# Characterization of the spoIVB and recN Loci of Bacillus subtilis<sup>†</sup>

BARBARA E. VAN HOY AND JAMES A. HOCH\*

Division of Cellular Biology, Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037

Received 16 October 1989/Accepted 29 November 1989

Two independent genes, *recN* and *spoIVB*, along with their respective promoter and termination regions, were discovered and sequenced in the 3.4-kilobase region between the *ahrC* and *spo0A* genes at map position 216 in the *Bacillus subtilis* chromosome map. The gene encoding a 576-amino-acid protein, which maintains a high homology with the *Escherichia coli recN* gene product, was adjacent to *ahrC*. The sequence revealed a 64,472-dalton polypeptide which contained a conserved ATP-binding site and possible *lexA*-type regulatory binding sequences in its promoter region. A second open reading frame identified as the *spoIVB* gene was directly downstream of *recN*. It consisted of 1,275 nucleotides which coded for a 425-amino-acid polypeptide with a molecular weight of 45,976. Phenotypic, genetic, and transcriptional analyses confirmed that this gene was *spoIVB*. Although no chloroform-resistant spores were produced by *spoIVB*-inactivated strains, under microscopic examination, phase-gray forespores were visible. The *spoIVB*165 mutation was localized to a 200-base-pair region in the amino-terminal portion of the polypeptide. *spoIVB* was not transcribed until hour 2 of sporulation in wild-type *B*. *subtilis* cells, as determined by  $\beta$ -galactosidase activity assays from *lacZ* transcriptional fusion constructions. We found no amino acid sequence homology between the *spoIVB* gene product and other known bacterial proteins.

Sporulation in Bacillus subtilis may be interrupted genetically by mutations in a large number of genes (10, 12). These genes are defined phenotypically by the stage of sporulation at which they stop when mutationally inactivated. Although many sporulation genes have been defined and mapped, only a few have been assigned an enzymatic or regulatory function. In the case of those genes in which a mutation gives rise to a stage 0 phenotype, we now know that two of the genes, spo0F and spo0A, code for proteins with homology to the regulator components of two-component regulatory systems (7, 19, 23). It is generally agreed that the regulator component of these systems is phosphorylated by the sensor component, which acts as a kinase after stimulation by its specific effector molecule (15). These two components are usually the products of linked, coregulated chromosomal genes. The spo0A and spo0F genes differed from this pattern in that no sensor component gene was found genetically linked to either locus. Sequencing for a substantial distance upstream and downstream of the spo0F gene did not reveal a possible kinase gene for Spo0F (23).

Very little of the sequence surrounding the spo0A locus has been reported (7). It seemed possible that a gene for a sensor component acting on the Spo0A protein could be located in the unsequenced region surrounding it. We report here the sequence upstream of the spo0A gene that contains a possible *recN* gene and the *spoIVB* locus.

## MATERIALS AND METHODS

**Bacterial strains, growth conditions, and transformation.** *B.* subtilis strains used in this study are shown in Table 1. Selection for Erm<sup>r</sup> was as previously reported by Youngman et al. (25). *B. subtilis* strains were transformed by the method of Anagnostopoulos and Spizizen (1). Plasmid DNA (1 to 2  $\mu$ g) was used in each transformation. Cm<sup>r</sup> selection was on Schaeffer medium (22) supplemented with 5  $\mu$ g of chloram-phenicol per ml.

Sporulation efficiency was assayed by growing the strains in 5 ml of Schaeffer medium at  $37^{\circ}$ C for 24 h. The 5-ml culture was exposed to 1 ml of CHCl<sub>3</sub> for 30 min to kill any vegetative cells present (9). Serial dilutions were then plated onto Schaeffer plates, and the colonies were counted as an indication of the number of viable spores present in the original 5-ml culture.

*Escherichia coli* DH5 $\alpha$  competent cells (Bethesda Research Laboratories, Inc.) were used for plasmid construction and propagation. *E. coli* cultures were grown in Luria-Bertani medium supplemented with 100  $\mu$ g of ampicillin per ml.

**DNA manipulations.** Plasmid DNA was prepared by the method of Birnboim and Doly (2). Rapid plasmid DNA preparation was accomplished by the method of Holmes and Quigley (11). The plasmids used in this study were pJH1408 (7) and pJM102 and pJM103 (M. Perego and J. A. Hoch, unpublished data); the last two were used as vectors in the cloning and sequencing of recN, spoIVB, and pJM783, an integrative lacZ transcriptional fusion vector (Perego and Hoch, unpublished data).

Exonuclease III digestion of plasmid pJB2001 was carried out as described by the manufacturer of the enzyme, Boehringer Mannheim Biochemicals. Plasmid pJB2001 was cut with *SmaI* and *SstI* (two sites unique to the multiple cloning region of plasmid pJB2001), which resulted in unidirectional digestion of the inserted fragment from the *SmaI* site. Digestion of plasmid pJB2001 by Exonuclease III for 90 s or 2 min at 37°C produced plasmids pJB2018 and pJB2019, respectively.

Sequence analysis. The sequence analysis of both strands was accomplished by the supercoil sequencing method of Chen and Seeburg (3). The dideoxy chain termination and elongation reactions were performed by the method of Sanger et al. (20a) with Sequenase (United States Biochemical Corp.) or T7 polymerase (Pharmacia LKB) in accordance with the appropriate instructions. The M13 sequenc-

<sup>\*</sup> Corresponding author.

<sup>&</sup>lt;sup>†</sup> Publication no. 6014-MEM from the Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic.

TABLE 1. B. subtilis strains used in this study

Strain	Genotype		
JH642	trpC2 phe-1		
JH12719	trpC2 phe-1 spoIVB (ermG)		
JH12720	trpC2 phe-1 spoIVB (ermG)		
SL765 <sup>a</sup>	trpC2 spoIVB165		
165.1 <sup><i>a</i></sup>	met thr spoIVB165		
JH12713	trpC2 phe-1::pJB2026		

<sup>a</sup> Obtained from P. Piggot.

ing and reverse-sequencing primers from New England BioLabs, Inc., were used, as were the oligonucleotides supplied by the Scripps Clinic and Research Foundation Core Lab, *spo0A* 5'-CTTGCTACATGTTTACA-3' and *spo1VB* 5'-GCGGTTTGCATAAACCT-3'. The *spo1VB* primer was also used in the determination of the transcriptional initiation start site.

**Transcriptional initiation site determination.** The mRNA isolation and primer extension experiments were accomplished as described previously by Perego et al. (17) with a few modifications. mRNA from strain JH642 was isolated at four different times during sporulation,  $T_0$ ,  $T_1$ ,  $T_2$ , and  $T_3$ .

The primer extension analysis utilized the *spoIVB* number 1 17-mer primer for the reverse transcriptase reaction. Approximately 80 µg of RNA isolated at  $T_0$ ,  $T_1$ ,  $T_2$ , and  $T_3$ was used in each reverse transcription reaction. Reverse transcriptase (0.5 U/µl; Life Sciences, Inc.) was added to each reaction, and reaction mixtures were incubated at 37°C for 1 h. The reverse transcripts were then run on a 6% (24:1 cross-link) polyacrylamide–8 M urea sequencing gel next to the sequence of the promoter region to determine the actual start site of transcription. The sequencing reactions were carried out with the same primer as was used in the reverse transcription reaction.

 $\beta$ -Galactosidase assay. The *B. subtilis* strains carrying the integrated *spoIVB* promoter-*lacZ* fusion were assayed for their  $\beta$ -galactosidase activities as described previously (6). Activity was measured in Miller units (14).

# RESULTS

Sequencing studies. The original Charon 4A bacteriophage containing the spo0A locus consisted of several EcoRI fragments, of which a 5.3-kilobase fragment transformed for spo0A, ahrC, and strC mutations (7). This fragment has recently been shown to be 6.2 kilobases, and a section of it encoding the ahrC gene has been sequenced (16). We determined the complete sequence of the region between the ahrC and spo0A genes of the 6.2-kilobase EcoRI fragment in our search for kinase genes with specificity for Spo0A. The sequencing strategy is shown in Fig. 1, and the complete sequence is shown in Fig. 2. The amino terminus of recN agreed with that of North et al. (16) and was not sequenced on both strands. Two complete open reading frames could be found in the region between the ahrC and spo0A genes.

**Characterization of** *recN* gene. An open reading frame lying just downstream of the *ahrC* gene was identified as a possible *recN* gene by a computer homology search. The putative *recN* gene encoded a 576-amino-acid polypeptide with a deduced molecular weight of 64,472. Similarities between the entire sequences of the potentially identical *E. coli* and *B. subtilis* proteins were consistent with the hypothesis that these proteins are highly related and probably identical (Fig. 3). The greatest conservation of amino acids observed between the two proteins occured at the aminoterminal portion of the protein and in a region towards the carboxy terminus. The amino terminus of this protein contained a potential ATP-binding site, as expected for a RecN protein (Fig. 3).

The promoter regions of the *E. coli recN*, *lexA*, and *recA* genes possess consensus *lexA* box sequences that serve as regulatory binding sites for the LexA protein (20). We observed what could be similar *lexA*-type consensus sequences in the upstream regions of the *B. subtilis recN* gene (Table 2).

**Characterization of** *spoIVB* gene. An open reading frame located downstream of the recN gene consisted of 1,275 nucleotides encoding a 425-amino-acid polypeptide with a



FIG. 1. Restriction map, plasmids, and sequencing strategy for the *spoIVB* locus. Arrows at the top of the figure show the extent and direction of each sequencing reaction used to obtain the final sequence. The inserts in the various plasmids described in the text are shown. The approximate locations of the identified genes are indicated with boldface arrows.

