Nucleic Acids Research, 1995, Vol. 23, No. 13 2351–2360

Compilation and analysis of *Bacillus subtilis* σ^A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA

John D. Helmann

Section of Microbiology, Cornell University, Ithaca, NY 14853-8101, USA

Received March 3, 1995; Revised and Accepted May 24, 1995

ABSTRACT

Sequence analysis of 236 promoters recognized by the *Bacillus subtilis* σ^A-RNA polymerase reveals an extended promoter structure. The most highly conserved bases include the –35 and –10 hexanucleotide core elements and a TG dinucleotide at position –15,–14. In addition, several weakly conserved A and T residues are present upstream of the –35 region. Analysis of dinucleotide composition reveals A_2- and T_2-rich sequences in the upstream promoter region (–36 to –70) which are phased with the DNA helix: A_n tracts are common near –43, –54 and –65; T_n tracts predominate at the intervening positions. When compared with larger regions of the genome, upstream promoter regions have an excess of A_n and T_n sequences for $n > 4$. These data indicate that an RNA polymerase binding site affects DNA sequence as far upstream as –70. This sequence conservation is discussed in light of recent evidence that the α subunits of the polymerase core bind DNA and that the promoter may wrap around RNA polymerase.

INTRODUCTION

To define the DNA sequence features associated with promoter recognition by *Bacillus subtilis* RNA polymerase (RNAP), I have analyzed a compilation of 236 σ^A-dependent promoters (Table 1; refs 1–190). As expected, these studies confirm the presence of highly conserved –35 and –10 hexamers (positions are relative to the transcriptional start site), but they also reveal several more subtle features of promoter structure. In *Escherichia coli* both sequence comparisons (191–193) and genetic studies (194,195) indicate that the –35 and –10 hexamers contain the bases most critical for promoter function. However, bases outside these classically defined core elements are also important: upstream regions may enhance promoter activity by binding the α subunits of RNAP (196–198) or by facilitating DNA bending (199–201), while downstream sequences can affect promoter clearance (202–204).

RNAP from diverse bacteria recognize the same set of strong phage T7 promoters (205) which suggests that the TTGACA (–35) and TATAAT (–10) consensus elements, as defined for *E.coli* RNAP, are conserved features of eubacterial promoters. However, promoter strength cannot be inferred from sequence inspection alone (195,206). For example, the *E.coli lacUV5* promoter is used very poorly by *B.subtilis* RNAP, despite its close fit to consensus (207). Previous studies identified sequence elements, including the dinucleotide TG (at –15,–14) and an A-rich region near –43, which are conserved in promoters from gram positive bacteria but not from *E.coli* (208,209). Both of these elements contribute to promoter function: mutations which introduce a TG dinucleotide upstream of the *E.coli lacUV5* –10 element have a much stronger effect on transcription in *B.subtilis* than in *E.coli* (207) and upstream A-rich regions stimulate transcription both *in vivo* and *in vitro* (196,210–211). In the present study I have aligned 142 known and 94 putative σ^A-dependent promoters from *B.subtilis* to analyze the extent to which these and other sequence features are conserved.

MATERIALS AND METHODS

Promoter sequence alignments

All computer analysis was performed on a Power Macintosh 6100/60 computer using readily available software packages (Microsoft Word, 5.1, DNA Strider 1.2 and Deltagraph Professional). Promoter sequences were obtained from GenBank (212) via the National Institutes of Health gopher server (gopher.nih.gov). Annotations indicating promoter regions were noted and the relevant DNA sequences were copied into DNA Strider 1.2, to remove extraneous characters, and then into Microsoft Word to generate Table 1. References, either cited in GenBank or found by bibliographic database searches, were consulted to verify start sites and the method of transcript mapping. Sequences, aligned by their –10 and –35 regions, were sorted into two groups: (i) 125 chromosomal and 17 strong phage promoters (Table 1A) supported by experimental transcript mapping; (ii) 94 putative σ^A-dependent promoters (Table 1B) for which supporting experimental data were not found. Table 1 is available through the World Wide Web (URL http://www.bio.cornell.edu/microbio/helmann/helmann.html) under the heading of 'papers and publications' or from the author via e-mail (jdh9@cornell.edu).

Analysis of base frequencies

To determine base frequencies (percent of total) at each position, single character columns from Table 1 (from –100 to +15) were copied into DNA Strider as 'protein' sequences and enumerated

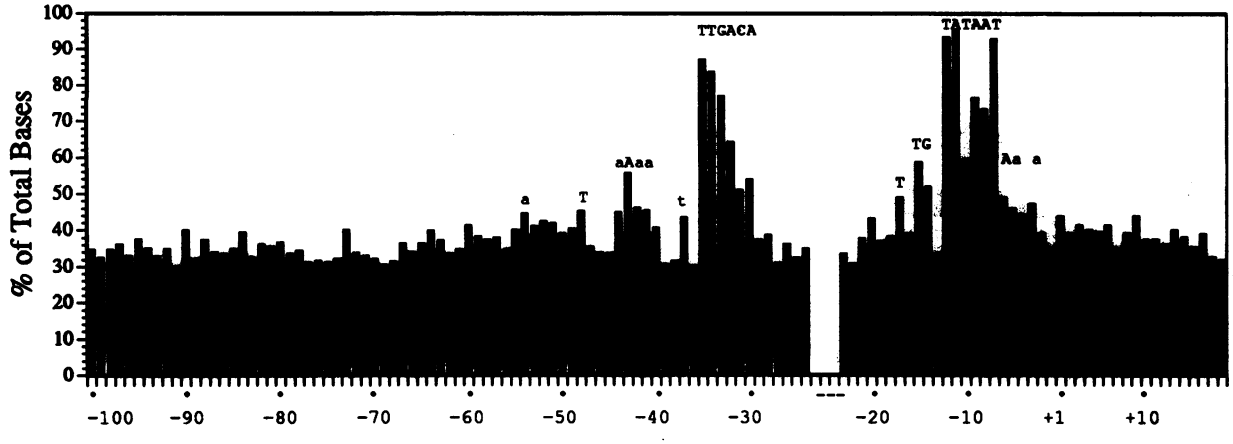

Figure 1. Base conservation within the 236 aligned promoter sequences. The abundance of the most frequent base is plotted as a function of position. Those bases found at a given position >3 SD above the expected occurrence are in capital letters and those occurring between 2 and 3 SD above the expected occurrence are in lower case letters.

with the 'amino acid analysis' function. These data were copied into Deltagraph Professional and the highest value at each position was used to generate Figure 1. The 236 aligned sequences (Table 1) have an overall base composition of 35% A, 15% C, 18% G and 32% T. Therefore, these promoter regions are significantly more AT-rich (67%) than the *Bacillus* genome (~ 57%). Poisson statistics were applied as described previously (213) to assess the statistical significance of conserved bases. Briefly, the expected number of occurrences (and the corresponding standard deviation, SD) for each base was calculated by multiplying the base frequencies noted above by 236 and determining the corresponding square root (1 SD).

Analysis of dinucleotide frequencies

To analyze the frequencies of all 16 DNA dinucleotides as a function of sequence position the data of Table 1 were converted into a 16 letter code. First, Table 1 was stripped of all unnecessary text and the global replace command of Microsoft Word was used to substitute each character with the same character followed by a comma (e.g. G → G,). Every other column of commas was deleted to generate a text file of the form AA,CC,GG,TT. This text file was duplicated and, for the first copy, 16 global replace steps allowed all the dinucleotides (beginning at every other position) to be replaced with a one letter code. For example, every occurrence of 'AA,' was coded as 'D,' and 'CC,' was coded as 'P,' and so forth. This procedure was repeated on the second file, except every occurrence of 'A,A' was coded as 'D,' and 'C,C' was encoded as 'P,' and so forth. Each resulting file was imported into Deltagraph Professional (as comma-delimited text) to generate two 56 × 236 matrices. Dinucleotide frequencies at each position were enumerated using DNA Strider, as described above, and the two resulting matrices were combined into a single 112 × 16 matrix representing the frequency (%) for each dinucleotide at positions between –100 and +13.

Analysis of oligo(dA) and oligo(dT) tracts

The occurrence of A_n and T_n tracts upstream of the –35 region (e.g. bases –36 to –80) was tabulated using Microsoft Word and the global replace function. A_n and T_n tracts were replaced in descending order beginning with A_{12} and T_{12}. After each replacement the number of text substitutions, corresponding to the number of words of length n, was noted. This procedure only enumerates words of the form BA_nB and VT_nV (where B designates 'not A' and V designates 'not T'). To calculate the expected number of occurrences, the frequencies of A and T within the region analyzed were used (e.g. A = 0.345 and T = 0.332 within the –36 to –80 interval). For example, the expected frequency for BA_6B equals $(1 - 0.345) \times (0.345)^6 \times (1 - 0.345)$ or 7.23×10^{-4}. Although 5978 bases were searched for this pattern, these bases were divided into 142 segments of length ≤44 (–36 to –80). Therefore, a correction for end effects has been employed (214). The actual number of positions at which the pattern A_6 could have initiated is $5978 - [142 \times (n - 1)]$, where $n = 6$ (therefore, 5268 nt). As a control, the expected and observed numbers of A_n and T_n tracts were obtained for several 30 kb segments from the *B.subtilis* genome (bases 1–30 kb and 150–180 kb from D26185 and bases 1–28 206 from L09228). The results for the three segments were similar and the data in Figure 5 are from the 150–180 kb region of sequence D26185. For Figure 6 the frequency of A_n and T_n sequences ($n > 4$) per kilobase of sequence were tabulated for five different 30 kb regions (two from D26185 and one each from L09228, D32216 and X73124) and the average and SD were determined. These values were compared with the upstream promoter DNA analyzed in overlapping 20 bp windows. In this case, the SD was calculated as the square root of the observed number of A_n and T_n sequences within the 20 bp window (the range was 23–70).

RESULTS AND DISCUSSION

Table 1 contains 236 σ^A-dependent promoters from *B.subtilis* aligned based on known start sites, when available, and optimized for the best match to the highly conserved –35 and –10 regions. DNA sequences extending to –100 relative to the transcriptional start site have been included to detect any sequence conservation in the upstream region. Recent data indicate that RNAP can contact promoter DNA to at least –60 (198) or even –70 (196). In addition, the upstream DNA provides a sample of the statistical background important for evaluating conserved sequence features.

