Sequencing and Characterization of a Gene Cluster Encoding the Enzymes for L-Rhamnose Metabolism in *Escherichia coli*

PILAR MORALEJO,¹ SUSAN M. EGAN,² ELENA HIDALGO,¹ AND JUAN AGUILAR^{1*}

Department of Biochemistry, School of Pharmacy, University of Barcelona, Diagonal 643, 08028 Barcelona, Spain,¹ and Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218²

Received 9 March 1993/Accepted 22 June 1993

The sequencing of the *Eco*RI-*Hin*dIII fragment complementing mutations in the structural genes of the L-rhamnose regulon of *Escherichia coli* has permitted identification of the open reading frames corresponding to *rhaB*, *rhaA*, and *rhaD*. The deduced amino acid sequences gave a 425-amino-acid polypeptide corresponding to rhamnulose kinase for *rhaB*, a 400-amino-acid polypeptide corresponding to rhamnulose kinase for *rhaB*, a 400-amino-acid polypeptide corresponding to rhamnulose kinase for *rhaB*, a 400-amino-acid polypeptide corresponding to rhamnulose labeled and a 274-amino-acid polypeptide corresponding to rhamnulose-1-phosphate aldolase for *rhaD*. Transcriptional fusions of the three putative promoter regions to *lacZ* showed that only the *rhaB* leader region acted as a promoter, as indicated by the high β -galactosidase activity induced by rhamnose, while no significant activity from the *rhaA* and *rhaD* constructions was detected. The *rhaB* transcription start site was mapped to -24 relative to the start of translation. Mutations in the catabolic genes were used to show that L-rhamnose may directly induce *rhaBAD* transcription.

L-Rhamnose, a methylpentose, is metabolized in *Escherichia coli* by a set of enzymes encoded by genes constituting the rhamnose regulon, which maps at 88.4 min in the chromosome (2). Four structural genes have been described: *rhaA*, encoding rhamnose isomerase; *rhaB*, encoding rhamnulose kinase; *rhaD*, encoding rhamnulose-1-phosphate aldolase (32); and *rhaT*, encoding the rhamnose transport system (17). The *rhaT* gene has been mapped in the *rha* locus, separated from *rhaA*, *rhaB*, and *rhaD* by the regulatory operon *rhaC*, which has been found to be formed by two partially overlapping genes, *rhaR* and *rhaS* (40). The gene order of the region, counterclockwise, is *glpK. . .sodA*-*rhaT-rhaR-rhaB-rhaA-rhaD*.

In *E. coli, rhaT*, encoding the transporter (17, 39), and *rhaR* and *rhaS*, governing expression (40, 41), have been sequenced and extensively analyzed. In this species, another methylpentose, L-fucose, is metabolized by a parallel metabolic pathway integrated by a set of specific enzymes encoded by the *fuc* gene cluster, which has been located at 60.2 min (23) and completely sequenced (11, 25). In *Salmonella typhimurium* LT2, *rhaB*, for rhamnulose kinase; *rhaC2*, one of the regulatory genes (31); and *rhaT*, encoding the transporter (39) have also been sequenced.

Here we present a sequence analysis of three structural genes of the rhamnose pathway, some experiments involving their expression, and a comparison with the corresponding gene sequences of the L-fucose system.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The bacterial strains used in this study are listed in Table 1. Cells were grown aerobically as described previously (7) on LB or minimal medium. For growth on minimal medium, L-rhamnose, glucose, or L-fucose was added to 0.2%. When indicated, X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopy-ranoside) was added to 40 µg/ml. Ampicillin was used at 100 µg/ml, kanamycin was used at 30 µg/ml, and streptomycin was used at 25 µg/ml. For primer extension analysis, the

strains were grown in M10 medium (33) containing 0.4% glycerol, 0.2% Casamino Acids, and 50 μ M thiamine, in the presence or absence of 0.2% rhamnose.

Preparation of cell extracts and enzyme assay. For enzyme assay, the cells were harvested at the end of the exponential phase and the cell extract was prepared as described previously (7) with 10 mM Tris-HCl buffer (pH 7.0). The β -galactosidase activity in strains grown under specified conditions was assayed as described by Miller, and the values are reported in the units defined by Miller (28).

The protein concentration in cell extracts was determined by the method of Lowry et al. (24) with bovine serum albumin as the standard.

DNA manipulation. Plasmid DNA was routinely prepared by the boiling method (34). For large-scale preparation, a crude DNA sample was subjected to purification by cesium chloride-ethidium bromide density gradient centrifugation or on a column (Qiagen GmbH, Düsseldorf, Germany). DNA manipulations were performed essentially as described by Sambrook et al. (34). The DNA sequence was determined by using the dideoxy-chain termination procedure of Sanger et al. (36). Double-stranded plasmid DNA was used as the template. Plasmid purified with a Qiagen column was used for the construction of ordered deletions with the Erase-a-Base system (Promega Biotec, Madison, Wis.). We resolved the numerous sequencing gel compressions as described previously (20).

Transcriptional fusions were constructed by inserting the DNA fragments into plasmid pRS550 of Simons et al. (37). Plasmid pRS550 carries a cryptic *lac* operon and genes that confer resistance to both kanamycin and ampicillin. After introduction of the recombinant plasmid into the streptomycin-resistant strain MC1061, blue colonies on X-Gal plates containing ampicillin, kanamycin, and streptomycin were isolated, and plasmid DNA was sequenced by using the M13 primer to ensure that the desired fragment was inserted in the correct orientation.

For the *rhaB* promoter, a fragment starting with a *Bgl*II site upstream of *Eco*RI (position 560 in the *rhaR-rhaS* sequence in reference 40) and ending with a *ClaI* site (position 392 in Fig. 2) was prepared by digestion of pJB3.1

^{*} Corresponding author.

TABLE 1. E. coli strains used in this work

Strain	Genotype	Source or reference
ECL1	HfrC phoA8 relA1 tonA22 T2 ^r (lambda)	13
DH5α	supE44 ΔlacU169(φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1	BRL
MC1061	hsdR mcrB araD139 ∆(araABC-leu) 7679 lacX74 galU galK rpsL thi	37
JA121	MC1061(pRS550)	This study
JA123	MC1061(pRS550-PrhaB-lacZ)	This study
JA124	MC1061(pRS550-PrhaA-lacZ)	This study
JA125	MC1061(pRS550-PrhaD-lacZ)	This study
JA126	MC1061(pRS550-PrhaBA-lacZ)	This study
ECL116	$\mathbf{F}^{-} \Delta \mathbf{a} \mathbf{c} \mathbf{U} 169$ end A hsd R thi	3
ECL339	As ECL116 but Δ(<i>rha-pfkA</i>)15 <i>zig-1</i> ::Tn10	10
ECL714	As ECL116 but rhaB101	12
ECL715	As ECL116 but rhaA502	12
ECL716	As ECL116 but rhaD701	12
ECL717	As ECL116 but <i>rhaR702</i>	12

(see Fig. 1), blunted, and ligated to vector pRS550 (37). Analogously, the *rhaA* fusion contained a *PvuI-NheI* fragment (positions 1178 to 1666 in Fig. 2), and the *rhaD* fusion contained a *SacII-Eco*RV fragment (positions 2888 to 3538 in Fig. 2) (strains JA123, JA124, and JA125, respectively). Another fusion contained the *BgIII-NheI* fragment encompassing the whole *rhaB* gene and the *rhaB-rhaA* intergenic space up to the 5' end of *rhaA* (strain JA126). A control with the vector pRS550 containing no insertion was strain JA121.

RNA preparation and Northern (RNA) blot hybridization. For the preparation of total RNA, cells of a 25-ml culture grown up to an A_{650} of 0.5 or 0.2 as indicated were collected by centrifugation and resuspended in 125 µl of ice-cold 0.3 M sucrose-0.01 M sodium acetate solution (pH 4.5). To this suspension, 125 µl of 0.01 M sodium acetate (pH 4.5) with 2% (wt/vol) sodium dodecyl sulfate (SDS) was added. The extract was then heated at 70°C for 3 min. The DNA and proteins were extracted with 1 volume of phenol at 70°C for 3 min and then cooled for 15 s in a -80° C bath by means of a dry ice-acetone freezing mixture. After centrifugation at $15,000 \times g$ for 5 min, the supernatant was extracted with phenol at 70°C two times more. The RNA was precipitated by addition of 1 ml of 100 mM sodium acetate in ethanol and then washed with 70% ethanol. After centrifugation, the total cellular RNA was resuspended in 50 µl of 20 mM sodium phosphate buffer (pH 6.5) containing 1 mM EDTA.

For the Northern blot hybridization, each RNA sample (10 μ g) was electrophoresed on a 1% agarose-formaldehyde gel and transferred to a nylon membrane filter (Schleicher & Schuell) in 10× SSPE (34). Prehybridization and hybridization were carried out in 50% formamide-5× Denhardt reagent-50 mM sodium phosphate (pH 6.5)-10 mM NaCl-0.1% SDS-125 μ g of sonicated salmon sperm DNA per ml at 42°C. The restriction nuclease fragments used as probes were 3' end labeled with the random-primed DNA labeling kit (Boehringer, Manheim, Germany). Filters were washed two times at room temperature and two times at 65°C in 2× SSC (34)-0.5% SDS and then washed once at 50°C and once at 55°C in 0.1× SSC. The filters were exposed to X-ray films at -70°C with an intensifying screen.

Primer extension analysis. RNA was isolated from cells harvested at an A_{600} of 0.2 as previously described (33),

except that 10 ml of cells was used and an ethanol precipitation was performed following the isopropanol precipitation. One microgram of RNA was mixed with 2.5 ng of ³²P-labeled primer (5'-AAAACGATGGATTTCGCGCAGC GTCAGGCT-3'), and primer extension reactions were performed as described previously (33).

Nucleotide sequence accession number. The nucleotide sequence of the *Eco*RI-*Hind*III fragment encompassing *rhaB*, *rhaA*, and *rhaD* has been deposited in the GenBank data library under accession no. X60472.

RESULTS AND DISCUSSION

Nucleotide sequence analysis and identification of *rhaB*, *rhaA*, and *rhaD*. In a previous study we used plasmid pJB4.1 (Fig. 1) to analyze the products of genes *rhaB*, *rhaA*, and *rhaD* (5). The *Eco*RI-*Hind*III fragment was inserted in Bluescript and designated plasmid pPM.2. This insert was subcloned in plasmids pPM.2.1 and pPM.2.2, containing fragments *Eco*RI-*Bam*HI and *Bam*HI-*Hind*III, respectively. Serial deletions of the three plasmids were obtained and sequenced by the strategy presented in Fig. 1. The DNA was sequenced at least once on each strand. Figure 2 depicts the 5,677 bp of DNA sequenced between the *Eco*RI and *Hind*III restriction sites as well as the proteins encoded by the five open reading frames (ORFs) found in the sequence.

