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^a, Alain Dupuis^{a,*}, Agnès Chapel^b, Silvie Kieffer^b, Joël Lunardi^a, Jean-Paul Issartel^a

^aLaboratoire de Bioénergétique Cellulaire et Pathologique, EA 2019 UJF, DBMS, CEA Grenoble, 38054 Grenoble Cedex 9, France ^bLaboratoire 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

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

Abbreviations: NUO, NADH-ubiquinone oxidoreductase; FP, flavo-protein fraction; MALDI, matrix-assisted laser desorption/ionization; PNE, synthetic peptide corresponding to the NUOE subunit N-terminus

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-

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 β -mercaptoethanol, 2 μ g/ml aprotinin, 1 μ g/ml antipain, 1 μ g/ml leupeptin, 1 μ g/ml chymostatin, 0.5 μ g/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

	atcgatgccgtcgccaagg	racaccaactggcggat	gtgtccgccatca	teggeacgatggatgtee	gtgtteggggagattgaeegetg	ATG CTG 95
1 96	CGC CGC TTG CAC GCC AC				NUOI CGC GCC TGG GCG GAA GC	M L 2
3	RRLHAT	Q P D S	F A F	T P A N	R A W A E A	Q I 27
171 28	ACC AAA TAC CCC GAG GO T K Y P E G	GC CGT CAG GCC TO R O A S	G GCG GTG ATO		CGC GCG CAG GAA CAG GA	AA GGC TGG 245 G W 52
246	CTG ACG AAG CCC GCG A	rc gaa Tac gtg go	CC GAC CTG CTC	GAC ATG CCC TAT	ATC CGC GTG CTC GAA G	rg gcg Tcc 320
53 321	L T K P A I	E Y V A AG CTG CAG CCC AC	D L L OG GGC TCG GTC	D M P Y	I R V L E V GTC TGC GGC ACC ACC AC	A S 77 SC TGC ATG 395
78	F Y F M F Q	L Q P T	g s V	A H I Q	V C G T T S	C M 102
396 103	ATC ATG GGC TCG GAA A	AC CTG ATC GAG GT	C TGC AAG CGC		CAT CCG CAT GCG CTT TC	CC GCC GAT 470 A D 127
471	GGC AAG TTC AGC TGG G	AA GAG GTC GAA TO	T CTG GGC GCC	TGT GCC AAT GCG	CCG ATG GCC CAG ATC GC	
128 546	G K F S W E	E V E C	L G A	C A N A ATC GAC GAT CTG	P M A Q I G GCG GCG GGC AAG GCG CC	K D 152 CG GTG CCG 620
153	Y F E D L T	E A G L	E R L AA CCG GCC ACG		A A G K A P TGC ACG ACG ACC TCC GC	V P 177
621 178	G G F G GAA ATC GGG C	GC TTC GGG GCG GA F G A E	P A T	G L T S	TGC ACG ACG ACC TCC GC	A R 202
696 203	GAG GCG CAG AAC GGC TO E A O N G S	CG GTC GCG CGG GC V A R A	CG GTT GCG GTG	GGC GAT GGC ACG	CGC CGG ATC GAC GGG ACR R I D G T	CC GAG GTG 770 E V 227
771		GG CAG GTC TCG GC			ATG ACG GCG CGG CCG CC	GC TCG GCG 845
228 846	R L S A P W GTG CAG CAG CAG GCG G	Q V S G CG GCC GAG GCG GC	Q M G CG AAA GCG GCG	A I G L G GCG GCA GAG	M T A R P R TCG GCC GGG GGG ACG TC	S A 252
253	V Q Q Q A A	AEAA	K A A	AAAE	S A G G T S	T P 277
921 278	GCT GCA CCG GTG GCC G	GG GCG GCG AA A A G K	AA GCC GCA CAZ A A O	A ACG GCT GAA ACC	GCT GCG CAG AAA CCG C'A A A O K P L	PT GCG CTG 995 A L 302
996	GTT GCA CCC CGC GAC G	GC AAG GCC GAT GA	AG CTG CAG ÃAC	ATC GAG AGC ATC	GGA CCG GCG CTC GAG G	
303 1071	V A P R D G CAC GAT CTG GGC ATC T	K A D E TC CAT TTC GAC CA	L Q K AG ATC GCC GCC	I E S I TGG GGC GCG GGC	G P A L E A GAA ATT GCC TGG ATG G	L L 327 AC GGC AAT 1145
328	H D L G I F	H F D Q	I A A	WGAG	E I A W M D	G N 352
1146 353	CTG AAG GGC TTT CCG G	GG CGC GCC AGT GC R A S G	D R W	GTG GTT CAG GCG V V Q A	AAG ATC ATC GTC GAG GA	AA GGG CTT 1220 G L 377
1221 378		GG GCG AAG ACA A	AC AAT TAC TO N N Y *	A agacggaatactgat	ctgaagacttgaacgacaacct	acagaggccgg 1307
1308	gacggcgtgacgtccgccaccd		aaggaaaggcaaa			G AAA GCC 1394
1 1395	CGC AAG CTG GGT GGT T	GG CTC GGA GCC G	Orfa G ATT TTC TT	M S T N	D T D C Q T	K A 12 FC TGG ATG 1469
13	R K L G G W	L G A A	I F F	ALWL	L G S V G I	W M 37
1470 38	GCG GCG ATC GCC GGG G A A I A G V	TG ATC TCT GCG CT	TG ATC CTG GCA	A CGG CTG ATC TGC R L I C	TGG GCG ACC TGC GGT C' W A T C G L	rg ccc gag 1544 P E 62
1545	CCG AAA GCG CCG ATG T	CG GCC GAG GAA GO	GC GCC GCG CTC	G CGT CGG GCC GTT	CCG GCC GCG GCG CCG A	CG GCG GCG 1619
63 1620	P K A P M S GTG AAA GTG GCC GAG G	A E E G CG CCG AAA CCG GO	A A L	R R A V	P A A A P T AAG GCC GCG CCC GTG G	A A 87
88	V K V A E A	PKPA	P K A	A E P V		
					K A A P V A	A P 112
1695 113						A P 112 CG CCC GCG 1769 P A 137
1695 113 1770	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A	CG GTG GCC GCG G V A A V AAG CCC CGT CGC (rg GCT GAA GCC A E A GGC TGA gacaga	CCG AAA CCG GCG P K P A	GCG AAG CCC GCC CCG G	CG CCC GCG 1769 P A 137 aaccatcccga 1858
