Volume 15 Number 5 1987 Nucleic Acids Research

Analysis of *E. coli* promoter sequences

Calvin B.Harley* and Robert P.Reynolds

Department of Biochemistry, McMaster University, Hamilton, Ontario L8N 3Z5, Canada

Received August 15, 1986; Revised and Accepted January 26, 1987

ABSTRACT

We have compiled and analyzed 263 promoters with known transcriptional start points for E. coli genes. Promoter elements (-35 hexamer, -10 hexamer, and spacing between these regions) were aligned by a program which selects the arrangement consistent with the start point and statistically most homologous to a reference list of promoters. The initial reference list was that of Hawley and McClure (Nucl. Acids Res. 11, 2237-2255, 1983). Alignment of the complete list was used for reference until successive analyses did not alter the structure of the list. In the final compilation, all bases in the -35 (TTGACA) and -10 (TATAAT) hexamers were highly conserved, 92% of promoters had inter-region spacing of 17±1 bp, and 75% of the uniquely defined start points initiated 7±1 bases downstream of the -10 region. The consensus sequence of promoters with inter-region spacing of 16, 17, or 18 bp did not differ. This compilation and analysis should be useful for studies of promoter structure and function and for programs which identify potential promoter sequences[1].

INTRODUCTION

Promoters are DNA sequences which affect the frequency and location of transcription initiation through interaction with RNA polymerase (1,2). Two conserved regions about 35 and 10 base pairs (bp) upstream from the transcription start (-35 and -10 regions, respectively) were identified by comparison of relatively few promoters (3-6). More extensive compilations and comparisons of promoters for genes of E. coli and its phage and plasmids supported and extended the concept of a "consensus" promoter sequence: a -35 (TTGACA) and -10 (TATAAT) region separated by 17 bp with transcription initiating at a purine about 7 bp downstream from the 3' end of the -10 region (7-9). While the -35 and -10 regions show the greatest conservation across promoters and are also the sites of nearly all mutations which affect transcriptional strength, other bases flanking the -35 and -10 regions, in addition to the start point also occur at greater than random frequencies and sometimes affect promoter activity (9-12). In addition, variation in spacing between the -35 and -10 regions plays a role in promoter strength (13-16).

Promoter compilations and analyses have led to computer programs which

predict the location of promoter sequences on the basis of homology either to the consensus sequence or to a reference list of promoters (17-19). Such programs are of practical significance in searching new sequences (2,20); thus promoter compilations are important beyond providing data regarding promoter structure. However, current compilations are based on sequences aligned by eye in attempts to maximize homology to the consensus sequence. Unfortunately, sequences closer to the consensus sequence may be missed thus weakening the homology between promoters and consequently reducing the predictive power of algorithms. Although promoter elements can be identified by biochemical or genetic evidence that pin-point bases which interact with RNA polymerase, such data is unavailable for most genes.

We have updated the compilation of E. coli promoter sequences and have reiteratively aligned them on the basis of a computer program which finds the sequence with greatest homology to the reference set. This compilation and reanalysis of 263 promoters should be useful in studies of promoter structure and function and in promoter search algorithms.

METHODS

Promoter Compilation

The starting point for analyses described below was the Hawley and McClure (9) compilation of 112 E. coli promoters with known transcriptional start points. Three resources were used to extend and update this compilation: Index Medicus, Dialog, and the National Institutes of Health GENBANK database on the National Biomedical Research Foundation Protein Identification Resource. Following Hawley and McClure, only promoters in which a transcriptional startpoint has been identified by biochemical or genetic means are used in the analysis. We included promoters whose start points were identified by S1 nuclease mapping (21) if additional evidence such as high resolution in vitro transcript run-off size or the site of polymerase binding supported the S1 data.

Analysis

DNA sequences from about -50 to +10, with respect to known transcriptional start points for genes of E. coli and its plasmids and phage were analyzed for promoter signals by a modification of the algorithm described by Staden (19). This algorithm utilizes the frequency of all bases at each position in the conserved areas of the promoter and therefore derives near maximal information about the similarity of any test sequence to the reference set of sequences. In brief, the test sequence is analyzed in all possible alignments of promoter

elements to determine the arrangement of -35 and -10 elements which maximizes similarity to known promoters on a strictly statistical basis. Each alignment yields a "promoter homology index" (PHI) derived from the weight matrix of the reference set of promoters. The weight matrix contains log frequencies for each base at each position in the -35 and -10 hexamers and log frequencies of the occurrence of -35 and -10 hexamers separated by 15-21 base pairs. PHI for a given alignment is the sum of log frequencies taken from the weight matrix for the elements of the test sequence. Staden's algorithm has been shown to be operationally similar in prediction of promoter strength to an alternative algorithm of Mulligan et al. (18) which includes data on cumulative deviations from the consensus sequence (20). We chose Staden's algorithm because it seemed less arbitrary in assessment of homology.

Our program finds for each DNA sequence the 10 (or more) highest ranking alignments of all possible -35 and -10 hexamers with a spacing of 15-21 base pairs, and flags those consistent with the transcription start data. A promoter sequence was deemed consistent with start data when the initiation point was between 4 and 12 bases from the -10 hexamer (see Results and Discussion).

The initial weight matrix was derived from the compilation and promoter alignment of Hawley and McClure (9). Null frequencies were replaced by the reciprocal of the number of entries in the weight matrix at that point to avoid complete exclusion of certain bases in, or spacing between, the -35 and -10 regions (19). Following analysis of the new promoter compilation, the weight matrix was updated using new alignments. This process was repeated until consecutive reiterations yielded identical highest ranking promoters for each sequence. To avoid chance fixation on extreme patterns in the weight matrix, frequencies were periodically smoothed artificially by reducing the frequencies of highly "conserved" bases and increasing the frequencies of highly excluded bases. This procedure was repeated on several promoter lists, including subdivisions of all promoters with 16, 17, or 18 bp spacing between the -35 and -10 regions.

RESULTS and DISCUSSION

Promoter Compilation

Table 1 shows 288 *E. coli* promoters aligned by reiterative application of the modified algorithm of Staden (19) (Methods). Although most of these promoters are wild type bacterial, plasmid, or phage promoters (type "b", "p", "f", column b, respectively), some mutant promoters (type "M" or "m", column b)

are also included. Mutations which generate an entirely new promoter (type "M") are included among 263 promoters with known transcription start points used for analyses as described below. Mutants of naturally occurring promoters (type "m") are not; transcription start data are often not available for these mutants and their inclusion would bias the weight matrix for base frequencies at the non-mutated positions. The list includes 112 promoters compiled by Hawley and McClure (9), which can be identified by reference "9" in column j. Analysis of these promoters separately or together with additional E. coli promoters yielded essentially identical results.

The algorithm makes no use of previously identified -35 and -10 regions for a given promoter; it identifies the statistically best -35 and -10 regions consistent with transcription start data using the weight matrix of 263 promoters listed in Table 1. Columns (c), (d), and (e) indicate the stable alignment of -35 and -10 regions and the spacing between them. Column (f) gives the relative promoter homology index (PHI) of the selected -35 and -10 regions: this value is the sum of the appropriate weight matrix values for each base in the -35 and -10 hexamers, plus the value for their spacing, minus the unnormalized index value of the consensus sequence (TTGACA...17...TATAAT). PHI values are from a logarithmic scale and can be interpreted loosely in terms of probability: for example, PHI = 0 indicates that the promoter elements are identical to consensus sequence elements, i.e. the most probable arrangement of bases and spacing, while PHI = -2 indicates that the probability of occurrence of bases in these regions and the spacing between them is theoretically 100 times smaller than that of the consensus sequence. Such interpretations may not be justified since they assume that gap penalties and bases at each position are independent and that these are the only conserved elements in promoter structure. Interestingly, a correlation exists between promoter strength and homology index (18). Thus promoter strength generally decreases as PHI values become more negative. Some promoters, however, do not follow this generalization (11,12).

Column (g) signals significant discrepancies between the best promoter alignment consistent with the transcription start data and the overall best alignment (indicated with double underlines) independent of transcription start data. The number in this column is the PHI value of the overall best alignment. Only discrepancies in PHI greater than 0.5 are shown. Column (h) signals discrepancies between published -35 and -10 regions (single underlines) and those selected by our analysis. The number in this column is the PHI value of the published alignment. These PHI values will be less negative than that in

