Pea chloroplast DNA encodes homologues of *Escherichia coli* ribosomal subunit S2 and the β' -subunit of RNA polymerase

Alison L. COZENS and John E. WALKER*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

The nucleotide sequence has been determined of a segment of 4680 bases of the pea chloroplast genome. It adjoins a sequence described elsewhere that encodes subunits of the F_0 membrane domain of the ATP-synthase complex. The sequence contains a potential gene encoding a protein which is strongly related to the S2 polypeptide of *Escherichia coli* ribosomes. It also encodes an incomplete protein which contains segments that are homologous to the β' -subunit of *E. coli* RNA polymerase and to yeast RNA polymerases II and III.

INTRODUCTION

The chloroplast genome is a circular double-stranded DNA molecule, on average about 150 kb in size (for reviews see Bonhert et al., 1982; Whitfeld & Bottomley, 1983; Dyer, 1984, 1985). It codes for about 70 proteins. Some of these have been identified and are involved in photosynthesis. They include the large subunit of ribulose 1,5-bisphosphate carboxylase, a soluble enzyme located in the stroma, and components of the multi-subunit thylakoid membrane complexes, photosystems I and II, cytochrome b-f complex and ATP-synthase (reviewed by Walker & Tybulewicz, 1985). These protein components are made in the chloroplast by a transcriptional and translational machinery localized in the stroma. This apparatus is distinct from that used for production of transcripts and proteins from nuclear genes, and components of it are also encoded in the plastid DNA. For example, chloroplast ribosomal 23 S, 16 S and 5 S RNA species are transcribed from chDNA, as are all tRNAs required for chloroplast protein synthesis, although the number of different tRNAs involved has not been clearly established (review by Dyer, 1985).

Other evidence indicates that some of the ribosomal protein subunits are chloroplast gene products. In Chlamydomonas reinhardtii five or six out of 31-33 proteins of the large ribosomal subunit and 14 of 31 small subunit proteins are estimated to be products of chloroplast protein synthesis (Schmidt et al., 1983). In spinach at least seven to ten 30 S and six to thirteen 50 S ribosomal proteins (Dorne et al., 1984; Posno et al., 1984) and in pea at least six 30 S and five 50 S ribosomal components are thought to be chloroplast coded (Eneas-Filho et al., 1981). The presence in the chloroplast genomes of genes encoding homologues of E. coli ribosomal proteins has been demonstrated by hybridization and DNA sequence analysis. For example, homologues of E. coli ribosomal protein S19 have been found in chDNA from two species of Nicotiniana (Sugita & Sugiura, 1983; Zurawski et al., 1984). Moreover, the gene for a homologue of the L2 subunit is found associated with it in Nicotiniana debreyi and spinach (Zurawski et al., 1984). Also, a homologue of the *E. coli* S4 protein is encoded in maize chDNA (Subramanian *et al.*, 1983) and homologues of S12 and S7 are co-transcribed in *Euglena* gracilis with a homologue of *E. coli* elongation factor Tu (Montadon & Stutz, 1984; Passavant *et al.*, 1983).

As described in the present paper, during the analysis of a fragment of pea chDNA containing the gene for a subunit of chloroplast ATP-synthase, we have discovered that adjacent DNA sequences encode homologues of *E*. *coli* ribosomal protein S2 and part of the β' -subunit of RNA polymerase. This strongly suggests that these proteins are also produced from the chloroplast DNA.

MATERIALS AND METHODS

Isolation of the fragment of chDNA

A 6.2 kb SalI-PstI fragment derived from plasmid pPscS6, a pBR322 plasmid containing a 10.6 kb SalI fragment of pea chDNA (see Fig. 1; Huttly & Gray, 1984) was prepared as described previously (Cozens et al., 1986).

DNA sequence analysis

Random fragments 300-800 bp long were produced from a 6.2 kb SalI-PstI fragment by sonication. Then they were cloned into the M13 mp8 vector that had previously been digested with SmaI (Bankier & Barrell, 1983). Sequences in the resultant M13 clones were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) as modified by Biggin et al. (1983).

Data analysis

DNA sequences were compiled in a database with the computer program DBAUTO (Staden, 1982a). Each nucleotide in the sequence was determined six times on average and at least once on each strand of the DNA. The DNA sequences were analysed with ANALYSEQ (Staden, 1984). A number of options in this program were used to predict protein coding regions. Because of the high A+T content of the DNA the most effective procedure was that based on the positional base

Abbreviations used: chDNA, chloroplast DNA; kb, kilobase (pair); bp, base pair.

^{*} To whom correspondence and reprint requests should be sent.



Fig. 1. Physical and genetic maps of the pea chloroplast genome

The outer arcs indicate genes that have been identified by hybridization and DNA sequencing. The arrows show the direction of transcription. Gene names (see Dyer, 1985): 16 S and 23 S, ribosomal RNAs; P700 is the P700/chlorophyll *a* apoprotein of photosystem 1; 32 kDa, 44 kDa, D2 and cyt b_{559} are components of photosystem II; cyt *f*, cyt b_{563} and 15 kDa are subunits of the cytochrome *b*-*f* complex; LS is the large subunit of ribulose bisphosphate carboxylase; α , I, III, a, β and ϵ are components of ATP synthase (Cozens *et al.*, 1986); S2 and r pol β' are homologues of *E. coli* ribosomal subunits S2 and the β' subunit of RNA polymerase respectively.

preference method (Staden, 1985). This exploits the preference for bases to occupy particular positions in codons, this being a function of the frequencies with which particular amino acids are found in proteins; thus it can be used to distinguish coding and non-coding regions. The calculations were weighted for a strong preference for A or T in the third codon position. The predicted protein sequences were analysed with ANALY-SEP (R. Staden, unpublished work). This computer program contains a number of options for analysis of protein sequences including one used in the present study which calculates the M_r of a protein from its sequence. The protein sequences also were compared with database of the Protein Information Resource (PIR; National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road NW, Washington, DC 20007, U.S.A.) by using the program FASTP (Lipman & Pearson, 1985). Homologous sequences were analysed further by pairwise comparison with DIAGON (Staden, 1982b).

RESULTS AND DISCUSSION

DNA sequence

The DNA sequence presented in Fig. 2 is 4680 bases long. It extends from the SalI site and runs into the sequence encoding a membrane component of ATPsynthase which occupies most of the remainder of the SalI-PstI fragment (Cozens et al., 1986). The A+T content of this sequence is 66.05%, corresponding to overall values of 61-67.3% for pea chDNA (Herrmann & Possingham, 1980). Open reading frames (potential genes) could be recognized in non-coding regions in both strands of the DNA. Their locations were confirmed by calculations employing the positional base preference method as explained above (see Fig. 3).

One complete open reading frame (bases 3749-4459) encodes a protein of 236 amino acids with an M_r of 26842. It is preceded by an incomplete open reading frame (bases 1-3493) and is followed by the gene coding for a protein identified as a component of chloroplast ATP-synthase (base 4669 onwards; Cozens *et al.*, 1986).

The ribosome component

By comparison of the predicted protein sequences, with protein sequences in the Protein Information Resource database, the complete open reading frame was found to encode a protein which is closely related in sequence to that of the component S2 of the small subunit of the E. coli ribosome (Wittmann-Liebold & Bosserhoff, 1981; An et al., 1981). Their close relationship is emphasized by pairwise comparison of their sequences with the aid of the computer program DIAGON (Fig. 4). It is supported by alignment of the two protein sequences (Fig. 5) in which the two proteins are identical in about 40% of their residues. This close relationship in primary structure and the conservation of 16 S rRNA (Schwartz & Kössel, 1980; Tohdoh & Sugiura, 1982; Graf et al., 1982) and other ribosomal subunits in bacteria and chloroplasts implies that the tertiary structures of the two proteins and their arrangements in the ribosome will also be similar. E. coli ribosomal protein S2 is an elongated molecule (Engleman et al., 1975; Tischendorf et al., 1975; Georgalis et al., 1981). It can be cross-linked to ribosomal proteins S1, S3, S5 and S8 (Traut et al., 1980) and also to translational initiation factor IF-2 (Bollen et al., 1975). Also, it can bind to 16 S rRNA (Littlechild et al., 1977). Immuno-electron microscopic studies have located S2 near the cleft between head and body of the 30 S subunit (Tischendorf et al., 1975; Lake, 1978). Several mutants in S2 have been isolated, some being linked to resistance to the antibiotic kasugamycin (Okuyama et al., 1974; Yoshikawa et al., 1975). The gene is located near minute 4 in the E. coli chromosome and is co-transcribed with elongation factor EF-Ts (An et al., 1981).

The component of RNA polymerase

The 3.5 kb region of DNA sequence at the 5' end of the fragment presented in Fig. 2 contains an open reading frame which starts beyond the 5' end and terminates at base 3493. It encodes a fragment $(M_r \ 133591)$ of a protein. Segments of this sequence (boxed in Fig. 2 and summarized in Fig. 6) are related particularly to the sequences of the β' -subunit of E. coli RNA polymerase (Squires et al., 1981; Ovchinnikov et al., 1982) and also more weakly to the related yeast RNA polymerases II and III (Allison et al., 1985). As summarized in Figs. 1 and 6, these relationships extend over five regions, I–V, of the pea chloroplast open reading frame. Between amino acids 150 and approx. 900, no relationship is found between the putative chloroplast protein and either the RNA polymerases or any other proteins in the database. The alignment with the E. coli RNA polymerase β' -subunit and the pea chloroplast protein requires that a large

RN	Ap	01y	meı	as	e su	bun	it	bet	ta'-	;	>]	[
R	L	V	E	V	V	Q	B	I	V	V	R	R	T	D	C G	T	I	R	G	I	S	V	N	T	R	N G	M	M	P	E	I	I	LI	Q	TL
TCGAC	TTG	TTG	AAG	GTG	STTC	AAC	CACA	\TT	FT	GTA	CGGG	CGA/	CAG	ATI	GCGG	TAC	CAT	COGI	GGG	ATT	ICTG	TA/	AC/	ACCC	GAA	ATGO	TAA	GAT	GCCA	GAA	ATA/	ATTT	TGAT	ACAA	ACATT
		10			20)			30			40			50			60			70			80			90			100			110		120
I	G	R	V	V	A	E	N	I	Y	I	G	S	R	с	1 1	v	R	N	Q	D	I	G	I	G	L	IN	R	F	I	Т	F	Q	тQ	Р	IF
AATTG	GTC	GTO	TAC	TA	GCOG	AGA	AT/	ATA:	FAT.	ATA	GGT	rcad	GGT	GCA	TTGT	CGT	TAG	AAAJ	CAA	GAT	ATTG	GA/	ATT	GGAC	TTA	TCA/	TCG	ATT	CAT/	ACT	TTT	CAAA	CACA	ACCA	ATATT
	1	130			140)		1:	50			160			170			180			190			200)		210			220	,		230		240
]	Ι						
Γ	R	т	P	F	т	C	R	N	т	s	W	т	<u> </u>	D	1 0	- v	6	P	6	D	T	п	G	n	T	VE			-	A	v	G	TT	-	6 0
TATTA	GAA	СТО	cc	TTT	ACCI	GT/	NGG/	AAT	ACG'	TCT	TGG/	ATC	GCC	GAI	TGT	TTA	TGG	GCGO	AGT	CCT	ATTO		iGG	GACO	TGG	TAG/	ATT	GGG	GA/	GCT	GTAC	GGGA	TTAT	TOT	GGTCA
	2	250			260)		2'	70		:	280			290			300			310			320)		330			340			350		360
	-											_																		٦	-	_			
ATOTA	1	G	E	P	G	T	Q		T	L	R	T	F	H	T (i G	V	F	T	G	G	T		E	Y			<u> </u>	S	J N	G	K	IK	L	NE
AICIA	3	10AC	JAA		3 80			2 (2 (ACA 00	IIA	AU A	400	iiic	AIP	410	166	AG I	AII(420	AUA	666	430	ici (scit	3881 44(AIG	160	450 450		LIC	460	GGA	A ALG A	470	ALIG	480
					500				~			100			410			420			450				, 		450			400			470		400
D	L	V	H	P	Т	R	T	R	H	G	Y	P	A	F	IO	C N	I	D	L	Y	V	Т	I	E	S	DI) I	I	H	N	v	I	I P	Р	K S
GGATI	TAG	TT	CAC	CCL	ACAC	GT/	ACAO	CGT	CAT	GGA'	TAT	CCTO	GCTT	TT/	TAT	TAA	TAT	AGA	TTG	TAT	GTA/	ACT /	ATT	GAA	GTG	ACG/	TAT	TAT	ACAT	TAAC	GTC	ATTA	TTCC	ACCA	AAAAG
	4	190			500)		э.	10			520			530			540			550			200	,		570			280			590		600
F	L	L	v	Q	N	D	Q	Y	v	K	S	Ε	Q	v	I	E	I	R	A	G	Т	Y	Т	F	N	LI	КΕ	R	v	R	K	H	I Y	S	DS
TTTTC	TAT	TAC	TT	CAA	AATG	ATC	CAAT	FAT	GTA	AAA	TCA	GAA	CAAG	TG/	TTG	œga	GAT	CŒG	CG CG	GGA	ACAT	rat/	ACT	TTT/	ATT	TG A /	A GA	G AG	GGT	rcga	AAA	CATA	TTTA	ттст	GACTC
	6	510			620)		6	30		0	640			650			660			670			680)		690			700	l .		710		720
EO	G	E	M	H	W	S	T	D	V	Y	H	A	S I	E _ :	FM	Y	S	N	V	H	I	L	P	K	T S	S H	L	W	I	L	S	G I	K S	С	RS
AGAAG	GGG. 7	20	TGC	ATI	GGA	GTA	CCG	ATG	TGI	racc	ATG	CAT	CAG	A T	TTAT	GTAI	TAGT	AAT	GTAC	ATA	TCT	TAC	CAA		CAAC	TCA	TTT	ATGO	ATA	TTG	ſĊAG	GAA	AGTCO	TGC	CGGTC
	'	30			740			13	U		'	00			//0			80			90			800			810			820			830		840
N C	r	I	H	F	LI	L	R	K	D.	Q	D	Q	I 1	Г	M D	S	L	S	N	G	ĸ	т	N	I	s 1	I L	L	Е	R	Ν	D	Q	V K	H	KL
TAATA	CAA	TCC	ATT	TT	TAC	TTC	GCA	AGG	ATC	CAAG	ATC	AAA	TTA	CCA	TGGA	TTC/	\CT1	TCG	AATG	GA/		CCA	ATA	TTT	CGA/	ATCT	TTT/	AGA/	AGG	AAT	FATC	'A AG'	ГААА/	CAT/	AAATT
	8	50			860			87	0		8	80			890		9	000		9	10			920			930			940		1	950		960
FI	R	F	N	т	F	G	т	ĸ	E	ĸ	G	T	S I	. .	y s	т	F	N	E	T	т	С	т	D	R 9	x x	р		т	F	н	л,	r F	v	FI
ATTTC	TT	TCA	ATA	CTI	TTG	GTA	CAA	AAG	AAA	AG G	- GGA	TTA	GCG/	\TT	ATTC	AAT/	- \TT	TAAT	GĂAA	TCA	TAT	GTA	CGG	ATC	ATTO	GTA	тсс	сст	ATT	TTT	CACG	ATA	CTTT	TAT	гтстт
	9	70			980			99	0		10	00		1	010		10	20		10	30		1	.040		1	050		1	060		1	070		1080
		D	D	ъ	N 1	D	F		Ŧ	n	P	n	F (•		~								~			••	~	-	-	-			~	
		л Пол	R G & A	R G A I		R Gat	г ТТС	і. Тта	1 TTC	ראחי ראחי	r TTY	Р 'САТ	r (TCC/	Х 1 а т	С 4 Т СС 4 Т	Q TCA	E	K	A AGA			K GAA	M TCT	5 222	р 2 (тт(s G	V TCT/	3 	ב דיד אי		1	P		G	I F
000014	10	90 90	UAA	1	100	UNI	110	111	0	-un i	11	20	100	1	130	ICA	11	40	1.101	11	50	044	101	160		1	170		1	180	11 AC	1	190	0017	1200
																										-			-			-			1200
_	_		_	_	_																														
HI	R I	N	S	I	F	A	Y	F	D	D	P	Q	Y I	R	RH	S	S	G	I	T	K	Y	R	T	IC	} I	H	S	I	F	Q	K	E D	F	IE
TCATA	344	ATA	GTA	TTI	TIG	СТТ	ATT	TCG	ATG	ATC			ACAG	GAC 1	GACA	TAG	TCA	GGA	ATTA		AAT	ATA	GAA	CTA	TAGO	GAAT	TCAT	FTCO	TTA	TTTO	CAAA	AAG	A AGA 1	TTC/	TTGA
	12	10		1	.220			123	v		12	40		1	230		12	.00		12	.70		1	. 2 80		1	290		1	300		1	910		1320
Y I	R (G	I	K	E I	L	K	Р	K	S	Q	I	Q	V	DR	F	F	F	Ι	Р	E	E	V	H	II	P	K	S	S	S	L	M	V R	N	N S
GTATCO	GAG	GAA	TCA	AAG	AAT	TAA	AG C	CAA	LAA'I	ICTO	AAA	TTC	AAG	ГAG	ATCG	ATT	TT	TTT	ATTO	cα	AAG	AAG	TGC	ATA	TTT	LVCC	CAA	ATCI	TCT	TCC	CTAA	TGG	FACGO	AAT	ATAG
	13	30		1	.340			135	0		13	60		1	370		13	80		13	90		1	400		1	410		1	420		1	430		1440