	30				^	~~~~ ~~~	~~ ^	^	^^^^					120
GGGGACCATTTGCGGAGATGATACGA	TTTAATT	ATTTGCCG	GACTCCTG	AAGATAC	AGAAGGCGT	AAAAACAGGO	CTCCTTGAA	CTGCTG	TAACAG		CTAAC	AAAGA	TAA <u>GAU</u>	<u>GGT</u>
GTICGDDT	I L I	ICR	TP	EDI	180 E G V	KNK		210	- ei	u <u>ann</u>	<u> </u>			240
GTAATGCTGTTTGTTAGCTGAACTAT	CGATTAAA	AACTTTGC	CATTATTG	AGGAACT	GACGGTTTC	TTTGAACGCO	GGGCTTACG	GTTTTA	ACAGGA	GAAAC	CGGAC	SCCGGC	AAATCI	CAT
recN M L A E L	SIK	NFA	II	EEL	. T V S	FER	GLT	V_L	Τ <u>G</u>	ΕI	G	<u>A</u> <u>G</u>	<u>k</u> s	I
	270			TTOTTAC	300			066 070000	TTTCTG	TTGGA	GAGCO	GCCAT	: [10100	
	V C C		SF		Y G E		ELE	GL	FL	LE	S	GH	ΡV	Ľ
1 1 0 4 1 3 2 2	390				420			450						480
AGGCGTATGCGCTGAGCAGGGAATCG	ACGTTTCA	GACGAAAT	GATCGTCA	TGAGAAG	GGATATCAG	ACGAGCGGG	AAAAGCGTT	TGCCGT	GTCAAT	GGCAA	GCTTO	STCACG	ATTGC	
GVCAEQGI		DEM	IV	MRH	540 I S	156	K S V	570	V M	U N		• •		600
ACTAAGAGAGATCGGCAGGCTTTTAT	TAGATATT	CACGGGCA	GCACGATA	ACCAGCI	GTTAATGGA	GACGAAAAT	CATCTCCAG	стттт	GACAAG	TTTGC	CGGGG	GCCGAA	GTGGA	AAG
LREIGRLL	LDI	HGQ	H D	NQL	L M E	DEN	HLQ	LL	DK	FA	G	A E	V E	S
-	630				660			690	CATTOT	CTTC	TOTT		TTTCA	720 601
CGCGCTCAAAACGTATCAGGAAGGGT	ACCAGCGG V O P	V V K		K K I	KOI	SES	F Q E	M A	H C	LD	L	IQ	FQ	Ľ
ALKIIYEG	750				780	••••		810						840
CGAAGAAATCGAATCGGCAAAGCTTG	AGCTTAAC	GAAGATGA	ACAGCTCC	AAGAAG	GCGCCAGCA	ATTTCAAAC	TTTGAGAAG	AȚATĂT	GAGTCG	CTGCA	AAATO	GCATAT	AACGC	GCT
EEIESAKL		EDE	QL	Q E E	: R 4 4	1 S N	r e K	1030	E 3			A 1	" "	960
TCGCAGTGAGCAGGGGGGGATTGGACT	GGGTTGGA	ATGGCATC	AGCGCAGC	TTGAAG	TATATCTGA	ATAAACGAA	CCGCTGAAA	AAÁATG	TCTGAA	AGTGT	TTCG/	ACTCC	TACTA	TTT
RSEQGLD	W V G	MAS	A Q	LEC	) I S D	INE	PLK	KM	SE	S V	S	N S	YY	L
	990			******				1050	CAAATC		CCTC		ן 141444	080
GCTTGAGGATGCCACATTTCAGATGA	GAAA IA IG D N M	I D F	GUIIGAAI	FDE	F P I		E T R		EI	KG		KR	KY	G
LEDATIER	<b>î</b> 110				1140			1170			-		1	200
TGCCACTGTTGAAGACATTTTGGAAT	ATGCTTCT	AAAATTGA	AGAAGAGA	TTGACC	MATTGAAAA	CAGGGACAGC	CACCTTCAA	AGCTTO	<b>AAAAA</b> G	GAGCI	TGAC	ICCGTC	GGCAA	AGA
ATVEDILE	Y A S	KIE	ΕE	IDQ	1 I E N	RDS	нгч	S L 1200	ĸĸ	ει		5 V	<u>ы к</u> 1	320
TOTOCOAGTOGAAGCOGCAAATGTGT	CACAAATC	CGAAAAAC	ATGGGCGA	AAAAGCI	GGCTGATGA	ATTCATCGG	GAATTAAAA	AGCCTG	TACATG	GAAAA	ATCA	ACTITI	GATAC	GGA
V A V E A A N V	S Q I	RKT	W A	ККІ	. A D E	IHR	ELK	S L	Y M	EK	S	ΤF	DI	E
	1350		****	~~~~~				1410	AACTT	TTAA1	TTCA	TAAT	4[  00808	44U CGA
ATTTAAGGTCAGAACAGCGTCACGCA	ATGAAGAG N E E	GCGCCGCI	IGIAAACG V N	GCCAGCI C 0 1		T F Q	GID	LV	KF	1	S	TN	TG	E
FKVRIAJR	1470	A F L			1500		• • •	1530		• •	•	• ••	1	56Ō
GCCGCTGAAATCCCTTTCGAAAGTCG	CTTCAGGT	GGAGAGCT	GTCACGGG	TGATGC	TGCCATAAA	AGCATTTTT	TCCTCACA	CAGGAT	GTGACA	TCGAT	CATT	TTGAC	GAGGT	TGA
PLKSLSKV	A S G	GEL	SR	VMI		SIF	s s q	Q D	V I	5 1	1	FU	E V	U 183
TACCCCTCTAACCCCCACCGTCCCA	ACCCAATT		GATCCATA	AGGTGTG	CAATAGGATC		TGCATTAC	CACCTO	CCTCAG	GTGGC	TGCC	ATGGCC	GATAC	GCA
T G V S G R V A	Q A I	AEK	IH	KV	SIGS	Q V L	CIT	HL	PQ	V	A	MA	DT	H
														~~~
	1710				1740			1770					1	800
TCTTTATATTGCGAAGGAGTTAAAAG	1710 ACGGCAGA	AÇGAÇGAÇ	ACGTGTCA		1740 ICTCTAAGCA		GCGGAAATO	1770 GAGCGC	AGTATT	GCAG	TGTC	GAAGTA	ACCGA	800 CTT
TCTTTATATTGCGAAGGAGTTAAAAG L Y I A K E L K	1710 Acggcaga D g r 1830	ACGACGAC T T T	ACGTGTCA R V	AGCCGCI K P I	1740 ICTCTAAGCA L S K Q 1860	GGAAAAGGTA E K V	GCGGAAATO A E I	1770 GAGCGC E R 1890	CAGTATT S I	GCAGO	TGTC V	GAAGTA E V	ACCGA T D 1'	800 CTT L 920
TCTTTATATTGCGAAGGAGTTAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTI	1710 ACGGCAGA D G R 1830 TAAAACAA	ACGACGAC T T T	ACGTGTCA R V AGTCAAAA	AGCCGCI K P I	1740 ICTCTAAGCA L S K Q 1860 GGTAAGCTGC	GGAAAAGGTA E K V GCGAGAAGCG	GCGGAAATO A E I CAGCTTATI	1770 GAGCGO E R 1890 TTTTTC	CAGTATT S I CGTGCAC	GCAGO A G	TGTC V	GAAGTA E V TTCATC	ACCGA T D 1' Agtat	800 CTT L 920 ATC
TCTTTATATTGCGAAGGAGTTAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTI T K R H A K E L	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q	ACGACGAC T T T IGCGGATCA A D Q	ACGTGTCA R V AGTCAAAA V K	AGCCGCT K P I Caactgi T T I	1740 ICTCTAAGCA L S K Q 1860 GGTAAGCTGC G >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	GGAAAAGGTA E K V GCGAGAAGCG >> <<<	GCGGAAATO A E I CAGCTTATI	1770 GAGCGC E R 1890 TTTTTC	CAGTATT S I CGTGCAC	GCAGO A G	ITGTCI V ITTCG'	GAAGTA E V TTCATC P2	ACCGA T D 1 Agtat	800 CTT 920 ATC
TCTTTATATTGCGAAGGAGTTAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTI T K R H A K E L	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950	ACGACGAC T T T AGCGGATCA A D Q	ACGTGTCA R V AGTCAAAA V K	AGCCGC K P I CAACTG T T I	1740 ICTCTAAGCA L S K Q 1860 Ggtaagctgc G >>>>>>> 1980 Catgtagcagg	GGAAAAGGTA E K V GCGAGAAGCG >> <<<	GCGGAAAT( A E I CAGCTTATI <<<<<<	1770 GAGCGC E R 1890 TTTTTC 2010	CAGTATT S I CGTGCAC	GCAGO A C	TGTC V TTCG	GAAGTA E V Itcatc P2 Aagtag	ACCGA T D 1' AGTAT 2 TGAAT	800 CTT 920 ATC 040 GCC
TCTTTATATTGCGAAGGAGTTAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTI T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAAATAAG	ACGACGAC T T T AGCGGATCA A D Q GCCGTCAGA	ACGTGTCA R V AGTCAAAA V K AGGCAAAA	AGCCGC K P L CAACTG T T L TTAAAT	1740 ICTCTAAGCA L S K Q I860 Ggtaagctgc G >>>>>> 1980 Gatgtagcag	GGAAAAGGTAI E K V SCGAGAAAGCG >> <<< CAAGTCATAA	GCGGAAAT( A E I CAGCTTATI <<<<<< AGAAGGTGI	1770 GAGCGC E R 1890 TTTTTC 2010 GGGATA	CAGTATT S I CGTGCAC	GCAGO A O ATCCA	TGTC V TTCG MGTG	GAAGTA E V ITCATC P2 AAGTAG <u>spoi</u>	ACCGA T D 1 <sup>4</sup> AGTAT 2 TGAAT <u>VB</u> M	800 CTT 920 ATC 040 GCC P
TCTTTATATTGCGAAGGAGTTAAAAG LYIAKELK AACGAAACGCCATGCGAAAGAACTTI TKRHAKEL CAATGTTTTCTTCATATGACAGTTA P1	1710 ACGGCAGA D G R 1830 TAMACAA L K Q 1950 TAMATAAG 2070	ACGACGAC T T T AGCGGATCA A D Q GCCGTCAGA	ACGTGTCA R V AGTCAAAA V K AGGCAAAA	AGCCGC K P I CAACTG T T I TTAAAT	1740 ICTCTAAGCA ISGO GGTAAGCTGC G>>>>>> I980 GATGTAGCAG DLIgO	GGAAAAGGTAI E K V GCGAGAAGCG >> <<< CAAGTCATAA	GCGGAAATO A E I CAGCTTATT <<<<<< AGAAGGTGT	1770 GAGCGC E R 1890 TTTTTC 2010 GGGATA 2130	CAGTATT SI CGTGCAC	GCAGO A O ATCCA	ITGTC( V ITTCG ATTCG	GAAGTA E V ITCATC P2 AAGTAG <u>spoi</u>	ACCGA T D 1' AGTAT 2 TGAAT <u>VB</u> M 2	800 CTT 920 ATC 040 GCC P 160
TCTTTATATTTGCGAAGGAGTTAAAAG LYIAKELK AACGAAACGCCATGCGAAAGAACTTT TKRHAKEL CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAMATAAG 2070 TAMTTCTC	ACGACGAC T T T AGCGGATCA A D Q ACCGTCAGA	ACGTGTCA R V AGTCAAAA V K AGGCAAAA	AGCCGCT K P I CAACTGG T T Q TTAAATG	1740 ICTCTAAGCA LSK 1860 GGTAAGCTGC S>>>>>> 1980 GATGTAGCAG DLIGO GTTTATGCAA	GGAAAAGGTA EKV SCGAGAAAGCG >> <<< CAAGTCATAA	GCGGAAATC A E I CAGCTTAT1 <<<<<< AGAAGGTG1 GAATATTT/	1770 GAGCGC E R 1890 TTTTTC 2010 GGGATA 2130 CTGAT1	CAGTATT SI CGTGCAC AGGAGCG	GCAGC A C ATCCA	TGTCI V TTCG ATTCG	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT	1; ACCGAA T D 1' AGTAT 2 TGAAT <u>VB</u> M 2 GAAAC E T	800 CTT 920 ATC 040 GCC P 160 CCA
TCTTTATATTGCGAAGGAGTTAAAAA LYIAKELK AACGAAACGCCATGCGAAAGAACTTT TKRHAKEL CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT DNIRKAVG	1710 ACGGCAGA D G R 1830 TAMACAA L K Q 1950 TAMATAAG 2070 TAMTTCTC L I L 2190	ACGACGAC T T T AGCGGATCA A D Q GCCGTCAGA CCTTGTTTC L V S	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA	AGCCGCT K P J CAACTG T T J TTAAAT GTGTAG	1740 ICTCTAAGCA LSKQ B60 GGTAAGCTGC GSTSAGCAG GATGTAGCAG GIIGO GIIGO GIIGO GIICATGCAA GLCK 2220	GGAAAAGGTA E K V Scgagaagcgg >> <<< Caagtcataa Accgctaaaa P L K	GCGGAAATC A E I CAGCTTAT1 <<<<<< AGAAGGTG1 GAATATTT/ E Y L	1770 GAGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CTGATI L I 2250	CAGTATT S I CGTGCAC AGGAGC <u>G</u> ICCCAACG P T	GCAGG A G ATCC/ AGGA( GCAAAT Q P	itgtci V Ittcg I <u>ia</u> gtgi I R	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F	1 ACCGA T D 1 AGTAT TGAAT <u>VB</u> M 2 GAAAC E T 2	800 CTT 920 ATC 040 GCC P 160 CCA 280
TCTTTATATTGCGAAGGAGTTAAAAA L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT	1710 ACGGCAGA D G R 1830 TAMACAA L K Q 1950 TAMATAAG 2070 TAATTCTC L I L 2190 CGGTAMAT	ACGACGAC T T T IGCGGATCA A D Q GCCGTCAGA CTTGTTTC L V S IGCTCAGAC	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA L L XATCAGAAT	AGCCGCT K P I CAACTGC T T ( TTAAAT( GTGTAGG S V ( CCTCAG	1740 ICTCTAAGCA LSKQ BGTAAGCTGC GGTAAGCTGC GATGTAGCAG Digo Digo GTTTATGCAA GLCK 2220 AAGCGTTTAC	GGAAAAGGTA E K V SCGAGAAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA	GCGGAAATC A E I CAGCTTATI <<<<<< AGAAGGTGI GAATATTT/ E Y L GATCCGCAI	1770 GAGCGC E R 1890 TTTTTC GGGAT/ 2130 CTGATI L I 2250 GAAATC	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CCAACGG	GCAGG A G ATCCA AGGA( GCAAAT Q P GACGGG	ITGTCI V ITTCG I AGTGI GAGAI I R I R	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F	ACCGA T D 1' AGTAT ZGAAT VB M 2 GAAAC E T 2 GGTGA	800 CTT 920 ATC 040 GCC 920 CCA 280 GTC
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT T Q A I E T S L	1710 ACGGCAGA D G R 1830 TAMACAA L K Q 1950 TAMATAAG 2070 TAATTCTC L I L 2190 CGGTAAAT S V N	ACGACGAC T T T IGCGGATCA A D Q GCCGTCAGA CTTGTTTC L V S IGCTCAGAC A Q T	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA L L ATCAGAAT S E	AGCCGCT K P I CAACTGG T T I TTAAATG GTGTAGG S V I CCTCAGG S S I	1740 ICTCTAAGCA L S K Q 1860 GATAAGCTGC G >>>>>> GATGTAGCAG GATGTAGCAG G L C K AGCGTTTAC E A F T	GGAAAAGGTAA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA V K K	GCGGAAATC A E I CAGCTTATI <<<<<< AGAAGGTGI GAATATTT/ E Y L GATCCGCAI D P H	1770 GAGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CTGATI L I 2250 GAAAT( E I 2270	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CAAGGTG KV	GCAGC A C ATCC/ AGGAC CAAAT Q P ACGGC T (	ITGTCI V TTCG I ATTCG I AGTGA GAGAA I R SCAAA S K	GAAGTA E V ITCATC <sup>P</sup> 2 MAGTAG <u>SDOI</u> GTATTT V F MAATCA K S	ACCGA T D 1 AGTAT CGAAT TGAAT TGAAT C GAAAC E T 2 GGTGA G E	800 CTT 920 ATC 040 GCC P 160 CCA 280 GTC 280 GTC
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT T Q A I E T S L	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAMATAAG 2070 TAMTTCTC L I L 2190 CGGTAMAT S V N 2310	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA CCTTGTTTC L V S IGCTCAGAC A Q T	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA L L ATCAGAAT S E	AGCCGCT K P I CAACTGU T T C TTAAATU GTGTAGG S V C CCTCAGL S S I	1740 ICTCTAAGCA L S K Q 1860 GATAAGCTGC G >>>>>> D1980 GATGTAGCAG D1190 GATGTAGCAG G L C K AGCGTTTAC E A F T 2340 UTCCTGATTT	GGAAAAGGTA EKV GCGAGAAAGCG >> <<< CAAGTCATAA PLK AGTAAAGAAA VKK	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H	1770 GAGCGC E R 1890 111110 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC E I 2370 CAATC	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CAAGGTG KV	GCAGC A C A C A C A C A C A C A C A C G A A C G A A A C A C	TGTCI V TTCG ATTCG AGTG GAGAI R GAGAA CAAA	GAAGTA E V ITCATC P2 MAGTAG <u>SPOI</u> GTATTT V F MAATCA K S CATTCC	1 ACCGA T D 1 AGTAT 2 TGAAT VB M 2 GAAAC E T 2 GGTGA GGTGA G E 2 GTCGG	800 CTT 920 ATC 040 GCC P 160 CCA 280 GTC S 400 TGT
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAMTTAAG 2070 TAMTTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCTTGTTTC L V S IGCTCAGAC A Q T IAAAAAAAC K K T	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA C L L ATCAGAAT S E AAAAGTGC K V	AGCCGCT K P L CAACTGC T T C TTAAATC GTGTAGG S V C CCTCAGG S S L ATGTTC H V	1740 (CTCTAAGCA L S K Q 1860 GATAAGCTGC G >>>>>> GATGTAGCAG GATGTAGCAG G L C K 2220 MGCGTTTAC E A F T 2340 TTCCTGATTT, L P D L	GGAAAAGGTA EKV GCGAGAAGCG >> <<< CAAGTCATAA PLK AGTAAAGAAA VKK AAAAGTTATAA KVI	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGG/ P G G	1770 GAGCGC E R 1890 111110 2010 GGGAT/ 2130 CTGATI L I 2250 GAAAT(C E I 2370 CAATC/ Q S	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CAAGGTG KV AATCGGT IG	GCAGC A C A C A C A GCAAAT Q P A C A C GTAA/ V P	TGTCI V TTCG ATTCG AGTGI GAGAI I R GCAAA S K VACTTI	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S	1: ACCGA T D 1: AGTAT 2: TGAAT VB M 2 GAAAC E T 2 GGTGA G E 2 GGTCGG V G	800 CTT L 920 ATC 040 GCC P 160 CCA Q 280 GTC S 400 TGT V
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAMTTAAG 2070 TAMTTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA C L V S IGCTCAGAC A Q T IGAAAAAAC K K T		AGCCGCT K P I CAACTGC T T C TTAAATC GTGTAGG S V I CCTCAGG S S I ATGTTC H V I	1740 ICTCTAAGCA L S K Q 1860 GATAAGCTGC G >>>>>>> 01190 GATGTAGCAG 01190 GATGTAGCAA G L C K 2220 MGCGTTTAC E A F T 2340 ITCCTGATTT. L P D L 2460	GGAAAAGGTA E K V GCGAGAAGCG >> ~~~ CAAGTCATAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGG/ P G G	1770 GGGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC E I 2370 CAATC/ Q S 2490	CAGTATT SI CGTGCAC AGGAGC <u>G</u> CCCAACG PT CAAGGTG KV AATCGGT IG	GCAGG A C ATCC/ AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AGGA( AG	TGTCG V TTCG ATTCG A AGTG GAGAA I R GAGAA I R GCAAA C TT C L	GAAGTA E V ITCATC. P2 AAGTAG SDOI GTATTT V F AAATCA K S CATTCC H S	1 ACCGA T D 1 AGTAT Z TGAAT VB M 2 GAAAC E T 2 GGTGA G E 2 GGTGA G E 2 GGTCGG V G 2	800 CTT L 920 ATC 040 GCC P 160 CCA Q 280 GTC S 280 GTC S 520
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGICGGATTCATCAAATCAATCAAT	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CAAGTGAA	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA C L V S IGCTCAGAC A Q T IGCTCAGAC A Q T IGCACAAAAC K K T		AGCCGCC K P L CAACTGC T T C TTAAATC GTGTĀG S V CCTCAG S S L CCTCAG S S L CCTCAG S S L CCTCAG S S L CCTCAG S S L CCTCAG	1740 ICTCTAAGCA L S K Q 1860 GATAAGCTGC G >>>>>>> 01190 GATGTAGCAG 01190 GATGTAGCAG G L C K MGCGTTTAC E A F T 2340 ITCCTGATTT. L P D L 2460 CGGCAGGAAT	GGAAAAGGTAX E K V GCGAGAAGCG CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAAA V K K AAAAGTTATAA K V I IGAAGGGGGGC	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGG/ P G G GACATCATT	1770 GGGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CCGATI L I 2250 GAAATC E I 2370 CAATCS Q S2490 ATTGAC	CAGTATT S I CGTGCAC AGGAGCG P T CAAGGTG K V AATCGGT I G GATGAAT	GCAGG A C ATCC/ AGGA( AGGA( AGGA( C ACGGG T C T C GGAC/ C	TGTCC V TTCG GAGAI GAGAA CAAAA S K ACTTC C L	GAAGTA E V ITCATC. P2 AAGTAG GTATTT V F AAATCA K S CATTCC H S ATTGAA	1: ACCGA T D 1' AGTAT ZGAAT VB M 2: TGAAT E 2 GGAAAC E 2 GGTGA G E 2 GGTCGG V G 2 GTCGG V 2 AAAAT	800 CTT 920 ATC 040 GCC P 160 CCA 280 GCC S 280 CCA S 280 CTS V 520 GAA