TABLE 1

Alignment of E. coli Promoter Sequences

SEQUENCE (a)	TYPE (b)	-35 (c)	-10 (d)	SP (e)	PHI (f)	DISCREP. (g)	DISCREP. (h)	TS (i)	REF (j)
aceEF	b	ACGTAGACCTGT CTTATT GAGCTTTC	CGGCGACAG TTCAAT GGGaCAGGTCCAG	17	-4.3		-4.4	4	24
ada	b	AAGATTGTTGGTTT TTGCGT GATGGTGA	CCGGGCAGC CTAAAG GCTaTCGTTAACC	17	-5.5	-3.4	-4.6	4	25,26
alaS	b	AACGCATACGGTAT TTTACC TTCCCAGTC	AAGAAAACT TATCTT ATTCCCaCTTTTCAGT	18	-3.1				9
ampC	b	TGCTATCCTGACAG TTGTCA CGCTGATT	GGTGTCGT TACAAT GTAACGCaTCGCCAATG	16	-1.5				9
ampC/C16	b	GCTATC TTGACA GTTGTCAC	GCTGATTGG TATCGT TACAATCTaACGTATCG	17	-1.3			1,3	25
araBAD	b	TTAGCGGATCCTAC CTGACG GTTTTTAT	CGCAACTC TCTACT GTTTCTCCATaCCCGTT	16	-3.6		-3.7		9
araC	b	GCAAATAATCAATG TGGACT TTTCTGCC	GTGATTATA GACACT TTTGTTACgCGTTTTTG	17	-3.6				9
araE	b	CTGTTTCCGAC CTGACA CCTGCGTGA	GTTGTTCACG TATTTT TTCACTATgTCTTACTC	19	-3.2			4	28
araI(c)	m	AGCGGATCCTAC CTGCCG CTTTTTAT	CGCAACTC TCTACT GTTTCTCCATaCCCGTTT	16	-4.3			4	29
araI(c)X(c)	m	AGCGGATCCTAC CTGCCG CTTTTTATC	GCAACTCTC TACTAT TTCTCCATaCCCGTTTT	18	-3.8			4	29
argCBH	b	TTTGTTTTTCATTG TTGACA CACCTCTGG	TCATGATAG TATCAA TATTCaTCCAGTATT	18	-2.4		-2.6		9
argCBH-P1/6-	m	TTTGTTTTTCATTG TTGACA CACCTCT	GGTCATAA TATTAT CAATATTCaTGCAGTAT	15	-2.0				30
argCBH-P1/LL	m	TTTGTTTTTCATTG TTGACA CACCTCT	GGTCATGA TATTAT CAATATTCaTGCAGTAT	15	-2.0				30
argE-P1	b	TTACGGCTGGTGGG TTTTAT TACGCTCA	ACGTTAGTG TATTTT TATTCaTAAATACTGCA	17	-2.6			4	31
argE-P2	b	CCGCATCATTGCTT TCCCCT GAAACAGT	CAAAGCGGT TATGTT CATaTGCGGATGGCG	17	-3.9		-3.9	4	31
argE/LL13	m	CCGCATCATTGCTT TCGCCT GAAACAGT	CAAAGCGGT TATATT CATaTGCGGATGGCG	17	-3.3				31
argF	b	ATTGTGAAATGGGG TTGCAA ATGAATAA	TTACACATA TAAAGT GAATTTTaATTCAATAA	17	-1.7			4	31,32
argI	b	AGAC TTGCAA ATGAATAA	TCATCCATA TAAATT GAATTTTaaTTCATTGA	17	-1.5			4	31
argR	b	TCGTCGCCCGCG TTGCAG GAGCAAGG	GTTTGACAA TATTAA TCAGTCTaaaGTCTCGG	17	-3.2		-5.9	2,4	31
aroF	b	TACGAAAATATGGA TTGAAA ACTTTACT	TTATGTGT TATCGT TACGTCaTCCTCGCTG	16	-1.9			2,4	33
aroG	b	AGTGTAAAACCCCG TTTACA CATTCTGA	CGGAAGATA TAGATT GGAAGTaTTGCATTCA	17	-1.6			2,4	33
aroH	b	GTACTAGAGAACTA GTGCAT TAGCTTAT	TTTTTTGT TATCAT GCTAaccaCCCGGCCAG	16	-3.1				9
bioA	b	GCCTTCTCCAAAAC GTGTTT TTTGTTGTT	AATTCGGTG TAGACT TGTaaaCCTAAATCT	18	-3.8	-3.4			9
bioB	b	TTGTCATAATCGAC TTGTAA ACCAAATT	GAAAAGATT TAGGTT TACAAGTctACACCGAA	17	-2.2				9
bioP98	M	TTGTTAATTCGGTG TAGACT TGTAAACC	TAAATCTTT TAAATT TGGTTaCAAGTCGAT	17	-2.0				9
C62.5-P1	b	CACCTCGTCTCGC TTGAAA TTATTCTC	CCTTGTCCC CATCTC TCCCAcatCCTGTTTT	17	-3.3		+	4	34
carAB-P1	b	ATCCCGCATTAAG TTGACT TTTAGCGC	CCATATGTC CAGAAT GCCGCCgTTTGCCAGA	17	-1.9			4	35
carAB-P2	b	TAAGCAGATTTGCA TTGATT TACGTCATC	ATTGTGAAT TAATAT GCAAaTAAAGTGAG	18	-2.4			4	35
cat	b	ACGTTGATCGGC ACGTAA GAGGTTCC	AAGTTTCAC CATAAT GAAATAAgATCACTACC	17	-4.2	-2.4	-5.3		9
cit.util-379	p	AAACAGGCCGGG GTCTCA GGCGACTAA	CCCGCAAAC TCTTAC CTCTATACaTAATTCTG	18	-5.6	-5.2	%	3,4	36-38
cit.util-431	p	GACAGGCACAGCA TTGTAC GATCAACTG	ATTTGTGCC AATAAT TAaaTGAAATCAC	18	-3.4			3,4	36-38
ClaDFclocain	p	TCATATATTGACAC CTGAAA ACTGGAGG	AGTAAGGT AATAAT CATACTgTGTATATAT	16	-2.9	-1.5	-3.5	3	39
ClaDFrnaI	p	ACACCCGGTTGCTC TTGAAG TGTCCGCCA	AAGTCCGGC TACACT GGAAGGACAGATTTGG	18	-2.2				9
colE1-B	p	TTATAAAATCCTCT TTGACT TTTAAAA	CAATAAGT TAAAAA TAAaTACTGTAA	15	-3.4		-4.4	1,3	40
colE1-C	p	TTATAAAATCCTCT TTGACT TTTAAAAC	AATAAGTT AAAAAT AAATACTgTACATATAA	16	-2.4			1,3	40
ColE1-P1	p	GGAAGTCCACAGTC TTGACA GCGAAAAT	GCAGCGCCG TAGCTT TTATGCTgTATATAAAA	17	-1.7				9
ColE1-P2	p	TTTTTAACTTATTG TTTTAA AAGTCAAA	GAGGATTT TATAAT GGAAACCgCGGTAGCGT	16	-1.7		-1.9		9
colE110.13	p	GCTACAGAGTTC TTGAAG TAGTGGCCC	GACTACGGC TACACT AGAAGGACaGTATTTGG	18	-2.2			1,3	41
colicinE1 P3	p	TTTTTAACTTATTG TTTTAA AAGTCAAA	GAGGATTT TATAAT GGAAACCgCGGTAGCGT	16	-1.7				42
crp	b	AAGCGAGACACCAG GAGACA CAAAGCGA	AAGCTATGC TAAAAC AGTCAGgATGCTACAG	17	-3.2		%	2,3	43
cya	b	GTAGCGCATCTTTC TTTACG GTCAATCA	GCAAGGTGT TAAATT GATCACgTTTTACACC	17	-1.8			1-3	44
dapD	b	AAGTGCATCAGCGG TTGACA GAGCCCCTC	AATCCAAAC GATAAA GGGTGatgTGTTTACTG	18	-2.8			4	45
deo-P1	b	CAGAAACGTTTTA TTCGAA CATCGATCT	CGTCTTGTGT TAGAAT TCTAACaTACGGTTGC	19	-3.5				9
deo-P2	b	TGATGTGTA TCGAAG TGTGTTGCG	GAGTAGATGT TAGAAT ACTAACaAACTCGCAA	19	-3.9				9
deo-P3	b	ACACCAACTGTCTA TCGCCG TATCAGCG	AATAACGG TATACT GATCTGaTCATTTAAA	16	-3.2			2,4	46
divE	b	AAACAAATTAGGGG TTTACA CGCCGCAT	CGGGATGTT TATAGT GCGCGTCaTTCCGGAAG	17	-1.2			1,2	47
dnaA-1p	b	TGCGGCGTAAATCG TCCCCG CCTCGCGGC	AGGATCGTT TACACT TAGCGAgTTCTGGAAA	18	-4.4		-4.9	4	48,49
dnaA-2p	b	TCTGTGAGAAACAG AAGATC TCTTCCGC	AGTTTAGGC TATGAT CCGcggtccGGATCG	17	-4.5			4	48
dnaK-P1	b	TTTGCATGTCCCCC TTGATG ACGTGGTTT	ACGACCCCA TTTAGT AGTcaaCCGCAGTG	18	-3.2		-8.2	2,4	34
dnaK-P2	b	ATGAAATTGGGCAG TTGAAA CCAGACGT	TTCGCCCC TATTAC AGACtccCAACCACA	16	-2.4		-9.3	2,4	34
dnaQ-P1	b	GCCAGCGCTAAACG TTTTCT CCGGTCCG	CGATAGCG TAAAAT AGCgccGTAACCCC	16	-2.1			2-4	50,51
Fpla-oriTpK	p	GAACCACCAACCTG TTGAGC CTTTTTGT	CGAGTGGGT TAAATT ATTDaCGGATAAAG	17	-2.5			2	52
Fplas-traM	p	ATTAGGGGTGCTGC TAGCGG CCCCGTGT	GTTTTTTTA TAGGAT ACCGCTaCGGGCCGCTG	17	-4.0		-5.7	2	52
Fplas-traY/Z	p	GCGTTAATAAGGT GTTAAT AAAATATA	GACTTTCCG TCTATT TaccttttctgaTTATT	17	-3.9	-3.0	-4.1	3	53
frdABCD	b	GATCTCGTCAA ATTTCA GACTTATC	GATCAGAC TATACT GTTGTACCTATaAAGGA	16	-3.2		-3.9	4	54
fumA	b	GTACTAGTCTCAGT TTTTGT TAAAAAG	TGTGTAGGA TATTGT TACTCGCTtttAACAGG	17	-3.5		-3.8	4	55-57
λ-δ-tnpA	p	ACACATTAACAGCA CTGTTT TTATGTGT	GCGATAATT TATAAT ATTTCGgACCGTTGCA	17	-2.4				9
λ-δ-tnpR	p	ATTCATTAACAAT TTTGCA ACCGTCCG	AAATATTA TAAATT ATCGCACaCATAAAAAC	16	-2.4		-3.0		9
gal-P1	b	TCCATGTCACAGTT TTGGCA TCTTTGTT	ATGCTATGG TTATTT CATaCCATAAG	17	-3.8	-2.9	-4.0		9
gal-P2	b	CTAATTTATTCCAT GTCACA CTTTTCGC	ATCTTTGT TATGCT ATGGTTaTTTCATACC	16	-2.9		-3.1		9
gal-P2/mut-1	m	TAATTTATTCCAT GTCACA CTTTTCGC	ATCTTTGT TATACT ATGGTTaTTTCATAC	16	-2.3		-4.0	3	58

