# Immuno-purification of a dimeric subcomplex of the respiratory NAHD-CoQ reductase of *Rhodobacter capsulatus* equivalent to the FP fraction of the mitochondrial complex I

Hervé Duborjal<sup>a</sup>, Alain Dupuis<sup>a,\*</sup>, Agnès Chapel<sup>b</sup>, Silvie Kieffer<sup>b</sup>, Joël Lunardi<sup>a</sup>, Jean-Paul Issartel<sup>a</sup>

<sup>a</sup>Laboratoire de Bioénergétique Cellulaire et Pathologique, EA 2019 UJF, DBMS, CEA Grenoble, 38054 Grenoble Cedex 9, France <sup>b</sup>Laboratoire de Chimie des Protéines, DBMS, CEA Grenoble, 38054 Grenoble Cedex 9, France

Received 13 February 1997

Abstract The *Rhodobacter capsulatus* genes encoding the NUOE and NUOF subunits, equivalent to the 24 kDa and 51 kDa subunits of the mammalian mitochondrial complex I, have been sequenced. According to the nucleotide sequence, the NUOE subunit is 389 amino acids long and has a molecular mass of 41.3 kDa. In comparison to the mitochondrial equivalent subunit, NUOE is extended at the C terminus by more than 150 amino acids. The NUOF subunit is 431 amino acids long and has a molecular mass of 47.1 kDa. A subcomplex containing both the NUOE and NUOF subunits was extracted by detergent treatment of *R. capsulatus* membranes and immuno-purified. This subcomplex is homologous to the mitochondrial FP fragment. Mass spectrometry after trypsin treatment of the NUOE subunit validates the atypical primary structure deduced from the sequence of the gene.

© 1997 Federation of European Biochemical Societies.

*Key words: Rhodobacter capsulatus*; Complex I; Mitochondrion; Gene sequence; Immunopurification; MALDI mass spectrometry

## 1. Introduction

NADH-CoQ reductase (or complex I) is the first component of the bacterial or mitochondrial respiratory chains. This enzyme catalyzes the oxidation of NADH and the reduction of ubiquinone with the concomitant ejection of 4 protons outside the bacterial cell or mitochondria [1,2]. In bovine mitochondria, structural analysis has revealed this complex to be composed of 43 different subunits. A number of cofactors (one FMN and 4-6 iron-sulfur clusters) are key components in the redox process. The precise assignment of these bound cofactors to the different subunits is still a matter of debate. Years ago, the subfractionation of the mitochondrial complex I led to the identification of three different subcomplexes called HP, IP and FP with the following features [3,4]. The HP subcomplex (hydrophobic proteins) is a non-soluble multimeric aggregate of hydrophobic subunits. It is partly built up of the 7 subunits encoded by the mitochondrial episome (also called ND subunits: ND1 through ND6 and ND4L). This hydrophobic fraction plays an important role in the anchoring

0014-5793/97/\$17.00 © 1997 Federation of European Biochemical Societies. All rights reserved. PII S 0 0 1 4 - 5 7 9 3 ( 9 7 ) 0 0 2 1 2 - 3

of complex I to the internal mitochondrial membrane and also in the proton channeling through the membrane. The second subcomplex, called IP fragment (iron-sulfur proteins), is a soluble domain made up of numerous subunits and contains some of the iron-sulfur clusters of complex I. The third subcomplex has been called FP fragment (flavoprotein) as it was proved to contain the FMN molecule and iron-sulfur clusters [5,6]. This soluble domain, when purified from bovine complex I, is composed of three subunits with molecular masses of 51, 24 and 9 kDa and displays NADH-ferricyanate oxidoreductase activity. The 51 kDa subunit is believed to be the catalytic subunit since it binds NADH [7,8]. One 4Fe-4S cluster and one FMN are probably liganded to this subunit. The second iron-sulfur cluster borne by the FP fragment is a 2Fe-2S cluster and is localized to the 24 kDa subunit. Complex I from bacteria, a more simple model of complex I, is being studied in different laboratories. In Escherichia coli and Paracoccus denitrificans, complex I is believed to be an assembly of at least 14 different subunits [9,10]. The genes encoding these subunits are clustered in one operon (so-called nuo for NAHD-ubiquinone oxidoreductase in E. coli and ngo for NADH-quinone oxidoreductase in P. denitrificans). Primary structure analysis has indicated that the 14 bacterial subunits are homologous to 14 subunits which are part of the bovine mitochondrial complex I. Seven bacterial subunits are equivalent to the mitochondrial ND subunits and seven others are the counterparts of the 75 kDa, 51 kDa, 49 kDa, 30 kDa, 24 kDa, 23 kDa and 20 kDa subunits. Hence, two of the bacterial subunits are highly similar to the two main subunits which compose the mitochondrial FP fragment (51 kDa and 24 kDa subunits). The third subunit of the mitochondrial FP fragment called 9 kDa has no counterpart amongst the subunits encoded by the bacterial operons. The present report shows that an FP-like subcomplex made up of two subunits homologous to the mitochondrial 51 kDa and 24 kDa subunits can be immuno-purified from the R. capsulatus membranes. Strikingly enough, the R. capsulatus NUOE subunit has a very long C terminus extension which is not found in the equivalent subunits of the different bacterial and mitochondrial complexes I.

#### 2. Materials and methods

An *R. capsulatus* B10 genomic DNA library constructed in lambda was screened with a bovine cDNA fragment coding for the 51 kDa subunit. Inserts from the positive clones were subcloned in plasmids and sequenced as described previously [11]. A 12 amino acid peptide called peptide PNE corresponding to the N terminus of the NUOE subunit was synthesized on an Applied Biosystems 432A peptide syn-

<sup>\*</sup>Corresponding author. Fax: (33) 4 76 88 51 87. E-mail: Dupuis@aup.ceng.cea.fr

*Abbreviations:* NUO, NADH-ubiquinone oxidoreductase; FP, flavoprotein fraction; MALDI, matrix-assisted laser desorption/ionization; PNE, synthetic peptide corresponding to the NUOE subunit Nterminus

thesizer (the PNE sequence was MLRRLHATQPDSC, in which the C terminus cysteine was an additional residue added for cross-linking to BSA). The HPLC purified peptide was cross-linked to maleimide-activated BSA and the final conjugate was used to raise antibodies in rabbits. Cytoplasmic membranes from aerobically grown wild type *R* capsulatus B10 strain were prepared as follows. Cells were grown in the dark in RCV medium [12] containing 30 mM lactate as a carbon source and harvested by centrifugation. They were resuspended in one tenth the initial volume in buffer A (50 mM HEPES, pH 7.5, 1 mM EDTA, 1 mM  $\beta$ -mercaptoethanol, 2 µg/ml aprotinin, 1 µg/ml lantipain, 1 µg/ml leupeptin, 1 µg/ml chymostatin, 0.5 µg/ml pepstatin) and disrupted by passing through a cell disrupter (Constant Cell, War-

wick, UK) at a pressure of 1 kbar. Subcomplex extraction was performed by treatment of the membranes with a stripping solution composed of 50 mM HEPES pH 7.5, 20 mM NaCl, 1 mM EDTA and 2% Glucopon (Glucopon 215 CS UP is a proprietary detergent from Henkel, France). After centrifugation (15 min, 90 000 rpm, 4°C) the supernatant was diluted five-fold in phosphate saline buffer and loaded onto a 1 ml immunoaffinity column of immobilized anti-PNE antibodies. The immunoaffinity column was generated essentially as described by Haines et al. [13]. Very briefly, the anti-PNE antibodies were immobilized on a protein A crosslinked Sepharose CL6B (Pharmacia) then chemically crosslinked with dimethylpimelimidate. The Glucopon solubilized R. capsulatus membrane extract was loaded