1695 113 1770 138 1859	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S gccctgcggggctcgggatgt	CG GTG GCC GCG GC V A A V AAG CCC CGT CGC (K P R R cccggaccagatccgga	rg GCT GAA GCC A E A GGC TGA gacaga G * aagcgccgggctgc	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg	GCG AAG CCC GCC CCG G A K P A P A ggagagggggtgacggggaga cgtgtgtatcggacagccgatq	CG CCC GCG 1769 P A 137 aaccatcccga 1858 148 geegegeggge 1958
1695 113 1770 138	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S	CG GTG GCC GCG GT V A A V AAG CCC CGT CGC (K P R T CCCGgacCagatCcgga CCCgagggaatggacAT(rg GCT GAA GCC A E A GGC TGA gacaga G * aagegeegggetge G CTC AAC GAT	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg ggcctgacgagggagac CAC GAT CGC ATC	GCG AAG CCC GCC CCG GA K P A P A A A A B A B A B A B A B A B B A B	CG CCC GCG 1769 P A 13.7 aaccatcccga 1858 148 gccgcgcgggc 1958 GG ATG CAC 2042
1695 113 1770 138 1859 1959 1	GTC AAG ACG CCC GAG C V K T P E CCG AAA GGC AAG TCG A F K G K S Gccctgogggctcgggatgt cggacaccggccccgccgaaa GAC CGC TCG CTG AAG G	CG GTG GCC GCG GCV A A V AAG CCC CGT CGC (K P R R CCCGGGACCAGALCCGG2 ACCGAGGGGACAGALCCGGGGGGGACAT GC GCG ATG GCG CC	rg GCT GAA GCC A E A GGC TGA gacaga G * aagcgecgggetgc G CTC AAC GAT L N D GT GGT CAC TGC	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg CGC GAT CGC ATC H D R I G GAT GGC ACG GCG	GCG AAG CCC GCC CCG GA K P A P A GA GAGAGAGAGAGAGAGAGAGAGAGAGAGA	CG CCC GCG 1769 P A 137 aaaccatcccga 1858 148 gccgcgcggggc 1958 3G ATG CAC 2042 M H 16 GC CGC GAC 2117
1695 113 1770 138 1859 1959	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S gccctgcggggctcgggatgt cggacaccggccccccgcgaac GAC CGC TCG CTG AAG G D R S L K G	CG GTG GCC GCG GCV A A V V A A V V A A V V A A V V A GCC CGT CGC (K P R R CCCggaccagadacagadacagagagadacagagagadacagagagadacagagagadacagagagadacagagagaatagaacagagagaatagaacagagagaatagaatagaatagagagaata	rg GCT GAA GCC A E A GGC ** G ** Aagegeegggetge G CTC AAC GAT L N D GT GGT CAC TGC G H W	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg Gggcctgacgaggagac CAC GAT CGC ATC H D R I G GAT GGC ACG GCG D G T A	GCG AAG CCC GCC CCG GA K P A P A P A GO GOOD GOOD GOOD GOOD GOOD GOOD GOOD	CG CCC GCG 1779 P A 137 aaaccatcccga 1858 148 gccgcgcggggc 1958 G ATG CAC 2042 M H 16 G CGC GAC 2117 R D 41
1695 113 1770 138 1859 1959 1 2043 17 2118	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S gccctgggggctcgggatgt cggacaccggcccccccccgaac GAC CGC TCG CTG AAG G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V	CG GTG GCC GCG GTG AA V AAG CCC CGT CGC (K P R R CCCGGACCAGACCAGGACCAGGACCAGGACCAGGACCAGAGGACCAGAGGACCAGAGACAGAAAAAGACCC TCC GCAAAAAAGACCC TCC GCCAAAAAGACCC TCCG GCCAAAAAGACCC TCCG GCCAAAAAAGACCC TCCG GCCAAAAAAAGACCC TCCG GCCAAAAAAAAAA	rg GCT GAA GCC A E A GGC TGA gacaga G * aagegeegggetge G CTC AAC GAT L N D GT GGT CAC TGC G H W GG CTG CGC GGC L R G	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg cggcctgacgagggagac CAC GAT CGC ATC H D R I G GAT GGC ACG GCG D G T A C CGC GGC GCG R G G A	GCG AAG CCC GCC CCG GA K P A P A P A GO	CG CCC GCG 1769 P A 1378 aaccatcccga 1858 486 ATG CAC 2042 M H 16 GC CGC GAC 2117 R D 41 TC AA G TGG 2192 K W 66
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S gccctgggggctcgggatgt cggacaccggcccccccccgaac GAC CGC TCG CTG AAG G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V	CG GTG GCC GCG GTG AA V AAG CCC CGT CGC (K P R R CCCGGACCAGACCAGGACCAGGACCAGGACCAGGACCAGAGGACCAGAGGACCAGAGACAGAAAAAGACCC TCC GCAAAAAAGACCC TCC GCCAAAAAGACCC TCCG GCCAAAAAGACCC TCCG GCCAAAAAAGACCC TCCG GCCAAAAAAAGACCC TCCG GCCAAAAAAAAAA	TG GCT GAA GCC A E A GGC TGA gacaga G * aagcgccgggctgc G CTC AAC GAT L N D GT GGT CAC TGC G H W GG CTG CGC GGC	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg cggcctgacgaggagac CAC GAT CGC ATC H D R I G GAT GGC ACG D G T A C CGC GGC GGC GGG R G G A	GCG AAG CCC GCC CCG GA K P A P A P A GO	CG CCC GCG 1769 P A 1378 aaaccatcccga 1858 GC ATG CAC 2042 M H 16 GC CGC GAC 2117 R D 41 TC AAG TGG 2192
