Nucleic Acids Research, 1993, Vol. 21, No. 7 **1507–1516**

Compilation of *E.coli* mRNA promoter sequences

Shlomit Lisser and Hanah Margalit*
Department of Molecular Genetics, Hebrew University—Hadassah Medical School, Jerusalem 91010, Israel

Received November 12, 1992; Revised and Accepted March 5, 1993

ABSTRACT

An updated compilation of 300 *E.coli* mRNA promoter sequences is presented. For each sequence the most recent relevant paper was checked, to verify the location of the transcriptional start position as identified experimentally. We comment on the reliability of the sequence databanks and analyze the conservation of known promoter features in the current compilation. This database is available by E-mail.

INTRODUCTION

We present an extensive compilation of *E.coli* mRNA promoter sequences, based on experimentally identified transcriptional start positions (1–200). Previous compilations of *E.coli* promoter sequences (201–203) comprise one of the most extensively studied datasets of sequences with a common function. These sequences have been used in studies attempting to extract sequence dependent features that correlate with promoter function (204–207) and with its specific recognition by RNA polymerase (208–210). In many studies, *E.coli* promoter database has been used as a test case to evaluate the efficiency of various algorithms aimed to identify consensus sequences (211–215). Our interest is in identifying sequences dependent characteristics that correlate with promoter function. For such computational studies, based on sequence data alone, the accuracy of the sequence data is essential. We therefore carried out a careful screening of the databanks and relevant literature and organized an expanded and updated database of experimentally identified mRNA transcriptional start positions and their corresponding promoter sequences, to be used in further studies. We believe that this updated database of *E.coli* promoter sequences will be useful to other researchers as well. The sequences can be obtained by E-mail upon request to HANAH@HUMD2.HUJI.AC.IL

COMPILATION PROCEDURE

The compilation includes 300 DNA sequences extracted from Genbank release 71 and EMBL release 30. For every mRNA promoter the relevant paper was checked to verify that the transcriptional start position as annotated in the databanks had been determined experimentally. When there were recent papers on promoters reported in previous compilations (202–203), or in previous versions of the databank, we checked if a more accurate identification procedure of the transcriptional start position had been carried out, and, when necessary, put in the appropriate corrections. Most of the promoters had been identified either by S1 nuclease or primer extension mapping strategies. For some promoters, sequencing of the mRNA or in vitro determination of the transcript size had been performed. In cases where there were previous S1 nuclease mapping and more recent mapping by primer extension, we referred to the position determined by the last procedure. When there were discrepancies between the literature and the information in the databanks, we used the information from the literature.

The sequences listed in Fig. 1 are aligned by the transcriptional start positions (indicated by lower case letters), and extend from −75 to +26 relative to the transcriptional start position. In sequences with several identified start positions, all the positions indicated in the experimental paper were marked, and the most prominent one was used for the alignment. Where no major start position was proposed in the experimental paper, the alignment was done by one nucleotide, preferably A. The sequences are listed alphabetically by the promoter name. Usually, the name of the promoter indicates the name of its gene. In several cases, the experimental papers assign to promoters names with no reference to their genes. These names were used in the compilation and are marked by a ~. Also listed are the corresponding references and the transcriptional start positions on *E.coli* chromosome in centisomes. Locations in centisomes are based on the file EcoSeq5 that contains the *E.coli* genes and their locations on the chromosome in base pairs (216). To get the centisome coordinates we divided each location in 46722 (1% of the genome size in EcoSeq5).

COMPILATION REMARKS

The ultimate goal of the sequence databanks is to serve as reliable sources of information regarding biological sequences. It should be possible to automatically extract sets of sequences of interest from the databanks and use them as databases for statistical and computational analyses, aimed to reveal new biological and structural facts that are encoded in the sequence. While generation of such databases is technically easy, we found that their reliability is not satisfactory. Especially, we would like to consider two points.

* To whom correspondence should be addressed

Gene	Transcriptional start position ↓	Reference	centisomes
aceB	CAACAAGTTATCAAGTATTTTTAATTAAAATGGAAATTGTTTTTGATTTTGCATTTTAAATGAGTAGTCTTAGTTgTGCTGAACGAAAAGAGCACAACGAT	(1)	90.87
aceE	ATCAAGCGCATATAAAAGCGCGGCAACTAAACGTAGAACCTGTCTTATTGAGCTTTCCGGCGAGAGTTCAATGGGaCAGGTTCCAGAAAACTCAACGTTAT	(2)	2.64
ada	TTCCTTGTCAGCGAAAAAAATTAAAGCGCAAGATTGTTGGTTTTTGCGTGATGGTGACCGGGCAGCCTAAAGGCtaTCCTTAACCAGGGAGCTGATTATGA	(3)	49.71
alaS	AGATTTGGCGAAATTTTGCCGACTGAACGCATACGGTATTTTACCTTCCCAGTCAAGAAAACTTATCTTATTCCCaCTTTTCAGTTACCAGCCCGGCGGTT	(4)	58.10
alkA	AGCGAAATGTTGCCGTCGCGACAACCGGAATATGAAAGCAAAGCGCAGCGTCTGAATAACGTTTATGCTGAAAGCgGATGAATAAGGAGATGCGATGTATA	(5)	46.21
amn-P1	GCTTGATAGCCTGGCGAAGATCATCCGATCTTCGCCTTACACTTTTGTTTCACATTTCTGTGACATACTATCGGAtGtGCGGTAATTGTATGGAACAGGAG	(6)	44.25
amn-P2	GGCGAAGATCATCCGATCTTCGCCTTACACTTTTGTTTCACATTTCTGTGACATACTATCGGATGTGCGGTAATTgtatGGAACAGGAGACACACATGAAT	(6)	44.25
ampC	GGCAAATGGGTTTTCTACGGTCTGGCTGCTATCCTGACAGTTGTCACGCTGATTGGTGTCGTTACAATCTAACGCaTCGCCAATGTAAATCCGGCCCGCCT	(7)	94.40
ansB	TTTAATTCTTCGTTTTGTTACCTGCCTCTAACTTTGTAGATCTCCAAAATATATTCACGTTGTAAATTGTTTAACgTCAAATTTCCCATACAGAGCTAAGG	(8)	66.77
araBAD	CCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATaCCCGTTTTTTTGGATGGAGTGAAAC	(9)	1.51
araC	TGGCATAGCAAAGTGTGACGCCGTGCAAATAATCAATGTGGACTTTTCTGCCGTGATTATAGACACTTTTGTTACgCGTTTTTGTCATGGCTTTGGTCCCG	(10)	1.52
araE	GTCGCAGCAATTTAATCCATATTTATGCTGTTTCCGACCTGACACCTGCGTGAGTTGTTCACGTATTTTTTCACTaTGTCTTACTCTCTGCTGGCAGGAAA	(11)	64.22
araFGH	GCACGTTCTCACTGTAATTCTGCGATGTGATATTGCTCTCCTATGGAGAATTAATTTCTCGCTAAAACTATGTcaaCACAGTCACTTATCTTTTAGTTAAA	(12)	42.81
araJ	TAACTATTCAGCAGGATAATGAATACAGAGGGGCGAATTATCTCTTGGCCTTGCTGGTCGTTATCCTGCAAGCTAtCACTTTATTGGCTACGGTGATTGGT	(12)	8.97
argCBH-P1	CGGTAAATCTCGATAAATGGCGGTAATTTGTTTTTCATTGTTGACACACCTCTGGTCATGATAGTATCAATATTCaTGCAGTATTTATGAATAAAAATACA	(13)	89.55
