Volume 11 Number 8 1983 Nucleic Acids Research

Compilation and analysis of *Escherichia coli* promoter DNA sequences

Diane K.Hawley and William R.McClure*

Department of Biological Sciences, Carnegie-Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Received 18 January 1983; Accepted 8 March 1983

ABSTRACT

The DNA sequence of 168 promoter regions (-50 to +10) for Escherichia coli RNA polymerase were compiled. The complete listing was divided into two groups depending upon whether or not the promoter had been defined by genetic (promoter mutations) or biochemical (5′ end determination) criteria. A consensus promoter sequence based on homologies among 112 well-defined promoters was determined that was in substantial agreement with previous compilations. In addition, we have tabulated 98 promoter mutations. Nearly all of the altered base pairs in the mutants conform to the following general rule: down-mutations decrease homology and up-mutations increase homology to the consensus sequence.

INTRODUCTION

Promoters for Escherichia coli RNA polymerase have been shown to contain two regions of conserved DNA sequence, located about 10 and 35 base pairs upstream of the transcription startsite (75,107,112,204). Although promoters are expected to share common structural features reflecting a similar interaction with RNA polymerase, comparison of promoter sequences has also revealed considerable sequence diversity. This diversity is related both to the wide range of initiation frequencies and to the partial overlap of binding sites for transcriptional control proteins. In this paper, we have analyzed the sequence homologies among 112 well-defined promoters. We have also tabulated the locations of 98 promoter mutations. This information supports the importance to promoter function of the conserved sequence elements proposed by Rosenberg and Court (205) and Siebenlist et al. (206).

© IRL Press Limited, Oxford, England.

BACTERIAL PROMOTERS	TTGACA	TATAAT	+1	I	II	III
araBAD	TTAGCGGATCCTACCTGACGCTTTT	TATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTT		1,2	3	
araC	GCAAATAATCAATGTGGACTTTTCT	GCCGTGATTATAGACACTTTTGTTACGCGTTTTTGT		2,4	4	4
galP1	CTAATTTATTCCATGTCACACTTTTCG	CATCTTTGTTATGCTATGGTTAT	TTCATACCATAAG	5	6	7
galP2	CACTAATTTATTCCATGTCACACTT	TTCGCATCTTTGTTATGCTATG	GTTATTTCATACC	5	6	7
lacP1	TAGGCACCCCAGGCTTTACACTTTA	TGCTTCCGGCTCGTATGTTGTG	TGGAATTGTGAGC	8	9	
lacP2	TTAATGTGAGTTAGCTCACTCATTA	GGCACCCCAGGCTTTACACTTTA	TGCTTCCGGCTCG	8	10,195	
lacI	GACACCATCGAATGGCGCAAAACCT	TTCGCGGTATGGCATGATAGC	GCCCGGAAGAGAGT	11	12	
malEFG	AGGGGCAAGGAGGATGGAAAGAGGT	TGCCGTATAAAGAAACTAGA	GTCCGTTTAGGTGT	13		14
malK	CAGGGGGTGGAGGATTTAAGCCATC	TCCTGATGACGCATAGTCAG	CCCATCATGAATG	13		14
malT	AAAAACGTCATCGCTTGCATTAGAA	AGGTTTCTGGCCGACCTTATA	ACCATTAATTACG	15		16
tnaA	AAACAATTTCAGAATAGACAAAAAC	TCTGAGTGTAATAATGTAGC	CTCGTGTCTTGCG	17	18	
deoP1	CAGAAACGTTTTATTCGAACATCGA	TCTCGTCTTGTGTTAGAATTCT	AACATACGGTTGC	19	19	19
deoP2	AATTGTGATGTGTATCGAAGTGTGT	TGCGGAGTAGATGTTAGAATACT	AACAAACTCGCAA	19	19	19
trp	TCTGAAATGAGCTGTTGACAATTAA	TCATCGAACTAGTTAACTAGT	ACGCAAGTTCACGT	20	21	21
trpR	TGGGGACGTCGTTACTGATCCGCAC	GTTTATGATATGCTATCGTACT	CTTTAGCGAGTACA	22,23	22	
aroH	GTACTAGAGAACTAGTGCATTAGCT	TATTTTTTTGTTATCATGCT	AACCACCCGGCGAG	24	24	
trpP2	ACCGGAAGAAAACCGTGACATTTTA	ACACGTTTGTTACAAGGTAAA	GGCGACGCCGCCC	25	26	
his	ATATAAAAAAGTTCTTGCTTTCTAA	CGTGAAAGTGGTTTAGGTTAAA	AGACATCAGTTGAA	27	28	28
hisA	GATCTACAAACTAATTAATAAATAG	TTAATTAACGCTCATCATTGTACA	ATGAACTGTACAAA	27	28	
leu	GTTGACATCCGT	TTTTGTATCCAGTAACTCTAA	AAGCATATCGCATT	29		
ilvGEDA	GGCCAAAAAATATCTTGTACTATTT	ACAAAACCTATGGTAACTCTTT	AGGCATTCCTTCGA	31,32	33	
argCBH	TTTGTTTTTCATTGTTGACACACCT	CTGGTCATGATAGTATCAATATTCATGCAGTATT		34		34
thr	AAATTAAAATTTTATTGACTTAGGT	CACTAAATACTTTAACCAATATAGGCATAGCGCACA		35	35	
bioA	GCCTTCTCCAAAACGTGTTTTTTGT	TGTTAATTCGGTGTAGACTTGT	AAACCTAAATCT	36,37	36	
bioB	TTGTCATAATCGACTTGTAAACCAA	ATTGAAAAGATTTAGGTTTACAAGTCTACACCGAAT		36	36	
fol	CATCCTCGCACCAGTCGACGACGGT	TTACGCTTTACGTATAGTGGC	GACAATTTTTTTT	38		
uvrB P1	TCCAGTATAATTTGTTGGCATAATT	AAGTACGACGAGTAAAATTAC	ATACCTGCCCGC	39,40	40	
uvrB P2	TCAGAAATATTATGGTGATGAACTG	TTTTTTTATCCAGTATAATTTG	TTGGCATAATTAA	39,40	40	
uvrB P3	ACAGTTATCCACTATTCCTGTGGAT	AACCATGTGTATTAGAGTTAG	AAAACACGAGGCA	40	40	
recA	TTTCTACAAAACACTTGATACTGTA	TGAGCATACAGTATAATTGC	TTCAACAGAACAT	41,42	41	
lexA	TGTGCAGTTTATGGTTCCAAAATCG	CCTTTTGCTGTATATACTCAC	AGCATAACTGTAT	43,44	43,44	
ampC	TGCTATCCTGACAGTTGTCACGCTG	ATTGGTGTCGTTACAATCTA	ACGCATCGCCAATG	45	45	
lpp	CCATCAAAAAAATATTCTCAACATA	AAAAACTTTGTGTAATACTTGT	AACGCTACATGGA	46		47
hisJ (S.t.)	CAAGGTAGAATGCTTTGCCTTGTCG	GCCTGATTAATGGCACGATAGT	CGCATCGGATCTG	48		
Pori-r	GATCGCACGATCTGTATACTTATTT	GAGTAAATTAACCCACGATCCC	AGCCATTCTTCTGC	49	49	
Pori-l	CTGTTGTTCAGTTTTTGAGTTGTGT	ATAACCCCTCATTCTGATCCC	AGCTTATACGGT	49	49	
spot 42 RNA	ATTACAAAAAGTGCTTTCTGAACTG	AACAAAAAAGAGTAAAGTTAG	TCGCGTAGGGTACA	50		51
M1 RNA	ATGCGCAACGCGGGGTGACAAGGGC	GCGCAAACCCTCTATACTGCG	CGCCGAAGCTGACC	52		
alaS	AACGCATACGGTATTTTACCTTCCC	AGTCAAGAAAACTTATCTTATT	CCCACTTTTCAGT	53	53	
trpS	CTACGGCGAGGCTATCGATCTCAGC	CAGCCTGATGTAATTTATCAGTCTATAAATGACC		54	54	54
glnS	TAAAAAACTAACAGTTGTCAGCCTG	TCCCGCTTATAAGATCATACG	CCGTTATACGTT	55	55	

