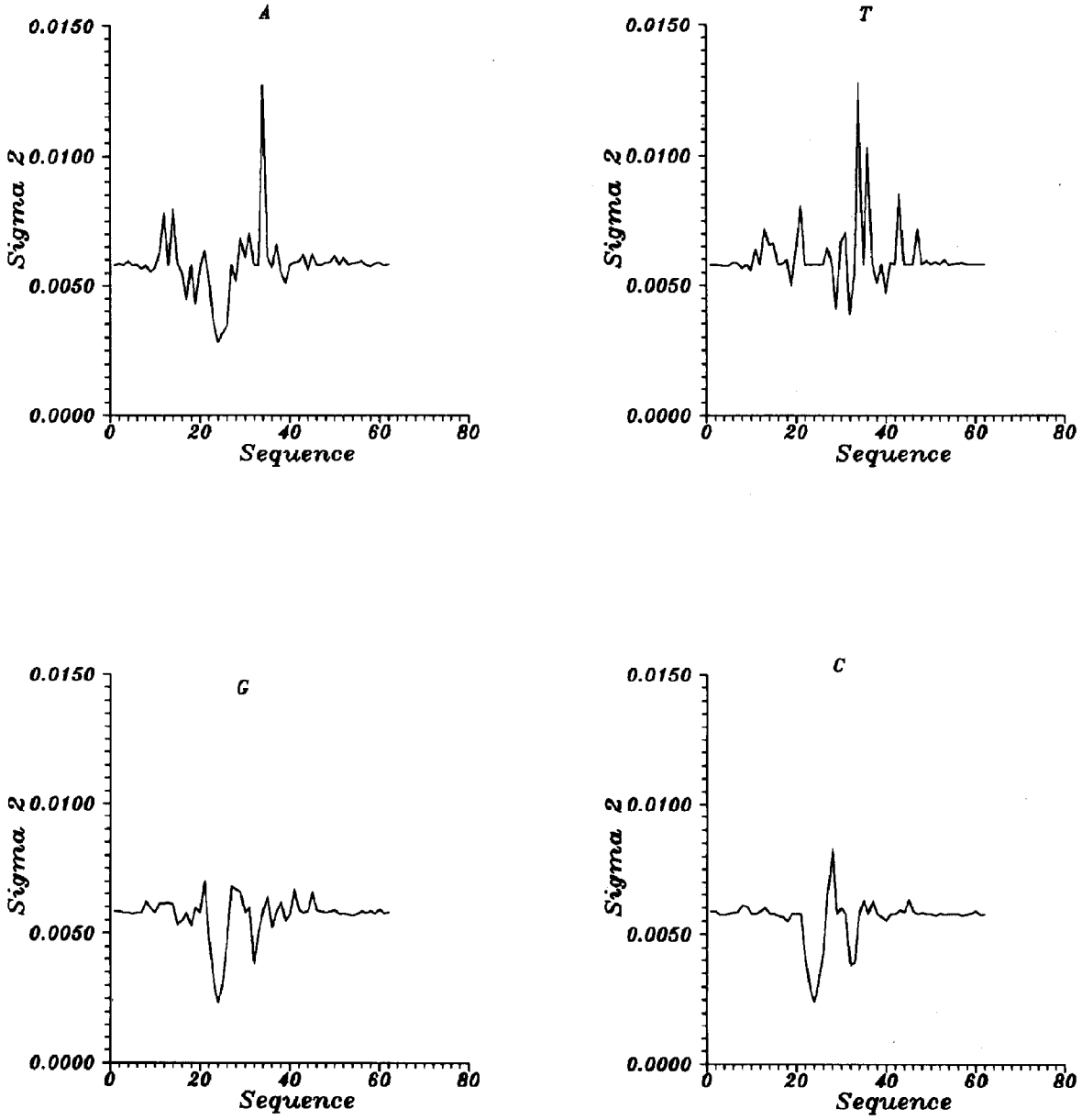

Fig. 3. The graphs represent the change in curvature dispersion for a single base change in the case of *araBAD* promoter.

a result of single base substitutions in the case of the 173 bp fragment of SV40 [33]. The experimental investigations for the electrophoretic retardation of this fragment was done by Milton et al. [46].