Table 1. Compilation of *B.subtilis* σ^A-dependent promoter elements

GENE	SEQUENCE -100 -90 -80 -70 -60 -50 -40 -30 -20 -10 +1 +15	SL	SS	REF
(A)	**PROMOTERS WITH BIOCHEMICAL OR GENETIC SUPPORTING DATA**			
abrB	TTCGGTAGTTTCCAAGACATTACTGACTATAAGAACTAATTCTTACAATCAATAGTAAACAAAATGATTGACGATTATT---GGAAACCTTGTTATGCTATGAAGgTAAGGATTTTGTCGAAT	17	PE	(1)
acka	AACCTATAGTGAATGTGTCTGAAAATAACGACTTCTTATTGTAAGCGTTATCAATACGCAAGTTGACTTGAAAAGCCGA---CATGACAATGTTTAAATGGAAAAgTCAGATATTTTTCGGAG	17	PE	(2)
acsA	AAAATACATGCCTTCTTTAAGAATACCAAATTATAAAGCGTTTTCAACATGGTTTATATTTAAAAATTGAGAAGAATA---TGAATATATACTATAATAATTGTgACAACTTCAGCAAAGGG	17	PE	(3)
acuA	GCTGAAGTTGTCACAATTATTATAGTATATATTCATATTCTTCTCAATTTTTTAAAATATAAACCATGTCCGGGT----TATAATTTGGTATTCTTAAAAGAaGGCATGTATTTTTGATA	16	PE	(3)
ada	TTGAAATTAGTTTATCACATAATCATCTTTACTCATTTTGTTATCTTGCTATTACATTCTAACCACCTTATTTTTTCT---ATTTATGAGGTTATAGTGTAGTTaTCAAGAATGCTAAACGG	17	PE	(4)
addA	GGGAGAGATGTGCGGAGATAATCAGCTTTTTATATGTGAAAAGGCCGTTTTTACCAATAGATCAGATTGGTCATTTTCG---TCAACATTCGATAAAATATAGaGAGATAAAAAGAGAGGGGT	17	PE	(5)
ald	AAACCTTCCGGCACATGGATTTGTGAAATTTCACAAATCCATGTTTTTTTATTTTCTTAATCAAACAAAGAATTTTCCA---AAATATCAAGCTACACTAAAAATaTCACATATACAGGAGGA	17	PE	(6)
alkA	TAACTACACTATAACCTCATAAATAGAAAAAAATAAGGTGGTTAGAATGTAATAGCAAGATAACAAAATGAGTAAAGAT----GATTATGTGATAAACTAATTTCaACCAGCGTGAGTTTTCT	16	PE	(4)
alsS	TCAATATGCATTCCTTTCCATAGGTTAATAATTCGTATTACATATTAATCATAAGGTCGAATCGATATTGGAAGGTCAAT--TTCCAAAGAGTGTATAGTGAAACTTaTCACAAGATATTTAAA	18	PE	(7)
amyLY	AGGGAAGCGTTCACAGTTTCGGGCAGCTTTTTTTATAGGAACATTGATTTGTATTCACTCTGCCAAGTTGTTTTGATAG---AGTGATTGTGATAATTTTAAATGTaAGCGTTAACAAAATTC	17	S1	(8)
ansA	TTTTTCACCTTGAGCTGATTATATCATATAGGTATTAGGAAGTGTACGCGAATTGAATAATAATGTCTTGCGTATTTTA---TTTTCCGTGTCTATAATTGCATtAAAAGTATGCGAAAAGCG	17	PE	(9)
aprE	GAGTCTCTACGGAAATAGCGAGAGATGATATACCTAAATAGAGATAAAATCATCTCAAAAAAATGGGTCTACTAAAATA-TTATTCCATCTATTACAATAAATTCaCAGAATAGTCTTTTAAG	19	PE	(10)
argC	GCTGTCAAACATGAGAATTCCCATTGAACAGAAAATTTATAAATGAATAAAAATATTAAATATACGATTGAATTAATTT----TTATTCATGTTATAATGTTAAATaATTTCACAAAGACCAA	16	PE	(11,12)
asd	TCTAGTTGAGAATTATCAGGGAAATGACTAAATTTACAGAAAATCTTTCCCAAACTAAAAAATTCGGTCATCACCCGCA----TATTCTATGGTAAAATAAAACGtAAAATCATAGTCAAATC	16	PE	(13)
cdd	TTATTTTTTTGCCAAAGCTGTAATGGCTGAAAATTCTTACATTTATTTTACATTTTTAGAAATGGGCGTGAAAAAAAGC---GCGCGATTATGTAAAATATAAaGTGATAGCGGTACCATTAT	17	S1	(14,15)
citA	GAAGCCATTTGAAATCCATTTCTATTCTCCCTCTGATTAATATTTTTAATTAATTCCCTTTAAAATATTGATTATTTTT---TAAATATTATATTTACTATAATAaCAGAAAAGGATAGGGGG	17	PE	(16)
citB	ACAGTAAAAATTACCGGTACATTTTTCTCATAAGTCGAACTTATTGTATTTAATAAAAACATTGATTTTACTATATGT--ATGATTTTGTTTTAATATGAAATTgTGAGAAAATTGTGATGG	18	RN	(17,18)
citC	CCGATTGAAGAAAGAGCCTAAAGAACCATTGGAGGCTGGCTTGCGTCAATGAGGTTAGCCTCTTGTTCAGACATCAAAA---TTGGGTTACACTTTAAATTGAATgTTAGGAAAAATCATTTT	17	PE	(16)
citR	TGGATGTTTCTACACATGTAATTCCCTTTAATCCGTAATGTACCATTTGTATTCCCCCTATCCTTTTCTGTTATTATAG---TAAATATAATATTTAAAAAATAaTCAATATTTTAAAGGGAA	17	PE	(16)
citZ	AACCCGCAAATAAAGGCGTTTTTTCGGGGAAATCACAGAATGGAAAAAAAAAATTATAGGTAAACATTAACAAAATGTC---TGATTTATTGTTATAATGAGAATaGGCTTCAAACTTAAATAA	17	PE	(16)
comA	AAGTGAAGGAAAGGGCTTTAAGGCTGATATTGAAATCGAATTGTAATGGATTTATAACGGAAACGACTTGGCACAGGCC----AAGTCTTTTTTATAAAATGGAaaAGAGTGAGTAAAAGGGA	18	S1	(19)
comB	TAAAGAAAAACTGGCCTTTCATTCGCGAAGGGCCAGTTTTCTTTATGTATATATTATTTTTCATGTTTAGACAATTTTCGTCAAATTATTTGATATACTTAGGGGTgAAAGCCGCGCGTATTG	20	PE	(19)
comC	TCCGGCTCCGGCAGAATCAAAAAAAGATTCTGCCGTTTTTTTCATGTGTAAAATTGAATATTCCCAATTGGAGCTCAGG-ATGTCCATTTTGCTACAATCCATACCAgAACTCAATGAAGTAAA	19	PE	(20,21)
comE	CAAGGCGCAGAAATCAAAAACCATCGTTTCCTAAAACGATGGTTTTTTATAAAATGCTTTTTTATGCTTTTGCAGTACAGA---CGAACGTATGACATACTCGTCTaCACATGAAACTGCTTTTT	17	PE	(22)
comF	GTGTTTTAAATAAGGCCAAATCTCCGTTTTTAGAGCGGAGATTTTTTTATATTCTTATTTTAATAGTTGGACAGAAAAT---ATTCATTCAGGCATACTGTTTCgAAAGGAGGCGTGCTATGT	17	PE	(23)
comG	CTTTTTTCTTGCCAGAAAGAATTGGTTTTTCAGCATATAACATCTCACAAAATCACGTTTCCCTGTTGAATTACCTTT---TCTTCTTTTTCTACAATATGCGTTgAAAGGAGAGGGAATCA	17	PE	(24)
comK	TATTACTAGTCATTTAGTACCATTAAATATCATTAAAAGATGATTTTATCTTAAATGTTAAAAAAACCTGTCGTTTTAC--AAAAACAGATGATAGATTATTAgTATAAATTTTGCAGAAAAA	18	PE	(25)
comQ	TTTTTTCGTTCTACCGATACAATAAATGGATAAAGTATTATATGATTGTTAAAAACGAAAACCTGCTGTCCTTTAAA--TGTCCCATTTAGTAAAATGGAATgGGAGGGGGGAAGTCGTT	18	PE	(26)
cspB	TTCCGTAAATAAATATCAAATTGAAAAATAAATATGGTATTTTCCGGAAAAAAATTTCAATAAGCAGTTGTTTTTTCTG---AAGATTACTGGTGGAGTAAAGGTaATTATTTTTGTTCGAAC	17	PE	(27)
ctaA	TCCAGAAGCCGTCATTCTATTATATTTTTGTGAACAAAAGGCTCTGGGAATTGCCACAAAACACATATTCGCTTACACT---TGGAACGTATATATAAAATTGCaGCAGTATGTTAAGAAGGTGA	17	S1	(28)
dciA	CAAACATGAATCCTTGAAAGAGGATTCTTTTTTTATCACTGAATGATTGGAAGTTTCCCAGTTATATTGCATTTTTCC---TCTTTTTTTAATAATAATTTGTTaGAATATTCATAATTTAGT	17	PE	(29)
degQ	ATTCCTTTGAGTATAAGGACCTATTGAGATTTGCGGTGTCACGCAGGACTTTTTTGCATACTTTTCGGTGAAAAATGAG---CCGAAAGCAGACACACTATTAgTAACAGATCAAATACCTAG	17	S1	(30)
degSU	TGTGCAGCATGCTAGCTGACCCTCCTGCTAAGCATAAAAGACTGCCTATACAAATTCGTACAGTCTTTAGAATTTTTGT---GCGTATTTTGGTATCATAAAGAGTAGaTAGTATATAAAAAT	17	PE	(31)
divIB	GTGAAAAATACGAGATTGATATGCACACACAGAGGTTGAAATCATCGGCGGAAATCGCTGATTCAAGTTCTGACTGAAGCT----GTTCATATGATATACTGTAAGCaAACGACAAACGGCATCA	16	PE	(32)
dnaA	TACCAACCCCTGTGGACAAGGTTTTTTCAACAGGTTGTCCGCTTTGTGGATAAGATTGTGACAACCATTGCAAGCTCTC---GTTTATTTTGGTATTATATTTGTgTTTTAACTCTTGATTAC	17	S1	(33)
dnaE P2	AACTCAGGACGCTCTATCCTGGGTTTTTGGCTGTCGCCAAAAGGGAATAATGAAAAACAATAGCATCTTTG--TGAAGTTTGTATTATAATAAAAAATTgTGATAAAATGATAAT	18	S1	(34)
dnaG	ACAGTCTGTCCACATGTGGATAGGCTGTGTTTCCTGTCTTTTTCACAACTTATCCACAAATCCACAGGCCCTACTATTACTTCTACTATTTTTTATAAATATATaTATTAATACATTATCCGT	—	S1	(33)
dnaK	TTATTTGACTCATCAAGGAAATTATTAGGCAATGAAGTTTTTGGCGCTTTTTTTGGGTGAGTTATAATTGACATTTTTC---TTGTGGTTTGATACTTTTGTTATaGAATTAGCACTCGCTTA	17	PE	(35)
ffhrpsp	CGGAATAAATTGGATGATATTTAGCGTATTTTGGAAAAGTTAATCGCCGCTTTGACAAGATAAAACCTTGACAGATGTCA---TTAAAACCGTGTAAACTCAAGTTATCaGTAAAGGATTTGACT	17	PE	(36)
ftsA P1	AAATGTGAAAAGCACATAAAAATATTCTGTTGTTATTTTTTGTTACACACTTGTAAAGCCACATTCATTGTATTGTTGT---TCCGCAAATAATAGAAATgATCGAAATGTGAGGAGG	17	PE	(37)
ftsA P3	GTACATATGATGAAATGGTATTTGTTTCCGGTTTCTTTTTTAATATTATATTGGCAATAAGTTTAGCTTTTCTGGGAGT---CCATCTTGGTGTAGACTTGTATTTAGCAGGTATATTCGCAT	17	PE	(37)
gcaD(tms)	TTTTTAATTCTGATTTTTCAAACTTAGTTGCACTCATAAGAAAATTCTTGCACTTCATGAAGTCTCCTTGAAATCAGAA---GATATTTAGGATATATTTTTCTATgGATAAAAGGGATATTG	17	RN	(17,38,39)
glnR	GAAAAGCAATTAATTTAAATTTTTTTAAAATTTCACTGAATTTGATGTTAAGAATCCTTACATCGTATTGACACAAAAT---ATAACATCACCTATAATGAAACTAaGTTAAGAAAAGGAGGA	17	RN	(40)
glpD	ATTGGAAGAAGAAAACGAAATGAGCTGTATAAAGCTGCTGGCAAAAAGCCGTGAAAGCAGCTATGGCTTTAAAATAAAGT----AATACTATGGTATAATGGTTACaAGTTAATAAGAACGGTC	16	PE	(41)
glpFK	CTGTAGCTGTCACAACATCTAATACCAAATTGTGGAAAAAATATGAAAACTTTTTGACGGAAAGTGATTGACACCGCTT---TCATGCACTGATACAATTGCACTaGGTTAATACATTGTGAT	17	PE	(42,43)