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S gccctgoggggctcgggatgt cggacaccggccccgccgaae GAC CGC TCG CTG AAG G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C S P M P C TGC AAG GAC CAG A	CG GTG GCC GCG GTG V A A V AGG CCC CGT CGC (K P R R CCcggagcaatcggatcggaatggaatggaatggaatgga	rG GCT GAA GCC A E A 3GC TGA gacagg G * * GAC GAT GGT CAC GAT L N D ET GGT CAC TGC G H W GG CTC GCG GGC L R G GC CCC TCG TTC AT CCC CAT ACC	G CCG AAA CCG GCG P K P A accccgcccggcgcctgg cggcctgacgagggagac CAC GAT CGC ATC H D R I G GAT GGC ACG GCG D G T A C CGC GGC GCG GCG R G G AC I CTG GTG ATC AAC L V I N G CTG ATC GAA GGC	GCG AAG CCC GCC CCG GA K P A P A P A P A P A P A P A P A P A	CG CCC GCG 1769 P A 1378 aaccatcccga 1858 486 ATG CAC 2042 M H 16 GC CGC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 CG GGC ACC 2267 G T 91 TC GCC ATG 2342
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67	GTC AAG ACG CCC GAG C	CG CTG GCC CGG GCC AA V AAG CCC CGT CGC CCC K P R R CCCGGGCGAGACCGGGACCGGGACCGGAGCAGCCGGGACCGGGGC A MOOF M GC GCG ATG GCG CC TC AAG GCC TCG GC K A S G A TCC GAC GGG CC ATC GAC AGG CC TC ATG CGC ACC GCC M R H D	rg GCT GAA GCC A E A GGC TGA gacagg G * aggcgcegggetgg G CTC AAC GAT L N D GT GGT CAC TGG G H W GG CTG CGC GG L R G GC CCC TCG TTC P S F AT CCG CAT ACC P H T	G CCG AAA CCG GCG P R P A ACCCGGCCCGGGCCCGGCCCGGCCCGGCCCGGCCC	GCG AAG CCC CCG GC A K P A P A R P A	CG CCC GCG 1769 P A 13769
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117	GTC AAG ACG CCC GAG CV K T P E CCG AAA GGC AAG TCG F K G K S CCC tGAG CG AAG CGC GAG CGC GAG CGC GGC G	CG CTG GCC GCG GCG V A A V VAGC CCC CGT CGC CK F R R CCCGGGACCGGGACCGGGACCGGGACCGGGACCGGGACCGGGACCGGGACCGGACGGACGGACGGACGGACGAC	TG GCT GAA GCC A E A GGC TGA gacagi G TG ACC GGC G H GG CTC GGC G CCC TCG F H GG CCC TCG F H GG CGC GAT GG F H GG GG GAA TA: G E Y	G CCG AAA CCG GCG P R P A ACCCGGCCTGGCGCTGGCGCTGGCGCTGGCGCGCGCGCGC	GCG AAG CCC CCG GC A K P A P A P A GCG AGG CC CGG CG CG CGG CG CGG CG CGG CG	CG CC GCG 1779 P A 137 137 148 148 148 148 148 148 148 148 148 148
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A P K G K S gccctgggggccccgcgaae GAC CGC TCG CTG AAG G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C S P M P K G CGC GAA A C K D R E I GGC GGG GGG GCC T G A R A A Y GAA TGC TAC GAG GAG GCC T G A R A A Y GAA TGC TAC GAG GAG GCC T G A R A A Y GAA TGC TAC GAG GGG GGC T G A R A A Y GAA TGC TAC GAG GGC GG E C Y E A G	CG GTG GCC GCG GTG AA A V AAG CCC CGT CGC (K P R R CCCcggaggaatggaatggaatggaatggaatggaatgga	TG GCT GAA GCC A E A GGC TGA gacag G * AGGC TGA gacag G CTC AAC GAT L N D GG TG CGC GGC G H W GG CTG CGC GGC L R G GC CCC TCG TT P S TF AT CCG CAT ACC P H T GGC GAA TAC G E Y GG AAA GCC N A C	G CCG AAA CCG GCG P R P A ACCCGGCCTGGCGCTGGCGCTGGCGCTGGCGCGCGCGCGC	GCG AAG CCC CCG GC A K P A P A P A GCG AGG CC CGG CG CG CGG CG CGG CG CGG CG	CG CGC GCG 1769 P A 1378 aaaccatcccga 1858 486 ATG CAC 2042 M H 16 GC CGC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 CG GGC ACC 2267 G T 91 TC GCC ATG 2342 A M 116 CG ATG GAC 2417 I D 141 GC ATG GAC 2417 I D 141 GC ATG GAC 2492 H H 166
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 142 2493	GTC AAG AGG CCC GAG C V K T P E C CCG AAA GGC AAG TCC F K G K S F K G K S GCCCtgoggggctcggggatgt cggacaccggccccgcgaaa GAC CGC TCG CTG AAG G D R S L K G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C S P M P K Q TCC AAG GAC GC GAA A C K D R E I GGC GG GG GG GC GCC TAT A GAT TGC TAC GAG GCC G E C Y E AA G GGT GG GGG GGC CTAT A	CG GTG GCC GCG GCG A A V AGG CCC CGT CGC CCC K P R R CCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	TG GCT GAA GCC A E A GGC TGA gacagi G * asgcgccgggctgc G CTC AAC GAT L N D GT GGT CAC TGC G H W GG CTG CGC GGC G CCC TCG TT P S GT CGC CAT AC P H T GGC GCG GAA TAC G E Y AG AAC GCC TGC AA ACC GCC CTG AA ACC GCC CTG	G CCG AAA CCG GCG P R P A ACCCGGCCCGGGCCCGGGCCCGGCCCGGCCCGGCC	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CC GCG 1769 P A 137 A 148 A