argCBH-P2	ATCAGAGCGCGGTAAATCTCGATAAATGGCGGTAATTTGTTTTTcATTGTTGACACACCTCTGGTCATGA	(13)	89.55
argE-P1	TGCCAGGTTAAACGTAAAACATTCACCTTACGGCTGGTGGGTTTTATTACGCTCAACGTTAGTGTATTTTTATTCaTAAATACTGCATGAATATTGATACT	(13)	89.55
argE-P2	CTGCGGATGCAAATCGGAGATTAACTTTCCCGCATCATTGCTTTGCGCTGAAACAGTCAAAGCGGTTATGTTCATaTGCGGATGGCGATTTACATAGGTCA	(13)	89.55
argF	TTGCCTCCACCTTTGTAAGAAAGAATTGTGAAATGGGGTTGCAAATGAATAATTACACATATAAAGTGAATTTTaaTTCAATAAGTGGCGTTCGCCATGCG	(14)	6.38
argR-P1	CTTTCATAACATTATTTCAGCCTTCTTCAGGGCTGACTGTTTGCATAAAAATTCATCTGTATGCACAATAATGTTgTATCAACCACCATATCGGGTGACTT	(15)	72.85
argR-P2	ATGCCGTGACGCAGGCATGTTTCTCAATAACGAAATTTGATAAAATCCCGCTcTTTCATAACATTATTTCAGCCTTCT	(15)	72.85
aroF	GACTAAATTGCCTGTGTAAATAAAAATGTACGAAATATGGATTGAAAACTTTACTTTATGTGTTATCGTTACGTCaTCCTCGCTGAGGATCAACTATCGCA	(16)	58.95
aroG	GCTAATTCTAGGATTAATCCGTTCATAGTGTAAAACCCCGTTTACACATTCTGACGGAAGATATAGATTGGAAGTaTTGCATTCACTAAGATAAGTATGGC	(17)	17.01
aroH	CTCGATATCATGGGCCTTAGTCGCCGAATGTACTAGAGAACTAGTGCATTAGCTTATTTTTTTGTTATCATGCTAaCCACCCGGCGAGGTGTGACACACCT	(18)	38.46
aroP	TTTAAGCAACTCATCTTCAACCATGCATAAAGCGGGTGCATTCGCTGCCGCATACCATTATTCTTGATCTGACGGaAGTCTTTTTGTAACAATTCAAACTT	(19)	2.62
asnB	ATCACCATTACGTTTTTTATTTTTTCACCGCCAAATCGCTTGCGCACAGTGCTGCTTTGCGGCATTTTTTTAAACaAGCAAACACAACAAGCAACAAATAC	(20)	15.18
aspC	ACGTCTTGCAAAAACAGCCTGCGTTTTCATCAGTAATAGTTGGAATTTTGTAAATCTCCCGTTACCCTGATAGCGgaCTTCCCTTCTGTAACCATAATGGA	(21)	21.24
atpB-P1	AGACCCCTTCTGGGGCGTCTCTGCAATAAGCGGGGGCCTGGCAGTCTTTCTGCCTAACGTTTTGTTTATGATATTtGCCTGGCGTCACCAGGCGCATACAC	(22)	84.53
atpB-P2	GGCGAAGCTTTCAAAGTTCTGGCGATGTTGGTGTTACTGGTGGTGGCGTTGGCGGTTTTAAAGGCGGTATTCTTGcCGCTGATCGTTACGTGGGTTTTGGT	(22)	84.53
atpI	AAAATATTTAAACATTTATTCACCTTTTGGCTACTTATTGTTTGAAATCACGGGGGCGCACCGTATAATTTGACCgCTTTTTGATGCTTGACTCTAAGCCT	(22)	84.54
bioA	AATGTCCAGCGTGGGCGGTGAGCCATGGGGCTTCCAAAACGTGTTTTTTGTTGTTAATTCGGTGTAGACTTGTAAaCCTAAATCTTTTCAATTTGGTTTAC	(23)	17.52
bioB	GTTGGTCAAAGGCAAGATCGTCCGTTGTCATAATCGACTTGTAAACCAAATTGAAAAGATTTAGGTTTACAAGTCtACACCGAATTAACAACAAAAAACAC	(23)	17.52
btuB	TGTCCAGATCTTGATGAATTCCTATTTGTGAGCTACGTCTGGACAGTAACTTGTTACAACCTGTAGCATCCACTTgCCGGTCCTGTGAGTTAATAGGGAAT	(24)	89.74
c62.5	AGCAAACGAGAGCAGGATCACCTGCTCTCGCTTGAAATTATTCTCCCTTGTCCCCATCTCTCCCAcAtCCTGTtttTAACCTTAAAATGGGCATTATTGAG	(25)	10.74
carAB-P1	GCATGAGCCACAAAATAATATAAAAAATCCCGCCATTAAGTTGACTTTTAGCGCCCATATCTCCAGAATGCCGCCgTTTGCCAGAAATTCGTCGGTAAGCA	(26)	0.64
carAB-P2	TGCCGCCGTTTGCCAGAAATTCGTCGGTAAGCAGATTTGCATTGATTTACGTCATCATTGTGAATTAATATGCAAaTAAAGTGAGTGAATATTCTCTGGAG	(27)	0.64
carB	CCAGGGGCACCCTGAAGCCAGCCCTGGTCCACACGACGCCGCGCCGTTGTTCGACCACTTTATCGAGTTAATTGAgCAGTACCGTAAAACCGCTAAGTAAT	(28)	0.67
cdh	GGATAATTATCAAACCCGGTGGTTTCTCGCGACCGGGTTTTTTATTTGTCACGATTTTGCGTTACCCTTGCATCTcTTTGAGGTACAGGGAAAAAAAGATG	(29)	88.59
cir-P1	TCAATTCCATTTCCCTGACAAATCATGCGGATATAAAATTTAACATTTGGATTGATAATTGTTATCGTTTGCATTaTCGTTACGCCGCAATCAAAAAAGGC	(30)	48.33
cir-P2	CCTGACAAATCATGCGGATATAAAATTTAACATTTGGATTGATAATTGTTATCGTTTGCATTATCGTTACGCCGCaATCAAAAAAGGCTGACAAATCAGAG	(30)	48.33
clcA	ATCCCGTATTGCAAAGGCTAAAAAAAGGTATTGGACCGCATGACACGCGAATCTTAGCATTCATGTTTGAAGCACcAACTCATCGGTGTTTCAACCATCAG	(31)	-
clpB	GTACGCGTGATTTTCTTTTCACATTAATCTGGTCAATAACCTTGAATAATTGAGGGATGACCTCATTTAATCTCCaGTAGCAACTTTGATCCGTTATGGGA	(32)	58.75
clpP-P1	TGAATTTTGCATGGAACCGTGCGAAAAGCCTCTTTCGGTGTTAGCGTAACAACAAAAGATTGTTATGCTTGAAATaTGGTGATGCCGTACCCATAACACAG	(33)	9.92
clpP-P2	TGCTTGAAATATGGTGATGCCGTACCCATAACACAGGACTAGCTGATAATCCGTCCATAAGGTTACAATCGGTACaGCAGGTTTTTTCAATTTTATCCAGG	(33)	9.92
crp	AACGGAAGGCGACCTGGGTCATGCTGAAGCGAGACACCAGGAGACACAAAGCGAAAGCTATGCTAAAACAGTCAGgATGCTACAGTAATACATTGATGTAC	(34)	75.05
crr-P2-I	ATTTCTTTAGTATCGGCACCAATGATTTAACGCAGTACACTCTGGCAGTTGACCGTGGTAATGATATGATTTCACaCCTTTACCAGCCAATGTCACCGTCC	(35)	54.64
crr-P2-II	TGCTTCTCATGCTGAAGGCAAATGGACTGGCATGTGTGGTGAGCTTGCTGGCGATGAACGTGCTACACTTCTGTTgCTGGGGATGGGTCTGGACGAATTCT	(35)	54.65
cspA-P1	GTTTTCCAACCGATTAATCATAAATATGAAAAATAATTGTTGCATCACCCGCCAATGCGTGGCTTAATGCACATCaaCGGTTTGACGTACAGACCATTAAA	(36)	80.10
cspA-P2	CCAATGCGTGGCTTAATGCACATCAACGGTTTGACGTACAGACCATTAAAGCAGTGTAGTAAGGCAAGTCCCTTCaAGAGTTATCGTTGATACCCCTCGTA	(36)	80.10
cya	GAGAATAAACGGTGCTACACTTGTATGTAGCGCATCTTTCTTTACGGTCAATCAGCAAGGTGTTAAATTGATCACgTTTTAGACCATTTTTTCGTCGTGAA	(37)	86.02
cyoA	GACCTGGCAGCCAAATCCAAGTAACAGGAATTTAATCATGTTTACAGTAATGTAACCTTCCCGTAAAATGCCCACaCACTTTAAACGCCACCAGATCCCGT	(38)	9.82
cysD	GTTATTCCGTTAAGGAACTACTCATTCTAATTGGTAATTTCATTCGTTCTCTTACGCTCCCTATAGTCGAAACATcTGATGGCAAGAAAATAGCGGTATTG	(39)	61.96
cysE	TGAAATGCCAATAACCGAGGAAATTTATCAAGTATTATATTGCGGAAAAAACGCGCGCGAGGCAGCATTGACTTTaCTAGGTCGTGCACGCAAGGACGAGC	(40)	81.47
cysJIH	CAGGTTAGTCGATTTGGTTATTAGTTATCGCTATCCCGTCTTTAATCCACACCGTTTGCCCCGTTAACCTTACCTtCTCTTCTGTTTTATGGGCGCTGACA	(41)	62.31
cysK-P1	CATTATTTCCCTTCTGTATATAGATATGCTAAATCCTTACTTCCGCATATTCTCTGAGCGGGTATGCTACCTGTTgTATCCCAATTTCATACAGTTAAGGA	(42)	54.58
cysK-P2	GTGGCTTATGCCGCCCCTTATTCCATCTTGCATGTCATTATTTCCCTTCTGTATATAGATATGCTAAATCCTTACtTCCGCATATTCTCTGAGCGGGTATG	(42)	54.58
dapA	AAAAAACATGCATACAACAATCAGAACGGTTCTGTCTGCTTGCTTTTAATGCCATACCAAACGTACCATTGAGACaCTTGTTTGCACAGAGGATGGCCCAT	(43)	55.82
dapB	CCCTGTTTATCATTAATTTCTAATTATCAGCGTTTTTGGCTGGCGGCGTAGCGATGCGCTGGTTACTCTGAAAACgGTCTATGCAAATTAACAAAAGAGAA	(44)	0.61
dapD	TGCGCTTAATAACGAGTTGCAGCAGGAAGTGCATCAGCGGTTGACAGAGGCCCTCAATCCAAACGATAAAGGGTGaTgtGTTTACTGATATGAAAAGAGTT	(45)	4.11
deo-P1	TGGCAGGGTTCTGTCGAGGTAACGCCAGAAACGTTTTATTCGAACATCGATCTCGTCTTGTGTTAGAATTCTAACaTACGGTTGCAACAACGCATCCAGTT	(46)	99.49
deo-P2	GCAAACTTGTAAGTAGATTTCCTTAATTGTGATGTGTATCGAAGTGTGTTGCGGAGTAGATGTTAGAATACTAACaAACTCGCAAGGTGAATTTTATTGGC	(46)	99.50
deo-P3	AAACTTGCCGATAAAGCACCGGAAAGCACACCAACTGTCTATCGCCGTATCAGCGAATAACGGTATACTGATCTGaTCATTTAAATTTGAAGCACTGAGTA	(47)	99.55
div	AACAAGGATTTTCACGTTTGTGTTACCTGTATGAGACGAGAGTTAACCGGACAAGTGTGCCATAATCTCGCGGCCaGGCATACTTGCGAAGATTTCAGGTA	(48)	52.57
dnaA-P1	TCTATGGTCATTAAATTTTCCAATATGCGGCGTAAATCGTGCCCGCCTCGCGGCAGGATCGTTTACACTTAGCGAgTTCTGGAAAGTCCTGTGGATAAATC	(49)	83.65
dnaA-P2	GAAAGTCCTGTGGATAAATCGGGAAAATCTGTGAGAAACAGAAGATCTCTTGCGCAGTTTAGGCTATGATCCGcggtcccGATCGTTTTGCAGGATCTTGA	(50)	83.65
dnaK-P1	AAAAGCACAAAAAATTTTTGCATCTCCCCCTTGATGACGTGGTTTACGACCCCATTTAGTAGTcaaCCGCAGTGAGTGAGTCTGCAAAAA	(25)	0.26
dnaK-P2	AACCGCAGTGAGTGAGTCTGCAAAAAAATGAAATTGGGCAGTTGAAACCAGACGTTTCGCCCCTATTACAGACtcaCAACCACATGATGACCGAATATATA	(25)	0.27
dnaN-P1	AAAAAGGCCGACGAAAACGACATTCGTTTGCCGGGCGAAGTGGCGTTCTTTATCGCCAAGCGTCTACGATCTAACgTaCGTGAGCTGGAAGGGGCGCTGAA	(51)	83.63
dnaN-P2	CATCGACTTCGTGCGTGAGGCGCTGCGCGACTTGCTGGCATTGCAGGAAAAACTGGTCACCATCGACAATATTCAgAAGACGGTGGCGGAGTACTACAAGA	(51)	83.62
dnaN-P3	GACTTGCTGGCATTGCAGGAAAAACTGGTCACCATCGACAATATTCAGAAGACGGTGGCGGAGTACTACAAGATCaAAGTCGCGGATCTCCTTTCCAAGCG	(51)	83.62
dnaN-P4	CGGTGCTTCATGCCTGCCGTAAGATCGAGCAGTTGCGTGAAGAGAGCCACGATATCAAAGAAGATTTTTCAAATTtAATCAGAACATTGTCATCGTAAACC	(51)	83.62
dnaN-P5	AGAAGATTTTTCAAATTTAATCAGAACATTGTCATCGTAAACCTATGAAATTTACCGTAGAACGTGAGCATTTATtAAAACCGCTACAACAGGTGAGCGGT	(51)	83.62

