Photosynthetic regulatory gene cluster in an aerobic photosynthetic bacterium, *Roseobacter denitrificans*¹

Kohji Nishimura,[†] Hiroshi Shimada, Takashi Shinmen, Takeshi Obayashi, Tatsuru Masuda, Hiroyuki Ohta, and Ken-ichiro Takamiya^{*}

Department of Biological Sciences, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226–8501, Japan

(Received February 4, 1999; Accepted July 2, 1999)

Key Words—aerobic photosynthetic bacteria; bacteriochlorophyll; photosynthetic genes; photosynthetic regulatory genes; *Roseobacter denitrificans*; signal transduction; two-component regulatory system

Aerobic photosynthetic bacteria such as Roseobacter denitrificans grow chemoheterotrophically under aerobic dark conditions; however, unlike purple nonsulfur bacteria such as Rhodobacter sphaeroides and Rhodobacter capsulatus, aerobic photosynthetic bacteria do not grow phototrophically under anaerobic conditions (Shiba and Simidu, 1982). Nevertheless, Ros. denitrificans forms almost the same photosynthetic apparatus as those of purple nonsulfur bacteria even under highly aerobic conditions in darkness (Harashima et al., 1980; Iba et al., 1988). In fact, we demonstrated that in Ros. denitrificans, the levels of photopigments (bacteriochlorophyll and carotenoids) and mRNA levels of the *puf* operon encoding the pigment-binding polypeptides, both of the reaction center and light-harvesting complexes I, were very high even at nearly oxygen saturation of the growth culture (Nishimura et al., 1996). In purple nonsulfur bacteria, the photopigment formation and the expression of the puf operon is repressed under high oxygen tension

(for reviews, Bauer, 1995; Drews and Golecki, 1995). These results suggest that the mechanism for regulation by oxygen of the expression of the *puf* operon and of photopigment formation differs between aerobic photosynthetic bacteria and purple nonsulfur bacteria.

It is now widely accepted that in Rba. capsulatus and Rba. sphaeroides, a "two-component regulatory system" is involved in the signal transduction for the regulation of gene expression (e.g., *puf* operon) by oxygen tension (Eraso and Kaplan, 1994, 1995; Mosley et al., 1994; Sganga and Bauer, 1992). The regulatory system mainly consists of a sensor kinase (RegB or PrrB) and a response regulator (RegA or PrrA), and they are highly homologous between the two bacteria, probably reflecting a very similar response of both bacteria to oxygen tension. Downstream of the two-component regulatory gene cluster are several genes (spb, hvrA, hvrB, orf5, orf318, and ahcY), presumably involved in the signal transduction of oxygen or light (Buggy et al., 1994a, b; Mizoguchi et al., 1997; Sganga et al., 1992; Shimada et al., 1996). These additional genes are not always homologous in both bacteria, probably reflecting minor differences in the response to external environmental changes. Since the effect of oxygen tension on the formation of the photosynthetic apparatus is quite different between aerobic photosynthetic bacteria and purple nonsulfur bacteria as described above, it is anticipated that component(s) or their properties in the signal transduction of Ros. denitrificans differs from those of the purple nonsulfur bacteria. However, little is known about the regulatory components and their properties in Ros. denitrificans.

¹ The nucleotide sequence reported in this paper has been deposited in the DDBJ data base under the accession number AB020211.

^{*} Address reprint requests to: Dr. Ken-ichiro Takamiya, Department of Biological Sciences, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226–8501, Japan.

E-mail: ktakamiy@bio.titech.ac.jp

[†] Present address: Department of Life Science and Biotechnology, Faculty of Life and Environmental Science, Shimane University, Nishikawatsu-cho, Matsue 690–8504, Japan.

On the occasion of his retirement, this paper is dedicated to Professor Keiji Harashima who accomplished an epoch-making contribution to the research field of aerobic photosynthetic bacteria.

 Table 1.
 Identities of regulatory genes among photosynthetic bacteria.

Oharia	Identity of amino acid sequence (%)					
Strain	Ros. denitrificans	Rba. sphaeroides				
PrrA (RegA)						
Rba. sphaeroides	83	_				
Rba. capsulatus	79	83				
PrrB (RegB)						
Rba. sphaeroides	62	_				
Rba. capsulatus	57	60				
PrrC (SenC)						
Rba. sphaeroides	54	_				
Rba. capsulatus	48	49				
Orf5						
Rba. sphaeroides	73	_				
Rba. capsulatus	71	74				
Orf329						
Rba. sphaeroides	55	_				
AhcY						
Rba. sphaeroides	84	_				
Rba. capsulatus	84	92				

In this study, we identify the photosynthetic regulatory gene cluster in *Ros. denitrificans* and reveal that the organization of genes in the cluster is similar to that of *Rba. sphaeroides* and *Rba. capsulatus*. Furthermore, we investigate the expression of these genes under various growth conditions and discuss the possible mechanism for regulating the pigmentation and the expression of *puf* operon in *Ros. denitrificans*.

An aerobic photosynthetic bacterium *Ros. denitrificans* (ATCC 33942) was cultured as described previously (Nishimura et al., 1996, 1998). A genomic library was constructed with *Bam*HI-digested genomic DNAs and screened for *prrA* of *Ros. denitrificans* by using a *Rba. sphaeroides prrA*-specific probe. The preparation of RNA and Northern hybridization analysis was performed by using specific probes for each gene and 16S rRNA as described by Nishimura et al. (1996, 1998).

Ros. denitrificans had homologs of *prrA*, *prrB*, and *prrC* as reported by Matsumoto et al. (1997). As shown in Table 1, the identity of PrrAs (RegA) among these three bacteria was about 80%; for PrrBs (RegB), about 60%; and for PrrCs (SenC), about 50%, indicating that these three genes of *Ros. denitrificans* do not especially differ from the corresponding genes of *Rba. sphaeroides* and *Rba. capsulatus*. As with PrrA (or RegA) of *Rba. sphaeroides* (Eraso and Kaplan, 1994) and *Rba. capsulatus* (Du et al., 1998), a putative DNA binding domain, helix-turn-helix, was in the C-terminal region of PrrA of *Ros. denitrificans* (data not shown, Matsumoto et al., 1997), suggesting that PrrA of *Ros. denitrificans* was also a *trans*-factor.

As in purple nonsulfur bacteria, no distinct oxygenbinding domain or redox-sensing domain was found in PrrB of *Ros. denitrificans*, suggesting that PrrB of this bacterium did not also directly sense molecular oxygen or the redox state of the environment.

In purple nonsulfur bacteria, other additional regulator genes are in the flanking region downstream of *prrA*, such as *spb* and *orf318* in *Rba. sphaeroides*, *hvrA* and *hvrB* in *Rba. capsulatus*, and *orf5* and *ahcY* in both bacteria (Buggy et al., 1994a, b; Mizoguchi et al., 1997; Shimada et al., 1996). We therefore further analyzed the flanking region downstream of *prrA* of *Ros. denitrificans* and determined nucleotide and amino acid sequences of *orf5*, *orf318*, and *ahcY* homologs of *Rba. sphaeroides* (Mizoguchi et al., 1997) (Fig. 1).

Orf5 of 195 amino acid residues exhibited 73 and 71% amino acid identity to those in Rba. sphaeroides and Rba. capsulatus, respectively (Fig. 1; Table 1). No potential Shine-Dalgarno sequence was found upstream of orf5 in Ros. denitrificans, as in Rba. sphaeroides (Mizoguchi et al., 1997). Orf329 of 329 amino acid residues exhibited 55% amino acid identity to Orf318 in Rba. sphaeroides (Fig. 1; Table 1). Orf329 also exhibited 27 and 22% amino acid identities to Rfal and RfaJ of E. coli, respectively, which are involved in the biosynthesis of lipopolysaccharide (Pradel et al., 1992), as Orf318 of Rba. sphaeroides (Mizoguchi et al., 1997). Although the structure of lipopolysaccharide moiety of Ros. denitrificans has not been reported, Orf329 is possibly responsible for the biosynthesis of lipopolysaccharide or involved in the transcriptional regulation of photosynthesis genes, since the orf318 is found in the regulatory gene cluster of Rba. sphaeroides. The potential Shine-Dalgarno sequence upstream of orf329 is underlined in Fig. 1.

AhcY of 462 amino acid residues exhibited 84% amino acid identity to AchYs in Rba. sphaeroides and Rba. capsulatus (Fig. 1; Table 1). ahcY, which encodes S-adenosyl-L-homocysteine hydrolase, is thought to be involved in the regulation of bacteriochlorophyll biosynthesis (Sganga et al., 1992). The putative NAD⁺-binding domain (Fig. 1, boxed amino acids) was highly conserved among these bacteria (data not shown). The potential Shine-Dalgarno sequence upstream of ahcY is underlined in Fig. 1. As shown in Table 1, the identities of Orf5 and AhcY among the three bacteria are 71-74% and 84-92%. respectively, indicating that these genes of Ros. denitrificans do not especially differ from the corresponding genes of the other two bacteria.