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 142 2493 167 2568	GTC AAG AGG CCC GAG C V K T P P CCG AAA GGC AAG TCG F F K G K S GCCCtgagggctcggggtctgggacaccggccccgccgaaa GAC CGC TCG CTG AAG G D R S L K G TGG ATC ATC GAG GAG C TCC TCC TC AG CGA GC G W I I E Q V TCC TTC ATG CGA GAG C S F M P K Q TGC AAG GAG CGC GAA GC S F M P K Q TGC AAG GAC CGC GAA GC S F M P K Q GAA TGC TAC GAG GCC GG GA A A Y GAA TGC TAC GAG GCC G GG GGG GGC TAT A G A G A Y I TATG AAA CGC CCT TTC C	CG CTG GCC GCG GCG V A A V VAGC CCC CGT CGC C K P R R CCCGGGGCGGGACCGGGACCGGGACCGGGACCGGGC CGGGCGACGGGCGACGGGACGGGACGGAC	TG GCT GAA GCC A E A GGC TGA gacagig C * GGC TGAC GAC GGC TGAC GAC GGC TGAC GGC GCC TGG TT P S F GGC CCC TGG TT P T GGG GAA TA: GG E Y GAA ACC GCC CT AA ACC GCG CT AA ACC GCG CT AA ACC GCG CT AG CTT TAT GGC	G CCG AAA CCG GCG P R P A ACCCGGCCGGGCGCGGCGCGCGCGCGCGCGCGCGCGC	GCG AAG CCC CCG GC A K P A P A A C GCC ACG GC ACG CCC ACG GC ACG ACG	CG CC GCG 1779 P A 1378 aaaccatccga 1858 GG ATG CAC 2042 M H 16 GC CGC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 GG GGC ACC 2267 G T 91 TC GC ATG 2342 A M 116 CC ATG GAC 2417 I D 141 TG CAC CAC 2492 H H 166 TG CCG CGG 2567 F R 191 TC GCC CGG 2567 F R 191 TC GCC CGG 2567 F R 191 TC GCC CGG 2567 F R 191 TC GC CGC CGG 2567 F R 191 TC GC CGC CGG 2567 F R 191 TC ATG GCC 22642
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 142 2493 167 2568	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A F K G K S GCCCtgogggctccgggatgt cggacaccggcccccgccgaaa GAC CGC TCG CTG AAG G D R S L K G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C S P M P K Q TGC AAG GAC GC GAA A C K D R E I GGC GCG GCG GCC TG G A R A A Y GAA TGC TAC GAG GCC G E C Y E A G GGT GGG GGC GCC TAT A G A G A C C CTAT A G A G A C C C CTAT A G A G A C C C C TAT A G A C C C C TAT A G A C C C C C TAT A G A C C C C TAT A G C C C TAT A G C C C C C TAT A G C C C TAT	CG GTG GCC GCG GTG AA A V AAG CCC CGT CGC CK F R R CCCGgaccapatccggacapatccggacapatccggacapatccggacafc A M A R TC AAG GCC TCG GC K A A S A M A R TC AAG GCC TCG GC S D G R TC ATG CGC CAC GG M R H D AT ATC TAT ATT CC AT ATC TAT ATT CC GG CTG ATC GGC GG CTG GC T T T T T T T T T T T T T T T T T T T	TG GCT GAA GCC A E A GGC TGA gacag G * agcgccgggctgc G CTC AAC GAT L N D F GGT CAC TGC G H W GG CTG CGC GGC L R G GT GCC TCG TT P S F AT CCG CAT ACC P H T GC GGC GAA TAC G E Y GA AAC GCC CT A ACC GCG CT A ACC GCG CT T AT LGG L Y GG L Y GGC L Y GGC L Y GGC L GAA TAC G E Y GAA CGC GCG CT T A LGG GGC CT T TAT GGGC L Y GGC L Y GGC	G CCG AAA CCG GCG P K P A A A CCG GCG ACC GAT CGC ATC ATC A CCC GCG GCG GCG GCG GCG GCG GCG GCG	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CCC GCC 1769 P A 1378 aaaccatccga 1858 GG ATG CAC 2042 M H 16 GC CCC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 CG GG ACC 2267 G T 91 TC GCC ATG 2342 A M 116 CC ATC GAC 2417 I D 141 TG CAC CAC 2492 H H H 166 TG CAC CAC 2492 H H H 166 TG CCC GCG 2567 P R 191 CC ATC GCC 2642 I A 216
1695 113 17700 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 23443 117 2418 142 2493 167 2568 192 2643 217	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC ACG TC A F K G K S GCC-toggagactegogacggacggacggacggacaccggc-cccgcgaaa GAC CGC TCG CTG AAG G D R S L K G D R S L K G TGG ATG ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C TTC ATG CCG AAG C S P M P K Q TCC AAG GAC GC GAA A C K D R E I GGC GG GGG GCC GAC GAA TGC TAC GAG GCC T GAA TGC TAC GAG GCC G E C Y E A G GT GG GGC GCC TAT A G A G A Y I ATG AAA CCG CG TTC C M K P P GTC GTG CCG ACG ACT TC C M K P P GTC GTG CCG ACG ACT TC C M K P P GTC GTG CCG ACG ACC ACC V V V P T I L	CG GTG GCC GCG GCG A A V AGG CCC CGT CGC CCC K P R R CCCGGGGGGATGGCAT A M A R TC AAG GCC TCG GCC K A S GC A TC GAC GGG CCC MTC GAC GGG CCC MTC ATG CGC ACC MTC ATG ATG ACC MTC ATG ATG ACC MTC ACC	TG GCT GAA GCC A E A GGC TGA gacagi G * agcgccgggctgc G CTC AAC GAT L N D ET GGT CAC TGG G H W G CTG CGC GG L C GG GG L C TG TG P S F AT CGC CAT ACC P H T GG GGC GAA TAC G G AAC GCC TG T A L GG AAC GCC GCT T A L GG CTT TAT GG ET GGC TGG TG CTT TAT GG ET GGC TGG TGG TG TGG TGG TGG TG TGG TGG TGG TG TGG TG	G CCG AAA CCG GCG P R P P A ACCCGGCCCGGGCCTGGGCCTGGCCGGCGGCGCTGGGCACGCGCGCG	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CC GCG 1769 P A 1378 aaaccatccga 1858 GG ATG CAC 2042 M H 16 GG CC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 GG GGC ACC 2267 G T 91 TC GC ATG 2342 A M 116 CC ATC GAC 2417 I D 141 TG CAC CAC 2492 H H 166 TG CG CGG 2567 P R 191 CC ATG GC 2642 GG GT AGG 2567 P R 191 CC ATG GC 2642 GG GT AGG 2717 V K 241