dnaQ-P1	CTCCATACGGTTGTTGGTGGTGCGGGTGTAGCCAGCGCTAAAGGTTTTCTCGCGTCCGCGATAGCGTAAAATAGCgCCGTAACCCCCAGGTCCTGGATTGC	(52)	5.23
dnaQ-P2	CCCCAGGTCCTGGATTGCCCAGACACGAACCATCGGTGAAAATTTCTACCTGTTTAAGCATCTCTGGTAGACTTCcTGTAATTGAATCGAACTGTAAAACG	(52)	5.23
dnaT	TGGTGCGTTATCCATCGGTCTTTCCATTCCTGGATTATGGTTGTACCGCAAGCCCCTCGCGTATAAAATTGCCTCgCTTAAGCACACGGATGAGAGACAGC	(53)	99.18
drpA	AAGTCTTTGCGCTGGAACCGCGTTAAATTCACGCCCTTCTCTTTTGACATTTCTTTTGCACTGGTAAACTAAATCaCTTTTTTTTGTCCCAGGCTCGCCTT	(54)	4.85
dsda	TACGGACTCGCTTTTCTGGCTAGTGAAGCAATATTGCGGCGCTACGCTCAATGAAACATTTAAATACTATACGACaGCGACATTTATCGCTTCAGTCGTCG	(55)	53.48
fabB	TAAGGCTGCGCAAATTTCTCTATTAAATGGCTGATCGGACTTGTTCGGCGTACAAGTGTACGCTATTGTGCATTCgAAACTTACTCTATGTGCGACTTACA	(56)	52.65
fadL	GTTTCTTAGATCATATTTGAAAAAAGATAGAAACATACTTGCAACATTCCAGCTGGTCCGACCTATACTCTCGCCaCTGGTCTGATTTCTAAGATGTACCT	(57)	53.07
fdhF	TTTGAATAAATAGTGCCCGTAATATCAGGGAATGACCCCACATAAAATGTGGCATAAAAGATGCATACTGTAGTCgAGAGCGCGTATGCGTGATTTGATTA	(58)	92.64
fepA	GCCGTCACACCATAACCCCATGTTTACTGTGCAATTTTTCATTGATTGCAGAAATATATTGATAATATTATTGATaaCTATTTGCATTTGCAATAGCGTAA	(59)	13.27
fes	TAATTAATGTCCGCGCTTCCCACGGCGCGCCATTACGCTATTGCAAATGCAAATAGTTATCAATAATATTATCaAtATATTTCTGCAATCAATGAAAAATT	(59)	13.27
flaA	AACGTCAGAGGTAGCACCGTAATCCGCGTCTTTTCCCCGCTTTGTTGCGCTCAAGACGCAGGATAATTAGCCGATaAGCAGTAGCGACACAGGAAGACCGC	(60)	43.51
frd	AGCCCTCTTGCGCACTAAAAAAATCGATCTCGTCAAATTTCAGCATTATCCATCAGACTATACTGTTGTACCTATaAAGGAGCAGTGGAATAGCGTTCGCA	(61)	94.48
fumA	AGTAACCTGGAGCCGCAAAAAGTCGTACTAGTCTCAGTTTTTGTTAAAAAAGTGTGTAGGATATTGTTACTCGCTttAACAGGGCAACGGAACACCCGCC	(62)	36.32
fur-Pa	GCCACGTTTTTATTAACAATATTTGCCAGGGACTTGTGGTTTTCATTTAGGCGTGGCAATTCTATAATGATACGCaTTATCTCAAGAGCAAATTCTGTCAC	(63)	15.44
fur-Pb	AAGCTGTGCCACGTTTTTATTAACAATATTTGCCAGGGACTTGTGGTTTTCATTTAGGCGTGGCAATTCTATAATgATACGCATTATCTCAAGAGCAAATT	(63)	15.44
g30k-rpmF-P1	AGGCTAAAATAAGTTTAGGCATATTTTTTTCCATCAGATATAGCGTATTGATGATAGCCATTTTAAACTATGCGCtTCGTTTTGCAGGTTGATGTTTGTTA	(64)	24.79
g30k-rpmF-P2	ATTATTGGCAAAAGGCAACCACAGGCTGCCTTTTTCTTTGACTCTATGACGTTACAAAGTTAATATGCGCGCCCTaTGCAAAAGGTAAAATTACCCCTGAC	(64)	24.78
gal-P1	AATTCTTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTATGCTATGGTTaTTTCATACCATAAGCCTAATGGAGCGAATT	(65)	17.14
gal-P2	GGCTAAATTCTTGTGTAAACGATTCCACTAATTTATTCCATGTCACACTTTTCGCATCTTTGTTATGCTATGGTTaTTTCATACCATAAGCCTAATGGAGC	(65)	17.14
gene-x	TCACTTCATCACATTCTTTCTGAAAAACACCAAAGAACCATTTACATTGCAGGGCTATTTTTTATAAGATGCATTtGAGATACATCAATTAAGATGCAAAA	(66)	57.69
glnA-P1	TCGCTTTCACGGAGCATAAAAAGGGTTATCCAAAGGTCATTGCACCAACATGGTGCTTAATGTTTCCATTGAAGCaCTATATTGGTGCAACATTCACATCG	(67)	87.49
glnA-P2	GCACGGATGGTTGCGCATGATAACGCCTTTTAGGGGCAATTTAAAAGTTGGCACAGATTTCGCTTTATCTTTTTaCGCGACACGGCCAAAATAATTGCAG	(67)	87.49
glnH-P1	TTTTTGTATCCACATCATCACACAATCGTTACATAAAGATTGTTTTTTCATCAGGTTTTACGCTAAATAATCACTgTGTTGAGTGCACAATTTTAGCGCAC	(68)	18.34
glnH-P2	GTGCATCACGTTTTTGCCGCATCTCGAAAAATCAAGGAGTTGCAAAACTGGCACGATTTTTTCATATATGTGAATgTCACGCAGGGGATCGTCCCGTGGAT	(68)	18.34
glnL	TCTCTGCATGGGCTTTTTTCTCCGTCAATTCTCTGATGCTTCGCGCTTTTTATCCGTAAAAAGCTATAATGCACTaAAATGGTGCAACCTGTTCAGGAGAC	(69)	87.45
glnS	TAATCCTTGTTAGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCGCTTATAAGATCATACGCCgTTaTACGTTGTTTACGCTTTGAGGAATC	(70)	15.33
glpD	ATAATGTTATACATATCACTCTAAAATGTTTTTTCAATGTTACCTAAAGCGCGATTCTTTGCTAATATGTTCGATaACGAACATTTATGAGCTTTAACGAA	(71)	76.72
gltA-P1	ATCATCTAATTTGACAATCATTCAACAAAGTTGTTACAAACATTACCAGGAAAAGCATATAATGCGTAAAAGTTAtGAAGTCGGTATTTCACCTAAGATTA	(72)	16.37
gltA-P2	CTTATGTAACAGTGTGGAGTATTGACCAATTCATTCGGGACAGTTATTAGTGGTAGACAAGTTTAATAATTCGGAtTGCTAAGTACTTGATTCGCCATTTA	(72)	16.37
gltB	ATTTTGCGCTAAAGCACATTTCTGTACCAATAAGCTTGCCATTTGACCTGTATCAGCTTTCCCGATAAGTTGGAAaTCCGCTGGAAGCTTTCTGGATGAGC	(73)	72.20
glyA	TAGCCTGAAGGTAATCGTTTGCGTAAATTCCTTTGTCAAGACCTGTTATCGCACAATGATTCGGTTATACTGTTCgCCGTTGTCCAACAGGACCGCCTATA	(66)	57.68
gnd	GTAATATAAAGCCGTAAGCATATAAGCATGGATAAGCTATTTATACTTTAATAAGTACTTTGTATACTTATTTGCgAACATTCCAGGCCGCGAGCATTCAG	(74)	45.22
groE	CCCGATTTTGTGCTGATCAGAATTTTTTTTCTTTTCCCCCTTGAAGGGGCGAAGCCATCCCCATTTCTCTGGTCaCCAGCCGGGAAACCACGTAAGCTCC	(25)	94.21
grpE	GGTCCTTGTTGCGAAGATTGATGACAATGTGAGTGCTTCCCTTGAAACCCTGAAACTGATCCCCATAATAAGCGAaGTTAGCGAGATGAATGCGAAAAAAA	(75)	59.13
guaBA	TGGGCGGTGTCGTCTTTGAGTGTAAAGTACCAGTGACCGGAAGCTGGTTGCGTGAAATTAGAAATTTCGCCGctgaTCCAAACCTGTCCCATCTCATGCTC	(76)	56.55
gut	CAAATTAAAATATTTATCTTTCATTTTGCGATCAAAATAACACTTTTAAATCTTTCAATCTGATTAGATTAGGTTgCCGTTTGGTAATAAAACAATAAATC	(77)	60.87
gyrA	AGTATATCAGGCATTGGATGTGAATAAAGCGTATAGGTTTACCTCAAACTGCGCGGCTGTGTTATAATTTGCGACcTTTGAATCCGGGATACAGTAGAGGG	(78)	50.37
gyrB	AGTGCTGAACACGTTATAGACATGTCGGACGAAAATTCGAAGATGTTTACCGTGGAAAAGGGTAAAATAACGGATtAACCCAAGTATAAATGAGCGAGAAA	(79)	83.57
hemA-P1	ACAAGTCCTTGAGATACGTTGCAGTTATAACCCTTAATGCTAGCGTTACCGTCCGCTATCGTCTATGTTCAAGTTgTCTTAATTGCCAGAATCTAACGGCT	(80)	27.20
hemA-P2	TAACGGCTTTCGGCAATTACTCCAAAAGGGGGCGCTCTCTTTTATTGATCTTACGCATCCTGTATGATGCAAGCAgACTAACCCTATCAACGTTGGTATTA	(80)	27.20
hisA	TTCGTCTCACTGACAATACCTTGATCTACAAACTAATTAATAAATAGTTAATTAACGCTCATCATTGTACAATGAaCTGTACAAAAGTGGATTGACATGCG	(81)	44.98
hisB	AACCAACTACATTCTGGCGCGCTTTAAAGCCTCCAGTGCGGTGTTTAAATCTTTGTGGGATCAGGGCATTATCTTaCGTGATCAGAATAAACAACCCTCTT	(82)	45.06
hisG	TAAACAAATTCATTGTCATAAAATATATAAAAAAGTTCTTGCTTTCTAACGTGAAAGTGGTTTAGGTTAAAAGACaTCAGTTGAATAAACATTCACAGAGA	(83)	44.99
hisS	GTCGAATTGACGTTCAGCAGGTTGAAAAATAATAACGTGATGGGAAGCGCCTCGCTTCCCGTGTATGATTGAACCaGCATGGCTCCCGAAACATTGAGGGA	(84)	56.69
hsd	CAAACCTCTAGGCAGGTCATTCGTAGCCTAATGTCCGAACTGCTAAAGCATCCAAGTTGCTGTAGAATCACCGCCaATTACATAAGCCTGAAATAAGTGGA	(85)	98.44
htrA	AGACGAACAATAAATTTTTACCTTTTGCAGAAACTTTAGTTCGGAACTTCAGGCTATAAAACGAATCTGAAGAACaCAGCAATTTTGCGTTATCTGTTAAT	(86)	3.88
ileR-P1	CTCACGGTGCTACAGGAAACGTTTTACCCCAGCGCATTTGGGAATGCCAGCAGAGTTTTTTATGCGTTAATAGATaTGGAAAGCGGCTCGGAGAGAAAAGC	(87)	95.88
ileR-P2	CAGCAGAGTTTTTTATGCGTTAATAGATATGGAAAGCGGCTCGGAGAGAAAAGCAAAAGGTGAGGGAATTACAAAaCAGAATGCGAGTCGTCTCGCAAAAA	(87)	95.88
ilvBN	AAACGTGATCAACCCCTCAATTTTCCCTTTGCTGAAAAATTTTCCATTGTCTCCCCTGTAAAGCTGTGCTTGTATaAATATTGTTAAACACAAAACCAACA	(88)	82.98
ilvC	GTTGACGTTGCAAAAATTGCAATGTGACGTTGTGAATATATCAATTTCCGCAATAAATTTCCTGTCATATAGTGAaTTCAATCTCGCAAACGCGAACCGAA	(89)	85.31
ilvGMEDA-P1	ACAACATTCATACTGAAATTGAATTTTTTTCACTCACTATTTTATTTTTAAAAAACAACAATTTATATTGAAATTaTTAAACGCATCATAAAAATCGGCCA	(90,91)	85.14
ilvGMEDA-P2	ATTATTAAACGCATCATAAAAATCGGCCAAAAAATATCTTGTACTATTTACAAAACCTATGGTAACTCTTTAGGCaTTCCTTCGAACAAGATGCAAGAAAA	(90,91)	85.15
ilvGMEDA-Pe	CCGCCAGCGATGCACAAAATATAAATATCGAATTGACCGTTGCCAGCCCACGGTCGGTCGACTTACTGTTTAGTCaGTTAAATAAACTGGTGGACGTCGCA	(92)	85.19
ilvIH-P1	GTTTATTCTTATTACCCCGTGTTTATGTCTCTGGCTGCCAATTGCTTAAGCAAGATCGGACGGTTAATGTGTTTTaCACATTTTTTCCGTCAAACAGTGAG	(93)	1.84
ilvIH-P2	ATTTATTATCAATTTAATCCTCTGTAATGGAGGATTTTATCGTTTCTTTTCACCTTTCCTCCTGTTTATTCTTATtACCCCGTGTTTATGTCTCTGGCTGC	(93)	1.84
ilvIH-P3	GGTTTATTCTGCATTTTTTATTGAATGTAGAATTTTATTCTGAATGTGTGGGCTCTCTATTTTAGGATTAATTAAaAAAATAGAGAAATTGCTGTAAGTTG	(93)	1.83
ilvIH-P3'	TTTATTCTGAATGTGTGGGCTCTCTATTTTAGGATTAATTAAAAAAATAGAGAAATTGCTGTAAGTTGTGGGATTcAGCCGATTTATTATCAATTTAATCC	(93)	1.83
ilvIH-P3''	GGCTCTCTATTTTAGGATTAATTAAAAAAATAGAGAAATTGCTGTAAGTTGTGGGATTCAGCCGATTTATTATCAaTTTAATCCTCTGTAATGGAGGATTT	(93)	1.84
ilvIH-P4	GCAGAATAGGTCAGACATGAATGTCTGGTTTATTCTGCATTTTTTATTGAATGTAGAATTTTATTCTGAATGTGTgGGCTCTCTATTTTAGGATTAATTAA	(93)	1.83
ilvY	TATTGTTCGGTTCGCGTTTGCGAGATTGAATTCACTATATGACAGGAAATTTATTGCGGAAATTGATATATTCACaACGTCACATTGCAATTTTTGCAACG	(89)	85.31
infC	GCGCATATCTTCTGTACTGAAGAACAAATTCGCGATGAAGTTAACGGATGTATCCGTTTAGTCTATGATATGTACaGCACTTTTGGCTTCGAGAAGATCGT	(94)	38.74
katE	CTGGCTTCACTAAACGCATATTAAAAATCAGAAAAACTGTAGTTTAGCCGATTTAGCCCCTGTACGTCCCGCTTTgCGTGTATTTCATAACACCGTTTCCA	(95)	39.00
lac	AATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGaATTGTGAGCGGATAACAATTTCACA	(96)	7.99
lacI	GACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCgGAAGAGAGTCAATTCAGGGTGGTGA	(97)	8.01
lep	ATACCACAATCCATACGGTATGAACATTTCCTCGCCTCAATGTTGTAGTGTAGAATGCGGCGTTTCTATTAATACaGACGTTAAGCTCAGAACAGCGACTT	(98)	58.15
leu	CAGAACTCACGCTGGCGGGACGTTTTTATTGCGTCAGGGTTGACATCCGTTTTTGTATCCAGTAACTCTAAAAGCaTATCGCATTCATCTGGAGCTGATTT	(99)	1.80
lexA	GAATTCGATAAATCTCTGGTTTATTGTGCAGTTTATGGTTCCAAAATCGCCTTTTGCTGTATATACTCACAGCaTAACTGTATATACACCCAGGGGGCG	(100)	91.74
livJ	CAATCCCCACGCAGATTGTTAATAAACTGTCAAAATAGCTATTCCAATATCATAAAAATCGGGATATGTTTTAGCaGAGTATGCTGCTAAAGCACGGGTAG	(101)	77.52
livK	CGCCCCTTTGTGCCGTTATTTTATGCTGACAAAGGCACTTTTTTCTGTTTGTCTATCAATAAATTCGGAATATTAtCTGTTCTTAATCGACTGAAAAATGG	(101)	77.48
lpd	GTTATTAGCGAATAGACAAATCGGTTGCCGTTTGTTGTTTAAAAATTGTTAACAATTTTGTAAAATACCGACGGAtAGAACGACCCGGTGGTGGTTAGGGT	(2)	2.74
lpp	CGAACGATCAAAAATAAGTGCCTTCCCATCAAAAAAATATTCTCAACATAAAAAACTTTGTGTAATACTTGTAACgCTACATGGAGATTAACTCAATCTAG	(102)	37.81
malE	GATCACACAAAGCGACGGTGGGGCGTAGGGGCAAGGAGGATGGAAAGAGGTTGCCGTATAAAGAAACTAGAGTCCgTTTAGGTGTTTTCACGAGCACTTCA	(103)	91.51
malI	AGAAGATAAAACGTTTTATCAAATTTTAGTGAGGCATAAATCACATTACGCAACGATAATAGCGGGTATAAGATAaATAAAAGGTAAAACGTTTTATCTGT	(104)	36.56
malK	CTCCCGCCTCCTCCCCCATAAAAAAGCCAGGGGGTGGAGGATTTAAGCCATCTCCTGATGACGCATAGTCAGCCCaTCATGAATGTTGCTGTCGATGACAG	(103)	91.52