glpTQ	AAAATATAATTGCATTTAGTAAAAGGATAATATTCACTTCCCTAAGGCCCAATTTGAAAATTACCCGTTTACATTCGCT---CCGGACTATGATAATTTAAGAAAGCGCTATCATGATAAG	17	S1	(44)
gltA	GCGCACGTCCATGTTTGTCTCACATCCATCTATCTCATTTTGAGATTCTTTTGATCTAAATTATATATTGTTTATATCG---TTTTGAAAACCTACAATGATTATaGAGTTGTTACaAGTTAATAAGGTTC	17	S1	(45)
gltC	TCCTCTCCCCCGATCAATTTCCGATAATACCGGTCATAAAATCTAACAACTCTATAATCATTGTAGGTTTTCAAAACGA---TATAAACAATATATAATTTAGATCaaAAGAATCTCAAAATG	17	S1	(45)
gltX	ATTTACAGGCCGGGCGGAAGGAATAGCGGCTCAGGCGACAGTACTGATACAAAAAGGCTAACTTGTTTGATGCCGGGT---CTATTTGGTGGTAGAAATAGATTCaTACATTTTTGCCTGA	17	S1	(46)
gluB	CTTATTCAATAGAGATAGAGCAAATTGACAGGCTTTTACACCTCCCAAAAACAAGAAATTAGGTTGATAGACAATCATG---AGAAAGATTTTTACAATGAGTTCgaGCTCATAAGAAAGAGC	17	S1	(47)
gntR	TAATTGATCTGGAAATACATACCATGCAATATGGTAAAAATTTAAATAAAAATTAGAAATGAAGTGTTTGCATAAAAGA---AATATTCACGTTATCATACTTGTATaCAAGTATACTCCTTG	17	S1	(48)
groESL	GAAAAAGCTAACGGAAAAGGGAGCGGAAAAGAATGATGTAAGCGTGAAAAATTTTTATCTTATCACTTGAAATTGGAA---GGGAGATTCTTTATTATAAGAATTgTGTTAGCACTCTTTAG	17	PE	(49)
gsiA	GGTGTTTGAAAAACCCGTGCTTTTTTGTTGCGGTTAGCCGAAATTCGACAATTGCGGTTATTTTGCGTTCTTCTTTTTC--TTGTAAATATGATAAAAATATGACaTaTCTCGGGTAATTCAAA	18	PE	(50)
gsiB	GAAAGCAGACGGACACCGCGATCCGCCTGCTTTTTTTAGTGGAAACATACCCAATGTGTTTTGTTTGTTTAAAAGAATT---GTGAGCGGGAATACAACAACCAaCACCAATTAAAGGAGGAA	17	PE	(50)
guaA	AAGATTCTTCGGCGCTATGGAATGATTCGAGAGAGTTATGGGAAAATTTAATTAAAAGAAGATGGTCTTGACCGCTTAT---CGACGTGTTGTTAGAATTAGTGAaTATTATCGGAGTCTGGG	17	PE	(51)
gutB	AGCGTTTTTTGTATATGAAAAAATAAAAGTACAGTGCCGCTGTCCTTTTATACAGCAGGAAAGGCTGTTGAACGTGTTA-----AAAAGCAGATAAAATGGGGGCaGTAACAGAGAAAACAAA	15	PE	(52)
gyrB	AGGCTTTTTTCTGAACAAGATCAGATGCAGCGATGCCTGCAATACCTATATATTCTAGCAATTTAATGTGTATAATCA--TAAGTTTATTGATAATAGGAGAatagTGAAATCGTATTGAA	—	S1	(53)
hbs	AATGCTTGATATGGCTT---TTTATATGTGGTTACTCTACATACaGAAATTCTTCACTTTGT	17	PE	(54)
hemA	ATTGTCAATAGGAATGCTTCTTTTCCCTGAAGCTTTTTCTAATATAGCATAAGAATTTTAAAATCTGTTCACATTTTGTGAAAGAAACTATGTTATAATTATTATAAATAATGAGTTCTATGT	20	PE	(55)
hsmB1	AAAGAACATGTCTAGACTATCTCTAATTGTGATAAACCCTGCTGAAAAGTACCAGCAGGGTTTATTGTGTTAATCA-TTACAAAAAATGATAAAATAAAAAAGcAAGACGTATCAATATT	19	PE	(56)
hut	TCTTCTGCTTTCTGCGTCACGCTATTACAATAGCAATCTAACTTGTTAAAGACTATAAAAAAACCTTTGACTTCTGCT---GCTGAACCAATTAATATAATACTCaGTTAATAGTTATCAGA	17	PE	(57)
ilvB	CAATAATGAAAGCGTATACAATATAGATTGATTAATCAAAATTGTCTAATAATTTTAAAAAATGCTGTTGACACTGCGT---CCAAAGCGGCGTAATATGAGTTCaaCAAAAGATAAATGCAA	17	PE	(58)
kinC	TTACCGCCGCCTAAGATAAGACATCAAGATATTTGGAAGTTGAATGATTTGGGATACTTTACATATTTTACTCAATTATT-TGTCGAAGAATGGTACAATAAGTTAGaGAAACACAAGCGGCAGG	19	PE	(59)
lon S1	TCAATAAAGAATCCGTTTAAACCCTATTTTGATAATAGGGTTTTTTTCATGAAGGACATCTTTTCCTTTTCATATCAG----AAAGAAAGGGTATACTACGAGGAgACTGTTTTATAGAAAA	16	PE	(60)
lon S2	CATATCAGAAAGAAAGGGTATACTACGAGGAGACTGTTTTATAGAAAATGGAACCATGCCGTTTATTGTACAAGGATG---AAAAAGTTGTATTAATGGTTCATaCTAAAGTCACGGAGGT	18	PE	(60)
lysC	GACATCAAAAAAGCCGGTGTTCCGCAGCGGCGGCTCAGCTCCTTTACGACAAATTGCAAAAATAATGTTGTCCTTTTAA---ATAAGATCTGATAAAATGTGAACTaaTTTCATAGTTAGATC	17	PE	(61)
mecA	TGTGACTGTTTTATCATAAAATAGAAATACAAAGGAATTCACACTGGCCTTGGTTAAGGTTAAGATGTGGACGGAATGG---GTAAAGTGTAGTAAAGTACAATTAaTCGGGAGCTTAGATGT	17	PE	(62)
men	AAGCTTGTTCTTTTTCATTTCTG----AAATTAGGTTTATAATAGGTAAgCGAGCCATTTGGACTG	16	PE	(63)
menB	TGGAGCAGCCGCGATTGGTTTGGTAAAATAGTGAGTGAGTTTTGACAAGCATCTGACTCATTCACATAGAGAATAAAA---GGAGGTCATCATATGGCTGAATGGaAAACAAAACGGACATA	17	PE	(63)
menE	TTCCGTGATCAGCAATATCTAGTAAACCAACAGCTTGAGACTTTGCGGTCCAAGCTGTTTTCTTTTCAATACAGACATT----TTACCTCGGAGATGATGACATGCTGACAGAACAGCCCAAC	16	PE	(63)
nadB	GGCGTTGTCGCTGCATAATCTAAATAAATCATCCGGTCTTCCCTCCATCCGTTCTCCATAAAAACTCTTGAGTTTATTT---TATCCTTGTGTAAATATAGGTGTCaAGACAGGTGTAAACA	17	PE	(64)
nasA	CTATTAAAAATTATGTCACAATGCATTGTTAACGCATTAAACGTGTCACAAAAACTTACACATGTCTTTCCAGAAAAT----AATGGTCCTATATCCTTGATTCaGAAAATGTAAAATAATG	16	PE	(65)
nasB	TTTCTGAATCAAGGATATAGGACCATTATTTTCTGGAAAAGACATGTGTAAGTTTTGTGACACGTTTAATGCGTTAACAATGCATTGTGACATAATTTTTAATaGGAGAAAACTTACGAG	20	PE	(65)
nifS	CCCCTGAACCGATGACTGCAATCGTCTTTTTAGACATATGCCATCCTCCTGTTGTTTACACCTGTCTTGACACCTATA---TTTACACAAGGATAAAATAAACTCaAGAGTTTTTTATGGAGA	18	PE	(64)
nrgA	TCGATAACATTTCTCAAAAACCATGTCAGGAAATCTTACATGAAAATGTTTTATCAT---TCTTTTTTCTCTATAATGAAGAAaTaTAATAATTGCTTTTT	17	PE	(66)
nusA	AAAAAGTGCCATCGTAATATTAGAGTTTCTGTCACTCTGCTTAGGTATGAAGGTAAGCGTATATCCATTTGCAATAAAAA---TATGGTTATGGTATAGTTTTATTcgGAAATGCTAACGATTA	16	PE	(67)
odhA	ATGCTTTTTGGCGTTTGTTTTTTCGAATGATTAAATTTTTTGTTTTTTATAAAGGTTTTTTACTATTTTGTGAACAATC-----AAGGTAGAATCAAATTGCAAACAGtGGTAAAATATCGTTG	16	PE	(68)
orfS	AGTATTGTATGTATTCTGTTTGATTTTCCTATTTCCTTTAATTATAATAGTCTACTTTACGACATTTTCTGAGCATTT--TCTCTTTTGTTGTATACTGATATTGTACGTTATAAAGGAGGA	18	S1	(69)
pbpD	CGGAAAGATTCTTATGCCAAAGCAAGCTGATTCCGAGAAAAAACTAAGGAATCTTCTAAATTTAACCTTCTCGTAATCT---CAAAAGAATGGTACGATATGGCTaGAATTTAGGAGAAAAGA	17	PE	(70)
pbpE	AATAGTGAACGGCAGGTATATGTGATGGGTTAAAAAGGATCTCTAACTGAAGGATTTTTCAAAATATTTGAAACGTTAG-TAGGTTAGTAACGTACAGAGATATgGGAGGTGGAGGAATGA	19	PE	(71)
pbpF	TACGATCGTTTTTAAAAAATCGGCTGTCCCATATGTAATAACCTTCATATCAGCCACCTCCTGCTTGCTAGTAATGTTTCTATTgGCGAGTGCTTCGAAC	17	PE	(72)
phoA	AGGCAAGATAACGAAAAGCGTTTTTTCATTTCCTTACAAGGCTTTCATTATTGTTTACATGATCAACAGCCGCATTTAA---CAAAGTTTCCCTAACATGATAAACgGAATACATTAAAGGAG	—	PE	(73)
ppiB	TTATTTTTTAGTAAAGCTCGGGAGATTCGTTTTCCCCCGAAAGGTGTGGTATAATAGCCGCGACAAGGAGAGACAGGCT-----ATAACCTATACATAATCTCCTTTgAAAATTAAAATAAAGA	16	PE	(74)
ptsXHI	GGTTCAGTCAACAGAGAACAAGAAGATATTGTGAAGATTGAAAATAAGGGTGTTAGTACGCCGTGCTTGTCAGATGAC---AAGTACGGTTGTATGATAATATTGTGAAGTAATAAGCT	17	PE	(75)
purA	AACGCTTATAAAACGAATGGAAGCGAACGAATATAGATTTACAAATAAATTAATGTTCGGATTTACAATTGACTTTCTGT---TTCTTCACTGATAAACTTGATTTgTTTGAATAGAATCGTTT	17	PE	(51)
purF	TTGAAGGGAAATTGATCTAAAACACGAACATTAGTAGAATGAATTTTTGTATCGTTCGATAAATATCGTTGACGATTATCC---ATGTCCGTTGTTAAGATAAACaTGAAATCAAAACACGACCT	17	S1	(76)
pyr	TCGAATTTTGAAGCGTCGCGTTCCCGAGGATATGGCAGAATTAATCGAAAACCTCAGAAAAAACGGTTGACAGAGGGT---TTCTTTTCTGAAATAATAAACGAAgCTGAATAGATTCTTTA	17	S1	(77)
rbs	GAGTGAAAACCTTAAATTTTTCAATTATATATACAATTTACAATTAGATTTCTTTTGATATTTTTATTGCTAACTTCG--GATTGTTCATGATAATCTATCTATgTAAACGGTTACATAAAC	18	PE	(11)
recA	TGATACATTATGATATTTTGATAGGAATCACGCCAAGAAAAAATCCGAATATGCGTTCGCTTTTTTCTTGGCAAATCCC---TTCAAACAGGGTATAGTATATgtAGTGGTAACATAAAGGAG	17	S1	(78)
ribG	ATTAAAAAACATCACTTTCGGATCGAAGGGTGATGTTTTGTTTTTCTCAAATTGTAAGTTTATTTCATTGCGTACTTTA---AAAAGGATCGCTATAATAACCAATaAGGACAAATGAATAAA	17	PE	(79)
rpmH	AGGTTTCGAAAGTTGAAAAGGTATGGTATCCTATTATGGTTGCAAGAAATAAAAGCACTAGTGAAGTTGACAATGAAT--AGGTAACGCAAATATAATAAGTAAGACtgCTTTAACAGCTA	18	S1	(80)
rpsd	ACCTAAAAGTTTACCACTAATTTTTGTTTATTATATCATAAACGGTGAAGCAATAATGGAGGAATGGTTGACTTCAAAA---CAAATAAATTATATAATGACCTTTgTGTGAAATATTGCAGC	17	PE	(81)