According to previous mapping employing complementation (5), the first ORF downstream of the *Eco*RI site corresponded to *rhaB*, encoding rhamnulose kinase. The second one, which has the *Bam*HI site, corresponded to *rhaA*, encoding rhamnose isomerase, and the third one, including the *SalI* restriction site, corresponded to *rhaD*, encoding rhamnulose-1-phosphate aldolase. Two additional unidentified ORFs designated URF1 and URF2 were found between the end of *rhaD* and the *Hind*III restriction site.

rhaB corresponds to a 1,276-nucleotide ORF which can encode a 425-amino-acid polypeptide with a calculated molecular mass of 47,708 Da. No -10 or -35 boxes were apparent upstream from rhaB, but a catabolite repression protein recognition inverted repeat could be identified between positions 50 and 65. rhaA corresponds to a 1,200nucleotide ORF which can encode a polypeptide of 400 amino acids with a calculated molecular mass of 44,246 Da, and rhaD corresponds to an 822-nucleotide ORF which can encode a polypeptide of 274 amino acids with a calculated molecular mass of 30,149 Da. No evident transcription initiation regulatory signal was apparent in the rhaB-rhaA and rhaA-rhaD intergenic spaces. The N-terminal amino acid sequences of rhamnose isomerase and rhamnulose-1phosphate aldolase (4) were in agreement with the assigned ORFs. Unfortunately, the N-terminal end of rhamnulose kinase was blocked, and the sequence is not known.

Two sequences seem to fulfill the requirements of a transcription terminator (dashed arrows in Fig. 2) (14, 29). The first one has a 14-bp inverted repeat separated by 4 bp that could form a stable stem-loop structure with a calculated free energy of stabilization of 24 kcal (ca. 100.4 kJ) and is followed by two thymidine residues. The second one, downstream of *rhaD* and more likely to be functional, has a 10-bp inverted repeat separated by 4 bp, which would correspond to a stem-loop with a calculated free energy of stabilization of 14 kcal (ca. 58.5 kJ), followed by four thymidine residues. This terminator is located 330 bp downstream from the TAA stop codon, an unusually long distance between translation and transcription termination; perhaps there is posttranscriptional processing, as is the case for the



FIG. 1. Restriction map of the genomic region containing the rhamnose system. The thin line represents the genomic DNA, while the thick lines correspond to the fragments subcloned into plasmid pUC18 (pUC) or Bluescript (BS). Thick arrows at top show the directions of transcription and the positions of *rha* genes. The sequencing strategy for the *Eco*RI-*Hind*III fragment encompassing the *rhaB*, *rhaA*, and *rhaD* genes is also presented. Thin arrows indicate the start point, direction, and extent of sequence determined from each subclone. If a particular sequence was obtained more than once, this is indicated by a number over the arrow.

lactose operon of *E. coli* (27), or rho-dependent termination (16, 29).

Analysis of DNA sequences of the *rhaA-rhaD* intergenic fragment showed five 88- to 92-bp repetitions plus an incomplete sixth one (Fig. 2). Each complete element was a combination of the motifs described by Gilson et al. (18, 19) which appeared sequentially as follows: an REP (PU) palindromic unit of 34 bp (Y motif) highly homologous to the REP consensus previously described (15, 38), a right internal segment (S motif), a second PU sequence in the opposite orientation (Z^2 motif), and a left internal segment (s motif). The sixth repetition was a Y motif followed by a B-like external fragment. This combination of motifs is in agreement with what is generally known as the BIME (bacterial interspersed mosaic elements) family (18, 19), although the number of repetitions found in other intergenic regions is usually lower than that in the one described here. The function of these short, interspersed repetitive DNA sequences in the regulation of gene expression has been widely discussed (26) but remains uncertain. No repetitive sequences were found in the other intergenic fragments of the rhamnose regulon.

Transcription. Total RNA of cells of strain ECL1 grown aerobically on L-rhamnose was prepared as indicated in Materials and Methods. Northern blot hybridizations showed a major RNA of ca. 1.4 kb for an *rhaB* probe and transcripts of ca. 1.6 and 2.5 kb for both *rhaA* and *rhaD* probes. A larger probe containing *rhaB*, *rhaA*, and part of *rhaD* also gave (but only for some mRNA preparations from early exponential-phase growth) a minor band which might correspond to a 3.8-kb transcript (Fig. 3). RNA prepared from cells of the same strain grown on L-fucose or glucose gave no band of hybridization with any of the probes used (not shown), indicating the specificity for L-rhamnose of the RNA analysis performed.

Strains carrying transcriptional fusions (see Materials and Methods) were grown under different conditions, and β -galactosidase activity in their extracts was determined. As shown in Table 2, the promoter of *rhaB* yielded high activity in growth on L-rhamnose, while the promoters of *rhaA* and *rhaD* yielded very low activities. Thus, only the *rhaB* leader region appears to contain an L-rhamnose-inducible promoter. The strain containing the *rhaBA-lacZ* fusion also displayed high activities, indicating that no strict termination occurs between these two genes. Growth on glucose gave very low activities, close to basal levels. Similarly, growth on the isomer sugar L-fucose, which differs only in the stereoconfiguration at carbons 2 and 4, yielded undetectable activity, indicating no cross induction with this sugar.

In view of these results, we conclude that *rhaBAD* is probably transcribed as a single transcription unit and that the smaller RNAs observed (Fig. 3) may result from degradation or in vivo processing. The role of the differential mRNA stability in the regulation of gene expression of *rhaBAD*, as described for other systems (9, 30, 42), deserves more study. Moreover, intercistronic transcription terminators, such as the one proposed to exist between *rhaA* and *rhaD*, could also be acting as gene expression regulators (30). Alternatively, this stem-loop structure could simply

J. BACTERIOL.

10 Сааттттсассааа	20 TGCGGTGAGCA'	30 TCACATCAC	40 САСААТТСАС	50	CRP 60 GAACATCATCAC	70 GTTCATCTTT	
			<u> </u>				rhaB
100	110	120 Eco 1	RI130	140	150▼ ▼	v 160	170 180
CCCATTTTCCTGTC	AGTAACGAGAA	GGTCGCGAA	ITCAGGCGCI	TTTTAGAC	TGGTCGTAATGA	AATTCAGC <u>AG</u>	GATCACATTATGACCT
							M T
190	200	210	220	230	240	250	260 270
F R N C V	A V D L	G A S	S G R	V M L	A R Y E	R E C	R S L T L
280	290	300	310	320	330	340	••••••••••••••• 350 260
GCGAAATCCATCGT	TTTAACAATGG	GCTGCATAG	ICAGAACGG	TATGTCAC	CTGGGATGTGGA	TAGCCTTGAA	AGTGCCATTCGCCTTG
REIHR	FNNG	LHS	QNG	YVТ	WDVD	SLE	SAIRL
370	380	390 ClaI	400	410	420	430	440 450
GATTAAACAAGGTG	TGCGAGGAAGG C E E G	GATTCGTAT(I R I	CGATAGCATI DSI	GGGATTGA	TACCTGGGGGCGT T W G V	GGACTTTGTG	CTGCTCGACCAACAGG
460 GTCAGCGTGTGGGGC	470 CTGCCCGTTGC	480 TTATCGCGA	490 TAGCCGCACO	500 AATGGCCT	510 AATGGCGCAGGC	520 ACAACAACAA	530 540 CTCGGCAAACGCGATA
GQRVG	LPVA	YRD	SRT	NGL	MAQA	Ω Q Q	LGKRD
550	560	570	580	590	600	610	620 630
TTTATCAACGTAGC	GGCATCCAGTT	TCTGCCCTT	CAATACGCT	TATCAGTT	GCGTGCGCTGAC	GGAGCAACAA	CCTGAACTTATTCCAC
IIYKS	GIQF	ы г г	ИТП	түг		ΡŲŲ	F B D I F
640 እር እጥጥር ርጥር እርርርርጥ	650 ഗന്ദാനമാനദാന	660 നേരമന്നമരനസ്ത	670 നമരണ്ണമന്റരം	680 200006970	690 CAACATCAACTC	700 2023 3 T 3 T 3 C C	
H I A H A	L L M P	D Y F	S Y R	L T G	K M N W	E Y T	N A T T T
730	740	750	760	770	780	790	800 810
AACTGGTCAATATC	AATAGCGACGA	CTGGGACGA	GTCGCTACTO	GCGTGGAG	CGGGGCCAACAA	AGCCTGGTTI	GGTCGCCCGACGCATC
QLVNI	NSDD	WDE	SLL	AWS	GANK	. A W F	GRPTH
820	830	840	850	860	870	880	890 900
P N V I G	H W I C	P Q G	N E I	P V V	A V A S	H D T	A S A V I
910	920	930	940	950	960	970	980 990
CCTCGCCGTTAAAC	GGCTCACGTGC	TGCTTATCT	CTCTTCTGG	CACCTGGTC	ATTGATGGGCTI	CGAAAGCCAG	ACGCCATTTACCAATG
ASPLN	GSRA	AYL	SSG	TWS	LMGF	'ESQ	TPFTN
1000	1010	1020	1030	1040	1050	1060	1070 1080
ACACGGCACTGGCA	GCCAACATCAC	CAATGAAGG N E G	CGGGGGGGGAA G A E	AGGTCGCTA G R Y	TCGGGTGCTGAA R V L K	AAATATTATG	GCTTATGCCTCCTTC G L W L L
1090 AGCGAGTGCTTCAG	1100 GAGCAGCAAAT	1110 CAACGATCT	1120 TCCGGCGCT	1130	1140 GACACAGGCACI	1150 TCCGGCTTGC	1160 1170 CGCTTCATTATCAATC
QRVLQ	EQQI	NDL	PAL	ISA	TQAL	PAC	RFIIN
1180PvuI	1190	1200	1210	1220	1230	1240	1250 1260
CCAATGACGATCGC	TTTATTAATCC	TGAGACGAT	GTGCAGCGA	ATTCAGGC	TGCGTGTCGGGA	AACGGCGCAA	CCGATCCCGGAAAGTG
PNDDR	FINP	ETM	CSE	IŲA	ACRE	таў	PIPES
1270	1280	1290	1300		1320	1330	1340 1350 COCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
D A E L A	R C I F	D S L	A L L	Y A D	V L H E	L A H	V R G E D
1260	1370	1380	1390	1400	1410	1420	1430 1440
TCTCGCAACTGCT	TATTGTCGGCG	GAGGCTGCC.	AGAACACGC	IGCTCAACC.	AGCTATGCGCCG	ATGCCTGCGG	TATTCGGGTGATCGCC
FSQLL	YCRR	RLP	ЕНА	A Q P	AMRF	CLR	YSGDR
1450	1460	1470	1480	1490	1500	1510	1520 1530
GGGCCTGTTGAAGC R A C *	CTCGACGCTCG	GCAATATCG	GCATCCAGT	I'AA'IGACGC	TGGATGAACTCA	ACAATGTGGA	TIGATITICCGTCAGGTC