malPQ	CATCCCCGCAACTCCTCCCTGCCTAATCCCCCGCAGGATGAGGAAGGTCAACATCGAGCCTGGCAAACTAGCGATaACGTTGTGTTGAAAATCTAAGAAAA	(105)	76.51
malT	GACACAGTGCAAATTCAGACACATAAAAAAACGTCATCGCTTGCATTAGAAAGGTTTCTGGCCGACCTTATAACCaTTAATTACGAAGCGCAAAAAAAATA	(106)	76.52
manA	AAACTGAGACTAGTACGACTTTTTGCGGCTCCAGGTTACTTCCCGTAGGATTCTTGCTTTAATAGTGGGATTAATtCCACATTAAAACAGGGATTGATCA	(62)	36.33
manX	TCTGCTAATCGAAAGTTAAATTACGGATCTTCATCACATAAAATAATTTTTCGATATCTAAAATAAATCGCGAAaCGCAGGGGTTTTTGGTTGTAGCCCT	(107)	40.95
melA	CAGGATCTGAGTTTATGGGAATGCTCAACCTGGAAGCCGGAGGTTTTCTGCAGATTCGCCTGCCATGATGAAGTTaTTCAAGCAAGCCAGGAGATCTGCAT	(108)	93.57
metA-P1	TAGTCTTTTCCTTGAAACTTTCTCACCTTCAACATGCAGGCTCGACATTGGCAAATTTTCTGGTTATCTTCAGCTaTCTGGATGTCTAAACGTATAAGCGT	(109)	90.84
metA-P2	CCCCTTCCTCATGCCGTGAGCGGCGAAGACTAATTACCATTTTCTCTCCTTTTAGTCATTCTTATATTCTAACGTaGTCTTTTCCTTGAAACTTTCTCACC	(109)	90.84
metBL	CTCAATCTATACGCAAAGAAGTTTAGATGTCCAGATGTATTGACGTCCATTAACACAATGTTTACTCTGGTGCCTgACATTTCACCGACAAAGCCCAGGGA	(110)	88.99
metj-P1	GCACCAGAGTAAACATTGTGTTAATGGACGTCAATACATCTGGACATCTAAACTTCTTTGCGTATAGATTGAGCAaATCCCAAATAGCCGTTAAAATTATA	(110)	88.99
metj-P2	GTCAATACATCTGGACATCTAAACTTCTTTGCGTATAGATTGAGCAAATCCCAAATAGCCGTTAAAATTATATGCaTTATCACGCCGACAGGTGCATTACA	(110)	88.99
metj-P3	AAAATTATATGCATTATCACGCCGACAGGTGCATTACACGATGTCACGGTAACGCCTGTACGGTAAACTATGCGGgTTTACGGTCAGTACCCACATCAACT	(110)	88.99
mtl	TATATTTATGTGATTGATATCACACAAAAGGCCGTCGACTGGACAGTTAACCGATTCAATGCCAGATTTCGCAGtatCTACAAAGGTCCGGCTACCTCTGC	(111)	81.24
nagB-P1	TCGGATTGACGCCTGTCTGCGCAAATCCAGGTTACGCTTAAAGATGCCTAATCCGCCAACGGCTTACATTTTACTtATTGAGGTGAATAATGAGACTGATC	(112)	15.28
nagB-P2	AAACGCGTGTCTTTTGTGAGTTTTGTCACCAAATATCGTTATTATCACTCCCTTTTACTGGCTAAACCAGAAAACtTATTTTATCATTCAAAAAATCAGGT	(112)	15.28
nagD	GTAAAACGCGCCATGCTCAACGGTATTTTGCTCCAGCATTTGCTGGAAAATTAATGTGCTTTTATAGTGGCGCTTaTTGTTGTCAATATTCTGGGTAGTCC	(113)	15.21
nagE	ACTCACAAAAGACACGCGTTTAATTTGCGATACGAATTAAATTTTCACACACTCTGTAGCAGATGATCTAACAATcTGATTACAGAACATCGGCAGTACAA	(112)	15.28
narG	ATTTTTCATCCCCATCACTCTTGATCGTTATCAATTCCCACGCTGTTTCAGAGCGTTACCTTGCCCTTAAACATTaGCAATGTCGATTTATCAGAGGGCCG	(114)	27.55
nrd	GGCCACTTTTTCCTTCCTGAGTTATCCACAAAGTTATGCACTTGCAAGAGGGTCATTTTCACACTATCTTGCAgtgaatCCCAAACATACCCCCTATATAT	(115)	50.50
nupG	TGCAAATTGGGAATGTTTGCAATTATTTGCCACAGGTAACAAAAAACCAGTCCGCGAAGTTGATAGAATCCCATCatCTCGCACGGTCAAATGTGCTTTTT	(116)	66.88
nusA	ATTTTACGTCCGTCTCGGTACACCAAATCCCAGCAGTATTTGCATTTTTTACCCAAAACGAGTAGAATTTGCCACgTTTCAGGCGCGGGGTGGAGCAGCCT	(117)	71.43
ompA-P1	ACCTTATACAAGACTTTTTTTTCATATGCCTGACGGAGTTCACACTTGTAAGTTTTCAACTACGTTGTAGACTTTaCATCGCCAAGGGTGCTCGGCATAAG	(118)	22.00
ompA-P2	TATACAAGACTTTTTTTTCATATGCCTGACGGAGTTCACACTTGTAAGTTTTCAACTACGTTGTAGACTTTACATcGCCAAGGGTGCTCGGCATAAGCCGA	(118)	22.00
ompB	ACGTGATCATATCAACAGAATCAATAATGTTTCGCCGAATAAATTGTATACTTAAGCTGCTGTTTAATATGCTTTgTAACAATTTAGGCTGAAATTCATAC	(119)	76.17
ompC	CGATAAATGAAACATCTTAAAAGTTTTAGTATCATATTCGTGTTGGATTATTCTGCATTTTTGGGGAGAATGGACttGCCGACTGATTAATGAGGGTTAAT	(120)	49.76
ompF	TTTTTGAAACCAAATCTTTATCTTTGTAGCACTTTCACGGTAGCGAAACGTTAGTTTGAATGGAAAGATGCCTGCaGACACATAAAGACACCAAACTCTCA	(121)	21.27
orf83-P1	CGAGCCGCTTTCCATATCTATTAACGCATAAAAAACTCTGCTGGCATTCCCAAATGCGCTGGGGTAAAACGTTTCcTGTAGCACCGTGAGTTATACTTTGT	(87)	95.88
orf83-P2	ATAAAAAACTCTGCTGGCATTCCCAAATGCGCTGGGGTAAAACGTTTCCTGTAGCACCGTGAGTTATACTTTGTAtAACTTAAGGAGGTGCAGATGCGTAT	(87)	95.89
oxyR	GCCGCTCCGTTTCTGTGAGCAATTATCAGTCAGAATGCTTGATAGGGATAATCGTTCATTGCTATTCTACCTATCgCCATGAACTATCGTGGCGATGGAGG	(122)	89.63
p-Int~	GTCGCATTACGCTGCATGGCCCGCTCGATCAACCGACGCTGAAAAGGCTGGTGCATTTGGTGTATGATGTGCGCCgCGATGACGCACCGCTGCGTAAAGTC	(123)	52.55
pabA-P1	GGGGATCTCACCGCACTGCAGATGATATTTAGCAAAGTGGTAAGCGAAGCCGGGGAATCTGAGTAAAATAGCGCGgTTCTTTTGTACCGGAGCCGCCATGA	(124)	75.16
pabA-P2	ATTTCTCGCGCGGGCAATCTTGCCCGCGCTTCTGCTCTCCCGGCGTAACCCGGATTTGCCGCTTATACTTGTGGCaAATGGACACGTTCAGGGAGGCATCA	(124)	75.18
pac	GTTATCGCGCTCACAGTTCATAATGAAACAATTCTCTGCAAATAGATAAACCGAAGCTTCGTTGCTAGTATCAAttcGCTAATTATACACCTGCCAGAAGG	(125)	90.17
pckA	CGCCATATAAACCAAGATTTAACCTTTTGAGAACATTTTCCACACCTAAAATGCTATTTCTGCGATAATAGCAACcGTTTCGTGACAGGAATCACGGAGTT	(126)	76.08
pdx	GTTGTATCATTACGTATCCTTATACCTGAAATCTTCGCAAGTATGCCTGGCCGCGAGATTATGGCACACTTGTCCgGTTAACTCTCGTCTCATACAGGTAA	(48,123)	52.57
pepD-P1	CCAGTATTTCACATGAGTCCGCGTCTTTTTACGCACTGCCTCTCCCTGACGCGGGATAAAGTGGTATTCTCAAACaTaTCTCGCAAGCCTGTCTTGTGTTG	(127)	5.65
pepD-P2	CACTGCCTCTCCCTGACGCGGGATAAAGTGGTATTCTCAAACATATCTCGCAAGCCTGTCTTGTGTTGACAACATtTTCTGCTAACCCTGTGACCTGCAAT	(127)	5.65
pfkB	AATTTTAAATAAAGCTCCAATAAATCATATTGTTAATTTCTTCACTTTCCGCTGATTCGGTGCCAGACTGAAATCaGCCTATAGGAGGAAATGATGGTACG	(128)	38.84
pgsA	GGGTATTTCGCAAGGTCTGGCAGAAAAGATCTTCTGGTCGTTGAAACATTGATGTCTCTGTAGCAACATAGGGGTaATCTTACTGACAACAGATAGTTACC	(129)	42.95
phe	ATTGAGTGTATCGCCAACGCGCCTTCGGGCGCGTTTTTTGTTGACAGCGTGAAAACAGTACGGGTACTGTACTAAaGTCACTTAAGGAAACAAACATGAAA	(130)	58.87
pheST	GCTAGTCCTCCCTCTTTTCGTTTCAACGCCATCAAAACATTGACTTTTATCGCCGTAGCCTTTTCAATAAAGGTCtTTTGAAGAGTACCAAAAGGTAACGC	(131)	38.70
phoB	CGGAAATCAATAACCTGAAGATATGTGCGACGAGCTTTTCATAAATCTGTCATAAATCTGACGCATAATGACGTCgCATTAATGATCGCAACCTATTTATT	(132)	9.07
phoE	TAATAAATTACCACATTTTAAGAATATTATTAATCTGTAATATATCTTTAACAATCTCAGGTTAAAAACTTTCCTgTTTTCAACGGGACTCTCCCGCTGAA	(133)	5.72
pncB	AATGCCATCGACCAGAAAGGTGGCATATGGTGTGATCGGGGTTCAATAAATTGCGAAACAAGGTATACTCCAGCAgTTCCTGAAGATGTTTATTGTACTAA	(134)	21.34
pnp	CAGAAAAGGGGGCCTGAGTGGCCCCTTTTTTCAAGCTGACGGCAGCAATTCACTGGAAACTAATGTATTGTTGCTaTGAATGATCTTCCGTTGCAGAGGTT	(135)	71.28
ppc	ATTCCTGCTATTTATTCGTTTGCTGAAGCGATTTCGCAGCATTTGACGTCACCGCTTTTACGTGGCTTTATAAAAgaCGACGAAAAGCAAAGCCCGAGCAT	(136)	89.52
proU-P1	CCCGCAGCAGGGAAATAATTCCCGCCAAATAGCTTTTTATCACGCAAATAATTTGTGGTGATCTACACTGATACTctgTTGCATTATTCGCCTGAAACCAC	(137)	60.41
proU-P2	TTATCACGCAAATAATTTGTGGTGATCTACACTGATACTCTGTTGCATTATTCGCCTGAAACCACAATATTCAGGcgtTTTTTCGCTATCTTTGACAAAAA	(137)	60.41
proU-P3	TTGCTCGCATCAATATTCATGCCACATTTGCCATCAGGGGTTGCCTCAGATTCTCAGTATGTTAGGGTAGAAAAAaGTGACTAtTTCCATTGGGTAATATA	(137)	60.42
prs	ATGATTGCGAGGTTATCGCAAGAAAACGTTTTCGCGAGGTTGATGCGGTGCTTTCCTGGCTGTTAGAATACGCCCcGTCGCGCCTGACTGGGACAGGGGCC	(138)	27.17
ptr	CCGCTGTTTCGCTTTAATCAGTCATGAGTGCTGTATAAAAATTGCGCAATCTATCCGCTTACTTTATGATGCGCAcCAGTCACGGACTGATGGTTATATAA	(139)	63.72
ptsH	CGTTACAGGAAAAGCCAAAGCTGAATCGATTTTATGATTTGGTTCAATTCTTCCTTTAGCGGCATAATGTTTAATgACGTACGAAACGTCAGCGGTCAACA	(35)	54.60
purE	AAAACCCGCAACTTTGCTGATTTCACAGCCACGCAACCGTTTTCCTTGCTCTCTTTCCGTGCTATTCTCTGTGCccTCTAAAGCCGAGAGTTGTGCACCAC	(140)	11.97
purF-P1	TCAGCGCCATTTTTATTGATGCGCGGGAAGGAAATCCCTACGCAAACGTTTTCTTTTTCTGTTAGAATGCGCCCCgAACAGGATGACAGGGCGTAAAATCG	(141)	52.42
purF-P2	GGAAATCCCTACGCAAACGTTTTCTTTTTCTGTTAGAATGCGCCCCGAACAGGATGACAGGGCGTAAAATCGTGGgACACATATGGTCTGGATTGATTACG	(141)	52.42
purHD	CAGTCAAAATTGTGATCACCATTGAAAGAGAAAAATTCGCGAGCGTTGCGCAAACGTTTTCGTTACAATGCGGGCgAAAAATAAGGATGCCCCGTTAGGGG	(142)	90.70
purMN	CTGGGGAAAAGACGTGCAAAAAGGTTGTGTAAAGCAGTCTCGCAAACGTTTGCTTTCCCTGTTAGAATTGCGCCgaATTTTATTTTTCTACCGCAAGTAAC	(143)	56.27
purR	AATCTCCCGTCATTTATAATGATAAGTGTTTTTACCACTTCCCCTTTTCGTCAAGATCGGCCAAAATTCCACGCTtACACTATTTGCGTACTGGCCATTGA	(144)	37.40
putA	CATTCATTTTTAAGCTTGCTACGCATGTCACATTTAACATGGTTGCACAAAGTTGCAACATCATGGATATTTCACgATAACGTTAAGTTGCACCTTTCTGA	(145)	23.28
putP-P1	GCAAAAAATGTGAGAGAGTGCAACCTGATGAAAAATAGTGTCGCTGAGCACTAAAATTTAATGTAAATGGTGTGTtAAATCGATTGTGAATAACCAGCGCT	(145)	23.28
putP-P2	GAGAGAGTGCAACCTGATGAAAAATAGTGTCGCTGAGCACTAAAATTTAATGTAAATGGTGTGTTAAATCGATTGtGAATAACCAGCGCTTCCGGCAGGAT	(145)	23.28
putP-P3	GTGCAACCTGATGAAAAATAGTGTCGCTGAGCACTAAAATTTAATGTAAATGGTGTGTTAAATCGATTGTGAATAaCCAGCGCTTCCGGCAGGATACGGTC	(145)	23.28
putP-P4	GCTGAGCACTAAAATTTAATGTAAATGGTGTGTTAAATCGATTGTGAATAACCAGCGCTTCCGGCAGGATACGGTcGCCCTGGTAAAACATAAACTCTGTT	(145)	23.29
putP-P5	GTAAAACATAAACTCTGTTACCCCGTTCCGGTGGCAGATATAACGGCAAGTTTCGACATTGCCGATAATAATTTTtTGGAGACTTTAGATGGCTATTAGCA	(145)	23.29
pX~	CCAGACGGTGGCCGTCGGTTGCCACGGTGCGCAGTTCTTCACCTTCGGTTTCAAACAGCATACCATTTAAGTaataGCGAACGTCCTGATGCGCCATAGAA	(146)	83.61
pyrBI-P1	AGCAGTAAGCAATAGTGTTAGCCGTTCGCTTTCACACTCCGCCCTATAAGTCGGATGAATGGAATAAAATGCATAtctgATTGCGTGAAAGTGAAAAAGGA	(147)	96.37
pyrBI-P2	AAATAAACCGTTTGCGCTGACAAAATATTGCATCAAATGCTTGCGCCGCTTCTGACGATGAGTATAATGCCGGacaATTTGCCGGGAGGATGTATGGTTCA	(147)	96.37
pyrC	GATTAATCACGAGGGCGCATTCGCGCCCTTTATTTTTCGTGCAAAGGAAAACGTTTCCGCTTATCCTTTGTGTccgGCAAAAACATCCCTTCAGCCGGAGC	(148)	24.23
pyrD	AAACAGGTTCGGAAAACGTTTGCGTTTTTTTTGCCGCAGGTCAATTCCCTTTTGGTCCGAACTCGCACATAATACgcCCCCGGTTTGCACACCGGGAATCC	(149)	21.66
pyrE-P1	GGAGAGGTGGAAGGATTATAGCCATCGATGCCTTGTAAGGATAGGAATAACCGCCGGAAGTCCGTATAATGCGCAgCCACATTTGTTTCAAGCCGGAGATT	(150)	82.18
pyrE-P2	CCATTCACCCTTGATTTCACGCCGGGCGTAGGCGGTCATACTGCGGATCATAGACGTTCCTGTTTATAAAAGGAGaGGTGGAAGGATTATAGCCATCGATG	(150)	82.18
pyrF	GACGGCGCAGAAAAAGCCTGCCAGGGGAGAAATCGCAACTGTTAATTTTTTATTTCCACGGGTAGAATGCTCGCCgTTTACCTGTTTCGCGCCACTTCCGG	(151)	28.91
pyrG-P1	CATTTGTGTCATTTTTTAAATGACAAGCGCTTGATTTGCGTCAAAAACATTTACCCCAAAGGGGCTATTTTCTCActcctgaTTTCAATAGTGCGCTGGCG	(152)	62.68
pyrG-P2	GATTTCAATAGTGCGCTGGCGAAGAGGAGGGATAATGAAAGTTTGTGGCACAGGTCATGTTCGGGTATACTGCTTtcccgTCTTGGTTATTCCATCGTCTT	(152)	62.68
rbs	AACGTTTCGAGGTTGATCACATTTCCGTAACGTCACGATGGTTTTCCCAACTCAGTCAGGATTAAACTGTGGGTCaGCGAAACGTTTCGCTGATGGAGAAA	(153)	84.77

