Isolation and Characterization of the *nifUSVW-rpoN* Gene Cluster from *Rhodobacter sphaeroides*

WIM G. MEIJER[†] AND F. ROBERT TABITA*

Department of Microbiology and The Biotechnology Center, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210-1292

Received 10 February 1992/Accepted 3 April 1992

The *rpoN* gene from *Rhodobacter sphaeroides* was isolated from a genomic library via complementation of a *Rhodobacter capsulatus rpoN* mutant. The *rpoN* gene was located on a 7.5-kb *HindIII-EcoRI* fragment. A Tn5 insertion analysis of this DNA fragment showed that a minimal DNA fragment of 5.3 kb was required for complementation. Nucleotide sequencing of the complementing region revealed the presence of *nifUSVW* genes upstream from *rpoN*. The *rpoN* gene was mutagenized via insertion of a gene encoding kanamycin resistance. The resulting *rpoN* mutant was not impaired in diazotrophic growth and was in all respects indistinguishable from the wild-type strain. Southern hybridizations using the cloned *rpoN* gene as a probe indicated the presence of a second *rpoN* gene. Deletion of the *nifUS* genes resulted in strongly reduced diazotrophic growth. Two conserved regions were identified in a NifV LeuA amino acid sequence alignment. Similar regions were found in pyruvate carboxylase and oxaloacetate decarboxylase. It is proposed that these conserved regions represent keto acid-binding sites.

The sigma factor of RNA polymerase confers promoter specificity to the holoenzyme. After initiation of transcription, the sigma factor is released, and the core RNA polymerase proceeds to transcribe the DNA. In addition to the major sigma factor, other sigma factors are frequently employed by bacteria (20). One of these alternative sigma factors is RpoN (23). RNA polymerase containing RpoN initiates transcription at promoters containing the consensus sequence CTGG..8 bp..TTGCA 26 bp upstream from the transcription initiation site (9). Although RpoN was initially recognized for its role in nitrogen metabolism, it now is firmly established that RpoN is required for a wide range of metabolic activities, including hydrogen oxidation, use of C_4 dicarboxylic acids, assimilation and dissimilation of nitrate, flagellation, and formation of pili (37). The metabolic role of RpoN is species dependent. The rpoN gene has been cloned from a number of organisms and was found to be flanked by two conserved open reading frames (ORFs) of unknown function (6, 12, 27, 28, 35, 36, 41-43, 49a, 50, 58). In Rhodobacter capsulatus, the rpoN gene is not flanked by these conserved ORFs but is located between nifHDK, which encodes the Fe and FeMo proteins of nitrogenase, and nifA, which encodes a transcriptional activator of the nitrogen fixation (nif) genes (31). The nifA gene is duplicated in R. capsulatus. Upstream from the duplicated nifA gene, the nifUSVW genes are present instead of rpoN (34). NifV catalyzes the synthesis of homocitrate, an integral part of the FeMo cofactor of nitrogenase (24-26). The nifU, nifS, and nifW genes are involved in maturation or stability of nitrogenase (30, 46).

Rhodobacter sphaeroides, a metabolically versatile organism, can grow in five different growth modes. During autotrophic or photoheterotrophic growth, it employs two forms of ribulose bisphosphate carboxylase (RuBisCO). These forms of RuBisCO are encoded in two similar but not identical operons in which genes encoding other enzymes required for CO_2 fixation are also found (16–19). In the upstream region of the form I operon, there exists a sequence very similar to the RpoN consensus promoter sequence (17), suggesting that transcription of this operon may be dependent on RpoN (57). To probe the role of RpoN in the expression of the *cfx* operons and metabolism of *R*. *sphaeroides* in general, we set out to clone and inactivate the *rpoN* gene of *R*. *sphaeroides*. In this article, we report the cloning and sequence analysis of *rpoN* from *R*. *sphaeroides* and its organization in an operon with the *nifUSVW* genes. Evidence for the presence of a second copy of the *rpoN* gene is also presented.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are listed in Table 1.

Media and growth conditions. R. capsulatus and R. sphaeroides strains were grown photoheterotrophically in Ormerod's medium (45) supplemented with malate (MSM) at 30°C in screw-cap tubes or in GasPak anaerobic jars (American Scientific Products, McGaw Park, Ill.) as previously described (61). Nitrogen fixation capacity was tested by growth under photoheterotrophic conditions with MSM without NH₄Cl. Rhodobacter strains were also grown heterotrophically on peptone-yeast extract or minimal medium supplemented with fructose in Erlenmeyer flasks or on agar plates (61). Escherichia coli strains were grown on Luria-Bertani (LB) or M9 medium at 37°C (51). When appropriate, the following supplements were added (concentrations given in micrograms per milliliter except as otherwise noted): ampicillin, 50; 5-bromo-4-chloro-3-indolyl-β-D-galactoside, 20; isopropyl-β-D-thiogalactoside, 0.1 mM; kanamycin, 25; rifampin, 50; spectinomycin, 100; streptomycin, 50; and tetracycline, 12.5 (E. coli) or 1 (R. capsulatus and R. sphaeroides). Agar (1.5%) was added for solid media.

DNA manipulations. Plasmid DNA was isolated by the alkaline lysis method of Birnboim and Doly (10). Chromosomal DNA was isolated via the sarcosyl-lysate procedure

^{*} Corresponding author.

[†] Present address: Department of Microbiology, University of Groningen, 9751 NN Haren, The Netherlands.