gal-P2/mut-2	m	TAATTTATTCCAT GTCACA CTTTTCGC	ATTTTTGT TATGCT ATGGTTaTTTCATAC	16	-2.9			3	58
glnL	b	CAATTCTCTGATGC TTGGCG CTTTTTATC	CGTAAAAAGC TATAAT GCACTaAATGGTGC	19	-3.2			2,4	59
glnS	b	TAAAAAACTAACAG TTGTCA GCCTGTCC	CGCTATAAA GATCAT ACGCCgttaTACGTT	17	-2.1				9
gltA-P1	b	ATTCATTCGGGACA GTTATT AGTCGTAG	ACAAGTTT AATAAT TCGGAtTGCTAAGTA	16	-4.3	-4.4		4	57,60
gltA-P2	b	AGTTGTTACAAACA TTACCA GCAAAAGCA	TATAATGCG TAAAAG TTAtGAAGTCGGT	18	-4.0	-1.8	-2.5	4	57,60
glyA	b	TCGTTTGTCAAGAC CTGTTA TCGCACAA	TGATTCGGT TATACT GTTCgCCGTTGTCC	17	-2.4			2,4	63
glyA/geneX	b	ACACCAAAGAACCA TTTACA TTGCAGCG	CTATTTTTTA TAAGAT GCATTcGAGATACAT	18	-1.9			2,4	61
gnd	b	GCATGGATAAGCTA TTTATA CTTTAATA	AGTACTTTG TATACT TATTTGCgAACATTCCA	17	-1.7			4	62
groE	b	TTTTTCCCCC TTGAAG GGGCGAAG	CCATCCCCA TTTCTC TGGTCaCCAGCCCGCAA	17	-3.9	+		4	34
gyrB	b	CGGACGAAAA TTCGAA GATGTTACCGTGGAAAAGGG	TAAAAT AACGGATcAACCCAAGT	21	-3.2			4	63
his	b	ATATAAAAAGTTC TTGCTT TCTAACGTG	AAAGTCGTT TACGTT AAAAGACaTCAGTTGAA	18	-3.6				9
hisA	b	GATCTACAAACTAA TTAATA AATAGTTA	ATTAACGCT CATCAT TGTACAATGAaCTGTAC	17	-3.5	-2.7	-5.7		9
hisBp	b	CCTCCAGTGCGGTG TTTAAA TCTTTGTG	GGATCAGCG CATTAT CTTacGTGATCAG	17	-2.4			2,4	64
hisJ(St)	b	TAGAATGCTTTGCC TTGTCG GCCTGATT	AATGCCAC GATAGT CGCATCCGATCTG	16	-3.0	-3.6			9
hisS	b	AAATAATAACGTGA TGGGAA GCGCCTCG	CTTCCCGTG TATGAT TGAACccgCATGGCTC	17	-2.7			4	65,66
htpR-P1	b	ACATTACGCCACTT ACGCCT GAATAATA	AAAGCGTGT TATACT GTTTCCtGCAATGGTT	17	-3.8			4	67,68
htpR-P2	b	TTCACAAGCTTGCA TTGAAC TTGTGGATA	AAATCACGG TCTGAT AAAACAgTGAATG	18	-3.7	-2.3		4	67,68
htpR-P3	b	AGCTTGCATTGAAC TTGTGG ATAAAATC	ACGGTCTGA TAAAAC AGTGAATgATAACCTCGT	17	-3.2			4	67,68
ilvGEDA	b	GCCAAAAAATATCT TGTAGT ATTTACAA	AACCTATGG TAACTC TTTAGGCaTTCCTTGGA	17	-4.6	-3.9	-4.6		9
ilvIH-P1	b	CTCTGGCTGCCAA TTGCTT AAGCAAGA	TCGGACGGT TAATGT GTTttacacattttcTTC	17	-3.2			2,4	69
ilvIH-P2	b	GAGGATTTTATCGT TTCTTT TCACCTTT	CCTCCTGTT TATTCT TATcACCCCGTGT	17	-3.1		-3.1	2,4	69
ilvIH-P3	b	ATTTTAGGATTAA TTAAAA AAATAGAG	AAATTGCTG TAAGTT GTGGGATTcAGCCGATT	17	-2.7			2,4	69
ilvIH-P4	b	TGTAGAATTTTATT CTGAAT GTCTGGGC	TCTCTATTT TAGGAT TAATTAAaAAAATAGAG	17	-2.7			2,4	69
ISlins PL	p	CGAGCCCGGTGATG CTGCCA ACTTACTG	ATTTAGTG TATGAT GGTGcttcTTGAGGTGCT	16	-2.5			1,3,4	70
ISlins PR	p	ATATATACCTTA TGGTAA TGACTCCA	AGTTATTGA TAGTGT TTTATGTcCAGATAAT	17	-3.6	-3.3		1,3,4	70
IS2I-II	M	ATGTC TGGAAA TATAGGGG	CAAATCCAC TAGTAT TAAGACtaTCACTTATT	17	-2.6				9
lacI	b	GACACCATCGAATG GCGCAA AACCTTTC	GCGGTATGG CATGAT AGCGCCGgAAGAGAGT	17	-4.5				9
lacP1	b	TAGGCACCCCAGGC TTTACA CTTTATGCT	TCCGGCTCG TATGTT GTGTGGaATTGTGAGC	18	-2.0				9
lacP115	M	TTTACACTTTATG CTTCCG GCTCGTATG	TTGTGTGG TATTGT GAGcggatcaacaATTT	17	-3.9	-2.0	-4.2		9
lacP2	b	AATGTGAGTTAGCT CACTCA TTAGGCAC	CCCAGGCTT TACACT TTATGCcTCCGGCTCG	17	-4.0	-2.6	-4.3		9
lambdacl7	M	GGTGTATGCATTTA TTTGCA TACATTCA	ATCAATTGT TATAAT TGTTATcTAAGGAAAT	17	-1.4				9
lambdacin	M	TAGATAACAATTGA TTGAAT GTATGCAA	ATAAATGCA TACACT ATAGGTgTGGTTTAAT	17	-1.6				9
lambdaL57	M	TGATAAGCAATGC TTTTTT ATAATGCCA	AGTTAGTA TAAAAT AGCCAACcTGTTCGACA	17	-2.4		-2.5		9
lambdaPI	f	CGGTTTTTTCTTGC GTGTAA TTGCGGAG	ACTTTCCGA TGTAGT TGACACcTCAGGAGTG	17	-3.6				9
lambdaPL	f	TATCTGTGGCCGTG TTGACA TAAATACC	ACTGGCCGT GATACT GAGCACaTCAGCAGGA	17	-1.4				9
lambdaPo	f	TACCTCTGCCGAAG TTGAGT ATTTTTGC	TGTATTTGT GATAAT GACTCCTgTTGATAGAT	17	-2.1				9
lambdaPR	f	TAACACCGTGCGTG TTGACT ATTTTACC	TCTGGCCGT GATAAT GGTTGCaTGTACTAAG	17	-1.4				9
lambdaPR'	f	TTAACGGCATGATA TTGACT TATTGAAT	AAAATTGGG TAAATT TGACTCAaCGATGGGTT	17	-1.1				9
lambdaPRE	f	GAGCCTCGTTGCGT TTGTTT GCACGAACC	ATATGTAAG TATTTC CTTaGATAACAAT	18	-4.1		-5.7		9
lambdaPRM	f	AACACGCACGGTGT TAGATA TTTATCCC	TTGCCGTGA TAGATT TAACGTaTGAGCACAA	17	-2.6				9
lep	b	TCCTCGCCTCAATG TTGTAG TGTAGAAT	GCCGCGTT TCTATT AATAcaGACGTTAAT	16	-3.4			2,4	71
leu	b	G TTGACA TCCGTTTT	TGTATCCAG TAACTC TAAAAGCATATCGCATT	17	-2.5				9
leultRNA	b	TCGATAATTAACTA TTGAGG AAAAGCTG	AAAACCAC TAGAAT GCGCCTCCgTGGTAGCA	16	-1.5				9
lex	b	TGTGCAGTTTATGG TTCCAA AATCGCCT	TTTGCTGTA TATACT CACAGCaTAACTGTAT	17	-1.9				9
livJ	b	TGTCAAAATAGCTA TTCCAA TATCATAA	AAATCGGGA TATGTT TTAGCaGAGTATGCT	17	-2.5			1,4	67,68
lpd	b	TGTTG TTTAAA AATTGTTA	ACAATTTTG TAAAAT ACCGACGGAtagAACGA	17	-1.1			4	24,57
lpp	b	CCATCAAAAAAATA TTCTCA ACATAAAAA	ACTTTGTG TAATAC TTGTAACgCTACATGGA	17	-3.2		-3.3		9
lpp/P1	m	ATCAAAAAAATA TTCTCA ACATAAAAA	ACTTTGTGT TATACT TGTAACgCTACATGGA	18	-1.9				72
lpp/P2	m	ATCAAAAAAATA TTCTCA ACATAAAAA	ACTTTGTGT TATAAT TGTAACgCTACATGGA	18	-1.6				72
lpp/R1	m	ATCAAAAAAATA TTCACA ACATAAAAA	ACTTTGTG TAATAC TTGTAACgCTACATGGA	17	-2.7		-2.8		72
Mlrna	b	ATGCCCAACCCCGG GTGACA ACGCCCCG	CAAACCCTC TATACT GCGCCCCgAAGCTGACC	17	-1.2				9
mscl1	M	CCCCCCCAGGGAT GAGGAA GGTGGTGGA	CCGGGCTCG TATGTT GTGTGGaATTGTGAGC	18	-4.1			4	76
mscl2	M	CCCCCCCAGGGAT GAGGAA GGTCGGTCG	ACCGGCTCG TATGTT GTGTGGaATTGTGAGC	18	-4.1			4	76
msc21	M	CCCCCCCAGGGAT GAGGAA GGTCGACGT	TCCGGCTCG TATGTT GTGTGGaATTGTGAGC	18	-4.1			4	76
msc3	M	CCCCCCCAGGGAT GAGGAA GGTCGGTC	GACGGCTCG TATGTT GTGTGGAATTgTGAGCG	17	-3.7			4	76
msc31	M	CCCCCCCAGGGAT GAGGAA GGTCGGTC	GACGGCTCG TATATT GTGTGGAATTgTGAGCG	17	-3.1			4	76
malEFG	b	AGGGGCAAGGAGGA TGGAAA GAGGTTGC	CGTATAAA GAAACT AGAGTCCgTTTAGGTGT	16	-3.5				9
malK	b	CAGGGGGTGGAGGA TTTAAG CCATCTCC	TGATGACG CATAGT CAGCCCaTCATGAATG	16	-3.3				9
malPQ	b	ATCCCCCCAGGATG AGGAAG GTCAACAT	CGAGCCTGG CAAACT AGCGATaACGTTGTGT	17	-4.7			2	77
malPQ/A516P1	m	ATCCCCCCAGG ATGAGG AGCCTGGC	AAACTAGC GATGAT AACGTTGTGTTgAA	16	-4.6			2,4	78
malPQ/A516P2	m	ATCCCCCCAGGAGG ATGAGG AGCCTGGCA	AACTAGCGA TAACGT TGTGTTgAA	18	-4.6			2,4	78
malPQ/A517/A	m	CCCCCCAGGATGAG GTCGAG CCTGCCAA	ACTAGCGA TAACGT TGTGTTgAA	16	-4.9			2,4	78
malPQ/Ppl2	m	ATCCCCCCAGGAT GAGGAA GGTCAACAT	CGAGCCTG GAAAAC TAGCGATaACGTTGTGT	17	-5.2		-5.2		77
malPQ/Ppl3	m	ATCCCCCCAGGAT TAGGAA GGTCAACAT	CGAGCCTGG CAAACT AGCGATaACGTTGTGT	18	-3.9		-4.7		77
malPQ/Ppl4	m	ATCCCCCCAGGAT GAGGAA GGTCAACA	TCGAGCCTG GAAACT AGCGATaACGTTGTGT	17	-4.4				77
malPQ/Ppl5	m	ATCCCCCCAGGAT GAGAAA GGTCAACAT	CGAGCCTGG CAAACT AGCGATaACGTTGTGT	18	-4.0				77
malPQ/Ppl6	m	ATCCCCCCAGGATA AGGAAG GTCAACAT	CGAGCCTGG CAAACT AGCGATaACGTTGTGT	17	-4.7				77
malPQ/Ppl8	m	ATCCCCCCAGGATG GGGAAG GTCAACAT	CGAGCCTGG CAAACT AGCGATaACGTTGTGT	17	-4.3				77