recA	GTATGCTTGCAGACCTTGTGGCAACAATTTCTACAAAACACTTGATACTGTATGAGCATACAGTATAATTGCTTCaaCAGAACATATTGACTATCCGGTAT	(154)	60.84
recF-P1	AAGCGTCTGATTGAAGCGACCCAGTTTTCTATGGCGCATCAGGACGTTCGCTATTACTTAAATGGTATGCTGTTTgAAACCGAAGGTGAAGAACTGCGCAC	(146)	83.61
recF-P2	CCCTGCCGCAGGCAACGATGAAGCGTCTGATTGAAGCGACCCAGTTTTCTATGGCGCATCAGGACGTTCGCTATTaCTtAAATGGTATGCTGTTTGAAACC	(146)	83.61
relA	CGAGCAGGTGCCGCAGGTGTTATGCAGCAAATTATAAAACTGGAACCTATTCGTATAGTTTATGTATCCTGTAACcCTGCAACGCTGGCTCGGGATAGCGA	(155)	62.77
rep	TGATTTAAACATCCGCAGCCAACCGGTTAGCGGCTTACACGCGGTCACATTCAAATGCGATTCTGCTACAATCCTcCCCCCGTTCGAAGATTGAGCAATAC	(156)	85.36
rho	GTAAGGGAATTTCATGTTCGGGTGCCCCGTCGCTAAAAACTGGACGCCCGGCGTGAGTCATGCTAACTTAGTGTTgACTTCGTATTAAACATACCTTATTA	(157)	85.48
rnh	CCGGTGGTTTCGGTATCGAGAACGATCTGGCGTGTAATTGCAGTGCTCATAGCGGTCATTTATGTCAGACTTGTCgTTTTACAGTTCGATTCAATTACAGG	(52)	5.23
rplJ	GTTGACCAGGCTGGCCTGAGCGCTTCTGTAAACTAATGCCTTTACGTGGGCGGTGATTTTGTCTACAATCTTACCcCCACGTATAATGCTTAATGCAGACG	(158)	90.08
rplK	AAAAGCCTAACCCAGCGATCAAAAAAGCGGCGATTTAATCGTTGCACAAGGCGTGAGATTGGAATACAATTTCGCgCCTTTTGTTTTTATGGGCCTTGCCC	(158)	90.06
rplT	TCTGGGCCTGGTAATCGCGTGCCTGCCGTACGCATAAGCCGTTAACGTTTTTAACTTTTTAATTAGAATATAGATaCAGGAGAGCACATATGGCTCGCGTA	(131)	38.71
rpmH-P1	AAAAATTGGCTTAATCGATCTAATAAAGATCCAGGACGATCCTTGCGCTTTACCCATCAGCCCGTATAATCCTccacccCGGCGCGCCATGCTGGTTTCCAC	(50)	83.65
rpmH-P2	CTGCGAAAAGGTGCGGAAAAGCGCGGTAAATAAGGAAAGAGAATTGACTCCGGAGTGTACAATTATTACAATCCGgcctcTTTAATCACCCATGGCTTCGG	(50)	83.66
rpmH-P3	GGCGGGCACGATTTACGCCGCATATTGGAAAATTTAATGACCATAGACAAAAATTGGCTTAATCGATCTAATAAAgatcCAGGACGATCCTTGCGCTTTAC	(50)	83.65
rpoB	AGCCTTTTGCGCTGTAAGGCGCCAGTAGCGTTTCACACTGTTTGACTACTGCTGTGCCTTTCAATGCTTGTTTCTaTCGACGACTTAATATACTGCGACAG	(159)	90.11
rpoD-Pa	TGCCAAAAGCGGCAGAGAGCGGCGTTTCTCGCCCTGTTCCGCAGCTAAAACGCACGACCATGCGTATACTTATAGgGTTGCTGGTGCAAAATCCAGAATTA	(160)	69.18
rpoD-Pb	AGAGAACTGGTCAACACTTGTCTCTCCCAGCCAGGTCTGACCACCGGGCAACTTTTAGAGCACTATCGTGGTACAaATAATGCTGCCACCCTTGAAAAACT	(160)	69.19
rpoD-Phs	CTTTTAGAGCACTATCGTGGTACAAATAATGCTGCCACCCTTGAAAAACTGTCGATGTGGGACGATATAGCAGATaAGAATATTGCTGAGCAAACCTTCAC	(160)	69.19
rpoD-Phs/min	CACTATCGTGGTACAAATAATGCTGCCACCCTTGAAAAACTGTCGATGTGGGACGATATAGCAGATAAGAATATTgCTGAGCAAACCTTCACCGACTCACT	(160)	69.19
rpoH-P1	CTGCACGGATCAACATTACGCCACTTACGCCTGAATAATAAAAGCGTGTATACTCTTTcctGCAATGGGTTCCGTAGCAGGGAAAG	(161)	77.55
rpoH-P2	CACATTTTGTGCGTAATTTATTCACAAGCTTGCATTGAACTTGTGGATAAAATCACGGTCTGATAAAACAGTGAAtgATAACCTCGTTGCTCTTAAGCTCT	(161)	77.55
rpoH-P2*	CTATGTCACATTTTGTGCGTAATTTATTCACAAGCTTGCATTGAACTTGTGGATAAAATCACGGTCTGATAAAACaGTGAATGATAACCTCGTTGCTCTTA	(161)	77.55
rpsA-P1	GACGCCTCCTCGGAAGAACGTGCGCATCGCCGCATGCTACAGTTGCAGGTGAAGGGCTTTAGTGTTAACTTTGAGcgcCTTTTGGCCGAGATCAAAGAACG	(162)	20.76
rpsA-P2	CCTCGGAAGAACGTGCGCATCGCCGCATGCTACAGTTGCAGGTGAAGGGCTTTAGTGTTAACTTTGAGCGCCTTTtggCCGAGATCAAAGAACGCGACGAC	(162)	20.76
rpsA-P3	CGGCAGCCGATGCTTTAGTGTTGGATTCCACCACCTTAAGCATTGAGCAAGTGATTGAAAAAGCGCTACAATACGcgcGCCAGAAATTGGCTCTCGCATAA	(162)	20.76
rpsJ	TGAGAGATAACCCGAAGGCTGTTTACTTTACTAGCAATACGCTTGCGTTCGGTGGTTAAGTATGTATAATGCGCGgGCTTGTCGTAGTTGACAGCAGGTTC	(163)	74.37
rpsO	GTCGCCTGGTGGTTGAATACCCGGCGTAATGTTAACCGTCTTGCGATAACAGGTCGCTACGAGTAGAATACTgCCgCTTAACGTCGCGTAAATTGTTTAAC	(135)	71.29
rpsU-P1	AGTGTAGGACCAATGCGGGTTGATGTAAAACTTTGTTCGCCCCTGGAGAAAGCCTCGTGTATACTCCTCACCcttataaaAGTCCCTTTCAAAAAAGGCCG	(164)	69.15
rpsU-P2	CCCTTATAAAAGTCCCTTTCAAAAAAGGCCGCGGTGCTTTACAAAGCAGCAGCAATTGCAGTAAAATTccgCACcattTTGAAATAAGCTGGCGTTGATGC	(164)	69.15
ruv-P1	ATACTGTGCCATTTTTCAGTTCATCGAGACACCTCGCAAGTTTTCTTCATCCTTCGCTGGATATCTATCCAGCATtTTTTTATCATACAGCATTATCTTTG	(165)	41.92
ruv-P2	CCATTTTTCAGTTCATCGAGACACCTCGCAAGTTTTCTTCATCCTTCGCTGGATATCTATCCAGCATTTTTTTATcATACAGCATTATCTTTGATTCATTA	(165)	41.92
sbcB	CATGATCTATTGCCACTCGCTGCCAAATTGTGGCGCTAAAGCTGATTAGCACGGTGATATTTGATACTCTGGCaGaCAGCAGAAATAACGGATTTAACCTA	(166)	44.84
sdh-P1	TGTTCAGGATAAAACCCGACAAACTATATGTAGGTTAATTGTAATGATTTTGTGAACAGCCTATACTGCCGCCAGtCTCCGGAACACCCTGCAATCCCGAG	(167)	16.38
sdh-P2	CCTCTCTTCCCCTGAAGGTTTCGAGCAAGCTTCCGCGATTATGGGCAGCTTCTTCGTCAAATTTATCATGTGGGGcATCCTTACCGCTCTGGCGTATCACG	(167)	16.39
secE	GGGCGTTGTAGCAAAAGTTCTGAGCTAATTGCCGATAACATTTGACGCAATGCGCACTAAAGGGCATCATTTGATgCCCTTTTTGCACGCTTTCGTACCAG	(168)	90.03
serB	ATCTTCCACCCTTTGAAAATTTGAGACTTAATGTTGCCAGAAGCAATGGATACAAGGTAGCCTCATGCGTTATTTtCCCTGCTTCGAACGATTTTACAGGA	(169)	99.67
serC	TGCCCCTCAACGGTTTTACTCATTGCGATGTGTGTCACTGAATGATAAAACCGATAGCCACAGGAATAATGTATTaCCTGTGGTCGCAATCGATTGACCGC	(170)	20.67
smp	TCGAAGCAGGGAAAATAACGCATGAGGCTACCTTGTATCCATTGCTTCTGGCAACATTAAGTCTCAAATTTTCAAaGGGTGGAAGATGGCTCGCACAAAAC	(169)	99.67
sodB	TAAAACTTTACCCTGTTGTTACGGCAACAGGGTAAGTTCATCTTTTGTCTCACCTTTTAATTTGCTACCCTATCCaTACGCACAATAAGGCTATTGTACGT	(171)	37.35
spc	AATACGAATAAACGGCTCAGAAATGAGCCGTTTATTTTTTCTACCCATATCCTTGAAGCGGTGTTATAATGCCGCgCCCTCGATATGGGGATTTTTAACGA	(172)	74.25
speB	ACCGATCTTGATGCTGAACTGCAACAACAGTTCCTTGAAGAGTTCGAGGCAGGTTTGTACGGTTATACTTATCTTgAAGATGAGTAAGTCCTGTGTTACTT	(173)	66.40
speC-P1	CTGCAGGACAACAGtAAAATCAGAGCGTTTCTGCTTTTAC	(174)	66.95
speC-P2	CTGCAGGACAACAGTAAAATCAGAGCgTTTCTGCTTTTACTGATGTCTGGCG	(174)	66.95
speC-P3	CTGCAGGACAACAGTAAAATCAGAGCGTTTCTGCTTTTACTgATGTCTGGCGGTCGGAGCTGGTGAC	(174)	66.95
ssb-P1	TTAATTATGACACAAATTGACCTGAATGAATATACAGTATTGGAATGCATTACCCGGAGTGTTGTGTAACAATGTcTGGCCAGGTTTGTTTCCCGGAACCG	(175)	92.11
ssb-P2	GTCTGGCCAGGTTTGTTTCCCGGAACCGAGGTCACAACATAGTAAAAGCGCTATTGGTAATGGTACAATCGCGCGtTTACACTTATTCAGAACGATTTTTT	(175)	92.11
ssb-P3	TTTGTTTCCCGGAACCGAGGTCACAACATAGTAAAAGCGCTATTGGTAATGGTACAATCGCGCGTTTACACTTATtCAGAACGATTTTTTTCAGGAGACAC	(175)	92.11
str	CCAGATGGCCTGGTGATGATGGCGGGATCGTTGTATATTTCTTGACACCTTTTCGGCATCGCCCTAAAATTCGGCgTCCTCATATTGTGTGAGGACGTTTT	(172)	74.81
sucAB	GCGCAGCGCATCAGGCGTAACAAAGAAATGCAGGAAATCTTTAAAAACTGCCCCTGACACTAAGAGAGTTTTTAAaGGTTCCTTCGCGAGCCACTACGTAG	(176)	16.45
sulA	ACGACAAAATTTGCGAGGCTCTTTCCGAAAATAGGGTTGATCTTTGTTGTCACTGGATGTACTGTACATCCATaCaGTAACTCACAGGGGCTGGATTGATT	(177)	22.02
tnaA	CGAACGATTGTGATTCGATTCACATTTAAACAATTTCAGAATAGACAAAAACTCTGAGTGTAATAATGTAGCCTCgTGTCTTGCGAGGATAAGTGCATTAT	(178)	83.75
tonB	GCCAAAAATGACATTTTCACTGATCCTGATCGTCTTGCCTTATTGAATATGATTGCTATTTGCATTTAAAATCGAgACCTGGTTTTTCTACTGAAATGATT	(179)	28.22
treA	TTCTAAAATGCCTGACAGTTCGCAGAATGAGATTTCGATCATGCAGCTAGTGCGATCCTGAACTAAGGTTTTctgaTACTTGAATACCGTTTTATTCGGTT	(180)	26.84
trmA	ACTGTCCAGACGTAGCTCACAAATAGGAATTCATCAAGATCTGGACATCTGATGAGCAATCCCTACAATCGCCGCgTACTTTAATTTTTCAGGATACATCA	(181)	89.73
trp	ACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGAACTAGTTAACTAGTACGCaAGTTCACGTAAAAAGGGTATCGACA	(182)	28.48
trp-P2	CACCAGGAGCAACTGGCAGGCGGAACACCGGAAGAAAACCGTGACATTTTAACACGTTTGTTACAAGGTAAAGGCgACGCCGCCCATGAAGCAGCCGTCGC	(183)	28.41
trpR	CCGACGTTGATGAGCGCCACGGAATGGGGACGTCGTTACTGATCCGCACGTTTATGATATGCTATCGTACTCTTTaGCGAGTACAACCGGGGGAGGCATTT	(182)	99.84
trxA-P1	ACCTTCGGATGCAGGGCGAAGTCGGAAAACTTCTGTTCTGTTAAATGTGTTTTGCTCATAGTGTGGTAGAATATCaGCTTACTATTGCTTTACGAAAGCGT	(184)	85.47
trxA-P2	TTAAATGTGTTTTGCTCATAGTGTGGTAGAATATCAGCTTACTATTGCTTTACGAAAGCGTATCCGGTGAAATAAaGTCAACCTTTAGTTGGTTAATGTTA	(184)	85.47
tsx-P1	ATGCAGTAAACGTATTTCGGGACGATTTTGTGCGTCCCGCAACATCTTTCCCCGTCATTTTGTTACTCTGCTTACaTCACCTGGATTGATAGTAAAAGTTT	(185)	9.40
tsx-P2	CGCAATCGATTACGTAAATGATAGAACGTGGAAACGAAACATATTTTTGTGAGCAATGATTTTTATAATAGGCTCaTCTGTATACGAAATATTTAGAAACG	(185)	9.39
tufB	TAAAATGTGACCAATAAAACAAATTATGCAATTTTTTAGTTGCATGAACTCGCATGTCTCCATAGAATGCGCGCTaCTTGATGCCGACTTAGCTCAGTAGG	(186)	89.99
tyrB	TCACTCTGTTGCTAATTGCCGTTCGCTCCTGAACATCCACTCGATCTTCGCCTTCTTCCGGTTTATTGTGTTTTAaCCACCTGCCCGTAAACCTGGAGAAC	(187)	91.97
tyrP-P1	TTATTGTACATTTATATTTACACCATATGTAACGTCGGTTTGACGAAGCAGCCGTTATGCCTTAACCTGCGCCGCaGATATCACTCATAAAGATCGTCAGG	(188)	42.88
tyrP-P2	GCGTATATTGCTCGCCTGATTCAATCATCATATTTAACACTTCATGACCGTGAATAGAGTCCATCGTCCCTCCTCaAAAAAAGCCTAGCGTAGCGATTGCC	(188)	42.88
tyrR	GTTTTTCCGTCTTTGTGTCAATGATTGTTGACAGAAACCTTCCTGCTATCCAAATAGTGTCATATCATCATATTAaTTGTTCTTTTTTCAGGTGAAGGTTC	(189)	29.20
udf-Px~	TGAGGAGTATACACGAGGCTTTCTCCAGGGGCGAACAAAGTTTTACATCAACCCGCATTGGTCCTACACTGCGCGgtaatAAAGCGAGGTAAAACAAGTCA	(164)	69.14
umu	AGCAGCCTATGCAGCGACAAATATTGATAGCCTGAATCAGTATTGATCTGCTGGCAAGAACAGACTACTGTATATaAAAACAGTATAACTTCAGGCAGATT	(190)	26.48
ung	GCACTCGTTTAAGTCTAAAAAATGAGCATGATTTTGTTCTGTAGAAAGAAGCAGTTAAGCTAGGCGGATTGAAGAtTCGCAGGAGAGCGAGATGGCTAACG	(191)	58.37
uvrA	ATTGTTACACAACACTCCGGGTAATGCATTCCAATACTGTATATTCATTCAGGTCAATTTGTGTCATAATTAACCgTTTGTGATCGCCGGTAGCACCATGC	(192)	92.11
uvrB-P1	ATATTATGGTGATGAACTGTTTTTTTATCCAGTATAATTTGTTGGGATAATTAAGTACGACGAGTAAAATTACATaCCTGCCCGCCCAACTCCTTCAGGTA	(193)	17.61
uvrB-P2	GATTGACAGCGGAGTTTACGCTGTATCAGAAATATTATGGTGATGAACTGTTTTTTTATCCAGTATAATTTGTGgGATAATTAAGTACGACGAGTAAAAT	(193)	17.61
uvrB-P3	CGGCGGGGGATAGGGGCTGGACACAGTTATCCACTATTCCTGTGGATAACCATGTGTATTAGAGTTAGAAAaCACGAGGCAAGCGAGAGAATACGCG	(193)	17.60
uvrC-P2	ACAAAAACGAGAGAAAAATCGAATACCCACCATTTTTAACGTTTCAAAGTTGCAATAAAAACCGCTAATATACGaatGACTAACTATCAGTAGCGTTATCC	(194)	43.00
uvrC-P3	CAGATCGAACCAGAAAAAACAGAAAGCCCATTTGCCAGTTTGTCTGAACGTGAATTGCAGATTATGCTGATGaTCaCCAAGGGCCAGAAGGTCAATGAGAT	(195)	42.99
uvrD-P1	CCTGGGACAGCCGTTTACTGCCGCATCTGGAAATTTCCCGGTTGGCATCTCTGACCTCGCTGATATAATCAGCAAaTCTGTATATATACCCAGCTTTTTGG	(196)	86.17
uvrD-P2	ATAATCAGCAAATCTGTATATATACCCAGCTTTTTGGCGGAGGGCGTTGCGCTTCTCCGCCCAACCTATTTTTACgCGGCGGTGCCAATGGACGTTTCTTA	(197)	86.17
uxuAB	AGTGAAATTGTTGTGATGTGGTTAACCCAATTAGAATTCGGGATTGACATGTCTTACCAAAAGGTAGAACTTATaogCCATCTCATCCGATGCAAGCCACG	(198)	98.20
valS-P1	CGATATATTGATTAGTCTGCGAACAAGCTTTGCAGATTTTGCCACCGCTTTCACAGAAGTGGTAGACTTCGTTCCtTATGAAGATTCGTCCTGAAACAACT	(199)	96.60
valS-P2	TGATTAGTCTGCGAACAAGCTTTGCAGATTTTGCCACCGCTTTCACAGAAGTGGTAGACTTCGTTCCTTATGAAGaTTCGTCCTGAAACAACTGGCGCGCG	(199)	96.60
zwf	GCTTTTCCCGTAATCGCACGGGTGGATAAGCGTTTACAGTTTTCGCAAGCTCGTAAAAGCAGTACAGTGCACCGTaAGAAAATTACAAGTATACCCTGGCT	(200)	41.67