FIG. 2. Nucleotide sequence of the *Eco*RI-*Hin*dIII fragment. The coding regions of the ORFs present in the fragment have been translated and are indicated by the single-letter amino acid code. The restriction sites used in this work are indicated. The inverted repeat constituting a good catabolite repression protein (CRP) consensus binding site is indicated by heavy underlining. DNA sequences predicted to form hairpin loop structures are shown by dashed arrows. The putative Shine-Dalgarno sequences are underlined. Stop codons are indicated by asterisks. The major transcription start site is indicated by a closed triangle, while open triangles show the two minor or processing transcription start sites found. Dots indicate the position and length of the primer used in primer extension analysis. The locations of motifs in repetitive DNA sequences as described in the text are as follows: Y(>), $Z^2(<)$, S(#), s(*), and B-like (+).

1540	1550	1560	1570	1580	1590	1600	1610	1620
GTCAGCACCACCGCG	AATCTGACCA	CCTTTACCC	CTAATCCTGA	CAGTGAAATTG	CCCACTATG	TGCGCAGATI	CACTCTACACG	ACAGA
	rhaA							
1630	1640	1650	1660 NheI	1670	1680	1690	1700	1710
C <u>AAAGGAG</u> CTTIGCG	CATGACCACI	CAACIGGAA	CAGGCCTGGG	AGCTAGCGAAA	CAGCGITITC	GCGGCGGIGG	GGATIGATGTC	GAGGA
	мтт	QLE	QAWI	ELAK	QRF	AAV	GIDV	Е Е
1700	1720	1740	1750	1760	1770	1700	1700	1000
								TOOO
	DBI			J O C D	DVC	CEE	M D E C	
ADRQL		r v S	маст	v y u D	D V 3	GrE	NFEG	3 1
1810	1820	1830	1840	1850	1860	1870	1880	1890
GACCGGGGGGATTCA	222676762	DOOT TATTA	020006667000	TATATCCACT	GACCTACCT	ວລາວມີຮູ້	2000 Dramana	TODE
TGGIO	A T G	N Y P	GKAI	RNAS	E L R	A D L	EOAM	RL
1900	1910	1920	1930	1940	1950	1960	1970	1980
GATTCCGGGGCCGAA	ACGGCTTAAI	TTACATGCC	ATCTATCTGG	AATCAGATACG	CCAGTCTCG	CGCGACCAGA	TCAAACCAGAG	CACTT
IPGPK	RLN	LHA	IYLI	ESDT	P V S	RDQ	IKPE	H F
1990	2000	2010	2020	2030	2040	2050	2060	2070
CAAAAACTGGGTTGA	ATGGGCGAAA	GCCAATCAG	CTCGGTCTGG	ATTTTAACCCC	TCCTGCTTT	TCGCATCCGC	TAAGCGCCGAT	GCTT
K N W V E	W A K	A N Q	LGLI) F N P	SCF	SHP	LSAD	GF
2080	2090	2100	2110	2120	2130	2140	2150	2160
TACGCTTTCCCATGC	CGACGACAGO	CATTCGCCAG	TTCTGGATTG	ATCACTGCAAA	GCCAGCCGT	CGCGTTTCGC	SCCTATTTTGGC	GAGCA
TLSHA	DDS	IRQ	FWI	рнск	ASR	RVS	AYFG	ΕQ
0480								
2170	2180	2190BamHI	2200	2210	2220	2230	2240	2250
ACTEGGEACACCATE	GGIGAIGAA	CATCIGGATC	CCCGGATGGTA	IGAAAGATATC	ACCGIIGAC	CGTCTCGCCC	CGCGTCAGCGT	CIGCI
LGTРS	S V M N	1 W 1	PDGI	MKDI	TVD	RLA	PRQR	чг
2260	2270	2200	2200	2200	2210	2220	2220	2240
	22/0 	2200 2020 2000	2230 3300000000	2300 ACCAMAMCCAC	2310 CCCCMMC2C	2320 20022200000		2340
		FYI	N D A	U U T D	A V F	C V L	FGIG	λĒ
			NIA		A V D	5 K H	r G I G	л ц
2350	2360	2370	2380	2390	2400	2410	2420	2430
2350 GAGCTACACGTTG	2360	2370	2380	2390	2400	2410	2420 300066662	2430
2350 GAGCTACACGGTTGG S Y T V G	2360 CTCCAATGAC	2370 STTTTACATG F Y M	2380 GGGTATGCCA	2390 CCAGCCGCCAG	2400 GACTGCGCTG	2410 TGCCTGGACC	2420 GCCGGGCACTTC A G H F	2430 CACCC H P
2350 GAGCTACACGGTTGO SYTVO	2360 CTCCAATGAC S N E	2370 GTTTTACATG F Y M	2380 GGGTATGCCA G Y A	2390 CCAGCCGCCAG ISRQ	2400 ACTGCGCTG T A L	2410 TGCCTGGACC C L D	2420 SCCGGGCACTTC A G H F	2430 CACCC H P
2350 GAGCTACACGGTTGG SYTVC 2440	2360 CTCCAATGAC SNE 2450	2370 STTTTACATG F Y M 2460	2380 GGGTATGCCA G Y A 2470	2390 CCAGCCGCCAG ISRQ 2480	2400 SACTGCGCTG T A L 2490	2410 TGCCTGGACC C L D 2500	2420 GCCGGGCACTTC A G H F 2510	2430 CACCC H P 2520
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC	2360 CTCCAATGAC S N E 2450 CCGACAAGAT	2370 STTTTACATG F Y M 2460 FTCCGCCGCC	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG	2390 CCAGCCGCCAG ISRQ 2480 IGCCGCAGTTG	2400 SACTGCGCTG T A L 2490 SCTGCTGCAC	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG	2430 CACCC H P 2520 GACAG
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S	2360 CTCCAATGAC S N E 2450 CCGACAAGAT S D K I	2370 STTTTACATG F Y M 2460 S A A	2380 GGGTATGCCA GYA 2470 CATGCTGTATG MLY	2390 CCAGCCGCCAG F S R Q 2480 IGCCGCAGTIG V P Q L	2400 SACTGCGCTG T A L 2490 SCTGCTGCAC L L H	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R	2420 SCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W	2430 CACCC H P 2520 GACAG D S
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S	2360 CTCCAATGAC S N E 2450 CCGACAAGAT S D K I	2370 STTTTACATG F Y M 2460 FTCCGCCGCC S A A	2380 GGGTATGCCA GYA 2470 CATGCTGTATG MLY	2390 CCAGCCGCCAG T S R Q 2480 TGCCGCAGTTG V P Q L	2400 BACTGCGCTG T A L 2490 BCTGCTGCAC L L H	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W	2430 CACCC H P 2520 GACAG D S
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530	2360 CTCCAATGAC S N E 2450 CCGACAAGAT S D K I 2540	2370 STTTTACATG F Y M 2460 FTCCGCCGCC S A A 2550	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560	2390 CCAGCCGCCAG ISRQ 2480 IGCCGCAGTTG VPQL 2570	2400 BACTGCGCTG T A L 2490 BCTGCTGCAC L L H 2580	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600	2430 CACCC H P 2520 GACAG D S 2610
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT	2360 CTCCAATGAC S N E 2450 CCGACAAGATT S D K I 2540 CCTGGATGAT	2370 STTTTACATG F Y M 2460 FTCCGCCGCC S A A 2550 FGAAACCCAG	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA	2390 CCAGCCGCCAG I S R Q 2480 IGCCGCAGTIG V P Q L 2570 GIGAGATIGIG	2400 BACTGCGCTG T A L 2490 BCTGCTGCAC L L H 2580 BCGTCACGAT	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC	2430 CACCC H P 2520 GACAG D S 2610 GGCCT
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I	2360 CTCCAATGAC S N E 2450 CGACAAGAT S D K I 2540 VGCTGGATGAT L D D	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A	2390 CCAGCCGCCAG ISRQ 2480 IGCCGCAGTIG VPQL 2570 GIGAGATIGIG SEIV	2400 SACTGCGCTG T A L 2490 SCTGCTGCAC L L H 2580 SCGTCACGAT R H D	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I	2360 CTCCAATGAC S N E 2450 CGGACAAGAT S D K I 2540 CGCTGGATGAT L D D	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A	2390 CCAGCCGCCAG I S R Q 2480 IGCCGCAGTIG V P Q L 2570 GIGAGATIGIG S E I V	2400 SACTGCGCTG T A L 2490 SCTGCTGCAC L L H 2580 SCGTCACGAT R H D	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620	2360 CTCCAATGAC S N E 2450 CCGACAAGAT S D K I 2540 NCCTGCATGAT L D D 2630	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q 2640	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650	2390 CCAGCCGCCAG I S R Q 2480 IGCCGCAGTTG V P Q L 2570 GIGAGATIGIG S E I V 2660	2400 CACTGCGCTG TAL 2490 CTGCTGCAC LLH 2580 CGTCACGAT RHD 2670	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGG	2360 CTCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCTCGATGATGAT L D D 2630 CCTCTATCAAC	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 GCGTGGGTCA	2390 CCAGCCGCCAG ISRQ 2480 IGCCGCAGTIG VPQL 2570 GIGAGAIIGIG SEIV 2660 ITGGIACACGC	2400 CACTGCGCTG T A L 2490 CCTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGG	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTTG	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGG D F F D A	2360 CCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCCTGGATGAT L D D 2630 CCTCTATCAAC S I N	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V	2390 CCAGCCGCCAG ISRQ 2480 RGCCGCAGTTG VPQL 2570 GTGAGATTGTG SEIV 2660 TTGGTACACGC IGTR	2400 CACTGCGCTG T A L 2490 CCTGCTGCAC L L H 2580 CCGTCACGAT R H D 2670 CAATATGAAA N M K	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTTG L R A L	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGC D F F D Z	2360 CTCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCTGGATGATGAT L D D 2630 CCTCTATCAAC S I N	2370 STTTTACATG F Y M 2460 FTCCGCCGCC S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A	2380 GGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V	2390 CCAGCCGCCAG ISRQ 2480 RGCCGCAGTTG VPQL 2570 GTGAGATTGTG SEIV 2660 TTGGTACACGC IGTR	2400 CACTGCGCTG T A L 2490 CTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAGCCCTGC K A L	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGGC D F F D Z 2710	2360 CTCCAATGAC S N E 2450 CCGACAAGAT S D K I 2540 CCCTGGATGATGAT L D D 2630 CCTCTATCAAC S I N 2720	2370 STTTTACATG F Y M 2460 PTCCGCCGCC S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 GCGTGGGTCA A W V 2740	2390 CCAGCCGCCAG ISRQ 2480 IGCCGCAGTTG VPQL 2570 GTGAGATTGTG SEIV 2660 ITGGTACACGC IGTR 2750	2400 CACTGCGCTG T A L 2490 CTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K 2760	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGC D F F D A 2710 ACCTACCGCTGAGCT	2360 CTCCAATGAC S N E 2450 CGGACAAGAT D K I 2540 CGCTGGATGAT L D D 2630 CTCTTATCAAC S I N 2720 CGCGCAAGCTY	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCGC	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 GCGGTGGGTCA A W V 2740 GGGGGATTACA	2390 CCAGCCGCCAG I S R Q 2480 IGCCGCAGTIG V P Q L 2570 GIGAGATIGIG S E I V 2660 IGGTACACGC I G T R 2750 CIGGCGCGGTCI	2400 CAL A L 2490 CAL A L 2580 CATACACAA R H D 2670 CAATATGAAA N M K 2760 CACACTGCT	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGG L R A L 2780 GAAATCGTTGCC	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGC D F F D A 2710 ACCTACCGCTGAGCT P T A E I	2360 CCCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCTCGATGAT L D D 2630 CCTCTATCAAC A S I N 2720 CCCCCAAGCTC R K L	2370 STTTTACATG F Y M 2460 TCCGCCGCC S A A 2550 TGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCC E A A	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 GCGTGGGTCA A W V 2740 GGCGATTACA G D Y	2390 CCAGCCGCCAG T S R Q 2480 TGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGTCT T G A S	2400 TAL 2490 CTGCTGCAC LLH 2580 CGTCACGAT RHD 2670 CAATATGAAA NMK 2760 CGCACTGCT GTA	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A