L V G I G T P I T F N I R S R V G G L V R L D K K K K I E L K I F S G N I H F

P G E M D K I S R H S A I L I P P G T V K K K K C N K S K K I K N W I Y V Q W I

A T T K K K Y F V L V R P V I L Y E I P D S N N F V K L F P Q D L F Q E K D N L CGCAACTACCAAAAAAAGTATTTTGTTTTGGTTCGACCTGTCATTTTATATGAAATACCGGACAGTAACAATTTTGTAAAACTTTTTCCCCAAGATCTATTCCAGGAAAAGGATAATCT

ELKVVNYILYGNGKSIRGISDTRIQLVRTCLVFNWDDGKN GGAACTAAAAGTTGTCAATTATATTCTTTATGGAAAATGGAAAATCCATTCGGGGAATTTCCGACACGAGGATTCAATTAGTTCGGACTTGTTTAGTCTTCAATTGGGATGACGGCAAAAA

S S	S S	Ι	Е	E /	A P	A	S	F	I	E	VF	R I	N	G	LJ	E	Y I	ΓL	R	Ι	D	Ľ	V K	S	N	Т	S Y	I	R	K	RN
CAGTTO	TTCG	ATT	GAGO	AGGO	CCCC	жст	TCCI	TT/	TTG	AAG	TACO	GTAC	AAA	IGGT	TTGA	TGAG	TATT	TCT	AAGA	ATT	ACT	TAG	TAAA	ATCG.	AAT/	CTT	CATA	TAT	TAGA	AAAA	GAAA
	1930)	1	940		19	50		19	60		19	070		1980)	19	90		2000	D		2010		20)20		20	30		2040
FF	2 (G	F	с т	т	G	n	N	r	s	ח נ		N	P	F I	s	T	I S	ĸ	G	ĸ	т	0 0	S	L	s	0 N	н	G	т	IR
TGAAC	ATCT	GGT	TTCG	GAT	GAT	ngGG	GAT/	NATA	TAA.	n CGGG	ATCO	TAA7	CAA'	TCCA	TTTT	ттст	ATTC		CAAG	GGC/	1	TTC	ĂACĂ	ATCA	CTT/	GCC		TCA	CGGA	ACTA	TTCG
	2050)	2	060		20	70		20	080		20	90		210)	21	10		2120	0		2130		21	140		21	50		2160
ML	, L	N	R	N	E	С	R	S	W	I	II	L S	S	S	N C	F	Q	A R	P	F	N	N I	E K	S	H	N	GΙ	K	K	D	ΡI
TATGTI	GTTG	AAT/	AGAA	ATA/	AGA/	TGC	CGAT	гстт	IGGA	TAA	TTT	IGTO	ATC/	ATCT.	AATT	TTTT	'CAAA'	IG AG /	ACCA	TTC/	ACA	ATG	****	ATCT	CACA	ATG	GGAT	****	****	GATC	CTAT
	2170)	2	180		21	90		22	00		22	10		2220)	22	30		2240)		2250		22	260		22	70		2280
т с	. .	N	N	N (т	~				• •		. NT	P	v (, ,	v		т	T		м .	Λ Τ	c	т	т		т	^	т	n 7
) I 18 8 TT	אן א א די א אי	N 8 8 T 8			L TTA		1	A 2000	ь ттс		י א רידרים	\ IN YG & &1	r mmr	і і Татт <i>і</i>	, ст 1 (тт	L L TATC	I L	1 1		п Гата	ATC	ACAT	e Antro	1 1	ТТА		TTD	GCAA	стта стта	
	2290		2	300		23	10	1110	23	20		23	30		2340)	23	50		2360	0		2370		23	380		23	90		2400
			-																		-										
LI	Е	I	F	QI	/ I	K	Y	Y	L	M	DE	ΕN	I D	K	I	C K	P) L	Y	S	N	I	ΙL	N	P	F	HL	N	W	F	FL
ATTAAC	GGAA	ATT	TTTC	AAGI	TAAT	FAA A	TATI	FAT1	TTAA	TGG	ACG/	AAA /	TGA?	ГААА	ATTI	TAAA	CCCG.	ATCL	ATAC	CAGT	AATA	TCA	TTTT	GAAT	CCAI	TTCC	ATTI	GAA	TTGG	TTTI	TTCT
	2410)	2	420		24	30		24	40		24	150		246)	24	70		2480	0		2490		2:	500		25	10		2520
нн	F	Y	С	E I	ст	F	т	R	T	S	1. C	G () F	т	C I	2 N	T	ст	A	0	м	ĸ	NR	Р	н	L	K L	ĸ	S	G	Q V
CCATCA	TTT	TAT	IGTG	AAA		- TTT	- ACA/	MGA/	TA	GTC	TTG	GACA	ATT	- FATT	TGTG	TAAA	TATA	GTAT.	AGCI	CAA	ATGA		ATAG	ACCA	CACO	TAA	AACT	-	ATCG	GGTO	AAGT
	2530)	2	540		25	50		25	60		25	570		2580)	25	90		260	0		2610)	26	520		26	30		2640
							_																								
_ I]	[V	Q	M	DS	<u>s</u> v	I	I	R	S	A .	NI	P 3	L	A	TI	P G	A '	[]	H	G	н	Y	<u>G</u> E	I	L	S	QG	D	I	L	VT
TATAAT	AGTI	CAA	AIGO	ATT	TGT	ATA	ATA	AGAI	rcgo	CTA	ATCO		ATTR	GCA	ACTC	CAGGA	GCAA	CCAT	TCAC	JGGC		AIG	GGGA	GATC	CIT		CAAGO	iAGA	TATI	TIA	2760
	2030	,	4	000		20	70		20	080		20	90		270	,	21	10		212	0		2130		2	/40		21	30		2760
									Ш																						
FI	Y	E	K	S I	<u> </u>	G	D	I	Т	Q	GI	LI	ĸ	v	E () I	L	ΕI	R	S	I	D	S 1	S	M	N	LE	2 K	R	I	DA
ATTCAT	TATAT	GAA	AAAT	CGAC	ATC	CGGT	GAT/	ATA/	ACAC	AAG	GTC	гтсо	CAAA.	AGTG	GAAC	GATA	TTAG	AAAT	ACGT	TCG	ATTG	ATT	CAAT	ATCO	ATG	AATC	TAG/		AAGA	ATT	GATGC
	2770)	2	780		27	90		28	800		28	310		282	0	28	30		284	0		2850)	. 21	860		28	70		2880
W	NE	С	I	т	K I	I	G	I	Р	W	G	F	LI	G	A	EL	т	I /	A Q	S	R	I	s 🖸	LV	N	ĸ	I	0	K V	Y	RS
W TTGGA	N E Acga	C GTGI	I TATA	T ACA A	K I Aaat	I TATO	G CGGC	I ATT	P CCT	W TGG	G Ggat	F	L I Tgat	G TGG	A IGCIG	E L Agct	T AACT/	I A	A Q CGCA	S AAGT	R	I ATT	S [] ICIT	L V TGGT	N FAAT	K	I ATCC	Q 1	K V Aggt	Y TTAT	R S CGATC
W TTGGA	N E ACGA0 289	C GTGI 0	I TATA	T ACA A 2 90 0	K I AAAT	I TATO 29	G CGGC 910	I ATT	P CCT 2	W TGG 920	G GGAT	F TCT 2	L 1 TGA1 930	G TGG	A IGCIG 294	E L Agct	T AACT/ 29	I A ATCGO 950	A Q CGCA	S AAG1 296	R CGT	I ATT	S ICTT 297	L V TGGT 0	N ГААТ 2	K TAAA 980	I ATCC	Q 1 AAA 2	K V AGGT 990	Y TTA1	R S CGATC 3000
W TTGGA	N E ACGA 289	C GTGI 0	I TATA	T ACA A 2 90 0	K I AAAT	I TATO 29	G GGC 910 Π7	I ATT	Р ССТ 2	W TGG 920	G GGAT	F TCT 2	L 1 TGA1 930	G TGG	A IGCIG 294	E L Agct 0	T AACT/ 29	I A TCGO 50	A Q CGCA	S AAGT 296	R CGT	I ATT	S [] ICIT 297	L V TGGT 0	N ГААТ 2	K 7AAA 980	I ATCC	Q 1 AAA 2	<u>k v</u> Aggt 990	Y TTAT	R S CGATC 3000
W TTGGA	N E ACGA 289	C GTGI 0	I TATA	T ACAA 2900		I TATO 29	G GGC 910 IV	I ATT	Р ССТ 2	W TGG4 920	G GGAT	F TCT 2	L 1 TGA1 930	G TGG	A IGCIG 294	E L AGCT 0	T AACT/ 29	I A TCGO 50	A Q CGCA	S AAGT 296	R CGT	I ATT:	S [] ICIT 297	L V TGGT 0	<u>N</u> ГААТ 2	K 7444 980	I ATCC	Q 1 AAA 2	<u>K V</u> AGGT 990	TTAT	R S CGATC 3000
W TTGGA	N E ACGA 289 G V	C GTGI 0 H	I TATA	T ACAA 2900 <u>H</u>	K I AAAT N R	I TATO 29		I ATT E	P CCT 2 1	W TGG 920	G GGAT V	F TCT 2 R	L 1 TGAT 930	G TGG T	A rgCrg 294 S	E L AGCT 0 K V	T AACT/ 29	I A ATCG(950 V S	A Q CGCA S E	S AAGT 296	R CGT 50 G	I ATT: M	S ICIT 297 S	L V TGGT 0 N I	N F	<u>K</u> 7AAA 2980 L	I ATCC	Q 1 AAA 2 G 1	K V AGGT 990 E L	TTAT	R S CGATC 3000 G L
W TTGGA CCAGG	N E ACGA 289 <u>G V</u> GAGT 301	C GTGI 0 H ACAC	I TATA I CATC	T ACAA 2900 <u>H</u> CATA	K I AAAT N R ATAG	I TATO 29 H ACA	G GGC 910 IV I TATC	I ATT E GAG	P CCT 2 I ATT	W TGG 920 I ATT	G GGAT V GTGC	F TCT 2 R GTC	L 1 TGAT 930 Q 1 AAAT	G TGG TTGG T TAAC	A IGCTG 294 S ATCCA	E L AGCT 0 K V AAGT	T AACT/ 29] L CTTGO	I A TCG 50 V S TTTC	A Q CGCA S E CAGA	S AAGT 296 D AGAT	R CGT 60 G GGA	I ATT M ATG	S [] ICIT 297 S ICIA	L <u>V</u> TGGT 0 N I ATAT	N FAAT F ATTT	K FAAA 2980 L TTTA	I ATCC P CCTG	Q AAA 2 GCG	<u>k v</u> Aggt 990 E L AACT	TTAT TTAT	R S CGATC 3000 G L CGATT 3120
W TTGGA Q CCAGG	N E ACGA 289 <u>G</u> V GAGT 301	C GTG1 0 H ACAC	I ATA I CATC	T ACAA 2900 H CATA 3020	K I AAAT N R ATAG	I TATO 29 H ACA 30		I ATT E GAG	P CCT 2 1 ATT 3	W TGG 920 I ATT 040	G GGAT V GTGC	F TCT 2 R GTC 3	L 1 TGAT 930 Q 1 AAAT 050	G TGGT TTGGT	A rgctrg 294 <u>S</u> Atcca 306	E L AGCT 0 <u>K V</u> AAGT	T AACT/ 29] L CTTG(30	I A TCG(950 V S TTT(970	A Q CGCA S E CAGA	S AAGT 296 D AGAT 308	R CGT 60 G GGA	I ATT M ATG	S ICTT 297 S ICTA 309	L V TGGT 0 N I Atat.	N [AA1 2 F ATT] 3	K 2980 L TTTA 2100	I ATCC P CCTG	Q AAA 2 G GCG 3	<u>K V</u> AGGT 990 E L AACT 110	TTAT TTAT	R S CGATC 3000 G L CGGATT 3120
W TTGGA Q CCAGG	N E ACGA 289 <u>G</u> V GAGT 301	C GTGI 0 H ACAC 0	I TATA	T ACAA 2900 H CATA 3020	K I AAAT N R ATAG	I TATO 29 H ACA 30	GGC 910 IV I TATC 030	I ATT E GAG	P CCT 2 I ATT 3	W TGG0 920 <u>1</u> ATT0 040	G GGAT V GTGC	F TCT 2 R XGTC 3	L 1 TGAT 930 Q 1 AAAT 050	G TGGT TGGT	A IGCTG 294 S ATCCA 306	E L AGCT 0 <u>K V</u> AAGT 0	T AACT/ 29] L CTTGO 30	I A TCG(50 V S TTTC 070	A Q CGCA 5 E CAGA	S AAGT 296 D AGAT 308	R CGT 0 G GGA	I ATT M ATG	S ICTT 297 S ICTA 309	L V TGGT 0 N I Atat. 0	N FAAT F ATTT	K 7AAA 980 L TTTA 100	I ATCC P CCTG	Q AAA 2 G GCG 3	K V AGGT 990 E L AACT 110	Y TTAT	R S CGATC 3000 G L CGGATT 3120
W TIGGA Q CCAGG	N E ACGAG 289 <u>G V</u> GAGT 301 R A	C GTGI 0 H ACAC 0 E	I ATA I ATC R	T ACAA 2900 <u>H</u> CATA 3020 T	K I AAAT N R ATAG	I TATO 29 H ACA 30	G GGC 910 IV I TATC 030	I ATT E GAG E	P CCT 2 I ATT 3 E	W TGG 920 I ATTO 040 A	G GGAT V GTGC I	F TCT 2 R SGTC 3 C	L 1 TGAT 930 Q 1 AAAT 050 Y F	G TGG TTGG TAAC	A rgctrg 294 S Atcca 306	E L AGCT 0 <u>K V</u> AAGT 0 L L	T AACT/ 29] L CTTGO 30 G	I A TCG 50 V S TTTC 070	A Q CGCA S E CAGA	S AAGI 296 D AGAI 308	R CGT 50 G GGGA 30 S	I ATT M ATG	S [1CTT 297 S 1CTA 309 N	L V IGGT 0 N I Atat. 0 T Q	N 7AA7 2 F AT71 3 S	K FAAA 2980 L TTTA 100 F	I ATCC P CCTG	Q AAA 2 GCG 3 S	K V AGGT 990 E L AACT 110 E A	TTAT TTAT	R S CGATC 3000 G L CGGATT 3120 F Q
TTGGA Q CCAGG L ATTGO	N E ACGA 289 G V GAGT 301 R A GAGC	C GTGI 0 H ACAC 0 E AGAA	I ATA I ATC	T ACAA 2900 H CATA 3020 T ACGG	K I AAAT N R ATAG G R GGCG	I TATO 29 H ACA 30 TGC	G G G G G G G G G G G G G G	I ATT E GAG E GAA	P CCT 2 I ATT 3 E GAA	W TGG 920 I ATT 040 A GCA	G GGAT V GTGC I ATTT	F TCT 2 R SGTC 3 C C TT	L I TGAT 930 Q I AAAT 050 Y F ACAG	G TGGT TAACA	A rgCtg 294 s AtCCA 306 L	E L AGCT 0 K V AAGT 0 L L TATT	T AACTA 29] L CTTGO 30 GGGAQ	I A TCG 50 V S TTTC 070 V 2 TTAA	A Q CGCA S E CAGA <u>F K</u>	S AAGT 296 D AGAT 308 T	R CGT 0 G GGA 30 <u>S</u>	I ATT M ATG	S [] ICTT 297 S ICTA 309 N AATA	L V TGGT 0 N I ATAT 0 T Q CTCA	N F ATT1 3 AAG1	K FAAA 2980 L TTTA 100 F	P CCTG	Q AAAA 2 GCG 3 S CCG	K V AGGT 990 E L AACT 110 E A AAGC	TTAT TTAT AATT	R S CGATC 3000 G L CGATT 3120 F Q TTTCA
W TTGGA Q CCAGG L ATTGO	N E ACGA 289 <u>G V</u> GAGT 301 R A GAGC 313	C GTGI 0 H ACAC 0 E AGAA 0	I ATA I ATC	T ACAA 2900 H CATA 3020 T ACGG 3140	K I AAAT N R ATAG G R GGCG	I TATO 29 H ACA 30 TGCO 31	G GGC 910 IV I TATC 030 L CTTG 150	I E GAG E GAA	P CCT 2 I ATT 3 E GAA 3	W 1GG0 920 1 ATTO 040 A GCA 160	G GGAT V GTGC I ATTI	F TCT 2 R CTC 3 CC CTT 3	L 1 TGA1 930 Q 1 AAA1 050 Y F ACAG 170	G TGG TTGG T TAAC	A IGCIG 294 S ATCCA 306 L ITTAT 318	E L AGCT 0 K V AAGT 0 L L TATT	T AACTA 29] L CTTGC 30 <u>G</u> GGAC 31	I A TCG(50 V S TTTC 070 V T TAA(90	A Q CGCA S E CAGA <u>F K</u> CAAA	S AAGT 296 D AGAT 308 T AACA 320	R CGT 60 G GGGA 30 <u>S</u> ATCT 20	I M ATG	S ICTT 297 S ICTA 309 N AATA 321	L V TGGT 0 N I ATAT 0 T Q CTCA 0	N FAAT F ATTI 3 S AAGI	K 7AAA 2980 L 7TTA 100 F 7TTC 3220	P CCTG	Q 1 AAA 2 GCG 3 S CCG 3	<u>E</u> L 990 E L AACT 110 <u>E A</u> AAGC 230	Y TTAT AATT	R S CGATC 3000 G L GGATT 3120 F Q TTTCA 3240
W TTGGA CCAGG CCAGG	N E ACGA 289 <u>G V</u> GAGT 301 R A GAGC 313	C GTGI 0 H ACAC 0 E AGAA 0	I CATA I CATC R CGA	T ACAA 2900 H CATA 3020 T ACGG 3140	K I AAA1 N R ATAG G R GGCG	I TATO 29 H ACA 30 TGCO 31	G CGGC 910 IV I FATC 030 L CTTG 150	I ATT E GAG E	P CCT 2 I ATT 3 E GAA 3	W TGG0 920 1 ATTO 040 A GCA 160	G GGAT T GTGC I ATTT	F TCT 2 R GTC 3 C [C GTT 3	L 1 TGAT 930 Q 1 AAAT 050 Y F ACAG 170	G TGGT TAACA	A IGCIG 294 S ATCCA 306 L ITTAT 318	E L AGCT 0 K V AAGT 0 L L TATT 0 V	T AACT/ 29] L CTTGO 30 GGGAO 31	I A TCG(50 V S TTTC 070 V 1 TAA(90	A Q CGCA S E CAGA <u>F k</u>	S AAG1 296 AGA1 308 T AACA 320	R CGT 60 G GGGA 80 S ATCT 00	I ATT M ATG	S ICTT 297 S ICTA 309 N AATA 321	L V TGGT 0 N I ATAT 0 T Q CTCA 0	N FAAT 3 S AAGT 3	K 7AAA 2980 L TTTA 100 F TTTC 3220	P CCTG	Q 1 AAA 2 GCG 3 S CCG 3	<u>K V</u> AGGT 990 E L AACT 110 <u>E A</u> AAGC 230	TTAT TTAT	R S CGATC 3000 G L GGATT 3120 F Q TTTCA 3240
W TIGGA Q CCAGG L ATTGO	N E ACGA 289 GAGT 301 R A GAGC 313	C GTGI 0 H ACAC 0 E AGAA 0	I ATA I ATC R ACGA	T ACAA 2900 <u>H</u> CATA 3020 T ACGG 3140	K I AAAI N R ATAG G R GGCG	I TATO 29 H ACA 30 TGC(31	G CGGC 910 IV I TATC 030 L CTTG 150	I ATT E GAG E GAA	P CCT 2 I ATT 3 E GAA 3	W TGG 920 I ATT 040 A GCA 160	G GGAT V GTGC I ATTT	F TCT 2 R GTC 3 C [GTT 3	L 1 TGA1 930 Q 1 AAAA1 050 Y F ACAG 170	G TGGT TTGT TAAC/	A IGCTG 294 S ATCCA 306 L ITTAT 318	E L AGCT 0 $K V$ AAGT 0 $L L$ TATT 0 V	T AACT/ 29] L CTTGO 30 GGGAO 31	I A TCG(50 V S TTTC 070 V T TAA(90	A Q CGCA S E CAGA	S AAG1 296 AGA1 308 T AACA 320	R CCGT 60 G GGGA 30 <u>S</u> ATCT 00	I ATT M ATG	S ICIT 297 S ICTA 309 N AATA 321	L V TGGT 0 N I ATAT 0 T Q CTCA 0	N FAAT 3 S AAG 3	K 7AAA 2980 L TTTA 1100 F TTTC 3220	I ATCC P CCTG	Q 1 AAA 2 GCG 3 S CCCG 3 T	K V AGGT 990 E L AACT 110 E A AAGC 230 G E	TTAT TTAT AATT GAGT	R S CGATC 3000 G L CGATT 3120 F Q TTTTCA 3240