The organization of the regulatory gene cluster of *Ros. denitrificans, prrB-prrC-prrA-orf5-orf329-ahcY* was somewhat simpler than that of the purple nonsulfur bacteria because *spb, hvrA*, and *hvrB* are lacking

Photosynthetic regulatory gene cluster

195	aagcttgctg * P	atcttgcaga M A A	cgaaaggtcg L L E	gctttgaaaa N H R A	cgctgacctg T F A	aggcgcggtgc D R T E	gtggctactg A P P	cgtgccggct R L T	tggcccgtg I E L	90 169	
169	aggaTCACGG	CATGGCAGCC	AGCAATTCGT	TGTGACGCGC	GGTAAAGGCG	TCGCGCGTCTC	AGCCGGGGGC	CGCAGGGTAA	TTTCCAGCC	130	↑
100	CTTGCATGAT	CTTGTGATCC	GGTGTGCCAA	AGAACTTGCT	GGCCTCGGTT	TCTGAAAAACC	GGCGATCTGC	GTGGCTTCCA	TCCATGCGC	270	
138	S I K D	A K K	I Q R	K V A V	PIA	A P L G	FRL	H I A	A T L	109	
108	R E D L	A E Y	G P G	V A A K	V P S	I M D G	I V Y	E P A	D H L	79	
70	GCTCATCCAG	CGCCTCATAC	CCCGGCCCGA	CAGCGGCCTT	TACCGGGGAG	ATCATATCGCC	GATCACGTAT	TCCGGGGCGT	CATGCAAAA	450	
70	GGGCTGCCAA	TCGCCATTTC	GCGGGTGATT	TATCAGCGAT	GCGGGTAAAA	AGCGTCTCTAC	CAAAAGCGAA	TGTTCCGCAA	CTGAATAGG	49 540	
48	A H A G	S T Q	G N W	R A V F	A L G	R A I D	EIE	V D V	P T P	19	
18	D L L D	L R R	G S L	M R Q W	A R K	M	TTCAATTICT	ACGTCAACCG	GAGTCGGAT	1	orf5
329						* R H V	TLE	N P F R	ALY	317	▲
316	H A I	K R A	T I L H	K I L	M K T	M K M S	D P A	L R R G	G T E	287	
200	GTGCGCTATT	TTCCGGGCTG	TTATCAGATG	TTTGATCAGC	ATTTTCGTCA	TTTTCATACTG	TCCGGCGCGA	GTCTTCTACC	CCCCGTTTC	810	
280	CACTTGTCTG	GAACCTGAAA	AGTGTGTCCT	GATGAAACGA	TCAAATGAGT	D A F R CAGCAAAGCGC	GGTTCGAACT	GTCCGCTCTC	GTCTTTCCA	257 900	
256	PKA	PGI	FHII	N P Y	AFV	AFLR	TSW	S Y Q W	N W V	227	
226	P S I	E A W	D N Q L	V A N	Y L N	Q D H R	K L D	R R R A	R G F	990 197	
100	GGGGCTTATT	TCGGCCCAGT	CGTTTTGCAG	CACCGCATTG	TAAAGGTTCT	GATCATGCCGT	TTCAAATCCC	GACGGCGTGC	GCGGCCAAA	1080	
196	E V C TTCGACGCAA	CGTCGCATCA	GCTCCTGTTC	T Y A GGTATAGGCC	Q V D TGCACATCCA	M L M V TTAGCATTACT	G A N CCTGCGTTGA	F'YAS AGTATGCAGA	P P I TGGGGGTAT	167	
166	GKI	TNR	KNQR	K P T	RWQ	VNDR	VSA	I C H P	A V D	137	
136	I D F	L A N	F D G G	O V F	CGCCATTGCA I D S	CGTTGTCGCGA D L Y L	ACCGATGCGA I K D	TGCAATGCGG Y E G A	TGCTACATC F A T	1260	
100	AATATCAAAA	AGCGCGTTGA	AATCGCCACC	TTGAACGAAA	ATATCCGAAT	CCAGATAGAGA	ATTTTGTCAT	ATTCACCGGC	AAAAGCGGT	1350	
106	CGGCAAGGCG	ATGCGCAAAT	Y V D H ATACGTCATG	T K G CGTTTTGCCC	K D L TTATCCAAGC	R L G E GCAGGCCTTCA	F V D AAGACGTCGC	G V D V CGACATCTAC	H C L GTGACAAAG	1440	
76	RIG	LGA	LSDP	LVV	AQH	GYCI	CID	FEPK	EVL	47	
46	S A I	Q A A	A H A A	AAGCACAACT F V L	GCCTGATGAC Y N O	CGTAACAGATA D C C F	CAGATGTCGA V I A	ATTCCGGCTT R K H R	A P R	1530 17	
	CGATGCAATC	TECCCCCCCCC	CATGTGCCCC	AAAAACAACC	TAATTCTGGT	CCCAACAAAAG	ACGATTGCCC	TTTTTCTCCCC	TCCCCCCCC	1620	1
16		100000000		MMMCAA00	TIMITICIOUT	COCIMICITATIO	neomineece	1111010000	100000000	1	+220
16	S A T CGATGCCGTC	L R V AACCTGACGT	H F S E GAAATGACTC	A P A TGCGGGTGCT	T S M GTGGACATag	ccgggcgtttg	ctcgttgcgt	totgtaaatg	tgtcgctgg	1 0 1710	rf329
16	S A T CGATGCCGTC cggggctgta	L R V AACCTGACGT ctcaagacac	H F S E GAAATGACTC gcccacttct	A P A TGCGGGTGCT agagaactga	T S M GTGGACATag aagatattgt	ccgggcgtttg ccagcactcaa	ctcgttgcgt ctttttggag	totgtaaatg	tgtcgctgg	1 0 1710 1800	rf329
16	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc	L R V AACCTGACGT ctcaagacac tacgcttgta agagcctcgg	H F S E GAAATGACTC gcccacttct aagaggatgt agtgcacgcG	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC	ctcgttgcgt ctttttggag gccctgcaac CCTTGCCGAG	tctgtaaatg tcacacatgc gggtggtaga TTTGGCCGCA	tgtcgctgg agggcaaac gcagcggca AAGAGCTTG	1 0 1710 1800 1890 1980	rf329
16 1	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc	L R V AACCTGACGT ctcaagacac tacgcttgta agagcctogg	H F S E GAAATGACTC gcccacttct aagaggatgt agtgcacgcG M	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A	ctcgttgcgt ctttttggag gccctgcaac CCTTGCCGAG L A E	tctgtaaatg tcacacatgc gggtggtaga TTTGGCCGCA F G R K	tgtcgctgg agggcaaac gcagcggca AAGAGCTTG E L D	1 0 1710 1800 1890 1980 21 2070	nf329 ahcY
16 1 22	S A T CGATGCCGTC cggggotgta ccaatacgat aagatcattc ATATTGCCGA I A E	L R V AACCTGACGT ctcaagacac tacgcttgta agagcctcgg AACCGAAATG T E M	H F S E GAAATGACTC gcccacttct aagaggatgt agtgcacgcG M CCGGGGGCTGA P G L M	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GGCGACAGCAA G D S K	ctcgttgcgt ctttttggag gccctgcaac CCTTGCCGAG L A E GCCGCTGGCC P L A	totgtaaatg toacacatgo gggtggtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I	tgtcgctgg agggcaaac gcagcggca AAGAGCTTG E L D TTGTCGGCT V G S	1 0 1710 1800 1890 1980 21 2070 51	ahcY
16 1 22 52	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M	L R V AACCTGACGT otcaagacac tacgottgta agagcotogg AACCGAAATG T E M GACGATCCAG T I O	H F S E GAAATGACTC gcccacttct agggggggtgt M CCGGGGCTGA P G L M ACGGCAGTTC T A V L	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC	T S M GTGGACATag aagatattgt tggcattcgt F I V CGCGGAATAT A E Y GCTTGTTGCC L V A	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GGCGACAGCAA G D S K TTGGGTGCCGA	ctcgttgcgt ottttggag gccctgcaac CCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V P W	totgtaaatg toacacatgc gggtggtaga F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N	tgtcgctgg agggcaaac gcagcggca AAGAGCTTG E L D TTGTCGGCT V G S ATATCTTT T F S	1 0 1710 1800 1980 21 2070 51 2160 81	ahcY
16 1 22 52	S A T CGATGCCGTC cgggggtgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA	L R V AACCTGACGT ctcaagacac tacgcttgta agagcotogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG	H F S E GAAATGACTC gcccacttct aagaggatgt agtgcacgcG CCGGGGCTGA P G L M ACGGCAGTTC T A V L GCGGCCATCG	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCCGGCGG	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V GCGGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GCGCGACGCAA G D S K TTGGGGTCCGA L G A D TTCGCGATCAA	ctogttgcgt ottittggag gccctgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGGTCAGTCG	totgtaaatg tocacacatgo ggtggtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC	tgtcgctgg agggcaaac gcagcggca AAGAGCTTG E L D TTGTCGGCT V G S ATATCTTT I F S ATTGGGATT	1 0 1710 1800 1980 21 2 2070 51 2160 81 2250	ahcY
16 1 22 52 82	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG	L R V AACCTGACGT ctcaagacac tacgcttgta agagcctcgg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGCGCG H A A	H F S E GAAATGACTC goccaottot aagaggatgt agtgoacgcG CCGGGGCTGA P G L M ACGGCAGTTC T A V L GCGGCCATCG A A I A TTTGGAGGATG	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGGG A G G GTCCCAACCT	T S M GTGGACATag aagatattgt tggoattogt CTTCATCGTA F I V GCCGGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V A GACGCCTGTT	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GCGCACAGCAA G D S K TTGGGGTGCCGA L G A D TTCGCGATCAA F A I K GATGGCGGTGA	ctogttgcgt ottittggag gccctgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGCCGCCTGG V R W GGGTCAGTCG G Q S TGCGCCCCC	totgtaaatg todoacatgc gggtggtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC	tgtcgctgg agggcaaac gcagcggca AAGAGCTTG E L D TTGTCGGCT V G S ATATCTTT I F S ATTGGGATT W D Y TTGGTGCCC	1 0 1710 1800 1990 21 2070 51 2160 81 2250 111 2340	ahcY