$\begin{array}{cccc} TCTTTATATTGCGAAGGAGTTAAAAAG\\ L & Y & I & A & K & E & L & K \\ AACGAAACGCCATGCGAAAGAACTTT\\ T & K & R & H & A & K & E & L \\ CAATGTTTTTCTTCATATGACAGTTA\\ P1 \\ CGATAACATCAGAAAAGCAGTAGGTT\\ CGATAACATCAGGCGATTGAAAGCAGTAGGTT\\ D & N & I & R & K & A & V & G \\ AACACAAAGCGATTGAAAACGAGTTGATACGAGTTTAT\\ T & Q & A & I & E & T & S & L \\ AGAATTGGTATATGATCTTGCCGGAT\\ E & L & V & Y & D & L & A & G \\ TCTTGTCGGATTTCATCAAATCAAATCAATCAATCAATCA$	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTAAG 2070 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CAAGTGAA T S E 2550	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCTTGTTTC L V S IGCTCAGAC A Q T ICAAAAAAAC K K T IGGCAAAAA G K K	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA L L ATCAGAAT S E AAAAGTGC K V ATCTCCGG S P	AGCCGCC K P CAACTGC T T C TTAAATC GTGTĀG S V CCTCAGA S S I ATGTTC H V GAGAAAA G E	1740 ICTCTAAGCA L S K Q SGTAAGCTGC GGTAAGCTGC GGTAAGCTGC G >>>>>>> D1190 GATGTAGCAG D1190 GGTTATGCAA G L C K 2220 MAGCGTTTAC C A F T 2340 ITCCTGATTT. L P D L 2460 CGGCAGGAAT I A G I 2800	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAA D P H CCTGGCGGG/ P G G GACATCATT D I I	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L 1 2250 GAAATC G E 1 2370 CCAATC/ Q S 2490 ATTGAC I E 2610	CAGTATT SI CGTGCAC AGGAGCG PT CAAGGTG KV AATCGGT IG SATGAAT MN	GCAGG A C A C A C A GCAAAT Q P A C A C G A C G G A C G G A C C G G A C C C C	ITTCG I I I I I I I I I I I I I I I I I I I	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E	1 ACCGA T D 1 AGTAT TGAAT TGAAT VB M 2 GAAAC E T 2 GGAAC E 2 GGTGA G E 2 GGTGA G C 2 GTCGG V C C C C C C C C C C C C C C C C C C C	800 CTT 920 920 ATC 040 GCC P 160 CCA Q 280 CCA Q 30 CCA S 280 CCA S 280 CCA S 280 CCA S 280 CCA S 400 CTT L 920 ATC 920 ATC 920 ATC 920 ATC 920 920 ATC 920 920 920 920 920 920 920 920 920 920
TCTTTATATTGCGAAGGAGTTAAAAA L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTGAT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT/ L V G F H Q I N	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTAAG 2070 TAATTATC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CAAGTGAA T S E 2550 CCTGGGAAAA	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCTTGTTTC L V S IGCTCAGAC A Q T IAAAAAAAC K K T IGGCAAAAA G K K	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA ATCAGAAT S E AATCAGAAT S E AATCTCCGG S P AATCTTTAG	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V CCTCAG S S C CCTCAG S S C C CCTCAG S S C C CCTCAG S S C C CCTCAG S S C C CCTCAG S S C C CCTCAG S S C C CCTCAG S S C C C CCTCAG S S C C C CCTCAG S S C C C CCTCAG S S C C C CCTCAG S S C C C C CCTCAG S S C C C C C CCTCAG S S C C C C C C C C C C C C C C C C C C	1740 ICTCTAAGCA L S K Q 860 GGTAAGCTGC GGTAAGCTGC G >>>>>>> 00 00 00 00 00 00 00 00	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG	GCGGAAATC A E I CAGCTTATI <<<<<< AGAAGGTGI GAATATTT/ E Y L GATCCGCAI D P H CCTGGCGGG/ P G G GACATCATI D I I AAAATCAA/	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC E I 2370 CCAATC/ Q S 2490 ATTGAC I E 2610 OACGAAC	CAGTATT SI CGTGCAC AGGAGC <u>G</u> TCCAACGG PT CAACGTG KV NATCGGT IG GATGAAT MN SCTGATC	GCAGC A C A C A C A A C A A C A A C A A C A A C A A C A A C A A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C A C C A C A C A C C A C A C A C C A C C A C C A C C A C C C A C C A C C C A C C C C A C C C C C C C A C C C C C C C C C C C C C C C C C C C C	GTGTC( TTTCG TTTCG TGAGAGA TGAGAGA TR GCAAAA GCAAAA CL AGAAAA CL AGAAAA	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA	1 ACCGA T D 1 AGTAT TGAAT TGAAT TGAAT E 2 GGAAA G E 2 GGTGA G E 2 GGTGA G E 2 GGTGA G C 2 GGTGA G C 2 GGTGA G C 2 GGAGA	800 CTT 920 ATC 040 GCC P 160 CCA Q 280 GCC S 280 GTC S 280 GTC S 280 GTC S 280 GCA S 640 AGG
TCTTTATATTGCGAAGGAGTTAAAAA L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTGAT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT/ L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGC D V A P F I Q K	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTAAG 2070 TAATTATC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CAAGTGAA T S E 2550 CCTGGGAAA A G K	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCTTGTTTC L V S IGCTCAGAC A Q T IAAAAAAAC K K T IGGCAAAAA G K K IACTGGTGA T G E	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA AGGCAAAA GTTATTAA ATCAGAAT S E AATCAGAAT K V ATCTCCGG S P ATCTTCGG S L	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V C CCTCAG S S C CCTCAG S S C CCTCAG S S C CCTCAG S S C CCTCAG G C CCTCAG G C CCTCAG G C CCTCAG C CCTCAG C CCTCAG C CCTCAG C CCTCAG C CCTCAG C CCTCAG C CCTCAG C C CCTCAG C C CCTCAG C C CCTCAG C C C C C C C C C C C C C C C C C C	1740 ICTCTAAGCA L S K Q SGTAAGCTGC GGTAAGCTGC GGTAAGCTGC G >>>>>>> GGTTAGCAG G L C K C2220 AAGCGTTTAC E A F T 2340 TTCCTGATTT. L P D L 2460 CGGCAGGAATT T A G I 2580 L I K R	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q	GCGGAAATC A E I CAGCTTATT «<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGC P G G GACATCATT D I I AAAATCAAV K I K	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATIC L I 2250 GAAATC E I 2370 CCAATC/ Q S 2490 ATTGAC I E 2610 ACGAAC I E 2610 CACGAC C E R C R C C R C C R C C R C C R C C C C C	CAGTATT SI CGTGCAC AGGAGCG ICCAACG PT CAAGGTC KV AATCGGT IG GATGAAT MN SCTGATC LI	GCAGG A C ATCC/ ATCC/ AGGAA C AGGAA V J GGAAC/ G GGAC/ C C CCAG/ P I	GTGTCC TTTCG GAGTGJ GAGTGJ GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAA GCAAAAAA GCAAAAAAAA GCAAAAAAAAAA	GAAGTA E V ITCATC P2 AAGTAG <u>SDOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E	ACCGA T D AGTAT AGTAT TGAAT VB M GAAAC E 2 GGTGA G 2 GGTCGG G C G C G G C G G C G G C G G G C G G G G	800 CTT L 920 ATC 040 GCC P 160 AGC P 160 AGG S 520 GAA N 640 GAA GG G
TCTTTATATTGCGAAGGAGTTAAAAA L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTTTAT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT/ L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGC D V A P F I Q K	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATAAG 2070 TAATACTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGAA T S E 2550 CCTGGGAAA C S K 2670	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCTTGTTTC L V S IGCTCAGAC A Q T IAAAAAAAC K K T IGGCAAAAA G K K IACTGGTGA T G E	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA AGGCAAAA GTTATTAA ATCAGAAT S E AATCAGAAT S P ATCTCCGG S P ATCTTTAG S L	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V C CCTCAG S S C CCTCAG S S C CCTCAG S S C CCTCAG S S C CCTCAG C C C CCTCAG C C C C CCTCAG C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C	1740 ICTCTAAGCA L S K Q SGTAAGCTGC GGTAAGCTGC G >>>>>>> 1980 SATGTAGCAG G L C K 2220 AAGCGTTTAC E A F T 2340 ITCCTGATTTAC E A F T 2340 ITCCTGATTTAC CGGCAGGAAT T A G I 2580 L A K R 2700 CGGCAACG L I K R 2700	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA V K K AAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q	GCGGAAATC A E I CAGCTTATT «««««« AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGCGG P G G GACATCATT D I I AAAATCAAA K I K	1770 GAGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CTGATTC L I 2250 GAATC E I 2370 CAATC/ G S 2490 ATTGAC I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E CAATCA I E CAACCAAC I E I E I E I E I E I E I E I E I E I E	CAGTATT S I CGTGCAC AGGAGCQ ICCAACG P T CAAGGTG K V AATCGGT I G SATGAAT M N SCTGATC L I	GCAGG A C ATCC/ AATCC/ AAGGA( CAAAAT Q D AGGAA/ V J GGAA/ V J GGAC/ G C C CCCAG/ P I	GTGTCC TTTCG TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAGAGA TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGAT TGA	GAAGTA E V ITCATC P2 AAGTAG <u>SDOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E	T AGCGAA T D 1' AGTAT. 2' TGAATI GAAAC E T 2 GGAAC G E 2 GTCGG G E 2 GTCGG G C E GTCGG G E 2 GGAAA G E 2 GGAGA G E 2 GGAGA G E 2 GGAGA	800 CTT 920 ATC 040 GCC 160 CCA 280 GTC S20 AGG S400 TGT V 520 AGG GAA AGG G 760 C
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT L V G F H Q I N TGATGTAGCCCCATTTATTCAAAAGC D V A P F I Q K	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATAAG 2070 TAATACTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGAAA T S E 2550 CCTGGGAAA T S E 2550 CCTGGGAAA T S E 2550 CCTGGGAAA T S E 2570 CCTGGGAAA C S C S S S S S S S S S S S S S S S S S	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCTTGTTTC L V S IGCTCAGAC A Q T IAAAAAAC K K T IGGCAAAAA G K K IACTGGTGA T G E		AGCCGCC K P CAACTGC T T C TTAAATC GTGTAGG S V C CCTCAG S S C ATGTTC H V C GAGAAAA G E ACTTAC D L C CTATGA	1740 ICTCTAAGCA ICTCTAAGCAG ICTCTAAGCAG IS800 IGTAAGCTGC IS900 IS800 IS800 IGTTAAGCAA IGTTTATGCAA G L C K 22200 IGGCAGGAATTI L P D L 23400 ITCCTGATTIACAACG IGATCAAACG L I K R 2700 IGATCAAACG L I K R 2700 ICCTTTATGAACG I F Y E	GGAAAAGGTA E K V GCGAGAAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAAG D K Q ACCGAAAACAA P K T	GCGGAAATC A E I CAGCTTATT «<<<<< AGAAGGTGT GAATATTTI E Y L GATCCGCAT D P H CCTGGCGGC P G G GACATCATT D I I AAAATCAAA K I K AAAAAATAC K K Y	1770 GAGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CTGATTC I 2250 GAATC I 2370 CAATC/ G S 2490 ATTGAC I E 2610 ACGAAC T K 2730 GGGAGC/ G A	CAGTATT SI CGTGCAC AGGAGCQ ICCAACG PT CAAGGTG KV AATCGGT IG SATGAAT MN SCTGATC LI	GCAGG A C ATCC/ AGGAA AGGAA C AGGAA C G GTAAA C G G C C C C C C C C C C C C C C C	ITGTC() V ITTCG I GAGAGA( I R GCAAAA S K VACTT( C L AGAAAA S K VAAAG( I K V AAAAG( I X I V I I I I I I I I I I I I I I I I	GAAGTA E V TTCATC P2 AAGTAG <u>SDOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E TCCGAT S D	T CGAX ACCGAX T D 1' AGTAT. 2 TGAAT GAAAC E T 2 GGAGA G E 2 GTCGG G E 2 GTCGG G E 2 GTCGG G E 2 GTCGG G E 2 GTCGG G E 2 C C C C C C C C C C C C C C C C C C	800 CTT L 920 ATC 0400 GGCC P 160 GGCC P 160 CCA Q 280 CCA Q 280 CCA Q 400 S20 AGG G 760 CCA T
$\begin{array}{c} TCTTTATATTGCGAAGGAGTTAAAAGL Y I A K E L KAACGAAACGCCATGCGAAGAACTTTT K R H A K E LCAATGTTTTTCTTCATATGACAGTAP1CGATAACATCAGAAAAGCAGTAGGTTD N I R K A V GAACACAAGCGATTGAAACGAGTAGGTTT Q A I E T S LAGAATTGGTATATGATCTTGCCGGATE L V Y D L A GTCTTGTCGGATTTCATCAAATCAATL V G F H Q I NTGATGTAGCCCCCATTTATTCAAAAGCD V A P F I Q KCAAATACAGAATCGGGTTATATATC/K Y R I G L Y I$	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATAAG 2070 TAATATAAG 2070 TAATATCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGAAA T S E 2550 CCGGGAAA T S E 2550 CCGGGAAA T S E 2550 CCGGGAAA T S E 2570 GAGATTCT R D S 2790	ACGACGAC T T T GCCGGATCA A D Q GCCGTCAGA CCTTGTTTC L V S GCTCAGAC A Q T GAAAAAAC K K T AGGCAAAAA G K K T G E GCTGCTGG A A G	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA AGGCAAAA GTTATTAA ATCAGAAT S E AAAAGTGC K V ATCTCCGGC S P ATCTTTAG S L GCATCGGCA	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAGG S V C CCTCAGG S S C ATGTTC H V C GAGAAAA G E ACTTAC D L C CTATGA T M	1740 ICTCTAAGCA ICTCTAAGCAG ICTCTAAGCAG IS800 IGTAAGCTGC IS900 IS800 IS800 IS800 IS800 IS800 IGATGAAGCAG ICCTGATTTA ICCTGATTTAC ICCTGATTTAC ICCTGATTTAC ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGCAACG ICCGCAACG ICCGCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCAACG ICCCCAACG ICCCCAACG ICCCCAACG ICCCAACG ICCCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCAACG ICCCCAACG ICCCCAACG ICCCAACG ICCCCAAC	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T	GCGGAAATC A E I CAGCTTATT «««««« AGAAGGTGI GAATATTTI E Y L GATCCGCAI D P H CCTGGCGGC P G G GACATCATT D I I AAAATCAAA K I K AAAAATAG K Y	1770 GAGCGC E R 1890 TTTTTC 2010 GGGAT/ 2130 CTGATTC I 2250 GAATC E I 2250 CAATC C E I 2370 CAATC/ E S 2490 ATTGAC I E 2610 ACGAAC T K 2730 GGGGC/ G A C 2730 CGGGGC/ G A C 2850	CAGTATT SI CGTGCAC AGGAGC <u>G</u> TCCAACG PT CAAGGTG KV AATCGGT IG SATGAAT MN SCTGATC LI ACTTGGC LG	GCAGG A C ATCC/ AGGAA AGGAA AGGAA T C GTAAJ V J GGAA/ C GGAA/ C C CCAGG C C CCAGG H V	ITGTC() V ITTCG' I GAGAGA( I R GCAAAA C C CAAAA C C C C C C C C C C C C C	GAAGTA E V ITCATC P2 AAGTAG SDDI GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E ICCGAT S D	T AGCGA ACCGA T D T T AGTAT. 2 TGAAT T GAAA G T C G T C G G T C G G T C G G T C G G T C G G C G G C G C	800 CTT L 920 GGC P 160 GGC P 160 CCA Q 280 CCA Q 280 CCA S 400 T S 20 A G G C 520 A A C 520 A C 520 CA T 520 CA T 520 C T T 880 C T T 880 C T T 880 C T T 160 C C 7 C 7 C 7 C 7 C 7 C 7 C 7 C 7 C 7 C
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT L V G F H Q I N TGATGTAGCCCCATTTATTCAAAAGC D V A P F I Q K CAAATACAGAATCGGGTATATATC/ K Y R I G L Y I	1710 ACGGCAGA D G R 1830 TAAAACAA L K G 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CAAGTGAA T S E 2550 CGAGGAATCT R D S 2790 GAGGAATC	ACGACGAC T T T IGCGGATCA A D Q GCCGTCAGA CL V S IGCTCAGAC A Q T IAAAAAAC K K T IGGCAAAAA G K K IGCTGGTGG A A G IGCTGCTGG A A G	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA L L ATCAGAAT S E AATCTCCGG C S P ATCTTTAG S L GCATCGGCA I G CCACTGTAA	AGCCGCC K P CAACTGC T T C TTAAAT( GTGTAG S V C CCTCAG S S C ATGTTC H V C GAGAAAA G E GAGAAAA G E GAGAAAA G E CTATGA	1740 ICTCTAAGCA ICTCTAAGCAG ICTCTAAGCAG IS800 IGTAAGCTGC IS980 IS80 IS80 IS80 IS80 IS80 IGATGAAGCAG IGACAGGAAT ICCTGATTIA ICCTGATTIA ICCTGATTIA ICCTGATTIA ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGGCAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAAGGAAT ICCGCAGGAAGGAAT ICCGCAGGAAGGAAT ICCGCAGGAAGGAAT ICCGCAGGAAGGAAT ICCGCAGGAAGGAAT ICCGCAGGAAGGAAGGAAGGAAGGAAT ICCGCAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAGGAAGAA	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA ACCGCTAAAA P L K AGTAAAGAAA V K K AAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T GACAGGCGGGT.	GCGGAAATC A E I CAGCTTATT <<<<<<< AGAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGG/ P G G GACATCATT D I I AAAATCAA/ K I K AAAAATAA K K Y	1770 GAGCGC E R 1890 11111C 2010 GGAT/ 2130 CTGATIL 1 2250 GAATC E I 2250 CAATC E I 2370 CAATC I E 2490 CAATCA I E 2610 CGAATC E R 2490 CAATCA I E 2490 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CAATCA I E 2610 CACGAACA I E 2650 CAACCA I E 2650 CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CACCA CAC	CAGTATT SI CGTGCAC AGGAGC <u>G</u> TCCAACG PT CAAGGTG KV ATCGGT IG SATGAAT N N SCTGATC LI ACTTGGC ACTGGCG	GCAGG A C ATCC/ AGGAA AGGAA AGGAA T C GTAA/ V J GGAA/ V J GGAA/ C CCAGG H V SCACGI H V	ITGTCC	GAAGTA E V TTCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E TCCGAT S D	T AGCGAA ACCGAA T D T 1' TGAATAT. 2 TGAATA TGAATA G GAAA G C C C C C C C C C C C C C C C	800 CTTL 920 ATC 0400 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 060 P 000 P 060 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 000 P 00 P 000 P 00 P 000 P 00 P 00 P 0 P 0 P 00 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P 0 P P 0 P P 0 P P P P P P P P P P P P P P P P P P P P