malT	b	GTCATCGCTTGCAT	TAGAAA	GGTTTCTG	GCCGACGT	TATAAG	CATTAATTACG	16	-2.6	-3.9		9
manA	b	CGGCTCCAGGTTAC	TTCCCG	TAGCATTC	TTGCTTTAA	TAGTGG	GATTAATttCCACATTA	17	-5.0	-2.9 -2.9	4	56
metA P1	b	TTCAACATGCAGGC	TCGACA	TTGCCAAA	TTTTCTGGT	TATCTT	CAGCTaTCTGGATGT	17	-2.3		2,4	79
metA P2	b	AAGACTAATTACCA	TTTTCT	CTCCTTTT	AGTCATTGT	TATATT	CTAACGTaGTCTTTTCC	17	-1.8	-2.5	2,4	79
metBL	b	TTACCGTGACA	TCGTGT	AATGCACGT	GTCGCGCT	GATAAT	GCATATAActtTAACCG	17	-3.9	-3.3	2,4	80
metF	b	TTTTCGG	TTGACG	CCCTTCGG	CTTTTCCTT	CATCTT	TacaTCTGGACG	17	-2.5		2-4	81
micF	b	GCGGAATGGCGAAA	TAAGCA	CCTAACAT	CAAGCAAT	AATAAT	TCAAGGTcAAAATCAAT	16	-4.6	-2.9	2,4	82,83
motA	b	GCCCCAATCGCGCG	TTAACG	CCTGACGAC	TGAACATCC	TGTCAT	GGTCAaCAGTGGA	18	-4.5			84
MuPc-1	f	AAATT	TTGAAA	AGTAACTTTATAGAAAAGAAT	AATACT	GAAAAGTCAAtttGGTG		21	-3.3	-2.0 -4.0	2,4	85
MuPc-2	f	GGAACACA	TTTAAA	AACCCTCC	TAAGTTTTG	TAATCT	ATAAAGttAGCAATTTA	17	-2.1	-4.0	2,4	85
MuPe	f	TACCAAAAAGCACC	TTTACA	TTAAGCTT	TTCAGTAAT	TATCTT	TTTAGTaAGCTAGCTA	17	-1.7		2,4	85
NRlrnaC	p	GTCACAATTCTCAA	GTCGCT	GATTTCAAA	AAACTGTAG	TATCGT	CTGCgaaacGATCCCT	18	-4.1	-4.1	2-4	86
NRlrnaC/m	m	TCACAATTCTCAAG	TTGCTG	ATTTCAAA	AAACTGTAG	TATCGT	CTGCgaaacGATCCCT	17	-2.8			86
NTP1rna100	p	GGAGTTTGTC	TTGAAG	TTATGCACC	TGTTAAGGC	TAAACT	GAAAGAaCAGATTTTGT	18	-1.8			87
nusA	b	CAGTAT	TTGCAT	TTTTTACC	CAAAACGAG	TACAAT	TTGCCACgTTTCAGGCG	17	-1.8		1,3	88,89
ompA	b	GCCTGACGGAG	TTCACA	CTTGTAAG	TTTTCAAC	TACGTT	GTAGACtTTAC	16	-2.7	-2.0 -7.4	3,3	90
ompC	b	GTATCATATTCGTG	TTGGAT	TATTCTGC	ATTTTTGGG	GAGAAT	GGACcTGCCGACTG	17	-2.9		3,4	92,893
ompF	b	GGTAGG	TAGCGA	AACGTTAG	TTTGAATGG	AAACAT	GCCTGCaGACACATAAA	17	-4.6	-3.9	3,4	91
ompF/pKI217	m	GG	TAGCGA	AACGTTAG	TTTGCAAGC	TTTAAT	GCGGtaGTTTATCAC	17	-3.4	-2.6	3,4	91
ompR	b	TTTCGCCGAATAAA	TTGTAT	ACTTAAG	GTGCTGTT	TAATAT	GCTTTgTAACAATTT	15	-3.4	-2.4	4	92
pl5primer	p	ATAAGATGATCTTC	TTGAGA	TCGTTTTG	GTCTGCCGG	TAATCT	CTTGCTTgAAAAGCAAA	17	-2.1		1	93
pl5rnaI	p	TAGAGGAGTTAGTC	TTGAAG	TCATGCGCC	GGTTAAGGC	TAAACT	GAAAGGaCAAGTTTTG	18	-1.8		1	93
P22ant	f	TCCAAGTTAGTGTA	TTGACA	TGATAGAA	GCACTCTAC	TATATT	CTCAATaggTCCACGG	17	-0.4			9
P22mnt	f	CCACCGTGGACCTA	TTGAGA	ATATAGTA	GAGTGCTTC	TATCAT	GTCAATACACTAACTT	17	-1.5			9
P22PR	f	CATCTTAAATAAAG	TTGACT	AAAGATTC	CTTTAGTA	GATAAT	TTAAGTgTTCTTTAAT	16	-1.8			9
P22PRM	f	AAATTATC	TACTAA	AGGAATCT	TTAGTCAAG	TTTATT	TAAGATGACTTaACTAT	17	-3.7	-3.1 -3.9		9
pBR313Htet	m	AATTCTCATGT	TTGACA	GCTTATCA	TCGATAAGC	TAGCTT	TAATGCgGTAGTTTAT	17	-1.7		1,3	94
pBR322bla	p	TTTTTCTAAATACA	TTCAAA	TATGTATC	CCCTCATGA	GACAAT	AACCCTgATAAATGCT	17	-2.6			9
pBR322P4	p	CATCTGTGCGGTAT	TTCACA	CCGCATATGGTGCACTCTCAG		TACAAT	CTGCTCTgATGCCGCAT	21	-2.7			9
pBR322primer	p	ATCAAAGGATCTTC	TTGAGA	TCCTTTTT	TTCTGCGCG	TAATCT	GCTGCTTgCAAACAAAA	17	-2.1			9
pBR322tet	p	AAGAATTCTCATGT	TTGACA	GCTTATCA	TCGATAAGC	TTTAAT	GCGGTAgTTTATCACA	17	-1.0			9
pBRH4-25	M	TCG	TTTTCA	AGAATTCA	TTAATGCGG	TAGTTT	ATCAcagTTA	17	-2.7		4	95
pBRP1	p	TTCATACACGGTGC	CTGACT	CCGTTAGCAATTTAACTGTGA		TAAACT	ACCCCAttAAAGCTTA	21	-3.3			9
pBRRNAI	p	GTGCTACAGAGTTC	TTGAAG	TGGTGCCCT	AACTACGGC	TACACT	AGAAGGacaGTATTTG	18	-2.2			9
pBRtet-10	M	AAGAATTCTCATGT	TTGACA	GCTTATCA	TCGATCGGG	TAGTTT	ATCAcagTTA	17	-1.6		4	95
pBRtet-15	M	AAGAATTCTCATGT	TTGACA	GCTTATCA	TCGGTAGTT	TATCAC	AGTTAaatTGC	17	-1.8		4	95
pBRtet-22	M	AAGAATTCTCATGT	TTGACA	GCTTATCAT	CGATCACAG	TTAAAT	TGCTAacgCAG	18	-1.8		4	95
pBRtet/TA22	M	TTCTCATGT	TTGACA	GCTTATCA	TCGATAAGC	TAAATT	TTATATAaaATTTAGCT	17	-0.7		1	96
pBRtet/TA33	M	TTCTCATGT	TTGACA	GCTTATCA	TCGATAAGC	TAAATT	TATATAaaATTTTATAT	17	-0.7		1	96
pColViron-P1	p	TCACAATTCTCAAG	TTGATA	ATGAGAAT	CATTATTGA	CATAAT	TGTDaTTATTTTAC	17	-1.6		1,3,4	97
pColViron-P2	p	TGTTTCAACACC	ATGTAT	TAATTGTG	TTTATTTG	TAAAAT	TAATTTcctgacaATAA	16	-3.0		3,4	97
pEG3503	M	CGC	TGGACT	TCGAATTCA	TTAATGCGG	TAGTTT	ATCAcagTTA	18	-3.6		4	95
phiXA	f	AATAACCGTCAGGA	TTGACA	CCCTCCCA	ATTGTATGT	TTTCAT	GCCTCCaATCTTGGA	17	-1.7			9
phiXB	f	GCCAGTTAAATAGC	TTGCAA	AATACGTGG	CCTTATGGT	TACAGT	ATGCCCaTCGCAGTT	18	-2.6			9
phiXD	f	TAGAGATTCTCTTG	TTGACA	TTTTAAAAG	AGCGTGGAT	TACTAT	CTGAGTCCgATGCTGTT	18	-1.7			9
pori-I	b	CTGTTGTTCAGTTT	TTGAGT	TGTGTATA	ACCCCTCAT	TCTGAT	CCCAGCTTATACGGT	17	-3.2			9
Pori-r	b	GATCGCACGATCTG	TATACT	TATTTGAGT	AAATTAACC	CACGAT	CCCAGCCaTTCTTCTGC	18	-4.5			9
ppc	b	CGATTTCGCAGCAT	TTGACG	TCACCGCT	TTTACGTGG	CTTTAT	AAAAgaCGACGAAAA	17	-3.1		3,4	99
pSC101oriP1	p	T	TTGTAG	AGGAGCAAACAGCGTTTGCGA		CATCGT	TTTGTAATACTGCGGAA	21	-4.4	%	2,3	102,103
pSC101oriP2	p	ATTATCA	TTGACT	AGCCCATC	TCAATTGG	TATAGT	GATTAAAATCACCTAGA	16	-1.4	%	2,3	102,103
pSC101oriP3	p	ATACGCTCAGATGA	TGAACA	TCAGTAGG	GAAAATGCT	TATGGT	GTATTAGCTAAAGC	17	-3.6		2,3	102,104
pyrBl-P1	b	CTTTCACACTCCGC	CCTATA	AGTCGGAT	GAATGGAA	TAAAAT	GCATatcTGATTGCGTG	16	-4.2	-3.6	3	105
pyrBl-P2	b	TTGCATCAAATG	CTTGCG	CCGCTTGT	GACGATGG	TATAAT	GCCGgacAATTTGCCGG	17	-2.8		3	105
pyrD	b	TTGCCGCAGGTCAA	TTCCCT	TTTGGTCC	GAACTCGCA	CATAAT	ACgcccCGGTTTG	17	-2.6		3,4	106
pyrE-P1	b	ATGCCTTGTAAGGA	TAGGAA	TAACCGCC	GGAAGTCGG	TATAAT	GCGCAgCCACATTTG	17	-1.8		4	107,108
pyrE-P2	b	GTAGCCGGTCATA	CTGCGG	ATCATAGAC	GTTCCTGTT	TATAAA	AGGAGaGGTGGAAGG	18	-4.6		4	108
R100rna3	p	GTACCGGCTTACGC	CGGGCT	TCGGCGGTT	TTAGTCCTG	TATCAT	ATGAaACAACAGAG	18	-4.3			9
R100RNAI	p	CACAGAAAGAAGTC	TTGAAC	TTTTCCGG	GCATATAAC	TATACT	CCCCGCaTAGCTGAAT	17	-1.6			9
R100RNAII	p	ATGGGCTTACATTC	TTGAGT	GTTCAGAA	GATTAGTGC	TAGATT	ACTGATCgTTTAAGGAA	17	-2.2			9
R1RNAII	p	ACTAAAGTAAAGAC	TTTACT	TTGTGCCG	TAGCATGC	TAGATT	ACTGATCgTTTAAGGAA	16	-2.4			9
recA	b	TTTCTACAAAACAC	TTGATA	CTGTATGA	GCATACAG	TATAAT	TGCTTCaACAGAACAT	16	-1.1			9
rnh	b	GTAAGCGGTCATTT	ATGTCA	GACTTGTC	GTTTTACAG	TTCGAT	TcaaTTACAGGA	17	-4.0	-4.5	2,3,4	50,51
rnp(RNaseP)	b	ATGCGCAACGCGCG	GTGACA	AGGCCCCG	CAAACCCTC	TATACT	GCGCGCgAAGCTGACC	17	-1.2		1	109
rplJ	b	TGTAAACTAATGCC	TTTACG	TGGCCGGT	GATTTTGTC	TACAAT	CTTACcccCACGTATA	17	-1.8			9
rpmHlp	b	GATCCAGCAGCATC	CTTGCG	CTTTACCC	ATCAGCCCG	TATAAT	CCTccacccGGCGCG	17	-2.8	-2.9	4	48
rpmH2p	b	ATAAGGAAAGAGAA	TTGACT	CCGGAGTG	TACAATTAT	TACAAT	CCGgcctcTTTAATC	17	-1.0		4	48
rpmH3p	b	AAATTTAATGACCA	TAGACA	AAAATTGG	CTTAATCGA	TCTAAT	AAAgatcCCAGCAGG	17	-2.3		4	48
rpoA	b	TTCGCATATTTTTC	TTGCAA	AGTTGGGT	TGAGCTGGC	TAGATT	AGCCAgCCAATCTTT	17	-1.8			9