16 1 22 52 82 112	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R	L R V AACCTGACGT ctcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M	H F S E GAAATGACTC goccacttct aagaggatgt agtgoacgoG CCGGGGCTGA P G L M ACGGCAGTTC T A V L GCGGCCATCG A A I A TTTGAGATG F E D G	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGCCTTGCG A L R TGATCGAAAC I E T CGGCGGGCGG A G G GTCCCAACCT P N L	T S M GTGGACATag aagatattgt tggoattogt CTCATCGTA F I V GCCGGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GGCGACAGCAA G D S K TTGGGTGCCGA L G A D TTCGCGATCAA F A I K GATGGCGTGA D G G D	ctogttgcgt ottittggag gccctgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGCCGCCTGG V R W GGGTCAGTCG G Q S TGCGACCTC A T L	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L	tgtogotgg agggoaaac goagoggoa AAGAGCTTG E L D TTGTCGGCT V G S ATATCGTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R	1 0 1710 1800 1980 21 2 2070 51 2160 81 2250 111 2340 141	ahcY
16 1 22 52 82 112 142	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCCGACGA CA C CGACGCAGGA C D R GTGCCGAGGG A E A	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GCGTTCATG S F M GGGGTGAAGAG G E E	H F S E GAAATGACTC goccacttct aagaggatgt agtgcacgcG M CCGGGGCTGA P G L M ACGGCAGTTC T A V L GCGGCCATCG A A I A TTTGAGGATG F E D G ATTATCCCCG I I P V	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGGCGG A G G GTCCCAACCT P N L TGCCCACATC P T S	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V GCCTGGTGTGCC L V A GACGCCTGTT T P V GACGCCTGTT T L D CGAAGAGGAA E E E	ccggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GCGCACAGCAA G D S K TTGGCTGCCGA L G A D TTCCGCGATCAA F A I K GATGGCCGCTGA D G G D GAGGCCATCAA E A I K	ctagttgcgt attittggag gccttgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGGTCAGTCG G Q S TGCGACCTC A T L GCCGCAGATC A O I	totgtaaatg gggtggtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGCC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M	tgtogotgg agggoaaac goagoggoa AAGAGCTTG E L D TTGTCGGCT V G S ATATCGTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R TGGCGCGCT A A S	1 0 1710 1800 1890 1980 21 2070 51 2160 81 2250 111 2340 141 2430	ahcY
16 1 22 52 82 112 142	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGAGGC A E A CTCCCGGTGG	L R V AACCTGACGT otcaagacac tacgottgta agagootogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA	H F S E GAAATGACTC goccacttct aagaggatgt agtgcacgcG P G L M ACGGCGCTGA P G L M ACGGCCATCG A A I A TTTGAGGATG F E D G ATTATCCCG I I P V GTGCGTGATC	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGGCGG A G GTCCCAACCT P N L TGCCCACATC P T S AGATCAAAGG	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V GCGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GGCGACAGCAA G D S K TTGGGTGCCGA L G A D TTCGCGATCAA F A I K GATGGCGGTGA D G G D GAGGCCATCAA E A I K GAGACGACAAC	ctagttgcgt attittggag gcctgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGGTCAGTCG G Q S TGCGACCTCC A T L GGCGCAGATC A Q I TGGCGCTCAC	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCC X K R M CGTCCTCTATG	tgtogotgg agggoaaac goagoggoa AAGAGCTTG E L D TTGTCGGCT V G S ATATCGTTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R TGGCGCGGT A A S ACCTCCTCA	1 0 1710 1800 1800 1980 2070 51 2160 81 2250 111 2340 141 2430 171 0 0 0 0 0 0 0 0 0 0 0 0 0	ahcY
16 1 22 52 82 112 142 172	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGAGGG A E A CTCCCGGTG P G W ACCAGGGGGCA	L R V AACCTGACGT ctcagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC	H F S E GAAATGACTC goccacttct aagaggatgt agtgcacgcG P G L M ACGGCAGTCG T A V L GCGGCCATCG A A I L TTTGAGGATG F E D G ATTATCCCCG I I P V GTGCGTGATC V R D Q CCTGCGATCA	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAAC I E T CGGCGGGCGG A G GTCCCAACCT P N L TGCCCACACCT P N L TGCCCACACC P T S AGATCAAAGG I K G ATGTGAATGA	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGATATA A E Y GCTTGTTGCC L V A GACGCCTGTT T P V CGAAGAGGA I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC	ccggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GGCGACAGCAA G D S K TTGGCGGCGCGA L G A D TTCGCGATCAA F A I K GATGGCGCGTGA D G G D GAGGCCATCAA E A I K GAGACGACAAC E T T T	ctagttgcgt cttttggag gcctgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGTCAGTCG G Q S TGCGACCTCC A T L GGCGCAGATC A Q I TGGCGTTCAC G V H TGACAACAAA	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGCCGCA	tgtogotgg agggoaaco goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R TGGCGCGCT A A S ACCTCGTCA L V K AGGAATCGC	1 0 1710 1800 1800 1980 2070 51 2160 81 2250 111 2340 141 2430 171 2520 201 2610	ahcY
16 1 22 52 82 112 142 172 202	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGAGGG A E A CTCCCGGTG P G W ACCAGGGGCA Q G Q D	L R V AACCTGACGT ctcagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F	H F S E GAAATGACTC goccacttct aagaggatgt agtgcacgcG P G L M ACGGCAGTCG T A V L GCGGCCATCG A A I A TTTGAGGATG F E D G ATTATCCCCG I I P V GTGCGTGATC V R D Q CCTGCGATCA P A I N	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAACC I E T CGGCGGGCGG A G GTCCCAACCT P N L TGCCCACACCT P N L TGCCCACACC P T S AGATCAAAGG I K G ATGTGAATGA V N D	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGATATA A E Y GCTTGTTGCC L V A GACGCCTGTT T P V CGAAGAGGA I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T	ccgggcgtttg ccagcactcaa tctgttgcagc AAAGATATTGC K D I A GGCGACAGCAA G D S K TTGGCGGCGCGA L G A D TTCGCGATCAA F A I K GATGGCGCGTGA D G G D GAGGCCATCAA E A I K GAGACGACAAC E T T T AAGTCCAAGTT K S K F	ctagttgcgt ctttttggag gccttgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGTCAGTCG G Q S TGCGACCTCC A T L GGCGCAGATC A Q I TGGCGTTCAC G V H TGACAACAAA D N K GGCGCGCC	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGCCTGCA Y G C K	tgtogotgg agggoaaco goagoggoa AAGAGCTTG E L D TTGTCGGCT V G S ATATCGTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R TGGCGCGGT A A S ACCTCGTCA L V K AGGAATCGC E S L	1 0 1710 1800 1800 1980 21 2 2160 81 2250 111 2340 141 2430 171 2520 201 231	ahcY
16 1 22 52 82 112 142 172 202 232	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGACGG C A E A CTCCCGGTGG P G W ACCAGGGGCA Q G Q TGGTGGATGG V D G	L R V AACCTGACGT otcaagacac tacgottgta agagootogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCCATTC L P F TATCCGCCGC I R R	H F S E GAAATGACTC gcccacttct aagaggatgt agtgcacgcG P G L M ACGGCAGTCC T A V L GCGGCCATCG A A I A TTTGAGGATG F E D G ATTATCCCCG I I P V GTGCGTGATC V R D Q CCTGCGATCA P A I N GCCACGGATCA A T D T	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGGG A G GTCCCAACCT P N L TGCCCACATCC P T S AGATCAAAGG I K G ATGTGAATGA V N D CGATGATGACGC M M A	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTT G K V	$\begin{array}{c} \mbox{coggcqsttg} \\ \mbox{coggcqsttg} \\ \mbox{cogqcactcaa} \\ \mbox{cogqcactcaa} \\ \mbox{cogqcactcaa} \\ \mbox{cogqcactcaa} \\ \mbox{GCQACCAA} \\ \mbox{G} & D & S & K \\ \mbox{TTGGCGCGCCQA} \\ \mbox{G} & D & S & K \\ \mbox{TTGGCGCGCGACAA} \\ \mbox{F} & A & I & K \\ \mbox{GATGGCGCGTGA} \\ \mbox{D} & G & G & D \\ \mbox{GAGGCCATCAA} \\ \mbox{E} & A & I & K \\ \mbox{GAGGCCATCAA} \\ \mbox{E} & A & I & K \\ \mbox{GAGGCCACACAC} \\ \mbox{E} & T & T & T \\ \mbox{AAGTCCAAGGTT} \\ \mbox{K} & S & K & F \\ \mbox{GCCGCGCGGCGACAA} \\ \mbox{A} & V & W \\ \end{array}$	ctagttgcgt ctttteggag gccctgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGTCAGTCG G Q S TGCGACCTCC A T L GGCGCAGTCC A Q I TGGCGTTCAC G V H TGACAACAAA D N K <u>GGCTACGGC</u> G Y G	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGCCTGCA Y G C K CATGTGGGCA D V G K	tgtogotgg agggoaaco goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R TGGCGCGT A A S ACCTCGTCA L V K AGGAATCGC E S L AAGGATCCG G S A	1 0 1710 1800 1800 1980 21 2 2160 81 2250 111 2340 141 2430 171 2520 201 2610 2700 261	ahcY