$\begin{array}{c} {} TCTTTATATTGCGAAGGAGTTAAAAG  \\ L & Y & I & A & K & E & L & K \\ {} AACGAAACGCCATGCGAAAGAACTTT \\ T & K & R & H & A & K & E & L \\ {} CAATGTTTTTCTTCATATGACAGTAT \\ P1 \\ {} CGATAACATCAGAAAAGCAGTAGGTT \\ D & N & I & R & K & A & V & G \\ {} AACACAAGCGATTGAAAAGCAGTAGGTT \\ D & N & I & R & K & A & V & G \\ {} AACACAAGCGATTGAAACGAGTAGGTT \\ T & Q & A & I & E & T & S & L \\ {} AGAATTGGTATATGATCTTGCCGGATT \\ \mathsf{E & L & V & Y & D & L & A & G \\ {} TCTTGTCGGATTTCATCAAATCAAT  \\ L & V & G & F & H & Q & I & N \\ {} TGATGTAGCCCCCATTTATTCAAAAGCA \\ \mathsf{D & V & A & P & F & I & Q & K \\ \\ CAAATACAGAAATCGGGTTATATATC \\ K & K & R & I & G & L & Y & I \\ \\ AAAGAAACCAATTGTAGTGGGAGAATCG \\ K & K & P & I & V & V & E & N \end{array}$	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGGAAA T S E 2550 CCTGGGAAA A G K 2670 GAGAATTCT R D S 2790 GAGAAATC G E I 2010	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCGTCAGAC L V S IGCTCAGAC A Q T IAAAAAAC K K T IGGCAAAAA G K K IGCTGGTGA T G E IGCTGCTGG A A G ICGTTAAATC V K S	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA C L CATCAGAAT S E AATCAGAAT S C ATCTCCGGCA ATCTTTAG S L GCATCGGCA I G CACTGGCA T V	AGCCGCC K P I CAACTGG T T ( TTAAAT( GTGTÅG S V ( CCTCAG S S ( ATGTTC H V I GAGAAAA G E ACTTAC D L CTATGA T M CATCAA T S	1740 ICTCTAAGCAG ICTCTAAGCAG ISCTAAGCTGC GATAAGCTGC GATGTAGCAG DIITATGCAA GATGTAGCAA GATGTAGCAA GACGTTTAC CACAACGAATT ICCTGATTTA L P D L 2460 CGGCAGGAAT I K R 2700 CCTTTTATGAA I K R 2700 CCTTTTATGA I F Y E 2820 ITGAAAAAGG I E K G 2020	GGAAAAGGTAA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T GACAGGCGGGT. T G G	GCGGAAATC A E I CAGCTTATT <<<<<<< AGAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGG P G G GACATCATT D I I AAAATCAAA K I K AAAAAATAA K K Y AATCCGGGG N P G	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAT( E I 2370 CAATC/ Q S 2490 CAATC/ Q S 2490 CAATC/ Q S 2490 CAATC/ Q S 2490 CAATC/ Q S 2490 CAATC/ Q S 2490 CAATC/ Z 2610 CAATC/ C R E R 2730 CGAAC/ C R E R 2850 CGAAC/ C R E R 2850 CGAAC/ C R E R 2850 CAATC/ C R E R 2850 CAATC/ C R C R C R C R C R C R C R C R C R C R	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CAAGGTG KV ATCGGT IG SATGAAT MN SCTGATC LI SCTGACC LG ACTGGCG LA	GCAGG A C ATCC/ AAGGA( AAGGA( C CAAAAT A G GTAA/ V J C GTAA/ V J C GTAA/ V J C GTAA/ V J C GTAA/ C C C C C C C C C C C C C C C C C C	ITGTC( V ITTCG I GAGTG GCAAA C C C C C C C C C C C C C C C C	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E TCCGAT S D TCAGAA S E	Т АСССВА Т СССВА Т СССВА Т ССССБА С ССССБА С ССССБА С ССССБА С ССССБА С ССССБА С СССССА С С СССССА С С ССССА С С ССССА С С ССССА С С С С С С С С С С С С С	800 CTTL 920 ATC 040 CC 940 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CCC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 70 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 2 CC 90 CC 90 CC 2 C CC 2 C CC 2 C C CC 2 C CC 2 CC 2 C C
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT L V G F H Q I N TGATGTAGCCCCATTTATTCAAAAGG D V A P F I Q K CAAATACAGAAACGGGTATATATC/ K Y R I G L Y I AAAGAAACCAATTGTAGTGGAGAATC K K P I V V E N	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGGAAAT CGGTAAAT S E 2550 CCGGTAGGAAATCC R D S 2790 GAGAATTCT R D S 2790 GAGAAATCC G E I 2910 CG E I 2910	ACGACGAC T T T GCCGGATCA A D Q GCCGTCAGA CLTTGTTTC L V S GCTCAGAC A Q T GAAAAAAC K K T AGGCAAAAA G K K T G E GCTGCTGG A A G CGTTAAATC V K S GATTTTCGG	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA C L L ATCAGAAT S E AAAAGTGC K V ATCTCCGGCA S C CACACGGCA G T V CACACTGG	AGCCGCC K P I CAACTGG T T ( TTAAAT( GTGTĀG S V ( CCTCAG S S ( ATGTTC H V I GAGAAAA G E ACTTAC D L CTATGA T M CATCAA T S	1740 ICTCTAAGCA ICTCTAAGCAG ICTCTAAGCTGC GOTAAGCTGC GOTAAGCTGC GOTTAGCAG ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO ISTO	GGAAAAGGTAA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGCAAA V K K AAAAGTTATA K V I TGAAGCGGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T GACAGGCGGGT. T G G CAACATATCA	GCGGAAATC A E I CAGCTTATT <<<<<<< AGAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGG/ P G G GACATCATT D I I AMAATCAAJ K I K AAAAATAA K K Y AATCCGGG/ N P G GATCAAGC/	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAT( E I 2370 CAATCC Q S 2490 CATCGAC T K 2730 GAAC C A 2610 CGAAC C R 2490 CATCGAC C R 2490 CAATCC Q S 2490 CAATCC Q S 2490 CAATCC C C 270 CAATCC Q S 2490 CAATCC C C C C C C C C C C C C C C C C	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CAAGGTG KV ATCGGT IG SATGAAT MN SCTGATC LI SCTGGCG LA SGTTGCG	GCAGG A C ATCC/ AGGAA AGGAA AGGAA A ACGGA C C C C C C C	ITGTCC ITTCG ITTCG IGAGAGA I I I I I I I I I I I I I	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E TCCGAT S D TCAGAA S E GAAGTC	Т. АССGАЙ Т. D. Т. Т. АССGАЙ Т. D. Т. Т. Т. АСТСАЙ Т. Т. АСТАТ. Т. С.	800 CTTL 920 ATC 040CC P 16CA 02 280 CC 280 CC 280
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAATCA L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGG D V A P F I Q K CAAATACAGAAACGGGTTATATATCC K Y R I G L Y I AAAGAAACCAATTGTAGTGGAGAATC K K P I V V E N	1710 ACGGCAGA D G R 1830 TAMAACAA L K G 1950 TAMATAAG 2070 TAATTCTC L I L 2190 CGGTAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGGAAA A G K 2670 CGGGAAATCCT R D S 2790 GAGAATTCT R D S 2790 GAGAATTCT G E I 2910 CGTTTGGG P F G	ACGACGAC T T T GGCGGATCA A D Q GCCGTCAGA CTTGTTC L V S GGCTCAGAC K K T GGCCAAAAA G K K T G E GGCTGCTGG A A G CGTTAAATC V K S GATTTTCGG I F G	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA GTTATCAGAAT S E AAAAAGTGC K V ATCTCCGGCA S P ATCTTCCGGCA G S C CACTCGGCA G T V GCACACTGC	AGCCGCC K P I CAACTGC T T ( TTAAAT( GTGTĀG S V ( CCTCAG S S ( CCTCAG S S ( CCTCAG G E ATGTTC H V I GAGAAAA G E ACTTAC D L I CTATGA T M CATCAAC T S ATCAGCC H Q I	1740 (CTCTAAGCA (CTCTAAGCAG (CTCTAAGCTGC (CTCTGAGCAG (CTCTGAGCAG (CTCTGATTA (CTCTGATTA (CTCTGATTA (CTCTGATTA (CTCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATTA (CTCCTGATA (CTCCTGATA (CTCCTGATA (CTCCTGATA (CTCCTGATA (CTCCTGATA (CTCCTGATA (CTCCTGATA (CTCCTAAAA (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAGG (CTCCTAAAAAAAAGG (	GGAAAAGGTAA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T GAACAGGCGGT. T G G CAACATATCA N I S	GCGGAAATC A E I CAGCTTATT <<<<<<< AGAATATTT/ E Y L GATCCGCGAT D P H CCTGGCGGG/ P G G GACATCATT D I I AAAATCAAA K I K AAAAAATAG K K Y AATCCGGGG/ N P G GATCAAGC/ D Q A	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC Q S 2490 CAATC/ Q S 2490 CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S 2490 CAATC/ C S C C C C C C S C C S C C S C C S C C S C C S C S	CAGTATT SI CGTGCAC AGGAGC <u>G</u> PT CAAGGTG KV ATCGGT IG SATGAT LI SCTGATC LI ACTTGGC LA SGTTGCC VA	GCAGC A C ATCC/ AGGAC AGGAC C CAAAAT G A C GTAAA C GTAAA C GGAC/ C C CCAGG C C C CCAGG C C C C C C C C C	ITGTC( V ITTCG I GAGTG GAGAA C C C C C C C C C C C C C C C C	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E TCCGAT S D TCAGAA S E GAAGTC E V	Т С GA A C C GA T D 1 1 1 1 1 1 1 1 1 1 1 1 1	800 CTL 920 ATC 94 CCA 920 94 CCA 920 94 CCA 920 94 CCA 920 94 CCA 920 94 CCA 920 94 CCA 920 94 CCA 94 CCA 920 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 9 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 94 CCA 9 CCA 9 CCA 9 CCA 9 CCA 9 C C 9 C CCA 9 C C CCA 9 C C C C
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAATA L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGGA D V A P F I Q K CAAATACAGAATCGGGTTATATATCA K Y R I G L Y I AAAGAAACCAATTGTAGTGGAGAATC K K P I V V E N GATCGGGGATATTAACAGAAACAGCC I G D I N R N S	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAMATAAG 2070 TAMTTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGTGAAA T S E 2550 CCAGGGAAATC R D S 2790 GAGAATTCT R D S 2790 GAGAATCCT G E I 2910 CGTTTGGG P F G 3030	ACGACGAC T T T GCGGATCA A D Q GCCGTCAGA CTTGTTC L V S GCTCAGAC K K T GCTCAGAC K K T GGCAAAAA G K K ACTGGTGA T G E GCTGCTGG A A G CGTTAAATC V K S GATTTTCGG I F G	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA C L CATCAGAAT S E AATCAGAAT S P ATCTCCGGCA S P ATCTTCCGGCA S C CACTCGGCA G I G CACCTGTAA T V CACACTGC	AGCCGCC K P I CAACTGC T T ( TTAAAT( GTGTĀG S V CCTCAG S S I CCTCAG S S I CCTCAG G E ATGTTC D L CAGAAAA T S ATCAGCC H Q I	1740 (CTCTAAGCA (CTCTAAGCAG SCTAAGCTGC GTTAAGCTGC GTTAGCAG GTTAGCAG GATGTAGCAG GATGTAGCAG GATGTAGCAG GATGTAGCAG GATGTAGCAG CGCCAGGAAT T A G I 2580 CGCCAGGAAT T A G I 2580 CCCTTTAAGAG CGCCAGGAAT T CAAAAAGG I E K G 2940 CGCATTCAAAA P I Q N 3060	GGAAAAGGTAA E K V GCGAGAAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA P L K AGTAAAGGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T GAACAGGCGGT. T G G CAACATATCA N I S	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCGAT D P H CCTGGCGGG/ P G G GACATCATT D I I AAAATCAAA K I K AAAAAATA K K Y AATCCGGGG/ N P G GATCAAGC/ D A A	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CCTGATI L I 2250 GAAAT( E I 2370 CCAATC/ Q S 2490 ACCGAAC I E C 2490 ACCGAAC I K 2730 GGAGC/ G A 2850 GGAAA/ E K 2970 CCGCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CAGTATT S I CGTGCAC GGAGCG P T CAAGGTG K V ATCGGT I G SATGAAT M N SCTGATC L I ACTTGGC L G ACTGGCG L A SGTTGCG V A	GCAGC A C ATCC/ AGGA( AGGA( C C C C C C C C C C C C C C C C C C C	ITGTC() ITTCG ITTCG IGAGAGA IR GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA GCAAAA IR ITTCC I I I I I I I I I I I I I	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S CATTCC H S CATTCCA H S CATTCCA H S CATTCCA H S CATCCA CATCCA CATCCA CAGAA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S C CATCA S CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C C CATCA S C C CATCA S C C CATCA S C C CATCA S C C CATCA S C C CATCA S C C C C C C C C C C C C C C C C C C	Т С GAX Т D 1 1 1 1 1 1 1 1 1 1 1 1 1 1	800 CTTL 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 920 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04CC 90 04 00 04 0 0 0 0 0 0 0 0 0 0 0 0 0 0
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAATA L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGG D V A P F I Q K CAAATACAGAAACGGGTTATATATCA K Y R I G L Y I AAAGAAACCAATTGTAGTGGAGAATC K K P I V V E N GATCGGGGATATTAACAGAAACAGCC I G D I N R N S GCCGGCTGAAATTTTAACGGTTATTCA	1710 ACGGCAGA D G R 1830 TAMAACAA L K G 1950 TAMATAAG 2070 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CCAGGGAAAT CAGGGAAAT CAGGAAATCC G E I 2910 CGGTTGGGAAATCC G E I 2910 CGGTTGGGAAATCC CG C E I 2910 CGGTTGGGAAATCC CG C E I 2910 CGGTTGGACAACC CG C E I 2910 CGGTTGGACAACC CG C E I 2910 CGGTTGGGAAATCC CG C E I 2910 CGGTACC CG C E I 2910 CGGTTGGCAAATCC CG C E I 2910 CGGTTGGCAAATCC CG C E I 2910 CGGTTGGCAAATCC CG C E I 2910 CGGTTGGCAAACC CG C E I 2010 CGGTTGGCAAACC CG C E I 2010 CGTTGGCAACC CG C E I 2010 CGTTGGCAACC CG C E I 2010 CGTTGGCACC CG C E I 2010 CGTTGGCACC CG C E I 2010 CGTTGGCACC CG C E I 2010 CG CG C E I 2010 CG CG C E I 2010 CG CG C E I 2010 CG CG CG CG C C C C C C C C C C C C C	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA ICCGTCAGAC L V S IGCTCAGAC A Q T IAAAAAAAC K K T IGCTGCTGG A A G I F G I F G ICCAAGTAGA K V F	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA ATCAGAAAT S L AATCTCCGGCA S P ATCTTCCGGCA S L GCACTGCACTGC T L AAAATTCG K F	AGCCGCC K P I CAACTGC T T ( TTAAAT( GTGTĀG S V ( CCTCAG S S I CCTCAG S S I CCTCAG C E CAGAGAAA G E CATCTAC D L I CTATGA T M CATCAGC H Q I CATCAGC H Q I CATCAGC	1740 ICTCTAAGCA ICTCTAAGCA ICTCTAAGCTGC GATAAGCTGC GATGTAGCAG D1980 GATGTAGCAG D1980 GATGTAGCAG GATGTAGCAG GATGTAGCAG CGCAGGATTAC CGCAGGATTAC CGCAGGAAT T A G I 2340 ICCTGATTTA CGCAGGAAT T A G I 2460 CGCAGGAAT T A G I 2580 ICCTCTTATGAA T A G I 2820 ICCTTTTATGAAAGG I E K G 2940 CGATTCAAAAG F I Q N 3060 MAATCGTCAG E I V S	GGAAAAGGTAX E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA P L K AGTAAAGAAA V K K AAAAGTTATAA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACAG D K Q ACCGAAAACAG C G GACAGGCGGGT T G G CAACATATCA N I S CACAACGCCCG T T P	GCGGAAATC A E I CAGCTTATT <<<<<< AGAACTTATT GAATATTT/ E Y L GATCCGCGAT D P H CCTGGCGGC P G G GACATCATT D I I AAAATCAAA K I K AAAAAATAG K K Y AATCCGGGG N P G GATCAAGC/ D Q A CAAAAATTC Q K F	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC Q S 2490 ACTGAT I E 2370 ACAATC/ Q S 2490 ACTGAC I E 2610 ACGAAC I E 2610 CGAATC/ Q S 2490 ACTGAC I E 2730 GGAGC/ G R 2730 CGGAGC/ G R 2730 CGGAGC/ C R 2730 CGGAGC/ C R 2730 CGGAGC/ C R 2730 CGGAGC/ C R 2730 CGGAGC/ C R 2730 CGGAGC/ C R 2850 CGGAGC/ C R 2730 CGGAGC/ C R 2730 C C C C C C C R C R C R C R C R C R C R	CAGTATT SI CGTGCAC GGGAGCG PT CAAGGTG KV ATCGGT IG SATGAAT MN SCTGATC LI ACTTGGC LG ACTGGCG LA SGTTGCG VA SACAAAA TK	GCAGC A C ACCA AGGA( AGGA( C C C C C C C C C C C C C C C C C C C	ITGTC() ITTCG ITCGG IGAGA() IR IGCAAA IR IGCAAA IR IGCAAA IR IGCAAA IR IGCAAA IR IGCAAA IR IGCAAA IR IGCAAAA IR IR IGCAAAA IR IR IR IR IR IR IR IR IR IR	GAAGTA E V ITCATC P2 MAGTAG SDOI GTATTT V F MAATCA K S CATTCC H S CATTCC S CATTCC S CATTCC S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S CATCCA S C CATCCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C C CATCA S C C CATCA S C CATCA S C CATCA S C C CATCA S C C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C CATCA S C C CATCA S C C CATCA S C C C CATCA S C C CATCA S C C C C C C C C C C C C C C C C C C	Т. С. GAX Т. D. T. T. AGTAT. 1 1 1 1 4 GTAT. 1 1 4 GTAT. 2 GAAAT. 2 GAAAT. 2 GGAAACC 2 GGAAACC 2 GGAAACC 2 GGGAGA 4 C 2 2 AAAAT K M D 2 CGCAA R S 3 AAAAA K K K 3 AATTACC 1 T	8000 CTTL 9200 ATC 0400 P1600 2800 2800 2800 2800 2800 2800 2800 2