rpoS	b	CGACTTAATATACT	GCGACA	GGACGTCC	GTTCTGTG	TAAATC	GCAATGAAATGGTTTAA	16	-4.4				9
rpoD-Pa	b	CGCCCTGTTCCG	CAGCTA	AAACGCAC	GACCATGCG	TATACT	TATAgggTT	17	-3.5			2,4	110
rpoD-Pb	b	AGCCAGGT	CTGACC	ACCGGGCAA	CTTTTAGAG	CACTAT	CGTGGTACaaaT	18	-4.6		-5.9	2,4	110
rpoD-Phs	b	ATGCTGCCACCC	TTGAAA	AACTGTCG	ATGTGGGAC	GATA	TAGCAG	ATAAGAATATTgcT	17	-2.9			4	110
rpoD-Phs/stn	b	CCC	TTGAAA	AACTGTCG	ATGTGGGAC	GATA	TAGCAG	ATAAGAATATTgcT	21	-4.2	-2.9	-4.7	4	110
rrn4.5S	b	GGCACGCGATGGG	TTGCAA	TTAGCCGG	GGCAGCAGT	GATAAT	GCGCCTCCgCGTTCGTT	17	-1.9				1	111
rrnABP1	b	TTTTAAATTTCGTC	TTGTCA	GGCCGGAA	TAACTCCC	TATAAT	GCGCCACCaCTGACACG	16	-0.8					9
rrnABP2	b	GCAAAAATAAATGC	TTGACT	CTGTAGCG	CGAAGGCG	TATTAT	GCACAcccCGCGCCGC	16	-1.4					9
rrnB-P3	b	CTATGATAAGGAT	TACTCA	TCTTATCCTT	ATCAAACCGT	TAAAAT	GGGCGgtgTGAGCTTG	20	-4.1				2,4	112-114
rrnB-P4	b	GCCTATCCCGTCAC	CTCTCA	CCTGACA	GTTCGTGC	TAAAAT	AGCCAAccTGTTCGACA	15	-3.8				2,4	112-114
rrnDEXP2	b	CCTGAAATTCAGGG	TTGACT	CTGAAAGA	CGAAAGCG	TAATAT	ACGCCACCcTCGCGACAG	16	-1.7					9
rrnD-P1	b	GATCAAAAAAATAC	TTGTGC	AAAAAATT	GGGATCCC	TATAAT	GCGCCTCCgTTGAGACG	16	-2.7					9
rrnE-P1	b	CTGCAATTTTTCTA	TTGCGG	CCTGCGGA	GAACTCCC	TATAAT	CGCCCTCCaTCGACACG	16	-2.3					9
rrnG-P1	b	TTTATATTTTTCGC	TTGTCA	GGCCGGAA	TAACTCCC	TATAAT	GCGCCACCACTGACACG	16	-0.8					9
rrnG-P2	b	AAGCAAAGAAATGC	TTGACT	CTGTAGCG	CGAAGGCG	TATTAT	GCACACCGCCGCCGCG	16	-1.4					9
rrnXl	b	ATGCATTTTTCCGC	TTGTCT	TCCTGAGC	CGACTCCC	TATAAT	GCGCCTCCaTCGACACG	16	-1.2					9
RSFprimer	p	GGAATAGCTGTTCG	TTGACT	TGATAGAC	CGATTGATT	CATCAT	GTCATaAATAAAGAA	17	-2.0					9
RSFrnaI	p	TAGAGGAGTTTGTC	TTGAAG	TTATGCACC	TGTTAAGCC	TAAACT	GAAAGAaCAGATTTTG	18	-1.8					9
S10	b	TACTAGCAATACGC	TTGCGT	TCGGTGGT	TAAGTATG	TATAAT	GCGCgggGTTGTCGT	16	-2.2					9
sdh-P1	b	ATATGTAGGTTAA	TTGTAA	TGATTTTG	TGAACAGCC	TATACT	GCCGCCAGtcTCCCGGAA	17	-1.0				4	57,115
sdh-P2	b	AGCTTCCGCGATTA	TGGCCA	GCTTCTTC	GTCAAATT	TATCAT	GTGGGGCaTCGTTACCG	16	-2.9				4	57,115
spc	b	CCGTTTATTTTTTC	TACCCA	TATCCTTG	AAGCGGTGT	TATAAT	GCCGCgCGCCTCGATA	17	-2.2					9
spotA2r	b	TTACAAAAAGTGCT	TTGTGA	ACTGAACA	AAAAAGAG	TAAAGT	TAGTCCCgTAGCGTACA	16	-3.2		-3.3			9
ssb	b	TAGTAAAAGCGCTA	TTGGTA	ATGGTACAA	TCGCCGGTT	TACACT	TATTCAgAACGATTTT	18	-2.9					116,117
str	b	TCGTTGTATATTTC	TTGACA	CCTTTTCG	GCATCGCCC	TAAAAT	TCGGCgTCCTCATAT	17	-0.3					9
sucAB	b	AAATGCAGGAAATC	TTTAAA	AACTGCCCC	TGACACTAA	GACAGT	TTTAAAaGGTTCCTT	18	-3.6				4	22
supB-E	b	CGTTGAAAAAGAGG	TTGACG	CTGCAAGG	CTCTATACG	CATAAT	GCGCCCCgCAACGCCGA	17	-1.4					9
T7-A1	f	TATCAAAAAGAGTA	TTGACT	TAAAGTCT	AACCTATAG	GATACT	TACAGCCaTCGAGAGGG	17	-1.8					9
T7-A3	f	GTGAAACAAAACGG	TTGACA	ACATGAAG	TAAACAGCG	TACGAT	GTACCACaTGAAACGAC	17	-1.2					9
T7-C	f	CATTGATAAGCAAC	TTGACG	CAATGTTA	ATGGGCTCA	TAGTCT	TATCTTaCAGGTCATC	17	-2.1					9
T7-D	f	CTTTAAGATAGCCG	TTGACT	TGATGGT	CTTTAGGTG	TAGGCT	TTAGGTgTTGGCTTTA	17	-1.9					9
T7A2	f	ACGAAAAACAGGTA	TTGACA	ACATGAAGT	AACATGCAG	TAAGAT	ACAAATGgCTAGGTAAC	18	-1.3					9
T7E	p	CTTACGGATG	ATGATA	TTTACACA	TTACAGTGA	TATACT	CAAGCCCaCTACAGATA	17	-2.4				1,3	118
TAC16	M	AATGAGCTG	TTGACA	ATTAATCA	TCGGCTCG	TATAAT	GTGTGGAaTTGTG	16	-0.4					119,120
Tn10Pin	p	TCATTAAG	TTAAGG	TGGATACAC	ATCTTGTCA	TATGAT	CAAATGGTTTCgCGAAA	18	-3.5		-5.0			9
Tn10Pout	p	AGTGTAATTCGGGG	CAGAAT	TGGTAAAG	AGAGTCGTG	TAAAAT	ATCGAGttCGCACATC	17	-2.7					9
Tn10tetA	p	ATTCCTAATTTTTG	TTGACA	CTCTATCAT	TGATAGAGT	TATTTT	ACCACTCCCTATCAGT	18	-1.4					9
Tn10tetR*	p	TATTCATTTCACTT	TTCTCT	ATCACTGAT	AGCGAGTGG	TAAAAT	AACTCTATCAATGATA	18	-2.2					9
Tn10tetR*	p	TGATAGGGAG	TGGTAA	AATAACTC	TATCAATGA	TACAGT	GTCAACaAAAATTAGG	17	-3.0				4	122
Tn10xxxP1	p	TTAAAATTTTCTTG	TTGATG	ATTTTTAT	TTCCATGA	TAGATT	TAAAATaACATACC	16	-2.6				4	123
Tn10xxxP2	p	AAATGTTCTTAAGA	TTGTCA	CGACCACA	TCATCATGA	TACCAT	AAACaTACTGACGG	17	-1.8				4	123
Tn10xxxP3	p	CCATGATAGA	TTTAAA	ATAACATA	CCGTCAGTATGTT	TATGGT	ATCATGATGaTGTGGTC	21	-3.3		-4.6	4		123
Tn2660bla-P3	p	TTTTTCTAAATACA	TTCAAA	TATGTATC	CGCTCATGA	CACAAT	AACCCTgATAAATGCT	17	-2.6				2,4	124
Tn2661bla-Pa	p	GGTTTATAAAATTC	TTGAAG	ACGAAAGG	GCCTCGTGA	TACGCT	TATCtttcATAGGTTAA	17	-2.3				2,4	124
Tn2661bla-Pb	p	CCTC	GTGATA	CGCTTATT	TTTATAGGT	TAATGT	CATGatcaaTAATGGTTT	17	-3.1				2,4	124
Tn501mer	p	TTTTCCATATCGC	TTGACT	CGTACATG	AGTACGGAAG	TAAGGT	TACGCTaTCCAATTTC	19	-3.2				3,4	125-127
Tn501merR	p	CATGCGCTTGTCCT	TTCGAA	TTGAAATT	GCATAGCG	TAACGT	TACTcCCGTACTCA	16	-3.3		-3.8		3,4	125-127
Tn5IR	p	TCCAGGATCTGATC	TTCCAT	GTGACGTC	CTAACATCG	TAACGT	TCATGATaACTTCTGCT	17	-3.4					9
Tn5neo	p	CAAGCGAACCGGAA	TTGCCA	GCTCGGGC	GCCCTCTGG	TAAGGT	TGGGAAGCCCTGCAA	17	-2.1					9
Tn7-PLE	p	ACTAGACAGAATAG	TTGTAA	ACTGAAAT	CAGTCCAGT	TATgcc	gtgaaaaaGCAT	17	-1.6				4	128
tnaA	b	AAACAATTTCAGAA	TAGACA	AAAACTCT	GAGTGTAA	TAATGT	AGCCTCgtgTCTTGCG	16	-2.8					9
tonB	b	ATCGTCTTGCCTTA	TTGAAT	ATGATTGCT	ATTTGCATT	TAAAAT	CGAgACCTGGTTT	18	-1.3				4	129
trfA	p	AGCCGCTAAAGTTC	TTGACA	GCGGAACCA	ATGTTTAGC	TAAACT	AGAGTCtCCT	18	-1.1				4	130,131
trfB	p	AGCGGCTAAAGGTG	TTGACG	TGCGAGAA	ATGTTTAGC	TAAACT	TCTCTCaTGT	17	-1.1				4	130
trp	b	TCTGAAATGAGCTG	TTGACA	ATTAATCA	TCGAACTAG	TTAACT	AGTACGCaAGTTCACGT	17	-1.7					9
trpP2	b	ACCGGAAGAAAACC	GTGACA	TTTTACA	CGTTTGTTA	CAAGGT	AAAGCGgACGCCCCCC	17	-3.3					9
trpR	b	TGGGGACGTCGTTA	CTGATC	CGCACGTTT	ATGATATGC	TATCGT	ACTCTTTaGCGAGTACA	18	-4.3	-2.8				9
trpS	b	CGGCGAGGCTATCG	ATGTCA	GCCAGCCT	GATGTAATT	TATCAG	TCtatAAATGACC	17	-4.5		-5.7			9
trxA	b	CAGCTTACTATTGC	TTTACG	AAAGCGTAT	CCGGTGAAA	TAAAGT	CAACTaGTTGGTTAA	18	-2.5				3	132
tufB	b	ATGCAATTTTTTAG	TTGCAT	GAACTCGC	ATGTCTCCA	TAGAAT	GCGCGCTaCTTGATGCC	17	-1.8					9
tyrT	b	TCTCAACGTAACAC	TTTACA	GCGGCGCG	TCATTTGA	TATGAT	GCGCCCCgGTTCCCCAT	16	-1.6					9
tyrT/109	b	ACAGCCGCTGTTTG	TTTACG	GTAATCGAA	CGATTATTC	TTTAAT	CGCCAGcAAAAATAA	18	-2.6				2-4	131
tyrT/140	b	TTAAGTCGTCACTA	TACAAA	GTACTGCCA	CAGCGGGTC	TTTGTT	TACgGTAATCG	18	-4.2		-5.2		2-4	131
tyrT/178	b	TGCGCGCAGGTC	GTGACG	TCGAGAAAA	ACGTCT	TAAGTC	GTGCAcTATACA	15	-5.2		-4.9		2-4	131
tyrT/212	b	G	ATCATA	CCTACACAG	CTGAAGA	TATGAT	GCGCGCAgGTCGTGACG	16	-3.6				2-4	131
tyrT/6	b	ATTTTTCTCAAC	GTAACA	CTTTACA	GC	GCCGCGTCA	TTTGAT	ATGATGcGCCCCCGTTC	16	-4.1	-1.6	-1.6	2-4	131
tyrT/77	b	ATTATTCTTTAA	TCGCCA	GCAAAAATAA	CTGGTTACC	TTTAAT	CCGTTACGgATGAAAAT	19	-4.3	-4.2			2-4	131
uncI	b	TGGCTACTTATTGT	TTGAAA	TCACGGGG	GCGCACCG	TATAAT	TTGACCgGTTTTTGAT	16	-0.6		-1.6		3,4	132,133

uvrB-P1	b	TCCAGTATAATTTG	TTGGCA	TAATTAAG	TACGACGAG	TAAAAT	TACATaCCTGCCCGC	17	-1.0		9
uvrB-P2	b	TCAGAAATATTATG	GTGATG	AACTGTTTT	TTTATCCAG	TATAAT	TTGTTGgCATAATTAA	18	-2.5		9
uvrB-P3	b	ACAGTTATCCACTA	TTCCTG	TGGATAAC	CATGTGTAT	TAGAGT	TAGAAAaCACCACGCA	17	-3.7		9
uvrC	b	GCCCATTTGCCAGT	TTGTCT	GAACGTCA	ATTGCAGAT	TATGCT	GATGatccaCCAAGG	17	-1.8	4	136
uvrD	b	TGGAAATTTCCCGC	TTGGCA	TCTCTGAC	CTCCCTCA	TATAAT	CAGCAAaTCTGTATAT	16	-1.1	3	137
434PR	f	AAGAAAAACTGTAT	TTGACA	AACAAGAT	ACATTGTAT	GAAAAT	ACAAGAAAgTTTGTTGA	17	-1.3		9
434PRM	f	ACAATGTATGTTGT	TTGTCA	AATACAGT	TTTTCTTGT	GAAGAT	TGGGGGTAAATAACAGA	17	-2.4		9