2350 GAGCTACACGGTTGG S Y T V C 2440 GACTGAAGTGATTTC T E V I S 2530 CGATCACGTAGTGCT D H V V I 2620 TGACTTCTTCGATGG D F F D Z 2710 ACCTACCGCTGAGCT P T A E I	2360 CTCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCTCTGGATGAT L D D 2630 CCTCTATCAAC A S I N 2720 CCCCCAAGCTC R K L	2370 STTTTACATG F Y M 2460 S A A 2550 GAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 SGAAGCGGCC E A A	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y	2390 CCAGCCGCCAG F S R Q 2480 FGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGCGTCT T G A S	2400 CACTGCGCTG T A L 2490 CCTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K 2760 CGCACTGCT G T A	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 CGAAGAGCAC G R A	2420 CCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 SAAATCGTTGCC E I V A	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A
$\begin{array}{c} 2350\\ \text{GAGCTACACGGTTGG}\\ \text{S} & \text{Y} & \text{T} & \text{V} & \text{O}\\ 2440\\ \text{GACTGAAGTGATTTGT}\\ \text{T} & \text{E} & \text{V} & \text{I} & \text{S}\\ 2530\\ \text{CGATCACGTAGTGCT}\\ \text{D} & \text{H} & \text{V} & \text{V} & \text{I}\\ 2620\\ \text{TGACTTCTTCGATGG}\\ \text{D} & \text{F} & \text{F} & \text{D} & \text{A}\\ 2620\\ \text{TGACTTCTTCGATGG}\\ \text{D} & \text{F} & \text{F} & \text{D} & \text{A}\\ 2710\\ \text{ACCTACCGCTGAGCT}\\ \text{P} & \text{T} & \text{A} & \text{E} & \text{I}\\ 2800\\ \text{ACCTTCTTCGATA}\\ \end{array}$	2360 CCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCCTCGATGATGAT L D D 2630 CCTCTATCAAC S I N 2720 CCCCCAAGCTC CCCCAAGCTC R K L 2810	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 IGAAGCGGCC E A A 2820	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830	$\begin{array}{c} 2390\\ \text{CCAGCCGCCAG}\\ \textbf{\Gamma} & \textbf{S} & \textbf{R} & \textbf{Q}\\ 2480\\ \text{ReccecAGTTG}\\ \textbf{V} & \textbf{P} & \textbf{Q} & \textbf{L}\\ 2570\\ \text{STGAGATTGTG}\\ \textbf{S} & \textbf{E} & \textbf{I} & \textbf{V}\\ 2660\\ \textbf{TGGTACACGC}\\ \textbf{I} & \textbf{G} & \textbf{T} & \textbf{R}\\ 2750\\ \text{CTGGCGCGCGTCT}\\ \textbf{T} & \textbf{G} & \textbf{A} & \textbf{S}\\ 2840\\ \text{CCTACCCANTCACCCANTCACCC}\\ \end{array}$	2400 CACTGCGCTG T A L 2490 CCTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CATATGAAA N M K 2760 CGCACTGCT G T A 2850	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880
$\begin{array}{c c} 2350\\ GAGCTACACGGTTGG\\S Y T V C\\ 2440\\ GACTGAAGTGATTTCTTCT T E V I S 2530\\ CGATCACGTAGTGGTTGGT D H V V I I 2620\\ TGACTTCTTCGATGCC D F F D Z 2710 ACCTACCGCTGAGCCT P T A E I 2800 AGGCCGTCTGGGAAA$	2360 CCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCTCGATGATGAT L D D 2630 CCTCTATCAAC S I N 2720 CCCCCAAGCTC C R K L 2810 CCTCTATCGCCA	2370 STTTTACATG F Y M 2460 S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCC E A A 2820 AACGTCACGA	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 TACGCCAGCA X A S	2390 CCAGCCGCCAG ISRQ 2480 RGCCGCAGTTG VPQL 2570 GTGAGATTGTG SEIV 2660 TTGGTACACGC IGTR 2750 CTGGCGCGCTCI IGAS 2840 GGTAGCGAATG	2400 CACTGCGCTG T A L 2490 CTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CATATGAAA N M K 2760 CGCACTGCT G T A 2850 CGCTGGAGAG	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860 CCGTGCGGGCC	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 TTATGAGAAAGA	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	2360 CTCCAATGAC S N E 2450 CGACAAGAT S D K I 2540 CGCTGGATGAT L D D 2630 CCTCTATCAAC S I N 2720 CCCCCAAGCTC R K L 2810 ATGTATTGCCA V L P	2370 F Y M 2460 FCCGCCGCC S A A 2550 GAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 CGAAGCGCGCG E A A 2820 AACGTCACGA T S R	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 TACGCCAGCA Y A S	2390 CCAGCCGCCAG F S R Q 2480 RGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGCTCT T G A S 2840 GGTAGCGAATC R *	2400 CACTGCGCTG TAL 2490 CTGCTGCAC LLH 2580 CGTCACGAT RHD 2670 CATATGAAA NMK 2760 CGCACTGCT GTA 2850 CGTGGAGAGAG	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860 GCGTGCGGGCC	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 TTATGAGAAAGA	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT
$\begin{array}{c} 2350\\ \text{GAGCTACACGGTTGGC}\\ \text{S} & \text{Y} & \text{T} & \text{V} & \text{C}\\ 2440\\ \text{GACTGAAGTGATTTC}\\ \text{T} & \text{E} & \text{V} & \text{I} & \text{S}\\ 2530\\ \text{CGATCACGTAGTGGTT}\\ \text{D} & \text{H} & \text{V} & \text{V} & \text{I}\\ 2620\\ \text{TGACTTCTTCGATGC}\\ \text{D} & \text{F} & \text{F} & \text{D} & \text{A}\\ 2620\\ \text{TGACTTCTTCGATGGC}\\ \text{D} & \text{F} & \text{F} & \text{D} & \text{A}\\ 2710\\ \text{ACCTACCGCTGAGCT}\\ \text{P} & \text{T} & \text{A} & \text{E} & \text{I}\\ 2800\\ \text{AGGCGGTCTGGGAAAGC}\\ \text{G} & \text{G} & \text{L} & \text{G} & \text{M}\\ \end{array}$	2360 CTCCAATGAC S N E 2450 CGGACAAGAT D K I 2540 CGCTGGATGAT L D D 2630 CTCTATCAAC S I N 2720 CGCGCAAGCTC R K L 2810 TGTATTGCCI V L P 2900	$\begin{array}{c} 2370\\ \textbf{F} \textbf{Y} \textbf{M}\\ 2460\\ \textbf{FCCGCCGCC}\\ \textbf{S} \textbf{A} \textbf{A}\\ 2550\\ \textbf{GAAACCCAG}\\ \textbf{E} \textbf{T} \textbf{Q}\\ 2640\\ \textbf{CCGCATTGCC}\\ \textbf{R} \textbf{I} \textbf{A}\\ 2730\\ \textbf{GAAGCGGCGC}\\ \textbf{E} \textbf{A} \textbf{A}\\ 2820\\ \textbf{AACGTCACGA}\\ \textbf{T} \textbf{S} \textbf{R}\\ 2910 \end{array}$	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 GCGTGGGTCA A W V 2740 GGCGATTACA G D Y 2830 TACGCCAGCA Y A S 2920	2390 CCAGCCGCCAG F S R Q 2480 IGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGTCT T G A S 2840 GGTAGCGAATG R *	2400 CACTGCGCTG TAL 2490 CTGCTGCAC LLH 2580 CGTCACGAT RHD 2670 CAATATGAAA NMK 2760 CGCACTGCT GTA 2850 CGCTGGAGAG	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860 CGTGCGGGCC	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 ITATGAGAAAGA 2960	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGGC\\S&Y&T&V\\ \hline \\ 2440\\ \hline \\ GACTGAAGTGATTTC\\T&E&V&I\\ \\ 2530\\ \hline \\ CGATCACGTAGTGCT\\D&H&V&V&I\\ \hline \\ 2620\\ \hline \\ TGACTTCTTCGATGCT\\D&F&F&D&A\\ \hline \\ 2620\\ \hline \\ TGACTTCTTCGATGCT\\CGATGCCTGAGCT\\P&T&A&E&I\\ \hline \\ 2800\\ \hline \\ AGGCGGTCTGGGAAA\\G&G&L&G&N\\ \hline \\ 2890SacII\\ \hline \\ TGAGTCGCCGCCGCCGCCGCCCCCCCCCCCCCCCCCCCC$	2360 CCCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCTCGATGAT L D D 2630 CCTCTATCAAC S I N 2720 CCGCCAAGCTC R K L 2810 CCGCCAAGCTC V L P 2900	2370 STTTTACATG F Y M 2460 TCCGCCGCC S A A 2550 TGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCG E A A 2820 AACGTCACGA T S R 2910 CGGATGCGCG	2380 GGGTATGCAA G Y A 2470 ATGCTGTATG M L Y 2560 GGCAATTGCCAA A I A 2650 GGCGTGGGTCAA A W V 2740 GGCGATTACAA G D Y 2830 ATACGCCAGCA Y A S 2920 GGCGAGCGCCT	$\begin{array}{c} 2390\\ CCAGCCGCCAGTGCCCAGCCGCCCCAGTF S R Q2480FGCCGCAGTTGV P Q L2570GTGAGATTGTGS E I V2660TTGGTACACGCI G T R2750CTGGCGCGCTCTT G A S2840GGTAGCGAATCR *$	2400 ACTGCGCTGCACG T A L 2490 CCTGCTGCACC L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K 2760 CGCCACTGCT G T A 2850 CGCTGGAGAGAG 2940 ACCGGTCGCC	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860 CCGTGCGGGCC	2420 3CCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGG L R A L 2780 SAAATCGTTGCC E I V A 2870 TTATGAGAAAGA 2960 AGGCCTGATAAC	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACCCG
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGAC\\S Y T V C\\ 2440\\ \hline \\ GACTGAAGTGATTACTCTCACGTAGTGATTACTCTCACGTAGTGCCTCD H V V I I\\ 2620\\ \hline \\ TGACTTCTTCGATGCTCTCCGATGACTCTCCTCCGATGACCTD F F D A ACCTACCGCTGAGCCTGAGCCTCACGCTGAGCTCTCGGAAAACG G G L G N 2890SacII TGAGTCGCCGCCGCGCTCTGGCAAAACG G G L G N 10000000000000000000000000000000000$	2360 CCCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 NGCTGGATGAT L D D 2630 CCTCTATCAAC A S I N 2720 NGCGCAAGCTC R K L 2810 NGCTATTGCCZ V L P 2900 TAAACACTGCC	2370 STTTTACATG F Y M 2460 TCCGCCGCC S A A 2550 TGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCG E A A 2820 AACGTCACGA T S R 2910 CCGCATGCGGC	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 TACGCCAGCA Y A S 2920 CGCGAGCGCCT	2390 CCAGCCGCCAG F S R Q 2480 FGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGGTCT T G A S 2840 GGTAGCGAATC R * 2930 TATCCGGCCTA	2400 ACTGCGCTG T A L 2490 CCTGCTGCACC L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K 2760 CGCACTGCT G T A 2850 SGCTGGAGAG 2940 ACGGGTCGGC	2410 TGCCTGGACG C L D 2500 GTCAGCCGTQ V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 CGAAGAGCAC G R A 2860 CGTGCGGGCC 2950 CAACAGTTGT	2420 3CCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CCGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 SAAATCGTTGCC E I V A 2870 TTATGAGAAAGA 2960 AGGCCTGATAAG	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCG C>
$\begin{array}{c} 2350\\ \text{GAGCTACACGGTTGG}\\ \text{S} & \text{Y} & \text{T} & \text{V} & \text{O}\\ 2440\\ \text{GACTGAAGTGATTTG}\\ \text{T} & \text{E} & \text{V} & \text{I} & \text{S}\\ 2530\\ \text{CGATCACGTAGTGCT}\\ \text{D} & \text{H} & \text{V} & \text{V} & \text{I}\\ 2620\\ \text{TGACTTCTTCGATGG}\\ \text{D} & \text{F} & \text{F} & \text{D} & \text{A}\\ 2620\\ \text{TGACTTCTTCGGAGGGGT}\\ \text{P} & \text{T} & \text{A} & \text{E} & \text{I}\\ 2800\\ \text{AGGCGGTCTGGGAAAGGG}\\ \text{G} & \text{G} & \text{L} & \text{G} & \text{N}\\ 2890\text{SacII}\\ \text{TGAGTCGCCGCGGGGGT}\\ \end{array}$	2360 CTCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 VGCTGGATGAT L D D 2630 CTCTTATCAAC A S I N 2720 VGCGCAAGCTC R K L 2810 VGCGCAAGCTC V L P 2900 CAAACACTGCC	2370 STTTACATG F Y M 2460 TCCGCCGCC S A A 2550 TGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 SGAAGCGGCG E A A 2820 AACGTCACGA T S R 2910 CGGATGCGGC	2380 GGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 TACGCCAGCA Y A S 2920 CGCGAGCGCCT	2390 CCAGCCGCCAG F S R Q 2480 RGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGGTCT T G A S 2840 GGTAGCGAATG R * 2930 TATCCGGCCTA	2400 CACTGCGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CATATGAAA N M K 2760 CGCCACTGCT G T A 2850 CGCTGGAGAG CGCTGGAGAG 2940 ACCGGTCGGC >>>##################################	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 CGAAGAGCAC G R A 2860 CGTGCGGGCC 2950 CAACAGTTGT	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 ITATGAGAAAGA 2960 AGGCCTGATAAG	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCC >