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAATCA L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGG D V A P F I Q K CAAATACAGAAACCGGGTTATATATCC/ K Y R I G L Y I AAAGAAACCAATTGTAGTGGAGAATC K K P I V V E N GATCGGGGGTATTTAACAGAAACAGCC I G D I N R N S GCCGGCTGAAATTTTAACGGTTATTCC/	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAATT CGGTAATT S V N 2310 TTCCAATT F P I 2430 CTGGGAAATC GAGATTCT R D S 2790 GGAGAATCCT R D S 2790 3030 ATGATGACGAC D D D D 3150	ACGACGAC T T T GGCGGATCA A D Q GCCGTCAGA CTTGTTCC L V S GGCTCAGAC A Q T GGCTCAGAC K K T AGGCAAAAA G K K T G E GGCTGCTGG A A G CGTTAAATC V K S GGTTAAATC V K S GATTTTCGG I F G CAAAGTAGA K V E	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA ATCAGAAAT S L ATCTCCGGCA S P ATCTTCCGGCA S L GCACTGCCACTGC T L AAAATTCG K F	AGCCGCC K P I CAACTGC T T C TTAAATC GTGTĀG S V CCTCAG S S I CCTCAG S S I CCTCAG G E ACTTAC D L I CATATAG T S ATCAGCC H Q I CATATAG D I I	1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740 1740	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGTAAAA P L K AGTAAAGTAAA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACA P K T GACCGAAAACA P K T GACAGGCGGGT. T G G CAACATATCA N I S CACAACGCCGG T T P	GCGGAAATC A E I CAGCTTATT <<<<<< AGAACTTATT GAATATTT/ E Y L GATCCGCCAT D P H CCTGGCGG/ P G G GACATCATT D I I AAAATCAAA K I K AAAAATAG K K Y AATCCGGG/ N P G GATCAAGC/ D Q A CAAAATTC Q K F	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC/ G SAAATC/ Q S 2490 ACTGAC I E 2610 ACGAAC I E 2610 ACGAAC I E 2630 ACGAAC I E 2730 ACGAAC I E 2730 ACC I E 270 ACC I E 270 ACC I E 270 ACC I E 270	CAGTATT S I CGTGCAC AGGAGCG P T CAAGGTG K V ATCGGT I G SATGAAT M N SCTGATC L I ACTTGGCC L A SGTTGCC V A SACAAAA T K	GCAGC A C ACCA AGGAC CAAAAT Q D ACCGC G G G C C C C C C C C C C C C C C	ITGTCCI ITTCG ITTCG IGAGAGA I R GCAAAA I R GCAAAAA I R GCAAAAAA I R GCAAAAAA I R GCAAAAAA I R GCAAAAAA I R GCAAAAAA I R GCAAAAAA I R GCAAAAAAAA I R GCAAAAAA I R GCAAAAAA I R GCAAAAAAAA I R GCAAAAAAA I R GCAAAAAAAAAA I R GCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	GAAGTA E V ITCATC P2 AAGTAG SDOI GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E ICCGAT S D ICCGAT S D ICCGAA S E GAAGTC E V ITGAAA L K	T CGAA ACCGAA T D I I AGTAT. 2 TGAATA 2 TGAATA 2 GGAAA C E T GGAAA C E T GGTGA C G G G G G G G G G G G G G G G G G G	800 CTTL 920 ATC 920 ATC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 920 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 20 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 90 CC 20 CC 20 CC 20 CC 20 CC 20 CC 20 CC 20 C 20 CC 20 CC 20 C
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L K AACGAAACGCCATGCGAAAGAACTTT T K R H A K E L CAATGTTTTTCTTCATATGACAGTA P1 CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V G AACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S L AGAATTGGTATATGATCTTGCCGGAT E L V Y D L A G TCTTGTCGGATTTCATCAAATCAAT/ L V G F H Q I N TGATGTAGCCCCCATTTATTCAAAAGG D V A P F I Q K CAAATACAGAAACCGGGTATTATATC/ K Y R I G L Y I AAAGAAACCAATTGTAGTGGAGAATC K K P I V V E N GATCGGGGATATTAACAGAAACAGGC I G D I N R N S GCCGGCTGAAATTTTAACGGTATTC P A E I L T V I	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAATT S V N 2310 TTCCAATT F P I 2430 CGAGTGAAA T S E 2550 CTGGGAAATC G A G 2550 CTGGGAAATC G G E I 2910 CGTTTGGG G S 3030 ATGATGAC D D D S150 SCACCGTA	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGA CLTTGTTTC L V S IGCTCAGAC A Q T IAMAAAAAC K K T AGGCAAAAAA G K K IACTGGTGA A A G I F G I F G IAAAAGTAGA K V E IAAAGTAGA	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA ATCAGAAAT S E AAAAGTGC K V ATCTCCGGCA S P ATCTTTAG S L GAACAGTGC A T C GAACACTGC K F GAAGCGGAA	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V CCTCAG S S I CCTCAG S S I CCTCAG S S I CCTCAG C CTCAG C CTCAG C CTCAG C CCTAG C CCTAG C CCCGA	1740 ICTCTAAGCA ISK SK GTAAGCTGC GTAAGCTGC GTAAGCTGC GTAAGCTGC GTAAGCAG DIGO GATGTAGCAA GLCK 2220 MAGCGTTTAC CACCA TCCTGATTT. LPDL 2460 CGGCAGGAAT TAGI 2580 TGATCAAACG LIKR 2700 CCTTTTATGA TGAAAAAGG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAAG LEKG 2940 CGATCAAAA	GGAAAAGGTA E K V GCGAGAAGCG CAAGTCATAA P L K AGTAAAGTAAAA P L K AGTAAAGTAAA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACAG C A CGAAAACAA P K T GACAGGCGGT. T G G CAACATATCA N I S CACAACGCCGG T T P TGGAAAAGTG	GCGGAAATC A E I CAGCTTATT <<<<<< AGAACTTATT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGG/ P G G GACATCATT D I I AAAATCAA/ K I K AAAAATAC K K Y AATCCGGGG/ D Q A CAAAATTC Q K F ATCCGGTGC	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAATC/ G SAATC/ G SAATC/ Q S 2490 ATTGAC I E 2610 ACGAAC I E 2630 ACGAAC I E 2730 ACGAAC I E 2630 ACGAAC I E 2730 ACGAAC I E 2900 ACGAAC I E 2730 ACGAAC I E 2900 ACGAAC I E 2900 ACGAAC I E 2900 ACGAAC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACC ACC I E 200 ACGAAC I E 200 ACGAAC I E 200 ACC I E 20 ACC I E 20 ACC 20 ACC I E 20	CAGTATT SI CGTGCAC CGGAGCG PT CAAGGTG KV ATCGGT IG SATGAAT MN SCTGATC LI ACTGGCG LA SGTTGCG VA SACAAAA TK CCATGTA	GCAGG A C ACCA AGGAC C AGGAAAT Q D ACCGGC C C C C C C C C C C C C C C C C	ITGTCCI ITCG ITCG IGAGAA IR IGAGAA IR IGAGAA IR IGGTG IV IGGTG IV I IGGTG IV I I I I I I I I I I I I I	GAAGTA E V ITCATC P2 AAGTAG SDOI GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA D E ICCGAT S D ICCGAT S D ICAGAA S E GAAGTC E V ITGAAA L K GACCCG	T AGCGAA T D I AGTAT. I GAAAT I GAAAC E T GGATGA G C GGGGA C GGGGA C GGGGA C GGGGA C GGGGAA C C GGAAA K C C GCAAA K C C GCAAA C A C C GAAAA C A C C GAAAA C A C	800 CTL 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 GC 920 0 C 920 0 C 920 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 C 90 0 0 0
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L KAACGAAACGCCATGCGAAAGAACTTT T K R H A K E LCAATGTTTTTCTTCATATGACAGTAGAT P1CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V GCGATAACATCAGAAAAGCAGTAGGTT D N I R K A V GAACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S LAGAATTGGTATATGATCTTGCCGGAT E L V Y D L A GTCTTGTCGGATTTCATCAAATCAATA L V G F H Q I NTGATGTAGCCCCCATTTATTCAAAAGCA D V A P F I Q KCAAATACAGAACCGGGTTATATATCAA K Y R I G L Y IAAAGAAACCAATTGTAGTGGGGAAATCAGAACAGCC I G D I N R N SGATCGGGGCTGAAATTTAACAGAAACAGGAC I G D I N R N SGCCGGCTGAAATTTTAACAGAAACAGGAA P R L L K E T G	1710 ACGGCAGA D G R 1830 TAMAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CAAGTGAA T S E 2550 CTGGGAAATC G E I 2910 CGATTCTGG G AAATTCG G E I 2910 CGGTATTCGG G E I 2910 CGGTATCGTA G I V 3270	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGAC C TTGTTTC L V S IGCTCAGAC A Q T IGCTCAGAC K K T AGGCAAAAA G K K IACTGGTGA A A G I F G I F G CAAAGTAGA K V E IACAGGGGAT Q G H	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATCAGAAT S E AATCTCCGGCA GTTCTCCGGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTCCCCCCCCCGCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCA GTTCTCCGCCCA GTTCTCCGCCCA GTTCTCCGCCCA GTTCTCCCCCCCCCC	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V CCTCAG S S I CCTCAG S S I CCTCAG G E CATGTTC H V GAGAAAA G E CATCAC T N CATCAAC T S CATCAGCC H Q I CATCAGC H Q I CATCAGC S P	1740 ICTCTAAGCA L S K Q SGTAAGCTGC GGTAAGCTGC GGTAAGCTGC GGTAAGCAG D1190 GATGTAGCAG GL C K 2220 MAGCGTTTAC CACCAGATT L C K 2340 ITCCTGATTT. L P D L 2460 ICGCAGGAAT I K R 2700 I I I I R 3180 I I I R 3100 I I I I I R 3100 I I I I I R 3100 I I I I I I I I I I I I I I I I I I I	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACAG D K Q ACCGAAAACAG T G G CAACATATCA N I S CAACATATCA N I S CACAACGCCG T T P TGGAAAAGTG G K V	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGA P G G GACATCATT D I I AAAATCAA/ K I K AAAAATAA K K Y AATCCGGGA D Q A CAAAAATTC Q K F ATCGGTGCT I G A	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAAT( E I 2370 CAAATC/ G SAAAT( E 1 2370 CAAATC/ Q S 2490 ATTGAC I E 2610 ACGAAC Z730 CGAGC/ G A 2850 ACGAAC E K 2970 CTGCCC P A 3210 GTCACC V T 3330	CAGTATT SI CGTGCAC AGGAGCG PT CAAGGTG KV ATCGGT IG SATGAAT MN SCTGATC LI ACTTGGCC LA SGTTGCC VA SGTTGCC VA SACAAAA TK CCATGTA	GCAGG A C ACCAAAT G AGGAC T C GGAAA Y J GGAAC G C CCCAG F E CCACG H Y GCGATI R I CCACG F S CCACG F S CCACG F S C CCACG F S C CCACG F S C CCACG F S C CCACA F S C C CCACAA T C C C C C C C C C C C C C C C C C C C	ITGTCCI ITTCC ITTCC IGAGAGA I R IGAGAGA I R IGCAGAA I R IGCAGAA I R IGCAGA I R I R IGCAGA I R I R IGCAGA I R I R IGCAGA I R I R IGCAGA I R I R IGCAGA I R I R I R I R I R I R I R I R	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K CATTCC H S ATTGAA I E GATGAA D E TCCGAT S D TCAGAA S E GAAGTC E V ITGAAAA L K GACCCG D P	T ACCGA ACCGA T D I I AGTAT. I GAAAT I GAAAC E T GGAAAC E T GGGGA C G G G G G G G G G G G G G G G G	800 CT L 920C 0 920C 0
TCTTTATATTGCGAAGGAGTTAAAAAG L Y I A K E L KAACGAAACGCCATGCGAAAGAACTTT T K R H A K E LCAATGTTTTTCTTCATATGACAGTAG P1CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V GAACACAAGCGATTGAAACGAGTAGGTT T Q A I E T S LAGAATTGGTATATGATCTTGCCGGAT E L V Y D L A GTCTTGTCGGATTTCATCAAATCAATTA L V G F H Q I NTGATGTAGCCCCCATTTATTCAAAAGCAGTAGGTT CAAATACAGAATCGGGTTATATATCAAATCAGAATCAGAATCGGGTATATATA	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CGAGTGGAAATC G E I 2910 CGGTAATTS E 2550 CTGGGAAATCG G E I 2910 CGAGATTCTG R D S 2790 CGAGATTCTG G I V 3270 CGCAACGAA	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGAC C TTGTTTC L V S IGCTCAGAC A Q T IGCTCAGAC K K T AGGCAAAAA G K K IACTGGTGA A A G I F G I F G CAAAGTAGA K V E IACAGGGGAT Q G H	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GAACAGGCA GAACACTGC GAACACTGC CAACACTGC GAACCGGAAA CGATATTT	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V CCTCAG S S I CCTCAG S S I CCTCAG G E CATGTTC H V GAGAAAA G E CATCAG C L CTATGAA T S CATCAGC H Q I GATCAGC H Q I GCCCGA S P ATGGAA	1740 ICTCTAAGCA L S K Q SGTAAGCTGC GGTAAGCTGC GGTAAGCTGC GGTAAGCAG D1190 GATGTAGCAG GL C K 2220 MAGCGTTTAC GL C K 2220 MAGCGTTTAC CAGCAGGAAT T A G I 2340 ICCCTGATTT. L P D L 2460 CGGCAGGAAT T A G I 2580 IGATCAAAAG L K R 2700 CGGTAGAAAAGG L K R 2700 CGATCAAAAG I E K G 2940 CGATCAAAAG I Q N 3060 MAACCGTCAG E I V S 3180 ICATCAAAAG I Q N 33300	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACAG D K Q ACCGAAAACAG T G G CAACATATCA N I S CAACATATCA N I S CACAACGCCGG T T P TGGAAAAGTG G K V AAGCTGACTG	GCGGAAATC A E I CAGCTTATT <<<<<< AGAAGGTGT GAATATTT/ E Y L GATCCGCAT D P H CCTGGCGGA P G G GACATCATT D I I AAAATCAA/ K I K AAAAATAA K I K AAAAATAA K Y AATCCGGG/ D Q A CAAAAATTC Q K F ATCGGTGCT I G A CCGGAGTTT	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATI L I 2250 GAAAT(C E I 2370 CAAAT(C) GAAAT(C) G SAAT(C) C SAAT(C)	CAGTATT S I CGTGCAC AGGAGCG P T CAAGGTG K V ATCGGT CAAGGTG K V ATCGGT G ACTGGCG L A SGTTGCG V A SACAAAA T K CCATGTA H V CTTTTTT	GCAGG A C ATCC/ AGGAA C AGGAA T C GTAA/ V J GGAA/ V J GGAA/ C C C C C C C C C C C C C C C C C C	ITGTCC ITCG ITCG ITCG ITCG ITCG ITCG ITCG ITCG ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCC ITCCCCC ITCCCCC ITCCCCC ITCCCCC ITCCCCC ITCCCCC ITCCCCC ITCCCCCC ITCCCCCC ITCCCCCCCCCC	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K CATTCC H S ATTGAA I E GATGAA D E ICCGAT S D ICCGAT S D ICCGAA S E GAAGTC E V ITGAAA L K GACCCG D P CCCTCT	T ACCGA ACCGA T D I AGTAT. Z GAAAT Z GAAAC E T Z GGGGA C GGGA C GGGA C GGGA C GGGA C GGGA C GGGA C GGGA C GGGA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C GGAAA C C C C	800 CT L 920 G C P 0 G C P 1 CC Q G C P 0 CC Q C Q C Q C Q C Q C Q C Q C Q C Q C
TCTTTATATTGCGAAGGAGTTAAAAG L Y I A K E L KAACGAAACGCCATGCGAAGAACTTT T K R H A K E LCAATGTTTTTCTTCATATGACAGTA P1CGATAACATCAGAAAAGCAGTAGGTT D N I R K A V GAACACAAGCGATCAGAAAAGCAGTAGGTT T Q A I E T S LAACACAAGCGATTGAAACGAGTTGAAACGAGTTGATTGGTATATGGTATATGATCTTGCCGGATT T Q A I E T S LAGAATTGGTATATGATCTTGCCGGAT E L V Y D L A GTCTTGTCGGATTTCATCAAATCAATT L V G F H Q I NTGATGTAGCCCCATTTATCAAAGCA C D V A P F I Q KCAAATACAGAACCAGCATTGATAGGGGATATCAATACAGAACCAGATCGGGTTATATATC/ K Y R I G L Y IAAAGAAACCAATTGTAGTGGAAAACAGCA I G D I N R N SGCCGGGCTGAAATTTTAACAGAAACAGCAGAACAGCA P A E I L T V ITCCCAAGACTGTTGAAAGAAACAGGAACAGCA P R L L K E T GCTACGGTGTTCATATTGAATGGATGGAAACAGGAAGAACAGGAACAGCAGTATTAACAGAAACAAGAAACAGGAACAGCAATATGAAACAAGAAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAAGAACAGAACAGAACAGAACAGAACAAGAAACAAGAACAAGAACAAGA	1710 ACGGCAGA D G R 1830 TAAAACAA L K Q 1950 TAATTCTC L I L 2190 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CGGTAAAT S V N 2310 CGGTAAAT S V N 2310 TTCCAATT F P I 2430 CGGTAAAT S C 2790 CGGGAAATC G E I 2910 CG E I 2910 CC E I 2910 CG E I 2910 CC E I 200 CC E I	ACGACGAC T T T IGCGGATCA A D Q ICCGTCAGAC C TTGTTTC L V S IGCTCAGAC A Q T IGCTCAGAC K K T IGCTGCTGG A A G I F G I F G I F G CAAAGTAGA K V E IACAGGGGAT Q G M	ACGTGTCA R V AGTCAAAA V K AGGCAAAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GTTATTAA GAACAGAAT S L GAACACTGCA G T L GAACACTGCAAA C T L GAACACTGCAAA C T L GAACACTGCAAA C T L GAACACTGCAAA C T L GAACACTGCAAA C T L GAACACTGCAAA C T L GAACACTGCAAAA C T L GAACACTGCAAAAA C T L GAACACTGCAAAAA C T L GAACACTGCAAAAA C T L GAACACTGCAAAAA C T L GAACACTGCAAAAA C T L GAACACTGCAAAAA C T L GAACACTGCAAAAAAAA C T L GAACACTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AGCCGCC K P CAACTGC T T C TTAAATC GTGTAG S V CCTCAG S S C CCTCAG S S C CCTCAG S S C CCTCAG C C CTCTAC G C CATCAC T M CATCAAC T S CATCAGCC H Q CATCAAC T S CATCAGCC H Q CATCAAC T S CATCAGCC S P CATCAGCA S P ATGGAAA Y G C	1740 ICTCTAAGCA L S K Q SGTAAGCTGC GGTAAGCTGC GGTAAGCTGC GGTAGCAG D1190 GATGTAGCAG D1190 GATGTAGCAG G L C K 2220 MAGCGTTTAC G L C K 2220 MAGCGTTTAC G L C K 2220 MAGCGTTTAC G L C K 2240 ICCCGATTACAATT I C N 1 C K 2700 I C K 2700 I C K 2700 I C K 2800 I C K 28	GGAAAAGGTA E K V GCGAGAAGCG >> <<< CAAGTCATAA P L K AGTAAAGAAA V K K AAAAGTTATA K V I TGAAGCGGGC E A G TGATAAACAG D K Q ACCGAAAACAA P K T GACAGGCGGGT T G G CAACATATCA N I S CACAACGCCGG T T P TGGAAAAGTG G K V AAGCTGACTG S *	GCGGAAATC A E I CAGCTTATI <<<<<< AGAACTTATI GAATATTT/ E Y L GATCCGCAA D P H CCTGGCGGG/ P G G GACATCATT D I I AAAATCAA/ K I K AAAAATAA K I K AAAAATAA K N P G GATCAAGC/ D Q A CAAAAATTI Q K F ATCCGGTGCI I G A CCGGAGTTI >>>>	1770 GAGCGC E R 1890 11111C 2010 GGGAT/ 2130 CTGATTI L I 2250 GAAATC G E I 2370 CCAGCAC I E 2490 ATTGAC I E 2610 ACGAAC I E 2610 ACGAAC I E 2630 ATTGAC I E 2630 ACGAAC I E 2630 CCC ACGAAC I E 2630 CCC ACGAAC I E 2630 CCC ACGAAC I E 2630 CCC ACC I E 300 CCCC CCC I E 3000 CCCC CCC CCC I E 3300 CCCCC CCC CCC CCC CCC CCC CCC CCC	CAGTATT S I CGTGCAC AGGAGCG P T CAAGGTG K V ATCGGT I G CATGGAT L I ACTTGGC L I ACTTGGC L A SGTTGCG V A SGTGCG V A SACAAAA T K CCATGTA H V CTTTTTT	GCAGG A C ATCC/ AGGAA GAGGAA G A GGAAA V D GGAA/ V D GGAA/ V D GGAA/ C GGAA R D GCAGG C CCCAG/ R D GCAGAT R D GTTTTC G G D GTTTTC G G D C CCAGA C C C CCAGA C C C C C C C C C C	ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC: ITTCC:	GAAGTA E V ITCATC P2 AAGTAG <u>SPOI</u> GTATTT V F AAATCA K S CATTCC H S ATTGAA I E GATGAA J E TCCGAT S D TCAGAA S E GAAGTC E V ITGAAA L K GACCCG D P CCCTCT	T ACCGA ACCGA T D 1' AGTAT. 2 TGAAT TGAAT C GGAAA C C C GGTGG C G G G G G G G G G G G G G	800 CTL 920 GCP 10 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 280 CC 200 CC 280 CC 280 CC 280 CC 200 CC 280 CC 280 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CCC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 200 CC 20 CC 20 CC 200 CC 20 CC 20 CC 20 CC 20 CC 20 CC 20 CC 20 CC CC