1	atogat	geeg	tcgc	caag	ggac	acca	actg	gcgg.	atgt	gtcc	gcca	tcate	cggci	acgat	tggal	tgtco	gtgtt	cddd	gaga	ittga	CCGC	tg 2	ATG I	CTG	95
96	CGC CGC	TTG	CAC	GCC	ACC	CAA	ccc	GAC	AGC	TTC	GCC	TTT	ACC	ccc	GCC	AAC	CGC	GCC	TGG	GCG	GAA	GCG	CAG	ATC	170
3	R R	L	H	A	T	Q	P	D	S	F	A	F	T	P	A	N	R	A	W	A	E	A	Q	I	27
28	T K	Y	P	E	G	R	Q	A	s	A	v	I	P	L	L	F	R	A	Q	E	Q	E	G	W	52
246	CTG ACG	AAG	CCC	GCG	ATC	GAA	TAC	GTG	GCC	GAC	CTG	CTC	GAC	ATG	CCC	TAT	ATC	CGC	GTG	CTC	GAA	GTG	GCG	TCC	320
321	TTC TAC	TTC	r ATG	TTC	CAG	ĊTG	CAG	čcc	ACG	GGC	TCG	GTC	GCG	CAT	ATT	CAG	GTC	TGC	GGC	ACC	ACC	AGC	TGC	ATG	395
78	F Y	F	M	F	Q	L	Q	P	T	G	S	V	A	H	I	0	V	C	G	T	T	S	C	M	102
103	I M	GGC	S	GAA E	AAC N	L	I	EAG	V	C	AAG K	R	K K	I	S	H	H	P	H	A	L	S	A	D	127
471	GGC AAG	TTC	AGC	TGG	GAA	GAG	GTC	GAA	TGT	CTG	GGC	GCC	TGT	GCC	AAT	GCG	CCG	ATG	GCC	CAG	ATC	GGC	AAG	GAT	545
128 546	G K TAC TTC	r GAA	GAC	W CTG	E ACC	E GAG	V GCG	EGGG	CTG	GAG	G CGG	A CTG	C ATC	A GAC	N GAT	A CTG	P GCG	M GCG	A GGC	Q AAG	GCG	G CCG	к G <b>T</b> G	CCG	152 620
153	Y F	Е	D	L	т	Е	А	G	L	Е	R	L	I	D	D	L	A	А	G	к	A	Ρ	V	Ρ	177
621 178	GGG TCG	GAA E	ATC I	GGG G	CGC R	TTC F	GGG G	GCG A	GAA E	CCG P	GCC A	ACC T	GGC G	CTG L	ACC T	AGC S	TGC C	ACG T	ACG T	ACC T	TCC S	GGC	GCG A	CGC R	695 202
696	GAG GCG	CAG	AAC	GGC	TCG	GTC	GCG	CGG	GCG	GTT	GCG	GTG	GGC	GAT	GGC	ACG	CGC	CGG	ATC	GAC	GGG	ACC	GAG	GTG	770
203	E A CGG CTT	Q TCC	N GCC	G CCA	S TGG	V CAG	A GTC	R TCG	A GGG	CAG	A ATG	V GGC	G GCG	D ATC	G GGT	T CTG	R ATG	R ACG	I GCG	D CGG	G CCG	T CGC	E TCG	V GCG	227 845
228	R L	S	A	Р	W	Q	V	S	G	Q	М	G	A	I	G	L	М	т	A	R	Ρ	R	S	A	252
846 253	GTG CAG V O	CAG	CAG	GCG A	GCG A	GCC A	GAG E	GCG A	GCG A	AAA K	GCG A	GCG A	GCG A	GCG A	GCA A	GAG E	TCG S	GCC A	GGG G	GGG G	ACG T	TCA S	ACC T	CCG P	920 277
921	GCT GCA	ĉcg	GTG	GCC	GGG	GCG	GCG	GGG	AAA	GCC	GCA	CAA	ACG	GCT	GAA	ACC	GCT	GCG	CAG	AAA	CCG	CTT	GCG	CTG	995
278	A A GTT GCA	P	V	A GAC	GGC	A AAG	A GCC	GAT	K GAG	A CTG	A CAG	Q AAG	T ATC	A GAG	E AGC	T ATC	A GGA	A CCG	Q GCG	K CTC	PGAG	L GCG	A CTT	L CTG	302
303	V A	P	R	D	G	к	A	D	Е	L	Q	К	I	E	S	I	G	Р	A	L	E	A	L	L	327
1071 328	CAC GAT	CTG L	GGC	ATC T	TTC	CAT H	TTC	GAC D	CAG	ATC T	GCC A	GCC A	TGG W	GGC	GCG A	GGC	GAA E	ATT	GCC A	TGG W	ATG M	GAC D	GGC	AAT N	1145 352
1146	CTG AAG	GGC	TTT	ĈĊĠ	GGG	CGC	GCC	AGT	ĞGC	GAC	CGC	TGG	GTG	GTT	CAG	GCG	AAG	ÂTC	ATC	GTC	GAG	GAA	GGG	CTT	1220
353	L K GAC CGT	G TTC	F	P GAA	G	R	A	S	G	D TAA '	R TAC	W TGB	Varia	V	Q atac	A	K	I agac	I ttma	V acrai	E	E	G adadd	reeda	377
378	D R	F	L	E	R	A	К	т	N	N	Y	*	. ugt	.0990	ucu.	. egue	oega	agae	cegu	acga	cauc	cuuci	19499	10099	390
1308	gacggcg	cgacg	tccg	ccac	ccgg	cacg	gaaa	gaggg	jaaaq	gaaa	aggea	aa A	ATG 1 M	CG 2	ACG 1	AAC (	GAC J	ACA	GAC	TGC	CAG	ACG T	AAA K	GCC A	1394
1395	CGC AAG	CTG	GGT	GGT	TGG	CTC	GGA	GCC	GCG	ATT	TTC	TTT	GCA	CTT	TGG	CTT	CTG	GGT	TCG	GTG	ĞGG	ATC	TGG	ATG	1469
13 1470	R K GCG GCG	L ATC	G GCC	G GGG	W GTG	L ATC	G TCT	A GCG	A CTG	I ATC	F CTG	F GCA	A CGG	L CTG	W ATC	L TGC	L TGG	G GĆG	S ACC	V TGC	G GGT	I CTG	W CCC	M GAG	37 1544
38	A A	I	A	G	v	I	S	A	L	I	L	A	R	L	I	С	W	A	т	С	G	L	P	Е	62
1545 63	CCG AAA P K	GCG A	CCG P	ATG M	TCG S	GCC A	GAG E	GAA E	GGC G	GCC A	GCG A	CTG L	CGT R	CGG R	GCC A	GTT V	CCG P	GCC A	GCG A	GCG A	CCG P	ACG T	GCG A	GCG A	1619 87
1620	GTG AAA	GTG	GCC	GAG	GCG	CCG	AAA	CCG	GCG	CCG	AAA	GCC	GCC	GAG	CCC	GTC	AAG	GCC	GCG	ccc	GTG	GCG	GCG	CCG	1694
1620 88 1695	GTG AAA V K GTC AAG	GTG V ACG	GCC A CCC	GAG E GAG	GCG A CCG	CCG P GTG	AAA K GCC	CCG P GCG	GCG A GTG	CCG P GCT	AAA K GAA	GCC A GCG	GCC A CCG	GAG E AAA	CCG P CCG	GTC V GCG	AAG K GCG	GCC A AAG	GCG A CCC	CCC P GCC	GTG V CCG	GCG A GCG	GCG A CCC	CCG P GCG	1694 112 1769
1620 88 1695 113	GTG AAA V K GTC AAG V K	GTG V ACG T	GCC A CCC P	GAG E GAG E	GCG A CCG P	CCG P GTG V	AAA K GCC A	CCG P GCG A	GCG A GTG V	CCG P GCT A	AAA K GAA E	GCC A GCG A	GCC A CCG P	GAG E AAA K	CCG P CCG P	GTC V GCG A	AAG K GCG A	GCC A AAG K	GCG A CCC P	CCC P GCC A	GTG V CCG P	GCG A GCG A	GCG A CCC P	CCG P GCG A	1694 112 1769 137
1620 88 1695 113 1770 138	GTG AAA V K GTC AAG V K CCG AAA P K	GTG V ACG T GGC G	GCC A CCC P AAG K	GAG E GAG E TCG S	GCG A CCG P AAG K	CCG P GTG V CCC P	AAA K GCC A CGT R	CCG P GCG A CGC R	GCG A GTG V GGC G	CCG P GCT A C TG2 *	AAA K GAA E A ga	GCC A GCG A cagao	GCC A CCG P CCCG	GAG E AAA K JCCCG	idede b CCC b CCC	GTC V GCG A ctgg	AAG K GCG A ggag	GCC A AAG K aggc	GCG A CCC P gctg	CCC P GCC A acgc	GTG V CCG P ygca	GCG A GCG A yaaa	GCG A CCC P ccato	CCG P GCG A coga	1694 112 1769 137 1858 148
1620 88 1695 113 1770 138 1859	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc	GTG V ACG T GGC G gggggc	GCC A CCC P AAG K tcgg	GAG E GAG E TCG S gatg	GCG A CCG P AAG K	CCG P GTG V CCC P cgga	AAA K GCC A CGT R caga	CCG P GCG A CGC R Atccg	GCG A GTG V GGC G Igaag	CCG P GCT A C TG2 (C TG2	AAA K GAA E A ga	GCC A GCG A cagao	GCC A CCG P CCG CCG	GAG E AAA K CCCCg	idede b CCC b CCC b CCC	GTC V GCG A ctgg	AAG K GCG A ggag cgtg	GCC A AAG K aggc	GCG A CCC P gctg	CCC P GCC A acgci	GTG V CCG P ggca	GCG A GCG A yaaad	GCG A CCC P ccato	CCG P GCG A cccga	1694 112 1769 137 1858 148 1958
1620 88 1695 113 1770 138 1859 1959 1	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacac	GTG V ACG T GGC ggggcc	GCC A CCC P AAG K tcgg cccg	GAG E GAG E TCG S gatg ccga	GCG A CCG P AAG K tccc aacc	CCG P GTG V CCC P cggao gagg	AAA K GCC A CGT R caga gaat	CCG P GCG A CGC R Atccg ggac/ DF	GCG A GTG V GGC G Igaag ATG M	CCG P GCT A TGI C TGI t CTC .	AAA K GAA E A ga Cggg AAC N	GCC A GCG A cagao ctgcg GAT D	GCC A CCG P CCCG CCC CCC CCC CAC	GAG E AAA K gcccg gacg GAT D	CCG P Igcgc raggg CGC R	GTC V GCG A setgg pagac ATC I	AAG K GCG A ggag cgtg TTC F	GCC A AAG K aggc tgta ACC T	GCG A CCC P gctg tcgg AAC N	CCC P GCC A acgci acag ATC I	GTG V CCG P ggca ccga TAC Y	GCG A GCG A gaaa tgcc GGG G	GCG A CCC P ccato gcgc ATG M	CCG P GCG A cccga gggc CAC H	1694 112 1769 137 1858 148 1958 2042 16
1620 88 1695 113 1770 138 1859 1959 1 2043	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacacc GAC CGC	GTG V ACG T GGC G ggggcc tggcc TCG	GCC A CCC P AAG K tcgg cccg CTG	GAG E GAG E TCG S ggatg ccga AAG	GCG A CCG P AAG K tccc aacc GGC	CCG P GTG V CCC P CCC gaga gagg GCG	AAA K GCC A CGT R CGT R CGT R CGT R CGT R CGT R CGT R CGT R CGT R CGT R CGT R CGT R CGT R C CGT R C CGT R C C C C C C C C C C C C C C C C C C	CCG P GCG A CGC R Atccg ggac/ DF GCG	GCG A GTG V GGC G G G G G ATG CGT	CCG P GCT A TGJ CTGJ CTC L GGT	AAA K GAA E A ga Agggo AAC N CAC	GCC A GCG A cagao ctgcg GAT D TGG	GCC A CCG P Secco Igoot CAC H GAT	GAG E AAA K gcccg GAT D GGC	CCG P CCG P ggcgc gaggg CGC R ACG	GTC V GCG A setgg Magac ATC I GCG	AAG K GCG A ggag cgtg TTC F GCG	GCC A AAG K aggc tgta ACC T ATC	GCG A CCC P gctg ACC AAC N CTT	CCC P GCC A acgc: acag ATC I GGC	GTG V CCG P ggca ccga TAC Y AAT	GCG A GCG A yaaad GGG GGG GGC	GCG A CCC P ccato gege ATG M CGC	CCG P GCG A ccoga gggc CAC H GAC	1694 112 1769 137 1858 148 1958 2042 16 2117
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacac GAC CGC D R TGG ATC	GTG V ACG T GGC gggggc cggcc TCG S ATC	GCC A CCC P AAG K tcgg cccg CTG L GAG	GAG E GAG E TCG S gatg ccga AAG K CAG	GCG A CCG P AAG K tccc aacc GGC GTC	CCG P GTG V CCC P Cggao gagg GCG A AAG	AAA K GCC A CGT R CCGT R CCGT R CCGT A C CGT A C CGT A C C C C C C C C C C C C C C C C C C	CCG P GCG A CGC R Atccg ggach <b>DF</b> GCG A TCG	GCG A GTG V GGG G G G ATG CGT R GGG	CCG P GCT A TGJ CTG CTC GGT GGT CTG	AAA K GAA E A ga AAC N CAC H CGC	GCC A GCG A cagad ctgcg GAT D TGG W GGC	GCC A CCG P CCCG GCAC H GAT D CGC	GAG E AAA K gcccg GAT D GGC GGC GGC	CCG P CCG P ggcgc cGC R ACG T GGT	GTC V GCG A sctgg Aggac ATC I GCG A GCG	AAG K GCG A cggag TTC F GCG A GGC	GCC A AAG K aggc tgta ACC T ATC I TTT	GCG A CCC P gctg AAC N CTT L CCG	CCC P GCC A acgc acag ATC I GGC G ACC	GTG V CCG P ggca tac Y AAT N GGG	GCG A GCG A yaaad GGG GGG GGC G CTC	GCG A CCC P ccato ATG M CGC R AAG	CCG P GCG A cccga CAC H GAC D TGG	1694 112 1769 137 1858 1958 2042 16 2117 41 2192
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2103	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacacc GAC CGC D R TGG ATC W I	GTG V ACG T GGC G GGC CGGC CGGC S ATC I	GCC A CCC P AAG K tcgg cccg CTG GAG E	GAG E GAG E TCG S gatg ccga AAG K CAG Q	GCG A CCG P AAG K tccc aacc GGC GGC GTC V	CCG P GTG V CCC gagg GCG A AAG K	AAA K GCC A CGT R CGT R CAGA ATG ATG A CCC A	CCG P GCG A CCGC R Atccg ggac/ DF GCG A TCG S	GCG A GTG V GGC G G G G G G G G G G G G G G G G	CCG P GCT A CTG2 CTC GGT GGT CTG L CTG	AAA K GAA E A ga AAC AAC N CAC H CAC R CAC	GCC A GCG A cagao ctgcg GAT D TGG W GGC G	GCC A CCG P Igoot CAC H GAT D CGC R	GAG E AAA K GCCCG GGAT D GGC G GGC G GGC G GGC	CCG P CCG P Igcgc CGC R ACG T GGT GGT	GTC V GCG A ctgg A TC I GCG A GCG A	AAG K GCG A ggag TTC F GCG A GGC G	GCC A AAG K aggc tgta ACC T ATC I TTT F	GCG A CCC P gctg AAC N CTT L CCG P	CCC P GCC A acgc: acag ATC I GGC G ACC T	GTG V CCG P gca tac TAC Y AAT N GGG G	GCG A GCG A yaaaa GGG G G G C T C L C T C	GCG A CCC P ccato ato ATG M CGC R AAG K	CCG P GCG A cccga gggc CAC H GAC D TGG W	1694 112 1769 137 1858 1958 2042 16 2117 41 2192 66
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacacc GAC CGC D R TGG ATC W I TCC TTC S F	GTG V ACG T GGC G G G G G G G G G G G G C G G C G C G C G C C C G C C C G C C C G C C C G C C C G C	GCC A CCC P AAG K tcgg cccg CTG L GAG E CCG P	GAG E GAG E TCG S gatg ccga AAG K CAG Q AAG K	GCG A CCG P AAG K Stccc aacc GGC GGC GTC V CAA O	CCG P GTG V CCC P Gggag GCG A AAG K TCC S	AAA K GCC A CGT R Caga gaat M GCC A GCC A GAC D	CCG P GCG A CGC R Atccg ggac/ DF GCG A TCG S GGG GGG G	GCG A GTG V GGC G G G G G G G G G G CGC R	CCG P GCT A C TG2 CTG GGT GGT CTG L CCC P	AAA K GAA E A ga AAC AAC N CAC H CAC R CAC S	GCC A GCG A cagao GAT D TGG W GGC G TTT F	GCC A CCG P CCCG CAC H GAT D CGC R CTG L	GAG E AAA K gcccg GAT D GGC G GGC G GGC G GTG V	CCG P CCG P ggogo CGC R ACG T GGT G ATC I	GTC V GCG A ctgg agac ATC I GCG A GCG A AAC N	AAG K GCG A gggag TTC F GCG A GGC G GCC A	GCC A AAG K aggc tgta ACC T ATC I TTT F GAT D	GCG A CCC P gctg AAC N CTT L CCG P GAA E	CCC P GCC A acago ATC I GGC G ACC T TCC S	GTG V CCG P ggca TAC Y AAT N GGG G GAG E	GCG A GCG A gaaaa GGG GGG GGC G CTC L CCG P	GCG A CCC P ccato ATG M CGC R AAG K GGC G	CCG P GCG A cccga CAC H GAC D TGG W ACC T	1694 112 1769 137 1858 148 1958 2042 16 2117 41 2192 66 2267 91
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacacc GAC CGC D R TGG ATC W I TCC TTC S F TGC AAG	GTG V ACG T GGC G G G G G G G G G C G G C G G C G C G C G C G C G C C G C C G C C G C C G C C G C C C G C C C G C C C G C C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C C G C C C G C C C G C C C G C C C G C	GCC A CCC P AAG K tcgg cccg CTG L GAG E CCG P CCG CCG CCG CCG	GAG E GAG E TCG S gatg ccga AAG K CAG Q AAG K GAA	GCG A CCG P AAG K Jtccc aacc GGC GGC GGC V CAA Q ATC	CCG P GTG V CCC P gggac gagg GCG A AAG K TCC S ATG	AAA K GCC A CGT R Caga gaat M GCC A GCC A GAC D CGC	CCG P GCG A CGC R Atccg ggac2 <b>DF</b> GCG A TCG S GGG G G GCG CAC	GCG A GTG V GGC G G G CGT R GGG G CGC R GAT	CCG P GCT A TGJ CTG CTG CTG CTG CCC P CCC	AAA K GAA E A ga AAC N CAC H CGC R TCG S CAT	GCC A GCG A cagao tgcg GAT TGG W GGC G TTT F ACG	GCC A CCG P CCCG CAC H GAT D CGC R CTG L CTG	GAG E AAA K gcccg GAT D GGC G GGC G GGC G GTG V ATC	CCG P CCG P ggcgc CGC R ACG T GGT G ATC I GAA	GTC V GCG A setgg ATC I GCG A GCG A AAC N GGC	AAG K GCG A ggag TTC F GCG A GGC G GCC A GCC	GCC A AAG K aggc T ACC T ATC I TTT F GAT D CTG	GCG A CCC P gctg AAC N CTT L CCG P GAA E ATC	CCC P GCC A acgc ATC I GGC G ACC T TCC S GCC	GTG V CCG P ggca TAC Y AAT N GGG GAG E GAG	GCG A GCG A gaaao GGG GGC G GGC G CTC L CCG P TTC	GCG A CCC P ccato ATG M CGC R AAG K GGC GCC	CCG P GCG A ccoga gggc CAC H GAC D TGG W ACC T ATG	1694 112 1769 137 1858 1958 2042 16 2117 41 2192 66 2267 91 2342
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacacc GAC CGC D R TGG ATC W I TGC ATC S F TGC AAG C K GGC GCG	GTG V ACG T GGC 33339C TCG S ATC I ATC M GAC D CGG	GCC A CCC P AAG K tcgg cccg CTG GAG E CCG P CCG R GCG	GAG E GAG E TCG S gatg ccga AAG K CAG Q AAG K GAA E GCC	GCG A CCG P AAG K Steed aacc GGC GGC V CAA Q ATC I TAT	CCG P GTG V CCC P GGG A AAG K TCC S ATG ATC	AAA K GCC A CCGT R CCGG ATG M GCC A GCC A CGC R CGC R TAT	CCG P GCG A CCGC R tcccg ggac/ DF GCG A TCG S GCG G G CAC H ATT	GCG A GTG V GGC G GGC ATG GGC CGC R GGG CGC R GAT D CGC	CCG P GCT A TGJ CTG CTG CTG CTG CTG CCC P CCC P CCG P GGC	AAA K GAA E A ga AAC N CAC H CGC R CAC R CGC R CGC S CAT H GAA	GCC A GCG A cagad tgcg GAT TGG W GGC G TTT F ACG T TAT	GCC A CCG P GCCCG GCC CAC H GAT D CGC R CTG L CTG L CTG GTC	GAG E AAA K gcccg GAT GGC GGC GGC GGC GTG V ATC I CGC	CCG P CCG P ggogo CGC R ACG T GGT GGT GAA E GAA	GTC V GCG A ctgg AGC ATC I GCG A GCG A AAC N GCC G AAG	AAG K GCG A ggag TTC F GCG A GGC G GCC A GAG	GCC A AAG K aggc tgta ACC T ACC T TTT F GAT D CTG L GCC	GCG A CCC P gctg AAC N CTT L CCG P GAA E ATC I CTG	CCC P GCC A acgc A T GGC G G C C C C C C C C C C C C C C C	GTG V CCG P ggca TAC Y AAT N GGG GAG GAG GGT GCC	GCG A GCG A gaaao GGG GGC GGC GGC CTC L CCG P TTC F GCC	GCG A CCC P ccato ATG ATG CGC R AAG K GGC GCC ATC	CCG P GCG A cccga gggc CAC H GAC D TGG W ACC T ATG M GAC	1694 112 1769 137 1858 148 2042 16 2117 41 2192 66 2267 91 2342 116 2417
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2218 42 2193 67 2268 92 2343 117	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacac GAC CGC D R TGG ATC W I TCC TTC S F TGC AAG C K GGC GCG G A	GTG V ACG GGC G 33339C TCG S ATC I ATC GAC CGG R	GCC A CCC P AAG K CCG CCG CCG CCG CCG R GCG A	GAG E GAG E TCG S GAG CCGA AAG K CAG Q AAG K GAA E GCC A	GCG A CCG P AAG K tccc G GC G GC G GC CAA Q ATC I TAT Y QC	CCG P GTG V CCC P gggag GCG A AAG K TCC S ATG M ATC I	AAA K GCC A CGT R CAGT ATG ATG ATG ATG ATG CGC A CGC R TAT Y	CCG P GCG A CCC R ggac/ DF GCG A TCG S GCG A TCG S GCG A TCG S GCG A TCG S GCG A TCG S GCG A TCG S GCG A TCG C C C C C C C C C C C C C C C C C C	GCG A GTG V GGC GGC ATG CGT R GGG CGC R GGT D CGC R CGC R CGC	CCG PGCT A CTGI CTGI GGT GGT CCC PCCG GGC GGC GGC	AAA K GAA E A ga Cac AAC N CAC H CAC R TCG S CAT H GAA E CAC	GCC A GCG A cagao tgcg GAT D TGG W GGC G TTT F ACG T ACG T AT Y	GCC A CCG P CCCC CAC H GAT D CAC R CAC CAC CAC CAC CAC CAC CAC CAC C	GAG E AAA K GCCCG GGC GGC GGC GGC GGC GTG V ATC I CGC R	CCG P CCG P ggcgc CGC R ACG T GGT GGT GAA E GAA E	GTC V GCG A sctgg Aggac ATC I GCG A GCG A AAC N GGC G AAG K	AAG K GCG A ggag TTC F GCG A GCC A GCC A GCC A GAG E	GCC A AAG K dggc tgta ACC T ATC I TTT F GAT D CTG L GCC A	GCG A CCC P gctg AAC N CTT L CCG P GAA E ATC I CTG L	CCC P GCC A acgc ACC G GCC G ACC T TCC S GCC A CAG Q C A	GTG V CCG P ggca TAC Y AAT N GGG GAG E GGT GCC A	GCG A GCG A gaaaa GGG GGC GGC GGC CTC L CCG P TTC F GCC A	GCG A CCC P Coato ATG ATG CGC R AAG K GGC GCC A ATC I	CCG P GCG A CCCG CAC H GAC D TGG W ACC T GAC D ACC T GAC D ACC	1694 112 1769 137 1858 148 2042 16 2117 41 2192 66 2267 91 2342 116 2417 141