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 142 2493 167 256 268 167 268 27 268 27 268 27 268 27 268 27 268 27 268 27 268 27 268 27 268 27 268 27 268 27 268 27 27 27 27 27 27 27 27 27 27 27 27 27	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC ACG TC A F K G K S GCC-toggagactegogacggacggacggacggacaccggc-cccgcgaaa GAC CGC TCG CTG AAG G D R S L K G D R S L K G TGG ATG ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C TTC ATG CCG AAG C S P M P K Q TCC AAG GAC GC GAA A C K D R E I GGC GG GGG GCC GAC GAA TGC TAC GAG GCC T GAA TGC TAC GAG GCC G E C Y E A G GT GG GGC GCC TAT A G A G A Y I ATG AAA CCG CG TTC C M K P P GTC GTG CCG ACG ACT TC C M K P P GTC GTG CCG ACG ACT TC C M K P P GTC GTG CCG ACG ACC ACC V V V P T I L	CG GTG GCC GCG GTG V A A V AND CCC CGT CGC CK F R R CCCGGGCAGATCGGG NUOF M GC GCA ATG GCG CC A A GC TCG GG A A A R A M A R TC AAG GCC TCG GG A A GC TCG GG A TC GAC GGC CC S D G R TC ATG CGC CAC GG M R H D AT ATC TAT ATT CC GG CTG ATC GGG GG TG TG GGC ATC GG CTG GG GG TG TG GGC ATC GG CTG GG GG GG CTG GG GG GG CTG GG GG TG GGC GGG GGG TG GGC TG GG	TG GCT GAA GCC A E A GGC TGA gacagic G * GGC TGAC GAT GGT CAC GAT L N D GT GGT CAC TGC G H W GG CTG CGC GGT F AT CGC CAT AC F AT CCG CAT AC F AT CCG CAT AC G G E Y AC GAA AC GCC TGC AA ACC GCC CT AA ACC GCC GCT T A L GG CTT TAT GGC GT GCC TGG TT T A L GG CTT TAT GGC GT GCC TGG TT T A C GG CTT TAT GGC GT GCC TGG TT T A C GG CTT TAT GGC GT GCC TGG TT TGCC TGC TGC TGC TGCC TGG TT TGCC TGC TGC TT TGCC TGG TT TGCC TGG TT TGCC TGC TGC TT TGCC TGC TT TGCC TGC TT TGCC TT TT TCCC TT TGCC TT TT TCCC TT TGCC TT TGCC TT TGCC TT TT TCCC TT TT TCCC TT TT TCCC TT TT TCCC TT TT TT TCCC	G CCG AAA CCG GCG P R P P A ACCCGGCCCGGGCCTGGGCCTGG ACG GCG ACG GCG ACG GCG ACG GCG ACG GCG ACG GCG ACG A	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CC GCG 1769 P A 1378 acaccatcccga 1858 GG ATG CAC 2042 M H 16 GC CG GAC 2117 R D 41 FC AAG TGG 2192 K W 66 GG T 91 FC GC ATG 2342 A M 11 GC ATC GAC 2417 I D 141 GC ATC GAC 2492 H H 166 GC ATC GC CGG CGG 2567 P R 191 CC ATC GCC CGG CGG 2667 I D 141 G CAC CAC 2492 H H 166 GC CGC GGG CGG CGG CGG CGG CGG CGG CGG
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 142 2493 167 2568 192 2643 247 2718 242 2793	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GGC AAG TCG A F K G K S GCCCtgogggctcgggatgt cggacaccggcccccgcgaaa GAC CGC TCG CTG AAG G D R S L K G D R S L K G TGG ATC ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C S F M P K Q TGC AAG GAC GG GAA C K D R E I GGC GGG GCG GCC T G AA G AC CGC GAA GAA TCC TAC GAG GCC G G A R A A A GAA TCC TAC GAG GCC G G A G G GC GCC TAT A G A G C C GAC ATC C G A G A C G GAC ATC C C V V P T I L CTG GTG CGG ACG ATC G L T A G AA CCC GC TTC G C TTC GCG ACG ATC C C TTC GCG ACG ATC C C TTC GCG ATG TCG G L T A G AA CCC GC TTC G C TTC GCG ATG TCG G L T A G AA CCC GC GTC G C TTC GCG ATG TCG G C TTC GCG ATG TCG G C TTC GCG ATG TCG G C TTC GCA AAG CCC GTC TCG G C TTC GCG ATG TCG G C TTC GCG ATG TCG G C TTC GCA AAG CAC GCC GTC G C TTC GAA AAG CAC GCC GTC G C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GATC C C TTC GAA AAG CAC GCC GTC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GATC C C TTC GAA AAG CAC GCC GAC GATC C C TTC GAA AAG CAC GCC GTC TCC GAC AAG CAC GCC GTC GAC GATC C C TTC GAA AAG CAC GCC GTC TCC GAC AACC GCC TTC AACC GAC AATC C C TTC GAA AAG CAC GCC GTC TCC GAC AACC GCC TTC AACC GAC AACC GCC TTC AACC GAC AACC C C TTC GAA AAG CAC GCC TTC AACC GAC AACC GCC TTC AACC AAC	CG GTG GCC GCG GCG A A V AGG CCC CGT CGC C K P R R CCC CGT CGC C K P R R CCC GGT GGC C K A R S GC GC GC TCC GC K A S G A TC AAG GCC TCC GC S D G R GC K A S G ATC ATC GAC GGC C M R H D ATT ATC TATT ATT C GC CAC GC GC CGC GGC GCC T GGC ATC GGC A G S G C G G C GGC GGC A G S G T G CGC GGC GGC GGC A G S G T G CGC CGT GGT GC R R G G G G