FIG. 2. Sequences of recN and spoIVB loci (GenBank accession no. M30297). Identified transcription start sites are indicated as P<sub>1</sub> and P<sub>2</sub>. Potential terminators are indicated by arrows, and potential ribosome-binding sites are underlined. The sequence overlined and labeled oligo shows the extent of the primer used for determining the transcription start sites. Carets indicate residues similar to those in *lexA* boxes (Table 2). Underlined amino acids at residues 29 to 35 of *recN* are the ATP-binding site.

E.C. PECN MLAQLTISNFAIVRELEIDFHSGMTVITGETGAGKSIAIDALGLCLGGRAEADMVRTG-CRADLCARFSLKDTPARLRWLEENQLEDGHE-CLLRRVISSDGRSRGFINGTAVPLSQ B.S. PECN MLAELSIKNFAIIEELTVSFERGLTVLTGETGAGKSIIIDAISLLVGGRGSSEFVRYGEAKAELEGLFLLESGHPVLGVCAEQGIDVSDEMIVMRRDISTSGKSVCRVNGKLVTIAS

E.C. PECN AVTKHHQQALEIARALHQQRQQYAEELAQLITDSMHALSMPHGQFTIDVKF------DEHHLGADGADRIEFRVTTNPGQPMQPIAKVASGGELSRIALAIQVITARKM B.S. PECN ELDSVGKDVAVEAANVSQIRKTWAKKLADEIHRELKSLYMEKSTFDTEFKVRTASRNEEAPLVNGQPVQLTEQGIDLVKFLISTNTGEPLKSLSKVASGGELSRVMLAIKSIFSSQQ