1620 88 1695 113 1770 138 1859 1959 2043 17 2118 42 2193 67 2268 92 2343 117 2418 42 2193 67 2243 117 2418 42 2193 67 2443 117 2418 42 2193 67 2443 117 2444 2193 117 2444 2193 2443 2443 2444 2444 2444 2444 2444 24	GTG AAA V K GTC AAG V K CCG AAA P K gccctgc cggacac CG AC CGC D R TGG AC CG V I TGG AG C K GC CG GC GCG G A GAA TGC E C	GTG V ACG GGC G GGG G GGG TCG S ATC I ATC D CGG R CGG R TAC Y	GCC A CCC P AAG K CCG CCG GAG E CCG CCG P CCG R GCG A GAG E	GAG E GAG E TCG S CCGa CCGa CAG CAG CAG CAG CAG CAG CAA E GCC A	GCG A CCG P AAG K CCG G C G G C CAA Q ATC I TAT Y GGG G	CCG P GTG V CCC P ggagg GCG A AAG K TCC S ATG M ATC I CTG L	AAA K GCC A CGT R CGT R GAC ATG M GAC D CGC R TAT Y ATC I	CCG P GCG A CCC R A CCC R GCG A TCG S GCG G GCG H TCG S GCG H TCG S GCG G G G G G G G G G G G G G G G G	GCG A GTG V GGC GGC M CGT R GGG CGC R GGG CGC R CGC R AAG K	CCG P GCT A C TGJ CTG CTG CTG CTG CTG CCG P CCG P GG AAC N	AAA K GAA E A ga A CAC N CAC H CGC R TCG CAT H GAA E GCC A	GCC A GCG A Cagao GAT D TGG W GGC G TTT F ACG TAT Y C C	GCC A CCG P GCCCG CAC H GAT D CGC R CTG L CTG L GTC V CGC R	GAG E AAA K GCCCG GAT D GGC GGC GGC GGC GGC GTG V ATC I CGC R TCG S	CCG P CCG P ggcgc CGC R ACG T GGT GGT GAA E GAA E GAA E GAA D	GTC V GCG A setgg ATC I GCG A GCG A AAC N GGC G AAG K TAT Y	AAG K GCG A gggag TTC F GCG A GCC A GCC A GCC A GCC A GAG E GAT D	GCC A AAG K aggc tgta ACC T ATC I TTT F GAT D CTG L GCC A TTC F	GCG A CCC P gctg AAC N CTT L CCG P GAA E ATC I GAC D	CCC P GCC A acage ATC I GGC G GCC T T CC S GCC A CAG Q GTC V	GTG V CCG P ggca TAC Y AAT N GGG GAG GAG GGT GCC A TAT Y	GCG A GCG A gaaaa GGG GGG GGC G CTC L CCG P TTC F GCC A CTG L	GCG A CCC P ccato ATG ATG CGC R AAG K GGC A ACC A CCC A CGC A CGC A CGC A CCC R AAG K GGC A CCC P P CCATO A CCC P P CCATO A CCC P P CCATO A CCC P P CCATO A CCC A CCC P P CCATO A CCC A CCC P P CCATO A CCC A CCC A CCC P P CCATO A CCC A C CCC A CCC A C CCC A C CCC A C CCC A C CCC A CCCC A CCCC A CCC A CCC A CCC A CCC A CCCC A CCC A CCC A CCC A CCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCC A CCCCCC	CCG P GCG A CCCG A CCC CAC H GAC D TGG W ACC T GAC D CAC H	1694 112 1769 137 1858 1958 2042 16 2117 41 2192 66 2267 91 2342 116 2417 141 2417 141 2492 166
1620 88 1695 113 1770 138 1859 1959 2043 17 2118 42 2193 67 2268 92 2343 117 2418 42 2343 117 2418 42 2343	GTG AAA V K GTC AAG V CCG AAA P K GCCCGCC GAC CGC D R TGG ATC TGG ATC TGG ATC S F TGC AAG C K GCC GCG G A GAA TGC E C GCT GCG	GTG V ACG T GGC GG2ggGC TCG S ATC I ATC GAC D CGG R TAC Y GGC	GCC A A CCC P AAG K trogg ccccg CTG L GAG E CCG P CGC R GCG A GAG E GCC	GAG E GAG E TCG S S CCGa CCGa CCG A AAG K GAA E GCC A TTAT	GCG A CCG P P AAG K GGC G GGC G V CAA Q ATC I TAT Y GGG G G ATC	CCG P GTG V CCCC P CGG GAG A AAG K TCC S ATG M ATC I CTG C TGC	AAA K GCC A CGT R Cagaate ATG M GCC ATG M GCC A CGC R TAT Y ATC I GGC	CCG P GCG A CGC R Atccg ggac/ <b>F</b> GCG A TCG S GGG GGG CAC H ATT I GGC GGAG	GCG A GTG V GGC G G G G G G G G G G G G G G G C G C G C G C G C G C G C G C G C G C G C G C G C G C G C G	CCG P GCT A TGJ CTC CTC CTG GGT CCG P CCG G CCG G GGC AAC N ACC	AAA K GAA E Sgggd AAC N CAC H CGC R CAC H CGC S CAT H GAA E GCC A GCC	GCC A GCG A Cagao Cagao D TGG W GGC G TTT F ACG T ACG T ACG T ACG T C TTT C C C TT	GCC A CCG P CAC H GAT D CGC R CTG L CTG L CTG CTG CTG CTG	GAG E AAA K GCCCG GAT D GGC GGC GGC GGC GGC GTG V ATC I CGC R TCG S GAA	CCG P CCG P gcgc CGC R ACG T GGT GAA E GAA E GAA E GAA E GAA	GTC V GCG A sctgg GCG A GCG A GCG A A A C GCG A A C T T T T T C T T	AAG K GCG A ggag TTC F GCG A GGC GCC A GCC A GCG A GCG A GAG E GAT D GAG	GCC A AAG K aggc tgta ACC T T TTT F ATC I TTT GAT D CTG L GCC A TTC F GGC	GCG A CCC P gctg AAC N CTT L CCG P GAA E ATC I CTG L GAC D AAG	CCC P GCC A acadge acadge ATC I GGC G GGC G CAC GGC C ACC A CAG GGC V AAG	GTG V CCG P ggca TAC Y AAT AAT N GGG GGG GGG GGG GGC A TAT Y GGG	GCG A GCG A gaaao GGG G GGC G GCTC L CCG F TTC F GCC A CTG L ATG	GCG A CCC P coato GCGC R AAG K GGC G GCC A ATC I CAC H CCG	CCG P GCG A Cccga gggc CAC H GAC D TGG M ACC T ATG GAC D CAC H CCG G	1694 112 1769 137 1858 1958 2042 16 2117 41 2192 66 2267 91 2342 116 2417 141 2492 166 2567 246 2417
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 2343 117 2448 2493 167 2493	GTG AAA V K GTC AAG V K CCG AAA P K gcctgcc GAC CGC D R TGG ATC TGG ATC GAC CGC W I TGC TTC TGC TTC TGC TCC TGC C G A GAA TGC E C GGT GCG G A ATG AAA	GTG V ACG T GGC GG S S ATC I ATC C GAC C GGC GGC C C GGC C C G C C G GC C C G G C C G	GCC A A CCCC P P AAG K trogg ccccg CTG L GAG E CCG C CCG C CCG C CCG A GAG E CCG A GAG C CCG C CCG C CCC C CCC C CCC C CCC C C C C	GAG E GAG E TCG S Gatç Ccga K CAG Q AAG K GAA E GAA GCC A TAT Y TTTC	GCG A AAG K tcccc GCC G GCC G GCC V CAA Q ATC I TAT TAT Y GGG G GCC Q CAA Q CCA Q CAA Q CCA Q CCA GCC G C C G C C G C C G C C G C C C C	CCG P GTG CCC PCCCC P CCCC P CCCC P CCCC C GCG A AAG AAG AAG AAG CCTG L CCTG L CC CCG GCG	AAA K GCC CGC CGC R R CCGG2 ATG M GCC A ATG M GCC A CGC C CGC R TAT Y ATC I GCC G G G G G G G G G G G G G G CC	CCG P GCG A CCGC R R Atccgc gggac2 DF GCG A TCG GCG GCG GCG GCAC H T TCG GCG GCAC H T T TCG GCG GCG CAC T TCG TCG T TCG T TCG C CAC T CCGC C C C	GCG A GTG V GGC G GGC G G CGT R GGG G G CGC R AAG GAA E GGG G GGG G GGG G GGG G GGG G GGG	CCG P GCT TG2 CTG2 CTC CTC CCC CCG CCG G GCT CCC CCC CCC C	AAA K GAA E A GGGGGA R CAC R CAC R CAC R CAC R CAC A GCG A GCC A GCC A TAT	GCC A GCG GCG GAT D D TGG GAT T TGG GGC C T T T T T T T C C C C T T L C GGC C GGC G C C S C S C S S C S C S C	GCC A CCG P GCCCG P GCCCG CAC H GAT D CCGC R CTG CTG CTG CTG CTG CTG TGC TGC TGC	GAG E AAA K GCCCCG GGC GGC GGC GGC GGC GGC CG CGC CG CG	CCG P CCG P ggcgc CGC R ACG G ACG G ATC I GAA E GAA E GAA E C G ACG D ACG C S ACG C C C C C C C C C C C C C C C C C C	GTC V GCG A cctgg GCG A C GCG A GCG A A GCG A A A C C T T T T T T L A C G C	AAG K GCG A ggag TTC F GCG A GCG A GCG A GCC A GCC A GCC A GCC G C G	GCC A AAG K K tgtaaggc tgtaa AACC T T T T T T T T C T G C T G C T G C T G C T G C T G C T G C T G C T G A C G A A G A A G C K K S C K S C S C S C S C S C S C S C	GCG A CCCC P gctg AAC CTT L CCG GAA E ATC L CTG GAA C T GAC C T GAC A AC C C C C C C C T L	CCC P GCC A acacgo acacgo ATC GGC G ACC T T TCC S GCC A CAG Q CCA Q CCA GTC V AAG GTC V A	GTG V CCG P ggcai ccga TAC Y AAT N GGG GGG GGG GGG GGG GGG QGG GGG GGG G	GCG A GCG GCG A Jaaac GGG G G C C C C C C C C C C C C C C C	GCG A CCCC P CCCC P CCCC A TG CCCC R A ATG A ATG GCC A A CCGC C C G GCC A C CCCC P CCCC P CCCC P C CCCC P P CCCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P P CCCC P CCCC P ATG A TG CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P ATG A ATG A AGG GCCC CCCC P ATG A AGG CCCC P A CCCC P A CCCC P A CCCC P A CCCC CCCC P A CCCC CCCC P A CCCC CCCC P A CCCC CCCC P A CCCC CCCC P C CCCC CCCC P C CCCC CCCC P C CCCC CCCC P C CCCC CCCC P C CCCC CCCC C CCCC C CCCC C CCCC C CCCC	CCG P GCG A gggc CAC H GAC D TGG GAC D TGG M GAC D CAC CAC CAC C CAC C CAC C CAC C C C	1694 1124 1769 137 1858 1958 2042 16 2192 66 22267 91 2342 116 22342 116 2417 141 2492 141 2492 141 2492 2567 191
1620 88 1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 42 2193 67 2249 117 2418 42 2193 67 2268 92 2343 117 0 2493 167 2493 167 2568 20 2493 2568 2568 2568 2568 2568 2568 2568 2568	GTG AAA V K GTC AAG V K CCG AAA P K gcctgcctgc ccggacacc GAC CGC D R TGG ATC TGG ATC TCC TTC TGG ATC GC K GC GCG G A TGC AAG C C GGT GCG G A ATG AAA M K C CGC CGC G A ATG AAA C CGC CGC C CGC CGC C C C C	GTG V ACG T GGC G G G G G G G G C G G C C G G C C G G C C G G C C G G C C G G C C G G C C C G C	GCC A AAG P P AAG K trogg ccccg CTG L GAG E CCCG P CCCG CCG CCG A GAG E CCCG A CCCG P	GAG E GAG E TCG S Gatç CCGa AAG K CAG Q AAAG K GAA E GCC A TAT TTC F	GCG A CCG P AAG K treece aacc GGC GGC V V CAA Q ATC I TAT TAT Y GGG GCC Q CCA Q CCA Q CCA Q CCA C C C C C C C	CCG P GTG CCC PGTG V CCCC P CGGGAGGG AAGGAGGGAAAGGAGGGAAAGG AAGGAGGGGAAGGGGAGGGAGGGAGGGGAGGGGAGGGGGAGGGG	AAA K GCC CGT CGT R CGGC CCGC A ATG GCC A A GGC CGC CGC CGC G GGC GGC CGC C	CCG P GCG A CCGC R CCGC R CCGC A CCGC A CCAC CAC H ATT I GCG G G G G G G G G G G G G G G G G C CAC C C C	GCG A GTG V GGC G G G G G G G G G G G G G G G C C G C C G C C G C C G C C G C C G C C G C	CCG P GCT TG2 CTG2 CTG G G CTG CCG CCG CCG CCG CCG CCG CCG	AAA K GAA E A gaa A CGC R CAC H CGC R CAC TCG GAA E GCC A CAC A CAC TTCG A CAC TTCG CAT TTCG CAT TCG CAT TCG CAC A CCCC CAC A A CCCCC A A CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC A A C CAC C A A C CAC C C CAC C A C CAC C CAC C CAC C CAC C CAC C C CAC C C C C C C C C C C C C C C C C C C C	GCC A GCG GCG GAT D D TGG GAT T T T T T T T T T T T T T T T T	GCC A CCG P GCCCG P CCCCG CAC CCCG CCCG	GAG E AAA K Joccog GAT D D GGC G GGC G GGC GGC GGC C G C C C C	CCG P CCG P gcgc CGC R ACG T GGT GGA ATC I GGA ATC D CGA ATC C S ACG T C GAA C GAA E CGA C GAA C GAA C GAA C G C G G T	GTC V GCG A setgg GCG A ATC I GCG GCG A A AACC G A A AACC T T CTT T A CGT T	AAG K GCG Qggag TTC F GCG GCC A GCC A GCC A GCC A GCG C C C C	GCC A AAG K K AAG T TTT F TTT T T TTT T F GGAT GCC A TTC F GGC A AAC N CTG A ACC	GCG A CCCC P gctg gctg CCG P CTT L CCG GAA E ATC L GAA CTG L GAA CTG K AACC	CCC P GCC A Aacgc I GCC G GCC J GCC G GCC T T T CCC S GCC A ACC Q GTC V AAG GTC V V ACG GTC V	GTG V CCG ggca ccga TAC V AAT N GGG G GGA G GGC A TAT Y GGG G G C A A TAT Y S GGA G G C C A A C C C C C C C C C C C C	GCG A GCG A gaaaa GGC G G G G C C C G G C C C C C C C	GCG A CCC P CCCC P CCCC A TG CCC C R A AAG GCC A A CGC G GCC A A TC CAC C A C C C C C P P C C C C C C C C	CCG P GCG A gggc CAC H GAC D TGG W ACC T GAC D CAC CAC CAC CAC C CAC C CAC C C CAC C C C C C C C C C C C C C C C C C C C	1694 1129 137 1858 1958 2042 16 2117 41 2192 66 2267 91 2342 116 2417 141 2492 166 2267 91 2342 166 2567 191
1620 88 1695 113 1770 138 18599 1 2043 17 2118 42 22193 67 22193 67 22418 12418 12493 1672 2493 1672 2418 12493 1672 2418 12493 1672 2568 292 2343 1272 2418 12493 1672 24213 12418 12493 1672 2568 2562 2343 1272 2418 12493 1672 2568 2243 1272 2418 12493 12592 1242 1242 1252	GTG AAA V K GTC AAG V K CCG AAA P K gcctgcc cgcccgc GCC Q C R TGG ATC TGG ATC C TC S F TGC AAG C K GCC GCG G A ATG AAA M K GTC GTG GTG V V	GTG ACG T GGC G GGC GGG GGC S ATC G ATG ATG ATG CGG CCG G CCG G CCG P P	GCC A CCC P P AAG K CTG GCC GCC GCC GCC GCC GCC CCC CCC CCC	GAG E GAG E TCG S gatç ccga AAG K CAG Q AAAG K GAA E GCC A GCC A TTTC F ATC I	GCG A CCG P AAG G G CCG G G G C C V CAA Q ATC I TAT T Y G G G G C C G P C C G C C C C C C C C C	CCG P GTG CCC P GTG A CCC P C GTG A AAG AAG AAG A ATG C CTG C C GCG A CGC R	AAA K GCC CGT R CGT R CGT R MUC ATG GCC A ATG CGC CGC CGT CGT R	CCG P GCG A CCGC R CCGC R CCGC A CCGC A CCGC A CCGC G G G G G G G	GCG A GTG GTG GTG GTG GTG CGC CGC CGC CGC CGC	CCG P GGT C CGGC CTC CTC CTG CTG CCG GGT CCCG P GGC CCCG P CCCG CCCC CCCC	AAA K GGAA E a cggggg AAC N CAC R CAC H CGC CAC H CGC S CAT H GAA E GCC A TCG S CAT TCG S CAT TCG S CAT TCG S CAT TCG S CA CAT CAC CAC N CAC CAC N CAC CAC N CAC CAC N CAC CAC	GCC A GCG GCG GCG GCG D TGG GGAT T TGG G G TTTT F F ACG G TTTT TATTY TGC C CTTL GGC C G TTTC F	GCC A CCG P gcctc CAC H GAT D CCGC R CTG C CTG C CTG C CTG C CTG C C CG C C G C C C C	GAG E AAA K GCCCG GAG GGC GGC GGC GGC GGC GGC CG CG CG CG CG	CCG P GCG P GCGC CCG P GGT GGT GGT GGA ACG GAA E GAA C GAA E GAA C G CGC I GAA C G CGC T C GC T C GC C G C C G C G C C G C G	GTC V GCG A sctgg GCG A ATC I GCG A A AAC G G G A AAC C T T L ACG T C C T C C T C C C C A C C C C C C C C	AAG K GCG gggag ggag TTC F GCG G GCC A GCC A GCC A GCC A GCC A GCC G C C C G C C C C	GCC A AAG K aggc tgta AACC T T T GAT C C T G C C G G C C G G C C G G C C G G C C G P	GCG A CCCC P gctg AAC N CTT L CCCG P GAA E AATC C T G GAC L CTG GAC AATC D AAG AAC N N	CCC P GCC A accag GACAG GACAG GGC G AACC T T CCC S GGC C AACC V AACC V A	GTG V CCG P ggca TAC Y AAT TAC GGG GGG GGG GGG GGG C GGG GGA A E C GGG GGA A E C C C C C C P P ggca C C C C C C C C C C C C C C C C C C C	GCG A GCG A gaaaa GGC G GGC G G CCC G CCC G CCC CCC A CCCG A CCCG A ATG S GGC G G CCC G G G CCC CCC G G G G G	GCG A CCC P CCCC P CCCC C A CCC C C C C C C	CCG P GCG A Cccga gggc CAC H GAC D T GG M CAC T ATG GAC D CAC C CAC H CGG CAC A AAG C A CAC K	1694 11769 1376 1858 1958 2042 16 2117 41 2192 66 2417 2192 66 2417 2141 2492 166 2417 141 2492 166 2417 191 2567 191 2567 191 2567 191 264 2717
1620 88 1695 113 1770 138 18599 1 2043 17 2118 42 22043 17 2118 42 22193 67 22268 92 2343 117 2418 142 2493 167 2493 167 2493 2643 2217 2718	GTG AAA V K GTC AAG V K CCG AAA P K gccctigc Gggcacac GGC CGC D R TGG ATC TGG ATC TGG ATC C TTC S F TGC AAG C K GGC GCG G A ATG AAA M K GTC GTG V V V CTG TTC	GTG V ACG T G G G G G G G G G C G G C G G C G G C C G G C C G G C C G G C C G C C G C	GCC A CCC P AAG K CTG CCG CCG CCG CCG CCG CCG CCG CCG CCG	GAG E GAG E TCG S ggatg cccga K CAG Q AAG GCC A GCC A TAT TTC F ATC TTCG	GCG A CCG P AAG K GCC G G GCC Q G G G C C Q AAC I TAT Y Y GGG G G C C G C C Q A C C A G C C C C C C C C C C C C C C C	CCG P GTG V CCC P CGG GGG A AAG GCG A AAG K CCG S A TGC C C GCG A CGC R CAC	AAA K GCC A CGCC R R CGG G A A TG M GCC A A TG C G C C R TAT Y A TC C G C C R C R TAT Y A TC C G C C R C R C C C R R C C C R R C C C C	CCG P GCG A CCC R Atccgggac D F GCG G GGG GGG GGG GGG GGG GGG GGG GGG	GCG A GTG GTG GTG GTG GT GTT CCG C R CGT R GGA CGC CGC R AAG GAA E GGA G G G G G CGC CGC C R CGC CGC C CGC C G CGC G CGT G CGT G C G C	CCG P GGT C CTG CTG CTG CTG CCG GGT CCCG P GGC CCCG P CCCG CCCG	AAA K GGAA E A GGAA A CAC CAC H CGC CAC H CGC CAC H GGAA E GCC A CAT TCG GCC A TAT TGG W TGC	GCCC A GCG GCG A C agac D TGG GAT T TGG G G TTTT F ACG T T TAT TGC C CTTL GGC C G TTCC F GGC G TTCC G GCC G GCG GA A A C A C A C A C A C	GCC A CCG CCG P GCCCG CAC H GAT D CCGC R CTG CTG CCG CTG CCGC C CCG CTG CCG CCG	GAG E AAA K gccccg GGT GGC GGC GGC GGC GGC CGC CGC CGC CGC	CCG P CCG P gcgc CCC R ACG T GAA CG T GAA CGA CGA CGA CGA CGA CGA CGA CGA	GTC V GCG A CCTgg GCG A CTG C A CCT A GCG A A A CCT C C C C C A A C C G C C C C C C C C C C	AAG K GCG A gggag TTC F GCG G GCC A GCC G GCC A GCC A GCC G C G	GCC A AAG K tgtaa AACC T T T T T T T T T T T T T T T T T	GCG A CCCC P gctg AAC N CTT L CCCG GAA CCTG L GAA CTG GAA CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	CCC P GCC A accag ATC I GGC G G ATC T T CC G G C A ACC T C CAG Q GTC V AAG K K GTC V AAGC	GTG V CCG P ggca TAC Y AAT TAC GGG GGG GGG GGG GGG GGC GGC A TAT Y GGG GGA A E ACC T ATG	GCG A GCG GCG GGC GGC GGC GGC CTC CCG G CTC CCG G CTC CCG G CTC CCG G CTC CCG G CTG CCG G CTG CCG G C C GCC G C C G C C C C	GCG A CCCC P CCCC P CCCC A TG CGC CGC CGC CGC CGC CGC CA A ATC I CCCC CGC CGC CGC CGC CGC CGC CGC CGC	CCG P GCG A SCCCGA GGC CAC H GAC D TGG CAC T TGG CAC T CAC C T GAC D CAC C C C C C C C C C C C C C C C C	1694 1129 13769 1378 1958 2042 16 2117 41 2192 66 2417 2192 66 2417 2192 2267 91 2342 116 2492 166 2411 2492 166 2411 2492 2567 191 266 2417 2492 2567 191 266 2717 266 2717 267 291 2792