GC CGT GGT GC R R G G G G GC CGT GGT GC R R G G G G GC CGT GGT GC R R G G G G GC CGT GGT GC R R G G G G G G GC CGT GGT GC R R G G G G G G GC CGT GGT GC R R G G G G G G GC CGT GGT GC R R C G G G G G GC R R C G G G G G GC R R C G G G G G C GC R C G G R R C G G G G G G G G G G R R C G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G G	TG GCT GAA GCC A E A GGC TGA gacagi G * GGCT AAC GAT L N L GGC TGC GAC GG H W GG CTG CGC GCC L C GGC TGC TGC AA ACC GCG CTT T A L GG CTT TA L GG CT TGC TGC AA ACC GCG CTT T A L GG CT TGC TGC AA ACC GCG CTT T A L GG CT TGC TGC ACC ACC ACC ACC ACC ACC ACC ACC ACC A	G CCG AAA CCG GCG P R P A P A P A P A P A P A P A P A P A	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CCC GCG 1769 P A 1378 aaaccatccga 1858 GG ATG CAC 2042 M H 16 GG CCC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 CG GC ATG 2267 A M 116 CC ATC GAC 2417 I D 141 TG CAC CAC 2492 H H H 166 TG CCC CAC 2492 GC GTC AAG 2717 V K 241 AG GAG CTG 2792 E L 266 CC TCC TGC 2866
1695 113 1770 138 1859 1959 1 2043 17 2118 42 2193 67 2268 92 2343 117 2418 42 2493 117 2568 192 2643 217 2718 242	GTC AAG ACG CCC GAG C	CG CTG GCC GCG V A A V AGG CCC CGT CGC C K P R R CCCggagcgacatcgggacaTt AUOF MOF GC GCG ATG GCC K A A A R TC AAG GCC TCG GC K A C GGC GC A TC GAC GGG C TC GAC GG C G G G G GAC C G G G G G GAC C G G G G G G G G G G G G G G G G G G	TG GCT GAA GCC A	G CCG AAA CCG GCG P R P A ACCCGGCCGGGCGCGGCGCGGCGCGGCGCGGCGCG	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CC GCG 1769 P A 137 A
1695 1138 1770 138 18599 1959 12043 177 22118 42 21193 67 2264 327 2418 142 2493 167 2568 192 2643 217 247 2568 2643 217 247 247 247 247 247 247 247 247 247 24	GTC AAG AGG CCC GAG CV K T P E CGG AAA GGC AAG TCG F K G K S GCC GAGG GGG GGG GGG GGG GGG GGG GGG GG	CG CTG GCC CGC GCC AV A A V V A A V V A A V V A A V V A A V V A A V V A A C A V A A C A V A A C A C	TG GCT GAA GCC A E A GGC TGA gacagic G * GGC TGAC GAT GGT GGC GGC GGC GGC GAT ACC GGC TGG ACC GGC GGC ACC ACC GGC ACC GGC GGC ACC ACC GGC ACC GGC GGC ACC ACC GGC ACC ACC GGC ACC ACC GGC ACC ACC ACC ACC ACC ACC ACC ACC ACC AC	G CCG AAA CCG GCG P R P A ACCCGGCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCG	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CC GCG 1769 P A 137 A
1695 1133 1770 1385 1959 1959 12043 17 2118 42 2193 67 2268 92 23443 117 2418 142 2493 167 25643 217 2718 2492 267 2868	GTC AAG AGG CCC GAG CV K T F CGG AAA GGC AAG TCG F K G K S GCC TGG AAG GC GG AAG AG CGG AAG AG CGG AAG AG CG AGG AG	CG CTG GCC CGC GCC AV A A V V A A V V A A V V A A V V A A V V A A V V A A C A V A A C A V A A C A C	TG GCT GAA GCC A E A GGC TGA gacagi G * asgcgccgggctgc G CTC AAC GAT H ND GT GGT CAC TGC G H W GG CTG CGC GT P F F AT CCG CAT AC G AAC GCC TGC AAC GCC TGC AA ACC GCC CT AAC GCC TGC AAC GCC TGC GT CCC TGC GT CCC CTGC GT CCC CTCC CT	G CCG AAA CCG GCG P R P A ACCCG ACC ATC CAT CGC ACC ACC ACC ACC ACC ACC ACC ACC ACC	GCG AAG CCC CCG GC A K P A A P A A P A A C P A A C P A A C P A A C P A A C P A A C A C	CG CC GCG 1769 P A 137 A 148 A
1695 1138 1770 1388 18599 1959 12043 17 2118 42 2193 67 2248 92 2343 167 2418 142 2493 167 257 2493 247 2718 242 2793 247 2718 242 2793 243 247 2718 2493 2493 2493 2493 2493 2493 2493 2493	GTC AAG ACG CCC GAG CV K T P EV GAA GCC AAG CCC AAG CCC GAG CCC AAG CCC CCC GAG CCC CCC GAG CCC	CG GTG GCC GCG CG GTG GCC GCG A A V AGG CCC CGT CGC CGC K P R R CCCGgacgaccagaccagaccagaccagaccagaccagac	TG GCT GAA GCC A	G CCG AAA CCG GCG P R P R P A ACCCGGCCGGGCGCGGCCGGCGCGCGCGCGCGCGCGC	GCG AAG CCC CCG GC A K P A A P A A GGGGC AAC ACC AAC AAC ATC TAC GC GC ATC CCG ATC ACC ACC ACC ACC ACC ACC ACC ACC ACC	CG CC GCG 1779 P A 1 1 1779 P A 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1695 1133 1770 138 1859 1959 1 2043 17 22193 67 22268 92 2343 117 22568 1249 2418 142 2493 247 2718 2493 267 267 2868 292 2643 2793 267 2868 292 2943 317	GTC AAG ACG CCC GAG C V K T P E P CCG AAA GCC ACG TCG A F K G K S GCCCtgogggctcgggccccgggatgt cggacaccggcccccgcgaaa GAC CGC TCG CTG AAG G D R S L K G TCG ATG CTG ACG GC W I I E Q V TCC TTC ATG CCG AAG CAG G W I I E Q V TCC