E.C. FECN ETPALIFDEVDVGISGPTAAVVGKLLRQLGESTQVMCVTHLPQVAGCGHQHYFVSKETDGAMTETHMQSLNKKARLQELARLLVAVKSHVIHWRMRKNCLQRKLFSCFTVRVNSKTP B.s. FECN DVTSIIFDEVDTGVSGRVAQAIAEKIHKVSIGSQVLCITHLPQVAAMADTHLYIAKELKDGRTTTRVKPLSKQEKVAEIERSIAGVEVTDLTKRHAKELLKQ------ADQVKT-TG

FIG. 3. Comparison of putative recN gene (B.s.) with recN gene of E. coli (E.c.). Sequences were aligned by using the CLUSTAL 4 program of Higgins and Sharp (8). Symbols: \*, identical residues; · , conserved residues.

molecular weight of 45,975. Studies were undertaken to determine whether this potential protein corresponds to the product of the spoIVB gene (18) known to be in this region of the chromosome (5). Plasmids were constructed with various restriction fragments from this region in the integrative vectors pJM102 or pJM103 (Fig. 1) and were used to transform strains SL765 and 165.1 containing the spoIVB165 mutation. Plasmid pJB2001 was capable of transforming the spoIVB165 mutation to prototrophy. The gene in which a mutation gives rise to this stage IV phenotype is therefore the spo0A gene, the recN gene, or the unknown open reading frame. The genetic analysis was continued by transforming plasmids pJB2001, pJB2015, pJB2018, pJB2019, pJB2023, and pJB2025 into strains SL765 and 165 with selection for Cm<sup>r</sup>. If the wild-type allele of the spoIVB165 mutation resided in the donor plasmid, Spo<sup>+</sup> and Spo<sup>-</sup> tranformants would occur after a Campbell-type recombination event. We obtained Spo<sup>+</sup> and Spo<sup>-</sup> colonies after transformation with plasmids pJB2001, pJB2015, pJB2018, pJB2023, and pJB2025. Plasmid pJB2019 as donor gave only Spo<sup>-</sup> transformants. It was therefore concluded that the spoIVB165 mutation must be in the 200-base-pair region common to plasmids pJB2001, pJB2015, pJB2018, pJB2023, and pJB2025 and not carried by pJB2019. This 200-base-pair region is wholly contained within the amino-terminal end of the polypeptide encoded by the unassigned open reading frame, suggesting that it is the spoIVB locus.