Table 1. continued

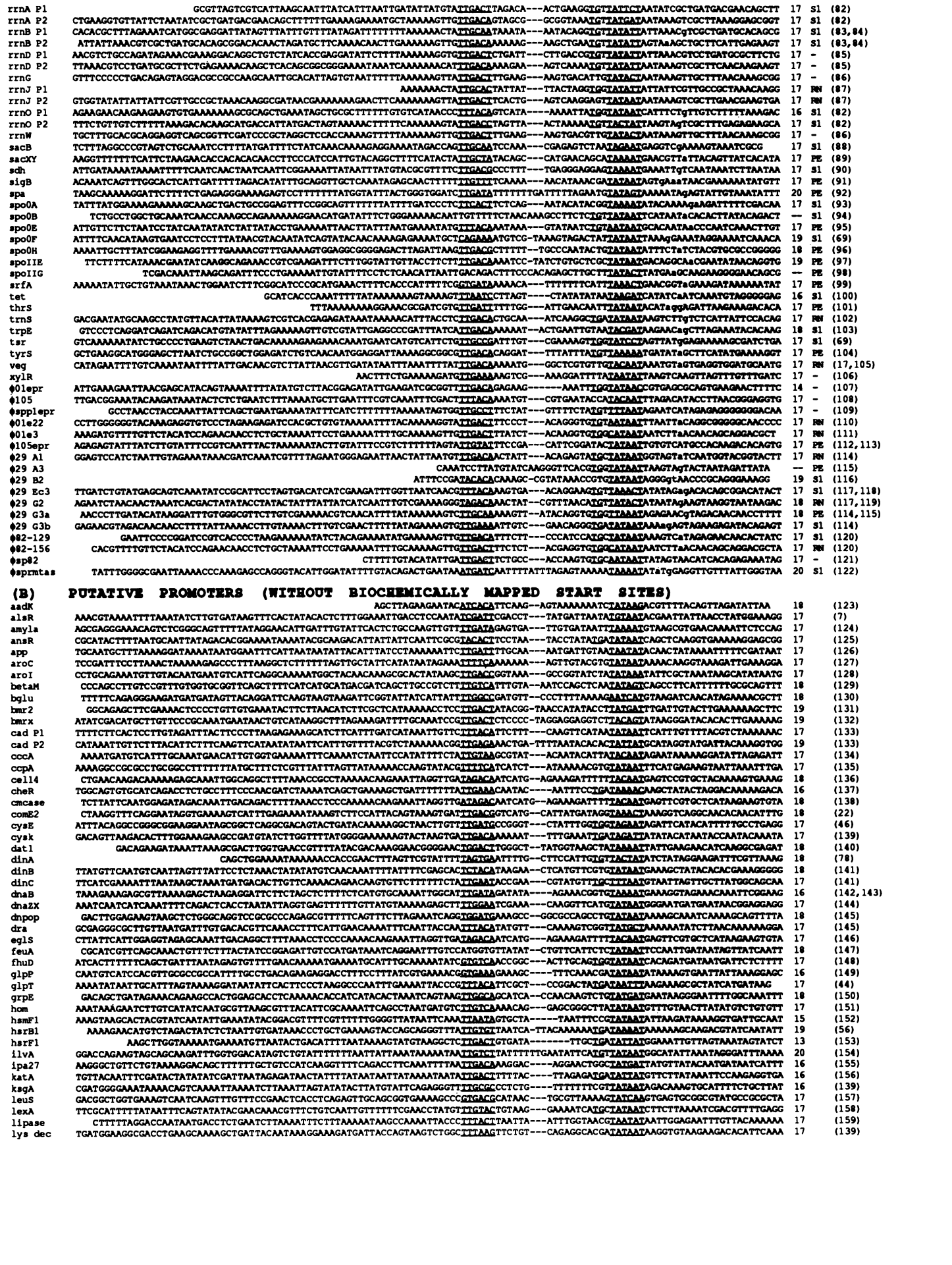

Table 1. continued

lysA	CAAAAGTAAGACCCAGCATCGTTCACCCCAAAAATCGCTTAAGGCAGCCTACCAATTCATAATCGCATTGAAACTGACT---GAAGAGTATGATAATGTATGTCTTAATTATGAAGGGACAGT	17	(160)
lysRhom	ATTCATAAACAATCATCCTTTATTTCTTACGAATTCAAACATCCTCGAGTCTGTTTATATTATTATCTTAGCACCGTTG----ATATATGTTGTATAATGCATATAAAATATCCTTTTCATGC	16	(139)
metS	ACTTTTTTTATTTTCATTAAAGATTTTTAATTTTAATTATTCTTTTTCAGGGCGTATGTATATATTCTTGATCTTAAAG----GCTAAGATGGTATCATAGATAAAGGATAAATATAAATAAT	16	(139)
mrgA	GCGTGGGAGTCTATCCTGAAGAAAAAGCTATTCAGCTGATCTAAATTATAATTATTATAATTTAGTATTGATTTTTATT----TAGTATATGATAATTAAGTCAACAGATCACAAGGAGAA	16	(161)
mrgC	GAATTCGCGCAGAAGACAAAGCATTACAGGAATACAGCGTAAAATAGCGATTGAAAATCATT---CTCAACCATGTTAAAATGGGTTTGAAAAAGATTCTC	17	(161)
mtrAB	AGAAAAGCCCCTTTCTAAGGGGCTTTTCATATTTCAAGAGCATGGGCTTCCTGACAGGGCATTCACTTTGCTTTAGCG--GGGCATATGTGCTAGAATCGAAATTAAATGTATTCATTGGTG	18	(162)
np	TGATATTATTCAACAAAAACAAACACAGGACAATACTATCAATTTTGTCTAGTTATGTTAGTTTTTGTTGAGTATTCCA---GAATGCTAGTTTAAATATAACAATATAAAGTTTTCAGTATTT	17	(163)
nprB	CGAAATATCTCTGCCCTTCTTTTTTGGGAAAATAAGAACGAAGCACCACATACAAGTTTTTGTATTTTTGATAGGATGA----AGAAAAATGGTAGATTCGGAAGGATGTGGTGTTTTGCGCA	16	(164)
orf17	TGATCATGTCTACAATGAGAAGGGCTTTATCTGCTTTGGACAATTGACAGTCTCC-TTTCCGTTTCAGTTATAGTTAATATGTAGCCTTTTTAGGCAAT	19	(165)
orfX14	AAGGCTCAGAGCACTGGCGGATCACGGTAAAAACGCTTTCTAAATTTCACATAACCTTCAAAAAGTAAGAAATGTGAAATGAACGTGCAATGATATAATTGAATGGATCTGCAATAGGGGGCA	--	(127)
pab	CAATTCGATTAAAAAAAGCCAAAACTCCCGGTTCGCCGGGAGTTTTTTATATTTCGTGCATCAAATATGCCATGTCTG----CGACGAATGTTATTCTTGGAAGAGAAATTCACTTTTTTCA	16	(166)
pcp	GGTTGGGAATAACTGGCGATACTGCTCCGGCAAATGCTCCATTATGATTCACTCCTTTACATTATTTTGATGAAAAACCCATAAATAAACAATAGAATAAAAGAAAAAGGAGATGGGAGA	20	(167)
pdhA	ATTAGTTTTTTACAACTTCGAAATAATACAGTTCAAAGCAGAGTGTAACTATCCTACATAAAAGGTTTGGCAAACTGG-GGATCGGCTAAAATATAATACGACTTACTGCTGATACTTTAGG	19	(168)
pdhC	CAGATACAGTATTTCCTTTCTCTCAAGCGGAGAGCGTATGGCTTCCAAACCATAAAGACGTTCTTGAAACAGCAAGAAAAGTGCTTGAATTTTAATCAAACTGCATAATCGAGAGGGAAG	21	(168)
pdhD	CGTGATGGCGAAATCGTAGCAGCTCCAGTCTTAGCTCTTTCTCTCAGCTTCGACCACCGTATGATCGACGGAGCAACTGCGCAAAATGCATTAAATCACATCAAGCGTTTACTGAACGAT	20	(168)
pel	TTTTTTTGCGGTCTTTGCGGTGGGATTTTGCAGAATGCCGCAATAGGATAGCGGAACATTTTCGGTTCTGAATGTCCCT---CAATTTGCTATTATATTTTTGTGATAAATTGGAATAAAATC	17	(169)
penP	TGTTGATTCTATTCTTATGCCCTTGTTCTCCCAACTTATAAAAATATAAGATCAATTGATTCTATGTTTACTAGCGGC--AAAGGGGTTTGTTACACTAACTGTCATGGGAAACATTTCAAA	18	(170)
pheS	TCACTAAACGTTGCGGTTGCGGCTGCCATCCTCGTGTATCACTTGCGAGGATAGTGTTGCGCCAAGCG---TTGGATTTCTCTATAATAGAACATAATATTTCAAAAACCGT	17	(171)
phoP	TTTTACTATAAATGAAAGCGCTATCATAAACGTCTTTATTTCTTTTAAAAATGATGTAAAAGGCGAATTGTCGGAATGT---CCTGCTTTCGCTAAAATAAAATCATGAAACATGTTAAGATG	17	(172)
prkA	TTTTTCCGTGCTGTCCGAGTCTGGATGATTCTGTATTTTTCTTTATTATGAACGGTCGCAGTCTGCCTTGTCAACGATG--TATGAGTATCTTTATTTTAAGAAACCAAAGCATGTATCA	18	(173)
pss	ACGCGTCTTGTCCTTTTTATGCGGGGGATAAAATAGTAAGGAACATGAAAAGATTGGCATTCATTTTTCAAAAGGTAA------CAAATGTTGTATAATAATAGAATTTGAATGCTGCTTTTT	15	(174)
pydhyo5	TGGTCATCTGTCCTGATTTTTCACTTTTTTGATTTAATTTCAGTATTGCATTTATACAGCACGGCTATTGCCAAATGTC---CCCATGGATTCTATAGTAGAAATTGTCAAGATTGGGGGATT	17	(168)
rib P2	GCAAACGGCCGTTTCGGAGGCCATTTCGTCTCAGGCCATGTCGACGGAACTGCGGAAATCACACGAATTGAAGAGAAAA---GCAACGCAGTTTACTATGATTTAAAAATGGACCCGTCATTA	17	(127)
rnpA	GAAGGGAAAGAATATGTGGTCCAAGACGGAGATGTTATTCATTTCCGATTTAATGTATAGGATGCAGTTGTAAAGGGAC---AAGAGCTTTGGTATAATATAAAATTGTGAGTAATAGAATTA	17	(139)
rodC	TTCAATATCTAATTAAAGTACTTCTTATTGAAATAAGAAGTACTTTTTGATGCTTACATTGATTTTTGTGACTAAAG--TATCCGGGTAGATATAAATTTCAATGATCATATATTCTAGAT	18	(175)
rplN	AAAATCGGTGACATCGTTAAGATCATGGAAACTCGTCCATTGTCTGCAACTAAGCGTTTCCGTCTAGTTGAAGTTGTCG----AAGAAGCTGTTATTATCTAATAAAGTTCGGAATTCTTTTT	16	(176)
rpoB	AACTGTTCGATGAAGTTTCCGTCGTTCAAAAAAAGAAAGGCTATTATATCATAAAAGCAAAAAAAGTTGACTCGGTAT---TTTAACTATGTTAATATTGTAAAATGCCAATGTATATATTT	17	(177)
rpsF	GAAAGCATAAAGCGTGCTTGAATTTTCAATTATTTCCCTTCTGATTCGGAACAAAATGATGAAAAAAATGAAACGCTTT-AATCTTTTTTTGATAGAATAGAAGCAAGATAGTACATGTTCAT	19	(139)
rtporf1	CCAAACCGAGTTTTTCAATTCAATGGACATAAAATAAAGAAAATGTAAATCGGACAATCGTTTCACATTGAAGATACGG---TGAAAATCGACTATAATAGAACTAAGTGAACTGTGAACCAA	17	(132)
scRNA	GTGCGTTTTTCAGAGAACTAAGAAAAAAGAAGAAAGCCGCCAGGAAAAACTTGTCTGAAATAGTACGGTTGCAATTTTTA---GGGGAAACAGATATACTTAAGTGTGCATCGTAATAGATGCA	17	(178)
secA	TGAATAATGAAGAGAAGCCTTCCGTGATGTCCGCGGAAGGTTTTTGTTTTTCTTATTTGCAAATTCTTTGGAAATAACA-----AAAGGTATGATATGATAATGAGAGGTATACATGGACTAGT	16	(179)
senS	GGAAAAGAGTGTATAAGGAAAAAGCGGGGATGCAATCTGATACAGTGTCAACACCCTCAAAAAATAGTTGACAGGTCGG--TATTGTATGAATTAACATGGTCAGTACAAATTTTTCAAATTT	18	(180)
serS	ACCGAGTGACGCAGCTGTTTGTTGAAATGGTTGAGGAATATAAGCAAAAGGCACTTGTATAAAACAGTTGAAAGCTGTG---GAAACTTATAGTACATTATAAGCACAAATAAAGATCGAAAA	17	(139)
sin	TTTTTTGAGAAAATACGATTATAATAAAGGTATATTGGAAAAAAATTCTGGTGATTTAATGGCAAATGACTTCCAGA-GACTAATGAAGCATACAATAAGTCATGGCCGGACTGGCTGAAA	19	(181)
sipS	GATTCATTTCTTTCAATATTTTCAGGCCTAGTTATGATCGACAAGAAAAAATAGGCATGATGTGGGTAGAAGATGAAG--ATTAACAGTTTGTAAAATCACTTCAATGGTCCTAATGGGATT	18	(182)
slp	ATAAAAAATAAAAAGCCATGCGGCTTAAGCCGCATGCTTTTTTTACTCTTCATCAACAGGACATTGACATTGTCC-GCAGCAAGTTCCATATAATGTTGTCACTTTCTCATCTTCAAAG	19	(168)
spo0K	AGAAGGATATTATGTTCATGGAAGAAAAACTAACGAAGTTTAAATATTTTAAATTGATAAAATAATATTGCAATAAATT---ATTTGTTTCATTATAATGAACTTGTTCACTCTATTGTTACA	17	(183)
tag	TGTAAAGATAATGTTAATCCTTTGTTGAAGATTTATGTTAAAATATAAGGTAGCTTGATTTTGCTTTTAGAAACTCTCG---GTTTTTAACTATAAAATAGAATCTTAAAGGAAGGAGGAGC	17	(184)
tenAI	CCGTTTAATTGTCACCGGATCTCGGGGGTACGTAACGGATCTTAAGAAAACATTGGAAGGATAAATCTTGACCAAATAA----GGATTCGTGTTAATATAGAGGTGCAGAACAATTCAATATG	16	(185)
tetBSM	ACTTGAGATAAATGCGAATAACGTTTTTCCTTGTTTCGCAATCTGGTCAAATTTTTATAAAAAAAGTAAAAGTTTAATCCTTAGTCTATATATAATAAGATCATATCAATCAAATGTAGG	--	(100)
thrZ	TGCTTGTATGGCTGTTTCTGGCTGCCCATAACGGGACTGCCGCGTCAGAAAAAATGTAAATCACTGATTGCGCTTTGTG---GGAATATATAGTAAGATATAAAAAATTATTGAAACAACAGG	17	(186)
thyB	AAGATGAAGAAAAAGAACAGGATGAAAAGCCTGTCGTCAGCTGACTATAAAAAAATCATTTCTGGGTTCAAAAATGAT--TTTTTATTGTGTTACACTACTAGAAGACTACTTTTAAAGGAT	18	(187)
trnJ	TATGCCGGTGTAGCTCAATTGGTAGAGCAACTGACTTGTAATCAGTAGGTTGGGGGTTCAAGTCCTCTTGCCGGCACCA---CTTTTATATGATATAAATACTCAAGTCTCTTGTAGAAGAGCC	17	(86)
trpS	AACGTCAGAGTGATTCCATTTTAATGGAATAATCAGGGTGGTACCACGGTTCATTCGTCCCTTTTTTACAGGGGAAGAATGAGCCTTTTTTATTATGTTTTAAGAAATGAGGTTGATGTT	20	(188)
tycA	TTGGATTTACGTAAAAAGGTTGTAAAAAAACTTGTTGAATTTTTTGCAAAATATCCCTATTTTTTAATCGATCTCCAAT-----TTTTCTCTGCTATAATGAGTTTCAGCGTCAGTAACCTAGT	16	(189)
valS	CATATGTAAACTTGGGCAATAGGTGCCTGCCCATCTGTATAGAAAGCGGTGTTTTTGGAAAAGATGATTGACGAATCAA---AAACACTTTTACTATAATAAAGAACATAAATAATGTATACAA	18	(190)
xpaC	CGCCATTATATTCATAGACCTGAAAAGGTCTTTTTTTGTACTCTTAATAATAAAAAGAAGATGAAACTTGTTTAAGGAT---TGAACGTAGTAGATAATAATAATAAAACTGAGTATAGACAC	17	(139)