TTC ATG CCG AAG CC F M P K Q TCC AAG GAC GC GAA A C K D R E I GGC GGG GGG GCC TG GAA TCC TAC AAG GCC G E C Y E A G GAA TCC TAC GAG GCC G E C Y E A G GAT GG GG GCC GCT TAT A G A G A Y I AAG AAA CCG CAG ATC C V V P T I L CTG TTC GCG ACG ATC C V V P T I L CTG TTC GCG ACG ATC C C T GAA AAG CAC GCC GT AAG CAC GCC GTC TCG AAG ATC GAA AAG CAC GCC GC T C GGG CGC GCC GCC GCC T C GGG CGC GCC GCC GCC T TC CGC GTC GCG GCC GCC T TC CGC GTC GCG GCC GCC T TC CGC GTC GCG GCC GCC T TC CGC GCC GCC GCC GCC GCC GCC GCC GCC	CG GTG GCC GCG CG GTG GCC GCG A A V AGG CCC CGT CGC CGC K P R R CCCGgacgaccagaccagaccagaccagaccagaccagac	TG GCT GAA GCC A E A GGC TGA gacagi G * AGC TGA gacagi G CTC AAC GAT L N D TG GGT CAC TGG GG CAC TGG CC TGG TT P S F AT CGC CAT ACC P H T GGC GGC GAA TAC GG AAC GCC TG T A L GG CTG CTG A ACC GCG CTT T A L GG CTG CTG CT TAT C C CAC TGC TGC C C C C TGC TGC C C TGC TGC C C TGC TGC C TGC C TGC TGC C TCC C	G CCG AAA CCG GCG P R P P A ACCCGGCCCGGGCCCGGGCCCGGCCCGGCCCGG	GCG AAG CCC CCG GC AA K P A P A P A P A P A P A P A P A P	CG CCC GCG 1769 P A 1378 aaaccatccga 1858 GG ATG CAC 2042 M H 16 GG CCC GAC 2117 R D 41 TC AAG TGG 2192 K W 66 CG GC ATG 2342 A 11 TC AAG TGG 2492 K W 66 CG GC ATG 2492 K W 66 CG GC ATG 2492 A M 116 CC ATC GAC 2417 I D 141 TG CAC CAC 2492 H H H 166 TG CAC CAC 2492 H H H 166 TG CCG CGG 2567 P R 191 CC ATC GAC 2492 H H H 166 TG CCG CGG 2567 CC ATC GCC 2662 I A 216 CC CGG CGG 2567 CC CGG 2567 CC CGG 2567 CC CGC 2662 I A 216 CC CGC 2567 CC CGC 2567 CC CGC 2567 CC CGC 2567 CC CC CC CC 2567 CC CC CC CC 2567 CC CC CC CC CC 2567 CC
1695 1133 1770 1385 1959 1959 12043 17 2218 42 2118 42 2118 42 22493 167 2418 1422 2493 167 2718 242 2793 267 2718 242 2793 217 2868 292 2943 317 317 317 317 317 317 317 317 317 31	GTC AAG AGG CCC GAG C V K T P E P CCG AAA GGC AGG GC F K G K S GCC-tygogggctcygggctggt cgga-accggcc-ccgcgaaac GAC CGC TCG CTG AAG G D R S L K G D R S L K G TGG ATG ATC GAG CAG G W I I E Q V TCC TTC ATG CCG AAG C GC ATG CGG GG GG GG GG GGC GAA A C K D R E I GGC GG GG GG GC GA GGT GG GG GG GC TAT A G A R A A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G E C Y E A GAA TGC TAC GAG GC G GT GT CCG GGC GC TC C M K P P F P GTC GTG GG GG GG GC GC GT TC GAG AGG ATC TCG ATC ATC GGC GG GG GC GG ATC ATT CCG ATC ATC GGC GG GG GC GG ATC ATT CCG ATC ATC GGC GG GG ATC ATT CGG GGG GGG GG ATC ATT CGG GGG GGG GGG GGG GGG GGG GGG GGG G	CG GTG GCC GCG GCG GCG GCC GTG GCC GT GCC GCG GCC GT GCC GCG GCC G	TG GCT GAA GCC A E A GGC TGA gacagi G * AGCC TGA gacagi G CTC AAC GAT GCT CAC GAT GGT CAC GAT GGT CAC GGC G H W GG CTG CGC GT P H T GGC GCA GAC G AAC GCC TGC AA ACC GCC CT AA CC GCC CT GT AC G CC TGC GT GCT GCC AA ACC GCC TGC AACC GCC TGC AACC GCC ACC AACC ACC GCC AACC ACC ACC AACC ACC ACC AACC ACC AC	G CCG AAA CCG GCG P R P A ACCCGGCCCGGGCCTGGGCCTGGCCGGCGCGGCGCGG	GCG AAG CCC CCG GC A K P A P A P A P A P A P A P A P A P A	CG CC GCG 1769 P A 137 A 137 A 148 A
1695 1133 1770 1385 1959 1959 12043 17 2218 42 21193 42 2193 367 2418 142 2493 167 2568 192 2493 127 2718 242 2793 217 2718 242 2793 2793 2793 2793 2793 2793 2793 279	GTC AAG ACG CCC GAG C	CG GTG GCC GCG GTG GCC A A V V A A V V A A V V A A V V A A V V A A GCC CGT CGC GCG GCG GCG GCG GCG GCG GCG	TG GCT GAA GCC	G CCG AAA CCG GCG P R P A ACCCGGCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	GCG AAG CCC CCG GC AA K P A P A P A P A P A P A P A P A P	CG CC GCG 1769 P A 137 aaaccatccga 1858 GA 700 CAC 2042 M H 16 GC CGC GAC 2117 R D 17 GC ACC 2042 A M 116 GC ACC ACC 2492 H 16 GC CC CC CGC 2567 R CC ACC 2667 CC CC CC CGC 2567 CC
1695 1138 1770 138 18599 1959 12043 177 22118 42 2193 67 22168 42 2193 117 2418 142 2493 167 2568 192 2443 217 2463 217 247 248 2493 267 268 2718 2493 267 268 2718 2718 2718 2718 2718 2718 2718 271	GTC AAG ACG CCC GAG C	CG CTG GCC CGC GCC CGC CGC CGC CGC CGC C	TG GCT GAA GCC	G CCG AAA CCG GCG P R P A ACCCGGCCGGGC GCG ACC GAT CGC ACC GCG CGC B GCG B B B B	GCG AAG CCC CCG GC A K P A A P A P A P A P A P A P A P A P	CG CC GCG 1769 P A 137 aaaccatccga 1858 GC ATC 2042 M H 16 GC CG GG ACC 2267 TC GC ATC ACC ACC ATC ACC ACC ACC ACC ACC AC