W TTGGA Q CCAGG L ATTGO	N E ACGA 2899 GAGT 301 R A GAGC 313 I A	C GTG1 0 H AACAC 0 E AAGAA 0 R TAGAA	I ATA ATC R ACGA	T ACAA 2900 <u>H</u> CATA 3020 T ACGG 3140 L	K I AAAT ATAG G R GGCG	I TATU 29 ACA 30 TGC 31	G CGGC 910 IV I FATC 030 L CTTG 150	I E GAG E GAA	P CCT 2 I ATT 3 E GAA 3 R	W TGG 920 I ATT 040 A GCA 160 G GT	G GGAT V GTGC I ATTT R CGTA	F TCT 2 R GTC 3 C [GTT 3 I	L 1 TGAT 930 Q 1 AAAT 050 Y F ACAG 170 D W	G G TGGT	A IGCTG 294 S ATCCA 306 L ITTAT 318 K GAAAG	E L $AGCT$ 0 $K V$ $AAGT$ 0 $L L$ $TATT$ 0 V $G L$ $GTCT$	T AACT/ 29] L CTTGC 30 GGGAC 31 K TAAA	I A TCGG 550 V S TTTC 070 V 1 TTAAC 190 E I	A Q CGCA S E CAGA <u>F K</u> CAAA	S AAGT 296 D AAGA1 308 T AAC/ 320	R CCGT 60 G GGGA 30 S ATCT 00 L	I ATT M ATG L CTG	S ICTT 297 S ICTA 309 N AATA 321 GGAA	L V TGGT 0 N I ATAT 0 T Q CTCA 0 M I TGAT	N FAAT 2 F ATTI 3 S AAG 3 P ACCC	K CAAA 2980 L TTTA 1000 F TTTC 3220 V	I ATCC P CCTG	Q 1 AAAA 2 GCG 3 S CCG 3 T	K V AGGT 990 E L AAACT 110 E A AAGC 230 G F	TTAT TTAT AAT GAGT	R S CGATC 3000 G L CGATT 3120 F Q TTTICA 3240 R I
TIGGA Q CCAGG ATTGO	N E ACGA 289 GAGT 301 R A GAGC 313 T A CTGC 325	C GTG1 0 H ACAC 0 E AGAA 0 R TAGA	I ATA I ATC R CGA V GTT	T ACAA 2900 <u>H</u> CATA 3020 T ACGG 3140 <u>L</u> ITAG 3260	K I AAAAT N R ATAG G R G G R G G CG A K CAAA	I TATO 29 H ACA 30 TGC 31 A AGCC 32	G CGGC 910 IV I FATC 030 L CTTG 150 A CTTG 150	I ATT E GAG E GAG	P CCT 2 I ATT 3 E GAA 3 R CGGG 3	W TGG(920 1 ATT 040 A GCA 160 GGT 280	G GGAT V GTGC I ATTT R CGTA	F TCT 2 GTC 3 C [GTT 3 I TTG 3	L 1 TGAT 930 Q 1 AAAT 050 Y F ACAG 170 D W ATTG 290	GTGGT	A IGCTG 294 S ATCCA 306 L ITTAT 318 K GAAAG 330	E L AGCT 0 K V AAGT 0 L L L TATT 0 V G L GTCT 0	T AACTA 29] L CTTGO 30 GGGAO 31 GGGGAO 31 K TAAAO	I A TCG(550 V S TTTC 70 70 TTAA 190 <u>E !</u> SAGAA	A Q CGCA S E CAGA CAGA N V ATGT	S AAGT 296 D AGAT 308 T T T T T T T T T T T T T T T T T T T	R CGT 50 G GGGA 30 <u>S</u> XTCT 00 <u>L</u> TTTA 20	I ATT M ATG CTG	S ICTT 297 S ICTA 309 N AATA 321 GGGAA 333	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0	N FAAT 3 S AAAGI 3 P P AACCO	K 7AAA 2980 L TTTA 3100 F TTTC 3220 V CGTT 3340	I ATCC P CCTG I ATAT	Q 1 AAA 2 GCG 3 S CCG 3 T CCG 3	K V AGGT 990 E L AACT 110 E A AAGC 230 G F GATI 350	Y TTAT AATT S GAGT	R S XCGATC 3000 G L XCGATT 3120 F Q TTTICA 3240 R I MGGATT 3360
TIGGA Q CCAGG ATIGO	N E ACGAA 289 G V GAGT, 301 R A GAGC, 313 <u>T A</u> CTGC 325	C GTG1 0 H ACAC 0 E AGAA 0 R TAGA	I ATA I ATC R CGA V GTT	T ACAA 2900 <u>H</u> CATA 3020 T T ACGG 3140 <u>L</u> TTAG 3260	K I AAAAT ATAG G R GGCG A K CAAA	H ACA 30 A AAGCO 31	G GGC 910 I I TATC 030 L CTTG 150 A CTTG 150	I ATT E GAG E GAA	P CCT 2 I ATT 3 E GAA 3 R CGG 3	W TGG 920 <u>1</u> ATTC 040 A GCA 160 <u>G</u> GGT 280	G GGAT FTGC I ATTT	F TCT 2 3 C [GTT 3 I VTTG 3	L 1 TGAT 930 Q 1 AAAT 050 Y F ACAG 170 D W ATTG 290	G TGGT TGGTT	A IGCTG 294 S ATTCCA 306 L L TTTAT 318 K GAAAC 330		T AACT/ 29] L CTTG(30 GGGA(31 GGGA(31 TAAM 33	I A TCGC 550 V S TTTC 770 TTAAC 90 <u>E 1</u> GAGA3	A Q CGCA S E CAGA <u>F K</u> CAAA	S AAGT 296 D AAGAT 308 T TAAC/ 320 7 V TGTT 332	R CCGT 50 G GGGA 30 S ATCT 50 C L TTTA 20	I M ATG <u>L</u> CTG	S [] ICTT 297 S ICTA 309 N AATA 321 GGAA 333	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0	N FAAT 3 S AAG 3 P AACCC	K CAAA 2980 L TTTA 3100 F TTTC 3220 V V CGTT 3340	I ATCC P CCCTG J ATAT	Q 1 AAAA 2 GCGG 3 S CCGG 3 T CCGG 3	K V AGGT 990 E L AACT 110 E A AAGC 230 G F GATT 350	Y TTAT AATT SGAGT	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MAGAAT 3360
TIGGA Q CCAGG ATIGO	N E ACGA 289 G V GAGT, 301 R A GAGC, 313 <u>T A</u> CTGC 325	C GTGI 0 H ACAC 0 E AGAA 0 R TAGA	I ATA I ATC R ACGA V GTT	T ACAA 2900 <u>H</u> CATA 3020 T T ACGG 3140 <u>L</u> ITTAG 3260	K I AAAAT ATAG G R GGCCG A K CAAA	I TATO 29 ACA 30 TGC 31	G GGGC 910 IV I FATC 030 L CTTG 150 A GGCT 270	I FATT GAG GAA E GAA	P CCT 2 I ATT 3 E GAA 3 R CGG 3	W TGG0 920 I ATT 040 A GCA 160 GGT 280	G GGAT FTGC I ATTT R CGTA	F TCT 2 R GTC 3 C [GTT 3 I VTTG 3	L 1 TGA1 930 Q 1 AAA1 050 Y B ACAG 170 D W ATTG 290	G TGGT TGGT AAGC	A rgCrg 294 S ATCCA 306 L TTTAT 318 K GAAAG 330		T AAACT/ 29] L CTTG(30 GGGA(31 GGGGA(31 S GGGA(31 S	I A TCGC 550 V S TTTC 070 V 7 TTAA 190 E I GAGA	A Q CGCA S E CAGA CAAA	S AAGT 296 D AAGAT 308 T T T T T T T T T T T T T T T T T T T	R CCGT 50 G G G G G G G S A TCT D 0 L TTTA 20	I ATT M ATG CTG	S [] ICTT 297 S ICTA 309 N AATA 321 GGAA 333	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 <u>M I</u> TGAT 0	N FAATI 3 S AAGI 3 P AACCC	K CAAA 2980 L TTTA 5100 F TTTC 3220 V V CGTT 3340	P CCCTG ATATAT	Q AAAA 2 GCCG 3 S CCCG 3 T CCCG 3	K V AGGT 990 E L AAGT 110 E A AAGC 230 G F GATI 350	I AATT GAGT	R S XCGATC 3000 G L YGGATT 3120 F Q TTTICA 3240 R I MAGAAT 3360
TIGGA Q CCAGG ATTGO AGAAA	N E ACGA 289 GAGT. 301 R A GAGC. 313 T A CTGC. 325 G	C GTGI 0 H ACAC 0 E AGAA 0 R TAGA 0 S	I ATA I ATC R ACGA V GTT	T ACAA 2900 <u>H</u> CATA 3020 T T ACGG 3140 <u>L</u> TTAG 3260 S	K I AAAT ATAG G R GGCG A K CAAA	I TATO 29 H ACA 30 31 TGC(31 31 AGC(32 32 H	G GGGC 910 IV I FATC 030 L CTTG 150 A CGCT 270 N	I EATT EGAG GAAG EGAA	P CCT 2 I ATT 3 E GAA 3 R CGG 3 I	W TGG0 920 I ATTO 040 A GCA 160 GGT0 280 T	G GGAT V FTGC I ATTT R CGTA	F TCT 2 3 C [3 TT 3 1 XTTG 3 K	L 1 TGAT 930 Q 1 AAAT 050 Y F ACAG 170 D W ATTG 290 K F	G TGGT TGGT AAC/ AGC	A rGCTG 294 S ATCCA 306 L TTTAT 318 K GAAAG 330 F	E L L AGCT 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	T AAACT/ 29] L CTTG(30 GGGA(31 GGGA(31 S GGGA(31 S C E	I A TTCG(550 V S TTTC(070 V 1 TTAA(90 E I AGAA 310 I I	A Q CGCA S E CAGA CAAA N V ATGT R N	S AAGT 296 D AAGAT 308 T T T T T T T T T T T T T T T T T T T	R CCGT 50 G GGGA 30 S XTCT D0 L TTTA 20 L	I ATT M ATG CTG	S [ICTT 297 S ICTA 309 N AATA 321 GGAA 333 H	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R	N FAATT 3 S AAGT 3 P P AACCC	K 7AAA 2980 L TTTA 3100 F TTTC 3220 V V CGTT 33340 L	I ATCC P CCTG ATATAT	Q AAAA 2 GCGG 3 S CCGG 3 T CCGG 3 D	K V AGGT 990 E L AAGT 110 E A AAGC 230 G F G F	Y TTAI AATI GAGI	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MAGAAT 3360 F K
TIGGA Q CCAGG ATTGO AGAAA	N E ACGA 2899 GAGT. 301 R A GAGC. 313 T A CTGC. 3250 H R	C GTGI 0 H ACACO 0 E AGAA 0 TAGA 0 S TTCA	I ATA I ATC R CGA V GTT R CGG	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S S CCAA	K I AAAAT N R G R GGCCG A K CAAAA R Q GGCA	I TATU 29 H ACA 30 TGCC 31 A AGCC 32 H ACA	G GGGC 910 IV I FATC 030 L CTTG 150 A CGCT 270 N CAAC	I E GAG E GAA E CTT K	P CCT 2 I ATT 3 E GAAA 3 CGGG 3 I ATT	W TGGG 920 1 ATTN 040 A GCA 160 GGTC 280 T T ACCC	G GGAT V FTGC I ATTT R CGTA R CGTA	F TCT 2 3 C [GTT 3 1 TTG 3 K	L 1 TGAT 930 Q 1 AAAAT 050 Y F AAAAG 170 D W ATTG 290 K F AAAA	G TGGT TAAC/ AGC	A rectro 294 s Attock 306 L L TTTAT 318 GAAAC 330 F F Attock	E L AGCT 0 <u>k V</u> AAAGT 0 <u>L L</u> GTCT 0 <u>G L</u> GTCT 0 0 E V AAGT	T AAACT/ 29] L CTTGG 30 GGGAA 31 K TAAAA 33 K E AGAAA	I A TCG(50 V S TTTC 070 V 1 TTAA 10 I I NTTAA	A Q CGCA S E CAGA CAGA N V ATGT R N GAAA	S AAGT 296 D AAGAT 308 T T T T T T T T T T T T T T T T T T T	R CCGT 50 G GGGA 30 S XTCT D0 L TTTA 20 L	I M ATT L CTG	S [] ICTT 297 S ICTA 309 N AATA 321 GGAA 333 H CATC	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG	N FAATTI 3 S AAAGI 3 P AACCO 5 K K AAAAA	K CAAA 2980 L TTTA 3100 F TTTC 3220 V V CGTT 33340 L L	I ATCC P CCCTG G G G G G G G G G G G G G T L	Q AAAA 2 GCCG 3 S CCCG 3 T CCCG 3 D D GATT	K V AGGT 990 E L AAACT 110 E A AAGC 230 G F GAT1 350 F A F A	TTAT TTAT AATT GAGT CAAA	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I VAGAAT 3360 F K TTTCAA
W TIGGA CCAGG ATTGO AGAAAA M AATGO	N E ACGA 289 G V GAGT 301 R A GAGC 313 T A CTGC 325 H R ACCG7 337	C GTGI 0 H ACACO 0 E AGAA 0 TAGA 0 S TTCA 0	I ATA I ATC R CGA V GTT R CGG	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S FCAA 3380	K I AAAAT N R ATAG GGCG GGCG A K CAAAA R Q GGCA	I TATU 29 H ACA 30 TGCC 31 A AGCC 32 H ACA 32 H	G CGGC 910 I I TATC 030 L CTTG 150 A CTTG 150 A CTTG 270 N FAAC 390	I E GAG E GAA L CTT K	P CCT 2 I ATT 3 E GAA 3 R CGG 3 I ATT 3	W TGG4 920 1 ATTX 040 A GCA 160 GGT0 280 T T ACCC	G GGAT V GTGC I ATTT R CGTA R CGGA	F TCT 2 GTC 3 C [GTT 3 I VTTG 3 K AAAA 3	L 1 TGAT 930 Q 1 AAAAT 050 Y F AAAA 170 D W ATTG 290 K F AAAA 410	G TGGT TAAC/ AGC	A rectro 294 s Attcca 306 L L TTTAT 318 GAAAC 330 F F Attcca 330 S Attcca 330 S Attcca 330 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S Attcca 5 S S Attcca 5 S S S S S S S S S S S S S S S S S S	$E L$ AGCT 0 $\frac{K V}{AAGT}$ 0 $\frac{L L L}{TATT}$ 0 $\frac{G L}{GTCT}$ 0 $E V$ AAGT 0	T AAACT/ 29] L CTTGG GGGAG 33 G GGAA 33 K TAAAA 33	I A TTCG 550 V S TTTC 770 V 1 TTAA (90 E 1 S AGAA 310 I I ATTAA	A Q GGCA S E CAGA CAGA N V ATGT R N GAAA	S AAAGT 296 D AAGA1 308 T T T T T T T T T T T T T T T T T T T	R CGT 60 G GGGA 30 S ATCT 20 L L TTTA 20 L L TTTG 40	I M ATG CTG GGGG F TTC	S ICTT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0	N FAAT 3 S AAAG 3 P AACCO 5 K AAAA	E TTTA 2980 L TTTA 3100 F TTTC 3220 V OGTT 3340 L ATTA 3460	I ATCC P CCCTG G G G G G G G G G G G G G G G G G	Q AAAA 2 GCCG 3 S CCCG 3 T CCCG 3 D D ATTT 3	K V AGGT 990 E L AAACT 110 E A AAGC 230 G F GATI 350 F A TTIGC 470	Y TTAT AATT AATT SGAGT	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480
W TTGGA CCAGG ATTGO AGAAA AGAAAA	N E ACGA 289 GAGT 301 R A GAGC 313 T A CTGC 325 H R ACCG 337(C GTGT 0 H ACAC 0 E AGAA 0 R TAGA 0 S TTCA 0	I ATA I ATC R CGA V GTT R CGG	T ACAA 2900 H CATA 3020 T ACGG 3140 L L TTAG 3260 S S CCAA 3380	K I AAAAT ATAG G R GGGCG A K CAAAA R Q GGCA	I TATO 29 H ACA 30 TGC 31 A AGCC 32 H ACA 33	G CGGC 910 I I TATC 030 L CTTG 150 A CTTG 150 A CTTG 270 N RAAC 390	I E GAG E GAA L CTT K	P CCT 2 I ATT 3 E GAA 3 R CGG 3 I ATT 3	W TGG4 920 1 ATT 040 A GGCA 160 GGT0 280 T ACCC 400	G GGAT V GTGC I ATTT R CGTA R CGGA	F TCT 2 GTC 3 C [GTT 3 I TTG 3 K MAAA 3	L] TGAT 930 Q] AAAAT 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410	G TGG TGG TGG T TAAC/ AGC	A rectro 294 306 <u>L</u> TTTAT 318 <u>K</u> GAAAC 330 F F ATTCC 342	E L L C C C C C C C C C C C C C C C C C	T AAACT/ 29] L CTTGG GGGAC 31 G GGGAC 31 K TAAAA 31 K AGAAA 34	I A TTCG 50 V S TTTC 70 V 1 TTAA (90 E 1 S AGAA 310 I I ATTAA	A Q GGCA S E CAGA CAGA N V ATGT R N GAAA	S AAAGT 296 D AAGA1 308 T T T T T T T T T T T T T T T T T T T	R CGT 60 G GGGA 30 <u>S</u> ATCT 00 <u>L</u> TTTA 20 L TTTG 40	I M ATG CTG GGGG F TTC	S [ICIT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345	L V TGGT 0 N I ATAT. 0 T <u>Q</u> CTCA 0 <u>M I</u> TGAT 0 H R ACAG 0	N FAAT 3 S AAG 3 P P AACCO 3 K K AAAA	E TTTA 3100 F TTTC 3220 V CGTT 33340 L ATTA 3460	I ATCC P CCCTG ATATAT	Q AAAA 2 G GCG 3 S CCCG 3 T CCCG 3 D CCCG 3 D CCCG 3 CCCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG CCCCG CCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCCC	K V AGGT 990 E L AAACT 110 E AACT 110 E G F G F ATTIGC 470	Y TTAT AATT AATT SGAGT CAAA	R S CGATC 3000 G L GGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480