The left column is the gene designation for each promoter sequence. Promoters are defined by the first downstream gene if an operon is present. In the case of multiple start sites, promoters are designated P1, P2, etc., as in the original citation. The spacer length (SL) and the method used to determine the start site (SS) are indicated to the right of each sequence. The abbreviations used are: PE, primer extension transcript mapping; S1, S1 nuclease transcript mapping; RN, RNA analysis by either high resolution run-off transcription or, in some cases, by RNA sequencing. The last column contains the reference to the promoter sequence or, when available, to determination of the start site.

Conserved bases in the promoter region

The most highly conserved bases have been identified for all 236 promoters (Fig. 1) and for a subset of 125 promoters with the most carefully defined transcription start sites (data not shown), but both analyses yielded similar results. Overall, the pattern of nucleotide conservation is reminiscent of that observed for *E.coli* promoters (190,191) and can be summarized as TTGaca ($N_{17\pm1}$) TAtAAT (where bases in capital letters are present in >70% of promoters). As inferred from biochemical studies (207,209), *B.subtilis* appears to be less tolerant of deviation from this 12 bp consensus than *E.coli*: on average *B.subtilis* promoters match consensus at 9.1 positions, compared with only 7.9 for *E.coli* (193,215). Perfect (12 out of 12) matches to this consensus are found in four out of the 125 chromosomal promoters in Table 1A (*glnR*, *rpmH*, *spoIIE* and *trnS*), but in none of the 298 tabulated *E.coli* promoters (193). In addition, relatively few *B.subtilis* promoters (seven out of 125 in Table 1A) lack an identifiable –35 region (<3/6 match to consensus), although not all of the assigned –35 regions are necessarily functional.

Many other positions within the promoter exhibit a lesser degree of sequence conservation. Statistical analysis reveals conservation of a T at –48, an A-rich region near –43, TnTG at –17 to –14 and a downstream extension of the –10 region (Fig. 1). Each of these features was noted previously based on an alignment of 29 promoters from several different gram positive organisms (208), but they are not prominent in alignments of *E.coli* promoters (191–193).

The conserved –35 and –10 elements are most frequently separated by a 17 base spacer region (Fig. 2A), as found for *E.coli* RNAP. There is no apparent correlation between the occurrence of the TG dinucleotide at –15,–14 (see below) and spacer length. The average distance between the –10 region and the start of transcription is seven bases, although values between four and 10 bases have been measured by primer extension start site mapping (Fig. 2B). *B.subtilis* RNAP appears to initiate transcription preferentially with a purine, as noted previously for *E.coli* RNAP (192). This is most apparent where transcription initiates at either of two purines while an intervening pyrimidine is largely ignored (e.g. *gsiA* and *nrgA*).

The conserved TG dinucleotide

Alignment of nine strong σ^A-dependent promoters led to a proposed consensus of RTRTG for the –18 to –14 positions of *B.subtilis* promoters (209). Recently a class of *E.coli* promoters lacking a conserved –35 region but containing a TG dinucleotide at –15,–14 has also been described. These 'extended' –10 promoters include a derivative of λ P_{re} (216), Gal P1 (217) and *cysG* (218). This element may play a role in promoter melting,

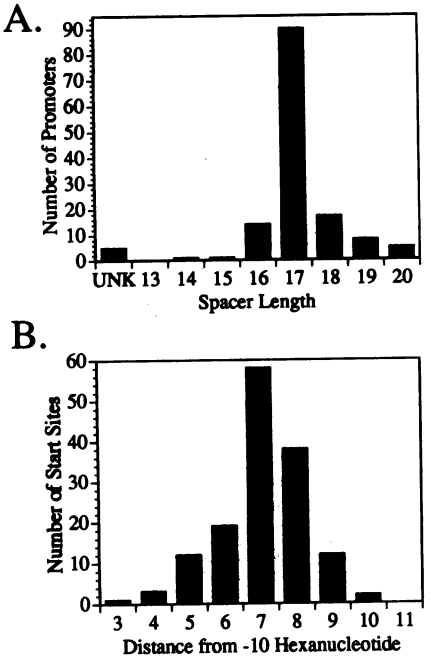

Figure 2. (**A**) Distribution of spacer lengths for the 142 promoters of Table 1A. Spacer lengths were assigned by optimizing the match to consensus for both the –35 and –10 regions. When no –35 region was discernible (<3/6 match to consensus) the spacer is assigned as UNK (unknown). (**B**) The distribution of mapped start sites as a function of distance from the last conserved base of the –10 hexamer (TATAAT).

since a single base change which creates a TG dinucleotide in a derivative of Gal P1 reduces the transition temperature by 20°C (219).

In the promoter alignments presented here, the T at –15 and the G at –14 are both conserved (58 and 52% respectively). Moreover, the T at –15 and G at –14 are positively correlated: TG occurs in 45% of promoters (Table 1A), significantly more than expected from the product of the individual base frequencies, 30% (0.58 × 0.52). This suggests that the important conserved element is at least a dinucleotide. Indeed, dinucleotide composition is a more precise indicator of DNA structural properties than base composition (220,221). The RTRTG motif proposed for this region (209) is supported by the presence of a weakly conserved T at –17 (Fig. 1) and the observation that T(–17) and R(–16) are 52 and 69% correlated with the presence of a TG dinucleotide.

Analysis of dinucleotide frequencies

Next, I analyzed the dinucleotide composition as a function of promoter position (Fig. 3). This analysis reveals the same overall pattern as when the most frequent base is plotted (Fig. 1): in each case the –35 and –10 regions contain the most conserved elements. However, since there are 16 dinucleotides, the background signal is reduced by at least a factor of two. In this analysis, 'TG' at –15 (position indicates the upstream base) clearly emerges as a conserved feature. Throughout the upstream promoter region (–100 to –36), 'AA' and 'TT' are the most frequent dinucleotides at 60 of the 64 positions. The exceptions are –92 and –48 (TA), –74 (GA) and –68 (AT). This is consistent with the AT-rich nature of this region and with the abundance of A_n and T_n tracts (see below).

Closer inspection of the dinucleotide frequencies upstream of –35 reveals a striking pattern (Fig. 4). There are regions enriched for 'AA' centered at –43, –54 and –64, with intervening regions enriched in 'TT.' This suggests that the binding of RNAP influences the DNA sequence as far upstream as –70. Non-randomness in dinucleotide composition upstream of *E.coli* promoters has also been reported (206), but no helical phasing was noted.

Statistical analysis of oligo(dA) and oligo(dT) frequencies

Inspection of the aligned promoter sequences suggests that the upstream regions are enriched for short A_n and T_n tracts, a feature noted also in promoters from *Lactobacillus* (222). Indeed, statistical analysis of A_n and T_n sequences within the upstream region (–36 to –80) of the 142 promoters of Table 1A reveals a substantial over-representation of longer repeats ($n > 4$) (Fig. 5A). In contrast, analysis of randomly chosen 30 kb regions of the *B.subtilis* genome revealed only a slight (2–3-fold) over-representation of longer A_n and T_n repeats (Fig. 5B). Note that since the relative AT-richness of the upstream region has been

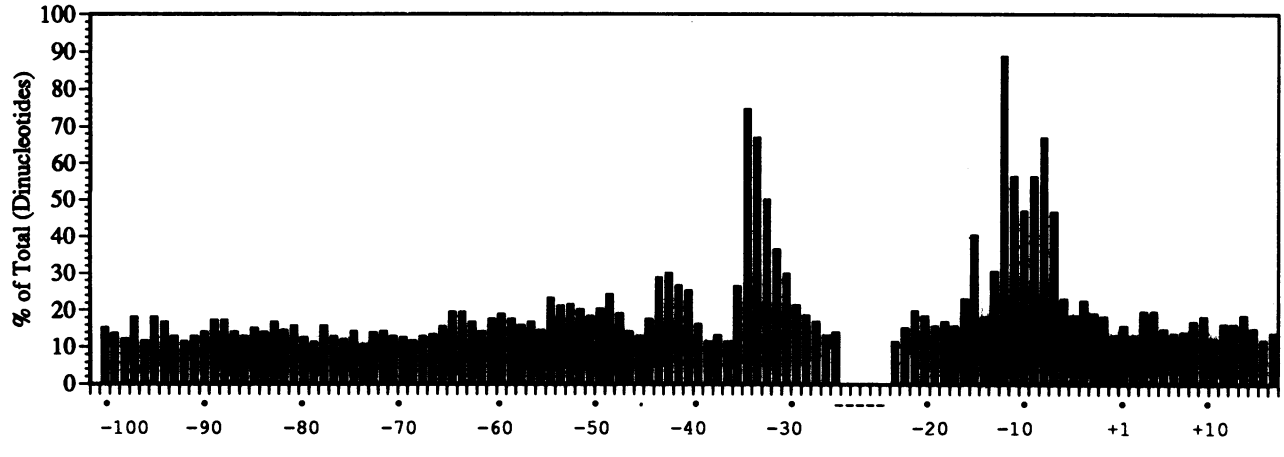

Figure 3. The frequency of the most commonly occurring dinucleotide at each position is indicated for the 236 promoters of Table 1 (the *x*-axis indicates the position of the upstream base of the dinucleotide).