5. Results

Table 1 lists the promoters as aligned by Hawley and McClure [5]. The sequences underlined correspond to the regions where the bend is localized (obtained by taking five bases from either side of the minima). The model has been successful in predicting the experimental results for other systems [30–33], while the promoters analyzed here have not been subject to any such experimental investigations and hence the theoretical predictions could not be tested. The third column corresponds to the point of minima and is given with respect to the start site. It is very interesting to note that almost all of the regions corresponding to the minima have bases that lie within the -10 and the -35 region. It has also been noted as a general rule that the mutations that decrease promoter's agreement with the consensus, decrease the frequency of transcription initiation, and the mutations that increase consensus agreement increase initiation frequency [5] and exception to this rule is infrequent. Since the minima obtained in almost all the cases lie within this region we would like to suggest that a mutation in this region may be changing the curvature dispersion which in turn affects the strength of the promoter. In order to further confirm this assumption each of the 112 promoters was point mutated and the changes in the curvature dispersion were noted. It is very evident from the graphs shown that the changes in the curvature dispersion due to a single base change are very significant and as has been reported earlier that the mutation in this region affects the strength of the promoter. We have not shown all the 448 graphs corresponding to the A, T, G, C base pair substitution for obvious reasons. The graph shown (Fig. 3) correspond to the A, T, G, C base pair substitution in case of the *araBad* promoter. Signifi-

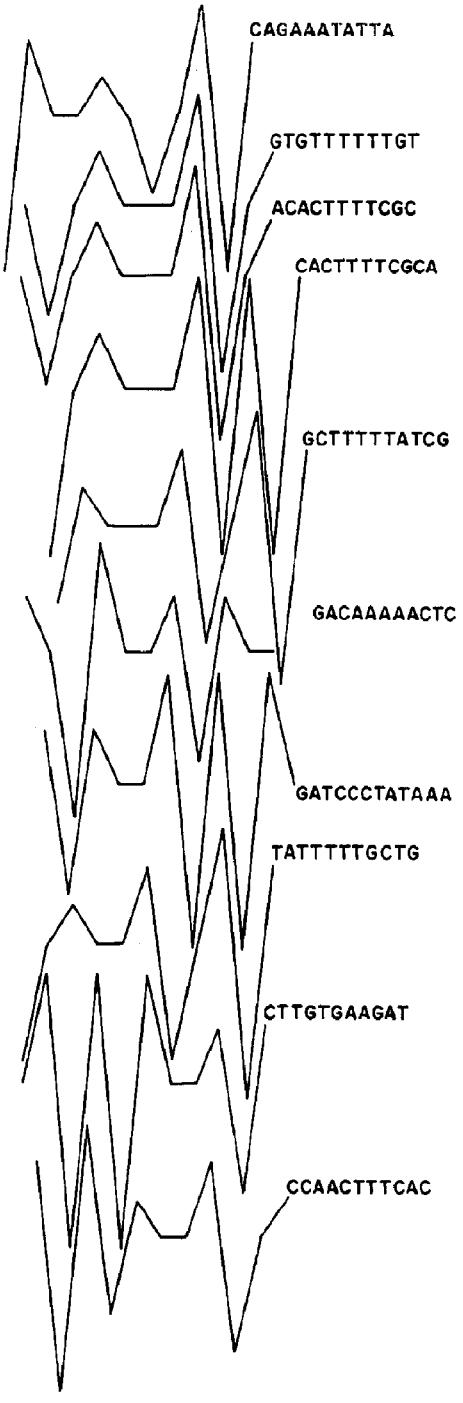

Fig. 4. Graphs for the Σ_1 for the following promoters: *uvrB P2*, *bioA*, *galP2*, *galP1*, *araBAD*, *tnaA*, *rrndP1*, *λPO 434 PRM*, *Cat*.