16 1 22 52 82 112 142 172 202 232 262	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGACGG A E A CTCCCGGTGG P G W ACCAGGGGCA Q G Q TGGTGGATGG V D G CTGCCAGTCG	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCGG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCCGTGCCCC	A I P S E GAAATGACTC gcccacttct aagaggatgt agtgcacgcG P G L M ACGGCAGTCC T A V L GCGGCCATCG A A I A TTTGAGGATG F E D G ATTATCCCCG I I P V GTGCGTGATC V R D Q CCTGCGATCA P A I N GCCACGATA A T D T GCCACGGATA	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGG A G GTCCCAACCT P N L TGCCCACATC P N L TGCCCACATC P T S AGATCAAAGG I K G ATGTGAATGA V N D CGATGATGAC M M A TAAAAGTCAC	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTT G K V CGAAGTGGAC	$\begin{array}{c} \mbox{coggcqtttg} \\ \mbox{coggcqctta} \\ \mbox{cogqcactcaa} \\ \mbox{cotgcactcaa} \\ \mbox{cotgcactcaa} \\ \mbox{dcactcaa} \\ \mbox{cogcactcaa} \\ \mbox{dcactcaa} \\ dc$	ctagttgcgt ctttteggag gccttgcaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGTCAGTCG G Q S TGCGACCTCC A T L GGCGCAGACC C V H TGCCGTTCAC G V H TGACAACAAA D N K <u>GGCTACGGC</u> <u>G Y G</u> <u>GCTGCAAGCC</u>	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGCCTGCA Y G C K GATGTGGGCA D V G K GCCATGGACG	tgtogotgg agggoaaco goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGATT W D Y TTGGTGCCC G A R TGGCGCGT A A S ACCTCGTCA L V K AGGAATCGC E S L AAGGATCGG G S A GTTCCAGG	1 0 1710 1800 1800 1980 2107 2160 81 2250 111 2340 141 2430 171 2520 201 2610 2700 261 2790	ahcY
16 1 22 52 82 112 142 172 202 232 262	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGAGGC A E A CTCCCGGTGG P G W ACCAGGGCA Q G Q TGGTGGATGG V D G CTGCCAGTCG TGCCAGTCG TGTCCAGCT	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCGAGGTGCC R G A CGAAGATGTG	$ \begin{array}{c} \text{GragGatcgeg} \\ \text{Gataatgagatgt} \\ \text{agaggatgt} \\ agagg$	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGGG G TCCCAACCT P N L TGCCCACATC P N L TGCCCACATC P T S AGATCAAAGG I K G ATGTGAATGA V N D CGATGATGGC M M A TAAAAGTCAC K V T CGGATATCTT	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTT G K V CGAAGTGGAC E V D TATCACAGACG	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgqcqcta} \\ \mbox{ccgqcqcta} \\ \mbox{ccgqcqcta} \\ \mbox{ccgcqcqc} \\ \mbox{A} \\ \mbox{G} \\ \mbox{C} \\ \mbox{C} \\ \mbox{C} \\ \mbox{ccgqcqc} \\ \mbox{ccgqcqc} \\ \mbox{ccgqcqc} \\ \mbox{ccqcqc} \\ \mbox{ccqcqc} \\ \mbox{ccqcqc} \\ \mbox{ccqcqc} \\ \mbox{ccqcqcqc} \\ \mbox{ccqcqcqc} \\ \mbox{ccqcqcqc} \\ \mbox{ccqcqcqcqc} \\ \mbox{ccqcqcqcqc} \\ \mbox{ccqcqcqcqc} \\ \mbox{ccqcqcqcqcqc} \\ \mbox{ccqcqcqcqcqc} \\ \mbox{ccqcqcqcqcqcqcqc} \\ ccqcqcqcqcqcqcqcqcqcqcqcqcqcqcqcqcqcqc$	$\begin{array}{c} \text{ctgttgcgt}\\ \hline \text{ctttteggag}\\ \hline \text{gcctgcaac}\\ \text{ccTrgcccgAg}\\ \text{L} & \text{A} & \text{E}\\ \hline \text{gccgcTgcgcc}\\ \text{P} & \text{L} & \text{A}\\ \hline \text{TgTccgcTgg}\\ \hline \text{V} & \text{R} & \text{W}\\ \hline \text{ggcctagtcg}\\ \hline \text{G} & \text{Q} & \text{S}\\ \hline \text{TgcgAcCcTC}\\ \hline \text{A} & \text{T} & \text{L}\\ \hline \text{ggcgCagAtc}\\ \hline \text{A} & \text{V} & \text{H}\\ \hline \text{TgcAcAacAaA}\\ \hline \text{D} & \text{N} & \text{K}\\ \hline \hline \text{ggcgTacGgc}\\ \hline \text{G} & \text{Y} & \text{G}\\ \hline \hline \text{gcgcTacgcc}\\ \hline \text{C} & \text{Y} & \text{G}\\ \hline \hline \text{gcgcacAacCa}\\ \hline \text{G} & \text{C} & \text{G}\\ \hline \text{gcgcacAacCa}\\ \hline \text{G} & \text{C} & \text{G}\\ \hline \text{gcgcacAacCa}\\ \hline \text{G} & \text{C}\\ \hline \text{gcgacTacggc}\\ \hline \text{G}\\ \hline \text{gcgacTacggc}\\ \hline \text{H}\\ \hline \end{array}$	totgtaaatg gggtgtaga F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GATGTGGGCA D V G K GCCATGGAGC	tgtogotgg agggcaaac goagoggca AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGACT T TGGTGCCC G A R TGGCGGCGT A A S ACCTCGTCA A A S ACCTCGTCA AAGGATCCC G S A GTTTCGAGG G S A GTTTCGAGG	1 0 1710 1800 1800 1980 21 2 2160 81 2250 111 2340 141 2430 171 2520 201 2610 2700 261 2790 291 2880	ahcY
16 1 22 52 82 112 142 172 202 232 262 292	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGACGG A E A CTCCCGGTGG P G W ACCAGGGCA Q G Q TGGTGGATGG V D G CTGCCAGTCT TTGTCCTGCT V L L CATCCAACA	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG G CACGCGGCG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCGGGGGCCC CGAAGATGTG E D V	A TITIGAGACTC GCCGCGCCTGA P G L M ACGGCAGTGA P G L M ACGCCAGTGA T A V L GCGGCCATCG A A I A TTTGAGGATG F E D G ATTATCCCCG I I P V GTGCGTGATCC V R D Q CCTGCGCATCA P A I N GCCACCGATA A T D T GGTGCCCCGCG G A R V GTGGGATCGG V G S A CTTCCACACA	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGGCGG A G GTCCCAACCT P N L TGCCCACATC P T S AGATCAAAGG I K G ATGTGAATGA V N D CGATGATGATGA C M M A TAAAAGTCAC CGGATATCTT D I F	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E GGCAAGGGT G K V CGAAGTGGAC E V D TATCACCAGACG I T T T	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgactcaa} \\ \mbox{ctgtgcagc} \\ \mbox{AAAGATATTGC} \\ \mbox{K} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	ctogttgogt ctttteggag gccttgoaac cCTTGCCGAG L A E GCCGCTGGCC P L A TGTCCGCTGG V R W GGTCAGTCG G Q S TGCGACCTC A T L GGCGCAGACC C V H TGACAACAAA D N K <u>GGCTCACGGC</u> <u>G Y G</u> GCTGCAAGCC L Q A GGACGTGATC D V I TGCCGCAACCA	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCGCA Y G C K GCCATGGACG A M D G GCCATGGACG R I E H	tgtogotgg agggoaaac goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGACT Y TGGTGCCC G A R TGGCGCGGCT A A S ACCTCGTCA A A S ACCTCGTCA AAGGATCGC G S A GTTTCGAGG F E V ATATCGGTG M R A CCCCGADATA	1 0 1710 1800 1800 1980 21 2 2070 51 2160 81 2250 111 2340 141 2430 171 2520 201 2610 2700 261 2790 291 2880 321 2800 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2110 2100 2110 2100 2110 2100 2110 2100 2110 2100 210 21	nf329