1620 88 1695 113 1770 138 91959 1 2043 17 22118 42 2193 22418 142 24193 167 22418 142 2493 167 22417 2418 142 2493 2417 2568 2247 2568 22718 24793	GTG AAA V K CCG AAA P K GCCCTGCC GAC CGC D R TGG ATC W I TCC TTC S F TGC AAG C K GGC GCG G A GGC GCG G A ATG GAA K C CTG TTC L F ATC GAA	GTG V ACG T GGC G GGC G GGC G G CCG G CCG G CCG G CCG G CCG G CCG G CCG G CCG G AAG AA	GCC A A CCC P P AAG CCCG CCCG CCCG CCCG	GAG E GAG E TCG S ggatg ccga ccga CCGA CAG CAG CAG GCC A CAG GCC A TAT TTC F F AATC I TCG S GGC	GCG A CCG P AAG K tcccca GCC G GCC Q GCC Q ATC I TAT Y GGG G GCC CCG P CCG G CCG Q CCG Q CCG GGG GGGC	CCG P GTG GTG GTG GTG P CCCC P ggaagg GCG A AAG K TCC S ATC C TCC C C C C C C C C C C C C C C C	AAA K GCC A CCA gaati NUC AATG AATG AATG AATG CAC CGC CGC CGC R TAT Y AATC I GGC CGT R GGC CGT R GGTC V GGTC	CCG P GCG A CCC R Atcccgggac/ DF GCG G GGG G GGG G GGG G GGG G GGG G GGG G	GCG A GTG GGC G G G G CGT R R GGG G CGC R R G G G G CGC R R A A G G G G G G G G G C G C T C C C C T R C G C G C G T G C G C G C G C G C G C G	CCGG P GGCT A TGJ CTG G GGT CTG G GGT CTG CTG CCG P GGC CCG P GGC CCG CCG CCG CCG CC	AAA K K GGAA E A GGGG A AAC C R CAT H CGC C A CAT H GCC A C GCC A C GCC A TGG W TGC C TGG C TGG C TGG C TGG C TGG	GCCC A GCG GCG Cagaac ttgcg GAT T TGG G TTGG G G TTT TAT Y Y GCC C TT T GC C G TTC G G C TTC C G T TCG G GCC A A A C A C A C A C A C A C A C	GCC A CCG P P CCCC C C C C C C C C C C C C C	GAG E AAA K gcccg GGC GGC GGC GGC GGC GGC GGC CGC CGC CG	CCCG P CCCG P gcgcgc CCC R ACG G T T GGT T GGT ATC I GAA E GGAA CG C G C G C G C G C G C G C G C G	GTC V GCG A CCtgg GCG A CCT C C C C C C C C C C C C C C C C C	AAG K GCG A gggag TTC F GCG GCG A GCC A GCG G GCC A GAG GAG E CGG CG CG CG CG CG CG CG CG CG CG CG CG	GCC A AAG AAG K tgta AACC T T T T T T T T T T T T T T T T C T G G C C T G G C C T G C C T G C C T G C T T T T	GCG A CCCC P gctg dAAC N CCTG CCTG CCTG CCTG CCTG CCTG CCTG C	CCC P GCC A acage ATC I GCC G GC C T T T C C S GCC A ACC S GCC A CAG Q C A CAG S C CA S C CA A A C S C C C C C C C C C	GTG V CCG P ggca TAC Y AAT TAC Y AAT GGG GGG GGG GGG GGA C C A TAT Y GGG G GAA TAT T ATG GGC C GGC C CCG P P ggca CCG P P ggca CCG P P ggca CCG P P CCG P P ggca CCG P P ggca CCG P P CCG P P CCG P P CCG P P CCG P P CCG P P CCG P P CCG TAC C C C C C C C C C C C C C C C C C C	GCG A GCG GCG GGC GGC GGC GGC CTC CCG F GCC A A CTG CTG CTG CTG GGC A A GGC CTG CTG CTG CTG CTG GGC CTG CTG CTG	GCG A CCCC P CCCC P CCCC A CCC A A CCCC R A AAG GCC C A A ACC C CCC C C C C C C C	CCG P GCG A Cccga GgGc CAC H GAC D TGG M GAC D CAC T GAC D CAC T GAC C C CAC C C C C C C C C C C C C C C C C C C C	1694 112 1769 13769 1858 192042 1652 2042 1662267 2117 41 2192 26627 2342 1162 2417 141 2492 16667 291 2642 2792 26667 2792 26667
1620 88 1695 113 1770 138 1859 1959 1 22043 2118 22043 2118 2218 2218 2218 2218 222343 117 2268 22418 142 2418 142 2493 167 2568 242 2643 22718 242 2718 247 2758 267 267 267 267 267 267 267 267 267 267	GTG AAA V K GTC AAG V CAAG V CCG AAA P K GCCCGCC GGC CGC D R TGG ATC C TGG ATC GGC GCG GGA AG GGA GCG GGA AG GGA GCG GGA AAA M K GGC GTG GTC GTG V V CTG TCC L FA ATC GAAC	GTG V ACG T GGC GGC GGC GGC G GGC G AATG G ACC G G CCG P CCG G CCG P CCG G CCG AAG CCG CCG CCG CCG CCG CCG C	GCC A A CCC P P AAG GAG E CCCG CCCG CCCG CCCG CCCG CCCG	GAG E GAG E TCG G G CCGa AAG K CAG Q AAG C CCG A AAG C C C G C C C G C C C G G C C C G G C C G G C C G G C C G AG C C S S C C G AG S S C C G AG S S C C C S S C C C G AG S S C C C C S C C C C S C C C C C C C	GCG A ACCCG P P AAGG G C C C C A C C C A C C C A C C C C	CCG P GTG GTG GTG GTG P CCC P P GTG A A A A C C C C C C C C C C C C C C C	AAA K GCC GCC CGT R CCagge CGT M GCC ATG ATG GCC CGC CGC CGC CGC CGC CGC CGC CGC C	CCG P GCG A CCGC R Atcccg ggacd A GCG CAC H TCG G G G G G G G G G G G G G G G G G G	GCG A GTG GTG GGC G G G G CGT R R GGG G G G G G CGC R R AAG G G G G G G G G G G G G G G G G	CCG P GGT TGJ TGJ CTG GGT CTG GGT CTG GGT CCG P CCG CCG P CCG CCG CCG CCG CCG CCG	AAA K GGAA E GGAA C GGGG AAAC C R C C C C T C G C A C A C C C T C G C C C T C G C C C T C G C C C T C G C C C T C G C C C C	GCC A GCG A cagad trgcog GAT T T GGC G G T T T T T T T T T T T T T	GCC A CCG P P CCCC CCG H H CCAC CCG C CTG L CTG CTG CCGC C GCG CTG CCGC C GCG GCG A GTC V AACC N N CCCC N P P CCCC S P P CCCC S P P CCCC S P P CCCC S P P CCCC S P P CCCC S P P CCCC S P P C CCC S P P C CCC S P P C CCC S S P P C CCC S S P P C CCC S S P P C CCC S S S C S C C S C C S C C S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S S C S	GAG E AAA K GCCCG GGC GGC GGC GGC GGC GGC GGC CGC G GGC CGC G GAA E CCCG G GAA E CCCG G CCCG CCC	CCCG P GCCG P gcgcgC CCC R ACG G ACG T T GGT T GAA CGAA E GAA CGAA C S ACCG T C GAA C S ACCG CGAA C S ACCG CGA C GAA C G C G C C C C C C C C	GTC V GCG A Ctgg GCG ATC I GCG A A AAC G G A A CTT L CTT L CTT CTT C G G G G G C G A AC C T A C C T A C C T A C C C C C C	AAG K GCG GCG A GCG F GCG GCG GCG GCG GCG GCG A GCG GAG GAG G	GCC A AAG AAG AAG T T T T T T T T T T T T T	GCG A CCCC P gctg daac N CCTG CCTG CCTG CCTG CCTG CCTG CCCG C AACC N AACC N AACC N AACC P P CCCG C CCCC P P CCTG CCCC P P CCTG CCCC P P CCCC P P CCCC CCCC	CCC P GCC A acage ATC I GCC G GCC T T TCC S CCC A CAG Q Q CAG GTC V AACC N AACC N AACC S GTC V AACGC GTC C C A	GTG V CCG P ggcai GGC A AT GGG GAG GGC GGC A AT GGG GGA C T TAT Y ATG GGG GGC C A ATG GGC C A C C GGC G C C C C C C C C C C	GCG A GCG GCG GCG GCG GCTC CCTC F TTC CCG GCC CCTC CTC CCTC C	GCG A CCCC P CCatc GCC A TG GCC A A GCC C A CGC C A CGC C C C C C	CCG P GCG A gggcccga CAC H GCC CAC H A CCG CAC H CAC CAC CAC CAC CAC CAC CAC CAC C	1694 112 1769 137 1858 2042 16 2117 4192 2267 91 2116 24192 2267 91 2342 241 24192 22667 12642 2251 241 241 2492 2567 12642 22717 241 22667 241 22791 2667 241
$\begin{array}{c} 1620\\ 88\\ 1695\\ 113\\ 1770\\ 138\\ 1959\\ 1959\\ 2043\\ 17\\ 2118\\ 42\\ 21193\\ 67\\ 2218\\ 22193\\ 2218\\ 22193\\ 2218\\ 2228\\ 22343\\ 142\\ 2493\\ 2568\\ 192\\ 2568\\ 192\\ 2568\\ 192\\ 2718\\ 242\\ 2793\\ 2568\\ 242\\ 2793\\ 2671\\ 2868\\ 242\\ 2793\\ 2867\\ 2862\\ 292\\ 2892\\ 292\\ 292\\ 292\\ 292\\ 292\\ 2$	GTG AAA V K GTC AAG V K GTC AAG V K P K GCC GCC D R TGG ATC C TGG ATC GAC CGC D R TGG AAG C S P GCC GCG G A GAA TGC CGC GG AAG GCA GGG G AAG ATG AAA M K GCC GGC GGG Q ATG AAA M K GTC GTG TC L F ATC GAA I E CCG GTC P V V	GTG V ACG GGC GGC GGC GGC GGC AATG GAC CCG G CCG CCG CCG CCG CCG CCG CCG	GCC A A CCC P AAG CCC CCC P CCC CCC CCC CCC CCC CCC CCC	GAG E GAG E CCG S GCC CCG A AAG K CAG Q AAG K CAG GAA E E GCC A TAT T T T C G GCC G GCC A ATC GCC A AC C A GCC A A C A G A C A G A C A C	GCG A ACCCG P P AAGG G CCG G CCG C CAA C CAA C CCG CCG	CCG P GTG GTG GTG C P GggaG GGG A AAG AAG C S AAG C C C GCG A CGC C C C C C C C C C C	AAA K GCC GCC R CCGT R CCGGC ATG GCC ATG GCC C C C C C C C C C C C C C C C C C	CCG P GCG A CCGC R R CCGC S GCG A A T C G G G G G G G G G G G G G G G G G G	GCG A GTG GGC G GGC G G GGC G CGC R R GAT D CGC R R GAA GGA CGC T GGG G GGC G GGC CGC C C C C C C C	CCG P GGT TGP CTGP GGT L GGT L GGT L GGT CTG CTG GGC CTG GGC CCG P CCG GGC CCG P CCG GGC A ACC CTT L GGC T GGC C GGC GGG A	AAA K K GGAA E CGGCA A CAC C C C C C C C C C C A T C GCA C A C A C C C C T C G C A T C G C A T C G C A T C G C A T C G C A C A C C C C C C C C C C C C C C	GCC A GCG A cagad trgcg GAT T D TGG G G T T T T T T T T T T T T T	GCCC A CCG CCG P P CCCC CAC CAC CAC CAC CGC CCG CCTG L CTG CCTG L CTG CCG CCG CCG CCG CCG CCG CCG CCG CCG	GAG E AAA K GCCCG GGC GGC GGC GGC GGC GGC GGC GGC	CCCG P ggcgcg CGC R ACG GGT GAA E GAT C GAA E GAA E GAA E C GAA CG C C C C C C R ACG GAA E C GAA C C C GAA C G C C C R C R C R C C C R C R C R C R	GTC V GCG A GCG ATC I GCG A A GCG A A AAC C T A C GC T A C G C T A C G C G G G G G G G G G G G G G G G G	AAG K GCG GCG A GCG F F GCG GCG GCG GCG A GCG GCG A GCG GAG GAG	GCC A AAG K aggc tgta AACC I TTTT F GAT D CTG GAT CTG G GCC A ACC N CCG G CCC S AACC R	GCG A CCC P P gctg CCC P CCT CCC CCC CCC CCC CCC CCC CCC C	CCC P GCC A accag ATC GGC G C G G C C A G C C A G C C A G C C A G C C A A C C S G C C A A C C S C C C C C C C C C C C C C	GTG V CCG P 9ggcai TAC V V AAT N GGG G GAG G GGC GGC GGC A A CA C GGC C A C GGC C A C C G G C C G G C C G C C G P P 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	GCG A GCG GCG GCG GCG GCG CTC CCG GCTC CCG GCTC CCG GCC CTC GCC CCG GCC CCC C	GCG A CCCC P CCatc GCC A TG GCC A A CGC A A CGC G GCC A A CCC G GCC A CCC C C C	CCG P GCG A gggcccqa CAC H GCC CAC H A CCG CAC H CAC T CAC CAC CAC CAC CAC CAC CAC CAC C	1694 112 1769 137 1858 1485 2042 2117 16 2117 2192 2247 2162 2247 2162 2247 22412 2267 2342 2166 22417 1262 22642 26642 2642 26444 264444 26444 264444 26444 26444 26444 26444 26444 26444 264444 2644
1620 1695 113 11770 138 1959 1959 2043 17 2043 17 2118 42 2193 67 2218 2193 67 2248 92 2343 142 2493 1656 142 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 1657 2418 2493 2418 2493 2418 2493 2418 2493 2418 2493 2418 2428 2493 2418 2428 2443 2477 2418 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2443 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2428 2477 2478 2478 2477 2478 2478 2478 2477 24788 2478 2478 24	GTG AAA V K GTC AAG V K CCG AAA P K gecctge: ggacac: GAC CGC D R TGG ATC TGG ATC GC TGG ATC GC GCG G A GAA TGC E C GGT GCG G A ATG AAA TGC GCG G A GAT GCA M K GTC GTG V V CTG TCC L F ATC GAA I E CCG GCC C C C C C C C C C C C C C C C C	GTG GGC GGC GGC GGC GGC GGC GGC GGC AATC I I ATC GAC D CCGG CCG G CCG G CCCG P P CCCG G CCCG A AAG CCC CCCG CCC G CCCG C CCCC CCCC CCCCCC	GCC A A CCCC P AAG CCC CCC CCC CCC CCC CCC CCC CCC CCC	GAG E GAG GAG E TCG S Gat C Ccga AAG K CAG GAA E CAG GCC A CAG GCC A TAT TTCC F C GGC GGC G GCC GCC GCC GCC GCC GCC	GCG A A CCG CCG P P AAG G G G C C C C G G C C C C C C C C C	CCG P GTG GTG V V CCCC P GggaG GCG A AAG X TCC S AATG C C C C C C C C C C C C C C C C C C C	AAA K GCC CGT CGT C R CCGG A ATG GCC A ATG GCC CGC R TAT Y ATC I GGC CGC CGC CGT R R CGC CGT C CGT C CGT C CGT C CGT C CGT C CGT CGT	CCG PGCG A CCGC R R tCCG GGCG CAC H ATTCG GGCG GGC GGC GGC GGC GGC GGC CAC H CCCC CAC H ATTCG GGC GGC GGC GGC GGC GGC GGC GGC GGC	GCG A GTG GGC G G G G G G C G C G C G C G C G	CCGG GGT X A CTG2 GGT L GGT L GGT L GGT C CCG G GGC CCG P P CCG G GGC CCC P CCG G GGC A ACC CTT L CCT L GGC C CCC G GGC GGC A CAG	AAA K K GGAA E CGGC A CAC C C C C C C C C C C C C C C	GCCC A GCG GCG Cagaco D TGG GCT TGG GC TTTT F ACG TTCC C TTTT TATT C G GTC C C TTCC G GTC V V AAG G GCC A A GCC G GTC C C TTTT T A C G G G G G G G G G G G G G G G G G G	GCCC CCG CCG P P CCCC CAC GCC GCC CAC CCG CCG CCG CCG CC	GAG AAA K gccccg Ggacg GGC GGC GGC GGC GGC GGC C C C C C C C	CCCG P ggcgcg CGC AGGG T GAT G ATCG GAT C I GAA E GAA E C GAA E C GAA C G C C C C R ATCG G ATCG G ATCG G ATCG G ATCG G ATCG G ATCG G ATCG G G ATCG G ATCG G ATCG G ATCG G ATCG G ATCG G ATCG G ATCG C C R ATCG C R ATCC C R A ATCC C R A ATCC C R A ATCC C R A ATCC C R A ATCC C R A ATCC C R A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC A ATCC	GTC V GCG A GCG GCG A GCG G GCG A AAC C T GCG G G G G G G G G G G G G G G G	AAG K GCG GGG GGG F GCG A GGC G GGC A GAG GAG C GAG C GAG C GAG C CGG A ATG M ATG V ATG M ATG	GCC A AAG K aggc tgta AACC I TTTT F GAT D CTG GAT CTG GCC A ACC N CCG G CAACC R GCC S AACC R CCG CCG CCG CCG CCG CCG CCG CCG CCG	GCG A CCC P P gctg CCC P CCT CCG P GAA CCTG L CCG CCG CCG CCG CCG CCG CCG CCC CCG CCC CCG CCC CCC CCC P P GAAAC CCC CCC P P GCC CCC P P CCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC P P GCC CCC C	CCC P GCC A accgc GCC G G G G G C C T C C C G G C C A G C C A G C C A C C S G C C A C C S G C C C S C C C C C C C C C C C C	GTG V CCG CCG P ggca CCG CCG P ggca V AAT N N GGG G G G G G G G G G G G C C G G G G	GCG A GCG GCG GCG GCC GCC GCC CCC F GCCC CCCG CCCG G CCCG G CCCG CCCG CCCG G CCCG CCCG CCCG CCCG CCCG CCCG CCCG CCCC	GCG A CCC P Coatc M CGC R AAG AAG G CGC R A AAG GGC CAC H GGC CAC I GTC V V GGG GGC I S TTC S	CCG P GCG CAC CAC CAC D TGG W AACC T TGG CAC D CAC CAC CAC CAC CAC CAC CAC CAC C	1694 112 112 137 1858 1485 2042 2117 41 2192 2267 91 2342 216 2417 116 2417 116 2417 116 2417 116 2567 191 2342 216 2577 241 2192 264 2267 291 2942 2192 2042 2192 2042 2197 2042 2197 2042 2197 2042 2197 2042 2197 2042 2197 2197 2197 2197 2197 2197 2197 219