tufB	ATGCAATTTTTTAGTTGCATGAACT	CGCATGTCTCCATAGAATGCG	CGCTACTTGATGCC	56	57	
tyrT	TCTCAACGTAACACTTTACAGCGGC	GCGTCATTTGATATGATGCG	CCCCGCTTCCCGAT	58		59
leu1 tRNA	TCGATAATTAACTATTGACGAAAAG	CTGAAAACCACTAGAATGCGCCTCCGTGGTAGCAA		60	60	
supB-E	CCTTGAAAAAGAGGTTGACGCTGCA	AGGCTCTATACGCATAATGCG	CCCCGCAACGCCGA	61	62	
rrnAB P1	TTTTAAATTTCCTCTTGTCAGGCCG	GAATAACTCCCTATAATGCGCCACCACTGACACGG		63-64b	65	
rrnG P1	TTTATATTTTTCGCTTGTCAGGCCG	GAATAACTCCCTATAATGCGCCACCACTGACACGG		66		
rrnD P1	GATCAAAAAAATACTTGTGCAAAAA	ATTGGGATCCCTATAATGCGCCTCCGTTGAGACGA		67	67	
rrnE P1	CTGCAATTTTTCTATTGCGGCCTGC	GGAGAACTCCCTATAATGCGCCTCCATCGACACGG		63	65	
rrnX P1	ATGCATTTTTCCGCTTGTCTTCCTG	AGCCGACTCCCTATAATGCGCCTCCATCGACACGG		67	67	
rrnAB P2	GCAAAAATAAATGCTTGACTCTGTA	GCGGGAAGGCGTATTATGCA	CACCCCGCGCCGC	63-64b	65	
rrnG P2	AAGCAAAGAAATGCTTGACTCTGTA	GCGGGAAGGCGTATTATGCA	CACCGCCGCGCCG	66		
rrnDEX P2	CCTGAAATTCAGGGTTGACTCTGAA	AGAGGAAAGCGTAATATACG	CCACCTCGCGACAG	63,67	65,67	
str	TCGTTGTATATTTCTTGACACCTTT	TCGGCATCGCCCTAAAATTCG	GCGTCCTCATAT	68	68	69
spc	CCGTTTATTTTTTCTACCCATATCC	TTGAAGCGGTGTTATAATGCC	GCGCCCTCGATA	68	68	69
S10	TACTAGCAATACGCTTGCGTTCGGT	GGTTAAGTATGTATAATGCG	CGGGCTTGTCGT	70		70
rpoA	TTCGCATATTTTTCTTGCAAAGTTG	GGTTGAGCTGGCTAGATTAGC	CAGCCAATCTTT	71	71	
rplJ	TGTAAACTAATGCCTTTACGTGGGC	GGTGATTTTGTCTACAATCTT	ACCCCCACGTATA	72	72	
rpoB	CGACTTAATATACTGCGACAGGACG	TCCGTTCTGTGTAAATCGCA	ATGAAATGGTTTAA	72,73		
PHAGE PROMOTERS						
T7 A1	TATCAAAAAGAGTATTGACTTAAAG	TCTAACCTATAGGATACTTAC	AGCCATCGAGAGGG	74	74	
T7 A2	ACGAAAAACAGGTATTGACAACATG	AAGTAACATGCAGTAAGATACA	AATCGCTAGGTAAC	74	75	
T7 A3	GTGAAACAAAACGGTTGACAACATG	AAGTAAACACGGTACGATGTA	CCACATGAAACGAC	74	75	
T7 C	CATTGATAAGCAACTTGACGCAATG	TTAATGGGCTGATAGTCTTAT	CTTACAGGTCATC	76	76	
T7 D	CTTTAAGATAGGCGTTGACTTGATG	GGTCTTTAGGTGTAGGCTTTA	GGTGTTGGCTTTA	77	78	
λPR	TAACACCGTGCGTGTTGACTATTTT	ACCTCTGGCGGTGATAATGGT	TGCATGTACTAAG	79,80	81	
λPRM	AACACGCACGGTGTTAGATATTTAT	CCCTTGCGGTGATAGATTTAA	CGTATGAGCACAA	82,83	83,84	
λPL	TATCTCTGGCGGTGTTGACATAAAT	ACCACTGGCGGTGATACTGAG	CACATCAGCAGGA	85	86	
λPo	TACCTCTGCCGAAGTTGAGTATTTT	TGCTGTATTTGTCATAATGAC	TCCTGTTGATAGAT	87	81	
λPR'	TTAACGGCATGATATTGACTTATTG	AATAAAATTGGGTAAATTTGA	CTCAACGATGGGTT	88	89	
λPRE	TAGAGCCTCGTTGCGTTTGTTTGCA	CGAACCATATGTAAGTATTTC	CTTAGATAACAAT	90	91	
λPI	CGGTTTTTTCTTGCGTGTAATTGCG	GAGACTTTGCGATGTACTTGA	CACTTCAGGAGTG	92-94	95	
434 PR	AAGAAAAACTGTATTTGACAAACAA	GATACATTGTATGAAAATACAAGAAAGTTTGTTGAT		96,97	97	
434 PRM	ACAATGTATCTTGTTTGTCAAATAC	AGTTTTTCTTGTGAAGATTGG	GGGTAAATAACAGA	96,97		
P22 PR	CATCTTAAATAAACTTGACTAAAGA	TTCCTTTAGTAGATAATTTA	AGTGTTCTTTAAT	98	99	
P22 PRM	AAATTATCTACTAAAGGAATCTTTA	GTCAAGTTTATTTAAGATGAC	TTAACTATGAAT	98	99	
P22 ant	TCCAAGTTAGTGTATTGACATGATA	GAAGCACTCTACTATATTCTC	AATAGGTCCACGG	100,101	102	
P22 mnt	CCACCGTGGACCTATTGAGAATATA	GTAGAGTGCTTCTATCATGTC	AATACACTAACTT	100,101	102	
ΦX A	AATAACCGTCAGGATTGACACCCTC	CCAATTGTATGTTTTCATGCC	TCCAAATCTTGGA	103	104,105	
ΦX D	TAGAGATTCTCTTGTTGACATTTTA	AAAGAGCGTGGATTACTATCTGAGTCCGATGCTGTTC		103	104	
ΦX B	GCCAGTTAAATAGCTTGCAAAATAC	GTGGCCTTATGGTTACAGTATG	CCCATCGCAGTT	103,106	104,105	
fd VIII	GATACAAATCTCCGTTGTACTTTGT	TTCGCGCTTGGTATAATCGC	TGGGGGTCAAAG	107-110	113,114	115
fd X	TCCTCTTAATCTTTTTGATGCAATT	CGCTTTGCTTCTGACTATAATAGA	CAGGGTAAAGACCT	108-112	116	

PLASMID & TRANSPOSON PROMOTERS						
pBR322 bla	TTTTTCTAAATACATTCAAATATGT	ATCCGCTCATGAGACAATAAC	CCTGATAAATGCT	117	118,119	
pBR tet	AAGAATTCTCATGTTTGACAGCTTA	TCATCGATAAGCTTTAATGCG	GTAGTTTATCACA	120	121	121
pBR Pl	TTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACC		GCATTAAAGCTTA	120	121	121
pBR RNAI	GTGCTACAGAGTTCTTGAAGTGGTG	GCCTAACTACGGCTACACTAGA	AGGACAGTATTTG	120	122,123	
pBR primer	ATCAAAGGATCTTCTTGAGATCCTT	TTTTTCTGCGCGTAATCTGCT	GCTTGCAAACAAAA	120	118	
pBR322 P4	CATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTG		CTCTGATGCCGCAT	124	124	
ColEl Pl	GGAAGTCCACAGTCTTGACAGGGAA	AATGCAGCGGCGTAGCTTTTA	TGCTGTATATAAAA	125	125	
ColEl P2	TTATTTTTAACTTATTGTTTTAAAA	GTCAAAGAGGATTTTATAATGGA	AACCGCGGTAGCGT	125	125	
RSF primer	GGAATAGCTGTTCGTTGACTTGATA	GACCGATTGATTCATCATCTC	ATAAATAAAGAA	126	126	
RSF RNAI	TAGAGGAGTTTGTCTTGAAGTTATG	CACCTGTTAAGGCTAAACTGAA	AGAACAGATTTTG	126	126	
CloDF RNAI	ACACGCGGTTGCTCTTGAAGTGTGC	GCCAAAGTCCGGCTACACTGGA	AGGACAGATTTGG	127	127	
R100 RNAI	CACAGAAAGAAGTCTTGAACTTTTC	CGGGCATATAACTATACTCCC	CGCATAGCTGAAT	128	128	
R100 RNAII	ATGGGCTTACATTCTTGAGTGTTCA	GAAGATTAGTGCTAGATTACT	GATCGTTTAAGGAA	128	128	
Rl RNAII	ACTAAAGTAAAGACTTTACTTTGTG	GCGTAGCATGCTAGATTACT	GATCGTTTAAGGAA	128	128	
R100RNAIII	GTACCGGCTTACGCCGGGCTTCGGC	GGTTTTACTCCTGTATCATATG	AAACAACAGAG	128	128	
cat	ACGTTGATCGGCACGTAAGAGGTTC	CAACTTTCACCATAATGAA	ATAAGATCACTACC	130	131	
Tnl0 Pin	TCATTAAGTTAAGGTGGATACACAT	CTTGTCATATGATCAAATGGT	TTCGCGAAAAATC	132	133	
Tnl0 Pout	AGTGTAATTCGGGGCAGAATTGGTA	AAGAGAGTCGTGTAAAATATC	GAGTTCGCACATC	132	133	
Tnl0 tetA	ATTCCTAATTTTTGTTGACACTCTA	TCATTGATAGAGTTATTTTACC	ACTCCCTATCAGT	134		
Tnl0 tetR	TATTCATTTCACTTTTCTCTATCAC	TGATAGGGAGTGGTAAAATAAC	TCTATCAATGATA	134		
γδ tnpA	ACACATTAACAGCACTGTTTTTATG	TGTGCGATAATTTATAATATT	TCGGACGGTTGCA	135	135	
γδ tnpR	ATTCATTAACAATTTTGCAACCGTC	CGAAATATTATAAATTATC	GCACACATAAAAAC	135	135	
Tn5 IR	TCCAGGATCTGATCTTCCATGTGAC	CTCCTAACATGGTAACGTTCA	TGATAACTTCTGCT	136,137	138	
Tn5 neo	CAAGCGAACCGGAATTGCCAGCTGG	GGCGCCCTCTGGTAAGGTTGG	GAAGCCCTGCAA	139		
PROMOTERS CREATED BY FUSION OR MUTATION						
lacPl15	TTTACACTTTATGCTTCCGGCTCGT	ATGTTGTGTGGTATTGTGAG	CGGATAACAATTT	138		
bioP98	TTGTTAATTCGGTGTAGACTTGTAA	ACCTAAATCTTTTAAATTTGG	TTTACAAGTCGAT	36	36	
λcl7	GGTGTATGCATTTATTTGCATACAT	TCAATCAATTGTTATAATTGT	TATCTAAGGAAAT	90	90	
λcin	TAGATAACAATTGATTGAATGTATG	CAAATAAATGCATACACTATA	GGTGTGGTTTAAT	90	141	
λL57	TTGATAAGCAATGCTTTTTTATAAT	GCCAACTTAGTATAAAATAGC	CAACCTGTTCGACA	64	142	
IS2 I-II	ATGTCTGGAAATATAG	GGGCAAATCCACTAGTATTAA	GACTATCACTTATT	143	143	