List of promoter sequences arranged alphabetically by name (a) and aligned with respect to optimal -35 (c) and -10 hexamer sequences (d) consistent with the transcriptional start. Column (b) designates promoter type: b, bacterial; p, plasmid or transposon; f, phage; M, mutation or fusion which generates a new promoter; m, point mutation in an existing promoter. The lower case base(s) downstream of the -10 region denotes experimentally determined transcriptional start point(s). Column (e) indicates spacing in base pairs between -35 and -10 hexamers. Column (f) reports relative promoter homology index (PHI) of promoter elements in columns c,d,e as described in the text. Column (g) signals discrepancies between the promoter elements consistent with transcriptional start data and the best promoter elements independent of start data (indicated by double underlines). Only discrepancies for which the PHI values of these promoters differed by at least 0.5 are shown. Column (h) signals discrepancies between the computer selected promoter elements and published -35 and -10 sequences (shown by single underlines). The figures in these columns are PHI values corresponding to the underlined promoter elements. Column (i) indicates the nature of experimental data defining the transcription start: 1, total or partial RNA sequence with identification of the 5' nucleoside triphosphate; 2, mutational or genetic identification of -35 and -10 regions; 3, high resolution sizing of in vitro transcripts; 4, high resolution S1 nuclease mapping. The 112 promoters documented by Hawley and McClure (9) are included in this compilation and can be identified by a 9 in reference column (j).

% Only one of the -35 or -10 promoter hexamers was unambiguously identified, thus no PHI value for the published promoter can be given.

+ Underlined -35 and -10 regions for these genes represent heat shock promoter elements which are apparently recognized by a distinct heat shock sigma factor (34).

column (f) whenever a combination of -35 and -10 elements found by the computer or in the literature is (i) more consensus-like than the elements our program finds, but (ii) inconsistent with the transcription start data.

Base Distributions

Figure 1 shows the distribution of bases for analyzed promoters and indicates positions at which bases occur more frequently than chance by greater than 6 standard deviations (highly conserved, upper case bases) or 3 standard deviations (weakly conserved, lower case bases) (9). The base distribution of a compilation of random sequences is multinomial with probabilities p_T, p_G, p_C, p_A, where p_T, p_G, p_C, p_A are the frequencies of occurrence of T,G,C, and A, respectively. The standard deviation for each base X is $\sqrt{(np_X(1-pX))}$ where n=number of bases at that position. This statistic applies strictly only to

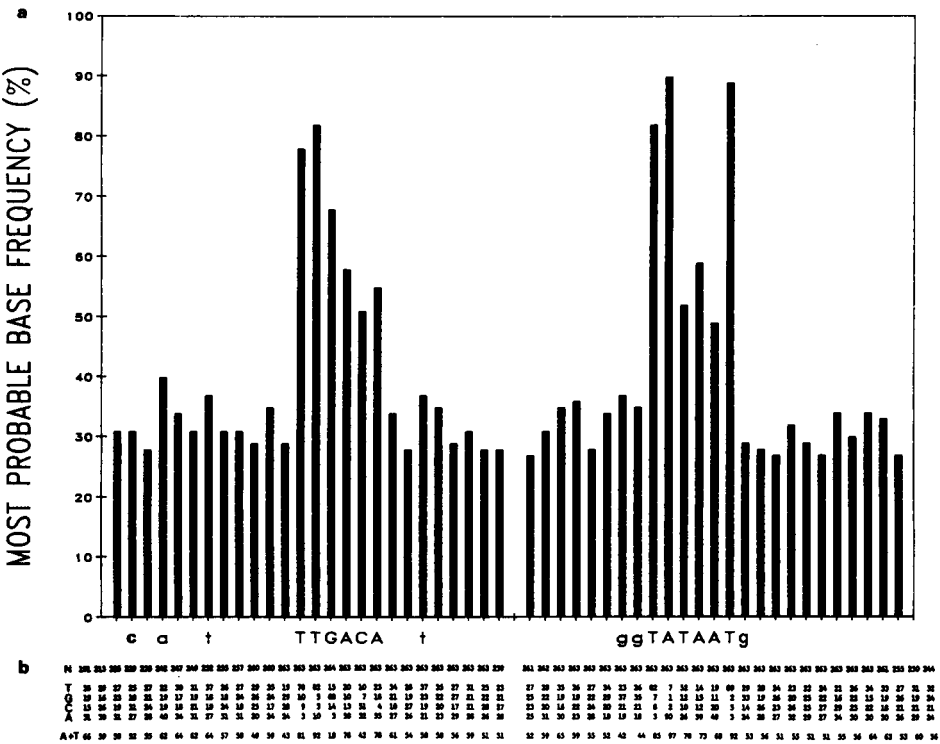

Figure 1. Base distribution of 263 analyzed promoters from Table 1. (a) Frequency histogram of the most highly conserved base on the non-template strand from 12 bp upstream of the -35 hexamer to 11 bp downstream of the -10 hexamer. Highly conserved (upper case) and weakly conserved (lower case) bases, as defined in the text, are shown below the histogram. (b) Frequency of bases (T,G,C,A and T+A) in aligned promoters as a percentage of total number of bases (N) at each position.

non-aligned positions. Frequencies T,G,C,A are 0.284, 0.225, 0.217, and 0.274, respectively, in non-aligned positions, yielding weakly conserved bases at -11, -9, -6, and +3 with respect to the -35 region, and -2, -1 and +1 with respect to the -10 region. Two of these bases (the A 9 bases upstream of the -35 and the G 2 bases upstream of the -10 region) were previously identified as weakly conserved by Hawley and McClure (9) using uniform base frequencies (.25,.25,.25,.25) and a Poisson approximation to the multinomial distribution. A similar consensus sequence was derived by Rosenberg and Court (7) from analysis of 46 promoters.

It is difficult to assign statistics to the conservation of bases in the aligned regions. However, using either the multinomial or Poisson distribution

TABLE 2
Base Distribution in -35 and -10 Regions

(a)		-35						-10					
		T	T	G	A	C	A	T	A	T	A	A	T
	T	78	82	15	20	10	24	82	7	52	14	19	89
All	G	10	5	68	10	7	17	7	1	12	15	11	2
Promoters	C	9	3	14	13	52	5	8	3	10	12	21	5
	A	3	10	3	58	32	54	3	89	26	59	49	3
Mean clonality				70						74			
(b)													
	T	78	85	22	27	11	25	84	2	65	9	11	93
Spacer = 16	G	9	4	67	9	7	13	5	0	7	9	11	2
(n=55)	C	7	5	9	9	58	5	4	2	5	9	15	5
	A	5	5	2	55	24	56	7	96	22	73	64	0
Mean clonality				69						81			
(c)													
	T	82	81	15	18	10	25	79	9	49	15	25	89
Spacer = 17	G	7	6	70	8	9	14	9	1	16	15	12	2
(n=140)	C	7	3	13	17	50	1	12	2	9	14	21	6
	A	4	10	2	57	32	60	1	88	26	56	43	3
Mean clonality				71						72			
(d)													
	T	75	82	12	14	14	18	88	10	49	18	18	86
Spacer = 18	G	18	6	69	14	4	29	4	2	6	20	12	2
(n=50)	C	8	0	12	8	47	12	6	4	22	11	25	4
	A	0	12	8	65	35	41	2	84	24	51	45	8
Mean clonality				69						72			

Frequency of bases in -35 and -10 hexamers for (a) all 263 analyzed promoters from Table 1 (a), and promoters with 16 (b), 17 (c) or 18 (d) bp separating the -35 and -10 regions. Mean clonality for each region is the arithmetic average of clonalities for each position within the region. Clonality of a base position is the square of the sum of squared frequencies at that position (138).

(which yields a larger standard deviation) and any of the base frequencies discussed above, all bases in the -35 hexamer and -10 hexamer appear highly conserved.

We did not align sequences with respect to transcription start point since in many cases this point is not precisely defined, due either to alternative initiation sites or experimental error in this determination. Nevertheless, the most probable bases 6-10 bp downstream of the -10 region, corresponding to the transcription start area of most promoters, reflect the sequence of bases in this region (CAT).

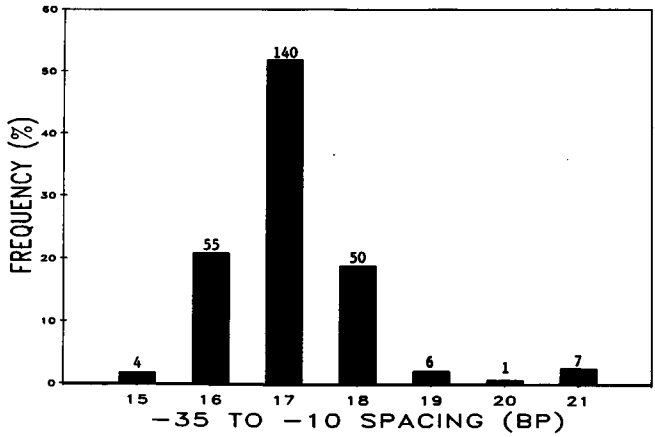

Figure 2. Distribution of promoters with 15-21 bp separating the -35 and -10 hexamers. The number of promoters in each group is indicated on top of the bars.

Base frequencies for -35 and -10 hexamers of all analyzed promoters are shown in Table 2a. Previous analysis of a limited compilation of promoter sequences suggested greater conservation of consensus-like sequences in promoters with -35 to -10 spacings of 16 or 18 bp than in promoters with the usual 17 bp spacing (J. McClarin and J. Hedgpeth, personal communication). To test this idea, subgroups of promoters with -35 to -10 spacing of 16, 17, or 18 bp were also tabulated (Table 2b-d). A composite measure of "clonality" for these regions (see Table legend) does not suggest an overall increase in conservation of bases in the -35 and -10 regions except in the -10 region of promoters with a 16 bp spacing. For these promoters, the -10 region is more consensus-like on average than the -10 region of other promoters. The statistical significance of these observations is difficult to determine since promoter sequences are not strictly independent.

Inter-region (-35 to -10) Spacing

Figure 2 shows the frequency of occurrence of promoters with 15-21 bp separating the -35 and -10 regions. As previously observed, this spacing is stringently constrained: 92% of all sequences are optimally aligned when 17±1 bp separate the -35 and -10 regions. This is consistent with known severe effects of spacer mutations (13-16) and our current understanding of RNA polymerase:promoter interaction in which the protein complex contacts one side of the DNA helix (8). Inter-region spacing outside the 16-18 bp range presumably requires unusual polymerase or DNA conformations since conserved

contact points would not lie on the same face of the DNA helix. Alternatively, the rarer inter-region distances may reflect interaction of regulatory proteins with RNA polymerase (1,2). It would be useful to obtain experimental data on interactions between RNA polymerase and DNA for promoters whose -35 to -10 spacing is thought to deviate significantly from 17 bp.