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGA\\S Y T V C\\ 2440\\ \hline \\ GACTGAAGTGATTTCTT E V I S\\ 2530\\ \hline \\ CGATCACGTAGTGATGCT\\D H V V I \\ 2620\\ \hline \\ TGACTTCTTCGATGCT\\D F F D Z\\ 2710\\ \hline \\ ACCTACCGCTGAGCT\\P T A E I\\ 2800\\ \hline \\ ACGCGGTCTGGGAAA\\G G L G N\\ 2890SacII\\ \hline \\ TGAGTCGCCGCCGCGGT\\ \end{array}$	2360 CCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 CCCTCGATGATGAT L D D 2630 CCTCTATCAAC S I N 2720 CCCCCAAGCTC C R K L 2810 VTGTATTGCCA V L P 2900 TAAACACTGCC	2370 STTTTACATG F Y M 2460 S A A 2550 GAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GAAGCGGCGC E A A 2820 AACGTCACGA T S R 2910 CCGCATGCGGC	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 TTACGCCAGCA Y A S 2920 CGCGAGCGCCT	2390 CCAGCCGCCAG F S R Q 2480 RGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGCTCT T G A S 2840 GGTAGCGAATG R * 2930 TATCCGGCCTP	2400 CACTGCGCTG T A L 2490 CTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CATATGAAA N M K 2760 CGCACTGCT G T A 2850 CGCTGGAGAG CGCTGGAGAG 2940 ACGGGTCGGC	2410 TGCCTGGACG C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860 CGTGCGGGCC 2950 CAACAGTTGT	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 TTATGAGAAAGA 2960 AGGCCTGATAAG	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCG > <<<<<
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGACSS Y T V C\\ 2440\\ \hline \\ GACTGAAGTGATTTCTT E V I S\\ 2530\\ \hline \\ CGATCACGTAGTGATGCTD H V V I \\ 2620\\ \hline \\ TGACTTCTTCGATGCD F F D Z\\ 2710\\ \hline \\ ACCTACCGCTGAGCTAGTCGCTGAGCCT P T A E I \\ 2800\\ \hline \\ AGGCGGTCTTGGAAAAGCCTAGGCAAAAGC G G L G N\\ 2890SacII \\ \hline \\ TGAGTCGCCGCCGCGGGT\\ \hline \\ 2980\\ \hline \end{array}$	2360 CTCCAATGAC S N E 2450 CGACAAGAT S D K I 2540 CGCTGGATGAT L D D 2630 CTCTATCAAC S I N 2720 CCCCCAAGCTC R K L 2810 CTGTATTGCCJ V L P 2900 CAAACACTGCC >>>>> 2990	2370 F Y M 2460 FCCGCCGCC S A A 2550 GAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GAAGCGCCGC E A A 2820 AACGTCACGA T S R 2910 CGGATGCGGC >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2380 GGGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 VTACGCCAGCA Y A S 2920 CGCGAGCGCCT >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2390 CCAGCCGCCAG F S R Q 2480 RGCCGCAGTTG V P Q L 2570 STGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGCTCI T G A S 2840 GGTAGCGAATC R * 2930 TATCCGGCCTA >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2400 CACTGCGCTG T A L 2490 CTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CATATGAAA N M K 2760 CGCACTGCT G T A 2850 CGCTGGAGAG 2940 ACGGGTCGGC >>######## 3030	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAC G R A 2860 CGTGCGGGCC 2950 CAACAGTTGT	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 TTATGAGAAAGA 2960 AGGCCTGATAAG 	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCG > <<<<< 3060
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGACSCGTTGACSSYTTCCSYTVVC 2440\\ \hline \\ GACTGAAGTGATTACTTEVIS 2530\\ \hline \\ CGATCACGTAGTGATGCTDHVVVI 2620\\ \hline \\ TGACTTCTTCGATGCDFFFDZ 2710\\ \hline \\ ACCTACCGCTGACGTCGCACAAAGGLGGTCGCCGCGGGT 2890 ACAGCGTCGCATCAC$	2360 CTCCAATGAC S N E 2450 CGACAAGAT D K I 2540 CGCTGGATGAT L D D 2630 CTCTATCAAC S I N 2720 CGCGCAAGCTC R K L 2810 TGTATTGCCI V L P 2900 CAACACTGCC >>>>>	2370 STTTTACATG F Y M 2460 TCCGCCGCCC S A A 2550 GAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCG E A A 2820 AACGTCACGA T S R 2910 CGGATGCGGC >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 GCGTGGGTCA A W V 2740 GGCGATTACA G D Y 2830 TACGCCAGCA Y A S 2920 GCGAGCGCCT ->>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2390 CCAGCCGCCAG T S R Q 2480 IGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGGTCT T G A S 2840 GGTAGCGAATG R * 2930 TATCCCGGCCTA >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2400 CACTGCGCTG TAL 2490 CTGCTGCAC LLH 2580 CGTCACGAT RHD 2670 CATATGAAAA NMK 2760 CGCACTGCT GTA 2850 CGCACTGCTG CACGGTCGGAGAG 2940 ACGGGTCGGAGAG 2940 ACGGGTCGGAGAG 2940 ACGGGTCGGAGAG 2940 ACGGGTCGGAGAG	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCACA G R A 2860 CGTGCGGGCC 2950 CAACAGTTGTA #######<<<<	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGGGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 ITATGAGAAAGA 2960 AGGCCTGATAAG 	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCG > <<<<< 3060 AGACG
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGACSSYTCOMSSYTCMSYTVVC \begin{array}{c} 2440\\ \hline \\ GACTGAAGTGATTACTTEVIS \\ 2530\\ \hline \\ CGATCACGTAGTGCTDHVVVI \\ 2620\\ \hline \\ TGACTTCTTCGATGCDFFFDZ \\ 2710\\ \hline \\ ACCTACCGCTGAGCTPTAETIC \\ 2800\\ \hline \\ ACGCGGTCTCGCGAAAGGLGGCCGCCGCGGT \\ \hline \\ 2890 \text{SacII}\\ \hline \\ TGAGTCGCCGCCGCGCTCAC \\ 2980\\ \hline \\ ACAGCGTCGCATCACCACCATCACCACCACCACCACCACCACCACCACC$	2360 CCCCAATGAC S N E 2450 CCGACAAGAT D K I 2540 NGCTGGATGAT L D D 2630 CCTCTATCAAC S I N 2720 NGCGCAAGCTC R K L 2810 NGCGCAAGCTC V L P 2900 TAAACACTGCC SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	2370 STTTTACATG F Y M 2460 TCCGCCGCC S A A 2550 IGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCG E A A 2820 AACGTCACGA T S R 2910 CCGGATGCGGC >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GGCAATTGCCA A I A 2650 GGCGTGGGTCA A W V 2740 GGCGATTACA G D Y 2830 ATACGCCAGCA Y A S 2920 GGCGAGCGCCT >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2390 CCAGCCGCCAG T S R Q 2480 TGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGCGTCT T G A S 2840 GGTAGCGAATC R * 2930 TATCCCGGCCTA >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2400 CACTGCGCTGCACG L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K 2760 CGCACTGCT G T A 2850 CGCTGGAGAG 2940 ACGGGTCGGC >>####### 3030 CTACAGGTCG	2410 TGCCTGGACC C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 GGAAGAGCAGCAG G R A 2860 CCGTGCGGGCC 2950 CAACAGTTGT CAACAGTTGT CAACAGTTGT	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGG L R A L 2780 GAAATCGTTGCC E I V A 2870 TTATGAGAAAGA 2960 AGGCCTGATAAG 3050 GTAGGCCTGATA	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCG > <<<<< 3060 AGACG
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGGC\\S&Y&T&V\\ \end{array}$	2360 CCCCAATGAC CCGACAAGATGAC CCGACAAGATG CCGACAAGATG CCCGACAAGATGAC CCCGACAAGATGAC CCCCCTGATGATGAC CCCCCTATCAAC CCCCCTATCAAC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCAAGCTC CCCCCCAAGCTC CCCCCCAAGCTC CCCCCCCAAGCTC CCCCCCCCCCCACGCC CCCCCCCCCCCCCCCCCCC	2370 STTTACATG F Y M 2460 TCCGCCGCC S A A 2550 TGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 GGAAGCGGCG E A A 2820 AACGTCACGA T S R 2910 CCGGATGCGGC >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2380 GGGTATGCCA G Y A 2470 ATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATGGCTCA A W V 2740 CGCGAGCGCCT 2830 ATACGCCAGCA Y A S 2920 CGCGAGCGCCT >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2390 CCAGCCGCCAG T S R Q 2480 TGCCGCAGTTG V P Q L 2570 STGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGGTCT T G A S 2840 GGTAGCGAATC R * 2930 TATCCGGCCTP >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2400 CACTGCGCTGCACG L L H 2580 CGTCACGAT R H D 2670 CAATATGAAA N M K 2760 CGCACTGCT G T A 2850 CGCACTGCT G T A 2850 CGCTGGAGAGAG 2940 ACGGGTCGGAGAGAG >>>####### 3030 CTACAGGTCG	2410 TGCCTGGACG C L D 2500 GTCAGCCGTQ V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 CGAAGAGCCCTGC G R A 2860 CGTGCGGGCC 2950 CAACAGTTGT CAACAGTTGT CAACAGTTGT CAACAGTTGT CAACAGTTGT CAACAGTTGT	2420 3CCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 TTATGAGAAAGA 2960 AGGCCTGATAAG 	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGCG > <<<<< 3060 AGACG <<<<<
$\begin{array}{c c} 2350\\ \hline \\ GAGCTACACGGTTGG\\S&Y&T&V\\ \end{array}$	2360 CCCCAATGAC S N E 2450 CCGACAAGATT D K I 2540 VCCTGGATGAT L D D 2630 CCTCTATCAAC A S I N 2720 VCCGCAAGCTC R K L 2810 VCTGTATTGCC2 V L P 2900 CAACACTGCC 2990 CCATGATTC	2370 STTTACATG F Y M 2460 TCCGCCGCC S A A 2550 TGAAACCCAG E T Q 2640 CCGCATTGCC R I A 2730 SGAAGCGCCGCG E A A 2820 AACGTCACGA T S R 2910 CGGATGCGGC >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2380 GGTATGCCA G Y A 2470 CATGCTGTATG M L Y 2560 GCAATTGCCA A I A 2650 CGCGTGGGTCA A W V 2740 CGCGATTACA G D Y 2830 TACGCCAGCA Y A S 2920 CGCGAGCGCCT >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	2390 CCAGCCGCCAG F S R Q 2480 RGCCGCAGTTG V P Q L 2570 GTGAGATTGTG S E I V 2660 TTGGTACACGC I G T R 2750 CTGGCGCGGTCT T G A S 2840 GGTAGCGAATG R * 2930 TATCCGGCCTA 3020 CTTATCCGGCC	2400 CACTGCGCTGCAC T A L 2490 CTGCTGCAC L L H 2580 CGTCACGAT R H D 2670 CATATGAAA N M K 2760 CGCCACTGCT G T A 2850 CGCTGGAGAG 2940 ACCGGGTCGGC >>>#######	2410 TGCCTGGACG C L D 2500 GTCAGCCGTC V S R 2590 CTGTTTGACC L F D 2680 AAAGCCCTGC K A L 2770 CGAAGAGCAC G R A 2860 CGTGCGGGCC 2950 CAACAGTTGTI #######<<<	2420 GCCGGGCACTTC A G H F 2510 CCGGTTCGCTGG P V R W 2600 CGGGTGCATATC R V H I 2690 CTGCGTGCGTGCGTTG L R A L 2780 GAAATCGTTGCC E I V A 2870 ITATGAGAAAGA 2960 AGGCCTGATAAG 	2430 CACCC H P 2520 GACAG D S 2610 GGCCT G L 2700 CTGGA L E 2790 GTGGC V A 2880 AATTT 2970 ACGGC > <<<<< 3060 AGACG