SEQUENCE HOMOLOGIES

We have used biochemical and genetic evidence to decide whether a DNA sequence should be included in our list of promoters. The 112 promoters in Figure 1 were defined by one or both of the following criteria: (i) The 5′ terminal nucleotides of the transcript have been determined, or (ii) one or more promoter mutations have been sequenced. Promoters have also been located less precisely by other techniques, including the measurement of run-off transcript lengths *in vitro*, S1 mapping of an RNA isolated *in vivo*, or detection of an RNA polymerase-DNA complex by filterbinding, protection from enzymatic digestion, or electron microscopy. All of these methods are useful for localizing a promoter to a limited region of DNA; however, the assumption that the promoter can then be more precisely located by finding the best match to a consensus sequence within such a region has proven unreliable. For this reason, those promoters for which the more stringent biochemical or genetic information is unavailable are listed separately in Figure 2. These proposed promoters have not been included in the analysis of sequence homologies.

The separation of promoter sequences into two categories is based on reasonable criteria, but we realize that making such a distinction is not without difficulties. One could argue that several of the proposed promoters in Figure 2 have, in fact, been located precisely by a combination of indirect evidence. For example, some of the fd promoters listed in Figure 2 have been characterized by filter-binding techniques, run-off

Figure 1. Promoters for *E. coli* RNA polymerase. The sense strand sequences of 112 promoters were aligned as described in the text. The consensus sequences for the -35 and -10 region hexamers are shown. The nucleotide corresponding to the major 5′ end of the transcript (+1) is underlined; a dashed line indicates that the RNA was sequenced, and the 5′ end occurs at one or more of the underlined bases. The numbers in the three columns at the right correspond to the references for the DNA sequence (column I), and the 5′ end of the RNA synthesized *in vitro* (II) and *in vivo* (III). The promoters are grouped into four categories: bacterial, phage, plasmid and transposon promoters, and promoters created by mutation. All of the bacterial promoters are from *E coli* K12, except for one *S. typhimurium* promoter (hisJ) that has no sequenced *E. coli* counterpart. Several *S. typhimurium* plasmid and phage promoters were also included. These have been shown to function as promoters in *E. coli*. The DNA from the filamentous phages fd, M13, and f1 have all been sequenced in their entirety. Although only the fd promoters are listed here and in Figure 2, references are given for the DNA sequences of the other two phage genomes.

	TTGACA	TATAAT	References
araFG	ACTGGAAAGTACGTTTGCAGTGAAA	TAACTATTCAGCAGGATAAT GAATACAGAGGGGCG	144,145
malP	TAATCCCCCGCAGGATGAGGAAGGT	CAACATCGAGCCTGGCAAACT AGCGATAACGTTGTG	15
pyrBI	TATTGCATCAAATCTTGCGCCGCTT	CTGACGATGAGTATAAT GCCGGACAATTTGCC	146
purF	TGGTTCGAATGGGATTGCCATCGCG	GTACTGTTTATCGCTACCCT GATCGTTGGTGCTAT	147
glyA	TCCTTTGTCAAGACCTGTTATCGCA	CAATGATTCGGTTATACT GTTCGCCGTTGTCCA	148
glyA region	ACACCAAAGAACCATTTACATTGCA	GGGCTATTTTTTTATAAGAT GCATTTGAGATACAT	148
argEp1	TTACGGCTGGTGGGTTTTATTACGC	TCAACGTTAGTGTATTTT TATTCATAAATACTG	34
argEp2	CCCGCATCATTGCTTTGCGCTGAAA	CAGTCAAAGCGGTTATGTT CATATGCGGATGGCG	34
argF	ATTGTGAAATGGGGTTGCAAATGAA	TAATTACACATATAAAGT GAATTTTAATTCAAT	149,150
argI	TTTATGCTTTAGACTTGCAAATGAA	TAATCATCCATATAAATT GAATTTTAATTCATT	149
aroG	AGTGTAAAACCCCGTTTACACATTC	TGACGGAAGATATAGATT GGAAGTATTGCATTC	151
ilvB	CTTTGCTGAAAAATTTTCCATTGTC	TCCCCTGTAAAGCTGTGCT TGTATAAATATTGTT	152
asnA	ATGCGGATTGATGATTCATTCTATT	TTAGCCTTCTTTTTTAAT GAATCAAAAGTGAGT	153
ompA	TATGCCTGACGGAGTTCACACTTGT	AAGTTTTCAACTACGTT GTAGACTTTACATCG	154-156
ompB	TTTCGCCGAATAAATTGTATACTTA	AGCTGCTGTTTAATAT GCTTTGTAACAATTT	157
argT	ACATCAGGACAATATTGCAACGTTT	TATTAACAAATTTAACGT CGAATCGTTTTGCTG	48
crp	AAGCGAGACACCAGGAGACACAAAG	CGAAAGCTATGCTAAAAC AGTCAGGATGCTACA	158
unc	TGGCTACTTATTGTTTGAAATCACG	GGGGCGCACCGTATAAT TTGACCGCTTTTTGA	159
gnd	GCATGGATAAGCTATTTATACTTTA	ATAAGTACTTTGTATACT TATTTGCGAACATTC	160
dnaA P1	ATGCGGCGTAAATCGTGCCCGCCTC	GCGGCAGGATCGTTTACACT TAGCGAGTTCTGGAA	161
dnaA P2	TCTGTGAGAAACAGAAGATCTCTTG	CGCAGTTTAGGCTATGAT CCGCGGTGCCCGATC	161
origin B	TTTCCACAGGTAGATCCCAAC	GCGTTCACAGCGTACAAT ACGCCACTCTTAATA	49
origin E	TCAAGCCGACAAAGTTGAGTAGAAT	CCACGGCCCGGGCTTCAAT CCATTTTCATAACGC	49
glyT	AAAGAGAGCTTCTCTCGATATTCAG	TGCAGAATGAAAATCAGGT AGCCGAGTTCCAGGA	56
hisS	TGGCTCCCGAAACATTGAGGGAAGCGTTGAGGGTTCATTTTTATATT	CAGAAAGAGAATAAA	162
rpsT P1	TTATCGCGGAAAAGCTGTATTCACA	CCCCGCAAGCTGGTAGAAT CCTGCGCCATCACTA	163
rpsT P2	TTTGCACAAATCCATTGACAAAAGA	AGGCTAAAAGGGCATATT CCTCGGCCTTTGAAT	163
rpsA	CACCACCTTAAGCATTGAGCAAGTG	ATTGAAAAAGCGCTACAAT ACGCGCGCAGAAATT	164
rpmB-rpmG	TGTCTGTTCGGGACTTGAGCACATC	GCTGAGTCAGCGTATACT ACGCACCTTTGAGAA	165
rpmH P1	ATCCAGGACGATCCTTGCGCTTTAC	CCATCAGCCCGTATAAT CCTCCAC	161
L11	CGGCGATTTAATCGTTGCACAAGGC	GTGAGATTGGAATACAAT TTCGCGCCTTTTGTT	72
T7 B	TTTATGATTATCACTTTACTTATGA	GGGAGTAATGTATATGCT TACTATCGGTCTACT	77
T5 25	CATAAAAAATTTATTTGCTTTCAGG	AAAATTTTTCTGTATAAT AGATTCATAAATTTG	166
T5 26	CTTAAAAATTTCAGTTGCTTAATCC	TACAATTCTTGATATAAT ATTCTCATAGTTTGA	166
T5 28	AGTTAAAATTGTAGTTGCTAAATGC	TTAAATACTTGCTATAAT ATTTATATAAATTGA	166
T5 207	TTTAAAAAATTCATTTGCTAAACGC	TTCAAATTCTCGTATAAT ATACTTCATAAATTG	166
T4 57	TGCTTTAGATTATCTTGATAAATTT	AACTCAGGTTATGATTATTAT ATCGTTCCTGATACC	167
T4 45	TTTAACGTTAATTGCTT	TATTAAATTAGTTATAAA ATTAAATCTCATTTG	168
fd I	AATTCTCCCGTCTAATGCGCTTCCC	TGTTTTTATGTTATTCT CTCTGTAAAGGCTGC	108-110
fd I'	GGCAAATTAGGCTCTGGAAAGACGC	TCGTTAGCGTTGGTAAGAT TCAGGATAATTGTAG	108-110
fd II	ACAAAACATTAACGTTTACAATTTA	AATATTTGCTTATACAAT CATCCTGTTTTTGGG	108-110
fd II'	TTTGAATCTTTGCCTACTCATTACT	CCGGCATTGCATTTAAAAT ATATGAGGGTTCTAA	108-110
fd III	TTAAGAAATTCACCTCGAAAGCAAG	CTGATAAACCGATACAAT TAAAGGCTCCTTTTG	108-110
fd IV	TGATAAATTCACTATTGACTCTTCT	CAGCGTCTTAATCTAAGCT ATCGCTATGTTTTCA	108-110
fd IV'	TAAAATTAATAACGTTCGCGCAAAGGATTTAATAAGGGTTGTAGAAT	TGTTTGTTAAATCTA	108-110
fd V	TTATTAACGTAGATTTTTCCTCCCA	ACGTCCTGACTGGTATAAT GAGCCAGTTCTTAAA	108-110
fd VI	CGCTGGTAAACCATATGAATTTTCT	ATTGATTGTGACAAAAT AAACTTATTCCGTGG	108-110
G4 A	GCTCCAAATAATGCTTGACTAATAC	TCAATCACCACTCTAATAT GCCTCCCATCAAACG	169
G4 B	GGCAAATAAATAGCTTGCAAAACAC	GTGGCCTTATGGTTACTCT ATGCCCATCGCAGTC	169
G4 D	TAAACAATCAATGCTTGACATACTG	AAAGAACGTGGCCTATTAT CCACATCGTCAACTG	169
Mu Pe	TACCAAAAAGCAGCTTTACATTAAG	CTTTTCAGTAATTATCTT TTTAGTAAGCTAGCT	170
pMB9	AATACGCTCAGATGATGAACATCAG	TAGGGAAAATGCTTATGGT GTATTAGCTAAAGCA	171
cloacin	TCATATATTGACACCTGAAAACTGG	AGGAGTAAGGTAATAAT CATACTGTGTATATA	172
traT	GATATCGGTGGTAATTCATATGGTT	ATAGTTCAAAACGATATGAT GAGTGAATCTTAATT	173
Tn3 tnpA	TGGACACTCAAACGAAGCCGTTTTA	CTATGTCTGATAATTTATAAT ATTTCGAACGGTTGC	174
Tn3 tnpR	CGGCTTCGTTTGAGTGTCCATTAAA	TCGTCATTTTGGCATAAT AGACACATCGTGTCT	174