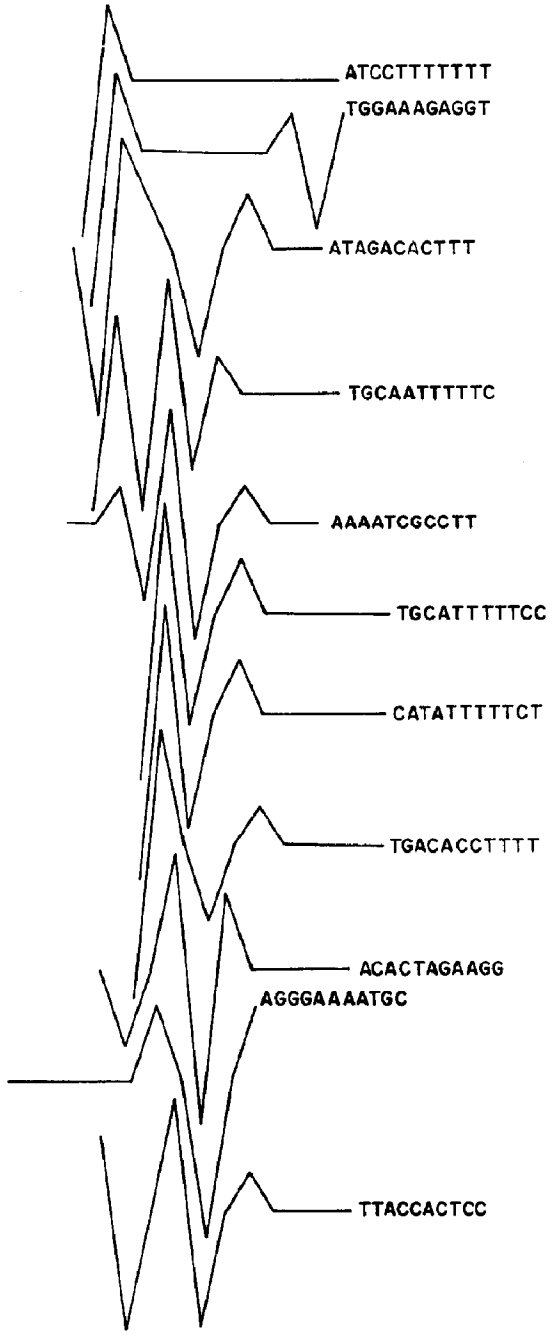

Fig. 5. Graphs for the Σ_1 for the following promoters: *pBR* primer, *malEFG*, *araC*, *rrne P1*, *lexA*, *rrnX P1*, *rpoA*, *str*, *pBR RNAI*, *Cole1 PI*, *Tn10teta*.

cant changes in curvature dispersion were also noted for the other promoters.

The promoter helical structure variation at the *E. coli* polymerase interaction site was reported