1620 1695 113 1770 138 1959 1959 2043 17 2118 42 2193 67 2218 42 2193 67 2248 92 2343 17 2418 42 2493 167 2268 92 2343 217 2418 2493 167 2268 92 2343 217 2418 2493 167 2268 92 2343 217 2418 2493 167 2268 92 2343 217 2418 2493 267 2864 292 2864 292 2748 277 2868 292 277 2868 292 277 2868 292 277 2788 292 277 2788 292 277 2788 292 277 2788 292 277 2788 292 277 2778 2779 2788 2797 2778 2797 2788 2797 2787 2797 2788 2797 2778 2797 2788 2797 2792 2797 2778 2797 2792 2797 2792 2797 2778 2797	GTG AAA V K GTC AAG V K CCG AAA P K gecctgec ggcccc D R TGG ATC TGG ATC TGG ATC S F GCC GCG G A GAA GCC C K GCC GCG G A GAT GCAA GCT GTG C L F ATC GAA I E CCG GTC P V CTG TTC L F ATC GAA I E CCG GTC P ATC AAG C C C C ATC AAG C C C C C A C C C C C C C C C C C C C C	GTG GGC GGC GGC GGC GGC GGC GGC GGC GGC	GCC A CCC P P AAG CCG GCG GCG CCG CCG CCG CCG CCG CCG	GAG E GAG GAG E TCG S Gat C CCG A AAG K K CAG GCC A AAG K CAG GCC A TTTC F F C I TCG GGC G GCC A TTTC TCC GCC GCC A TTCC TCC TCC C CCGA C C CCGA C C C C C C C C C C C	GCG A A CCG CCG P P AAG G G G C C C C G G C C C C C C C C C	CCG P GTG V CCCC P CCG P CCG C C C C C C C C C C C C C	AAA K GCC CGT R CCaga G ATG G G CGC CGC CGC CGC CGC CGT R ATC G G G CGC CGT C CGT C CGT C C TGT C C TGT C TGT C TGT C TGT C TGT C TGT TGT	CCG PGCG A CCGC R R CCGC GGCG CAC H ATTCG GGCG GGC GGC GGC GGC GGC GGC CAC H ATTCG GGC GGC GGC GGC GGC GGC GGC GGC GGC	GCG A GGC GGC GGC GGC GGC GGC CGC R R GGA CGC CGC CGC CGC CCC GGC GGC GGC GGC	CCG P GGT TG2 CTG2 CTG GGT CTC CTG CCG CCG CCG CCG CCG CCG	AAA K GAA GAA GAA COUC CAC R CAC R CAC R CAC R CAC CAC CAC CA	GCCC A GCG GCG Cagaco D TGG GCT TGG G GCC TTTT T TTTT TAT TGC C TTC C TTC C TTC G G C C TTC C F C G G C C TTC C G G C G T C G G G G G G	GCCC A CCG P P CCCCG P CCCC CCG CCTG CCTG CCTG	GAG AAA K Gcccg Ggacg GGC GGC GGC GGC GGC GGC C C C C C C C	CCGC P ggcgcc ggcgcc R GGT GGT GGT GGA CGC C GAA C GAA C GAA C C GAA C C GAA C C C GAA C C C C	GTC V GCG A GCG A ATC I GCG A GCG A A ACC N GCG A A ACC T T GCG G G G C T T GCG G G C T C CTT I L C CTT I C CTT I C C C T T G C G G G G C T G C G C C A T C C C C C C C C C C C C C C C C	AAG K GCG GGG GGG F F GCG A GGC G GGC A GGC G A GGC A GGC C G G G G	GCC A AAG K aggc ttgta AACC T T T T T T T T T T T T T T T C T G C T G C T G C T G C T G C T G C T G C T G C T G C T G C T G C T T T T	GCG A CCCC P gctg gctg CCG N CTT L CCG GAA CTT L CCG GAA CTL CCG GAA CTL CCG CCC P P ATC CTT L CCG CCC P P CTT CCC CCC P CCCC P C CCC P C CCC P C CCC P C CCCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC P C CCC CCC P C CCC CCC P P AAC CTT L CCCG C CCC CCC P P ACC CTT L CCCG C CCC CCC P P ACC CTT L CCCG CCC CCC CCC CCC P P ACC CTT L CCCG CCC CCC CCC CCC CCC CCC CCC CCC	CCC P GCC A acagc C GCC G G C C G G C C A G C C A G C C A G C C A G C C A C A	GTG V CCG CCG P P ggca CCG C C C C C C C C C C C C C C C C C	GCG A GCG GCG GCG GCC GCC GCC CCG GCTC L CCCG GCC CCG GCC CCG GCC CCG GCC CCG GCC CCG GCC CCG GCC CCG GCC CCG GCG CCC CCG CCCC	GCG A CCC P Coatc M CGC R AAG AATG AATG AATG CGC R AATG CGC R AATG GGC C A AATC CGC C C C C C C C C C C C C C C C C C	CCG P GCG GCG CAC CAC CAC D TGG W AACC T TGG CAC CAC CAC CAC CAC CAC CAC CAC CAC C	1694 112 112 1137 11858 1885 2042 2042 2117 16 2192 2192 2267 91 2342 216 2247 116 2342 2267 291 2342 216 2567 291 2717 241 2792 22642 2172 2912 2942 2912 2942 2951 2942 2951 2951 2951 2951 2951 2951 2951 295
1620 1695 113 1770 138 1959 1059 12043 17 2218 42 2343 17 2218 142 2343 17 2418 142 2343 17 2418 142 2343 17 2418 142 2343 17 2418 142 2343 17 2418 142 2343 17 2418 142 2493 167 2417 267 2643 27 28 267 27 28 27 28 27 28 27 28 27 28 27 28 27 28 27 28 27 28 27 28 28 27 28 28 27 28 28 27 28 28 27 28 28 29 27 28 28 29 27 28 28 28 29 27 28 28 29 27 28 28 28 29 27 28 28 28 27 28 28 27 28 28 28 28 28 27 28 28 28 28 28 28 28 28 28 28	GTG AAA V K GTC AAG V K CCG AAA P K gcctgcctgc GCC GCA D R TGG ATC TGG ATC TGG ATC TGG ATC GC K GC GCG G A GAA TGC E C GGT GCG G A ATG AAA M K GTC GTG V V CTG TTC L F ATC GAA I E CCG GTC CG A TC AAG CAC CCG GTC P V GGC ACC C GT C CG C T AACCAC C CGC CG C A CCG CTC C ACCAC CCG CTC C C C CGC CCC C C C C C C C C C C C C	GTG GGC GGC GGC GGC GGC GGC CGGC CG GGC CG CCC CCG CCC CCC CCC CCC CCC CCC CCC CCCC CCCC CCCC	GCC A CCC CCC P P AAG CCG GCG GCG GCG GCG CCG CCG CCG CCG	GAG E GAG E TCG S S GGC C CGA CAG Q AAG GCC A AAG GCC A AAG GCC A TTT F F ATC I I TCG S GCG G G G G G G C G C C C C C C C C	GCG A CCG F P AAG K treecc G G G G G G G G CAA Q ATC I CCG G G G G G G G G G G G G G G G C CAA G M C AAG C H P P AAG C S G C C G G C G G C G G C G G C G G C C G G C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C AAG C G C	CCG P GTG V CCCC P CCG G GCG A AAG AAG CTG GCG CTG CCC CCC CCCC CCCC P CCCCC P CCCCC P CCCCC P CCCCC P CCCCC P CCCCC P CCCCC P CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCCCC	AAA K GCC CGT R CC3T R R CC3G2 M GGC C A CGC R M GGC C A CGC R R TAT GGC CGT R CGT C V CGT C CGT C CGT TGC CC3T CGT CGT CGT CGT CGT CGT CGT CGT CGT CG	CCG P GCG A CCGC R R R CCGC R A TCG G G G G G G G G G G G G G	GCG A GGC GGC GGC GGC GGC GGC CGC R R GGA CGC CGC CGC C CGC C CCC C CCC C CCC C CCC C CCC C CCC C	CCG P GGT TGG CTG CGG CGG CGG CGG CCC CCC CCC	AAA K GAA GAA GAA COS CAC H CGC R CAC H CGC CAC H GAA CAC CAC H GAA CAC CAC CAC CAC CAC CAC CAC CAC CAC	GCC GCC Cagaac Cagaac GA C Cagaac GAT D D TGG G TTG G G TTT TGC C C TTTC C C TTTC C C TTTC C C TTTC C G G TTTC C C C	GCCC A CCG P CCCG P CCCG C CCG C CCG C CCG C CCG C CCG C CCG C CCG C CCG C CCG C CCG C CCG C C CCG C CCG C CCG C CCG C CCG C	GAG E E GGCT GGCT GGC GGC GGC GGC GGC GGC GGC G	CCCG P ggcgc ggcgc R AGGG GGT GGT GGA ATC I GAA E GAA E GAA E GAA E GAA E GAA C S ACG C T ATC GAA E TCG GAA S S TTCG GSA C T C GCA C S C C C C C C C C C C C C C C C C C	GTC V GGG A GC GGG A A GC G G G G C G A A A A	AAG GCG GCG GCG GCG F GCG GCG GCG	GCC A AAG K aggc ttgta AACC T TTT F F GCTG G GAT C C GG C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C C G G C C C C G C C C C C G C	GCG A CCCC P gctg gctg CCG P CTG CCCG CCCG CCCG CCCC CCCC CCCC	CCC P GCC A acagc GCC G C G G C C G G C C A C A C C A C A	GTG V CCGG P ggca TAC Y AAT TAC Y AAT GGG GGC GGC GGC GGC GGC GGC CAG GGC GGC	GCG A GCG GCG GCG GCT CCG GCC CCC CCC CCC GCC CCTG CCTG CCTG CCTG GCC CCTG GCC CCTG CCTG CCTG CCTG CCTG CCTG CCC CCTG CCC CCTG CCC CCC	GCG A CCCC P CCCCC P CCCCC P CCCCC R A ATG ATG CGC CGC CGC CAC H CCGC CAC H CCGC CAC H CCCCCC CAC H CCCC CCCC	CCG P GCGG A GGGCCAC H H TGG W W AACC T TGG C A CAC D CAC C A AACC C C C C C C C C	1694 112 112 112 112 112 115 115 115 115 115
16200 88 16955 113 1770 1388 1859 1959 1 22043 2118 42 2193 67 22181 42 23433 17 2418 142 2493 167 2418 142 2643 217 2718 2643 217 2718 2643 217 217 217 218 2643 217 217 218 2643 217 217 218 2643 217 217 218 2643 217 217 218 2643 217 217 218 218 22493 267 217 218 217 217 218 218 219 219 219 219 219 219 219 219 219 219	GTG AAA V K GTC AAG V K CCG AAA P K gcctgcc CG AAA P K gcctgcc CG AA TGG ATC C T GGA CGC V I TCC TC S F TGC AAG C K GC GGG GC G A TGC AAG GC CGC GCG G A ATG AAA M K GTC GTG V V CTG TTC L F ATG GAA I E CCG GTC CG C K I CCG GT CCG A GT CGT CTC L F ATG AAA CCG CGC CG C A GTC CGC CC C C C A GTC CGC CC C C C C C C C C C C C C C C C C	GTG GGC GGC GGC GGC GGC GGC ATC D CGG C C GGC C C G GGC C C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G C C G C C C G C	GCC A CCC P P AAG CTG CTG CTG CTG CCG CCG CCG CCG CCG CC	GAG E GAG E TCG S S GGC C CGA CAG Q AAAG CAG Q CAAAG CAG CCAG C	GCG A CCG CCG P AAG G K CCG G G C C C A A C C A T C T A T C T A T C T A T C T A T C T A G G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C G G C C G C G C C G C C G C C G C C G C C C G C	CCG GTG GTG V CCCC Pagaag GA GCG A AAG C C C C C C C C C C C C C	AAA K GCC CGT R CCGT R R CCGG ATG GCC ATG GCC CGC CGC CGC CGC CGC CGC CCGC C	CCG PGCG A CGCC R R HCCCC GGCG GGCG GGCG GGCG GGCG	GCGG A GGTG GGTG GGTG GGTG GGTG CGT CGT R R GGG GGT GGC GGC GGC GGC GGC GGC CGC GGC CGC C	CCG P GCT TGJ C CGC GGT C CCC C CCC C CCC C CCC C CCC C CCC C C	AAA K GAA A GAA C GAC N CAC CAC R TCG CAC R TCG CAC CAC CAC CAC CAC CAC CAC CAC CAC C	GCC GCC Cagaa Cagaa GA GCG Cagaa GAT D TGG G TT TGG G G TTT TGC C C TT TAT Y TGC C C TT TTC C C TT TTC C C TT C G GC GAG G GAC D GAC GAC CAGA	GCCC A CCG P P CCCC CCG C C C C C C C C C C C	GAG E   AAA K gcccgg   gcccgg gGAT    gGAT GGC GGC   GGC GGC GGC   GGC GGC GGC   GCCGC GGAA E   GGAA E CCGG   GGC GGC GGC   GGA E CCGG   GGA E CCGG   GGA E CTG   JATC I ATC   TCG S CTC	CCGC P GCGC P GGC CGC R CGC T GGT G G ATC G G ATC G G G ATC C G C G C G C C C C C C C C C C C C C	GTC V GCG A GCG GCG A AGC G GCG A A AAC C T L AAG G G G G G G G G G G G G G G G G G	AAG GCG GCG GCG GCG F GCG GCG GCG	GCC A AAG K AAG T T T T T T T T T T T T T T T G GAT D C C G G C C G G C C G G C C G G C C G C C G C C G C C G C C G C C G C	GCG A CCCC P gctg gctg Gatg CTT L CCG G A ACC T L CCG G A ATC CTG L CCG G A ATC CTG L CCG G CT C CTG C C C C C C C C C C C C	CCC P GCC A acage G G G G G G C C T T C C S G C C A C C A C Q G C C T T C C S G C C A A C Q G C C T T C C S C C C A C C C C C C C C C C C C C	GTG V CCGG P ggca TAC Y AAT N N GGG G GGG GGG GGG GGG GGC A ATG T ATG GGC G GAG GCC T T T T T T C GGG G G G G G G G G	GCG A GCG GCG GCG GGG GGC CTC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCC CCG CCCC	GCG A CCCC P CCCCC P CCCCC P CCCCC A ATG ATG CGC CGC CGC CGC CGC CGC CGC CGC CGC C	CCG P GCGG A ggggc CAC D D TGG W A A TG G A C C C A C C C C C C C C C C C C C	1694 112 112 112 1135 1135 148 12042 16 2042 16 22117 2342 2192 2247 2192 2247 2116 22417 1216 2247 2166 2717 2792 2264 22567 291 2116 2257 291 216 2264 2257 291 216 2264 2257 291 217 2792 2264 2257 291 217 2792 2264 2364 2364 2364 2364 2364 2364 236
1620 16920 113 113 113 1138 1138 1959 1043 105 107 2118 42 2193 67 2218 42 2268 2268 2268 2268 22718 2418 2492 24193 167 24188 242 2493 2568 292 2433 217 2458 242 2493 247 2468 247 247 248 2493 267 2718 242 2493 247 2468 247 247 247 2468 247 247 247 247 247 247 247 247	GTG AAA V K GTC AAG V K CCG AAA P K gcctgcc Gac CGC D R TGG ATC TGG ATC V I TCC TTC S F TGC AAG C K GC GGG G A TGC ACG C C GGA TGC ACG GGA TGC GGT GCG GGA GTC GTG V CTG TTC L F GTC GTG V C GGT GTG C K H GTC GAA I E CCG GTC CG P V GGC ACC C ACC C CG CC C CG CG C CG CG C CG C CG CG C CC C CG C CC C CG C CC C CG C C C C CG C CC C CG C C C C CG C C C C C	GTG GTG G GGC G GGC G GGC S A A TCG C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C G C C C G C C C G C C C C G C C C G C C C C G C C C G C	GCC A CCC P P AAG K ttcgQC CTG L L GAG GE CCG CCG CCG CCG CCG CCG CCG CCG CCG	GAG E GAG E TCG S gate ccga CaG CaG CaG CaG CaA CAG GAA CAG GAA CAG GAA TTTC F F CG GGC CA CTG GCC CCTG GCC CCTG GCC CAG GAC CAG CAG CAG CAG CAG CAG CA	GCG A CCG CCG P AAG K CCG G G CC C CCA A C CAA C CAA C CCG C CCG G CCTG G G G CCTG G G G CCTG G G G	CCG GTG GTG V CCCC Pggaag GCG A AAG C S ATG C C C C C C C C C C C C C C C C C C C	AAA K GCC CGT CGT CCaggaat ATG ATG AC GCC C C C C C C C C C C C C C C C C	CCGC P GCG A · CGCC R Licccg ggacJ <b>DF</b> GCG A CAC CAC C A CAC C A CAC C A CAC C A CAC C A CAC C C C C C C C C C C C C C C C C C C C	GCGG GGTG GGTG GGTG GGTG GGTG GGTG CGT R R GGG GGT GGC CGC CGC CGC CGC CGC CGC CGC	CCG P GCT TG GCT C CCG G GCTG G GCTG CCG G CTG CCG CC	AAA K GAA A GAA C GAC N CAC C CAC C CAC C CAC C CAC C C C C	$\begin{array}{c} {\rm GCC} \\ {\rm GCG} \\ {\rm a} \\ {\rm ccagac} \\ {\rm ccagac} \\ {\rm cd} \\ {\rm cd} \\ {\rm d} \\$	GCC A CCG P P CCCC CCG CCG CCG CTG CTG CTG CTG CTG CT	GAGG E AAAA K gcccg GAT D GGC G GGC G GGC G GGC C G GGC C G GGC C C G C C G C	CCGG P GGC GGC ACGG CGC CGC T T GGT ACG G ACG G ACG G ACG G ACG C G ACG C G ACG C CGC T T GAA E AAG G C C C G C A C G C G C C C C C C C	GTC V GCG A GCG A T C T C GCG G C G C G C G C G C G C G C	AAG GCG GCG GCG GCG TTC F GCG GCC GCC GCC GCC GCC GCC GCC A GCG GCC A GCG GCC CCG GCC CCG GCC CCC GCC G	GCC A AAG K AAG T T T T T T T T T T T T T T T T G A T C C T G G C T G G C T G G C T G G C T G G C T G G C T G G C T G G C T G C T C T	GCG A CCCC P gctg gctg CCTT L CCG GAA ACC T CCTG CCTG CCTG CCTG CCTG	CCC P GCC A acgc G G G G G G G C C G G C C C G C C C G G C C C G G C C C G C C C G C	GTG V CCCG P ggcaa CCGB G GTTAC Y AAT N GGG GGA GGA GGC A AT Y GGG GGA CCA G GCC GAA CC GCA G GCC GAA CC CCG CCG	GCG A GCG Jaaa GGG GGC GGC CTC CCG GGC CTC F TTCC GGC AAG AAG GGC CTG GGC AAG AAG CCTG GGC AAG CCTG GGC ATG GGC GGC GGC GGC GGC GGC GGC GGC GGC G	GCG A CCC P CCC CCC P CCC CCC R A ATG M CCC CCC R A ATG CCC CCC CCC CCC CCC CCC CCC CCC CCC C	CCG P GCG CAC GAC GAC D TGG W AACC T TGG W AACC T TGG AACC C C C C C C C C C C C C C C C C	1694 112 1769 137 18588 2042 2042 2166 2217 41 22267 91 2241 2241 2241 2241 2241 2241 2241 2
$\begin{array}{c} 1620\\ 88\\ 88\\ 113\\ 113\\ 113\\ 113\\ 1859\\ 1959\\ 2043\\ 17\\ 22118\\ 42\\ 2113\\ 17\\ 22118\\ 42\\ 2234\\ 17\\ 22418\\ 22268\\ 92\\ 2343\\ 17\\ 2568\\ 192\\ 2268\\ 192\\ 2268\\ 192\\ 2268\\ 2263\\ 217\\ 2568\\ 22643\\ 217\\ 2868\\ 292\\ 2943\\ 317\\ 3018\\ 367\\ 3016\\ 392\\ 3243\\ 3$	GTG AAA V K GTC AAG V C GCC AAA P K GCC CCC AAA P K GCC CCC A TGG ATC TGG ATC GAC CCC W I TGG ATC GC K GCC GCG GA GAA TGC E C GGT GCG GA ATC GAA TGC ACC V V GCC ACC GT ACC ACC ACC V TGC TGC C ACC ATC ACC V TGC TGC C ACC ATC ACC V TGC TGC C ACC C ACC ATC ACC V TGC TGC C ACC C ACC ATC ACC C C ACC C C C C	GTG GGC GGC GGC GGC GGC GGC GGC AATC D GGC CCG G GGC G GCCG A AGAC CCG G GCCG A AGAC CCG G GCCG G GCCG CCG	GCC A CCC CCC P AAG CCC CCC CTG L GAG CCC CCC CCC CCC CCC CCC CCC CCC CCC	GAG E GAG E C GAG C C C G C C C C C C C C C C C C C	GCG A CCG F AAG G CCG F AAG G G G C C C C C C C C C C C C C C C	CCG GTG GTG V CCC P Ggagg GGG A A A CCC P GGG A A CTG CTG C CTG C CCG G G GCG A CTGC C CCC C CCC P GCC A A CCC C CCC P G A A G C C C C C C C C C C C C C C C C	AAA K GCC CGT CGT CCage GCC CCage GCC CC C C C C C C C C C C C C C C C C	CCG GCG A CCGC R Lcccg ggac DF GCG GCG CAC CAC G GCG CAC CAC CAC CAC C	GCG A GTG GGC G GGC G G G G G G G G G G G G G	CCG P GCT I C C C C C C C C C C C C C C C C C C	AAA K K GAA E A GOO R CAC CAC CAC CAC CAC CAC CAC CAC CAC C	$\begin{array}{l} {\rm GCC}\\ {\rm GCC}\\ {\rm A}\\ {\rm GCG}\\ {\rm A}\\ {\rm Cagaac}\\ {\rm Cagaac}\\ {\rm TGG}\\ {\rm F}\\ {\rm TGG}\\ {\rm F}\\ {\rm TTT}\\ {\rm CC}\\ {\rm CTT}\\ {\rm TTT}\\ {\rm CC}\\ {\rm CTT}\\ {\rm CC}\\ {\rm CTC}\\ {\rm CT}\\ {\rm CC}\\ {\rm CTG}\\ {\rm C}\\ {\rm CTG}\\ {\rm CC}\\ {\rm C$	GCC A CCG P CCCC CAC H D CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCC C CCC C CCC C CCC C CCC C CCC C	GAG E AAA K gcccg GAT D GGC G GGC G GGC G GGC C G GGC C G GGC C C G G C C G C C G C C G C C G C C C G C	CCGG P CCGG P gcgcC CCC R ACG G ACG G ACG G ACG G ACG G ACG G ATG G ACG C G ATG S ACG C C C C C C C C C C C C C C C C C C	GTC V GCG A GCG A C C T GCG G C G C G C G C G C G C G C C T L C C T L C C T L C C T L C C T L C C T C C T C C T C C C G G G G G C G C	AAG GCG GCG GCG GCG CCG CCG GCG GCG CCG GCG GCG CCG GCG A GCG GCG	GCC A AAG AAG AAG K K aggc T T T T T T T T T T T T T T T T T T	GCG A CCC P getg getg AAC CTT L CCT CCTG CCG CCG CCG CCG CCG CCG CCG CC	CCC P GCC A acgci G GCC G A A C T T T C C G G C C A G C C A G C C A G C C A G C C A C A	GTG VCCG P ggcai CTAC VY AAT GGG GGG GGG GGC GGG GGC CAG GGG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG GGC CAG CAG	GCG A GCG Jaaa Ugec GGG G G C C C C C C C C C C C C C C C	GCG A CCC P CCC CCC P CCAC R ATG ATG GCC CGC R K GGC CGC A AG GCC CGC A ATG ATG GCC CGC CG C A ATG ATG ATG ATG ATG CGC C CGC C C C C C C C C C C C C C C	CCG P GCG CCAC CAC D GACC D GACC D T TGG W GACC D CAC C CAC C CAC C C C C C C C C C	1694 112 112 137 138588 148588 2042 2042 266 2117 4192 2042 2166 22417 241 2242 2166 22417 2642 2267 2642 22642 2142 2664 22942 2342 2166 2912 2662 2912 2672 2912 2673 2912 2663 2912 2673 2912 2664 2914 2664 2914 2664 2914 2665 2674 2674 2674 2674 2674 2674 2674 2674