Table I. Base distribution in the consensus hexamers

		-35						-10					
		T	T	G	A	C	A	T	A	T	A	A	T
all promoters (298)	A	10	6	9	56	21	54	5	76	15	61	56	6
	C	10	7	12	17	54	13	10	6	11	13	20	7
	G	10	8	61	11	9	16	8	6	14	14	8	5
	T	69	79	18	16	16	17	77	12	60	12	15	82
space = 16 (52)	A	19	4	6	52	19	60	6	75	13	69	62	6
	C	8	6	13	17	56	6	6	12	10	12	17	4
	G	4	12	60	6	13	15	8	0	15	12	10	6
	T	69	79	21	25	12	19	81	13	62	8	12	85
space = 17 (129)	A	9	9	9	56	23	50	6	82	19	53	55	4
	C	9	5	14	20	50	13	15	5	9	18	18	9
	G	8	9	57	11	10	18	8	4	16	17	9	3
	T	74	77	19	13	16	19	71	9	57	12	18	84
space = 18 (51)	A	4	6	10	59	20	57	0	71	14	61	61	12
	C	22	4	6	12	51	14	8	4	16	12	20	2
	G	8	6	73	18	6	16	6	8	18	14	8	2
	T	67	84	12	12	24	14	86	18	53	14	12	84

The frequencies (in %) of the four nucleotides A,C,G,T in each of the consensus positions is tabulated. For this calculation the promoter sequences, selected as described in the text, were aligned. Frequencies are reported for the whole database (298 sequences, as 2 sequences missed the region upstream from the transcriptional start position), and for three groups distinguished by the length of the spacer between the −35 and −10 regions. In parentheses are the numbers of sequences in each group.

Table II. The degree of consensus conservation in the current and previous compilations

	-35						-10					
	T	T	G	A	C	A	T	A	T	A	A	T
Hawley and McClure	82	84	79	64	54	45	79	95	44	59	51	96
Harley and Reynolds	78	82	68	58	52	54	82	89	52	59	49	89
Current Compilation	69	79	61	56	54	54	77	76	60	61	56	82

For each compilation the degree of conservation (in%) of the consensus nucleotides is tabulated. Note that the three compilations used different procedures for promoters alignment. Degrees of conservation in Hawley's and McClure's compilation (202) are based on 112 promoters and taken from O'Neil's calculations (208). Degrees of conservation in Harley and Reynolds' compilation are taken from that report and based on 263 promoters (203). These two compilations include all promoters of *E.coli* and its phages known at the time. Degrees of conservation in the current compilation are based on 298 mRNA *E.coli* promoters (updated to October 1992).

Table III. Distribution of promoters by −35 to −10 spacer length

-35 to -10 spacer length	frequency
15	35
16	52
17	129
18	51
19	31

Frequency is expressed in absolute numbers (out of a total of 298 promoters).

Table IV. Distribution of promoters by number of matches to the 12 consensus nucleotides

No. of matches	Frequency
5	2
6	27
7	93
8	86
9	71
10	15
11	4
Average no. of matches = 7.87	

Frequency is expressed in absolute numbers (out of a total of 298 promoters).

Duplicated sequences

The routine sequence analysis computer programs enable automatic extraction from the databanks, by searching for a specific query, such as the annotation for mRNA start site. Analyses based on such datasets may be biased, as the databanks include many duplications. The same gene and the same promoter may appear more than once under different code names and accession numbers, if reported to the databank by more than one research group. Therefore, at the present state of the databanks, generation of a sub-database out of a sequence databank must be accompanied by a search for and exclusion of duplications.

Gene features

The information in the databanks includes the locations of various gene features. In several cases, we found inaccuracies in the locations reported in the bank, in comparison to the experimental paper. In some cases the indicated locations were based on sequence data alone and not on experimental evidence. Another drawback regards updates of the gene features based on more recent procedures. For example, we ran into transcriptional start positions that were identified initially by S1 nuclease mapping and more recently by primer extension, resulting in different allocations, that are not reported in the databank. Therefore, to have the databank reliable and up to date, it may be necessary in the future to report not only newly discovered sequences and their features, but also newly identified gene features.

SURVEY OF PROMOTER SEQUENCES

Two consensus hexamers have been identified in *E.coli* promoters, TTGACA located about 35 base pairs upstream from the transcriptional start position (the −35 region), and TATAAT located about 10 base pairs upstream from the transcriptional start position (the −10 region), with an average separation of 17 nucleotides between them (201−203). The sequences in the current compilation are aligned by the experimentally identified transcriptional start positions. The promoter sequences upstream to these positions are not marked to allow unbiased use of this data in future analyses. Nevertheless, we performed several statistical analyses on the data to examine the conservation of the consensus hexamers and the spaces between them.

In the first extensive promoter compilation (202), Hawley and McClure searched for sequences that matched best the two consensus hexamers determined previously, TTGACA and TATAAT with a separation of 15−21 nucleotides between them. The best matches, with the space between the two hexamers closest to 17 were preferred. Several of the consensus nucleotides were identified as highly conserved, relative to others. The frequencies of the consensus bases in that compilation were used in subsequent analyses (203, 206) as weighting factors for different positions within the consensus.

For an unbiased allocation of the −35 and −10 regions in the sequences of the current database, we used a procedure that is similar to the one used by Hawley and McClure in their compilation (202). By this procedure we searched for the best match to the 12 consensus nucleotides, treating each position equally. We selected first regions with the highest degrees of similarity to the 12 nucleotide consensus, with a separation of 15−19 base pairs between the two hexamers. Different allocations that had equally high matches to the consensus were ranked according to several hierarchical criteria:

1) Homogeneous conservation in the two regions was preferred to a high conservation in one region and a low conservation in the second region.
2) Sequences with a space of 17 were preferred and spaces of 16 and 18 were chosen secondly.
3) Spaces of 6−8 nucleotides between the transcriptional start position and the −10 hexamer were preferred.

Table I demonstrates the distribution of bases in consensus hexamers based on this procedure. The conservation of consensus bases is very clear, though the degrees of conservation per position are somewhat different than in previous compilations, as demonstrated in Table II. For example, the conservation of the two first T's in the TTGACA hexamer is less strong in comparison to Hawley and McClure's (202) and Harley and Reynolds' (203) compilations, while conservation of other bases, such as the third T in the TATAAT hexamer is higher (44% in Hawley and McCLure's compilation, 52% in Harley and Reynolds' compilation and 60% here). Differences in the degrees of consensus base conservation in the different positions of the −35 and −10 hexamers were also evident in comparisons between the other compilations. Although the degree of conservation per position depends on the method of alignment, a general trend may be seen. In the compilation by Harley and Reynolds there is a decrease in the degree of conservation in 8 out of 12 positions, in comparison to the previous compilation, although their alignment used different weights for the different positions based on the nucleotide frequencies. It seems that with the increase in the database the degree of conservation is more uniformly distributed among the consensus positions. Positions that were identified before as highly conserved show lower degrees of conservation, and other positions, especially in the −10 region, show higher degrees of conservation.

There is a high frequency of sequences with a space of 17 nucleotides among the sequences with best matches to the 12 nucleotide consensus, as demonstrated in Table III, with a symmetric distribution around it. When the sequences are divided according to the spacer length and the degree of conservation is compared between the different compilations, the same trend is seen. The positions that were found to be very strongly conserved in the two previous compilations show a lower degree of conservation in the current compilation in all space groups.

Because of the increase in conservation in some positions and decrease in others, the average conservation does not change much in comparison to the other compilations. O'Neil (208) calculated the average conservation in Hawley and McClure's compilation to be 8.2 conserved nucleotides per promoter, and we find in the updated database of mRNA promoter sequences an average conservation of 7.9 nucleotides. However, as is demonstrated in Table IV, there are promoters that match the consensus in only about half the nucleotides and they serve successfully as the sites for transcription initiation. Conceivably, the RNA polymerase recognizes a specific configuration of the promoter that is dictated by the sequence but does not require complete conservation of the nucleotides. The compiled database may be used to reveal such sequence dependent characteristics in *E.coli* promoters.

ACKNOWLEDGEMENTS

We are thankful to Nicole Fleischer, Tamar Schure, Ora Schuller, and Itamar Simon for their help in checking the relevant literature

and sequences, and Yael Altuvia for her help in sequence analysis. This study was supported by the Israeli Cancer Research Fund.