16 1 22 52 82 112 142 172 202 232 262 292 322	S A T CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA I A E CGCTGCACAT L H M CGACGCAGGA T Q D ACCTCGACCG L D R GTGCCGAGGC A E A CTCCCGGTGG P G W ACCAGGGGCA Q G Q TGGTGGATGG V D G CTGCCAGTCT A S L TTGTCCTGCT V L L CGATGAAGA M K D	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K AATGCCATTC L P F TATCCGCCGC I R R GCGAGGAGCGCC R G A CGAAGATGTG E D V CATGGCGATC M A I	$\begin{array}{c} \text{Grader} \\ \text{Grader} \\$	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGGG G G G GTCCCAACCT P N L TGCCCACATC P N L TGCCCACATC P T S AGATCAAAGG I K G ATGTGAATGA V N D CGATGATGATGA V N D CGATGATGCC K V T CGGATATCTT D I F TTGGCCACTT G H F	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTT G K V CGAAGTGGAC E V D TATCACGACG I T T TGACAATGAA D N E	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgqcqcta} \\ \mbox{ccgqcqcta} \\ \mbox{ccgqcqcta} \\ \mbox{ccgcqcqc} \\ \mbox{A} \\ \mbox{G} \\ \mbox{C} \\ \mbox{C} \\ \mbox{C} \\ \mbox{ccgqcqc} \\ \mbox{ccgqcqc} \\ \mbox{ccgqcqcqc} \\ \mbox{ccgqcqcqc} \\ \mbox{ccgqcqcqc} \\ \mbox{ccgqcqcqcqc} \\ \mbox{ccgqcqcqcqc} \\ \mbox{ccgqcqcqcqc} \\ \mbox{ccgqcqcqcqcqc} \\ \mbox{ccgqcqcqcqcqc} \\ \mbox{ccgqcqcqcqcqcqc} \\ \mbox{ccgqcqcqcqcqcqcqc} \\ ccgqcqcqcqcqcqcqcqcqcqcqcqcqcqcqcqcqcqc$	Ctores and the construction of the constructio	totgtaaatg toacacatgc gggtgtaga F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GCCATGGACG A M D G CGCATCGACC R I E H AACCACAAGT N H K W	tgtogotgg agggoaaac goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGACT T GGCGCGT A A S ACCTCGTCA A A S ACCTCGTCA AAGGATCGC G S A GTTTCGAGG G S A GTTTCGAGG G S A GTTTCGAGG M R A GGACCAATA T N I	1 0 1710 1800 1800 1980 21 2 2160 81 2250 111 2340 141 2430 171 2520 201 2700 261 2790 2970 351	rf329
16 1 22 52 82 112 142 172 202 232 262 292 322	Solution of the second	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCGAGGATCG R G A CGAAGATGTG E D V CATGGCGATC M A I GGTAGACATG	$ \begin{array}{c} \text{Grass}_{F} \\ \text{GAAATGACTC} \\ \text{gcccacttct} \\ \text{acgaggatgt} \\ \text{acgacgcactcd} \\ \text{acgacgcactcd} \\ \text{gcccactcd} \\ \text{gcccactcacc} \\ \text{gcccactcacc} \\ \text{gcccactcacc} \\ \text{acgaccatcc} \\ \text{gcccaccacc} \\ \text{acgaccatcc} \\ \text{gcccaccacc} \\ \text{gccaccacc} \\ \text{gccaccaccacc} \\ \text{gccaccaccaccaccacc} \\ \text{gccaccaccaccacc} \\ \text{gccaccaccaccacc} \\ \text{gccaccaccaccaccaccaccaccacc} \\ gccaccaccaccaccaccaccaccaccaccaccaccacca$	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGGG G G CCCAACCT P N L TGCCCACATC P T S AGATCGAATGA V N D CGATGATGATGA V N D CGATGATGGC M M A TAAAAGTCAC K V T CGGATATCTT D I F TTGGTCACTT G H F CGAATGGCAA	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTGGAC E V D TATCACGACG I T T TGACAATGAA D N E TCGTCTCAT	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgqcactcaa} \\ \mbox{ccgqcactcaa} \\ \mbox{ccgqcactcaa} \\ \mbox{ccgcqcactcaa} \\ \mbox{GcQacCaA} \\ \mbox{G} & D & S & K \\ \\ \mbox{TrGGGGGCGCG} \\ \mbox{L} & G & D \\ \\ \mbox{GATGGCGGTGAA} \\ \mbox{CcGATGGTAAT} \\ \mbox{AcGGCGACAA} \\ \mbox{C} & F & K & F \\ \\ \mbox{GcGATGGCAACAA} \\ \mbox{C} & F & K & F \\ \\ \mbox{GcGATGGCAACAA} \\ \mbox{C} & G & N & K \\ \\ \mbox{AcGGGCACAAA} \\ \mbox{T} & G & N & K \\ \\ \mbox{AcGGGCACACAA} \\ \mbox{T} & G & N & K \\ \\ \mbox{AcGGCGACACAA} \\ \mbox{T} & G & N & K \\ \\ \mbox{AcGGCGCACACA} \\ \mbox{T} & G & N & K \\ \\ \mbox{AcGGCGCACACA} \\ \mbox{T} & G & N & K \\ \\ \mbox{AcGGCGCACACA} \\ \mbox{T} & G & N & K \\ \\ \mbox{AcGGCGTTGCTGC} \\ \mbox{L} & Q & V & A \\ \\ \mbox{CCGTTCTCTAA} \\ \mbox{C} & F \\ \\ \mbox{CCGTTCTCTGA} \\ \mbox{C} & F \\ \\ \mbox{C} & F \\ \\ CCGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT$	Ctore contraction of the contract of the contr	totgtaaatg toacacatgc gggtgtaga F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GATGTGGGCA D V G K GCCATGGACC A M D G CGCATCGACC R I E H AACCACAAGT N H K W CTGAACCTCG	tgtogotgg agggcaaac gcagcggca AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGACT T GGCGCGGCT A A S ACTCGTCA A A S ACCTCGTCA AAGGATCGC G S A GTTTCGAGG G S A GTTTCGAGG G S A GTTTCGAGG G S A GTTTCGAGG M R A GGACCAATA T N I GGAACCGTA N 2 7	1 00 1710 1800 1800 1980 2107 51 2160 81 2250 111 2340 141 2430 171 2520 201 2700 2610 2790 291 2880 321 3060 321 3060	rf329
16 1 22 52 82 112 142 172 202 232 262 292 322 352	Solution of the second	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCGAGGATCTG E D V CATGGCGATC M A I GGTAGACATG V D M GTCCTTCGTG	$ \begin{array}{c} \text{GARATGACTC}\\ \text{GCAAATGACTC}\\ \text{gcccacttct}\\ \text{acgaggatgt}\\ \text{acgccgcctca}\\ \text{acgagcactct}\\ \text{acgccactcc}\\ \text{gcccactcacc}\\ \text{gcccaccaccc}\\ \text{gcccaccaccc}\\ \text{gcccacccccc}\\ \text{gcccaccccccc}\\ \text{gcccaccccccc}\\ \text{gcccacccccccc}\\ \text{gccccccccccc}\\ \text{gccccccccccc}\\ \text{gcccccccccccc}\\ \text{gcccccccccccc}\\ \text{gcccccccccccc}\\ \text{gcccccccccccc}\\ \text{gcccccccccccccccccc}\\ \text{gcccccccccccccccccccccccc}\\ gcccccccccccccccccccccccccccccccccccc$	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGG G G CCCAACCT P N L TGCCCACATC P T S AGATCGAATGA V N D CGATGATGATGA V N D CGATGATGAC M M A <u>TAAAAGTCAC</u> K V T CGGATATCTT D I F TTGGTCACTT G H F CGAATGGCAA N G N CCTTTACCAA	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTGAC C E V D TATCACGACG I T T TGACAATGAA D N E TCGGCTCATGA	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccgcqcactaa} \\ \mbox{GCQaCAGCAA} \\ \mbox{G} D S K \\ \mbox{TrGGGGGCGCG} \\ \mbox{L} G A D \\ \mbox{TrGGGGGGGACAA} \\ \mbox{F} A I K \\ \mbox{GAGGCCGGTGAA} \\ \mbox{CcGATTGCGC} \\ \mbox{E} T T T \\ \mbox{AAGTCCAAGTT} \\ \mbox{K} S K \\ \mbox{F} \\ \mbox{GCCGTGGTAAT} \\ \mbox{A} \\ \mbox{ACGGCCACAA} \\ \mbox{C} P I C A \\ \mbox{ACGGCCACAA} \\ \mbox{T} G N K \\ \mbox{ATTCAGGTTGCGC} \\ \mbox{L} Q V A \\ \mbox{CGGCATCTCTGA} \\ \mbox{L} L S \\ \mbox{E} \\ \mbox{GCGCACCAA} \\ \mbox{C} \\ \mbox{CGCATCTCGA} \\ \mbox{C} \mbox{C} \\ \mbox{C} \ \mb$	$\begin{array}{c} \text{ctgttgcgt}\\ \hline \text{ctttteggag}\\ \hline \text{gtctteggag}\\ \text{gccctgcaac}\\ \text{ccTTGCCGAG}\\ \text{L} & \text{A} & \text{E}\\ \hline \text{gCCGCTGGCC}\\ \text{P} & \text{L} & \text{A}\\ \hline \text{TGTCGCTGG}\\ \hline \text{V} & \text{R} & \text{W}\\ \hline \text{GGTCAGTCG}\\ \hline \text{G} & \text{Q} & \text{S}\\ \hline \text{TGCGACCTC}\\ \hline \text{A} & \text{T} & \text{L}\\ \hline \text{GGCGCAGACC}\\ \hline \text{A} & \text{Q} & \text{I}\\ \hline \text{TGCACAACAAA}\\ \hline \text{D} & \text{N} & \text{K}\\ \hline \text{GGGCTACGGC}\\ \hline \text{G} & \text{V} & \text{H}\\ \hline \text{TGCACAACAAAC}\\ \hline \text{D} & \text{N} & \text{K}\\ \hline \text{GGGCTACGGC}\\ \hline \text{C} & \text{Q} & \text{A}\\ \hline \text{GGACGTGAACC}\\ \hline \text{D} & \text{V} & \text{I}\\ \hline \text{TGCCCCCAAG}\\ \hline \text{A} & \text{L} & \text{K}\\ \hline \text{GGGGCGCTAGGC}\\ \hline \text{G} & \text{R} & \text{L}\\ \text{ACTGTGGACC}\\ \end{array}{}$	totgtaaatg toacacatgc gggtgtaga F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GATGTGGGCA A M D G CGCATCGAGC R I E H AACCACAAGT N H K W CTGAACCTCG L N L G CGTGCGGATG	tgtogotgg agggoaaac goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTT I F S ATTGGGACT X D Y TTGGTGCCC G A R TGGCGGCGT A A S ACCTCGTCA A A S ACCTCGTCA AAGGATCGC G S A GTTTCAGG G S A GTTTCCAGG G S A GTTTCCAGG M R A GGACCAATA T N I GGAACGCGA N A T CGTATGACA	1 00 1710 1800 1800 1980 21 2070 51 2160 81 2250 111 2340 141 2430 171 2520 201 2700 2610 2790 291 2880 321 3060 381 3150	rf329