function in protection from degradation by 3' exoribonucleases (6).

The *rhaB* transcription start site was mapped by using primer extension analysis. We found in strain ECL116 a major rhamnose-inducible transcription start that mapped to

24 bp upstream of the *rhaB* translational start site (Fig. 4, lanes 1 and 2). Two minor RNA 5' ends were mapped 18 and 19 bp upstream of the *rhaB* translational start site. It is not clear whether these represented independent transcription start sites or degradation products of the larger transcript.

3070 3080 3090 3100 3110 3120 3130 CGACACGCATCAGGCATTGATTGCCGGATGCGGCACAAGTGCCTTATCCCGGCCTACAGGTCGGCAATAGTT	3140 3150 IGTAGGCCTGATAAGACGCG
<<<<<<<< >****	·····
3160 3170 3180 3190 3200 3210 3220 ACACCATCACCATCACCACTACCACACACACACACACAC	3230 3240
<pre></pre>	
3250 3260 3270 3280 3290 3300 3310	3320 3330
AGCATCAGGCATTGATTGCCGGATGCGGCACAAGTGCCTTATCAGGCCTACAGGTCGGCAATAGTTGTAGC <>>>>>>>++++++++++++++++++++++++++++++	JCCTGATAAGACGCGACACG
3340 3350 3360 3370 3380 3390 3400	rhaD 3410 3420
CATCAGGCATTGATTGCCGGATCGGCCGACGCTTATCCGGCCTACGGGTCGTGCATCGACAACACCGAATT	TTAC <u>AGGAA</u> CACAGAACATG
3430 3440 3450 3460 3470 3490 3490	2500 2510
CAAAACATTACTCAGTCCTGGTTTGTCCAGGGAATGATCAAAGCCACCACCGACGCCTGGCTGAAAGGCTC	GGATGAGCGCAACGGCGGC
Q N I T Q S W F V Q G M I K A T T D A W L K G V	N D E R N G G
3520 3530 3540Ecorv 3550 3560 3570 3580 AACCTGACGCTACGCCTGGATGACGCCGATATCGCACCATATCACGACAATTTCCACCAACAACCGCGCT	3590 3600 ATATCCCGCTCAGCCAGCCC
N L T L R L D D A D I A P Y H D N F H Q Q P R Y	YIPLSQP
3610 3620 3630 3640 3650 3660 3670	3680 3690
M P L L A N T P F I V T G S G K F F R N V Q L I	D P A A N L G
3700 sali 3710 3720 3730 3740 3750 3760	3770 3780
ATCGTAAAAGTCGACAGCGACGGCGCGCGCGCGCTACCACATTCTTTGGGGGTTAACCAACGAAGCCGTCCCCAC I V K V D S D G A G Y H I L W G L T N E A V P 7	CTTCCGAACTTCCGGCTCAC F S E L P A H
3790 3800 3810 3820 3830 3840 3850	3860 3870
TTCCTTTCCCACTGCGAGCGCATTAAAGCCACCAACGGCAAAGATCGGGTGATCATGCACTGCCACGCCAC	CCAACCTGATCGCCCTCACC
TATGTACTTGAAAACGACACCGCGGTCTTCACTCGCCAACTGTGGGAAGGCAGCACCGAGTGTCTGGTGGT	3950 3960 TATTCCCGGATGGCGTTGGC
YVLENDTAVFTRQLWEGSTECLVV	VFPDGVG
3970 3980 3990 4000 4010 4020 4030	4040 4050 CCTCCTCCTCCTCCCCTTC
I L P W M V P G T D E I G Q A T A Q E M Q K H S	SLVLWPF
4060 4070Avall 4080 4090 4100 4110 4120	4130 4140
H G V F G S G P T L D E T F G L I D T A E K S A	A Q V L V K V
4150 4160 4170 4180 4190 4200 4210	4220 4230
TATTCGATGGGCGGCATGAAACAGACCATCAGCCGTGAAGAGTTGATAGCGCTCGGCAAGCGTTCCGGCG' Y S M G G M K O T I S R E E I I A I G K R F G V	TTACGCCACTCGCCAGTGCG
	A210 A220
CTGGCGCTGTAAGGAGCAAACATGATCCGCAAAGCCTTTGTCATGCAGGTAAACCCCGACGCCACGAAG	AGTATCAGCGTCGGCATAAT
4330 4340 4350 4360 4370 4380 4390 CCCATCTGGCCAGAACTGGAAGCAGTGCTGAAATCTCACGGTGCGCATAACTACGCCATCTATCT	4400 4410 AAGCGCGTAATCTGCTGTTT
	4490 4500
GULTIGETAGANT IGANTO IGANGANUGUTGGGATIGUUGUTTGUUGUTTGUUGUTTGUUAUGITGGI	Somminini GACCOAIGII
4510 4520 4530 4540 4550 4560 4570	4580 4590
ATGCCCGCTAACCCCGGATAACAGCCCCGGTGAGTAGCGAGCTGCCAAGAAGTGTTTTACCTGCCGTAATTGC	CCAATGIGCCAGAGATGCCT