W TIGGA CCAGG L ATIGO AGAAA AATGO	N E ACGAM 2899 G V GAGT. 301 R A GAGC. 3130 <u>F A</u> CTGC. 3250 <u>H R</u> ACCCG 3370 F M	C GTGT 0 H ACAC 0 E AGAA 0 E TAGAA 0 S TTCA 0	I ATA I ATC R CGA V GTT R CGG	T ACAA 2900 <u>H</u> CATA 3020 T ACGG 3140 <u>L</u> TTAG 3260 S FCAA 3380	K I AAAAT N R G R GGCCG A K CAAAA R Q GGCA	I TATO 29 H ACA 30 TGC 31 A AGCC 32 H ACA 33	G CGGCC 910 IV I TATC 030 L CTTG 150 A CTTG 150 N FAAC 390	I ATT E GAA E GAA CTT K	P CCT 2 I ATT 3 E GAA 3 R CGG 3 I ATT 3	W TGG4 920 1 ATT 040 A GGCA 160 GGT0 280 T ACCC 400	G GGAT V GTGC I ATTT R CGTA R CGGA	F TCT 2 3 C [TCTT 3 TTTG 3 K AAAA 3	L] TGA1 930 Q 1 AAAA1 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410	G TGG TGG TGG T CAAC/ AGC	A rectro 294 306 <u>L</u> TTTAT 318 <u>K</u> GAAAC 330 F F ATTCC 342	$E L AGCT 0 $ $\frac{E V}{AAGT} 0 $ $\frac{L L L}{TATT} 0 $ $\frac{G L}{GTCT} 0 $ $E V AAGT 0 $	T AAACT/ 29] L CTTGG 30 G GGA0 31 K TAAAA 33 K AGAAA 34	I A TCGG 50 V S TTTC 770 V 7 TTAA 190 E N AGAA 310 I I ATTAG	A Q CGCA S E CAGA CAGA N V ATGT R N GAAA	S AAGT 296 D AAGAT 308 T T T T T T T T T T T T T T T T T T T	R CGT G G G G G G G G G G G G G G C G C G	I M ATG CTG	S [ICIT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0	N FAAT 3 S AAG 3 P AACCC 3 K K AAAA	E CAAA 2980 L TTTA 1000 F TTTC 3220 V CGT1 3340 L L ATTA 3460	I ATCC P CCCTG ATATAT	Q AAAA 2 G GCG 3 S CCCG 3 T CCCG 3 D SATT 3	K V AGGT 990 E L AAACT 110 E A AAGC 230 G F GATT 350 F A TTIGC 470	Y TTAI AATI GAGI CAAA	R S CGATC 3000 G L GGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480
W TIGGA Q CCAGG ATTGC AGAAAA AATGC	N E ACGAM 2899 GAGT. 301 R A GAGC. 3130 T A CTGC. 3250 H R ACCOT 3370 F M	C GTGI 0 H ACAC 0 E AGAA 0 TAGA 0 S TTCA	I ATA I ATC R CGA V GTT R CGG CGG	T ACAA 2900 H CATA 3020 T ACGG 3140 L L TTAG 3260 S FCAA 3380	K I AAAAT ATAG GG R GGGCG A K CAAAA CAAAA	I TATO 29 H ACAT 30 TGCC 31 ACAT 32 H ACAT 32 H ACAT	G CGGCC 910 IV I TATC 030 L CTTG 150 A AGGA	I ATT E GAG E GAA CTT K AAG	P CCT 2 I ATT 3 E GAA 3 R CGG 3 I ATT 3	W TGG 920 I ATTO 040 A GCA 160 GGT 280 T ACC 400	G GGAT V GTGC I ATTT R CGTA R CGGA	F TCT 2 3 C [TCTT 3 C [1 TTTG 3 K AAAA 3	L] TGAT 930 Q] AAAT 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410	GGGA	A rgCrg 294 S ATCCA 306 L TTTAT 318 K GAAAC 330 F ATTCC 342 CTCAT	$E L AGCT 0 $ $\frac{K V}{AAGT} 0 $ $\frac{L L L}{TATT} 0 $ $\frac{V}{G L} GTCT 0 $ $E V AAGT 0 $ $CCCCC 0 $	T AACT/ 29] L CTTG(30 G GGGA(31 K TAAAA 34 34 TTAAA	I / TCGG 50 V 2 TTTTG 770 V 7 TTAA 190 E 1 AGAA 310 I 1 ATTAG 30 ATCT	A Q CGCA S E CAGA CAGA N V ATGT R N GAAA	S AAGT 296 D AAGAT 308 T TGT 332 V V TGT 332 AATG	R CGT G G G G G G G G G G G G G G G C G C	I M ATG CTG G GGG F TTC	S [ICTT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TTIGA	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0 TTTG	N FAAT 3 F S AAAG 3 P AACCC 5 S AAAG 3 S S AAAG 3 S S S S S S S S S S S S S S S S S S	к ТАААА 2980 L TTTA 1000 F TTTC 3220 V CGT1 3340 L L ATTA 3460	I ATCC P CCTG ATATATATATATATATATATATATATATATATATATA	Q AAAA 2 G GCG 3 CCCG 3 T CCCG 3 D D ATT 3	K V AGGT 990 E L AAACT 110 E A AAAGC 230 G F GAT1 350 F A TTIGC 470	Y TTA AATT AATT CAAA CAAAA	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K CTTCAA 3480
W TIGGA CCAGG ATIGO AGAAA AATGO AGAAT	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 T A CTGC. 3250 H R ACCOT 3370 F M FTATC 3490	C GTGI 0 H ACAC 0 E AGAA 0 TAGA 0 S TTCA	I ATA ATC ATC R CGA V GTT R CGG CGG	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S S TCAA 3380 S ATTA 3500	K I AAAAT ATAG GG R GGGCG A K CAAA CAAA	I TATO 29 H ACAT 30 TGC0 31 ACAT 32 H ACAT 33 TAT/ 32	G GGGC 910 I I FATC 030 L CTTG 150 A CGCT 270 N FAAC 390	I E GAA E GAA L CTT K AAG	P CCT 2 I ATT 3 E GAA 3 CGGG 3 I ATT 3 AATT 3	W TGG 920 1 ATT 040 A GCA 160 GGT 280 T ACC 400 GGT 520	G GGAT FTGC I ATTT R CGTA R CGTA	F TCT 2 R GTC 3 C C TTTG 3 K AAAA 3	L 1 TGA1 930 Q 1 AAA1 050 Y E ACAG 170 D W ATTG 290 K F AAAA 410 AGCG 530	G GGGA	A rgCrg 294 S ATCCA 306 L TTTAT 318 K GAAAC 330 F ATTCCA 342 CTCAT 354	E L AGCT 0 $E V AAGT 0 $ $CCCCC 0$	T AACT/ 29] L CTTG(30 GGGAA 31 TAAAA 33 TTAAA 33	I A TCGG 50 V S TTTC 70 V 7 TTAA 190 E I AGAA 310 I I ATTAA 30 ATCT7 550	A Q CGCA S E CAGA CAAA CAAA ATGT R N GAAA	S AAGT 296 D AGAT 308 T TGT 332 TGT 332 TGT 332 TGT 332 AATG(350	R CGT G GGGA 30 S TTTA 20 L TTTA 20 L TTTG 40 CAGT	I M ATG CTG GGGG F TTC	S [ICTT 297 S ICTA 309 N AATA 321 GGAA 333 H CATC 345 TTIGA 357	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0 U	N FAAT 3 F S AAG 3 P ACCC 5 S AAG 3 3 S AAG 3 3 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C 5 S C C S C C S C S	K CAAA 2980 L TTTA B100 F TTTC 3220 V CGTT 3340 L ATTA 33460 ATTA 3580	I ATCC P CCTG ATATAT G G GGTA	Q 1 AAAA 2 GCGG 3 S CCGG 3 T CCGG 3 D CCCG 3 D CCCG 3 CCCCG 3 CCCG CCCCG CCCCCC	K V AGGT 990 E L AAACT 110 E A AAGC 230 G F GATI 350 F A TTIGC 470 AAAA 590	Y TTA , I AAT CAAT CAAT	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MAGAAT 3360 F K TTTCAA 3480 MTAATA 3600
W TIGGA CCAGG L ATIGO AGAAA AATGO AGAAT	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 <u>F A</u> CTGC. 3250 <u>H R</u> ACCO 3370 F M FTATC 3490	C GTGI 0 H ACAC 0 E AGAA 0 TAGAA 0 STGA	I ATA ATC R CGA V GTT R CGG	T ACAA 2900 H CATA 3020 T T ACGG 3140 L TTAG 3260 S F CAA 3380 S F CAA 3380	K I AAAAT M R GGGCG A K CAAAA R Q GGCA	I TATU 29 H ACA 30 TGC(31 AGC(32 AGC(32 AGC(32 AGC(32 AGC(33 AGC(AGC(G GGGC 910 I I FATC 030 L CTTG 150 A CGCT 270 N FAAC 390	I E GAAG E GAA L CTT K AAG	P CCT 2 I ATT 3 E GAA 3 CGGG 3 I ATT 3 AAT	W TGG(920 1 ATT 040 A GCA 160 GCA 160 GGT 280 T T ACC(400 GCT 520	G GGAT FTGC I ATTT R CGTA R CGTA	F TCT 2 R GTC 3 C C TTTG 3 K AAAA 3	L 1 TGA1 930 Q 1 AAA1 050 Y F ACAG 170 D W AAAA 410 AGCG 530	GGGA	A IGCTG 294 S ATCCA 306 L ITTTAT 318 K GAAAC 330 F ATTCC 342 CTCAT 354	$E L AGCT 0 $ $\frac{K V}{AAGT} 0 $ $\frac{L L L}{TATT} 0 $ $\frac{V}{G L} GTCT 0 $ $E V AAGT 0 $ $CCCCC 0 $	T AACT/ 29] L CTTG(30 GGGAA 31 TAAAA 33 TTAAA 33	I A TCGC 50 V S TTTC 70 TTAA 90 E I AGAA 310 I J ATTAA 30 ATCT: 550	A Q CGCA S E CAGA CAAA CAAA ATGT R N GAAA	S AAGT 296 D AGA1 308 T TGT1 332 V TGT1 332 V TGT1 332 AATG4 350	R CGGT G GGGA 30 S TTTA 20 L TTTA 20 L TTTG 40 CAGT	I M ATG CTG GGGG F TTC	S [ICITI 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TTIGA 357	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0 0	N TAAT 2 F ATTTI 3 S AAGT 3 P P AACCC 2 K AAAA 2 S AAGT 3 3 S AAGT 3 3 S AAGT 3 3 S AAGT 3 3 S AAGT 3 3 S AAGT 3 S A AAGT 3 S A AAGT 3 S A A A A A A A A A A A A A	K CAAA 2980 L TTTA B100 F TTTC 3220 V CGTT 3340 L ATTA 33460 ATTA 3580	I ATCC P CCTG ATATAT G GGTA	Q 1 AAAA 2 GCGG 3 S CCGG 3 D D ATT 3 CATA 3	K V AGGT 990 E L AACT 110 E A AAGC 230 G F GATI 350 F A TTIGC 470 AATAA 590	Y TTAJ , I AATJ SGAGJ 7 K CAAA	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480 ATAATA 3600
TIGGA Q CCAGG ATIGO ATIGO AGAAA AATGO AGAAT	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 T A CTGC. 3250 H R ACC07 3370 F M ITATC 3490	C GTGI 0 H ACAC 0 E AGAA 0 TAGAA 0 STGA	I ATA ATC R CGA V GTT R CGG	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S ICAA 3380 S ICAA 3380	K I AAAAT ATAG GGRG GGCG A K CAAA CAAAA	I TATU 29 H ACAT 30 TGC(31 AGC(32 AGC(32 AGC(32 AGC(32 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC(33 AGC) AGC(33 AGC(3	G CGGC 910 IV I TATC 030 L CTTG 150 A CGCT 270 N TAAC 390 A GGA 510	I E GAG E GAA CTT K AAG	P CCT 2 I ATT 3 E GAA 3 R CGG 3 I ATT 3 AATT 3	W TGG6 920 1 ATT 040 A GCA 160 GCA 160 GGT 280 T ACC 400 GCT	G GGAT V GTGC I ATTTI R CGTA R CGTA	F TCT 2 C C C C C C C C C C C C C C C C C	L 1 TGA1 930 Q 1 AAA1 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410 AGCG	GGGA	A rgCrg 294 S ATCCA 306 L TTTAT 318 K GAAAG 330 F F ATTCC 342 CTCAT 354	E L AGCT 0 $K V AAGT 0 $ $L L TATT 0 $ $M G GTCT 0 $ $E V AAGT 0 $ $CCCCC 0 $ $CCCCC 0$	T AAACT/ 29] L CTTGG 30 GGGAA 31 TAAAA 32 TTAAA 32 TTAAA 32	I A TCGC 50 V S TTTC 70 TTAA 90 E N AGAA 310 I J ATTAA 30 ATCT: 550	A Q CGCA S E CAGA CAGA CAAA CAAA ATGT R N GAAA	S AAGT 296 D AGAT 308 T T AAAC/ 320 V T TGTT 332 V T TGTT 344 AATG(350	R CGT 50 G GGA 30 S ATCT 00 L TTTA 20 L TTTG 40 CAGT 50	I M ATG CTG GGGG F TTC	S [ICITI 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TIGA 357	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0	N FAAT 3 S AAGT 3 P AACCC 3 K AAAA 3 S C C C C C C C C C C C C C C C C C C	K TAAA 2980 L TTTA 3100 F TTTC 3220 V CGTT 3340 L ATTA 3460	I ATCC P CCTG ATATAT GGTA	Q 1 AAAA 2 GCGG 3 S CCGG 3 D D ATT 3 CATA 3	K V AGGT 990 E L AAACT 110 E A AAGC 230 G F GATT 350 F A TTIGC 470 AATAA 590	Y TTA , I AAAT , S GAGT , K CAAA	R S CGATC 3000 G L GGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480 ATAATA 3600
W TIGGA Q CCAGG ATIGO ATIGO AGAAA AATIGO AGAAA	N E ACGAM 2899 G V GAGT. 301 R A GAGC. 3130 T A CTGC. 3250 H R ACCOT 33770 F M TTATC 3490	C GTGI 0 H AACAC 0 E AGGAA 0 E TAGAA 0 STGA	I ATA I ATC R CGA V GTT R CGG TAC	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S S CCAA 3380 ATTA 3500 GAA	K I AAAAT ATAG G R GGCG A K CAAA CAAA CAAA	I TATU 29 H ACAT 30 AGC(31 AGC(32 AGC(32 AGC(32 AGC) 32 TAT/ 32 TCAA	G CGGC 910 IV I TATC 030 L CTTG 150 A CGCT 270 N TAAC 390 A A A A A A A A A A A A A	I E GAAG E GAAG L CTT K AAG	P 2 1 ATT 3 E GAAA 3 R CGG 3 I ATT 3 AATT 3	W TGG 920 I ATTO 040 A GCA 160 G GGT 280 T A ACCC 400 GCT 520	G GGAT V GTGC I ATTTI R CGTA R CGTA R CGGA	F TCT 2 C C C C C C C C C C C C C C C C C	L 1 TGA1 930 Q 1 AAA1 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410 AGCG 530	GGGA	A rgCrg 294 S ATCCA 306 L TTTAT 318 K GAAAG 330 F F ATTCCA 342 CTCAT 354	E L AGCT 0 XAAGT 0 L L TATT 0 V G L L C TATT 0 0 E V AAGT 0 0 CCCCC	T AACT/ 29] L CTTG(30 GGGAA 31 TAAA 33 TTAAA 34 34 TTAAA 34 34 TTAAA	I A TCGC 50 V S TTTC 70 TTAA 90 E I SAGAA 310 I I ATTTAC 130 ATCT: 550	A Q CGCA S E CAGA CAGA CAAAA N V ATGT R N GAAA	S AAAGT 296 D AAGAT 308 T T T 308 T T T 312 320 T T G T 332 T C T G T 332 C T G T 332 C T G T 332 C C AAGAT 308 C AAGAT 307 C C AAGAT 307 C C C C C C C C C C C C C C C C C C C	R CGT 50 G GGA 30 S ATCT D0 CAGT 50 CAGT	I M ATG CTG GGGG F TTC	S ICTT 297 S ICTA 309 N AATA 321 GGAA 333 H CATC 345 TIGA 357	L V TGGT 0 N I ATAT. 0 T Q CTCA 0 M I TGAT 0 H R ACAG 0 TTTTG	N FAATT 3 S AAGT 3 P AACCO 3 K AAAG 3 S C C C C C C C C C C C C C C C C C C	K TAAA 2980 L TTTA 3100 F TTTC 3220 V CGTT 3340 L ATTA 33460 ATTA	I ATCC P CCTG J I ATATAT C G GGTA J L L L TTGG	Q 1 AAAA 2 GCCG 3 CCCG 3 T CCCG 3 D CCCG 3 D CATA 3 CATA 3	K V AAGGT 990 E L AAACT 110 E A AAAGC 230 G F GATI 350 F A TTIGC 470 AATAA 590	Y TTA I AAATI SGAGI	R S CGATC 3000 G L CGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480 ATAATA 3600