16 1 22 52 82 112 142 172 202 232 262 292 352 352 382	Solution of the second state of the second st	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCGAGGATCTG E D V CATGGCGATC M A I GGTAGACATG V D M GTCCTTCGTG S F V CATGGCGCC	H F S E GAAATGACTC gcccacttct acgaggatgt agtgcacgcG P G L M ACGGCAGTCG T A V L GCGGCCATCG A A I A V L GCGGCCATCG F E D G ATTATCCCCG I I P V GTGCGTGATC V R D Q CCTGCGGATCA A T D T GGTGGCATCA G A R V GTGGGATCGG C A R V GTGGGATCGG C A R V GTGGGATCGG C A R V GTGGGATCGG C N I ACTGCCCGCGT I E M P ATGTCCCGCGT M S A S A S	A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGGG G G CCCAACCT P N L TGCCCACATC P N L TGCCCACATC P T S AGATCGAATGA V N D CGATGATGATGA V N D CGATGATGCA M M A <u>TAAAAGTCAC</u> K V T CGGATATCTT D I F TTGGTCACTT G H F CGAATGGCAA N G N CCTTTACCAA F T N CTATACCAA	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTGGAC E V D TATCACGACG I T T T GGCCAAGTGCAC G N E TCGGCCACGTTC C Q V L	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{GCQaccaca} \\ \mbox{GCQacCacA} \\ \mbox{G} & \mbox{D} & \mbox{S} & \mbox{K} \\ \mbox{TGGGGGCGGCGA} \\ \mbox{D} & \mbox{G} & \mbox{G} & \mbox{D} \\ \mbox{GAGGCCGGTGAA} \\ \mbox{AAGTCGAAGTTA} \\ \mbox{A} & \mbox{CCGATTGCGC} \\ \mbox{P} & \mbox{I} & \mbox{C} & \mbox{CGATTGCGC} \\ \mbox{P} & \mbox{I} & \mbox{C} & \mbox{CGATTGCGC} \\ \mbox{P} & \mbox{I} & \mbox{C} & \mbox{C} \\ \mbox{CGATTGCGCC} \\ \mbox{P} & \mbox{I} & \mbox{C} & \mbox{C} \\ \mbox{CGATTGCGCC} \\ \mbox{P} & \mbox{I} & \mbox{C} & \mbox{C} \\ \mbox{CGATTGCGCC} \\ \mbox{P} & \mbox{I} & \mbox{C} & \mbox{C} \\ \mbox{CGCATTGCGCC} \\ \mbox{P} & \mbox{C} & \mbox{C} & \mbox{C} \\ \mbox{CGCATTGCGCC} \\ \mbox{L} & \mbox{L} & \mbox{S} & \mbox{E} \\ \mbox{CGCGTTGCTGC} \\ \mbox{L} & \mbox{L} & \mbox{S} & \mbox{E} \\ \mbox{CGCGATCGAA} \\ \mbox{A} & \mbox{Q} & \mbox{L} & \mbox{L} & \mbox{S} & \mbox{E} \\ \mbox{CGCGTTGCGAACTACA} \\ \mbox{C} & \mbox{CGCATCGACAAA} \\ \mbox{L} & \mbox{L} & \mbox{S} & \mbox{E} \\ \mbox{CGCGAACCAA} \\ \mbox{A} & \mbox{Q} & \mbox{L} & \mbox{L} & \mbox{S} & \mbox{CGCAACCAA} \\ \mbox{A} & \mbox{Q} & \mbox{L} & \mbox{L} & \mbox{S} & \mbox{CGCAACCAA} \\ \mbox{A} & \mbox{Q} & \mbox{L} & \mbox{L} & \mbox{S} & \mbox{CGCAACTGA} \\ \mbox{A} & \mbox{A} & \mbox{Q} & \mbox{L} & \mbox{L} & \mbox{C} & \mbox{CGCAACTGA} \\ \mbox{A} & \mbox{A} & \mbox{A} & \mbox{C} $	Ctortector Ctortector CCTTGCCCGAG CCTTGCCCGAG CCTTGCCCGAG C A E GCCGCTGGCC P L A TGTCCGCTGG C Q S TGCGACCTCC A T L GGCGCAGACC C Q I TGCGCAGACC C V H TGACAACAAA D N K GGCCTAAGGA C Q A GGCCTGCAAGCC L Q A GGACCGCGATCC D V I TGCCCCCAAG A L K GGGCGCTAGG C R L ACTGTGGGACCTCC L W T CATCGCTCTCC	totgtaaatg toacacatgc gggtgtaga F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GATGTGGGCA A M D G CGCATCGAGC R I E H AACCACAAGT N H K W CTGAACCTCG L N L G CGTGCGGATG R A D A AACCCCCCGC	tgtogotgg agggoaaac goagoggoa AAGACTTG E L D TTGTCGGCT V G S ATATCTTT I F S ATTGGGACT A A S ACTCGTCA A A S ACCTCGTCA A A S ACCTCGTCA A A S ACCTCGTCA AAGGATCGC G S A GTTTCGAGG G S A GTTTCCAGG G S A GGACCAATA T N I GGAACGCGA N A T CGTATGACA Y D N	1 200 1710 1800 1800 1980 2107 51 2160 81 2250 111 2430 141 2520 201 2700 2610 2710 2610 2790 2910 2970 3016 3150 411 3240	rf329
16 1 22 52 82 112 142 172 202 232 262 292 352 382 412	Schröchten SAT CGATGCCGTC cggggctgta ccaatacgat aagatcattc ATATTGCCGA IAE CGCTGCACAT LHM CGACGACGA GAT GGCCGAGGC AEA CTCCCGGTCG CAC GACGAGGGCA CG CTGCCGGCG VDG CTGCCGGCG VDG CTGCCGGCG VDG CTGCCGGCTG CAS CAS CAS CAS CAS CAS CAS CAS CAS CAS	L R V AACCTGACGT otcaagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R GCGAGGATCTG CGAAGATGTG E D V CATGGCGATC M A I GGTAGACATG V D M GTCCTTCGTG S F V CATTCTGCCC I L P		A P A TGCGGGTGCT agagaactga ggacaattgc TGACCAAAGA T K D TGGCCTTGCG A L R TGATCGAAAC I E T CGGCGGCGG G TCCCAACCT P N L TGCCCACATC P T S AGATCGAAAGG I K G ATGTGAATGA V N D CGATGATGAC M M A TAAAAGTCAC K V T CGGATATCTT D I F TTGGTCACTT G H F CGAATGGCAA N G N CCTTTACCAA F T N ATGAAAAGGT E K V	T S M GTGGACATag aagatattgt tggcattcgt CTTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTGGAC E V D TATCACGACG I T T TGACAATGAA D N E TCGGCTCATGA R L I TCAGGTGCTG Q V L CGCACGTTTG A R L	$\begin{array}{c} \mbox{ccggcqtttg} \\ \mbox{ccggcqctta} \\ \mbox{ccgqcactaa} \\ \mbox{ccgqcactaa} \\ \mbox{ccqactaa} \\ \mbox{ccqactaa} \\ \mbox{ccqcactaa} \\ \mbox{GCQACAGCAA} \\ \mbox{G} D S K \\ \mbox{TrGGCGCGCGA} \\ \mbox{L} G A D \\ \mbox{TrGGCGCGGTGA} \\ \mbox{D} G G D \\ \mbox{GAGGCCACAA} \\ \mbox{E} A I K \\ \mbox{GAGGCCACAA} \\ \mbox{E} T T T \\ \mbox{AAGTCCAAGTT} \\ \mbox{K} S K F \\ \mbox{GCGATGTGCCA} \\ \mbox{CCGATTGCGC} \\ \mbox{P} I C A \\ \mbox{ACGGCCACAC} \\ \mbox{ACGGCCACAA} \\ \mbox{T} Q V M \\ \mbox{CCGATTGCGC} \\ \mbox{CCGATTGCTC} \\ \mbox{L} L S E \\ \mbox{GCGCACACCA} \\ \mbox{A} Q I E \\ \mbox{CACTTGACCG} \\ \mbox{H} L D R \\ \end{array}$	$\begin{array}{c} \mbox{ctgt} \mbox{ctgt}$	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GATGTGGGCA D V G K GCCATGGAGC A M D G CGCATCGAGC R I E H AACCACAAGT N H K W CTGAACCTCG L N L G CGTGCGGATG R A D A AAACTGACCC K L T P	tgtogotgg aggocaac goagoggca AAGACTTG E L D TTGTCGGCT V G S ATATCTTTT I F S ATTGGGACT T W D Y TTGGTGCCC G A R TGGCGGCGT A A S ACCTCGTCA A A S ACCTCGTCA A A S ACCTCGTCA C S A GTTTCAGG G S A GTTTCCAGG G S A GTTTCCAGG M R A GGACCAATA T N I GGAACGCGA N A T CGTATGACA Y D N CGCTTGACC L D P	1 00 1710 1800 1980 2070 51 2160 81 2250 111 2340 141 2430 171 2520 201 2700 2610 2790 2910 2970 3060 381 3150 411 3240	rf329
16 1 22 52 82 112 142 172 202 232 262 292 352 382 412 442	Solution of the second	L R V AACCTGACGT ctcagacac tacgottgta agagoctogg AACCGAAATG T E M GACGATCCAG T I Q CCACGCGGCG H A A GTCGTTCATG S F M GGGTGAAGAG G E E GTTCACGAAA F T K ATTGCCATTC L P F TATCCGCCGC I R R CGAAGATGTG E D V CATGGCGGTCC M A I GGTAGACATG S F V CATGCCCTCGTG S F V CATTCTGCCC G I L P GGCATATATC		A hardward and a second state of the second s	T S M GTGGACATag aagatattgt tggcattcgt CTCATCGTA F I V CGCGGAATAT A E Y GCTTGTTGCC L V A GACGCCTGTT T P V TATCCTTGAC I L D CGAAGAGGAA E E E TGTCAGCGAA V S E TTCTGTCACC S V T GGCCAAGGTT G K V CGAAGTGGAC I T T TGACAAGGAC I T T TCACGACGCG I T T TCACGACGCG C V L CCCACGTTG Q V L CCCACGCTTG A R L CTTCAAGCCG F K P	$\begin{array}{c} \text{ccggcgctttg}\\ \text{ccagcactaa}\\ \text{ccagcactaa}\\ \text{ccgcaccaca}\\ \text{cagcactaa}\\ \text{ccgcaccaca}\\ \text{K} D I A \\ \text{GCGACAGCAA}\\ \text{G} D S K \\ \text{TTGGCGTGCCGA}\\ \text{L} G A D \\ \text{TTGGCGTCCAA}\\ \text{F} A I K \\ \text{GAGGCCGCTCAA}\\ \text{Cagcaccaccac}\\ \text{Cagcacaccaccac}\\ \text{Caccaccaccac}\\ \text{Caccaccaccac}\\ \text{Caccaccaccacc}\\ \text{H} L D \\ \text{Cacaccaccaccacc}\\ \text{H} L D \\ \text{Cacaccaccaccaccaccaccacc}\\ \text{H} L D \\ Caccaccaccaccaccaccaccaccaccaccaccaccacc$	$\begin{array}{c} \mbox{ctgtgcgt}\\ \hline \mbox{ctgtggg}\\ \hline \mbox{ctgtggg}\\ \mbox{ctgtgggg}\\ \mbox{ctgcaac}\\ \mbox{ccgtggaac}\\ \mbox{ccgtggaac}\\ \mbox{ccgtgggc}\\ \mbox{cgctgggg}\\ cgtggggggggggggggggggggggggggggggggggg$	totgtaaatg gggtgtaga TTTGGCCGCA F G R K GGTGCGCGCA G A R I GCGTCCTGCA A S C N CTGGAAGAGC L E E H TATGTTCTGC Y V L L AAGAAGCGCA K K R M CGTCTCTATG R L Y D TACGGCTGCA Y G C K GATGTGGGCA A M D G CGCATCGACC R I E H AACCACAAGT N H K W CTGAACCTCG L N L G CGTGCGGGATG R A D A AAACTGACCC K L T P aaaaccggta	tgtogotgg aggocaac goagoggca AAGACTTG E L D TTGTCGGCT V G S ATATCTTT I F S ATTGGGACT T W D N AGGATCGCC G A R TGGCGGCG G S A GTTTCAGG G S A GTTTCCAGG G S A GTTTCCAGG F E V ATATCCGTG M R A GGACCATA T N I GGAACCGTA N A T CGTATGACA Y D N tgcgaccag	1 200 1710 1800 1980 2070 51 2160 81 2250 111 2430 141 2430 171 2520 201 2610 2790 291 2880 321 3060 381 3150 411 3240 441 3330	rf329