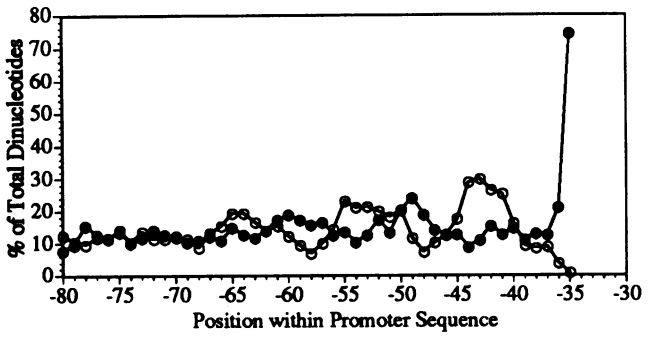

Figure 4. Frequency of AA (○) and TT (●) dinucleotides in the region between –35 and –80. Three regions where 'AA' is the most abundant dinucleotide (–40 to –45, –50 to –56, and –62 to –67) alternate with regions where 'TT' is most abundant.

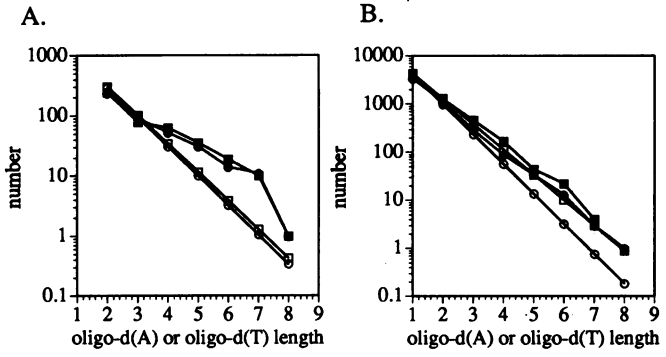

Figure 5. Enumeration of A_n and T_n sequences. The expected (open symbols) and actual (closed symbols) occurrence of A_n (squares) and T_n (circles) are plotted for (**A**) the upstream promoter regions (–36 to –80; Table 1A) and (**B**) a 30 kb segment of the *B.subtilis* genome. The expected occurrence of each dinucleotide is a decreasing exponential function of sequence length (n) calculated using the actual base frequencies within the corresponding DNA region.

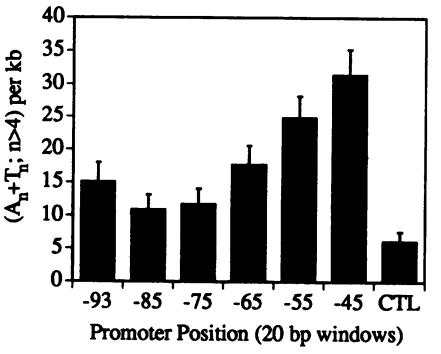

Figure 6. Enumeration of A_n and T_n sequences for $n > 4$. The occurrence of A_n and T_n sequences ($n > 4$) was determined for 20 bp windows centered at the indicated base in the upstream promoter regions (Table 1A). As a control (CTL), A_n and T_n sequences ($n > 4$) per kb of sequence were enumerated for five regions of 30 kb each from the *B.subtilis* genome. The error bars represent 1 SD based on Poisson statistics (for the promoters) and on the five independent determinations of A_n and T_n frequencies for the control. Note that no correction has been applied to account for the difference in AT-richness between the promoter DNA and *B.subtilis* genomic DNA.

taken into account in calculating the expected number of repeats, the abundance of A_n and T_n ($n > 4$) in the –36 to –80 region is not simply a consequence of high AT content (Fig. 5A). Next, I measured the total number of A_n and T_n repeats ($n > 4$) in overlapping 20 bp windows throughout the upstream promoter region. This analysis revealed that the over-representation of longer A_n and T_n tracts ($n > 4$) noted for the –36 to –80 interval decreases with increasing distance from the start site (Fig. 6).

The significance of these periodically repeated A_n and T_n tracts is not known, but two possibilities warrant consideration. First, this upstream region may contact the α subunits of RNAP, as demonstrated for both the *E.coli rrnB* P1 and the *B.subtilis* flagellin promoters (196–198). Secondly, the upstream region of the promoter may wrap around RNAP during transcription initiation (reviewed in 201). In fact, there is a very good correlation between those A_n and T_n tracts which are over-represented in the upstream region and those which induce a static DNA bend: A_n tracts of at least four bases in length are needed for production of a DNA bend and bending is maximal when n = 6 (223). Since the relationship between α binding and the writhe of the DNA in the promoter complex is not known, these two explanations are not mutually exclusive.

The observation that the average DNA sequence inferred from alignment of 236 promoters contains alternating A and T tracts in the region between –70 and –36 (Fig. 4) predicts that an upstream region conforming to this pattern might be stimulatory for transcription. This appears to be the case: the upstream regions of both the σ^H-dependent *spoVG* promoter (210) and the ϕ82alu156 promoter (120; Table 1) closely match the derived consensus, contain intrinsic DNA bends (211,224) and activate transcription *in vivo* and *in vitro* (210,211,225).

ACKNOWLEDGEMENTS

I thank Kurt Fredrick for pointing out the similarity between the *spoVG* UAS and the upstream promoter consensus. This work was supported by National Institutes of Health grant GM-47446.

REFERENCES

1 Perego,M., Spiegelman,G.B. and Hoch,J.A. (1988) *Mol. Microbiol.* **2**, 689–699.
2 Grundy,F.J., Waters,D.A. and Henkin,T.M. (1993) *J. Bacteriol.* **175**, 7348–7355.
3 Grundy,F.J., Turinsky,A.J. and Henkin,T.M. (1994) *J. Bacteriol.* **176**, 4527–4533.
4 Morohoshi,F., Hayashi,K. and Munakata,N. (1993) *J. Bacteriol.* **175**, 6010–6017.
5 Haijema,B.J., Hamoen,L.W., Kooistra,J., Venema, G. and van Sinderen, D. (1995) *Mol. Microbiol.* **15**, 213–223.
6 Siranosian,K.J., Ireton, K. and Grossman,A.D. (1993) *J. Bacteriol.* **175**, 6789–6796.
7 Renna,M.C., Najimudin, N., Winik, L.R. and Zahler,S.A. (1993) *J. Bacteriol.* **175**, 3863–3875.
8 Nicholson,W.L., Park,Y.K., Henkin,T.M., Won,M., Weickert,M.J., Gaskell,J.A. and Chambliss,G.H. (1987) *J. Mol. Biol.* **198**, 609–618.
9 Sun,D. and Setlow,P.L. (1991) *J. Bacteriol.* **173**, 3831–3845.
10 Henner,D.J., Ferrari,E., Perego,M. and Hoch,J.A. (1988) *J. Bacteriol.* **170**, 296–300.
11 O'Reilly,M., Woodson,K., Dowds,B.C. and Devine,K.M. (1993) *Mol. Microbiol.* **11**, 87–98.
12 Smith,M.C.M., Mountain,A. and Baumberg, S. (1986) *Gene* **49**, 53–60.