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 α -cyano-4-hydroxy-trans-cinnamic acid and analyzed by MALDI mass spectrometry. 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

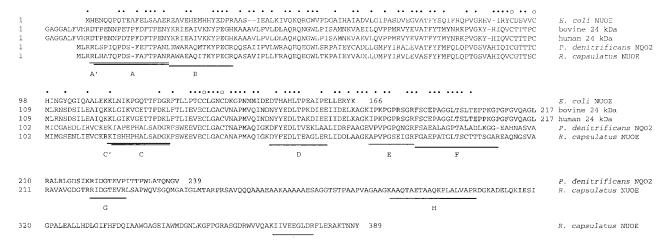


Fig. 2. Comparison of the polypeptide sequences of the *R. capsulatus* NUOE subunit, the *E. coli* NUOE subunit, the bovine and human 24 kDa subunits (as processed mitochondrially imported subunits) and the *P. denitrificans* NQO2 subunit. Identical residues between the *R. capsulatus* NUOE and the human 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.

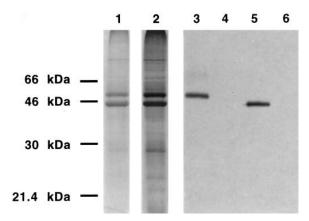


Fig. 3. SDS polyacrylamide gel and Western blot analysis of the fractions eluted from the anti-PNE affinity column. Proteins recovered from the affinity column by acidic elution were precipitated overnight with 5 volumes of ethanol at -20° C (lanes 1, 2, 3 and 5). Membrane lysates collected in the flow 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).

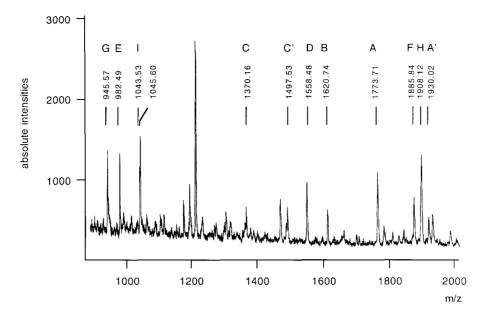
overlapping this gene. No error was found in our consensus sequence. Accordingly, we set up immunoaffinity experiments to check the size of the expressed subunit in the *R. capsulatus* membranes so as to definitively ascertain the existence of an elongated C terminus in the *R. capsulatus* NUOE subunit.

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 pI 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-



Peptide	Position	Sequence	Theoretical MH+
G	221-228	RIDGTEVR	945.51
E	174-183	AP VPGSE I GR	982.53
1	371-379	IIVEEGLDR	1043.57
С	117-129	ISHHPHALSADGK	1369.70
C'	116-129	KISHHPHALSADGK	1497.79
D	152-164	DYFEDLTEAGLER	1557.71
В	21-34	AWAEAQITKYPEGR	1619.82
Α	5-20	LHATOPDSFAFTPANR	1772.87
F	184-202	FGAEPATGLTSCTTTSGAR	1884.88 (a)
Н	288-306	AAQTAETAAQKPLALVAPR	1907.07
Α'	4-20	RLHATQPDSFAFTPANR	1928.97

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

- 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.

- [4] C.I. Ragan, Curr. Topics Bioenerg. 15 (1987) 1-36.
- [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. Vinçon, 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-
- [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.