Neither transcript was detected in strain ECL339, which carries a deletion of the *rha* region (Fig. 4, lane 3). Prior to this sequencing of *rhaBAD*, Tobin and Schleif used S1 mapping to identify three transcription start points (p1, p2, and p3) in the *rhaBAD* region (40). The transcription start point upstream of the *rhaB* open reading frame identified here (pB) is distinct from all of the previously identified start

points. It seems likely that, at least under most conditions, pB defines the major promoter for *rhaBAD* transcription. The p1 and p2 promoters defined by Tobin and Schleif (40) may correspond to the 5' ends of the same breakdown products identified in the Northern blot experiments in this work. The locations of p1 and p2 suggest that *rhaBAD* transcripts may have been processed between the *rhaB* and

4600 4610 4620 4630 4640 4650 4660 4670 4680 СССВСАСААТАТТТТАААТТТССАТТТСАААААТТТАААСАТАТТАСАТСТССССВСТААСТТАТАТТТСССТСАТАСАТТСТСАСДАДААААТС

4690	4700	4710	4720	4730	4740	4750	4760	4770
Асалалатаатдаата	AACCCATTAA	IGATTCATGIN	GGTTTATTTA	ААТААССАТТИ	ATGTGCATTA	CTCCGCAAAT	CTGACCTTTC	ACTCT
4780	4790	4800	4810	4820	4830	4840	4850	4860
GTCTCAAATTGTCAT	IGTTTACCGA	CATATCGGCC	ICTCATCACTI	FAATCGCTCC	ACCATTACAA	TTGATTAAAA	ГАТАТТТТАТ <u>(</u>	<u>3GAG</u> T
URF1 4870 CCGATTAATGGCAGC' M A A	4880 ICTTACTGCA L T A	4890 AGCTGTATTG S C I I	4900 ACCTGAATATI D L N I	4910 ICAGGGCAATO Q G N	4920 GCGCTTATT G A Y	4930 CCGTTCTGAA SVLK	4940 SCAGTTGGCG Q L A	4950 ACAAT T I
4960	4970	4980	4990	5000	5010	5020	5030	5040
Agcgttacaaaacgg	FTTATCACCO	GACTCACACC	AGTTCCTGCAA	AACCCTGCTGC	CTGCGCGAAA	AAATGCACTC	FACTGGATTTO	GGTTC
A L Q N G	FIT	D S H	Q F L Q	T L L	L R E	K M H S	T G F	G S
5050	5060	5070	5080	5090	5100	5110	5120	5130
CGGTGTCGCCGTGCCG	SCACGGTAAA	AGCGCCTGCG	TTAAACAACCO	STTCGTATTA	TTCGCCCGCA	AAGCGCAGGC	TATTGACTGG	AAAGC
G V A V P	H G K	SAC	V K Q P	F V L	F A R	K A Q A	I D W	K A
5140	5150	5160	5170	5180	5190	5200	5210	5220
CAGCGATGGCGAAGA	CGTCAATTGC	TGGATCTGCC	TCGGCGTGCCC	GCAAAGCGGCG	GAAGAGGATC	CAGGTCAAAAT	CATCGGCACA	CTGTG
S D G E D	VNC	W I C	L G V P	Q S G	E E D	Q V K I	I G T	L C
5230	5240	5250	5260	5270	5280	5290	5300	5310
TCGCAAAATTATTCA	CAAGGAATTT	АТТСАТСААС	TGCAACAGGGG	CGATACCGAC	CAGGTGCTTG	CCTTGTTAAA	TCAAACCCTC	AGCTC
R K I I H	K E F	I Н Q	L Q Q G	D T D	Q V L	A L L N	Q T L	S S
5320 ATAAGGAAGTGGCGA *	5330 TGGAGTCATC	5340 CTTACGTATT	5350 GTCGCGATCAG	5360 CCAACTGCCC	5370 GCCGGGATCG	5380 СТСАСАССТА	URF2 5390 CATGGTGGCG M V A	5400 GAAGC E A
5410	5420	5430	5440	5450	5460	5470	5480	5490
CCTGGAACAGAAAGC	CCGTTCTCTC	GGTCATACCA	TAAAAGTGGAJ	AACTCAAGGG	TCCAGTGGCG	TTGAAAACCG	CTTATCCAGC	GAAGA
L E O K A	R S L	G H T	I K V E	T O G	SSG	V E N R	L S S	E E
5500 GATTGCCGCTGCCGA I A A A D	5510 TTACGTCATT Y V I	5520 CTCGCTACCG L A T	5530 GGCGTGGCCTC G R G L	- 5540 GAGCGGTGAT S G D	5550 GATCGCGGGC D R G	5560 CGGTTTGCCGG R F A G	5570 GAAGAAAGTT K K V	5580 TATGA Y E
5590	5600	5610	5620	5630	5640	5650	5660	5670
GATTGCCATCTCCCA	GGCGTTGAAA	AATATCGACC	AGATTTTCAG	CGAATTACCG	ACAAACTCGC	CAGCTTTTTGC	CGCAGATAGC	GGCGT
I A I S Q	A L K	N I D	Q I F S	E L P	T N S	Q L F A	A D S	G V
HindIII GAAGCTT 5677 K L								

FIG. 2—Continued.

the *rhaA* genes. The origin of the p3 transcript (40), which would begin after the 3' end of *rhaD* but with opposite transcriptional polarity, is unknown at this time.

To determine whether L-rhamnose or a metabolite of L-rhamnose was the direct inducer of the rhaBAD operon, we also determined whether L-rhamnose induced rhaB transcription in strains carrying point mutations in *rhaA*, *rhaB*, rhaD, or rhaR (Fig. 4, lanes 4 to 7). Mutations in rhaA are expected to block any catabolism of L-rhamnose, while rhaB mutations prevent formation of L-rhamnulose-1-phosphate and *rhaD* mutations prevent its further catabolism to dihydroxy acetone phosphate and L-lactaldehyde. rhaB transcription was detected in strains carrying mutations in each of the structural genes, indicating that L-rhamnose may be the direct inducer of *rhaBAD* transcription. The lower level of transcription in the *rhaD* mutant strain is likely due to the fact that this strain grew very poorly in the presence of L-rhamnose. This poor growth is presumably a consequence of the accumulation of the phosphorylated intermediate

L-rhamnulose-1-phosphate (1). As expected (12), a point mutation in *rhaR* abolished *rhaBAD* transcription (Fig. 4, lane 7).

Sequence similarity with other proteins. The reported sequences of the L-fucose regulon gene cluster (11, 25) permitted a comparison with the functionally analogous enzymes encoded by the L-rhamnose regulon gene cluster whose sequences are presented here. Rhamnulose kinase was found to have 25% identity with *E. coli* fuculose kinase, 18% identity with *E. coli* xylulose kinase, and 66% identity with the rhamnulose kinase of *S. typhimurium* (31). No other kinase in the EMBL-GenBank data bank showed any homology with the *rhaB* product of *E. coli*, according to the TFASTA program.

In the cases of rhamnose isomerase and fucose isomerase of *E. coli*, homology was lower, with the alignment displaying only 15% identity. Likewise, in the cases of rhamnulose-1-phosphate aldolase and fuculose-1-phosphate aldolase, homology was very low, with an identity of 18%. As for



FIG. 3. Northern blots of mRNA from strain ECL1. RNA was isolated from cells grown on rhamnose to an A_{650} of 0.5 (lanes 1 to 3) or 0.2 (lane 4) and hybridized with the probes shown as thick lines in the upper part. A major transcript of 1.4 kb when the *rhaB* gene probe was used is apparent (lane 1). Two transcripts of 2.5 and 1.6 kb are present with the *rhaA* gene probe (lane 2) and with the *rhaD* gene probe (lane 3). A full-length transcript of 3.8 kb together with other RNA species appears when the probe used encompasses the three *rhaBAD* genes (lane 4).

rhamnulose kinase, no other isomerase or aldolase with significant homology to rhamnose isomerase or rhamnulose-1-phosphate aldolase was found.

In spite of the similarity between the reactions catalyzed by the corresponding enzymes in the parallel metabolic pathways for rhamnose and fucose, homologies between the sequences are rather low. Conservation is more stringent in specific short fragments, which are presumably involved in the active center, for the kinases but not for the isomerases or the aldolases. According to Sander and Schneider (35), the 25% identity between rhamnulose kinase and fuculose kinase is at the lower limit at which structural homology could be inferred, while sequence identities for the corre-

TABLE 2. β-Galactosidase activities in strains containing transcriptional fusions of the *rhaB*, *rhaA*, and *rhaD* genes and grown under different conditions

Ct	β-Galactosidase activity ^a of cells grown on:					
Strain	L-Rhamnose	L-Fucose	Glucose			
JA121 (control)	<100	<100	<100			
JA123 (rhaB-lacZ)	32,600	<100	<100			
JA124 (rhaA-lacZ)	1,330	<100	<100			
JA125 (rhaD-lacZ)	130	<100	<100			
JA126 (rhaBA-lacZ)	24,900	<100	<100			

^a Enzyme activities are given in Miller units (28).

sponding isomerases or aldolases are below the threshold permitting inference of structure homology.