Figure 2. DNA sequences of proposed promoters. The sense strand sequences of possible promoters are aligned to maximize homology to the conserved -35 and -10 region hexamers. References for the sequence and the evidence that indicates that the sequence contains a promoter are listed. For the reasons discussed in the text, the current genetic and biochemical information does not allow precise location of the promoters within the sequences.

transcription experiments, and protection experiments (108,110). Conversely, it is also possible that some of the promoters in Figure 1 that were defined biochemically *in vitro* do not function *in vivo*. Thus, the defining characteristics employed here were intended to limit the uncertainty in the alignment of promoters for determination of a consensus sequence, but are not viewed as the sole criteria of promoter function.

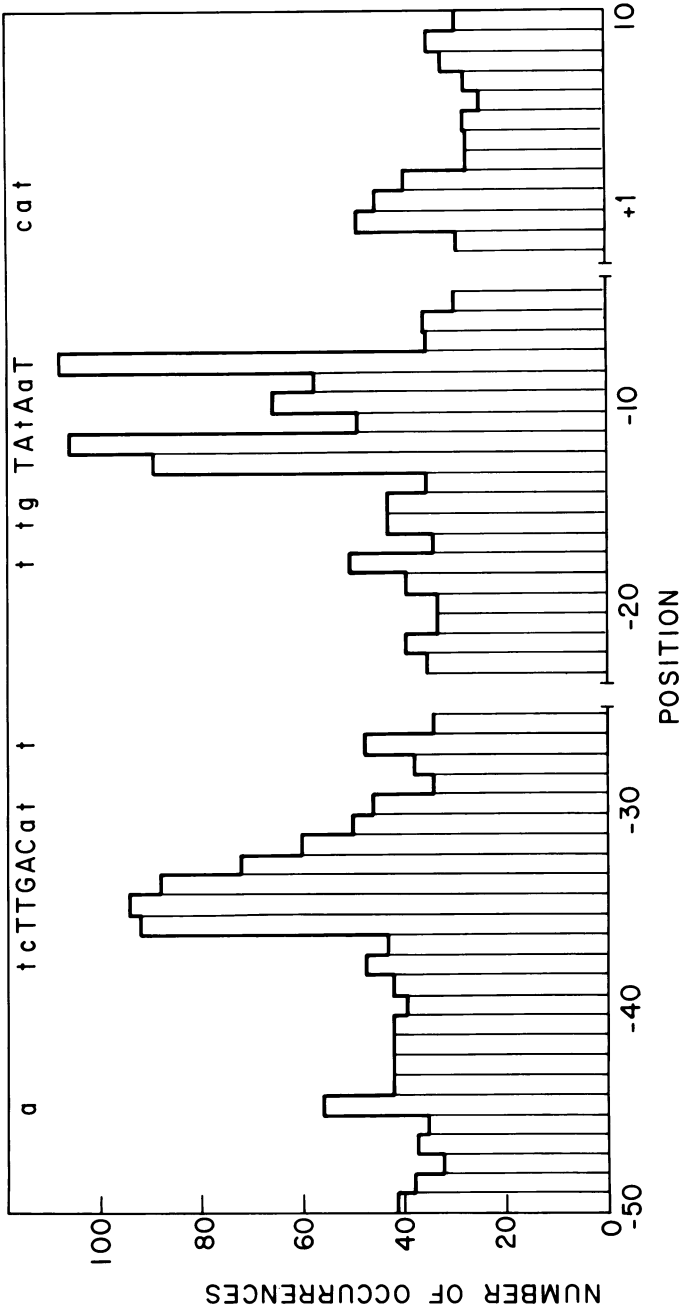
The alignment of the promoters in Figure 1 was based on several considerations. First, we tried to maximize homology to the 12 base pairs determined to be the most highly conserved among promoters previously compiled (205, 206). These bases were TTGACA around -35 and TATAAT around -10, with an allowed spacing of 15 to 21 base pairs between the two conserved sequences. When two alignments resulted in equal matches within the two hexamers, we assumed that a spacing of 17 base pairs was preferred. The optimum spacing and allowed range were selected on the basis of studies of promoters mutated by deletion or insertion between the -10 and -35 regions. Table 1 lists examples of spacer mutations. In all cases, the promoter was stronger if the spacing was closer to 17 base pairs. However, promoters with spacings of 15 and 20 base pairs were reported to retain partial function. All but 12 of the 112 promoters in Figure 1 could be aligned to maximize the homologies with spacings of 17 $\pm$ 1 base pairs.

In order to align promoters with different spacings, we placed two

Table 1
Mutations that change the spacing between the -35 and -10 regions.

Promoter	Mutation	New Spacing	Phenotype	Reference
ampC	+1	17	15 X ↑	45
$lacP^S$	-1	17	10 X ↑	207
$lacP^S$	-2	16	≃ w.t.	207
P22 ant	-1	16	weak ↓	100
tyrT	-1	15	50 X ↓	191
$lacP^+$	+2	20	10 X ↓	208

Promoters in which insertion (+) or deletion (-) mutations have been characterized are shown. All of the phenotypes correspond to estimates of the changes in levels of expression *in vivo*, except for the $lacP^S$ mutations, which were characterized *in vitro*.