REFERENCES

1. Chung, T., Klumpp, D.J. and Laporte, D. (1988) J. Bacteriol. **170**, 386–392.
2. Spencer, M.E. and Guest, J.R. (1985) Mol. Gen. Genet. 200, 145-154.
3. Nakabeppu, Y., Kondo, H., Kawabata, S., Iwanaga, S. and Sekiguchi, M. (1985) J. Biol. Chem. **260**, 7281–7288.
4. Putney, S.D., Melendez, D.L. and Schimmel, P.R. (1980) J. Biol. Chem. **256**, 205–211.
5. Nakabeppu, Y. and Sekiguchi, M. (1986) Proc. Natl. Acad. Sci. USA **83**, 6297–6301.
6. Leung, H.B., Kvalnes-Krick, K.L., Meyer, S.L., deRiel, J.K. and Schramm, V.L. (1989) Biochem. **28**, 8726–8733.
7. Jaurin, B., Grundstrom, T., Edlund, T. and Normark, S. (1981) Nature. **290**, 221–225.
8. Jennings, M.P. and Beacham, I.R. (1990) J. Bacteriol. **172**, 1491-1498.
9. Smith, B.R. and Schleif, R. (1978) J. Biol. Chem. **253**, 6931–6933.
10. Wallace, R.G., Lee, N. and Fowler, A.V. (1980) Gene **12**, 179-190.
11. Stoner, C. and Schleif, R. (1983) J. Mol. Biol. **171**, 369–381.
12. Hendrickson, W., Stoner, C. and Schleif, R. (1990) **215**, 497–510.
13. Cunin, R., Eckhard, T., Piette, J., Boyen, A., Pierard, A. and Glansdorff, N. (1983) Nucleic Acids Res. **11**, 5007–5019.
14. Piette, J., Cunin, R., van vliet, F.V., Charlier, D., Crabeel, M., Ota, Y. and Glansdorff, N. (1982) EMBO J. **1**, 853–857.
15. Lim, D., Oppenheim, J.D., Eckhardt, T. and Maas, W.K. (1987) Proc. Natl. Acad. Sci. USA **84**, 6697–6701.
16. Hudson, G.S. and Davidson, B.E. (1984) J. Mol. Biol. **180**, 1023-1051.
17. Baseggio, N., Davies, W.D. and Davidson, B.E. (1990) J. Bacteriol. **172**, 2547–2557.
18. Zurawski, G., Gunsalus, R.P., Brown, K.D. and Yanofsky, C. (1981) J. Mol. Biol. **145**, 47–73.
19. Chye, M. and Pittard, J. (1987) J. Bacteriol. **169**, 386–393.
20. Scofield, M.A., Lewis, W.S. and Schuster, S.M. (1990) J. Biol. Chem. **265**, 12895–12902.
21. Fotheringham, I.G., Dacey, S.A., Taylor, P.P., Smith, T.J., Hunter, M.G., Finlay, M.E., Primrose, S.B., Parker,D.M. and Edwards, R.M. (1986) Biochem. J. **234**, 593–604.
22. Nielsen, J., Jorgensen, B.B., von Meyenburg, K. and Hansen, F.G. (1984) Mol. Gen. Genet. **193**, 64–71.
23. Otsuka, A. and Abelson, J. (1978) Nature **276**, 689–694.
24. Aufrere, R., Tempete, M. and Bohin, J.P. (1986) Mol. Gen. Genet. **205**, 358–365.
25. Cowing, D.W., Bardwell, J.C.A., Craig, E.A., Woolford, C., Hendrix, R.W. and Gross, C.A. (1985) Proc. Natl. Acad. Sci. USA **82**, 2679-2683.
26. Bouvier, J. Patte, J-C. and Stragier, P. (1984) Proc. Natl. Acad. Sci. USA **81**, 4139–4143.
27. 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. USA **81**, 4134–4138.
28. Nyunoya, H. and Lusty, C.J. (1983) Proc. Natl. Acad. Sci. USA **80**, 4629–4633.
29. Icho, T., Bulawa, C.E. and Raetz, C.R.H. (1985) J. Biol. Chem. **260**, 12092–12098.
30. Griggs, D.W., Tharp, B.B. and Konisky, J. (1987) J. Bacteriol. **169**, 5343–5352.
31. Frantz, B. and Chakrabarty, A.M. (1987) Proc. Natl. Acad. Sci. USA **84**, 4460–4464.
32. Kitagawa, M., Wada, C,. Yoshioka, S. and Yura, T. (1991) J. Bacteriol. **173**, 4247–4253.
33. Maurizi, M.R., Clark, W.P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B. and Gottesman, S. (1990) J. Biol. Chem. **265**, 12536–12545.
34. Aiba, H., Fujimoto, S. and Ozaki, N. (1982) Nucleic Acids Res. **10**, 1345–1361.
35. De Reuse, H. and Danchin, A. (1988) J. Bacteriol. **170**, 3827-3837.
36. Goldstein, J., Pollitt, N.S. and Inouye, M. (1990) Proc. Natl. Acad. Sci. USA **87**, 283–287.
37. Aiba, H., Kawamukai, M. and Ishihama, A. (1983) Nucleic Acids Res. **11**, 3451–3465.
38. Minagawa, J., Nakamura, H., Yamato, I., Mogi, T. and Anraku, Y. (1990) J. Biol. Chem. **265**, 11198–11203.
39. Malo, M.S. and Loughlin, R.E. (1990) Gene **87**, 127–131.
40. Denk, D. and Bock, A. (1987) J. Gener. Microbiol. **133**, 515–525.
41. Ostrowski, J. and Kredich, N.M. (1989) J. Bacteriol. **171**, 130-140.
42. Byrne, C.R., Monroe, R.S., Ward, K.A. and Kredich, N.M. (1988) J. Bacteriol. **170**, 3150–3157.
43. Richaud, F., Richaud, C., Ratet, P. and Patte, J-C. (1986) J. Bacteriol. **166**, 297–300.
44. Bouvier, J., Richaud, C., Richaud, F., Patte, J-C. and Stragier, P. (1984) J. Biol. Chem. **259**, 14829–14834.
45. Richaud, C., Richaud, F., Martin, C., Haziza, C. and Patte, J-C. (1984) J. Biol. Chem. **259**, 14824–14828.
46. Valentin-Hansen, P., Aiba, H. and Schumperli, D. (1982) EMBO J. **1**, 317–322.
47. Valentin-Hansen, P., Hammer, K., Larsen, J.E.L. and Svendsen, I. (1984) Nucleic Acids Res. **12**, 5211–5224
48. Schoenlein, P.V., Roa, B.B. and Winkler, M.E. (1989) J. Bacteriol. **171**, 6084–6092.
49. Ohmori, H., Kimura, M., Nagata, T. and Sakakibara, Y. (1984) Gene **28**, 159–170.
50. Hansen, F.G. Hansen, E.B. and Atlung, T. (1982) EMBO J. **9**, 1043-1048.
51. Armengod, M.E., Garcia-Sogo, M. and Lambies, E. (1988) J. Biol. Chem. **263**, 12109–12114.
52. Nomura, T., Aiba, H. and Ishihama, A. (1985) J. Biol. Chem. **260**, 7122–7125.
53. Masai, H. and Arai, K. (1988) J. Biol. Chem. **263**, 15083–15093.
54. Zhou, Z. and Syvanen, M. (1990) J. Bacteriol. **172**, 281–286.
55. Bornstein-Forst, S.M., McFall, E. and Palchaudhuri, S. (1987) J. Bacteriol. **169**, 1056–1060.
56. Kauppinen, S., Siggaard-Andersen, M. and von Wettstein-Knowles, P. (1988) Carlsberg Res. Commun. **53**, 357–370.
57. Black, P.N. (1991) J. Bacteriol. **173**, 435–442.
58. Birkman, A., Zinoni, F., Sawers, G. and Bock, A. (1987) Arch. Microbiol. **148**, 44–51.
59. Pettist, G.S., Brickman, T.J. and McIntosh, M.A. (1988) J. Biol. Chem. **263**, 18857–18863.
60. Kuo, S. and Koshland Jr., D.E. (1986) J. Bacteriol. **166**, 1007–1012.
61. Jones, H.M. and Gunsalus, R.P. (1985) J. Bacteriol. **164**, 1100- 1109.
62. Miles, J.S. and Guest, J.R. (1984) Gene **32**, 41–48.
63. de Lorenzo, V., Herrero, M., Giovannini, F. and Neilands, J.B. (1988) J. Biochem. **173**, 537–546.
64. Tanaka, Y., Tsujimura, A., Fujita, N., Isono, S. and Isono, K. (1989) J. Bacteriol. **171**, 5707–5712.
65. Kuhnke, G., Krause, A., Heibach, C., Gieske, U., Fritz, H.-J. and Ehring, R. (1986) EMBO J. **5**, 167–173.
66. Plamann, M.D. and Stauffer, G.V. (1983) Gene **22**, 9–18.
67. Leon, P., Romero, D., Garciarrubio, A., Bastarrachea, F. and Covarrubias, A.A. (1985) J. Bacteriol. **164**, 1032–1038.
68. Nohno, T. and Saito, T. (1987) Nucleic Acids Res. **15**, 2777.
69. Ueno-Nishio, S., Mango, S., Reitzer, L.J. and Magasanik, B. (1984) J. Bacteriol. **160**, 379–384.
70. Cheung, A.Y. and Soll, D. (1984) J. Biol. Chem. **259**, 9953–9958.
71. Ye, S. and Larson, T.J. (1988) J. Bacteriol. **170**, 4209–4215.
72. Wilde, R.J. and Guest, J.R. (1986) J. Gen. Microbiol. **132**, 3239-3251.
73. Oliver, G., Gosset, G., Sanchez-Pescador, R., Lozoya, E., Ku, L.M., Flores, N., Becerril, B., Valle, F. and Bolivar, F. (1987) Gene **60**, 1-11.
74. Nasoff, M.S., Baker, H.V. and Wolf Jr., R.E. (1984) Gene **27**, 253-264.
75. Lipinska, B., King, J., Ang, D. and Georgopoulos, C. (1988) Nucleic Acids Res. **15**, 7545–7562.
76. Tiedeman, A.A. and Smith, J.M. (1985) Nucleic Acids Res. **13**, 1303–1316.
77. Yamada, M. and Saier, M.H. (1987) J. Biol. Chem. **262**, 5455- 5463.
78. Menzel, R. and Gellert, M. (1987) J. Bacteriol. **169**, 1272–1278.
79. Adachi, T., Mizuuchi, K., Menzel, R. and Gellert, M. (1984) Nucleic Acids Res. **12**, 6389–6395.
80. Verkamp, E. and Chelm, B.K. (1989) J. Bacteriol. **171**, 4728–4735.
81. Frunzio, R., Bruni, C.B. and Blasi, F. (1981) Proc. Natl. Acad. Sci.USA **78**, 2767–2771.
82. Grisolia, V., Riccio, A. and Bruni, C.B. (1983) J. Bacteriol. **155**, 1288–1296.
83. Verde, P., Frunzio, R., di Nocera, P.P., Blasi, F. and Bruni, C.B. (1981) Nucleic Acids Res. **9**, 2075–2086.
84. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J. and Schimmel, P. (1985) J. Biol. Chem. **260**, 10063-10068.
85. Price C., Lingner, J., Bickle, T.A., Firman, K. and Glover, S.W. (1989) J. Mol. Biol. **205**, 115–125.