In order to further prove that this open reading frame encoded the *spoIVB* gene, inactivation studies were carried

 TABLE 2. lexA box consensus sequences for three E. coli genes compared with B. subtilis recN gene

	lexA sequence						
Gene	Box 1	Box 2					
lexA	CTGTATATACTCACAG	CTGTAT ATACACCCAG					
recA	CTGTATGAGCATACAG						
recN	CTGTATATAAAACCAG	CTGTAC ACAATAACAG					
recNa	AA <u>G</u> GCGTA <u>AAAA</u> A <u>CAG</u> <sup>b</sup>	<u>TGAACT</u> G <u>C</u> TG <u>TAACAG</u>					

<sup>a</sup> Of B. subtilis.

<sup>b</sup> Underscoring indicates residues similar to those in at least one lexA box.

out by inserting the ermG gene into the open reading frame at two locations. The first insertional inactivation was made by placing the ermG gene into the unique BalI restriction site in plasmid pJB2001, which resulted in plasmid pJB2002 (Fig. 1). The same erythromycin gene was also inserted into the downstream *ClaI* restriction site in plasmid pJB2001, producing plasmid pJB2004 (Fig. 1). (Note that the upstream *ClaI* restriction site carried on plasmid pJB2001 is not cut by *ClaI* because of DNA methylation protection.) Both plasmids, pJB2002 and pJB2004, were linearized and transformed into strain JH642 with selection for erythromycin resistance, giving rise to strains JH12719 and JH12720. The Erm<sup>r</sup> transformants obtained from each transformation were the result of a double crossover event as determined by Southern blot analysis (data not shown).

Both strains carrying the insertional inactivations displayed the same phenotypic differences from the parental strain JH642. The mutant phenotype was that of a late-stage sporulation-defective strain as determined by observations of colonies on Schaeffer sporulation medium and by microscopic examination. Under the microscope, some phasegray forespores were evident after 48 h of growth at  $37^{\circ}$ C on Schaeffer sporulation medium. No phase-bright spores were observed in these strains. We also carried out a quantitative analysis of the sporulation efficiency in *spoIVB165* strains and our insertionally inactivated *spoIVB* mutant strains JH12719 and JH12720. Compared with the control strain, JH642, all *spoIVB* strains evaluated produce no viable chloroform-resistant spores (Table 3).

spoIVB transcriptional initiation site determination. The recN gene was followed by an inverted-repeat structure

TABLE 3. Efficiency of sporulation in various spoIVB strains

Strain	Total cells <sup>a</sup>	Spores <sup>a</sup>		
 JH642	$2.0 \times 10^{8}$	$1.2 \times 10^{8}$		
165.1	$6.8 \times 10^{7}$	0		
SL765	$4.0 \times 10^{7}$	0		
JH12719	$4.0 \times 10^{7}$	0		
JH12720	$1.1 \times 10^7$	0		

" Per milliliter of Schaeffer medium.



FIG. 4. Location of *spoIVB* transcription start site.  $P_1$  and  $P_2$  indicate the reverse transcripts obtained from mRNA prepared at 2 h ( $T_2$ ) or 3 h ( $T_3$ ) after the end of logarithmic growth. The primer extension experiments and the adjacent sequence ladder were both obtained by using the oligonucleotide primer shown in Fig. 2.

reminiscent of a terminator, which suggested that the *spoIVB* gene was not cotranscribed with *recN*. In order to determine the start site or sites of the *spoIVB* gene transcription, primer extension experiments using the oligonucleotide shown in Fig. 2 were performed on mRNA from strain JH642. mRNA was isolated from this strain at  $T_0$ ,  $T_1$ ,  $T_2$ , and  $T_3$  of sporulation. The precise location of the start site of transcription from each transcript was determined by running the reverse transcript next to a sequence of the promoter region with the same oligonucleotide primer. Two reverse transcripts were obtained (Fig. 4). Although both of these transcripts first appeared at  $T_2$ , only the smaller transcript remained visible by  $T_3$ .

Transcriptional analysis of spoIVB. In order to verify the point in the cell life cycle at which the spoIVB gene was transcribed, we constructed a lacZ transcriptional fusion with spoIVB. This was accomplished with the integrative lacZ transcriptional fusion vector pJM783. The 450-basepair StyI-SspI fragment containing the spoIVB promoter was inserted into the unique SmaI site adjacent to lacZ of pJM783. The resulting junctions were sequenced to ensure that the correct orientation of the spoIVB promoter with respect to the lacZ gene was obtained. A correctly constructed plasmid, pJB2026, was transformed into strain JH642 with selection for Cm<sup>r</sup>. The resulting Cm<sup>r</sup> colonies were light blue after 48 h at 37°C on Schaeffer sporulation medium containing 5-bromo-4chloro-3-indolyl-B-D-galactopyranoside (40 µg/ml). This strain, JH12713, was assaved for  $\beta$ -galactosidase production as a function of growth and sporulation. The results are shown in Fig. 5. The spoIVB gene was not transcribed until  $T_2$ , and transcription continued until at least  $T_4$ .

## DISCUSSION

Sequencing studies of the region of the chromosome between the *ahrC* and *spo0A* genes revealed the presence of two open reading frames of substantial size. The first of these coded for a 576-amino-acid protein with high homology to the *recN* gene product of *E. coli*. The highest homology was observed in the amino- and carboxyl-terminal portions of the proteins. The amino-terminal region also contained the GXXXXGK sequence characteristic of ATP-binding



FIG. 5. Time of expression of *spoIVB*.  $\beta$ -Galactosidase activity from the *lac* fusion to *spoIVB* and pJB2026 in strain JH12713 was determined as a function of time of sporulation.  $T_0$  is the end of the logarithmic phase.

sites (Fig. 2) which are present in RecN and several proteins involved in recombination and repair (20). The translation start site deduced from the sequence of the *recN* gene of *E*. *coli* was in some doubt, as two possible Met start codons were observed in the open reading frame (20). Our data suggest that the upstream start codon is the correct one, since the deduced *B. subtilis* protein was highly homologous to the deduced *E. coli* protein in the region between the two possible start codons. Furthermore, the downstream Met codon was not conserved in *B. subtilis*. The high homology and similar sizes of the two proteins are very suggestive that the *recN* equivalent in *B. subtilis* has been identified. However, no inactivation studies have been undertaken to prove this notion.

The recN gene of E. coli is under SOS control and regulated directly by LexA (24). The promoter for recN contains LexA-binding sites consistent with its control pattern (20). The B. subtilis recN gene has LexA-like boxes upstream of the coding region, and a system similar to SOS control has been observed in B. subtilis (13). Interestingly, the LexA-like boxes overlap the C-terminal reading frame of the ahrC gene, and no obvious inverted-repeat terminator structure is present. It seems possible that ahrB and recN are cotranscribed and that under SOS stress conditions, recN can be differentially expressed. No experiments were undertaken to identify the transcription start site of B. subtilis recN, so it is not known whether the LexA-like boxes are near the promoter for this gene.

Two sets of experiments identified the *spoIVB* gene. Transformation of the *spoIVB165* mutation by integrative plasmids carrying various restriction fragments located the mutation to within 200 base pairs in an open reading frame. Insertional inactivation of this open reading frame at two positions by an *erm* gene yielded Spo<sup>-</sup> transformants with the same characteristics as the strain carrying the *spoIVB165* mutation. The strains bearing the *spoIVB165* mutation and the *erm*-induced strains differ from the original description of the *spoIVB* mutant (4) in that they are not oligosporogenous (at least when tested by chloroform). We never found survivors to chloroform by using these strains under our cultural conditions. It seems possible that the original strain, P7, was partially suppressed for its sporulation defect, since the mutation gives complete sporulation deficiency when backcrossed to the sporulating parent of SL765. The classification of this mutation as one giving rise to a stage IV defect depends on the original classification of the P7 mutant, however.

Expression of the *spoIVB* locus occurs during sporulation, as determined from both mRNA isolation and *lacZ* fusion studies. mRNA for this locus first appeared at  $T_2$  of sporulation, when two transcripts were observed. The start sites are 22 base pairs apart in the region between *recN* and *spoIVB*. Neither promoter defined from the start sites had unequivocal homology to known promoter sequences, so it was not possible to assign a sigma factor responsible for this transcription. The upstream transcript was gone by  $T_3$ . A *lacZ* transcription fusion to this promoter region produced  $\beta$ -galactosidase beginning at  $T_2$  and continuing to at least  $T_4$ , which correlates directly with the mRNA studies. This timing of *spoIVB* expression is more characteristic of genes in which mutations cause a block earlier than stage IV of sporulation.

The spoIVB locus codes for a protein of 425 amino acids, with a molecular weight of 45,975. The protein is somewhat basic in that it has a calculated isoelectric point of 9.20. The function of the protein remains a mystery, and homology searches of sequenced proteins in the GenBank database did not reveal any homologies of significance. One-half of the first 20 residues at the amino terminus are hydrophobic, and this region lacks negatively charged residues. These properties are characteristic of signal sequences for secretion, or they might indicate the possibility of membrane association. The sequence I-K-V-T-G-K-K-S-G-E-S-E-L-V-Y beginning at codon 72 is predicted to form a helix-turn-helix structure (21), and it has some homology to other known regulatory proteins that use this structure for DNA binding. It seems fruitless to speculate further on function from primary sequence data without more knowledge of the properties of the protein. It will be of interest to see if this locus codes for a regulatory protein.

### ACKNOWLEDGMENTS

This research was supported in part by Public Health Service grants GM19416 and GM38843 from the National Institute of General Medical Sciences.

We thank Jonathan Day for performing the  $\beta$ -galactosidase assays and Marta Perego for reagents and advice.

### LITERATURE CITED

- Anagnostopoulos, C., and J. Spizizen. 1961. Requirements for transformation in *Bacillus subtilis*. J. Bacteriol. 81:741-746.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513–1523.
- Chen, E. Y., and P. H. Seeburg. 1985. Laboratory methods. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165-170.
- Coote, J. G. 1972. Sporulation in *Bacillus subtilis*. Characterization of oligosporogenous mutants and comparison of their phenotypes with those of asporogenous mutants. J. Gen. Microbiol. 71:1-15.
- Coote, J. G. 1972. Sporulation in *Bacillus subtilis*. Genetic analysis of oligosporogenous mutants. J. Gen. Microbiol. 71: 17-27.
- 6. Ferrari, E., D. J. Henner, M. Perego, and J. A. Hoch. 1988.

Transcription of *Bacillus subtilis* subtilisin and expression of subtilisin in sporulation mutants. J. Bacteriol. **170:289–295**.

- Ferrari, F. A., K. Trach, D. LeCoq, J. Spence, E. Ferrari, and J. A. Hoch. 1985. Characterization of the *spo0A* locus and its deduced product. Proc. Natl. Acad. Sci. USA 82:2647-2651.
- Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237-244.
- 9. Hoch, J. A. 1971. Selection of cells transformed to prototrophy for sporulation markers. J. Bacteriol. 105:1200–1201.
- Hoch, J. A. 1976. Genetics of bacterial sporulation. Adv. Genet. 18:69-99.
- Holmes, D. S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193–197.
- Losick, R., P. Youngman, and P. J. Piggot. 1986. Genetics of endospore formation in *Bacillus subtilis*. Annu. Rev. Genet. 20:625–669.
- Love, P. E., and R. E. Yasbin. 1984. Genetic characterization of the inducible SOS-like system of *Bacillus subtilis*. J. Bacteriol. 160:910-920.
- Miller, J. H. 1972. Experiments in molecular genetics, p. 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 15. Ninfa, A. J., and B. Magasanik. 1986. Covalent modification of the glnG product, NRI, by the glnL product, NRII, regulates the transcription of the glnALG operon in Escherichia coli. Proc. Natl. Acad. Sci. USA 83:5909-5913.
- North, A. K., M. C. M. Smith, and S. Baumberg. 1989. Nucleotide sequence of a *Bacillus subtilis* arginine regulatory gene and homology of its product to the *Escherichia coli* arginine repressor. Gene 80:29–38.
- 17. Perego, M., G. B. Spiegelman, and J. A. Hoch. 1988. Structure of the gene for the transition state regulator, *abrB*: regulator synthesis is controlled by the *spo0A* sporulation gene in *Bacillus subtilis*. Mol. Microbiol. 2:689–699.
- Piggot, P. J., and J. G. Coote. 1976. Genetic aspects of bacterial endospore formation. Bacteriol. Rev. 40:908–962.
- Ronson, C. W., B. T. Nixon, and F. M. Ausubel. 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell 49:579–581.
- Rostas, K., S. J. Morton, S. M. Picksley, and R. G. Lloyd. 1987. Nucleotide sequence and LexA regulation of the *Escherichia coli recN* gene. Nucleic Acids Res. 15:5041–5049.
- 20a.Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Sauer, R. T., R. R. Yocum, R. F. Doolittle, M. Lewis, and C. O. Pabo. 1982. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature (London) 298:447-451.
- Schaeffer, P., J. Millet, and J. Aubert. 1965. Catabolic repression of bacterial sporulation. Proc. Natl. Acad. Sci. USA 54:701-711.
- Trach, K., J. W. Chapman, P. Piggot, D. LeCoq, and J. A. Hoch. 1988. Complete sequence and transcriptional analysis of the spo0F region of the Bacillus subtilis chromosome. J. Bacteriol. 170:4194-4208.
- Walker, G. C. 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol. Rev. 48:60-93.
- 25. Youngman, P., J. B. Perkins, and K. Sandman. 1984. New genetic methods, molecular cloning strategies and gene fusion techniques for *Bacillus subtilis* which take advantage of Tn917 insertional mutagenesis, p. 103-111. In A. T. Ganesan and J. A. Hoch. (ed.), Genetics and biotechnology of the bacilli. Academic Press, Inc., New York.