A	28	38	30	37	35	56	42	42	37	42	39	18	25	26	2	6	2	72	26	50	26	34	25	26	31	20	39	33	33	39	23	29	16	23	19	2	106	29	66	57	1	35	23	31	20	9	45	16	24	25	28	24	24	32	35	26
T	41	33	32	25	34	22	35	35	42	27	32	42	47	14	92	94	11	19	15	37	46	34	38	48	34	35	28	28	27	39	51	34	43	26	31	89	3	49	15	19	108	31	29	21	16	22	2	42	27	23	20	25	27	15	16	29
G	16	11	29	19	21	18	13	21	9	19	24	26	29	29	11	6	88	3	11	17	15	21	26	21	27	23	24	27	25	22	13	29	28	43	35	11	1	18	17	14	0	33	24	30	21	8	37	5	12	27	22	17	20	24	21	16
C	22	27	18	29	20	14	20	12	22	23	16	25	10	43	7	6	11	18	60	8	25	23	23	17	20	34	21	24	27	12	25	20	25	20	27	10	2	16	14	22	3	13	36	30	29	49	4	25	25	13	18	22	17	17	16	17

breaks in the sequences. The break between the -10 region and the transcription startsite was arbitrarily placed three base pairs downstream of the conserved -10 region hexamer. We chose the position of the break between the -10 and -35 regions by comparing the sequences of six pairs of analogous E. coli and S. typhimurium promoters. Most of the sequence differences between these highly homologous promoters occurred approximately 8 to 15 base pairs upstream of the -10 region hexamer.

The distribution of specific bases at each position is displayed in Figure 3. We have numbered the positions relative to the startsite of a "standard" promoter with spacings of 7 and 17 base pairs between the startsite and the -10 region, and the -10 and -35 regions, respectively. For purposes of discussing conserved base pairs, we have used Poisson statistics to express standard deviations from the expected random (1/4) occurrence of a base pair. We have arbitrarily chosen to define a consensus sequence consisting of strongly conserved and weakly conserved base pairs, present at frequencies greater than the expected by 6 and 3 standard deviations, respectively. These are indicated in Figure 3 by upper and lower case letters above the histogram.

Within the -35 region, the TTGAC sequence is strongly conserved. The A at -31 occurs at approximately the same frequency as four other weakly conserved bases within the -35 region: a T at position -38, a C at -37, a T at -30 and a T at -27. When the base at position -31 is not an A, a T is present at significantly greater than random frequency. Upstream of the -35 region is an 8 to 10 base pair A-T-rich region, with a conserved A at position -45.

Figure 3. The distribution of bases at each position in the promoter. The histogram displays the number of occurrences of the most prevalent base in the sense stand at each position. The number of occurrences of each base is tabulated below each column of the histogram. The positions have been numbered according to a promoter with the most frequent spacing between the regions of homology (see text). The bases that occur in at least 39% of the promoters (three standard deviations above expected random occurrence) are listed above the histogram. Bases that are greater than 54% conserved (6 standard deviations) are capitalized. Standard deviations are based on Poisson statistics. In this compilation, equal occurrence of the four bases at each site corresponds to 28 $\pm$ 5.3. From position -2 to +10 only 88 sequences are tabulated. These correspond to promoters for which the 5' end of the RNA has been determined. In this region equal occurrence corresponds to 22 $\pm$ 4.7.

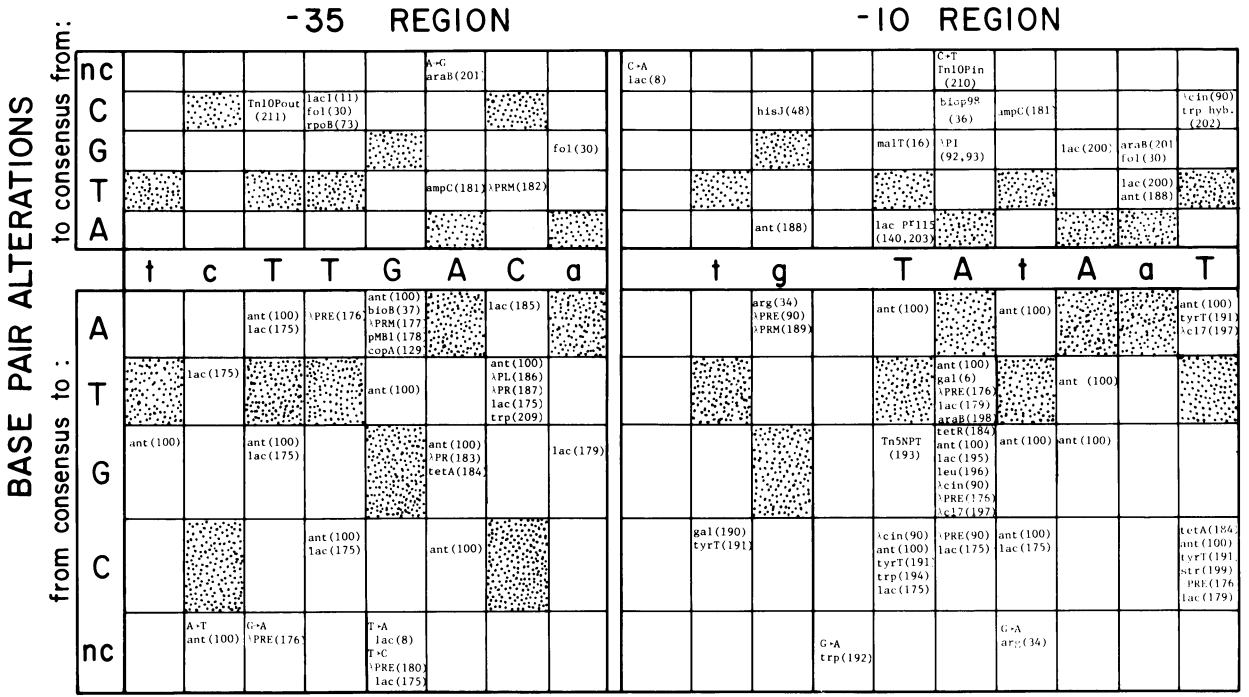


Figure 4. Location of promoter mutations. The base pair changes responsible for increasing or decreasing the initiation frequencies of promoters are shown. The consensus promoter is shown in the center of the grid. The mutations listed above this sequence are up-mutations that convert the indicated base to the consensus base. The mutations listed below are down-mutations that change the consensus base to the indicated base. Almost all of the mutations can be placed within this grid. The exceptions are the nonconsensus to nonconsensus changes (in the rows indicated "nc").

Four of the six base pairs in the -10 region hexamer are strongly conserved. The other two base pairs are weakly conserved, as are the T at -18, the T at -16, and the G at -15. The so-called "invariant T," while apparently not absolutely required for promoter function, is present in all but four of the wild-type promoters in this compilation. The A at -12 is present nearly as often; only six of the promoters do not share this homology.

Of the 112 promoters in Figure 1, only those 88 for which the 5′ end of the transcript has been precisely determined were used to examine the sequence homologies around the startsite of transcription. A preference for a C at position -1 and a T at +2 was observed. No significant homologies were found downstream of +2. The spacing between the -10 region and the startsite is usually 6 or 7 base pairs, but varies between 4 and 8 nucleotides. The presence of multiple starts for some promoters indicates that the RNA polymerase is somewhat flexible in the selection of a startsite. Initiation with a purine is highly preferred. For most promoters, transcription begins with an A if one occurs within the required region, or a G if an A is not present. However, this generalization is far too simplistic because the availability of an A or a purine does not always preclude initiation with a G or a pyrimidine, respectively.

PROMOTER MUTATIONS

The location of promoter mutations strongly suggests that the most highly conserved base pairs in the promoter are the main sequence determinants of promoter strength. About 75% of all sequenced mutations occur at the positions of the strongly conserved bases in the -10 and -35 regions of the promoter (Figure 4). Nearly all of the rest are located in positions of weakly conserved base pairs. In addition, all but seven of the mutations that decrease initiation frequency also decrease the homology of the promoter to the consensus sequence, while up-mutations increase the homology in all but three cases. This generalization is most strikingly illustrated by promoters for the phage P22 antirepressor and the *E. coli* lactose operon, for which many different single base pair mutations have been selected and sequenced. Only two of the mutations shown in Figure 4 occur at positions not included by our definition of consensus

sequence. The importance of some of the bases that are included (e.g., the A at -45 and the bases around the startsite) has not yet been demonstrated by genetic evidence.

The only up mutation that decreases homology to the consensus sequence is the A to G transition in the -35 region of araBAD. There are no down mutations that increase homology to the consensus sequence. However, seven of the mutations compiled in Figure 4 correspond to nonconsensus base pairs altered to other nonconsensus base pairs. These data suggest that a hierarchy of base pair preference could exist at some positions.

The demonstration that particular base pairs are highly preferred at some positions within the promoter and that mutations at these positions damage promoter function suggests that RNA polymerase specifically interacts with functional groups on these base pairs. It is also possible that certain bases at some positions interfere with promoter function. For example, at the following positions, one particular base is present at significantly lower frequency (by 3 standard deviations) than expected for random occurrence of three bases: a G at -42, a G at -33, and a C at -31. At most positions where strong homologies are observed, the other three bases all occur at such low frequency that a disfavored base cannot be assigned with certainty.