Fig. 1. Sequence of the *ClaI-MscI* restriction fragment of lambda clones, selected from an *R. capsulatus* genomic library by screening with a cDNA coding for the bovine 51 kDa subunit as a probe. The derived amino acid sequences of the three open reading frames *nuoE*, *orfA* and *nuoF* are indicated below the DNA sequence. Termination codons are marked by an asterisk. The nucleotide sequence data reported in this figure will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession number Y10142.

onto the column and allowed to recirculate through the column overnight at 4°C. The column was then washed extensively in buffer A. Finally, the immuno-adsorbed components were released by an acidic washing step with a solution composed of 100 mM HCl-glycine pH 2.8, 0.4% Glucopon. Collected fractions were immediately neutralized and analyzed on SDS-PAGE. Proteins were sequenced after electrophoretic transfer onto PVDF membranes as described [14].

The identification of the peptides generated by tryptic cleavage of NUOE was performed by mass spectrometry as follows. The silver stained NUOE spot was excised from the polyacrylamide gel and ingel digested with trypsin after iodoacetamide derivatization according to published procedures [15]. The resultant peptide mixture was embedded into a matrix of  $\alpha$ -cyano-4-hydroxy-*trans*-cinnamic acid and analyzed by MALDI mass spectrometer. The mass spectra were obtained on a Bruker BIFLEX mass spectrometer. For spectrum calibration, trypsin autolytic fragments were taken as standards.