Fig. 1. Nucleotide and amino acid sequences of *orf5*, *orf329*, and *ahcY*.

Putative coding sequences are indicated in capital letters. Arrows indicate the direction of transcription of these genes. A putative NAD⁺ binding site in *ahcY* is boxed. Nucleotide sequences underlined indicate potential Shine-Dalgarno sequences of each open reading frame.

in *Ros. denitrificans* (Fig. 2A). Moreover, in the region upstream of *ahcY* were σ^{70} -type promoter sequence and photopigment-like palindrome sequence (TGTGT-N₆₋₁₀-ACACA) (Fig. 2B), instead of 26 bp-dyad sequence in the region upstream of *ahcY* in *Rba. capsulatus* (Buggy et al., 1994a).

As shown in Fig. 3, *prrA*, *prrB*, and *prrC* were constitutively transcribed under semiaerobic darkness and light, and aerobic light conditions as in *Rba. sphaeroides* (Mizoguchi et al., 1997), although the mRNA levels of *prrB* were very low. By using the *prrA*-and *prrC*-specific probes, a similar size of transcripts (1.1 and 1.2 kb) was detected. In the *prrCA*-specific probe, the 1.2 kb transcript was also detected (data not shown). Since the predicted transcripts of *prrA* and *prrC* are 0.6 kb and 0.5 kb, respectively, *prrA* and

1999

Α



Fig. 2. Comparison of the photosynthetic regulatory gene clusters of photosynthetic bacteria (A) and nucleotide sequence for the *ahcY* promoter region (B).

(A) Arrows indicate the location of the genes and the direction of the transcription. (B) Asterisks indicate sequences identical to E. coli o⁷⁰ consensus promoter elements. Opposite arrows indicate a putative photopigment-like palindrome sequence (TGTGT-N₆₋₁₀-ACACA). -120 and -60 indicate the positions from the ATG initiation codon, and -35 and -10 indicate the positions from a putative transcriptional initiation site.

prrC may be transcribed polycistronically.

orf5, orf329, and ahcY were also constitutively transcribed under all the conditions tested as in Rba. sphaeroides (Mizoguchi et al., 1997), although the mRNA level of orf5 was very low (Fig. 3). Two species of orf329 transcripts were detected, and the longer one (1.1 kb) was close to the predicted 1.0 kb of orf329. Two species of transcript of ahcY also were detected, and the longer one (1.5 kb) was close to the predicted 1.4 kb of *ahcY*. The shorter transcripts may be caused by exo- or endonuclease digestion of the 1.5-kb transcript.