The degree of conservation between the genes of the two systems would be very low if one accepts a divergent evolution for the fucose and rhamnose genetic systems. However, the high homology between rhamnulose kinases of two different species, *E. coli* and *S. typhimurium*, seems to point to a convergent rather than to a divergent evolution.

G+C content and codon usage. The G+C content of rhaB, rhaA, and rhaD (55.3, 56.3, and 56%, respectively) is significantly higher than the approximately 50% G+C content which is the average of the whole genomes of E. coli K-12 and S. typhimurium. This could be interpreted to indicate that these genes were transferred to the enteric bacteria from an ancestor with a genome which was G+C rich, as has been proposed for the A+T-rich rfb region (8, 22). This possible horizontal acquisition of the rhamnose system would also hold true for the rhaB gene of S. typhimurium (31), which has a G+C content of 55%. The 50% G+C content of the fucose genes is interesting in view of the lack of homology between the corresponding proteins of the rhamnose and fucose systems presented above. If the high G+C content is indicative of an ancestral transfer of genes, the differences in G+C content suggest that each system was of a different origin.

The high G+C content of the three genes is reflected in the codon usage. At all codon positions, but particularly at the third one, there is a strong preference for G or C over A or



FIG. 4. Primer extension analysis of *rhaBAD* transcription. Primer extension reactions were performed as described in Materials and Methods. Sequencing reactions were performed by using the same ³²P-labeled oligonucleotide as used for the primer extension reactions (lanes A, C, G, and T). The wild-type strain was grown in the absence or presence of L-rhamnose, while all other strains were grown in the presence of L-rhamnose (see Materials and Methods). Lanes: 1, ECL116 (wild type) without rhamnose; 2, ECL116 with rhamnose; 3, ECL339 [Δ (*rha-pfk*)]; 4, ECL714 (*rhaB101*); 5, ECL715 (*rhaA502*); 6, ECL716 (*rhaD701*); 7, ECL717 (*rhaR702*). The arrow indicates nucleotide position at which major transcription start takes place.

U. The scores for the frequency of optimal codon usage (21) of 0.64 for *rhaB*, 0.61 for *rhaA*, and 0.65 for *rhaD*, close to the 0.59 reported for *rhaT* (39), are significantly lower than the scores of highly expressed proteins from genes such as *ompA* or *lpp*, which have scores of 0.92 and 0.98, respectively (21). These differences in codon usage could indicate that *rhaBAD* encodes proteins which are not highly expressed in *E. coli*. The genes *rhaR* and *rhaS* for the rhamnose regulatory proteins (40) have scores of 0.56 and 0.55, respectively, also corresponding to proteins with low expression levels, as do other regulatory proteins encoded by genes such as *trpR* (0.56) or *araC* (0.54) (21).

ACKNOWLEDGMENTS

This work was supported by grant PB91-437 from the DGICYT of Spain and by NIH grant GM18277 to R. Schleif. P.M. and E.H. were recipients of predoctoral fellowships (FPI) from the Ministerio de Educación y Ciencia of Spain, and S.M.E. was the recipient of NIH postdoctoral fellowship GM14364.

We thank M. Aldea for the gift of bacterial strains and plasmids and for helpful discussion and R. Schleif for critical reading of the manuscript.

ADDENDUM IN PROOF

While this article was being reviewed, we found that a nucleotide was erroneously inserted at position 2747. Deletion of this nucleotide changes the carboxyl-terminal part of rhamnose isomerase and adds 19 amino acids, yielding a molecular mass of 47,231 Da for the complete protein.

REFERENCES

- Al-Zarban, S., L. Hefferman, J. Nishitani, L. Ransone, and G. Wilcox. 1984. Positive control of the L-rhamnose genetic system in Salmonella typhimurium LT2. J. Bacteriol. 158:603-608.
- Bachmann, B. J. 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. 54:130–197.
- 3. Backman, K., Y.-M. Chen, and B. Magasanik. 1981. Physical and genetic characterization of the *glnA-glnG* region of the

Escherichia coli chromosome. Proc. Natl. Acad. Sci. USA 78:3743-3747.

- 4. Badía, J. (University of Barcelona). 1991. Personal communication.
- Badia, J., L. Baldoma, J. Aguilar, and A. Boronat. 1989. Identification of the *rhaA*, *rhaB*, and *rhaD* gene products from *Escherichia coli* K12. FEMS Microbiol. Lett. 65:253–258.
- Belasco, J. G., T. Beatty, C. W. Adams, A. von Gabain, and S. N. Cohen. 1985. Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic *rxcA* transcript. Cell 40:171– 181.
- Boronat, A., and J. Aguilar. 1979. Rhamnose-induced propanediol oxidoreductase in *Escherichia coli*: purification, properties, and comparison with the fucose-induced enzyme. J. Bacteriol. 140:320–326.
- Brown, P. K., L. K. Romana, and P. R. Reeves. 1992. Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. 6:1385–1394.
- Chen, C.-Y. A., J. T. Beatty, S. N. Cohen, and J. G. Belasco. 1988. An intercistronic stem-loop structure functions as an mRNA decay terminator necessary but insufficient for puf mRNA stability. Cell 52:609-619.
- Chen, Y.-M., and E. C. C. Lin. 1984. Dual control of a common L-1,2-propanediol oxidoreductase by L-fucose and L-rhamnose in *Escherichia coli*. J. Bacteriol. 157:828–832.
- Chen, Y.-M., Z. Lu, and E. C. C. Lin. 1989. Constitutive activation of the *fucAO* operon and silencing of the divergently transcribed *fucPIK* operon by an IS5 element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. J. Bacteriol. 171:6097-6105.
- Chen, Y.-M., J. F. Tobin, Y. Zhu, R. F. Schleif, and E. C. C. Lin. 1987. Cross-induction of the L-fucose system by L-rhamnose in *Escherichia coli*. J. Bacteriol. 169:3712–3719.
- Chen, Y.-M., Y. Zhu, and E. C. C. Lin. 1987. The organization of the *fuc* regulon specifying L-fucose dissimilation in *Escherichia coli* K12 as determined by gene cloning. Mol. Gen. Genet. 210:331-337.
- d'Aubenton Carafa, Y., E. Brody, and C. Thermes. 1990. Prediction of Rho-independent *Escherichia coli* transcription terminators. J. Mol. Biol. 216:835–858.
- Dimri, G. P., K. E. Rudd, M. K. Morgan, H. Bayat, and G. F.-L. Ames. 1992. Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. J. Bacteriol. 174:4583–4593.
- 16. Galloway, J. L., and T. Platt. 1988. Signals sufficient for rho-dependent transcription termination at *trp t'* span a region centered 60 base pairs upstream of the earliest 3' end point. J. Biol. Chem. 263:1761-1767.
- García-Martín, C., L. Baldomá, J. Badía, and J. Aguilar. 1992. Nucleotide sequence of the *rhaR-sodA* interval specifying *rhaT* in *Escherichia coli*. J. Gen. Microbiol. 138:1109–1116.
- Gilson, E., W. Saurin, D. Perrin; S. Bachellier, and M. Hofnung. 1991. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. 19:1375–1383.
- Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung. 1991. The BIME family of bacterial highly repetitive sequences. Res. Microbiol. 142:217-222.
- Hidalgo, E., Y. M. Chen, E. C. C. Lin, and J. Aguilar. 1991. Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 ald gene encoding aldehyde dehydrogenase. J. Bacteriol. 173:6118-6123.
- Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol. 151:389–409.
- Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves. 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. 5:695-713.

- Lin, E. C. C., and T. T. Wu. 1984. Functional divergence of the L-fucose system in mutants of *Escherichia coli*, p. 135–163. *In* R. P. Mortlock (ed.), Microorganisms as model systems for studying evolution. Plenum, New York.
- Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall. 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193:265-275.
- Lu, Z., and E. C. C. Lin. 1989. The nucleotide sequence of Escherichia coli genes for L-fucose dissimilation. Nucleic Acids Res. 17:4883–4884.
- Lupski, J. R., and G. M. Weinstock. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. J. Bacteriol. 174:4525–4529.
- McCormick, J. R., J. M. Zengel, and L. Lindahl. 1991. Intermediates in the degradation of mRNA from the lactose operon of *Escherichia coli*. Nucleic Acids Res. 19:2767–2776.
- Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Morgan, W. D., D. G. Bear, B. L. Litchman, and P. H. von Hippel. 1985. A sequence and secondary structure requirements for rho-dependent transcription termination. Nucleic Acids Res. 13:3739–3754.
- Newbury, S. F., N. H. Smith, and C. F. Higgins. 1987. Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell 51:1131–1143.
- Nishitani, J., and G. Wilcox. 1991. Cloning and characterization of the L-rhamnose regulon in Salmonella typhimurium LT2. Gene 105:37-42.
- Power, J. 1967. The L-rhamnose genetic system in *Escherichia coli* K12. Genetics 55:557–568.
- Reeder, T., and R. Schleif. 1993. AraC protein can activate transcription from only one position and when pointed in only one direction. J. Mol. Biol. 231:205–218.

- 34. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 35. Sander, C., and R. Schneider. 1991. Database of homologyderived protein structures and the structural meaning of sequence alignment. Proteins 9:56–68.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 174:5463-5467.
- Simons, R. W., F. Houman, and N. Kleckner. 1987. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. Gene 53:85–96.
- Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins. 1984. Repetitive extragenic palindromic (REP) sequences: a major component of the bacterial genome. Cell 37:1015-1026.
- 39. Tate, C. G., J. A. R. Muiry, and P. J. F. Henderson. 1992. Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H⁺ transport protein in *Salmonella typhimurium* and *Escherichia coli*. J. Biol. Chem. 267:6923-6932.
- Tobin, J. F., and R. F. Schleif. 1987. Positive regulation of the Escherichia coli L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J. Mol. Biol. 196:789– 799.
- 41. Tobin, J. F., and R. F. Schleif. 1990. Purification and properties of *rhaR*, the positive regulator of the L-rhamnose operons of *Escherichia coli*. J. Mol. Biol. 211:75–89.
- Ziemke, P., and J. E. G. McCarthy. 1992. The control of mRNA stability in *Escherichia coli*: manipulation of the degradation pathway of the polycistronic *atp* mRNA. Biochim. Biophys. Acta 1130:297-306.