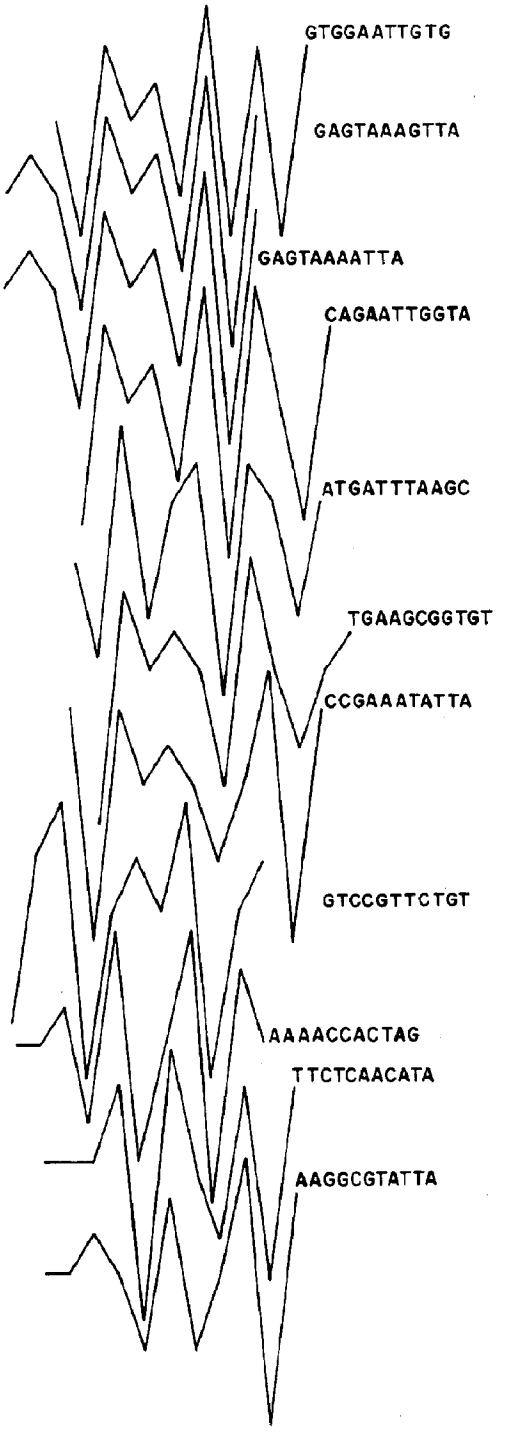

Fig. 6. Graphs for the Σ_1 for the following promoters: *lacP1*, *spot42*, *uvrB P1*, *Tn10Pout*, *malK*, *spc*, *GLTNPR*, *rpoB*, *leul tRNA*, *lpp*, *RRNG P2*.

by Ruth Nossinov [47] wherein he had evaluated the four sum functions for the three different consensus regions. Here we report the helical structure variations at the regions localized by the cyclic permutation of the sequence. The structural similarity among the promoters is displayed in Figs. 4, 5 and 6 which show the variation in twist (Σ_1) along the sequence. The other sum functions also upholds the similarities (not shown).

Most of the promoters analyzed here show the minima between −10 and −35 region. The analysis of the sequence at the minima reveals that there exist an homology among these sequences irrespective of the position of minima. The Calladine–Dickerson rules stress the importance of base pair steps and not of isolated base pairs. It is necessary to realize that the presence of a consensus sequence does not necessarily guarantee homologous structure. It is gratifying to show that apart from the structural similarity that exist in the consensus sequence [47] the regions that we have localized also show significant commonality in structure which is evident from the (Σ_1) function plot. We can therefore group the promoters on the basis of common structural features and advocate the notion of "consensus structure" suggesting their common biological significance.

Finally we can draw the following conclusions based on the studies carried out on the promoters using the theoretical model. There seems to exist some structural commonalities among the promoters which is recognized by the RNA polymerase. The variation in the structure can account for the varying strength of the promoters. The difference in the promoter strength due to mutations is evident from the variation in curvature dispersion. Further we hope that the current study will be useful in designing experiments that will help in defining the promoter location more precisely and also help in determining the promoter function.

Acknowledgements

We thank Prof. P. De Santis for his help in understanding the model. Thanks are also due to Jayanta K. Bandyopadhyay and Dr. Shirish Ranade for useful discussions. This work was supported by the Council of Scientific and Industrial Research, India.