We draw two general conclusions from the literature survey presented here. First, the genetic and biochemical criteria we have used are the least ambiguous defining characteristics for promoters. Second, the location and sequences of most promoter mutations suggest that the consensus promoter sequence corresponds to maximal function. In the future, it may be possible to locate promoters by considering DNA sequence alone. Such attempts are currently speculative because the relative contribution of each base pair to promoter function cannot be assigned on the basis of homology information and mutant data alone. We expect that the current compilations will be useful for designing experiments that will better define promoter location and the determinants of promoter function.

ACKNOWLEDGMENTS

We are grateful to many investigators who sent us DNA sequences and other information prior to publication. Research in this laboratory is supported by NIH grant GM30375.

*To whom correspondence should be sent

REFERENCES

1. Greenfield, L., Boone, T. and Wilcox, G. (1978) Proc. Natl. Acad. Sci. 75, 4724-4728.
2. Smith, B. R. and Schleif, R. (1978) J. Biol. Chem. 253, 6931-6933.
3. Lee, N. and Carbon, J. (1977) Proc. Natl. Acad. Sci. 74, 49-53.
4. Wallace R. G., Lee, N. and Fowler, A. V. (1980) Gene 12, 179-190.
5. Musso, R., DiLauro, R., Rosenberg, M., and deCrombrugghe, B. (1977) Proc. Natl. Acad. Sci. 74, 106-110.
6. Musso, R. E., DiLauro, R., Adhya, S., and deCrombrugghe, B. (1977) Cell 12, 847-854.
7. Aiba, H., Adhya, S., and deCrombrugghe, B. (1981) J. Biol. Chem. 256, 11905-11910.
8. Dickson, R. C., Abelson, J., Barnes, W. M., and Reznikoff, W. S. (1975) Science 182, 27-35.
9. Majors, J. (1975) Proc. Natl. Acad. Sci. 72, 4394-4398.
10. Malan, T. P., Jr. (1981) Ph.D. thesis. Harvard University, Cambridge, Massachusetts.
11. Calos, M. (1978) Nature 274, 762-765.
12. Steege, D. A. (1977) Proc. Natl. Acàd. Sci. 74, 4163-4167.
13. Bedouelle, H. and Hofnung, M. (1982) Mol. Gen. Genet. 185, 82-87.
14. Bedouelle, H., Schmeissner, U., Hofnung, M. and Rosenberg, M. (1982) J. Mol. Biol. 161, 519-531.
15. Debarbouille, M., Cossart, P. and Raibaud, O. (1982) Mol. Gen. Genet. 185, 88-92.
16. Chapon, C. (1982) EMBO J. 1, 369-374.
17. Deeley, M. C. and Yanofsky, C. (1981) J. Bacteriol. 147, 787-796.
18. Deeley, M. C. and Yanofsky, C. (1982) J. Bacteriol. 151, 942-951.
19. Valentin-Hansen, P., Aiba, H. and Schumperli, D. (1982) EMBO J. 1, 317-322.
20. Bennett, G. N., Schweingruber, M. E., Brown, K. D., Squires, C. and Yanofsky, C. (1978) J. Mol. Biol. 121, 113-137.
21. Squires, C., Lee, F., Bertrand, K., Squires, C. L., Bronson, M. J., and Yanofsky, C. (1976). J. Mol. Biol. 103, 351-381.
22. Gunsalus, R. P. and Yanofsky, C. (1980) Proc. Natl. Acad. Sci. 77, 7117-7121.
23. Singleton, C. K., Roeder, W. D., Bogosian, G., Somerville, R. L., and Weith, H. L. (1980) Nucleic Acids Res. 8, 1551-1560.
24. Zurawski, G., Gunsalus, R. P., Brown, K. D., and Yanofsky, C. (1981) J. Mol. Biol. 145, 47-73.
25. Horowitz, H., Christie, G. E., and Platt, T. (1982) J. Mol. Biol. 156, 245-256.
26. Horowitz, H. and Platt, T. (1982) J. Mol. Biol. 156, 257-267.
27. Verde, P., Frunzio, R., diNocera, P. P., Blasi, F. and Bruni, C. B. (1981) Nucleic Acids Res. 9, 2075-2086.
28. Frunzio, R., Bruni, C. B. and Blasi, F. (1981) Proc. Natl. Acad. Sci. 78, 2767-2771.
29. Wessler, S. R. and Calvo, J. M. (1981) J. Mol. Biol. 149, 579-597.
30. Smith, D. R., Rood, J. I., Bird, P. I., Sneddon, M. K., Calvo, J. M., and Morrison, J. F. (1982) J. Biol. Chem. 257, 9043-9048.
31. Lawther, R. P. and Hatfield, G. W. (1980) Proc. Natl. Acad. Sci. 77, 1862-1866.
32. Nargang, F. E., Subrahmanyam, C. S., and Umbarger, H. E. (1980) Proc. Natl. Acad. Sci. 77, 1823-1827.

33. Hatfield, G. W., personal communication.
34. Piette, J., Cunin, R., Boyen, A., Charlier, D., Crabeel, M., Van Vliet, F., Glansdorff, N., Squires, C., & Squires, C. L. (1982) Nucleic Acids Res. 10, 8031-8048.
35. Gardner, J. F. (1982) J. Biol. Chem. 257, 3896-3904.
36. Otsuka, A. and Abelson, J. (1978) Nature 276, 689-694.
37. Barker, D. F., Kuhn, J., and Campbell, A. M. (1981) Gene 13, 89-102.
38. Smith, D. R. and Calvo, J. (1980) Nucleic Acids Res. 8, 2255-2274.
39. van den Berg, E., Zwetsloot, J., Noordermeer, I., Pannekoek, H., Dekker, B., Dijkema, R. and van Ormondt, H. (1981) Nucleic Acids Res. 9, 5623-5643.
40. Sancar, G. B., Sancar, A., Little, J. W., and Rupp, W. D. (1982) Cell 28, 523-530.
41. Horii, T., Ogawa, T. and Ogawa, H. (1980) Proc. Natl. Acad. Sci. 77, 313-317.
42. Sancar, A., Stachelek, C., Konigsberg, W. and Rupp, W. D. (1980) Proc. Natl. Acad. Sci. 77, 2611-2615.
43. Horii, T., Ogawa, T. and Ogawa, H. (1981) Cell 23, 689-697.
44. Miki, T., Ebina, Y., Kishi, F., and Nakazawa, A. (1981) Nucleic Acids Res. 9, 529-543.
45. Jaurin, B., Grundstrom, T., Edlund, T. and Normark, S. (1981) Nature 290, 221-225.
46. Nakamura, K. and Inouye, M. (1979) Cell 18, 1109-1117.
47. Pirtle, R. M., Pirtle, I. L. and Inouye, M. (1978) Proc. Natl. Acad. Sci. 75, 2190-2194.
48. Higgins, C. F. and Ames, G. F.-L. (1982) Proc. Natl. Acad. Sci. 79, 1083-1087.
49. Lother, H. and Messer, W. (1981) Nature 294, 376-378.
50. Joyce, C. M. and Grindley, N. D. F., (1982) J. Bacteriol. 152, 1211-1219.
51. Sahagan, B. G. and Dahlberg, J. E. (1979) J. Mol. Biol. 131, 573-592.
52. Reed, R. E., Baer, M. F., Guerrier-Takada, C., Donis-Keller, H., and Altman, S. (1982) Cell 30, 627-636.
53. Putney, S. D., Melendez, D. L. D. and Schimmel, P. R. (1981) J. Biol. Chem. 256, 205-211.
54. Hall, C. V. and Yanofsky, C. (1981) J. Bacteriol. 148, 941-949.
55. Soll, D., personal communication.
56. An, G. and Friesen, J. D. (1980) Gene 12, 33-39.
57. Miyajima, A., Shibuya, M., Kuchino, Y. and Kaziro, Y. (1981) Mol. Gen. Genet. 183, 13-19.
58. Sekiya, T., Takeya, T., Contreras, R., Kupper, H., Khorana, H. G. and Landy, A. (1976) in RNA Polymerase, Losick, R. and Chamberlin, M., Eds., pp. 455- 472, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
59. Altman, S. and Smith, J. D. (1971) Nat. New Biol. 233, 35-39.
60. Duester, G., Campen, R. K. and Holmes, W. M. (1981) Nucleic Acids Res. 9, 2121-2139.
61. Nakajima, N., Ozeki, H., and Shimura, Y. (1981) Cell 23, 239-249.
62. Nakajima, N., Ozeki, H., and Shimura, Y. (1982) J. Biol. Chem. 257, 11113-11120.
63. deBoer, H. A., Gilbert, S. F. and Nomura, M. (1979) Cell 17, 201-209.
64a. Brosius, J., Dull, T. J., Sleeter, D. D., and Noller, H. F. (1981) J. Mol. Biol. 148, 107-127.
64b. Csordas-Toth, E., Boros, I., & Venetianer, P. (1979) Nucleic Acids Res. 7, 2189-2197.
65. Gilbert, S. F., deBoer, H. A., and Nomura, M. (1979) Cell 17, 211-224.
66. Shen, W.-F., Squires, C., and Squires, C. L. (1982) Nucleic Acids Res.