TIGGA Q CCAGG ATIGO ATIGO AGAAA M AATGO AGAAAT AGAAAT	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 T A CTGCC 3250 B R R CTGCC 33370 F M TTATC 3490 TAGA 361	C GTGI 0 H ACAC 0 E AGAA 0 F TAGA 0 S TTCA 0 S TTCA	I ATA I ATC R CGA TAC	T ACAA 2900 H CATA 3020 T ACGG 3140 L L TTAG 3260 S S CCAA 3380 S ATTA 3500 GAAT	K I AAAAT GGCCG A K CAAAA CAAAA CAAAA CCAAAA	I TATU 29 H ACAT 30 AGCO 31 AGCO 32 H ACAT 32 TTAT/ 35	G G G G G G G G G G G G G G	I E GAAG GAA L CTT K AAG TTA	P CCT 2 I ATT 3 GAA 3 CCGG 3 I ATT 3 AATT 3 AATT	W TGG(920 I ATTO 040 A GCA 160 G GGT 280 T T ACCC 400 GCT 520	G GGAT T GTGC I ATTT R R CGTA R CGTA R CGTA	F TCT 2 R GTC 3 C [] 3 TTG 3 K K AAAA 3 CAAAA 3 GATC	L 1 TGA1 930 <u>Q 1</u> AAA1 050 <u>Y F</u> AAA1 0 0 0 W ATTG 290 K F AAAA 410 AGCG 530	GGGA	A IGCTG 294 S ATCCA 306 L TTTAT 318 K GAAAG 330 F F AATCCA 342 CTCAT 354 AAAG/ 364	E L AGCT 0 \overline{K} V AAGT 0 \overline{L} L TATT 0 \overline{V} \overline{G} L GTCT 0 \overline{C} C CCCC 0 0 \overline{C} C CCCCC 0 0 \overline{C} C CCCCC 0 0 \overline{C} C 0 \overline{C} C 0 0 \overline{C} C 0 \overline{C} C 0 C 0 C 0 C 0 C 0 C 0 C 0 C 0 C 0 C 0	T AACT/ 29] L CTTG(30 GGGAA 31 E AGAAA 32 TTAAA 32 SGTTC	I A TCGG 50 V S TTTTC 70 V D TTAA 10 1 I AGAA 310 I I ATTAC 550 CATCC 670	A Q CGCA S E CAGA CAGA CAAA ATGT R N GAAA	S AAAGT 296 D AGAT 308 T T TAAC/ 320 T T T T T T T T T T T T T T T T T T T	R CGT 50 G GGA 30 S ATCT D0 L L TTTG 40 CAGT 50 TTTA 80	I M ATG CTG GGGG F TTC CAT	S [ICTT 297 S ICTA 309 N AATA 321 GGGAA 333 H CATC 345 TTIGA 357 ATTIGA	L V TGGT 0 N I ATAT. 0 TTCA 0 H R ACAG 0 TTTTG 0 TTTTG 0 TTTTG	N FAATT 3 S AAGT 3 P AACCC 3 K AAAA 3 GTAA 5 CTCA	K XAAA 2980 L TTTA TTTA 32200 V QGT1 33400 L ATTA 34600 ATTA GGA3300 GGA3200	I ATCC P CCTG G G G G G G G G G G G G G G G G G	Q 1 AAAA 2 GGCG 3 CCCG 3 T CCCG 3 T CCCG 3 D CCCG 3 CCCCG 3 CCCCG 3 CCCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCCG 3 CCCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCG 3 CCCGC 3 C C CCCG 3 CCCG 3 CCCG 3 CCCG 3 C C C C	K V AGGT 990 E L AACT 110 E A AAGC 230 G F G F ATTIGC 470 ATTAA 590 FCTCCC 3710	Y TTA AAT AAAT CAAT CAAA TCAAA	R S XCGATC 3000 G L XGGATT 3120 F Q TTTCA 3240 R I MAGAAT 3360 F K TTTCAA 3480 ATAATA 3600 TTTTTTT 3720
TIGGA Q CCAGG ATIGO ATIGO AGAAA M AATGO AGAAAT AGAAAT	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 T A GAGC. 3130 T A GAGC. 3250 3250 3370 F M TTATC 3490 TAGA 361	C GTGI 0 H ACAC 0 E AGAA 0 F TAGA 0 S TTCA 0 F TCA 0 F TGA	I ATA I ATC R CGA V GTT TAC	T ACAA 2900 H CATA 3020 T T ACGG 3140 L L T T CAA 3380 S S T CAA 3380 S ATTA 33500 GAA1 3620	K I AAAAT G R G R G G R G G C G A K C AAAA R Q G G C A C AAAA	I TATU 29 ACAT 30 AGCC 31 AGCC 32 H AAGCC 32 H ACAT 33 TAT/ 35 TTAT/ 35	G GGGC 910 IV I FATC 030 L CTTG 150 A CTTG 270 N FAAC 390 NGGA 510 FTCI 630	I E GAAG GAA CTT K AAG TTA	P CCT 2 I ATT 3 E GAA 3 CCGG 3 I ATT 3 AATT 3 AATT 3	W TGG(920 1 ATT 040 A GCA 160 GCT 280 T T ACC 400 GCT 520 ACG 6640	G GGAT T GTGC I ATTT R R CGTA R R CGTA	F TCT 2 R GTC 3 C [] 3 K MAAA 3 K AAAA 3 GATC 3	L 1 TGA1 930 <u>Q 1</u> AAA1 050 <u>Y F</u> ACAG 170 <u>D W</u> AATTG 290 K F AAAA 410 K F AAAA 410 CGTCC	GGGA	A IGCTG 294 S ATCCA 306 L TTTAA 318 K GAAAG 330 F F ATTCCA 330 CTCAT 354 AAAGG 36	E L AGCT 0 \overline{K} V AAGT 0 \overline{L} L TATT 0 \overline{G} L GTCT 0 \overline{G} L GTCT 0 0 \overline{G} L GTCT 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	T AACT/ 29] L CTTGG 30 GGGGAA 31 TAAAC 33 K AGAAA 34 34 TTAAA 34 35 CTTAAA 34 35 CTTAAA 34 35 CTTC 35 CTTC 35 CTTC	I A TCGG 50 V S TTTTC 70 V 1 TTAA 10 11 11 130 XTTA 130 XTCT 550 CATCC 670	A Q CGCA S E CAGA CAGA CAAA ATGT R N GAAA	S AAGT 296 D AGAT 308 T T T T T T T T T T T T T T T T T T T	R CGT 50 G GGA 30 <u>S</u> ATCT D0 <u>L</u> L TTTA 20 L L TTTG 40 CAGT 50 TTTA 80	I M ATG CTG G G G G G G G G C T T C A T	S [ICTT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TTIGA 357 ATTIGA	L V TGGT 0 N I ATAT. 0 T <u>T</u> Q CTCA 0 M <u>I</u> TGAT 0 H R ACAG 0 TTTTG 0	N FAAT 3 F AAG 3 P AACCC 3 K K AAAA 3 GTAA 5 CTCA	K CAAA 2980 L TTTA 32200 V CGT1 3340 L ATTA 3340 L ATTA 3340 GGA 3370	I ATCC P CCTG G G G G G G G G G G G G G G G G G	Q 1 AAAA 2 GGCG 3 S CCGG 3 T CCGG 3 T CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCGG 3 CCG CCG	K V AGGT 990 E L AACT 110 E A AAGC 230 G F GATI 350 F A TTIGC 470 AATAA 590 FCTCC 3710	Y TTA AATI AAATI CAATI CAATI CAAAI	R S XCGATC 3000 G L XGGATT 3120 F Q TTTCA 3240 R I MAGAAT 3360 F K TTTCAA 3480 ATAATA 3600 TTTTTTT 3720
TIGGA Q CCAGG ATIGO ATIGO AGAAA M AATGC AGAAAT ATAAA	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 T A GAGC. 3130 T A GAGC. 3250 3250 3370 F M TTATC 3490 TAGA 361	C GTGI 0 H ACAC 0 E AGAA 0 TAGAA 0 S TTCA 0 S TTCA 0 S TTGA	I ATA I ATC R CGA V GTT TAC	T ACAA 2900 H CATA 3020 T ACGG 3140 L L CAAA 3260 S S CCAA 3380 S CCAA 3380 S CCAA 3380 S CCAA 3380 S CCAA 3380 S CCAA 3380 S	K I AAAAT M R GGCG A K CAAAA R Q GGCA CAAAA	I TATU 29 ACAT 30 ACCAT 31 ACCAT 32 AAGCC 32 AAGCC 32 ACAT 33 TAT/ 35 TCGAT	G GGGC 910 IV I FATC 030 L CTTG L 50 A CGCT N FAAC 390 NGGA 510 FTCT 630 FTCT	I E GAAG E GAA CTT K AAG TTA	P CCT 2 I ATT 3 E GAA 3 CCGG 3 I ATT 3 AATT 3 Soma	W TGG6 920 1 ATTC 040 A GCA.1 160 GGT 280 T T ACCC 400 GCT. 520 ACCG 6640 .1 p	G GGAT T GTGC I ATTT R R CGTA R R CGTA R CGGA	F TCT 2 R GTC 3 C I I NTTG 3 K AAAA 3 GATC 3 ein	L 1 TGA1 930 Q 1 AAA1 050 V F ACAG 170 D W ATTG 290 K F AAAA 410 K F AAAAA 410 S30 S2 -	GGGA	A IGCTG 294 S ATCCA 306 L TTTTAT 318 K GAAAG 330 F F CTCAT 354 AAAGG 360	E L AGCT 0 \underline{K} V AAGT 0 \underline{L} L TATT 0 \underline{V} G C G C C C C C C C C C C C C C C C C	T AACT/ 29] L CTTGG 30 GGGGAA 31 K TTAAAC 32 GGGAA 32 C TTAAAC 32 C TTAAAC 32 C TTAAAC 32 C TTAAAC 32 C TTAAAC 32 C TTAAAC 32 C TTAAC 32 C TTAAC 32 C TTAAAC 32 C TTAAC 32 C TTAAAC 32 C TTAAC 3 C TTAAC 32 C TTAAC 3 C TTAAC 3 C TTAAC 3 C C C C C C C C C C C C C C C C C	I A TCGG 50 V S TTTTC 70 V D TTAAC 10 10 11 11 130 XTTTA 130 XTCT 550 CATCC 670	A Q CGCA S E CAGA CAAAA N V ATGT R N GAAA	S AAGT 296 AGAT 308 T T TAAC/ 320 T T T T T T T T T T T T T T T T T T T	R CGT 50 G GGA 30 <u>S</u> ATCT D0 <u>L</u> TTTA 20 L TTTA 50 TTTA 80	I ATT M ATG CTG G G G G G G G G G G G G G G G G G	S [ICTT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TTGA 357 ATTG(369	L V TGGT 0 N I ATAT. 0 TTCA 0 H R ACAG 0 TTTG 0 TTTG 0 TTTG 0 TTTG 0	N FAATTI 3 S AAGT 3 P AACCC 5 K K AAAA 5 GTAA 5 TTCA	K CAAA 2980 L TTTA 32200 V CGTT 3340 L ATTA 3460 ATTA 3580 GGA2 3700	I ATCC P CCTG G G G G G G G G G G G G G G G G G	Q 1 AAAA 2 GGCG 3 CCOG 3 T CCOG 3 T CCOG 3 T CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 C CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 C CCOG 3 CCOG CCOG	K V AGGT 990 E L AACT 110 E A AAGC 230 G F GATI 350 F A TTIGC 470 AATAA 590 FCTCC 3710	Y TTA AATI AAATI CAATI CAATI	R S XCGATC 3000 G L XGGATT 3120 F Q TTTCA 3240 R I MGGATT 3360 F K TTTCAA 3480 ATAATA 3600 TTTTTTT 3720
W TTGGA Q CCAGG ATTGO	N E ACGAU 2890 G V GAGT. 301 R A GAGC. 3130 T A GAGC. 3250 B R R CTGCC 3250 3250 CTGCC 3250 CTGCC 3490 TAGA 361	C GTGI 0 H ACAC 0 E AGAA 0 TTCA 0 STGA 0 STGA	I ATA I CGA R CGA TAC	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S S TCAA 3380 S ATTA 3500 GAAT 3620	K I AAAAT M R GGCG A K CAAAA R Q GGCA CAAAA	I TATU 29 H ACAT 30 AGCC 31 AGCC 32 H ACAT 33 TAT/ 35 TAT/ 35	G GGGC 910 IV I FATC 030 L CTTG L50 N FAAC 390 NGGA 510 FTCI 630 F1	I E GAAT E GAA E CTT K AAG TTA	P CCT 2 I ATT 3 E GAA 3 CCGG 3 I ATT 3 AATT 3 AATT 3 Soma 8	W TGG(920 1 ATTR 040 A GCA 160 GGT 280 T C ACC(400 GCT 520 ACG 640 1 p R	G GGAT T GTGC I ATTT R R CGTA R R CGTA R CGTA R CGTA	F TCT 2 R GTC 3 C I I TTG 3 K AAAA 3 GATC 3 cin W	L 1 TGA1 930 Q 1 AAA1 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410 K F AAAA 410 S30 S22 N 2	GGGA	A IGCTG 294 S ATCCA 306 L TTTAA 318 K GAAAG 342 CTCAT 354 AAAAG 36 F	E L AGCT 0 \underline{K} V AAGT 0 \underline{U} L G TATT 0 \underline{U} C G C C C C C C C C C C C C C C C C C C	T AACT/ 29] L CTTGG 30 GGGGAC 31 K K AGAAA 3' TTAAAC 3' C C C C C C C C C C C C C C C C C C	I A TCGG 50 V S TTTTC 70 TTAAC 10 10 11 11 10 10 10 10 10 10	A Q CGCA S E CAGA CAAAA N V ATGT R N GAAA FTTAA	S AAGT 296 AGAT 308 T T TAAC/ 320 T T T T T T T T T T T T T T T T T T T	R CGT 50 G GGA 30 S ATCT D0 L TTTA 20 L TTTA 50 TTTA 80 V	I ATT M ATG CTG G G G G G G G G G G G G G G G G G	S [ICTT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TTIGA 357 ATTIG 45 F	L V TGGT 0 N I ATAT. 0 T Q CTCA. 0 M I TGAT 0 H R ACAG 0 TTTG 0 TTTG 0 CTATT 90 G I	N TAAT 2 F AATTI 3 S AAGT 3 3 P ACCC 5 K K AAAA 5 GTAA 5 GTAA 1 D	K TAAA 2980 L TTTA 32200 V OGTT 3340 L ATTA 3340 L GGT 3340 L GGT 3340 L TTA GGA: 3700 T	I ATCC P CCTG G G G G G G G G G G G G G G G G G	Q 1 AAAA 2 GGCG 3 CCOG 3 T CCOG 3 T CCOG 3 T CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCO 3 C C C C	K V AGGT 990 E L AACT 110 E A AAGC 230 G F GATI 350 F A TTIGC 470 AATAA 590 TCTCC 3710	Y TTA A A A A A A A A A A A A A A A A A	R S XCGATC 3000 G L XGGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480 ATAATA 3600 TTTTTTTT 3720 R M
TIGGA Q CCAGG ATTGO ATTGO M ATTGO AGAAA ATAAA ATAAA TITTT	N E ACGAU 2899 G V GAGT. 301 R A GAGC. 3130 T A GAGC. 3250 H R ACCGT 3250 3370 F M ITATC 3490 F M ITATC 3490 TAGA. 361	C GTGI 0 H ACAC 0 E AGAA 0 TAGA 0 STGA 0 STGA 0 C TTCA	I ATA I CGA R CGA TAC CGG TAC	T ACAA 2900 H CATA 3020 T ACGG 3140 L TTAG 3260 S S TCAA 3380 S ATTA 3500 GAAT 3620 GTGT	K I AAAAT M R GGCG A K CAAA R Q GGCA CAAA CAAA	I TATU 29 H ACAT 30 TGCC 31 AGCC 32 H ACAT 33 TAT/ 35 TAT/ 31 CTGA' 31	G GGGC 910 IV I FATC 030 L CTTG L50 A GGCT N FAAC 390 NGGA 310 FTCI 630 FTCI	I E GGAG E GGAA L CTT K AAG TTA	P CCT 2 I ATT 3 E GAA 3 CCGG 3 I ATT 3 AAT 3 3 CCGG 3 A ATT 3 AAT 3 CCGG 3 A ATT 3 AAT 3 A AAT 3 A A 3 A A 3 A A A A	W TGG(920 1 ATTR 040 A GCAJ 160 C GGT 280 T ACCC 400 GGT 520 ACG 640 1 p R AGA	G GGAT T T T T C G T G C G C G C G C G C C C T A T T A T	F TCT 2 R GGTC 3 C GTT 3 C GTT 3 K AAAA 3 SATC 3 SATC 3 SATC 3 SATC 3 SATC	L 1 TGAT 930 Q 1 AAAT 050 Y F ACAG 170 D W ATTG 290 K F AAAA 410 K F AAAA 410 S30 S22 N S S22 N S	GGGA	A IGCTG 294 S ATCCA 306 L TTTAT 316 K GAAAG 330 F GAAAG 342 CTCAT 354 AAAG 36 F TTTTAT	E L AGCT 0 \overline{K} V AAGT 0 \overline{C} L \overline{L} L TATT 0 \overline{C} C \overline{C} C C	T AACT/ 29] L CTTGG 30 GGGAA 31 TAAAC 31 K TAAAC 32 AGAA 34 TTAAAC 34 C C C C C C C C C C C C C C C C C C	I A TCGG 50 V S TTTTC 70 V D TTAAC 190 E I AGAA 310 I I ATTAA 130 ATCT 550 CATCC 670 M G	A Q CGCA S E CAGA CAAAA N V ATGT R N GAAA GGGAA	S AAGT 296 AGAT 308 T T TAAC/ 320 T T T T T T T T T T T T T T T T T T T	R CGT 30 GGAA 30 <u>S</u> ATCT D0 <u>L</u> TTTA 20 L TTTG 40 CAGT 50 V V GGT	I ATT M ATG CTG G GGG F TTC CCAT	S [ICTT 297 S ICTA 309 N AATA 321 G GGAA 333 H CATC 345 TTIGA 357 ATTIG 45 F F	L V TGGT 0 N I ATAT. 0 T Q T Q T Q T Q T Q T Q T Q T Q T Q T Q	N TAAT 2 F AATTI 3 S AAGT 3 AAGT 3 A AAGT 3 A A A A A A A A A A A A A	K TAAA 2980 L TTTA 3220 V <	I ATCC P CCTG CCTG C CTG C G G G G G G G G G G G	Q J AAAA 2 G GCG 3 S CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCOG 3 CCO 3 C C CCO 3 C C C C	K V AGGT 990 E L AACT 110 E A AGC 230 G F GAT1 350 F A TTIGC 470 AATAA 590 FCTCC 3710 W IGGA	Y TTAI AATI AAATI CAATI CAAAI	R S XCGATC 3000 G L XGGATT 3120 F Q TTTCA 3240 R I MGAAT 3360 F K TTTCAA 3480 ATAATA 3600 TTTTTTT 3720 R M TCGAAT A