13 Chen,N.Y., Jiang,S.Q., Klein,D.A. and Paulus,H. (1993) *J. Biol. Chem.* **268**, 9448–9465.
14 Song,B.H. and Neuhard,J. (1989) *Mol. Gen. Genet.* **216**, 462–468.
15 Wang,P.-Z. and Doi,R.H. (1984) *J. Biol. Chem.* **259**, 8619–8625.
16 Jin,S. and Sonenshein,A.L. (1994) *J. Bacteriol.* **176**, 4680–4690.
17 Whipple,F.W. and Sonenshein,A.L. (1992) *J. Mol. Biol.* **223**, 399–414.
18 Fouet,A., Jin,S.-F., Raffel,G. and Sonenshein,A.L. (1990) *J. Bacteriol.* **172**, 5408–5415.
19 Weinrauch,Y., Guillen,N. and Dubnau,D.A. (1989) *J. Bacteriol.* **171**, 5362–5375.
20 Mohan,S. and Dubnau,D. (1990) *J. Bacteriol.* **172**, 4064–4071.
21 Mohan,S., Aghion,J., Guillen,N. and Dubnau,D.A. (1989) *J. Bacteriol.* **171**, 6043–6051.
22 Hahn,J., Inamine,G., Kozlov,Y. and Dubnau,D.A. (1993) *Mol. Microbiol.* **10**, 99–111.
23 Londono-Vallejo,J.A. and Dubnau,D. (1993) *Mol. Microbiol.* **9**, 119–131.
24 Albano,M., Breitling,R. and Dubnau,D.A. (1989) *J. Bacteriol.* **171**, 5386–5404.
25 Van-Sinderen,D., Ten-Berge,A., Jan-Hayema,B., Hamoen,L. and Venema,G. (1994) *Mol. Microbiol.* **11**, 695–704.
26 Weinrauch,Y., Msadek,T., Kunst,F. and Dubnau,D.A. (1991) *J. Bacteriol.* **173**, 5685–5693.
27 Willimsky,G., Bang,H., Fischer,G. and Marahiel,M.A. (1992) *J. Bacteriol.* **174**, 6326–6335.
28 Mueller,J.P. and Taber,H.W. (1989) *J. Bacteriol.* **171**, 4979–4986.
29 Slack,F.J., Mueller,J.P., Strauch,M.A., Mathiopoulos,C. and Sonenshein,A.L. (1991) *Mol. Microbiol.* **5**, 1915–1925.
30 Yang,M., Ferrari,E., Chen,E. and Henner,D.J. (1986) *J. Bacteriol.* **166**, 113–119.
31 Msadek,T., Kunst,F., Henner,D., Klier,A., Rapoport,G. and Dedonder,R. (1990) *J. Bacteriol.* **172**, 824–834.
32 Harry,E.J., Rowland,S.L., Malo,M.S. and Wake,R.G. (1994) *J. Bacteriol.* **176**, 1172–1179.
33 Moriya,S., Fukuoka,T., Ogasawara,N. and Yoshikawa,H. (1988) *EMBO J.* **7**, 2911–2917.
34 Wang,L.F. and Doi,R.H. (1987) *Mol. Gen. Genet.* **207**, 114–119.
35 Wetzstein,M., Voelker,U., Dedio,J., Loebau,S., Zuber,U., Schiesswohl,M., Herget,C., Hecker,M. and Schumann,W. (1992) *J. Bacteriol.* **174**, 3300–3310.
36 Honda,K., Nakamura,K., Nishiguchi,M. and Yamane,K. (1993) *J. Bacteriol.* **175**, 4885–4894.
37 Gonzy-Treboul,G., Karmazyn-Campelli,C. and Stragier,P. (1992) *J. Mol. Biol.* **224**, 967–979.
38 Hove-Jensen,B. (1992) *J. Bacteriol.* **174**, 6852–6856.
39 Nilsson,D., Hove-Jensen,B. and Arnvig,K. (1990) *Mol. Gen. Genet.* **218**, 565–571.
40 Strauch,M.A., Aronson,A.I., Brown,S.W., Schreier,H.J. and Sonenshein,A.L. (1988) *Gene* **71**, 257–26.
41 Holmberg,C. and Rutberg,B. (1991) *Mol. Microbiol.* **5**, 2891–2900.
42 Holmberg,C., Beijer,L., Rutberg,B. and Rutberg,L. (1990) *J. Gen. Microbiol.* **136**, 2367–2375.
43 Beijer,L., Nilsson,R.P., Holmberg,C. and Rutberg,L. (1993) *J. Gen. Microbiol.* **139**, 349–359.
44 Nilsson,R.P., Beijer,L. and Rutberg,B. (1994) *Microbiology* **140**, 723–730.
45 Bohannon,D.E. and Sonenshein,A.L. (1989) *J. Bacteriol.* **171**, 4718–4727.
46 Gagnon,Y., Breton,R., Putzer,H., Pelchat,M., Grunberg-Manago,M. and Lapointe,J. (1994) *J. Biol. Chem.* **269**, 7473–7482.
47 Robson,L.M. and Chambliss,G.H. (1987) *J. Bacteriol.* **169**, 2017–2025.
48 Fujita,Y. and Fujita,T. (1986) *Nucleic Acids Res.* **14**, 1237–1252.
49 Li,M. and Wong,S.-L. (1992) *J. Bacteriol.* **174**, 3981–3992.
50 Mueller,J.P., Bukusoglu,G. and Sonenshein,A.L. (1992) *J. Bacteriol.* **174**, 4361–4373.
51 Maentsaelae,P. and Zalkin,H. (1992) *J. Bacteriol.* **174**, 1883–1890.
52 Ye,R. and Wong,S.-L. (1994) *J. Bacteriol.* **176**, 3314–3320.
53 Ogasawara,N., Moriya,S. and Yoshikawa,H. (1985) *Nucleic Acids Res.* **13**, 2267–2279.
54 Micka,B., Groch,N., Heinemann,U. and Marahiel,M.A. (1991) *J. Bacteriol.* **173**, 3191–3198.
55 Petricek,M., Rutberg,L., Schroeder,I. and Hederstedt,L. (1990) *J. Bacteriol.* **172**, 2250–2258.
56 Xu,G.L., Kapfer,W., Walter,J. and Trautner,T.A. (1992) *Nucleic Acids Res.* **20**, 6517–6523.
57 Wray,L.V.,Jr and Fisher,S.H. (1994) *J. Bacteriol.* **176**, 5466–5473.
58 Grandoni,J.A., Zahler,S.A., and Calvo,J.M. (1992) *J. Bacteriol.* **174**, 3212–3219.
59 LeDeaux,J.R. and Grossman,A.D. (1995) *J. Bacteriol.* **177**, 166–175.
60 Riethdorf,S., Voelker,U., Gerth,U., Winkler,A., Engelman,S. and Hecker,M. (1994) *J. Bacteriol.* **176**, 6518–6527.
61 Chen,N.Y., Hu,F.-M. and Paulus,H. (1987) *J. Biol. Chem.* **262**, 8787–8798.
62 Kong,L., Siranosian,K.J., Grossman,A.D. and Dubnau,D. (1993) *Mol. Microbiol.* **9**, 365–373.
63 Driscoll,J.R. and Taber,H.W. (1992) *J. Bacteriol.* **174**, 5063–5071.
64 Sun,D. and Setlow,P. (1993) *J. Bacteriol.* **175**, 1423–1432.
65 Nakano,M.M., Yang,F., Hardin,P. and Zuber,P. (1995) *J. Bacteriol.* **177**, 573–579.
66 Wray,L.V., Atkinson,M.R. and Fisher,S.H. (1994) *J. Bacteriol.* **176**, 108–114.
67 Shazand,K., Tucker,J., Grunberg-Manago,M., Rabinowitz,J.C., and Leighton,T. (1993) *J. Bacteriol.* **175**, 2880–2887.
68 Resnekov,O., Melin,L., Carlsson,P., Mannerlov,M., von Gabain,A. and Hederstedt,L. (1992) *Mol. Gen. Genet.* **234**, 285–296.
69 Trach,K., Chapman,J.W., Piggot,P., Lecoq,D. and Hoch,J.A. (1988) *J. Bacteriol.* **170**, 4194–4208.
70 Popham,D.L. and Setlow,P. (1994) *J. Bacteriol.* **176**, 7197–7205.
71 Popham,D.L. and Setlow,P.L. (1993) *J. Bacteriol.* **175**, 2917–2925.
72 Popham,D.L. and Setlow,P. (1993) *J. Bacteriol.* **175**, 4870–4876.
73 Hulett,F.M., Lee,J., Shi,L., Sun,G., Chesnut,R., Sharkova,E., Duggan,M.F. and Kapp,N. (1994) *J. Bacteriol.* **176**, 1348–1358.
74 Herrler,M., Bang,H. and Marahiel,M.A. (1994) *Mol. Microbiol.* **11**, 1073–1083.
75 Gonzy-Treboul,G., Zagorec,M., Rain-Guion,M.C. and Steinmetz,M. (1989) *Mol. Microbiol.* **3**, 103–112.
76 Ebbole,D.J. and Zalkin,H. (1988) *J. Biol. Chem.* **263**, 10894–10902.
77 Quinn,C.L., Stephenson,B.T. and Switzer,R.L. (1991) *J. Biol. Chem.* **266**, 9113–9127.
78 Cheo,D.L., Bayles,K.W. and Yasbin,R.E. (1993) *J. Bacteriol.* **175**, 5907–5915.
79 Kil,Y.V., Mironov,V.N., Gorishin,I.Y., Kreneva,R.A. and Perumov,D.A. (1992) *Mol. Gen. Genet.* **233**, 483–488.
80 Moriya,S., Ogasawara,N. and Yoshikawa,H. (1985) *Nucleic Acids Res.* **13**, 2251–2265.
81 Grundy,F.J. and Henkin,T.M. (1990) *J. Bacteriol.* **172**, 6372–6379.
82 Ogasawara,N., Moriya,S. and Yoshikawa,H. (1983) *Nucleic Acids Res.* **11**, 6301–6318.
83 Stewart,G.C. and Bott,K.F. (1983) *Nucleic Acids Res.* **11**, 6289–6300.
84 Ogasawara,N., Moriya,S., von-Meyenburg,K., Hansen,F.G. and Yoshikawa,H. (1985) *EMBO J.* **4**, 3345–3350.
85 Jarvis,E.D., Cheng,S. and Rudner,R. (1990) *Genetics* **126**, 785–798.
86 Green,C.J. and Vold,B.S. (1992) *J. Bacteriol.* **174**, 3147–3151.
87 Widom,R. (1988) Heterogeneity of ribosomal RNA operons in *Bacillus subtilis*. PhD thesis. City University of New York, New York, NY.
88 Shimotsu,H. and Henner,D.J. (1986) *J. Bacteriol.* **168**, 380–388.
89 Crutz,A.M. and Steinmetz,M. (1992) *J. Bacteriol.* **174**, 6087–6095.
90 Melin,L., Rutberg,L. and von Gabain,A. (1989) *J. Bacteriol.* **171**, 2110–2115.
91 Wise,A.A. and Price,C.W. (1995) *J. Bacteriol.* **177**, 123–133.
92 Hansen,J.N. and Chung,Y.J. (1992) *J. Bacteriol.* **174**, 6699–6705.
93 Chibazakura,T., Kawamura,F. and Takahashi,H. (1991) *J. Bacteriol.* **173**, 2625–2632.
94 Trach,K. and Hoch,J.A. (1989) *J. Bacteriol.* **171**, 1362–1371.
95 Perego,M. and Hoch,J.A. (1991) *J. Bacteriol.* **173**, 2514–2520.
96 Weir,J., Predich,M., Dubnau,E., Nair,G. and Smith,I. (1991) *J. Bacteriol.* **173**, 521–529.
97 Kenney,T.J., York,K., Youngman,P. and Moran,C.P.Jr (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9109–9113.
98 York,K., Kenney,T.J., Satola,S., Moran,C.P.,Jr, Poth,H. and Youngman,P. (1992) *J. Bacteriol.* **174**, 2648–2658.
99 Nakano,M.M., Xia,L. and Zuber,P. (1991) *J. Bacteriol.* **173**, 5487–5493.
100 Sakaguchi,R., Amano,H. and Shishido,K. (1988) *Biochem. Biophys. Acta* **950**, 441–444.
101 Gendron,N., Putzer,H. and Grunberg-Manago,M. (1994) *J. Bacteriol.* **176**, 486–494.
102 Garrity,D.B. and Zahler,S.A. (1993) *J. Bacteriol.* **175**, 6512–6517.
103 Shimotsu,H. and Henner,D.J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6315–6319.
104 Henkin,T.M., Glass,B.L. and Grundy,F.J. (1992) *J. Bacteriol.* **174**, 1299–1306.
105 LeGrice,S.F.J. and Sonenshein,A.L. (1982) *J. Mol. Biol.* **162**, 551–564.
106 Kreuzer,P., Gaertner,D., Allmansberger,R. and Hillen,W. (1989) *J. Bacteriol.* **171**, 3840–3845.