Other Analyses

We did not include weakly conserved bases flanking the -35 and -10 regions in the weight matrix since this would limit the range of possible alignments for the -35 and -10 regions. The significance of weakly conserved bases has not been well studied and the apparent conservation of some of these bases may reflect chance. Furthermcre, an analysis of our compilation using a weight matrix based on an extended -35 and -10 region (the 9 most highly conserved positions in each region) produced results similar to those shown in Table 1 (unpublished data). Stronger homology might exist in these flanking bases if slight variability in their spacing from the -35 and -10 regions were allowed.

We also did not use weakly conserved bases near the transcription start in our weight matrix because mutation studies have not supported a role for this region in promoter recognition by RNA polymerase (22,23). However, initiation points were used to validate computer-selected -35 and -10 regions by disqualifying promoters whose -10 region was not within 4-12 bp upstream of the start point. A relatively wide range of separation between these regions was allowed since experimental error in determining the start point is often ± 2 bp and actual constraints dictated by promoter/polymerase interactions are not known. Despite the weak constraint on promoter position imposed by the program 75% of optimal promoter alignments were 7 ± 2 bp from the -10 hexamer (Fig. 3). This strengthens the notion that transcription initiation occurs 5-9 base pairs downstream from the -10 region. However, in 30 cases (column g), the program identified best-fit promoters inconsistent with the reported transcriptional start point. Such discrepancies have been noted for other, similar analyses (17,18,20) and have been attributed to either inadequacies in the computer algorithm for detecting promoters or inadequacies in experimental determination of transcriptional start points. These are likely explanations here as well, but since there have been few determinations of both polymerase contact points and sites of transcription initiation, a third possibility is that the true range of distance between the -10 and transcription start point has been underestimated.

McClure (2) outlined four generalizations of E. coli promoters from analysis of 112 promoters: (i) all promoters using sigma factor 70 have at least two of

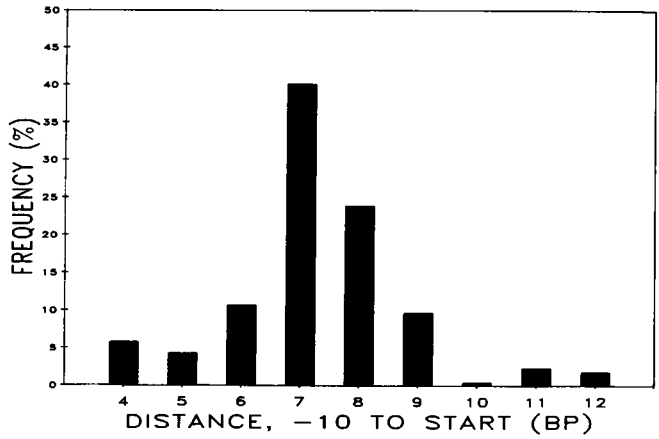

Figure 3. Distribution of promoters with transcription start points initiating 4-12 bases downstream of the -10 hexamer. Only promoters with uniquely defined start points are included in this analysis.

the three most highly conserved bases in the -10 region (TA...T), (ii) all promoters have at least one of the most highly conserved TTG residues in the-35 region, (iii) most promoters with poor homology to the consensus sequence in in the -35 region are positively regulated, and (iv) promoters using sigma factor 32 during heat shock have similar, non-consensus-like -10 regions. Our analysis supports these generalizations although some exceptions exist: 4 promoters (ada, cit.util-379, dapD, and ppc) listed in Table 1 break rule (i) and 2 promoters (lacP2 and pyrB1-P1) break rule (ii). Exceptions such as these are expected in larger compilations, but also might reflect differences in search algorithms. We have compared the ranking of the 112 promoters of Hawley and McClure (9) analyzed with the program of Mulligan et al. (16) with the ranking generated by our program. The correlation using Hawley and McClure's alignment was relatively high (Spearman rank-correlation coefficient = 0.81), but increased only slightly when our alignment was used (coefficient = 0.83). Therefore, there is no significant difference in the method by which the promoter homology score is derived.

SUMMARY

We have compiled and analyzed 263 promoter of E. coli including 112 studied by Hawley and McClure (9). The major difference in our approach is in the reiterative alignment of promoter regions to select -35 and -10 regions most consistent with the reference list of promoters and with known transcriptional

start points. The consensus sequence defined by this alignment (**c.a..t.....TTGACA..t.....ggTATAATg**) is identical in sequence to that of previous reports in the highly conserved -35 and -10 hexamer regions (7,9), but differs in some of the weakly conserved bases. Most aligned promoter elements are identical to those identified by Hawley and McClure (9) or the investigators reporting the promoter sequence. However, in 64 cases -35 and-10 regions were selected which were more consensus-like in sequence or inter-region spacing than those proposed in the initial publication. Of these, 15 differed from that of the computer-selected promoter by more than one PHI unit corresponding to a factor of 10 in statistical similarity to the consensus promoter. The computer generated alignment of promoter elements is derived from and consistent with our current knowledge of promoter sequence and thus should provide the best indication of promoter structure.

Although this compilation and analysis is an improvement over previous analyses, it too suffers the limitation that without experimental data confirming points of interaction between RNA polymerase and -35 and -10 regions, it is not possible to align these regions by existing methods without introducing bias from the initial alignment. Assuming promoter regions are defined by restricted sequence data, the consensus sequence should be identified by a program which examines all possible alignments of all sequences. Execution of an exhaustive alignment algorithm is not presently feasible for large sequence compilations such as E. coli promoters. However, we suspect that such an analysis would not significantly alter the consensus promoter sequence as defined here.

ACKNOWLEDGMENTS

We thank A.B. Futcher, J. Hedgpeth, and H. Ghosh for helpful discussion and reading of the manuscript. Supported by grants to CBH from the Medical Research Council and the National Cancer Institute of Canada.

[1]**The promoter compilation will be provided upon receipt of a blank 5¼" disk.**

*To whom correspondence should be addressed

REFERENCES

1. von Hippel, P.H., Bear, J.D., Morgan, W. and McSwiggen, J.A. (1984) Ann. Rev. Biochem. 53, 389-446.
2. McClure, W.R. (1985) Ann. Rev. Biochem. 54, 171-204.
3. Pribnow, D. (1975) Proc. Natl. Acad. Sci. USA 72, 784-788.

4. Schaller, H., Gray, C. and Herrmann, K. (1975) Proc. Natl. Acad. Sci. USA 72, 737-741.
5. Takanami, M., Sugimoto, K., Sugisaki, H. and Okamoto, T. (1976) Nature 260, 297-302.
6. Seeburg, P.H., Nusslein, C. and Schaller, H. (1977) Eur. J. Biochem. 74, 107-114.
7. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
8. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
9. Hawley, D.K. and McClure, W.R. (1983) Nucl. Acids Res. 11, 2237-2255.
10. Youderian, P. Bouvier, S. and Susskind, M. (1982) Cell 30, 843-853.
11. Deuschle, U., Kammerer, W., Gentz, R. and Bujard, H. (1986) EMBO J. 5, 2987-2994.
12. Kammerer, W., Deuschle, U., Gentz, R. and Bujard, H. (1986) EMBO J. 5, 2995-3000.
13. Mandecki, W. and Reznikoff, W.S. (1982) Nucl. Acids Res. 10, 903-912.
14. Stefano, J.E. and Gralla, J.D. (1982) Proc. Natl. Acad. Sci. USA 79, 1069-1072.
15. Russell, D.R. and Bennett, G.N. (1982) Gene 20, 231-243.
16. Aoyama, T., Takanami, M., Ohtsuka, E., Yaniyama, Y., Marumoto, R., Sato, H. and Ikehara, M. (1983) Nucl. Acids Res. 11, 5855-5864.
17. Harr, R., Haggstrom, M., and Gustafsson, P. (1983) Nucl. Acids Res. 11, 2943-2957.
18. Mulligan, M., Hawley, D., Entriken, R., and McClure, W. (1984) Nucl. Acids Res., 12, 789-800.
19. Staden, R. (1984) Nucl. Acids Res. 12, 505-519.
20. Mulligan, M. and McClure, W. (1986) Nucl. Acids Res. in press.
21. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
22. Aoyama, T. and Takanami, M. (1985) Nucl. Acids Res. 13, 4085-4096.
23. Tachibana, H. and Ishihama, A. (1985) Nucl. Acids Res. 13, 9031-9042.
24. Spencer, M.E. and Guest, J.R. (1985) Mol. Gen. Genet. 200, 145-154.
25. Demple, B., Sedgwick, B. Robins, P., Totty, N., Waterfield, M.D. and Lindahl, D. (1985) Proc. Natl. Acad. Sci. 82, 2688-2692.
26. Nakabeppu, Y., Kondo, H., Kawabata, S-i, and Sekiguchi, M. (1985) J. Biol. Chem. 260, 7281-7288.
27. Olsson, O. Bergstrom, S. and Normark, S. (1982) The EMBO J. 1, 1411-1416.
28. Stoner, C. and Schleif, R. (1983) J. Mol. Biol. 171, 369-381.
29. Horwitz, A.H., Garrett Miyada, C. and Wilcox, G. (1984) J. Bacteriol. 158, 141-147.
30. Piette, J., Cunin, R., Boyen, A., Charlier, D., Crabeel, M., Van Vliet, F., Glansdorff, N., Squires, C. and Squires, C. (1982) Nucl. Acids Res. 10, 8031-8049.
31. Cunin, R., Eckhardt, T., Piette, J., Boyen, A. and Glansdorff, N. (1983) Nucl. Acids Res. 11, 5007-5019.
32. Moore, S.K., Garvin, R.T. and James, E. (1981) Gene 16, 119-132.
33. Hudson, G.S. and Davidson, B.E. (1984) J. Mol. Biol. 180, 1023-1051.
34. Cowing, D.W., Bardwell, J.C.A., Craig, E.A. Woolford, C., Hendrix, R.W. and Gross, C.A. (1985) Proc. Natl. Acad. Sci. 82, 2679-2683.
35. Piette, J., Nyunoya, H., Lusty, C.J., Cunin, R., Weyens, G., Crabeel, M., Charlier, D., Glansdorff, N. and Pierard, A. (1984) Proc. Natl. Acad. Sci. 81, 4134-4138.
36. Sasatsu, M., Misra, T.K., Chu, L., Laddaga, R., and Silver, S. (1985) J. Bacteriol., 164, 983-993.
37. Ishiguro, N. and Sata, G. (1985) J. Bacteriol. 164, 977-982.
38. Ishiguro, Sasatsu, Misra and Silver (1986) personal communication.
39. van den Elzen, P.J.M.,Maat, J., Walters, H.H.B. Veltkamp, E. and Nijkamp, H.J.J. (1982) Nucl. Acids Res., 10, 1913-1928.
40. Parker, R.C. (1983) Gene 26, 127-136.
31. Chan, P. T., Lebowitz, J. and Bastia, D. (1979) Nucl. Acids Res. 7, 1247-1262.
42. Chan, P.T. and Lebowitz, J. (1983) Nucl. Acids Res. 11, 1099-1116.
43. Aiba, H. Fujimoto, S. and Ozaki, N. (1982) Nucl. Acids Res. 10, 1345-1361.
44. Aiba, H., Kawamukai, M. and Ishihama, A. (1983) Nucl. Acids Res. 11, 3451-3465.