A GGC/	F ACC	р ЛТ	F TT	I ATA	s .Ta	A GGC		K A A (R CGT	K AA	G Aggʻ] TA:	I FTC	H At/	I ATT	T ACA	N AA7	L ICT	TAC	Г СТ <i>і</i>	K \\\\	T ACI	A GCI	R CG1	F TTI	TT	L Tat	S ICA	E Gaa	A GCI	C TGT	D FGA	L	A GCI	F	D GAJ	A GC/	A AGC/	S A AG	CAA	. G AGO	i K	Q ACA	F ATT
		38	350			38	60			3	B70			31	880			38	90			39	000			39	10			392	20		39	30		3	94()		39	50		3	960
L TTT	1 1 A A	r 1716 39	V STT 970	G 3 6 1	T AC	8 444 39	() (AA (80)	K AA/	K AAA	A GC 3	A Agc 990	l CG/) ATT	S °CA(4(V GTA 000	T ACA	R ACGO	A GGC 40	TG 10	A Ca/	I ATA	R Ag/ 4(A GC1 020	R CG	C ATG	TC 40	H AT1 30	Y TATI	V GTT	N AA 404	к Гал Ю	K AAA	W ATGO 40	L 6CT () 5 0	R COGA	G G G 1 4	M 7AT0 1060	L STT.)	1 AA(Г № Сбал 40	70	/ Y GTA	т Тас 4	TAC 080
E Aga.	1 AAC	r DG <i>A</i> 40	R AGA 990	L CT1	G GG	В Гал 41	C 1 16T 100	F TC/	R Agg	D GA 4	L CTT 110	l G A (R GAA	T .CG(4:	E GAG 120	Q CA/	K	T GAC 41	GG(30	G GG/	к Алл	L CT(41	N GAAC L40	S CTC	L TCI	, ТС 41	P CAA 50	K AA	R Aga	D GA1 410	A IGCO 50	A GGC1	M ATC 41	L TTC 70	K Faag	R FAGA	Q ACA 180	L ATT	S ATC	5 E CTCA 41	(1 .TT) .90	i E ICGA	т ААС 4	¥ ATA 200
L TCT.	G Ag(3 GCG 42	G GT 210	I ATC	K CAA	¥ AT/ 42	(M TG/	T ACA	G .GG 4	L GTT 230	AC	P CTG	D AT 4	I ATT 240	V GT/	I ATA	1 AA1 42	CG:	V TTC	D Gat	Q CA(42	Q GCA/ 260	K 1.4.4.	AG A	AT 42	Y At/ 70	T \CG4	A GCT	L CT1 428	Q [CA. 30	E Aga /	C ATG1 42	I TATO 290	T CACI	L TTI	G NGG/ 130	I AAT D	I TCC	2 CAAC 43	GA1 10	I C TTG	E L TTT 4	I AAT 320
D CGAI	T TAC	43	N AT 30	C IGT	D GAC	P CC 43	40) \TC	L TT	A GCA 43	D GA1 50	M FAJ	I (GT)	S CG <i>A</i> 43	1 1 60	P COG	A GCT	N 'AA' 43 '	D TGA 70) ATG	D ATC	A 9CT 43	I ATA 80	A GCI	S FTC] CAT 43 9	r 1 20 90	R GAT	L TA	I ATT 440	L CTI 0	N TAAC	к Алл 44	L TTA 10	V GTT	F TTT 4	A GCA 420	I ATI	C TTG	E TGA 44	G GGG 30	R TCG:	S FTC: 44	S [AG 440
s ctci	I TA:	AC	R GA/	N AT	Y TAJ	+ TG	ATI	CA.A	TA	AT/	AG/	\TA		гсо	AT	rtt	TAG	AT	FTG	GT	TG	GC	GGT	CAT	Γ A G.	AT 1	TTT	TGO	GAA'	TTG	GGT	ATI	ATA	GCA	TTA	CAA	AAT	TGI	GT		AAG	V V V .	TAT.	пт
		44	50			44	60			44	170			44	80			44	90			45	00			451	10			452	0		45	30		4	540			45 A	50 TPa	se :	4: •	560 >
GTGA	T	'AG' 45	TAC 70	GT	ATI	CA 45	AA/ 80	\TA	GA	۸۸/ 45	ATC/ 590		GT.	AAA 46	AT/	AA G	GAA	AT 46	3 G1 10	TG	AA1	ГСА 46	AAA 20	TA.	ATT	CC(463	стт 30	TC/	AG	ГТА 46 4	ТАТ 0	TTI	TTT 46	ATT 50	TTA	GAG 4	GAC 660	AGG	GC	AAT 46	M ATG 70	AAT	v 3TT(4(L TA 580
Fig.	2.	N	luc	leo	tid	e se	equ	en	ce	of	рея	ı cl	hD	NA	\ fr	om	a	Sal	/I s	ite	e to	th	e bi	egir	nniı	ng	of	the	ge	ne	for	the	A	[P- 4	syn	thas	se a	su	bu	nit	and	sea	uen	ices