86. Lipinska, B., Sharma, S. and Georgopoulos, C. (1988) Nucleic Acids Res. **16**, 10053–10067.
87. Weiss, D.L., Johnson, D.I., Weith, H.L. and Somerville, R.L. (1986) J. Biol. Chem. **261**, 9966–9971.
88. Friden, P., Newman, T. and Freundlich, M. (1982) Proc. Natl. Acad. Sci. USA **79**, 6156–6160.
89. Wek, R.C. and Hatfield, G.W. (1986) J. Biol. Chem. **261**, 2441-2450.
90. Adams, C.W., Rosenberg, M. and Hatfield, G.W. (1985) J. Biol. Chem. **260**, 8538–8544.
91. Lawther, R.P., Wek, R.C., Lopes, J.M., Pereira, R., Taillion, B.E. and Hatfield, G.W. (1987) Nucleic Acids Res. **15**, 2137–1155.
92. Wek, R.C. and Hatfield, G.W. (1986) Nucleic Acids Res. **14**, 2763-2777.
93. Haughn, G.W., Squires, C.H., DeFelice, M., Largo, C.T. and Calvo, J.M. (1985) J. Bacteriol. **163**, 186–198.
94. Mayaux, J.-F., Fayat, G., Fromant, M., Springer, M., Grunberg- Manago, M. and Blanquet, S. (1983) Proc. Natl. Acad. Sci. USA **80**, 6152–6156.
95. Von Ossowski, I., Mulvey, M.R., Leco, P.A., Borys, A. and Loewen, P.C. (1991) J. Bacteriol. **173**, 514–520.
96. Aldea, M., Claverie-Martin, F., Diaz-Torres, M.R. and Kushner, S.R. (1988) Gene **65**, 101–110.
97. Steege, D.A. (1977) Proc. Natl. Acad. Sci. USA **74**, 4163–4167.
98. March, P.E. and Inouye, M. (1985) J. Biol. Chem. **260**, 7206- 7213.
99. Wessler, S.R. and Calvo, J.M. (1981) J. Mol. Biol. **149**, 579–597.
100. Miki, T., Ebina, Y., Kishi, F. and Nakazawa, A. (1981) Nucleic Acids Res. **9**, 529–543.
101. Adams, M.D., Wagner, L.M., Graddis, T.J., Landick, R., Antonucci, T.K., Gibson, A.L. and Oxender, D.L. (1990) J. Biol. Chem. **265**, 11436–11443.
102. Nakamura, K., Pirtle, R.M., Pirtle, I.L., Takeishi, K. and Inouye, M. 1980) J. Biol. Chem **255**, 210–216.
103. Bedouelle, H., Schmeissner, U., Hofnung, M. and Rosenberg, M. (1982) J. Mol. Biol. **161**, 519–531.
104. Reidl, J., Romisch, K., Ehrmann, M. and Boos, W. (1989) J. Bacteriol. **171**, 4888–4899.
105. Raibaud, O., Debarbouille, M., Schwartz, M. (1983) J. Mol. Biol. **163**, 395–408.
106. Chapon, C. and Kolb, A. (1983) J. Bacteriol. **156**, 1135–1143.
107. Saris, P.E.J., Liljestrom, P. and Palva, E.T. FEMS Microbiol. Letters **49**, 69–73.
108. Webster, C., Kempsell, K., Booth, I. and Busby, S. (1987) Gene **59**, 253–263.
109. Michaeli, S., Mevarech, M. and Ron, E.Z. (1984) J. Bacteriol. **160**, 1158–1162.
110. Kirby, T.W., Hindenach, B.R. and Greene, R.C. (1986) J. Bacteriol. **165**, 671–677.
111. Lee, C.A. and Saier Jr., M.H. (1983) J. Biol. Chem. **258**, 10761-10767.
112. Rogers, M.J., Ohgi, T., Plumbridge, J. and Soll, D. (1988) Gene **62**, 197–207.
113. Plumbridge, J.A. (1989) Mol. Microbiol. 3, 505–515.
114. Li, S.F. and DeMoss, J.A. (1987) J. Bacteriol. **169**, 4614–4620.
115. Tuggle, C.K. and Fuchs, J.A. (1986) EMBO J. **5**, 1077–1085.
116. Munch-Peterson, A. and Jensen, N. (1990) J. Biochem. **190**, 547-551.
117. Ishii, S., Kuroki, K. and Imamoto, F. (1984) Proc. Natl. Acad. Sci. USA **81**, 409–413.
118. Cole, S.T., Bremer, E., Hindennach, I. and Henning, U. (1982) Mol. Gen. Genet. **188**, 472–479.
119. Wurtzel, E.T., Chou, M.-Y. and Inouye, M. (1982) J. Biol. Chem. **257**, 13685–13691.
120. Mizuno, T., Chou, M-Y. and Inouye, M. (1984) Proc. Natl. Acad. Sci. USA **81**, 1966–1970.
121. Inokuchi, K., Furukawa, H., Nakamura, K. and Mizushima, S.(1984) J. Mol. Biol. **178**, 653–668.
122. Christman, M.F., Storz, G. and Ames, B.N. (1989) Proc. Natl. Acad. Sci. USA **86**, 3484–3488.
123. Schoenlein, P.V., Roa, B.B. and Winkler, M.E. (1989) J. Bacteriol. **171**, 6084–6092.
124. Tran, P.V., Bannor, T.A., Doktor, S.Z. and Nichols, B.P. (1990) J. Bacteriol. **172**, 397–410.
125. Valle, F., Gosset, G., Tenorio, B., Oliver, G. and Bolivar, F. (1986) Gene **50**, 119–122.
126. Medina, V., Pontarollo, R., Glaeske, D., Tabel, H. and Goldie, H. (1990) J. Bacteriol. **172**, 7151–7156.
127. Henrich, B., Monnerjahn, U. and Plapp, R. (1990) J. Bacteriol. **172**, 4641–4651.
128. Daldal, F. (1983) J. Mol. Biol. **168**, 285–305.
129. Gopalanishnan, A.S., Chen, Y.-C., Temkin, M. and Dowhan, W. (1986) J. Biol. Chem. **261**, 1329–1338.
130. Hudson, G.S. and Davidson, B.E. (1984) J. Mol. Biol. **180**, 1023-1051.
131. Fayat, G., Mayaux, J.-F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M. and Blanquet, S. (1983) J. Mol. Biol. **171**, 239–261.
132. Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) J. Mol. Biol. **190**, 37–44.
133. Tommassen, J., Koster, M. and Overduin, P. (1987) J. Mol. Biol. **198**, 633–641.
134. Wubbolts, M.G., Terpstra, P., Van Beilen, J.B., Kingma, J., Meesters, H.A.R. and Witholt, B. (1990) J. Biol. Chem. **265**, 17665–17672.
135. Regnier, P. and Portier, C. (1986) J. Mol. Biol. **187**, 23–32.
136. Izui, K., Miwa, T., Kajitani, M., Fujita, N., Sabe, H., Isihama, A. and Katsuki, H. (1985) Nucleic Acids Res. **13**, 59–71.
137. Gowrishankar, J. (1989) J. Bacteriol. **171**, 1923–1931.
138. Hove-Jensen, B., Harlow, K.W., King, C.J. and Switzer, R.L. (1986) J. Biol. Chem. **261**, 6765–6771.
139. Claverie-Martin, F., Diaz-Torres, M.R. and Kushner, S.R. (1987) Gene, **54**, 185–195.
140. Watanabe, W., Sampei, G.-I., Aiba, A. and Mizobuchi, K. (1989) J. Bacteriol. **171**, 198–204.
141. Makaroff, C.A. and Zalkin, H. (1985) J. Biol. Chem. **260**, 10378-10387.
142. Aiba, A. and Mizobuchi, K. (1989) J. Biol. Chem. **264**, 21239-21246.
143. Smith, J.M. and Daum, H.A. (1986) J. Biol. Chem. **261**, 10632-10636.
144. Rolfes, R.J. and Zalkin, H. (1988) J. Biol. Chem. **263**, 19653-19661.
145. Nakao, T., Yamato, . and Anraku, Y. (1987) Mol. Gen. Genet. **210**, 364–368.
146. Armengod, M.E. and Lambies, E. (1986) Gene **43**, 183–196.
147. Donahue, J.P. and Turnbough, C.L. (1990) J. Biol. Chem. **265**, 19091–19099.
148. Wilson, H.R., Chan, P.T. and Turnbough, C.L. (1987) J. Bacteriol. **169**, 3051–3058.
149. Larsen, J.N. and Jensen, K.F. (1985) Eur. J. Biochem. **151**, 39–65.
150. Poulsen, P., Bonekamp, F. and Jensen, K.F. (1984) EMBO J. **3**, 1783–1790.
151. Turnbough, C.L., Kerr, K.H., Funderburg, W.R., Donahue, J.P. and Powell, F.E. (1987) J. Biol. Chem. **262**, 10239–10245.
152. Weng, M., Makaroff, C.A. and Zalkin, H. (1986) J. Biol. Chem. **261**, 5568–5574.
153. Bell, A.w., Buckel, S.D., Groarke, J.M., Hope, J.N., Kingsley, D.H. and Hermodson, M.A. (1986) J. Biol. Chem. **261**, 7652–7658.
154. Horii, T., Ogawa, T. and Ogawa, H. (1980) Proc. Natl. Acad. Sci. USA **77**, 313–317.
155. Metzger, S., Ben Dror, I., Aizenman, E., Schreiber, G., Toone, M., Friesen, J.D., Cashel, M. and Glaser, G. (1988) J. Biol. Chem. **263**, 15699–15704.
156. Bialkowska-Hobrzanska, H., Gilchrist, C.A. and Denhardt, D.T. (1985) J. Bacteriol. **164**, 1004–1010.
157. Matsumoto, Y., Shigesda, K., Hirano, M. and Imai, M. (1986) J. Bacteriol. **166**, 945–958.
158. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Natl. Acad. Sci. USA **76**, 1697–1701.
159. Fukuda, R. and Nagasawa-Fujimori, H. J. Biol. Chem. **258**, 2720-2728.
160. Taylor, W.E., Straus, D.B., Grossman, A.D., Burton, Z.F., Gross, C.A.and Burgess, R.R. (1984) Cell **38**, 371–381.
161. Fujita, N., and Ishihama, A. (1987) Mol. Gen. Genet. **210**, 10–15.
162. Pedersen, S., Skouv, J., Kajitani, M. and Ishihama, A. (1984) Mol. Gen. Genet. **196**, 135–140.
163. Olins, P.O. and Nomura, M. (1981) Cell **26**, 205–211.
164. Burton, Z.F., Gross., C.A., Watanabe, K.K. and Burgess, R.R. (1983) Cell **32**, 335–349.
165. Shinagawa, H., Makino, K., Amemura, M., Kimura, S., Iwasaki, H. and Nakata, A. (1988) J. Bacteriol. **170**, 4322–4329.
166. Phillips, G.J. and Kushner, S.R. (1987) J. Bacteriol. **262**, 455-459.
167. Wilde, R.J. and Guest, J.R. (1986) J. Gen. Microbiol. **132**, 3239-3251.
168. Downing, W.L., Sullivan, S.L., Gotessman, M.E. and Dennis, P.P (1990) J. Bacteriol. **172**, 1621–1627.
169. Neuwald, A.F. and Stauffer, G.V. (1989) Gene **82**, 219–228.
170. Duncan, K. and Coggins, J.R. (1986) Biochem. J. **234**, 49–57.
171. Carlioz, A., Ludwig, M.L., Stallings, W.C., Fee, J.A., Steinman, H.M. and Touati, D. (1988) J. Biol. Chem. **263**, 1555–1562.
172. Post, L.E., Arfsten, A.E., Reusser, F. and Nomura, M. (1978) Cell **15**, 215–229.
173. Szumanski, M.B.W. and Boyle, S.M. (1990) J. Bacteriol. **172**, 538–548.
174. Barroso, L., Moore, R., Wright, J., Patel, T. and Boyle, S.M. (unpublished) M33766.
175. Brandsma, J.A., Bosch, D., de Ruyter, M. and van de Putte, P. (1985) Nucleic Acids Res. **13**, 5095–5109.

176. Spencer, M.E. and Guest, J.R. (1985) Mol. Gen. Genetics. **200**, 145–154.
177. Cole, S.T. (1983) Mol. Gen. Genet., **189**, 400–404.
178. Deeley, M.C. and Yanofsky, C. (1982) J. Bacteriol. **151**, 942–951.
179. Postle, K. and Good, R.F. (1983) Proc. Natl. Aca. Sci. USA. **80**, 5235–5239.
180. Repoila, F. and Gutierrez, C. (1991) Mol. Microbiol. **5**, 747–755.
181. Gustafsson, C., Lindstrom, P.H.R., Hagervall, T.G., Esberg, K.B. and Bjork, G.R. (1991) J. Bacteriol. **173**, 1757–1764.
182. Gunsalus, R.P. and Yanofsky, C. (1980) Proc. Natl. Acad. Sci. USA. **77**, 7117–7121.
183. Horowitz, H. and Platt, T. (1982) J. Mol. Biol. **156**, 257–267.
184. Aldea, M., Claverie-Martin, F., Diaz-Torres, M.R. and Kushner, S.R. (1988) Gene **65**, 101–110.
185. Bremer, E., Middendorf, A., Martinussen, J. and Valentin-Hansen, P. (1990) Gene **96**, 59–65.
186. Miyajima, A, Shibuya, M., Kuchino, Y. and Kaziro, Y. (1981) Mol. Gen. Genet. **183**, 13–19.
187. Yang, J. and Pittard, J. (1987) J. Bacteriol. **169**, 4710–4715.
188. Kasian, P.A., Davidson, B.E. and Pittard, J. (1986) J. Bacteriol. **167**, 556–561.
189. Cornish, E.C., Argyropoulos, V.P., Pittard, J. and Davidson, B.E. (1986) J. Biol. Chem. **261**, 403–410.
190. Kitagawa, Y., Akaboshi, E., Shinagawa, H., Horii, T., Ogawa, H. and Kato, T. (1985) Proc. Natl. Aca. Sci. USA. **82**, 4336–4340.
191. Varshney, U., Hutcheon, T. and van de Sande, J. (1988) J. Biol. Chem. **263**, 7776–7784.
192. Sancar, A., Sancar, G.B., Rupp, W.D., Little, J.W. and Mount, D.W. (1982) Nature **298**, 96–98.
193. Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. (1982) Cell **28**, 523–530.
194. Moolenaar, G.F., van Sluis, C.A., Backendorf, C. and van de Putte, P. (1987) Nucleic Acids Res. **15**, 4273-4289.
195. van Sluis, C.A., Moolenaar, G.F. and Backendorf, C. (1983) EMBO J. **2**, 2313–2318.
196. Easton, A.M. and Kushner, S.R. (1983) Nucleic Acids Res. **11**, 8625–8640.
197. Arthur, H.M, Cavavagh, D.R., Finch, P.W. and Emmerson, P.T. (1987) J. Bacteriol. **169**, 3435–3440.
198. Blnaco, C., Ritzenthaler, P. and Kolb, A. (1986) Mol. Gen. Genet. **202**, 112–119.
199. Heck, J.D. and Hartfield, G.W. (1988) J. Biol. Chem. **263**, 857-867.
200. Rowley, D.L. and Wolf, Jr. R.E. (1991) J. Bacteriol. **173**, 968-977.
201. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. **13**, 319-353.
202. Hawley, D.K., and McClure, W.R. (1983) Nucleic Acids Res. **11**, 2237–2255.
203. Harley, C.B., and Reynolds, R.P. (1987) Nucleic Acids Res. **15**, 2343–2361.
204. Nussinov, R. (1984) J. Biol. Chem. **259**, 6798–6805.
205. Plaskon, R.R. and Wartell, R.M. (1987) Nucleic Acids Res. **15**, 785–796.
206. Mulligan, M.E., Hawley, D.K., Entriken, R. and McClure, W.R. (1984) Nucleic Acids Res. **12**, 789–800.
207. Margalit, H., Shapiro, B.A., Nussinov, R., Owens, J. and Jernigan, R.L. (1988) Biochemistry **27** ,5179–5188.
208. O'Neill, M.C. (1989) J. Biol. Chem. **264**, 5522–5530.
209. O'Neill, M.C. and Chiafari, F. (1989) J. Biol. Chem. **264**, 5531-5534.
210. Berg, O.G. and von Hippel, P.H. (1987) J. Mol. Biol.**193**, 723–750.
211. Galas, D.J., Eggert, M. and Waterman, M.S. (1985) J. Mol. Biol. **186**, 117–128.
212. Nakata, K., Kanehisa, M. and Maizel, J.V. (1988) CABIOS. **4**, 367–371.
213. O'Neill, M.C. (1989) J. Mol. Biol. **207**, 301–310.
214. Alexandrov, N.N. and Mironov, A.A. (1990) Nucleic Acids Res. **18**, 1847–1852.
215. Demeler, B. and Zhou, G. (1991) Nucleic Acids Res. **19**, 1593-1599.
216. Rudd, K.E., Miller, W., Werner, C., Ostell, J., Tolstoshev, C., and Satterfield, S.G. (1991) Nucleic Acids Res. **19**, 637–647.