10, 3303-3313.
67. Young, R. A. and Steitz, J. A. (1979) Cell 17, 225-234.
68. Post, L. E., Arfsten, A. E., Reusser, F. and Nomura, M. (1978) Cell 15, 215-229.
69. Ikemura, T., Itoh, S., Post, L. E., and Nomura, M. (1979) Cell 18, 895-903.
70. Olins, P. and Nomura, M. (1981) Cell 26, 205-211.
71. Post, L. E., Arfsten, A. E., Davis, G. R., and Nomura, M. (1980) J. Biol. Chem. 255, 4653-4659.
72. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. and Dennis, P. P. (1979) Proc. Natl. Acad. Sci. 76, 1697-1701.
73. An, G. and Friesen, J. D. (1980) J. Bacteriol. 144, 904-916.
74. Siebenlist, U. (1979) Nucleic Acids Res. 6, 1895-1907.
75. Pribnow, D. (1975) J. Mol. Biol. 99, 419-443.
76. McConnell, D. J. (1979) Nucleic Acids Res. 6, 525-544.
77. Dunn, J. J. and Studier, F. W. (1981) J. Mol. Biol. 148, 303-330.
78. Cech, C. L., Lichy, J. and McClure, W. R. (1980) J. Biol. Chem. 255, 1763-1766.
79. Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975) Proc. Natl. Acad. Sci. 72, 1184-1188.
80. Walz, A. and Pirrotta, V. (1975) Nature 254, 118-121.
81. Blattner, F. R. and Dahlberg, J. E. (1972) Natl. New Biol. 237, 227-232.
82. Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B., and Sauer, R. T. (1976) Science 194, 156-161.
83. Walz, A., Pirrotta, V., and Ineichen, K. (1976) Nature 262, 665-669.
84. Meyer, B. J. (1979) Ph.D. thesis. Harvard University, Cambridge, Massachusetts.
85. Maniatis, T., Ptashne, M., Barrell, B. G., and Donelson, J. (1974) Nature 250, 394-397.
86. Dahlberg, J. E. and Blattner, F. R. (1975) Nucleic Acids Res. 2, 1441-1458.
87. Scherer, G., Hobom, G. and Kossel, H. (1977) Nature 265, 117-121.
88. Sklar, J. L. (1977) Ph.D. thesis. Yale University, New Haven, Connecticut.
89. Lebowitz, P., Weissman, S. M., and Radding, C. M. (1971) J. Biol. Chem. 246, 5120-5139.
90. Rosenberg, M., Court, D., Shimatake, H., Brady, C., and Wulff, D. L. (1978) Nature 272, 414-423.
91. Schmeissner, U., Court, D., Shimatake, H. and Rosenberg, M. (1980) Proc. Natl. Acad. Sci. 77, 3191-3195.
92. Abraham, J., Mascarenhas, D., Fischer, R., Benedik, M., Campbell, A., and Echols, H. (1980) Proc. Natl. Acad.Sci. 77, 2477- 2481.
93. Hoess, R. H., Foeller, C., Bidwell, K., and Landy, A. (1980) Proc. Natl. Acad. Sci. 77, 2482-2486.
94. Davies, R. W. (1980) Nucleic Acids Res. 8, 1765-1782.
95. Schmeissner, U., Court, D., McKenney, K., and Rosenberg, M. (1981) Nature 292, 173-175.
96. Pirrotta, V. (1979) Nucleic Acids Res. 6, 1495-1508.
97. Johnson, A. D., Poteete, A. R., Lauer, G., Ackers, G. K., and Ptashne, M. (1982) Nature 294, 217-223.
98. Poteete, A. R., Ptashne, M., Ballivet, M. and Eisen, H. (1980) J. Mol. Biol. 137, 81-91.
99. Poteete, A. R. and Ptashne, M. (1982) J. Mol. Biol. 157, 21-48.
100. Youderian, P., Bouvier, S., and Susskind, M. (1982) Cell 30, 843-853.
101. Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P., and Susskind, M., J. Mol. Biol., in press.

102. Youderian, P. and Hawley, D. K., unpublished results.
103. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A. III, Slocombe, P. M., and Smith, M. (1977) Nature 265, 687-695.
104. Smith, L. H., Grohmann, K. and Sinsheimer, R. L. (1974) Nucleic Acids Res. 1, 1521-1529.
105. Grohmann, K., Smith, L. H., and Sinsheimer, R. L. (1975) Biochemistry 14, 1951-1955.
106. Godson, G. N., Fiddes, J. C., Barrell, B. G., and Sanger, F. (1978) in Single-stranded DNA Phages, Denhardt, D. T., Dressler, D. and Ray, D. S., Eds., pp. 51-86, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
107. Takanami, M., Sugimoto, K., Sugisaki, H., and Okamoto, T. (1976) Nature 260, 297-302.
108. Schaller, H., Beck. E. and Takanami, M. (1978) in Single-stranded DNA Phages, Denhardt, D. T., Dressler, D. and Ray, D. S., Eds., pp. 139-163, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
109. van Wezenbeek, P. M. G. F., Hulsebos, T. J. M., and Schoenmakers, J. G. G. (1980) Gene 11, 129-148.
110. Beck, E. and Zink, B. (1981) Gene 16, 35-58.
111. Okamoto, T., and Takanami, M. (1975) Nucleic Acids Res. 2, 2091-2100.
112. Schaller, H., Gray, C., and Herrman, K. (1975) Proc. Natl. Acad. Sci. 72, 737-741.
113. Sugimoto, K. Sugisaki, H., Okamoto, T. and Takanami, M. (1977) J. Mol. Biol. 111, 487-507.
114. Hulsebos, T. and Schoenmakers, J. G. G. (1978) Nucleic Acids Res. 5, 4677-4698.
115. Rivera, M. J., Smits, M. A., Quint, W., Schoenmakers, J. G. G., and Konings, R. N. H. (1978) Nucleic Acids Res. 5, 2895-2912.
116. Sugimoto, K., Okamoto, T., Sugisaki, H., and Takanami, M. (1975) Nature 253, 410-414.
117. Sutcliffe, J. G. (1978) Proc. Natl. Acad. Sci. 75, 3737-3741.
118. Itoh, T. and Tomizawa, J.-I. (1980) Proc. Natl. Acad. Sci. 77, 2450-2454.
119. Russell, D. R. and Bennett, G. N. (1981) Nucleic Acids Res. 9, 2517-2533.
120. Sutcliffe, J. G. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.
121. Brosius, J., Cate, R. L., and Perlmutter, A. P. (1982) J. Biol. Chem. 257, 9205-9210.
122. Morita, M. and Oka, A. (1979) Eur. J. Biochem. 97, 435-443.
123. Chan, P. T., Lebowitz, J. and Bastia, D. (1979) Nucleic Acids Res. 7, 1247-1262.
124. Queen, C. and Rosenberg, M. (1981) Nucleic Acids Res. 9, 3365-3377.
125. Ebina, Y., Kishi, F., Miki, T., Kagamiyama, H., Nakazawa, T. and Nakazawa, A. (1981) Gene 15, 119-126.
126. Som, T. and Tomizawa, J.-I. (1982) Mol. Gen. Genet. 187, 375-383.
127. Stuitje, A. R., Spelt, C. E., Veltkamp, E., and Nijkamp, H. J. J. (1981) Nature 290, 264-267.
128. Rosen, J., Ryder, T., Ohtsubo, H., and Ohtsubo, E. (1981) Nature 290, 794-797.
129. Stougaard, P., Molin, S., and Nordstrom, K. (1981) Proc. Natl. Acad. Sci. 78, 6008-6012.
130. Marcoli, R., Iida, S., and Bickle, T. A. (1980) FEBS Lett. 110, 11-14.
131. LeGrice, S. F. J. and Matzura, H. (1981) J. Mol. Biol. 150, 185-196.
132. Halling, S. M., Simons, R. W., Way, J. C., Walsh, R. B., and Kleckner, N. (1982) Proc. Natl. Acad. Sci. 79, 2608-2612.