107 Osburne,M.S. and Craig,R.J. (1986) *J. Gen. Microbiol.* **132**, 565–568.
108 Osburne,M.S., Craig,R.J. and Rothstein,D.M. (1985) *J. Bacteriol.* **163**, 1101–1108.
109 Tailor,R., Bensi,G., Morelli,G., Canosi,U. and Trautner,T.A. (1985) *J. Gen. Microbiol.* **131**, 1259–1262.
110 Lee,G. and Pero,J. (1981) *J. Mol. Biol.* **152**, 247–265.
111 Lee,G., Talkington,C. and Pero,J. (1980) *Mol. Gen. Genet.* **180**, 57–65.
112 Cully,D.F. and Garro,A.J. (1985) *Gene* **38**, 153–164.
113 Dhaese,P., Seurinck,J., De Smet,B. and Van Montagu,M. (1985) *Nucleic Acids Res.* **13**, 5441–5455.
114 Murray,C.L. and Rabinowitz,J.C. (1982) *J. Biol. Chem.* **257**, 1053–1062.
115 Rojo,F., Nuez,B., Mencia,M. and Salas,M. (1993) *Nucleic Acids Res.* **21**, 935–940.
116 Mellado,R.P., Barthelemy,I. and Salas,M. (1986) *Nucleic Acids Res.* **14**, 4731–4741.
117 Garvey,K.J., Yoshikawa,H. and Ito,J. (1985) *Gene* **40**, 301–309.
118 Dobinson,K.F. and Spiegelman,G.B. (1985) *J. Biol. Chem.* **260**, 5950–5955.
119 Dobinson,K.F. and Spiegelman,G.B. (1987) *Biochemistry* **26**, 8206–8213.
120 McAllister,C.F. and Achberger,E.C. (1988) *J. Biol. Chem.* **263**, 11743–11749.
121 Achberger,E.C., Hilton,M.D. and Whiteley,H.R. (1982) *Nucleic Acids Res.* **10**, 2893–2910.
122 Buhk,H.-J., Behrens,B., Tailor,R., Wilke,K., Prada,J.J., Guenthert,U., Noyer-Weidner,M., Jentsch,S.P. and Trautner,T.A. (1984) *Gene* **29**, 51–61.
123 Ohmiya,K., Tanaka,T., Noguchi,N., O'Hara,K. and Kono,M. (1989) *Gene* **78**, 377–378.
124 Emori,M., Takagi,M., Maruo,B. and Yano,K. (1990) *J. Bacteriol.* **172**, 4901–4908.
125 Sun,D. and Setlow,P.L. (1993) *J. Bacteriol.* **175**, 2501–2506.
126 Koide,A. and Hoch,J.A. (1994) *Mol. Microbiol.* **13**, 417–426.
127 Sorokin,A.V., Zumstein,E., Azevedo,V., Ehrlich,S.D. and Serror,P. (1993) *Mol. Microbiol.* **10**, 385–395.
128 Nakane,A., Ogawa,K., Nakamura,K. and Yamane,K. (1994) *J. Ferment. Bioengng* **77**, 312–314.
129 Mendoza,N.S., Ueda,M., Kawaguchi,T., Joson,L.M. and Arai,M. (unpublished) D37964.
130 Tezuka,K., Yuuki,T. and Yabuuchi,S. (1989) *Agric. Biol. Chem.* **53**, 2335–2339.
131 Ahmed,M., Lyass,L., Vazquez-Laslop,N. and Neyfakh,A.A. (unpublished) L32599.
132 Neyfakh,A.A., Bidnenko,V.E. and Chen,L.B. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4781–4785.
133 Hemilae,H., Koivula,T. and Paulin,L. (unpublished) X58433.
134 von Wachenfeldt,C. and Hederstedt,L. (1990) *J. Biol. Chem.* **265**, 13939–13948.
135 Henkin,T.M., Grundy,F.J., Nicholson,W.L. and Chambliss,G.H. (1991) *Mol. Microbiol.* **5**, 575–584.
136 Seo,Y.S., Lee,Y.H., Pek,U.H. and Kang,H.-S. (1986) *Korean J. Microbiol.* **24**, 236–242.
137 Kirsch,M.L., Zuberi,A.R., Henner,D., Peters,P.D., Yazdi,M.A. and Ordal,G.W. (1993) *J. Biol. Chem.* **268**, 25350–25356.
138 Park,S.H., Kim,H.K. and Pack,M.Y. (1991) *Agric. Biol. Chem.* **55**, 441–448.
139 Ogasawara,N., Nakai,S. and Yoshikawa,H. (1994) *DNA Res.* **1**, 1–14.
140 Morohoshi,F., Hayashi,K. and Munakata,N. (1989) *Nucleic Acids Res.* **17**, 6531–6543.
141 Cheo,D.L., Bayles,K.W. and Yasbin,R.E. (1991) *J. Bacteriol.* **173**, 1696–1703.
142 Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H. (1986) *Nucleic Acids Res.* **14**, 9989–9999.
143 Hoshino,T., McKenzie,T., Schmidt,S., Tanaka,T. and Sueoka,N. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 653–657.
144 Struck,J.C., Vogel,D.W., Ulbrich,N. and Erdmann,V.A. (1988) *Nucleic Acids Res.* **16**, 2720.
145 Saxild,H.H., Andersen,L.N. and Hammer,K. (unpublished) X82174.
146 MacKay,R.M., Lo,A., Willick,G., Zuker,M., Baird,S., Dove,M., Moranelli,F. and Seligy,V. (1986) *Nucleic Acids Res.* **14**, 9159–9170.
147 Quirk,P.G., Guffanti,A.A., Clejan,S., Cheng,J. and Krulwich,T.A. (unpublished) L19954.
148 Hantke,K. and Schneider,R. (1993) *Mol. Microbiol.* **8**, 111–121.
149 Beijer,L., Nilsson,R.-P., Holmberg,C. and Rutberg,L. (1993) *J. Gen. Microbiol.* **139**, 349–359.
150 Wetzstein,M. and Schumann,W. (1990) *Nucleic Acids Res.* **18**, 1289.
151 Parso,C. and Cohen,G.N. (1988) *J. Biol. Chem.* **263**, 14654–14660.
152 Walter,J., Noyer-Weidner,M. and Trautner,T.A. (1990) *EMBO J.* **9**,1007–1013.
153 Kapfer,W., Walter,J. and Trautner,T.A. (1991) *Nucleic Acids Res.* **19**, 6457–6463.
154 Armpriester,J.M.,Jr and Fink,P.S. (unpublished) M58606.
155 Glaser,P., Kunst,F., Arnaud,M., Coudart,M.P., Gonzales,W., Hullo,M.F., Ionescu,M., Lubochinsky,B., Marcelino,L., Moszer,I., Presecan,E., Santana,M., Schneider,E., Schweizer,J., Vertes,A., Rapoport,G. and Danchin,A. (1993) *Mol. Microbiol.* **10**, 371–384.
156 Bol,D.K. (1991) *Bacillus subtilis* responses and mechanisms of resistance to hydrogen peroxide. PhD thesis. University of Maryland Baltimore County, Baltimore, MD.
157 Vander Horn,P.B. and Zahler,S.A. (1992) *J. Bacteriol.* **174**, 3928–3935.
158 Raymond-Denise,A. and Guillen,N. (1991) *J. Bacteriol.* **173**, 7084–7091.
159 Dartois,V., Baulard,A., Schanck,K. and Colson,C. (1992) *Biochim. Biophys. Acta* **1131**, 253–260.
160 Yamamoto,J., Shimizu,M. and Yamane,K. (1989) *Nucleic Acids Res.* **17**, 10105–10105.
161 Chen,L., James,L.P. and Helmann,J.D. (1993) *J. Bacteriol.* **175**, 5428–5437.
162 Gollnick,P., Ishino,S., Kuroda,M.I., Henner,D.J. and Yanofsky,C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8726–8730.
163 Yoshimoto,T., Oyama,H., Takeshita,T., Higashi,H., Xu,S. and Tsuru,D. (1991) *J. Ferment. Bioengng* **70**, 370–375.
164 Tran,L., Wu,X.-C. and Wong,S.-L. (1991) *J. Bacteriol.* **173**, 6364–6372.
165 Struck,J.C., Far,R.K., Schroder,W., Hucho,F., Toschka,H.Y. and Erdmann,V.A. (1990) *Biochim. Biophys. Acta* **1050**, 80–83.
166 Slock,J., Stahl,D.P., Han,C.-Y., Six,E.W. and Crawford,I.P. (1990) *J. Bacteriol.* **172**, 7211–7226.
167 Awade,A., Cleuziat,P., Gonzales,T. and Robert-Baudouy,J. (1992) *FEBS Lett.* **305**, 67–73.
168 Hemila,H.O., Palva,A., Paulin,L., Arvidson,S. and Palva,I. (1990) *J. Bacteriol.* **172**, 5052–5063.
169 Nasser,W., Awade,A.C., Reverchon,S. and Robert-Baudouy,J. (1993) *FEBS Lett.* **335**, 319–326.
170 van Dijl,J.M., de Jong,A., Nauta,A., Venema,G. and Bron,S. (unpublished) Z35652.
171 Brakhage,A.A., Wozny,M. and Putzer,H. (1990) *Biochimie* **72**,725–734.
172 Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H. (1987) *J. Bacteriol.* **169**, 2913–2916.
173 Fischer,C. and Deutscher,J. (unpublished) X79388.
174 Okada,M., Matsuzaki,H., Shibuya,I. and Matsumoto,K. (1994) *J. Bacteriol.* **176**, 7456–7461.
175 Honeyman,A.L. and Stewart,G.C. (1989) *Mol. Microbiol.* **3**, 1257–1268.
176 Henkin,T.M., Moon,S.H., Mattheakis,L.C. and Nomura,M. (1989) *Nucleic Acids Res.* **17**, 7469–7486.
177 Price,C.W. (unpublished) L24376.
178 Struck,J.C., Vogel,D.W., Ulbrich,N. and Erdmann,V.A. (1988) *Nucleic Acids Res.* **16**, 2719.
179 Sadaie,Y., Takamatsu,H. and Yamane,K. (1991) *Gene* **98**, 101–105.
180 Wang,L.-F. and Doi,R.H. (1990) *J. Bacteriol.* **172**, 1939–1947.
181 Gaur,N.K., Dubnau,E. and Smith,I. (1986) *J. Bacteriol.* **168**, 860–869.
182 van Dijl,J.M., de Jong,A., Vehmaanpera,J., Venema,G. and Bron,S. (1992) *EMBO J.* **11**, 2819–2828.
183 Rudner,D.Z., LeDeaux,J.R., Ireton,K. and Grossman,A.D. (1991) *J. Bacteriol.* **173**, 1388–1398.
184 Mauël,C., Young,M. and Karamata,D. (1991) *J. Gen. Microbiol.* **137**, 929–941.
185 Pang,A.S.-H., Nathoo,S. and Wong,S.-L. (1990) *J. Bacteriol.* **173**, 46–54.
186 Putzer,H., Gendron,N. and Grunberg-Manago,M. (1992) *EMBO J.* **11**, 3117–3127.
187 Iwakura,M., Kawata,M., Tsuda,K. and Tanaka,T. (1988) *Gene* **64**, 9–20.
188 Chow,K.-C. and Wong,J.T.-F. (1988) *Gene* **73**, 537–543.
189 Furbass,R. and Marahiel,M.A. (1991) *FEBS Lett.* **287**, 153–156.
190 Putzer,H. and Leautey,J. (unpublished) X77239.
191 Harley,C.B. and Reynolds,R.P. (1987) *Nucleic Acids Res.* **15**, 2343–2361.
192 Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.* **11**, 2237–2255.
193 Lisser,S. and Margalit,H. (1993) *Nucleic Acids Res.* **21**, 1507–1516.
194 McClure,W.R. (1985) *Annu. Rev. Biochem.* **54**, 171–204.
195 Travers,A.A. (1987) *CRC Crit. Rev. Biochem.* **22**, 181–219.
196 Fredrick,K., Caramori,T., Chen,Y.F., Galizzi,A. and Helmann,J.D. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2582–2586.
197 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.

198 Ross,W., Gosink,K.K., Salomon,J., Igarishi,K., Zou,C., Ishihama,A., Severinov,K. and Gourse,R.L. (1993) *Science* **262**, 1407–1413.
199 Ellinger,T., Behnke,D., Knaus,R., Bujard,H. and Gralla,J.D. (1994) *J. Mol. Biol.* **239**, 466–475.
200 Lavigne,M., Herbert,M., Kolb,A. and Buc,H. (1992) *J. Mol. Biol.* **224**, 293–306.
201 Pérez-Martín,J., Rojo,F. and deLorenzo,V. (1994) *Microbiol. Rev.* **58**, 268–290.
202 Knaus,R. and Bujard,H. (1988) *EMBO J.* **7**, 2919–2923.
203 Kammerer,W., Deuschle,U., Gentz,R. and Bujard,H. (1986) *EMBO J.* **5**, 2995–3000.
204 Ellinger,T., Behnke,D., Bujard,H. and Gralla,J.D. (1994) *J. Mol. Biol.* **239**, 455–465.
205 Wiggs,J.L., Bush,J.W. and Chamberlin,M.J. (1979) *Cell* **16**, 97–109.
206 Lisser,S. and Margalit,H. (1994) *Eur. J. Biochem.* **223**, 823–830.
207 Henkin,T.M. and Sonenshein,A.L. (1987) *Mol. Gen. Genet.* **209**, 467–474.
208 Graves,M.C. and Rabinowitz,J.C. (1986) *J. Biol. Chem.* **261**, 11409–11415.
209 Moran,C.P.,Jr, Lang,N., LeGrice,S.F.J., Lee,G., Stephens,M., Sonenshein,A.L., Pero,J. and Losick,R. (1982) *Mol. Gen. Genet.* **186**, 339–346.
210 Banner,C.D.B., Moran.,C.P.,Jr and Losick,R. (1983) *J. Mol. Biol.* **168**, 351–365.
211 McAllister,C.F. and Achberger,E.C. (1989) *J. Biol. Chem.* **264**, 10451–10456.
212 Benson,D.A., Boguski,M., Lipman,D.J. and Ostell,J. (1994) *Nucleic Acids Res.* **22**, 3441–3444.
213 Mulligan,M.E., Hawley,D.K., Entriken,R. and McClure,W.R. (1984) *Nucleic Acids Res.* **12**, 789–800.
214 McGhee,J.D. and von Hippel,P.H. (1974) *J. Mol. Biol.* **86**, 469–489.
215 O'Neill,M.C. (1989) *J. Biol. Chem.* **264**, 5522–5530.
216 Keilty,S. and Rosenberg,M. (1987) *J. Biol. Chem.* **262**, 6389–6395.
217 Chan,B., Spassky,A. and Busby,S. (1990) *Biochem. J.* **270**, 141–148.
218 Belyaeva,T., Griffiths,L., Minchin,S., Cole,J. and Busby,S. (1993) *Biochem. J.* **296**, 851–857.
219 Burns,H. and Minchin,S. (1994) *Nucleic Acids Res.* **22**, 3840–3845.
220 Sarai,A., Mazur,J., Nussinov,R. and Jernigan,R.L. (1988) *Biochemistry* **27**, 8498–8502.
221 Peticolas,W.L., Wang,Y. and Thomas,G.A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2579–2583.
222 Matern,H.T., Klein,J.R., Henrich,B. and Plapp,R. (1994) *FEMS Microbiol. Lett.* **122**, 121–128.
223 Crothers,D.M., Haran,T.E. and Nadeu,J.G. (1990) *J. Biol Chem.* **265**, 7093–7096.
224 Strauch,M.A. and Ayazifar,M. (1995) *Mol. Gen. Genet.* **246**, 756–760.
225 Frizby,D. and Zuber,P. (1991) *J. Bacteriol.* **173**, 7557–7564.