References

1 D. Pribnow, J. Mol. Biol. 99 (1975) 419.
2 M. Takanami, K. Sugimoto, H. Sugisaki and T. Okamoto, Nature 260 (1976) 297.
3 H. Schaller, C. Gray and K. Herrman, Proc. Natl. Acad. Sci. USA 72 (1975) 737.
4 P.H. Seeburg, C. Nusslein and H. Schaller, Eur. J. Biochem. 74 (1977) 107.
5 D.K. Hawley and W.R. Mc Clure, Nucl. Acids Res. 11 (1983) 2237.
6 P. Youderian, S. Bouviers and M.M. Sussbind, Cell 50 (1982) 843.
7 W. Mandecki and W.S. Reznikoff, Nucl. Acids Res. 10 (1982) 903.
8 J.E. Stefano and J.D. Gralla, Proc. Natl. Acad. Sci. USA 79 (1982) 1069.
9 D.R. Russel and G.N. Bennette, Gene 20 (1982) 231.
10 T. Aoyama, M. Takanami, E. Ohtsuka, Y. Yaniyama, R. Marumoto, H. Sato and M. Ikehara, Nucl. Acids Res. 11 (1983) 5855.
11 M. Mulligan, D. Hawley, R. Entriken and W. McClure, Nucl. Acids Res. 12 (1984) 789.
12 R. Staden, Nucl. Acids Res. 12 (1984) 505.
13 L. Bossi and D.M. Smith, Cell 39 (1984) 643.
14 R.L. Gourse, H.A. de Boer and M. Nomura, Cell 44 (1986) 197.
15 P.J. Hagerman, Biochemistry 24 (1985) 7033.
16 S. Diekmann, FEBS Letters 195 (1986) 53.
17 H.S. Koo, H.M. Wu and D.M. Crothers, Nature 320 (1986) 501.
18 E.N. Trifonov, Nucl. Acids. Res. 8 (1980) 4041.
19 J.C. Marini, S.D. Levene, D.M. Crothers and P.T. Euglund, Proc. Natl. Acad. Sci USA. 79 (1982) 7664.
20 E.N. Trifonov, Cold Spring Harbor Symp. Quant. Biol. 47 (1983) 271.
21 E.N. Trifonov, CRC Crit. Rev. Biochem. 19 (1985) 89.
22 L.E. Ulanovsky and E.N. Trifonov, Nature 326 (1987) 720.
23 A.A. Travers, Annu. Rev. Biochem 58 (1989) 427.
24 D.M. Crothers, T.E. Haran and J.G. Nadeau, J. Biol. Chem. 265 (1990) 7093.
25 P.J. Hagerman, Annu. Rev. Biochem 59 (1990) 755.
26 J. Ramstein and R. Larvey, Proc. Natl. Acad. Sci. USA 85 (1990) 7231.
27 T. Nishi and S. Itoh, Gene 44 (1986) 29.
28 R.R. Plaskon and R.M. Wartell, Nucl. Acids. Res. 15 (1987) 785.
29 B.F. Baner, E.G. Kar, R.M. Elford and W.M. Holmes, Gene 63 (1988) 123.

30 P. De Santis, A. Palleschi, M. Savino, and A. Scipioni, Biophys. Chem. 32 (1988) 305.
31 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biochemistry 29 (1990) 9269.
32 P. De Santis, S. Morositti, A. Palleschi and M. Savino, in: Structures and dynamics of nucleic acids, proteins and membranes, eds. E. Clementi and S. Chin (Plenum Press, New York 1986) p. 31.
33 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biophys. Chem. 42 (1992) 147.
34 C.R. Calladine, J. Mol. Biol. 161 (1982) 343.
35 R.E. Dickerson, J. Mol. Biol., 166 (1983) 419.
36 H.M. Wu and D.M. Crothers, Nature 308 (1984) 509.
37 D.L. Milton and R.F. Gesteland, Nucl. Acids Res. 16 (1988) 3931.
38 K. Inokuchi, A. Nakayama and F. Hishinuma, Nucl. Acid Res. 16 (1988) 6693.
39 K. Zahn and F.R. Blattner, Nature 317 (1985) 451.
40 D. Boffelli, P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Int. J. Quant. Chem. 42 (1992) 1409.
41 R.E. Dickerson and H.R. Drew, J. Mol. Biol. 149 (1981) 761.
42 R.E. Dickerson and H.R. Drew, Proc. Natl. Acad. Sci USA 78 (1981) 7318.
43 H.R. Drew and R.E. Dickerson, J. Mol. Biol. 151 (1981) 535.
44 H.R. Drew, T. Takono, S. Tanaka, K. Itakura and R.E. Dickerson, Nature (London) 286 (1980) 567.
45 H.R. Drew, R.M. Wing, T. Takono, C. Broko, S. Tanaka, K. Itakura and R.E. Dickerson, Proc. Natl. Acad. Sci. USA 78 (1981) 2179.
46 D.L. Milton, M.L. Casper and R.F. Gesteland, J. Mol. Biol. 213 (1990) 135.
47 R. Nussinov, J. Biol. Chem. 259 (1984) 6798.