### 3. Results and discussion

Lambda clones, containing a DNA fragment encompassing the R. capsulatus gene coding for the 51 kDa catalytic subunit of the respiratory complex I, were recovered by screening and R. capsulatus DNA genomic library using a cDNA clone coding for the bovine mitochondrial 51 kDa subunit as a probe. Restriction digest fragments of the positive clones were cloned and sequenced. A ClaI-MscI fragment 3.3 kb long (Fig. 1) was proved to contain 3 putative open reading frames. The first open reading frame from base 90 to base 1259 encodes a 41 kDa protein which is related to the bovine 24 kDa subunit (Fig. 2). As a consequence this gene was called *nuoE* following the nomenclature of the E. coli nuo genes [10]. The second open reading frame from base 1359 to 1802 was called orfA and encodes a potential polypeptide with a molecular mass of 14.8 kDa. This putatively expressed polypeptide has no equivalent with either of the known subunits of the mitochondrial or bacterial complexes I. The last open reading frame from base 1995 to base 3290 encodes a 47 kDa protein which is equivalent to the bovine 51 kDa subunit. This gene was called nuoF. The overall organization of this R. capsulatus DNA fragment with the nuoE gene located upstream of the nuoF gene is similar to what is found in the E. coli nuo operon, except that in R. capsulatus a supernumerary open reading frame (orfA) is located between nuoE and nuoF. In the *P*. *denitrificans* operon, Yagi's group has discovered two open reading frames called *urf1* and *urf2* located between the *nqo2* and *nqo1* genes (*nqo2* and *nqo1* genes are the equivalents of the *R. capsulatus nuoE* and *nuoF* genes respectively) [9]. Neither of these two *urfs* looks similar to the *R. capsulatus orfA*. Analysis of our lambda clones showed that in *R. capsulatus the nuoE* and *nuoF* genes are located in the close vicinity and upstream of the *nuoH-nuoN* gene cluster previously sequenced (not shown) [11,16].