133. Simons, R. W., Hoopes, B. C., McClure, W. R., and Kleckner, N., unpublished observations.
134. Bertrand K., Postle, K., Wray, L., and Reznikoff, W., personal communication.
135. Reed, R. R., Shibuya, G. I., and Steitz, J. A. (1982) Nature 300, 381-383.
136. Auerswald, E. A., Ludwig, G., and Schaller, H. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 107-113.
137. Reznikoff, W. S., personal communication.
138. Johnson, R. C. and Reznikoff, W. S. (1981) Nucleic Acids Res. 9, 1873-1883.
139. Rothstein, S. J. and Reznikoff, W. S. (1981) Cell 23, 191-199.
140. Maquat. L. E., Thornton, K., and Reznikoff, W. S. (1980) J. Mol. Biol. 139, 537-549.
141. Shimatake, H. and Rosenberg, M., personal communication.
142. Kingston, R. E., personal communication.
143. Hinton, D. M. and Musso, R. E. (1982) Nucleic Acids Res. 10, 5015-5031.
144. Kosiba, B. E. and Schleif, R. (1982) J. Mol. Biol. 156, 53-66.
145. Stoner, C. and Schleif, R., personal communication.
146. Roof, W. D., Foltermann, K. F., and Wild, J. R. (1982) Mol. Gen. Genet. 187, 391-400.
147. Tso, J. Y., Zalkin, H. van Cleemput, M., Yanofsky, C., and Smith, J. M. (1982) J. Biol. Chem. 257, 3525-3531.
148. Plamann, M. and Stauffer, G. V., personal communication.
149. Piette, J., Cunin, R., Van Vliet, F., Charlier, D., Crabeel, M., Ota, Y., and Glansdorff, N. (1982) EMBO J. 1, 853-857.
150. Moore, S. K., Garvin, R. T., and James, E. (1981) Gene 16, 119-132.
151. Davies, W. D. and Davidson, B. E. (1982) Nucleic Acids Res. 10, 4045-4058.
152. Friden, P., Newman, T., and Freundlich, M. (1982) Proc. Natl. Acad. Sci. 79, 6156-6160.
153. Nakamura, M., Yamada, M., Hirota, Y., Sugimoto, K., Oka, A., and Takanami, M. (1981) Nucleic Acids Res. 9, 4669-4676.
154. Movva, R. N., Nakamura, K., and Inouye, M. (1980) Proc. Natl. Acad. Sci. 77, 3845-3849.
155. Movva, R. N., Green, P., Nakamura, K., and Inouye, M. (1981) Febs Lett. 128, 186-190.
156. Beck, E. and Bremer, E. (1980) Nucleic Acids Res. 8, 3011-3024.
157. Wurtzel, E. T., Chou, M.-Y., and Inouye, M. (1982) J. Biol. Chem. 257, 13685-13691.
158. Aiba, H., Fujimoto, S., and Ozaki, N. (1982) Nucleic Acids Res. 10, 1345-1361.
159. Gay, N. J. and Walker, J. E. (1981) Nucleic Acids Res. 9, 3919-3926.
160. Nasoff, M. S. and Wolf, R. E., Jr., personal communication.
161. Hansen, E. B., Hansen, F. G., and von Meyenburg, K. (1982) Nucleic Acids Res. 10, 7373-7385.
162. Eisenbeis, S. J. and Parker, J. (1982) Gene 18, 107-114.
163. Mackie, G. A. (1981) J. Biol. Chem. 256, 8177-8182.
164. Schnier, J. and Isono, K. (1982) Nucleic Acids Res. 10, 1857-1865.
165. Lee, J. S., An, G., Friesen, J. D., and Isono, K. (1981) Mol. Gen. Genet. 184, 218-223.
166. Bujard, H., Niemann, A., Breunig, K., Roisch, U., Dresel, A., von Gabain, A., Gentz, R., Stuber, D., and Weiher, H. (1982) in Promoters: Structure and Function, Rodriguez, R L. and Chamberlin, M. J., Eds., pp. 121-140, Praeger Publishers, New York.
167. Herrmann, R. (1982) Nucleic Acids Res. 10, 1105-1112.
168. Spicer, E. K., Noble, J. A., Nossal, N. G., Konigsberg, W. H., and

Williams, K. R. (1982) J. Biol. Chem. 257, 8972-8979.
169. Godson, G. N., Barrell, B. G., Staden, R., and Fiddes, J. C. (1978) Nature 276, 236-247.
170. Priess, H., Kemp, D., Kahmann, R., Brauer, B., and Delius, H. (1982) Mol. Gen. Genet. 186, 315-321.
171. Pannekoek, H., Maat, J., van den Berg, E., and Noordermeer, (1980) Nucleic Acids Res. 8, 1535-1550.
172. van den Elzen, P. J. M., Maat, J., Walters, H. H. B., Veltkamp, E., and Nijkamp, H. J. J. (1982) Nucleic Acids Res. 10, 1913-1928.
173. Ogata, R. T., Winter, C., and Levine, R. P (1982) J. Bacteriol. 151, 819-827.
174. Heffron, F., McCarthy, B. J., Ohtsubo, H., and Ohtsubo, E. (1979) Cell 18, 1153-1163.
175. LeClerc, J. E. and Istock, N. L. (1982) Nature 297, 596-598.
176. Wulff, D., Mahoney, M., Shatzman, A., and Rosenberg, M., personal communication.
177. Meyer, B. J., Kleid, D. G., and Ptashne, M. (1975) Proc. Natl. Acad. Sci. 72, 4785-4789.
178. Cesareni, G. (1982) J. Mol. Biol. 160, 123-126.
179. LeClerc, J. E., personal communication.
180. Wulff, D. L., Beher, M., Izumi, S., Beck, J., Mahoney, M., Shimatake, H., Brady, C., Court, D., and Rosenberg, M. (1980) J. Mol. Biol. 138, 209-230.
181. Jaurin, B., Grundstrom, T., and Normark, S., personal communication.
182. Meyer, B. J., Maurer, R., and Ptashne, M. (1980) J. Mol. Biol. 139, 163-194.
183. Hawley, D. K. (1982) Ph.D. thesis. Harvard University, Cambridge, Massachusetts.
184. Bertrand, K., personal communication.
185. Reznikoff, W. S. and Abelson, J. N. (1978) in The Operon, Miller, J. H. and Reznikoff, W. S., Eds, pp. 221-243, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
186. Kleid, D., Humayun, Z., Jeffrey, A., and Ptashne, M. (1976) Proc. Natl. Acad. Sci. 73, 293-297.
187. Rosenberg, M., personal communication.
188. Grana, D., Youderian, P., and Susskind, M., personal communication.
189. Rosen, E. D., Hartley, J. L., Matz, K., Nichols, B. P., Young, K., M., Donelson, J. E., and Gussin, G. N. (1980) Gene 11, 197-205.
190. Busby, S. J. W., Aiba, H., and deCrombrugghe, B. (1982) J. Mol. Biol. 154, 211-227.
191. Berman, M. L. and Landy, A. (1979) Proc. Natl. Acad. Sci. 76, 4303-4307.
192. Oppenheim, D. S., Bennett, G. N., and Yanofsky, C. (1980) J. Mol. Biol. 144, 133-142.
193. Rothstein, S. J. and Reznikoff, W. S. (1981) Cell 23, 191-199.
194. Miozzari, G. and Yanofsky, C. (1978) Proc. Natl. Acad. Sci. 75, 5580-5584.
195. Reznikoff, W. S., Maquat, L. E., Munson, L. M., Johnson, R. C., and Mandecki, W. (1982) in Promoters: Structure and Function, Rodriguez, R. L. and Chamberlin, M. J., Eds., pp.80-95, Praeger Publishers, New York.
196. Calvo, J., personal communication.
197. Mozola, M., Friedman, D. I., Crawford, C., Wulff, D. L., Shimatake, H., and Rosenberg, M. (1979) Proc. Natl. Acad. Sci. 76, 1122-1125.
198. Wilcox, G., Al-Zarban, S., Cass, L. G., Clarke, P., Heffernan, L., Horwitz, A. H., and Miyada, C. G. (1982) in Promoters: Structure and Function, Rodriguez, R. L. and Chamberlin, M. J., Eds., pp.183-194,

Praeger Publishers, New York.
199. Post, L. E., Arfsten, A. E., Nomura, M., and Jaskunas, S. R. (1978) Cell 15, 231-236.
200. Gilbert, W. (1976) in RNA Polymerase, Losick, R. and Chamberlin, M., Eds., pp. 193-205, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
201. Horwitz, A. H., Morandi, C. and Wilcox, G. (1980) J. Bacteriol. 142, 659-667.
202. Oppenheim, D. S. and Yanofsky, C. (1980) J. Mol. Biol. 144, 143-161.
203. Maquat, L. E. and Reznikoff, W. S. (1980) J. Mol. Biol. 139, 551-556.
204. Seeburg, P. H., Nusslein, C., and Schaller, H. (1977) Eur. J. Biochem. 74, 107-113.
205. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
206. Siebenlist, U., Simpson, R. B., and Gilbert, W. (1980) Cell 20, 269-281.
207. Stephano, J. E. and Gralla, J. D. (1982) Proc. Natl. Acad. Sci. 79, 1069-1072.
208. Mandecki, W. and Reznikoff, W. S. (1982) Nucleic Acids Res. 10, 903-912.
209. Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., VanCleemput, M., and Wu, A. M. (1981) Nucleic Acids Res. 9, 6647-6668.
210. Davis, M. and Kleckner, N., personal communication.
211. Way, J. and Kleckner, N., personal communication.