45. Richaud, C., Richaud, F., Martin, C., Haziza, C. and Patte, J-C. (1984) J. Biol. Chem. 259, 14824-14828.
46. Valentin-Hansen, P., Hammer, K., Larsen, J.E.L. and Svendsen, I. (1984) Nucl. Acids Res. 12, 5211-5224.
47. Tamura, F., Nishimura, S. and Ohki, M. (1984) EMBO J. 3, 1103-1107.
48. Hansen, F. G., Hansen, E.B. and Atlung, T. (1982) EMBO J. 1, 1043-1048.
49. Ohmori, H., Kimura, M., Nagata, T. and Sakakibara, Y. (1984) Gene 28, 159-170.
50. Maki, H., Horiuchi, T. and Sekiguchi, M. (1983) Proc. Natl. Acad. Sci. 80, 7137-7141.
51. Nomura, T., Fujita, N. and Ishihama, A. (1985) Nucl. Acids Res. 13, 7647-7661.
52. Thompson, R., Taylor, L., Kelly, K., Everett, R. and Willetts, N. (1984) EMBO J. 3, 1175-1180.
53. Fowler, T., Taylor, L. and Thompson, R. (1983) Gene 25-26, 79-89.
54. Jones, H.M. and Gunsalus, R.P. (1985) J. Bacteriol. 164, 1100-1109.
55. Miles, J.S. and Guest, J.R. (1984) Nucl. Acids Res 12, 3631-3642.
56. Miles, J.S. and Guest, J.R. (1984) Gene 32, 41-48.
57. Guest, J.R. personal communication.
58. Busby, S., Truelle, N., Spassky, A., Dreyfus, M. and Buc, H. (1984) Gene 28, 201-209.
59. Ueno-Nishio, S., Mango, S., Reitzer, L.J. and Magasanik, B. (1984) J. Bacteriol. 160, 379-384.
60. Hull, E.P., Spencer, M.E., Wood, D. and Guest, J.R. (1983) FEBS Letters, 156, 366-370.
61. Plamann, M.D. and Stauffer, G.V. (1983) Gene 22, 9-18.
62. Nasoff, M.S., Baker, H.V. and Wolf Jr., R.E. (1984) Gene 27-28, 253-264.
63. Adachi, T., Mizuuchi, K., Menzel, R. and Gellert, M. (1984) Nucl. Acids Res. 12, 6389-6395.
64. Grisolia, V., Riccio, A. and Bruni, C.B. (1983) J. Bacteriol. 155, 1288-1296.
65. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S. Parker, J. and Schimmel, P. (1985) J. Biol. Chem., 260, 10063-10068.
66. Eisenbeis, S.J. and Parker, J. (1982) Gene 18, 107-114.
67. Landick, R., Vaughn, V., Lau, E., VanBogelen, R.A., Erickson, J.W. and Neidhardt, F.C. (1984) Cell 38, 175-182.
68. Vaughn, V., personal communication
69. Haughn, G.W., Squires, C.H., DeFelice, M., Largo, C.T. and Calvo, J.M. (1985) J. Bacteriol. 163, 186-198.
70. Machida, C., Machida, Y. and Ohtsubo, E. (1984) J. Mol. Biol. 177, 247-267.
71. March, P.E. and Inouye, M. (1985) J. Biol. Chem. 260, 7206-7213.
72. Inouye, S. and Inouye, M. (1985) Nucl. Acids Res. 13, 3101-3110.
73. Kamio, Y., Lin, C-K., Regue, M. and Wu, H.C. (1985) J. Biol. Chem. 260, 5616-5620.
74. Miller, K.W. and Wu, H.C., personal communication.
75. Cassan, M., Ronceray, J. and Patte, J.C. (1983) Nucl. Acids Res. 11, 6157-6166.
76. Vidal-Ingigliardi, D. and Raibaud, O. (1985) Nucl. Acids Res., 13, 5919-5926.
77. Gutierrez, C. and Raibaud, O. (1984) J. Mol. Biol. 177, 69-86
78. Raibaud, O., Gutierrez, C. and Schwartz, M. (1985) J. Bacteriol. 161, 1201-1208.
79. Michaeli, S., Mevarech, M. and Ron E.Z. (1984) J. Bacteriol. 160, 1158-1162.
80. Duchange, N., Zakin, M.M., Ferrera, P., Saint-Girons, I., Park, I., Tran, S.V., Py, M-C. and Cohen, G.N. (1983) J. Biol. Chem. 258, 14868-14871.
81. Saint-Girons, I., Duchange, N., Zakin, M.M., Park, I., Margarita, D., Ferrera, P. and Cohen, G.N. (1983) Nucl. Acids Res. 11, 6723-6732.
82. Matsuyama, S-I., and Mizushima, S. (1985) J. Bacteriol. 162, 1196-1202.
83. Mizuno, T., Chou, M-Y, and Inouye, M. (1984) Proc. Natl. Acad. Sci. 81, 1966-1970.
84. Dean, G.E., MacNab, R.M., Stader, J., Matsumura, P. and Burks, C. (1984) J. Bacteriol. 159, 991-999.
85. Goosen, N., van Heuvel, M., Moolenaar, G.F. and van de Putte, P. (1984) Gene 32, 419-426.

86. Womble, D.D., Sampathkumar, P., Easton, A.M., Lucknow, V.A. and Rownd, R.H. (1985) 181, 395-410.
87. Grindley, J.N. and Nakada, D. (1981) Nucl. Acids Res. 9, 4355-4366.
88. Ishii, S., Kuroki, K. and Imamoto, F. (1984) Proc. Natl. Acad. Sci. 81, 409-413.
89. Ishii, S., Ihara, M., Mackawa, T., Nakamura, Y., Uchida, H. and Imamoto, F. (1984) Nucl. Acids Res. 12, 3333-3342.
90. Movva, R.N., Nakamura, G.K. and Inouye, M. (1981) FEBS Letters 128, 186-190.
91. Inokuchi, K., Furukawa, H., Nakamura, K. and Mizushima, S. (1984) J. Mol. Biol. 178, 653-668.
92. Wurtzel, E.T., Chou, M.-Y., and Inouye, M. (1982) J. Biol. Chem. 257, 13685-13691.
93. Selzer, G., Som, T., Itoh, T. and Tomizawa, J.-i. (1983) Cell 32, 119-129.
94. Rodriguez, R.L., West, R.W., Heyneker, H.L., Bolivar, F. and Boyer, H.W. (1979) Nucl. Acids Res. 6, 3267-3287.
95. Gragerov, A.I., Smirnov, O.Y,m Mekhedov, S.I., Nikiforov, V.G., Chuvpilo, S.A. and Korobko, V.G. (1984) FEBS Letters 172, 64-66.
96. Harley, C.B., Lawrie, J., Betlach, M., Crea, R., Boyer H.W. and Hedgpeth, J., submitted.
97. Bindereif, A. and Neilands, J.B. (1985) J. Bacteriol. 162, 1039-1046.
98. Daldal, F. (1983) J. Mol. Biol. 168, 285-305.
99. Daldal, F. (1984) Gene 28, 337-342.
100. Schwartz, I., Klotsky, R.A., Elseviers, D., Gallagher, P.J., Krauskopf, M., Siddiqui, M.A.Q., Wong, J.F.H. and Roe, B.A. (1983) Nucl. Acids Res. 11, 4379-4389.
101. Izui, K., Miwa, T., Kajitani, M., Fujita, N., Sabe, H., Ishihama, A. and Katsuki, H. (1985) Nucl. Acids Res. 13, 59-77.
102. Churchward, G., Linder, P. and Caro, L. (1983) Nucl. Acids Res. 11, 5647-5659.
103. Linder, P., Churchward, G. and Caro, L. (1983) J. Mol. Biol. 170, 287-303.
104. Pannekoek, H., Maat, J., van den Berg, E. and Noordermeer, I. (1980) Nucl. Acids Res. 8, 1535-1550.
105. Turnbough, C.L., Hicks, K.L. and Donahue, J.P. (1983) Proc. Natl. Acad. Sci. 80, 368-372.
106. Larsen, J.N. and Jensen, K.F. (1985) Eur. J. Biochem. 151, 59-65.
107. Poulsen, P., Jensen, F., Valentin-Hansen, P., Carlson, P. and Lundberg, L.G. (1983) Eur. J. Biochem. 135, 223-229.
108. Poulsen, P., Bonekamp, F. and Jensen, K.F. (1984) EMBO J. 3, 1783-1790.
109. Sakamoto, H., Kimura, N. and Shimura, Y. (1983) Proc. Natl. Acad. Sci. 80, 6187-6191.
110. Taylor, W.E., Straus, D.B., Grossman, A.D., Burton, Z.F., Gross, C.A. and Burgess, R.R. (1984) Cell 38, 371-381.
111. Hsu, L.M., Zagorski, J. and Fournier, M.J. (1984) 178, 509-531.
112. Boros, I., Csordas-Toth, E., Kiss, A., Kiss, I., Torok, I. Udvardy, A., Udvardy, K. and Venetianer, P.(1983) Biochim. Biophys. Acta 739, 173-180.
113. Csordas-Toth, E., Boros, I. and Venetianer, P. (1979) Nucl. Acids Res. 7, 2189-2197.
114. Venetianer, P., personal communication.
115. Wood, d., Darlison, M.G., Wilde, R.J., and Guest, J.R. (1984) Biochem. J. 222, 519-534.
116. Sancar, A., Williams, K.R., Chase, J.W. and Rupp, W.D. (1981) Proc. Natl. Acad. Sci. USA 78, 4274-4278.
117. Chase, J.W., Merrill, B.M. and Williams, K.R. (1983) Proc. Natl. Acad. Sci. USA 80, 5480-5484.
118. Prosen, D.E. and Cech, C.L. (1985) Biochem. 24, 2219-2227.
119. Mulligan, M.E., Brosius, J. and McClure, W.R. (1985) J. Biol. Chem. 260, 3529-3538.
120. Brosius, J., Erfle, M. and Storella, J. (1985) J. Biol. Chem. 260, 3539-3541.
121. Saint Girons, I. and Margarita, D. (1985) J. Bacteriol. 161, 461-462.
122. Bertrand, K.P., Postle, K., Wray,Jr., L.V. and Reznikoff, W.S. (1983) Gene 23, 149-156.

123. Schollmeier, K. and Hillen, W. (1984) J. Bacteriol. 160, 499-503.
124. Chen, S.T. and Clowes, R.C. (1984) Nucl. Acids Res. 12, 3219-3234.
125. Lund, P.A., Ford, S. and Brown, N.L. (1986) J. Gen. Microbiol. 132, 465-480.
126. Misra, T.K., Brown, N.L., Fritzinger, D., Pridmore, R.D., Barnes, W.M., Haberstroh, L., and Silver, S. (1984) Proc. Natl. Acad. Sci. 81, 5975-5979.
127. Misra, T.K., personal communication
128. Gay, N.J., Tybulewicz, V.L.J. and Walker, J.E. (1986) Biochem. J. 234, 111-117.
129. Postle, K. and Good, R.F. (1983) Proc. Natl. Acad. Sci. 80, 5235-5239.
130. Smith, C.A., Shingler, V. and Thomas, C.M. (1984) Nucl. Acids Res. 12, 3619-3630.
131. Thomas, C.M., personal communication
132. Wallace, B.J. and Kushner, S.R. (1984) Gene 32, 399-408.
133. Travers, A.A., Lamond, A.I., Mace, H.A.F. and Berman, M.L. (1983) Cell 35, 265-273.
134. Jones, H.M., Brajkovich, C.M. and Gunsalus, R.P. (1983) J. Bacteriol. 155, 1279-1287.
135. Walker, J.E., Gay, N.J., Saraste, M. and Eberle, A.N. (1984) Biochem. J., 224, 799-815.
136. van Sluis, G.A., Moolenaar, G.F. and Backendorf, C. (1983) EMBO J. 2, 2313-2318.
137. Easton, A.M. and Kushner, S.R. (1983) Nucl. Acids Res. 11, 8625-8640.
138. Smith, G.P. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 507-514.