Fig. 2. Nucleotide sequence of pea chDNA from a Sall site to the beginning of the gene for the ATP-synthase a subunit and sequences of encoded proteins

Transcription and translation are from left to right. Proposed ribosome binding sites are boxed. The boxed protein sequences are homologous to the β' -subunit of *E. coli* RNA polymerase.



Fig. 3. Gene predictions in the DNA sequenced

The positional base preference method (Staden, 1985) was employed. The three boxes represent the translational reading frames of the DNA. The scale is in kb. The top of the scale on the ordinate represents the level that would be reached in a gene encoding a protein of average amino acid composition. Points under the midline are likely to be non-coding. Vertical bars on the abscissae represent possible start codons (ATG), those on the midlines indicate potential stop codons. The positions of the potential genes for the β' -subunit of RNA polymerase and ribosomal protein S2 are shown. The ATPase a subunit has been described elsewhere (Cozens *et al.*, 1986).



Fig. 4. Comparisons of the encoded proteins with the aid of the computer program DIAGON

(a) E. coli ribosomal subunit S2 (ordinate) versus the protein predicted by the complete open reading frame in the pea chDNA (abscissa; bases 3749-4459 in Fig. 1); (b) the β' -subunit of E. coli RNA polymerase (ordinate) versus the protein predicted by the incomplete open reading frame (abscissa; bases 1-3493). The sequences of homologous segments I-V are shown in Fig. 6. The numbers on the axes correspond to the lengths of the amino acid chains. The calculations were made with a window of 25 amino acids and a score of 290.