In Rba. capsulatus, the light-responding expression of both orf5 and ahcY is thought to be regulated by the binding of HvrB to a 26-bp dyad symmetry region between orf5 and ahcY (Buggy et al., 1994a). By contrast, in the region upstream of ahcY in Ros. denitrifi*cans*, as described above, were σ^{70} -type promoter sequence and a photopigment-like palindrome sequence instead of the 26-bp dyad sequence in Rba. capsulatus (Buggy et al., 1994a) (Fig. 2B). Since no hvrB gene existed in the flanking region of the orf5 gene of Ros. denitrificans, it is possible that another transcriptional factor binds to this promoter region in Ros. denitrificans.

The presence of PrrA and PrrB and their constitutive expression (Figs. 2 and 3), the oxygen-insensitive transcription of the puf operon (Nishimura et al., 1996), the absence of a distinct oxygen-sensor domain in PrrBs (Matsumoto et al., 1997; this study), the putative DNA-binding activity of PrrA as an anaerobic activator for the *puf* and *puc* operons (Du et al., 1998), and the difference of the promoter region of the puf operon (Liebetantz et al., 1991; Zhu et al., 1986) all are clues to characterize and to elucidate the mechanism for the oxygen-insensitive expression of the puf operon in Ros. denitrificans. The cognate PrrBs/PrrAs/ PrrCs is common among these bacteria, whereas the primary oxygen-sensor (or redox-sensor) of Ros. denitrificans, if any, may have a quite different affinity to molecular oxygen or the redox state of the environment from those of the latter two bacteria. Alterna-



Fig. 3. Northern hybridization of *prrA*, *prrB*, *prrC*, *orf5*, *orf329*, and *ahcY* of *Ros. denitrificans* under various growth conditions.

Lane 1, aerobic (oxygen tension, 78% saturation) dark condition; Lane 2, semiaerobic (oxygen tension, 2% saturation) dark; Lane 3, semiaerobic $(2.5 \times 10^6 \text{ erg/cm}^2 \cdot \text{s})$ light. Twenty micrograms of total RNA for mRNAs of *prrA*, *prrC*, and *ahcY*, 30 μ g for *prrB*, *orf5*, and *orf329*, and 1 μ g for 16S rRNA were loaded in each lane. Arrows on the right side indicate the size of transcripts. Analyses of the hybridization were performed in triplicated independent experiments.

tively, direct evidence for or against involvement of PrrB/PrrA in the oxygen insensitivity may be found by the analysis of PrrA- or PrrB-mutant of *Ros. denitrificans*. For example, it is possible that PrrA of *Ros. denitrificans* is always bound to a promoter region of the *puf* operon, resulting in the constitutive expression of the *puf* operon at any oxygen tension, as in a PrrA-mutant of *Rba. capsulatus* (Du et al., 1998).

This work was supported financially by a Grant for Scientific Research (no. 09440265 to K.T.) from the Ministry of Education, Science, Sports and Culture, Japan.

References

- Bauer, C. E. (1995) Regulation of photosynthesis gene expression.
 In Anoxygenic Photosynthetic Bacteria, ed. by Blankenship, R.
 E., Madigan, M. T., and Bauer, C. E., Kluwer Academic Publishers, Dordreccht, The Netherlands, pp. 1121–1234.
- Buggy, J. J., Sganga, M. W., and Bauer, C. E. (1994a) Nucleotide sequence and characterization of the *Rhodobacter capsulatus hvrB* gene: HvrB is an activator of *S*-adenosyl-L-homocysteine hydrolase expression and is a member of the LysR family. *J. Bacteriol.*, **176**, 61–69.
- Buggy, J. J., Sganga, M. W., and Bauer, C. E. (1994b) Characterization of light-responding *trans*-activator responsible for differ-

entially controlling reaction center and light-harvesting-I gene expression in *Rhodobacter capsulatus. J. Bacteriol.*, **176**, 6936–6943.

- Drews, G. and Golecki, J. R. (1995) Structure, molecular organization, and biosynthesis of membranes of purple bacteria. *In* Anoxygenic Photosynthetic Bacteria, ed. by Blankenship, R. E., Madigan, M. T., and Bauer, C. E., Kluwer Academic Publishers, Dordreccht, The Netherlands, pp. 231–257.
- Du, S., Bird, T. M., and Bauer, C. E. (1998) DNA binding characteristics of RegA. J. Biol. Chem., 273, 18509–18513.
- Eraso, J. M. and Kaplan, S. (1994) prrA, a putative response regulator involved in oxygen regulation of photosynthesis gene expression in *Rhodobacter sphaeroides. J. Bacteriol.*, **176**, 32–43.
- Eraso, J. M. and Kaplan, S. (1995) Oxygen-insensitive synthesis of the photosynthetic membranes of *Rhodobacter sphaeroides*: A mutant histidine kinase. *J. Bacteriol.*, **177**, 2695–2706.
- Harashima, K., Hayasaki, J., Ikari, T., and Shiba, T. (1980) O₂-stimulated synthesis of bacteriochlorophyll and carotenoids in marine bacteria. *Plant Cell Physiol.*, **21**, 1283–1294.
- Iba, K., Takamiya, K., Toh, Y., and Nishimura, M. (1988) Roles of bacteriochlorophyll and carotenoid synthesis in formation of intracytoplasmic membrane systems and pigment-protein complexes in an aerobic photosynthetic bacterium, *Erythrobacter* sp. strain OCh114. *J. Bacteriol.*, **170**, 1843–1847.
- Liebetantz, R., Hornberger, U., and Drews, G. (1991) Organization of the genes coding for the reaction-centre L and M subunits and B870 antenna polypeptides α and β from the aerobic photosynthetic bacterium *Erythrobacter* species OCH114. *Mol. Microbiol.*, **5**, 1459–1468.
- Matsumoto, Y., Nagashima, K. V. P., Inoue, K., Bauer, C. E., Shimada, K., and Matsuura, K. (1997) Determination of nucleotide sequences of putative RegA and RegB; oxygen sensitive regulatory components related to the expression of photosystem in *Rhodovulum sulfidophilum* and *Roseobacter denitrificans. Plant Cell Physiol.*, **38**, s13.
- Mizoguchi, H., Masuda, T., Nishimura, K., Shimada, H., Ohta, H., Shioi, Y., and Takamiya, K. (1997) Nucleotide sequence and transcriptional analysis of the flanking region of the gene (*spb*) for the *trans*-acting factor that controls light-mediated expression of the *puf* operon in *Rhodobacter sphaeroides*. *Plant Cell Physiol.*, **38**, 558–567.
- Mosley, C. S., Suzuki, J. Y., and Bauer, C. E. (1994) Identification and molecular genetic characterization of a sensor kinase responsible for coordinately regulating light harvesting and reaction center gene expression in response to anaerobiosis. *J. Bacteriol.*, **176**, 7566–7573.
- Nishimura, K., Shimada, H., Hatanaka, S., Mizoguchi, H., Ohta, H., Masuda, T., and Takamiya, K. (1998) Growth, pigmentation, and expression of the *puf* and *puc* operons in a light-responding-repressor (SPB)-disrupted *Rhodobacter sphaeroides*. *Plant Cell Physiol.*, **39**, 411–417.
- Nishimura, K., Shimada, H., Ohta, H., Masuda, T., Shioi, Y., and Takamiya, K. (1996) Expression of the *puf* operon in an aerobic photosynthetic bacterium, *Roseobacter denitrificans. Plant Cell Physiol.*, **37**, 153–159.
- Pradel, E., Parker, C. T., and Schnaitman, C. A. (1992) Structures of the *rfaB*, *rfaJ*, *rfaJ*, and *rfaS* genes in *Escherichia coli* K-12 and their roles in assembly of the lipopolysaccharide core. *J. Bacteriol.*, **174**, 4736–4745.
- Sganga, M. W., Aksamit, R. R., Cantoni, G. L., and Bauer, C. E. (1992) Mutational and nucleotide sequence analysis of Sadenosyl-L-homocysteine hydrolase from *Rhodobacter capsulatus. Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6328–6332.
- Sganga, M. W. and Bauer, C. E. (1992) Regulatory factors controlling photosynthetic reaction center and light-harvesting gene expression in *Rhodobacter capsulatus. Cell*, **68**, 945–954.

- Shiba, T. and Simidu, U. (1982) *Erythrobacter longus* gen. nov., sp. nov., an aerobic bacterium which contains bacteriochlorophyll *a. Int. J. Syst. Bacteriol.*, **32**, 211–217.
- Shimada, H., Wada, T., Handa, H., Ohta, H., Mizoguchi, H., Nishimura, K., Masuda, T., Shioi, Y., and Takamiya, K. (1996) A transcription factor with a leucine-zipper motif involved in light-

dependent inhibition of expression of the *puf* operon in the photosynthetic bacterium *Rhodobacter sphaeroides*. *Plant Cell Physiol.*, **37**, 515–522.

Zhu, Y. S., Kiley, P. J., Donohue, T. J., and Kaplan, S. (1986) Origin of mRNA stoichiometry of the *puf* operon in *Rhodobacter sphaeroides. J. Biol. Chem.*, **261**, 10366–10374.