A closer look at the primary structures of the NUOE and NUOF proteins has given additional information. The R. capsulatus NUOF protein is 38.5% identical to the E. coli NUOF subunit, about 67% identical to the bovine 51 kDa subunit and up to 85% identical to the P. denitrificans NQO1 subunit. Four cysteine residues which are believed to be involved in the binding of a 4Fe-4S cluster (C346, C349, C352 and C392, R. capsulatus numbering) are strictly conserved in all the subunits. Four cysteine residues are also invariant in the NUOE equivalent subunits (Fig. 2). All the NUOE equivalent subunits are highly similar except for their C terminal parts which are heterogeneous in size (Fig. 2). The E. coli NUOE subunit is the shortest one with 166 amino acids and a molecular mass of about 19 kDa. The human and bovine subunits, as mitochondrially imported and processed forms, are 217 amino acids and close to 24 kDa, and the P. denitrificans with 239 amino acids is 26 kDa [9,10,17]. According to the DNA sequence, the R. capsulatus NUOE subunit would have 389 amino acids and a molecular mass of 41 264 Da. This subunit looks very atypical and is nearly twice as big as the corresponding mammalian mitochondrial subunit. In comparison to the P. denitrificans subunit, the R. capsulatus NUOE subunit appeared to be extended at the C terminus by 150 amino acids. This predicted extension is alanine rich, without specific motifs, and contains only one additional cysteine residue. It is not highly hydrophobic and is not predicted to form a transmembranous helix. To rule out the possibility of any DNA sequencing artifact, we paid careful attention to accurate sequencing of various subcloned fragments of the *nuoE* gene as well as a number of PCR fragments



Fig. 2. Comparison of the polypeptide sequences of the *R. capsulatus* NUOE subunit, the *E. coli* NUOE subunit, the bovine and human 24 kDa subunits (as processed mitochondrially imported subunits) and the *P. denitrificans* NQO2 subunit. Identical residues between the *R. capsulatus* NUOE and the human 24 kDa subunits are marked by dots, invariant cysteine residues are indicated by open circles. The tryptic peptides of the *R. capsulatus* NUOE subunit identified by MALDI mass spectrometry (see Fig. 4) are underlined and labelled A to I and A' to C'. Alignment was computerized using the Clustal method with the PAM 250 weight table.

348



Fig. 3. SDS polyacrylamide gel and Western blot analysis of the fractions eluted from the anti-PNE affinity column. Proteins recovered from the affinity column by acidic elution were precipitated overnight with 5 volumes of ethanol at  $-20^{\circ}$ C (lanes 1, 2, 3 and 5). Membrane lysates collected in the flow through of the column during sample loading were also run as controls after being ethanol precipitated (lanes 4 and 6). Samples were run through a 12% polyacrylamide gel then stained with Coomassie (lane 1) or silver (lane 2). After electrophoretic transfer to PVDF membranes, proteins were identified using an anti-NUOF antibody (lanes 3 and 4) or an anti-PNE antibody (lanes 5 and 6).

overlapping this gene. No error was found in our consensus sequence. Accordingly, we set up immunoaffinity experiments to check the size of the expressed subunit in the *R. capsulatus* membranes so as to definitively ascertain the existence of an elongated C terminus in the *R. capsulatus* NUOE subunit.

Antibodies were raised against a synthetic peptide (called PNE) corresponding to the deduced N terminus sequence of the NUOE subunit. Membrane samples from the R. capsulatus B10 strain were analyzed by 2D electrophoresis in polyacrylamide gel. The anti-PNE antibodies specifically revealed one polypeptide with an apparent molecular mass of 42 kDa and a pI value close to 5.5 (not shown). These features are very close to the theoretical values of 41 264 Da and pJ 5.42 deduced from the sequence of the gene. The selectivity of our antibody allowed us to set up experiments with the aim to immuno-purify the whole complex I or some fraction of it from R. capsulatus membrane fractions. Membranes were diluted in a 2% Glucopon solution (see Section 2) and soluble material was recovered after centrifugation. This soluble fraction was loaded onto a column of anti-PNE antibodies immobilized on a protein A-crosslinked Sepharose. After washing, the eluted fractions (see Section 2) were analyzed on SDS-PAGE. These fractions contained 2 major proteins in a grossly stoichiometric amount as revealed by Coomassie blue staining (Fig. 3, lane 1). This subcomplex fraction appeared at least 95% pure as shown by Coomassie and intense silver staining (Fig. 3, lanes 1 and 2). Apparent molecular masses of 49 and 43 kDa were estimated for these two proteins. The two proteins were specifically detected by an anti-NUOF subunit antibody or by the anti-PNE antibody respectively (Fig. 3). The identity of these two proteins was further confirmed by sequencing of their N termini. The N terminus of the biggest subunit was shown to be MLNDHDRIFT and that of the smaller one MLRRLHATQP. These N termini perfectly matched those deduced from the sequenced genes coding for the NUOF and NUOE subunits respectively (Fig. 1). This confirmed that the two newly sequenced nuoE and nuoF genes are truly expressed in *R. capsulatus* and that the NUOE and NUOF subunits, purified as a soluble heterodimer after detergent extraction of the membranes, are in close contact in complex I. We also checked on an aliquot of the particulate material from the column if the NUOE-NUOF subcomplex might have some catalytic activity. Indeed, some deamino-NADH dependent nitroblue tetrazolium reductase activity was present in the immobilized protein fraction on the column after the extensive washes.

To ascertain that the mass of the NUOE subunit was related to a C terminus extension in comparison to other species NUOE subunits, we analyzed by MALDI mass spectrometry the peptide mixture obtained after tryptic digestion of the NUOE spot in the polyacrylamide gel. Eleven different peptides generated by trypsin cleavage of the NUOE subunit were unambiguously identified (Fig. 4). Among these peptides, two (peptide H from amino acid 288 to 306 and peptide I from amino acid 371 to 379) are part of the C terminus extension which has no counterpart in the equivalent bacterial or mitochondrial subunits (Figs. 2 and 4). This analysis ascertains that the amino acid sequence deduced from the DNA sequence of the *nuoE* gene is accurate and that the *R. capsulatus* NUOE subunit is significantly bigger than the other equivalent subunits known so far.

In chloroplasts, three subunits of the NADH dehydrogenase domain of complex I (75 kDa, 51 kDa and 24 kDa) are apparently missing. This observation was taken by Friedrich et al. as an indication that in chloroplasts 'complex I' may work in a different way to the mitochondrial complex I [18]. Chloroplast complex I was hypothesized by these authors to be a NADPH or a ferredoxin:plastoquinone oxidoreductase. The discovery that in *R. capsulatus* NUOE and NUOF, which are the equivalents of the 24 kDa and 51 kDa subunits, are expressed is a clear indication that the *R. capsulatus* complex I is more of the mitochondrial type than of the chloroplast type.

Friedrich et al. have shown that stripping the E. coli cytoplasmic membranes under acidic conditions with a buffer containing the technical alkylglucoside APG 225 (a Glucopon equivalent manufactured by Henkel, Düsseldorf, Germany) led to the isolation of a whole active complex I [19]. Unlike this work, many different attempts at solubilizing the whole R. capsulatus complex I in our laboratory or others [20] have thus far proved unsuccessful probably because of the great instability of this complex I. The experiments we report here clearly show that in the case of R. capsulatus, Glucopon treatment of the membranes led to the isolation of a subcomplex composed of the NUOE and NUOF subunits only. This subcomplex was stripped from the membrane whatever the pH of the stripping solution in the range of 6-8. Indeed, our purification protocol yielded a subcomplex of complex I very similar to the mitochondrial FP fragment. The mitochondrial FP fragment is composed of 3 different subunits, 51 kDa, 24 kDa and 9 kDa. An equivalent of the 9 kDa subunit, the role of which is unclear in the mitochondrial complex I, is absent in our preparation. None of the genes or other ORFs of the bacterial NADH-ubiquinone oxidoreductase operons analyzed so far encode a protein similar to the mammalian 9 kDa ([9,10] and unpublished data for R. capsulatus). Furthermore, no significant homology was observed between the C terminal extension of R. capsulatus NUOE subunit and the bovine 9 kDa. Consequently, the possibility for the R. capsu-



Fig. 4. MALDI mass spectrum of a tryptic digest of the NUOE subunit. After in-gel tryptic digestion of the NUOE band, the eluted peptide mixture was analyzed as described in Section 2. The identified tryptic peptides are labelled A to I, and A' and C', their positions and sequences are given in the table. Peptides A', C' and G correspond to partially digested peptides. The peptide at m/z 1045.60 is a trypsin autolytic fragment which partially overlaps the signal of peptide I. These two peaks are clearly distinguishable when this doublet peak is magnified (not shown). The major peak of the spectrum was unidentified. The theoretical masses of the protonated peptides are given in the table. They correspond to the monoisotopic masses for peptides G, E and I and to average masses for peptides A–D, C', F and H. Positions of the peptides A to I, A' and C' are underlined in Fig. 2. (a) Mass of the Cys-195 carboxyamidomethylated peptide F.

latus NUOE subunit to behave as a fusion protein that would be the equivalent of both the bovine 24 kDa and 9 kDa seems to be ruled out. In the case of P. denitrificans attempts to purify complex I after detergent treatment of the membranes have led to the partial purification of a heterodimer with NADH-ubiquinone oxidoreductase activity [21]. The two subunits had molecular masses of 48 and 25 kDa. An antibody raised against this heterodimer was proved to crossreact with the bovine 51 kDa subunit of the bovine mitochondrial complex I suggesting that the two-subunit NADH-ubiquinone oxidoreductase component purified from the P. denitrificans membranes might also be an FP-like subcomplex I [22]. Furthermore, it has been recently reported by Yagi's group that the P. denitrificans NOO1 and NOO2 subunits tend to associate in a dimeric aggregate when co-expressed in E. coli [23]. The above data as well as our immuno-purification of an FPlike fragment from the R. capsulatus membrane are clear indications that the NUOE and NUOF subunits interact closely and strongly in the bacterial complex I.

The *R. capsulatus* NUOE subunit has a very peculiar feature because of a very long C terminal alanine rich extension which is not found in the equivalent subunits from other species. Because this extension is predicted not to contain any hydrophobic helix it is probably not involved in the membrane anchorage of complex I. At the present time, we do not know if this extended C terminus is related to a specific structural component in the *R. capsulatus* complex I or whether it is associated with the fact that *R. capsulatus* is a photosynthetic bacterium. Mutagenesis studies are under way to check the role, if any, of this C terminal extension.

Acknowledgements: We thank J.E. Walker for the gift of bovine cDNA; T. Rabilloud, M. Chevallet and V. Procaccio for their assistance in immunochemistry; M. Louwagie for amino acid sequence analysis; J. Garin for helpful discussion and P.J. Harms for reading the manuscript. Glucopon was a generous gift from Henkel, France. This work was supported by grants from the EU Biotech program (BIO2-CT93-0364), from the EPR Rhône Alpes; from INSERM (CRE 930805) and from the Commissariat à l'Energie Atomique.

#### References

- [1] I.M. Fearnley, J.E. Walker, Biochim. Biophys. Acta 1140 (1992) 105-134.
- [2] T. Yagi, Biochim. Biophys. Acta 1141 (1993) 1-17.
- [3] Y. Hatefi, Annu. Rev. Biochem. 54 (1985) 1015-1069.

- [5] Y.M. Galante, Y. Hatefi, Arch. Biochem. Biophys. 192 (1979) 559-568.
- [6] C.I. Ragan, Y.M. Galante, Y. Hatefi, T. Onishi, Biochemistry 21 (1982) 590–594.
- [7] S. Chen, R.J. Guillory, J. Biol. Chem. 256 (1981) 8318-8323.
- [8] T. Yagi, T.M. Dinh, Biochemistry 29 (1990) 5515-5520.
- [9] T. Yagi, T. Yano, A. Matsuno-Yagi, J. Bioenerg. Biomemb. 25 (1993) 339-345.
- [10] U. Weidner, S. Geier, A. Ptock, T. Friedrich, H. Leif, H. Weiss, J. Mol. Biol. 233 (1993) 109–122.
- [11] A. Dupuis, FEBS Lett. 301 (1992) 215–218.
- [12] P.F. Weaver, J.D. Wall, H. Gest, Arch. Microbiol. 105 (1975) 207–216.
- [13] A.M.R. Haines, J.M. Cooper, J.A. Morgan-Hughes, J.B. Clark, A.H.V. Schapira, Prot. Express. Purif. 3 (1992) 223–227.
- [14] C. Adessi, A. Chapel, M. Vincon, T. Rabilloud, G. Klein, M. Satre, J. Garin, J. Cell. Sci. 108 (1995) 3331–3337.

- [15] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (1996) 850–858.
- [16] A. Dupuis, A. Peinnequin, M. Chevallet, J. Lunardi, E. Darrouzet, B. Pierrard, V. Procaccio, J.-P. Issartel, Gene 167 (1995) 99– 104.
- [17] S.J. Pilkington, J.E. Walker, Biochemistry 28 (1989) 3257-3264.
- [18] T. Friedrich, K. Steinmüller, H. Weiss, FEBS Lett. 367 (1995) 107-111.
- [19] H. Leif, V.D. Sled, T. Onishi, H. Weiss, T. Friedrich, Eur. J. Biochem. 230 (1995) 538–548.
- [20] B.C. Berks, S.J. Ferguson, Biochem. Soc. Trans. 19 (1991) 581– 588.
- [21] C.L. George, S.J. Ferguson, Eur. J. Biochem. 143 (1984) 567-573.
- [22] C.L. George, S.J. Ferguson, M.W.J. Cleeter, C.I. Ragan, FEBS Lett. 198 (1986) 135–139.
- [23] T. Yano, V.D. Sled, T. Ohnishi, T. Yagi, J. Biol. Chem. 271 (1996) 5907–5913.