Pea chloroplast <u>E.coli</u>	1 MTKRYWNITF Atvsm	EEMMEAGVHF RDMLKAGVHF	G HDT RKW NPR GHQT RYW NPK	MAPFISAKRK MKPFIFGARN	50 GIHINNLIKT KVHINLEKT
Pea chloroplast <u>E.coli</u>	51 ARFLSEACDL VPMFNEALAE	AFDAASKGKQ LNKIASRKGK	FLIVGTEEEA ILFVGTEEAA	AD SVITRAAIR SEAVIK DAALS	100 ARCHYVNKKW CDQFFVNHRW
Pea chloroplast <u>E.coli</u>	101 LRGMLTNWYT LGGMLTNWKT	TETRLGKFRD VRQSIKRLKD	LRTEQKTGKL LETQSQDGTF	N SILPIKR DAAM DKILTIKK EALM	150 LKRQLSHFET RTRELEKLEN
Pea chloroplast <u>E.coli</u>	151 YLGGIKYMTG SLGGIKDMGG	LPDIVIIVDQ LPDALFVIDA	QKEYTALQEC DHEHIAIKEA	ITLGIPTICL NNLGIPVFAI	200 IDTNCDPDLA VDTNSDPDGV
Pea chloroplast <u>E.coli</u>	201 DMSTPANDDA DFVIPGNDDA	IASIRLIINK IRAVTLYLGA	LVFAICEGRS VAATVREGRS	SSIRNY QDLASQAEES	245 FVEAE

Fig. 5. Alignment of the sequences of the protein predicted from the pea chloroplast complete open reading frame with that of the *E. coli* ribosomal subunit S2

Identities are boxed.

insertion of about 250 amino acids be present in the chloroplast protein. This has two possible implications. Firstly, that the pea chloroplast enzyme differs substantially in structure from the bacterial enzyme, or secondly, that the pea chloroplast gene contains an intron of approx. 750 bp. In the absence of further experimental evidence it is not possible, at present, to know which interpretation is correct.

Chloroplasts have been shown to contain two, and possibly three, different RNA polymerase activities. In *E.* gracilis two types of activity have been isolated. One polymerase is tightly bound to chDNA and is selective for transcription of rRNA genes. It contains three major subunits of M_r 116000–118000, 83000–88000 and 24000–26000 (Narita *et al.*, 1985). The other is in the soluble fraction and is selective for transcription of tRNAs. It has not been determined whether chloroplast mRNA is transcribed by the soluble extract (Greenberg *et al.*, 1984). Higher plant chloroplast RNA polymerases that transcribe mRNA are composed of multiple subunits with a broad M_r range (Kidd & Bogorad, 1980). The pea enzyme may contain polypeptides of M_r 180000, 140000, 110000, 95000, 65000, 47000 and 27000 (Tewari & Goel, 1983), and it is conceivable that the fragment sequenced in the present work corresponds to one of these subunits.

```
Ι
```

Pea chloroplast <u>E.coli</u> Yeast II Yeast III	1 RLVEVVQHIV VRRTDCGTIR GISVNTRN.G MMPEIILIQT 40 799 RLVDVAQDIV VTEDDCGTHE GIMMTPVIEG GDVKEPLRDR 838 840 RLVKALEDIM VHYDNTTRNS LGNVIQFIYG EDGMDAAHIE 879 888 RLMKSLEDLS CQYDNTVRTS ANGIVQFTYG GDGLDPLEME 927
Pea chloroplast	41 LIGRVVAENI 50
E.coli	839 VLGRVITAEDV 848
Yeast II	880 HIELQSLDTI 889
Yeast III	928 GNAQPVNFNR 937
II	
Pea chloroplast	81 IRTPFTCRNT SWICKLCYGR SPIHGDLVEL GEAVGIIAGQ 120
	882 VESUUSCOTD EGUCLARCYGR DI ARGHIINK GEATIGUTAAO 921
	1021 BITTOLEDWU I SUTENDELOU BOUNDERWUG CEMBUCUIALA 1070
	1031 RUIBERFDWY LONIEREFLE SY VEFOERVG GERVGVUNLANG 10/0
Yeast III	1056 VSQLYRISEK SVRKFLEIAL FKYRKARLEP GTALGALGAQ 1095
Pea chloronlast	121 STGEPGTOLT LETERTGOVE TGGTARVIRA PS 152

Pea chloroplast	121	SIGEPGTQLT	LRTFHIGGVIF	TGGTAEYVRA	PS	152
E.coli	922	SIGEPGTQLT	MRTFHIGGAA	SRAAAESSIQ	VK	953
Yeast II	1071	SIGEPATOMT	LNTFHFAGVA	SKKVTSGVPR	LK	1102
Yeast III	1096	SIGEPGTOMT	LIKT FHIFAG VA	SMNVTLGVPR	IK	1127

III

E.coli 1111 DGVQISSGDT LARIPOESGG TKDITGGLPR VADLFEARR 11	$\frac{01i}{11}$	1 DGVQISSGDI	LARIPQESGG	TKDITGGLPR	VADLFEARR	1149
---	------------------	--------------	------------	------------	-----------	------

IV

Pea chloroplast	989	LVNKIQKVYR	SQGVHIHNRH	IEIIVRQITS	KVLV	1022
<u>E.coli</u>	1233	IVNEVQDVYR	LOGVKINDKH	IEVIVRQMLR	KATI	1266

V

Pea chloroplast	1056 YRALLLGVTK TSLNTQSFIS EASFQETARV LAKAALRGRI 10)95
<u>E.coli</u>	1302 YSRDLLGITK ASLATESFIS AASFQETTRV LTEAAVAGKR 13	341
Yeast II	1379 GGLTSVTRHG FNRSNTGALM RCSFEETVEI LFEAGASAEL 14	418
Yeast III	1367 GEVLGITRFG LSKMRDSVLQ LASFEKTTDH LFDAAFYMKK 14	406
Pea chloroplast <u>E.coli</u> Yeast II Yeast III	1096 DWLKGLKENV VLGGMIPVGT GFKRIMHRSR SR 1127 1342 DELRGLKENV IVGRLIPAGT GYAYHQDRMR RR 1373 1419 DDCRGVSENV ILGQMAPIGT GAFDVMIDEE SL 1450 1407 DAVEGVSECI ILGQTMSIGT GSFKVVKGTN IS 1438	

Fig. 6. Alignment of homologous segments of the protein predicted from the incomplete open reading frame in the pea chloroplast genome, the β' -subunit of *E. coli* RNA polymerase and yeast RNA polymerases II and III

Parts I-V corresponds to homologies found in Fig. 4(b). See also Fig. 2. Identities between the pea chloroplast protein and E. coli RNA polymerase and/or yeast RNA polymerases II and III are boxed.

Gene arrangements

The region of the pea chloroplast genome discussed in this paper has an arrangement of genes that is probably as follows: RNA polymerase β' -subunit:ribosomal subunit S2: ATPase a: ATPase III (see Fig. 1). Transcription of this region has not been studied closely, but it is known that a transcript extends at least from within S2, through ATPase a and ATPase III (Cozens et al., 1986). Moreover, the lack of extensive non-coding sequences around the S2 gene makes it somewhat unlikely that major promoters will lie there, and therefore it is possible that all of these genes are co-transcribed. Co-transcription of genes with unrelated function is known in E. coli. For example, the α -subunit of RNA polymerase is co-transcribed with several ribosomal proteins (Post et al., 1980). It is therefore of some interest to find that the chloroplast β' -subunit of RNA polymerase is associated with ribosome subunit S2, although the E. coli homologue is associated with a translational factor, EF-Ts (An et al., 1981).

We are grateful to Dr J. C. Gray for supplying plasmid pPscS6. A.L.C. is supported by an MRC Research Studentship.

REFERENCES

- Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985) Cell 42, 599–610
- An, G., Bendiak, D. S., Mamelak, L. A. & Friesen, J. D. (1981) Nucleic Acids Res. 9, 4163–4171
- Bankier, A. T. & Barrell, B. G. (1983) in Techniques in Nucleic Acid Biochemistry (Flavell, R. A., ed.), B508/1-B508/31, Elsevier Scientific Publishers, Ireland
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3963–3965
- Bohnert, H. J., Crouse, E. J. & Schmitt, J. M. (1982) Encycl. Plant Physiol. New Ser. 14B, 475-530
- Bollen, A., Heimark, R. L., Cozzone, A., Traut, R. R., Hershey, J. W. & Kahan, L. (1975) J. Biol. Chem. 250, 4310–4314
- Cozens, A. L., Walker, J. E., Phillips, A. L., Huttly, A. K. & Gray, J. C. (1986) EMBO J. 5, 217–222
- Dorne, A. M., Lescure, A. M. & Mache, R. (1984) Plant Mol. Biol. 3, 83–90
- Dyer, T. A. (1984) in Chloroplast Biogenesis (Baker, N. R. & Barker, J., eds.), pp. 23-69, Elsevier Science Publishers, Amsterdam
- Dyer, T. A. (1985) Oxford Surveys Plant Mol. Cell Biol. 2, 147-177
- Eneas-Filho, I., Hartley, M. R. & Mache, R. (1981) Mol. Gen. Genet. 184, 484-488
- Engleman, D. M., Moore, P. B. & Schoenborn, B. P. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3888-3892
- Georgalis, Y., Giri, L. & Littlechild, J. A. (1981) Biochemistry 20, 1061–1064
- Graf, L., Roux, E. & Stutz, E. (1982) Nucleic Acids Res. 10, 6369–6381
- Gray, J. C., Phillips, A. L., Howe, C. J., Willey, D. L., Huttly, A. K., Doherty, A., Bowman, C. M. & Dyer, T. A. (1984) in Biosynthesis of the Photosynthetic Apparatus (Thorber, J. P. et al., eds.), pp. 295–307, A. R. Liss, New York
- Greenberg, B. M., Narita, J. O., De Luca-Flaherty, C., Graissem, W., Rushlow, K. A. & Hallick, R. B. (1984) J. Biol. Chem. 259, 14880–14887

- Herrman, R. G. & Possingham, I. V. (1980) in Chloroplasts (Reinert, J., ed.), pp. 45–96, Springer-Verlag, Berlin
- Huttly, A. K. & Gray, J. C. (1984) Mol. Gen. Genet. 194, 402-409
- Kidd, G. H. & Bogorad, L. (1980) Biochim. Biophys. Acta 609, 14-30
- Lake, J. A. (1978) in Advanced Techniques in Biological Electron Microscopy II (Koehler, J., ed.), pp. 173–211, Springer-Verlag, Berlin, Heidelberg, New York
- Lerbs, S., Briat, J.-F. & Mache, R. (1983) Plant Mol. Biol. 2, 67-74
- Lipman, D. J. & Pearson, W. R. (1985) Science 227, 1435–1441
- Littlechild, J. A., Dijk, J. & Garrett, R. A. (1977) FEBS Lett. 74, 292–300
- Montadon, P. E. & Stutz, E. (1984) Nucleic Acids Res. 12, 2851–2859
- Narita, J. O., Rushlow, K. E. & Hallick, R. B. (1985) J. Biol. Chem. 260, 11194–11199
- Okuyama, A., Yoshikawa, M. & Tanaka, N. (1974) Biochem. Biophys. Res. Commun. 60, 1163–1169
- Ovchinnikov, Yu, A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Salomatina, I. S., Shuvaeva, T. M., Lipkin, V. M. & Sverdlov, E. D. (1982) Nucleic Acids Res. 10, 4035-4044
- Passavant, C. W., Stiegler, G. L. & Hallick, R. B. (1983) J. Biol. Chem. 258, 693-695
- Phillips, A. L. & Gray, J. C. (1984) Mol. Gen. Genet. 194, 477–484
- Posno, M., van Noort, M., Debise, R. & Groot, S. P. (1984) Curr. Genet. 8, 147–154
- Post, L. E., Arfsten, A. E., Davis, G. R. & Nomura, M. (1980)
 J. Biol. Chem. 255, 4653–4659
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467
- Schmidt, R. J., Richardson, C. B., Gillham, N. W. & Boynton, J. W. (1983) J. Cell Biol. 96, 1451–1463
- Schwartz, Z. & Kössel, H. (1980) Nature (London) 283, 739-742
- Shine, J. & Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342–1346
- Squires, C., Krainer, A., Barry, G., Shen, W. F. & Squires, C. L. (1981) Nucleic Acids Res. 9, 6827–6840
- Staden, R. (1982a) Nucleic Acids Res. 10, 4731-4751
- Staden, R. (1982b) Nucleic Acids Res. 10, 2951-2961
- Staden, R. (1984) Nucleic Acids Res. 12, 521-538
- Staden, R. (1985) Genet. Eng. 7, 67-114
- Subramanian, A. R., Steinmetz, A. & Bogorad, L. (1983) Nucleic Acids Res. 11, 5277–5286
- Sugita, M. & Sugiura, M. (1983) Nucleic Acids Res. 11, 1913-1918
- Tewari, K. K. & Goel, A. (1983) Biochemistry **22**, 2142–2148 Tischendorf, G. W., Zeichardt, H. & Stöffler, G. (1975) Proc.
- Natl. Acad. Sci. U.S.A. 72, 4820–4824
- Tohdoh, N. & Sugiura, M. (1982) Gene 17, 213-218
- Traut, R. R., Lambert, J. M., Boileau, G. & Kenny, J. W. (1980) in Ribosomes (Chambliss, G. et al., eds.), pp. 89–110, University Park Press, Baltimore
- Walker, J. E. & Tybulewicz, V. L. J. (1985) in The Molecular Biology of the Photosynthetic Apparatus (Arntzen, C., Bogorad, L., Bonitz, S. & Steinback, K., eds.), pp. 141–153, Cold Spring Harbor
- Whitfeld, P. R. & Bottomley, W. (1983) Annu. Rev. Plant Physiol. 34, 279-310
- Wittmann-Liebold, B. & Bosserhoff, A. (1981) FEBS Lett. 129, 10–16
- Yoshikawa, M., Okuyama, A. & Tanaka, N. (1975) J. Bacteriol. **122**, 796–797
- Zurawaski, G., Bottomley, W. & Whitfeld, P. R. (1984) Nucleic Acids Res. 12, 6547–6558