Strain or plasmid	Genotype or characteristics	Source or reference
E. coli		
JM107	endA1 gyrA96 thi hsdR17 supE44 reLA1 Δ (lac-proAB) [F' traD36 proAB lac1 ^a Z Δ M15	59
S17-1	thi pro $r^- m^+ Sm' Tp' recA RP4-2 (Tc::Mu, Km::Tn7)$	54
SM10	supE44 hsdR thi thr leuB6 lacY1 tonA21 recA (1c::Mu, Km::Tn7)	54
RM1628	pro met Km ⁴ (In5)	R. J. Meyer
R. sphaeroi-		
des		
HR	Sm ^r	61
HRI	Sm ^r Rif ^r	This study
EM1	Sm ^r <i>rpoN</i> ::Km ^r	This study
EM11	$\operatorname{Sm}^r \operatorname{Rif}^r \Delta nifUS::\operatorname{Sp}^r$	This study
R. capsulatus		
LJİ	rpoN	60
Plasmids		
pRK404	Tc ^r , broad-host-range vector	13
pSUP202	$Ap^{r} Tc^{r} Cm^{r} mob^{+}$	54
pUC1318	Ap ^r , pUC vector with hybrid cloning site	32
pTZ18U	Cloning vector	Bio-Rad
pTZ19U	Cloning vector	Bio-Rad
pHP45-Ω	Ap ^r Sp ⁷ /Sm ^r	48
pUC1318K	Km ^r Ap ^r , pUC1318K with a 1.5-kb <i>Hin</i> dIII-Sall fragment from Tn5 encoding Km ^r	This study
pNIT1	Tc^r , pVK102 with a 18-kb fragment of R. sphaeroides chromosomal DNA encoding rpoN	This study
pNIT2	Tc^r , pVK102 with a 20-kb fragment of R. sphaeroides chromosomal DNA encoding rpoN	This study
pNIT11	Tc ^r , pRK404 with a 9-kb <i>Hin</i> dIII fragment from pNIT1 encoding <i>rpoN</i>	This study
pNIT12	Tc ^r , pRK404 with a 9-kb <i>Hin</i> dIII fragment from pNIT2 encoding <i>rpoN</i>	This study
pNIT122	Tc ^r , pRK404 with a 7.5-kb <i>HindIII-Eco</i> RI fragment from pNIT12	This study
pSNT108	Ap ^r , pTZ18U with a 2.7-kb <i>Hin</i> dIII- <i>Sal</i> I fragment from pNIT122	This study
pSNT4	Ap ^r , pTZ18U with a 1.2-kb <i>Bam</i> HI fragment from pNIT122	This study
pKTN1	Ap ^r Tc ^r , pSUP202 with a 1.2-kb <i>Bam</i> HI fragment encoding <i>rpoN</i>	This study
pKTN2	Ap ^r Tc ^r Km ^r , pKTN4 with a 1.5-kb SacI fragment from pUC1318K encoding Km ^r	This study
pKTN3	Ap ^r Sp ^r , pSNT108 with a 1.2-kb StuI deletion and a 2.0-kb SmaI fragment from pHP45-Ω encoding Sp ^r	This study
pKTN4	Ap' Tc' Sp', pSUP202 with a 3.5-kb PvuII fragment from pKTN3	This study

TABLE 1. Bacteria and plasmids used in this stu	. Bacteria and plasmids used in this	study
---	--------------------------------------	-------

(15). DNA-modifying enzymes were obtained from Bethesda Research Laboratories (Gaithersburg, Md.) and were used according to the manufacturer's instructions. Analysis of restriction digests by gel electrophoresis and transformation of *E. coli* were performed as described by Sambrook et al. (51). DNA fragments were isolated from agarose gels by adsorption to glass (Geneclean kit; Bio 101, La Jolla, Calif.).

Southern hybridizations. DNA was transferred to nylon membranes (GeneScreen Plus; Du Pont) by electroblotting as specified by the manufacturer. Prehybridization, hybridization, and washing procedures were performed at 63°C according to protocols suggested by the manufacturer, unless specified otherwise. DNA fragments used as probes were labelled with [³²P]dCTP by using a random primed labelling kit supplied by Boehringer (Mannheim, Germany).

Mobilization of plasmids. Mobilization of plasmids using E. coli S17-1 or SM10 containing the appropriate plasmids as a donor was performed essentially as described by Simon et al. (54). Triparental mating procedures were described previously (62).

Tn5 mutagenesis. *E. coli* RM1628 containing pNIT122 was grown on LB medium containing tetracycline and kanamycin until the late exponential growth phase; 0.1 ml of the culture was used to inoculate 50 ml of LB medium, while the kanamycin concentration was increased to 300 μ g/ml. When the culture reached the late exponential phase, 0.1 ml was

used to inoculate the same medium and allowed to grow overnight. Plasmid DNA was subsequently isolated and used to transform *E. coli* S17-1 to kanamycin resistance (Km^r) and tetracycline resistance (Tc^r).

Construction of R. sphaeroides rpoN and nifUS mutants. A 1.4-kb HindIII-SalI fragment from Tn5, encoding kanamycin resistance, was treated with Klenow enzyme and ligated into HindIII-digested, Klenow enzyme-treated pUC1318, generating a Km^r cartridge flanked by five unique restriction sites (pUC1318K). The 1.2-kb BamHI fragment internal to rpoN was treated with Klenow enzyme and ligated into EcoRI-digested, Klenow enzyme-treated pSUP202, generating pKTN1. A 1.4-kb SacI fragment containing the Km^r cartridge derived from pUC1318K was cloned into the unique SacI site within the rpoN gene, generating pKTN2.

A 1.4-kb StuI fragment from pSNT108 was replaced with the 2.0-kb SmaI fragment (Ω element) from pHP45- Ω , generating pKTN3. pKTN3 was digested with PvuII, and the fragment containing the Ω element was cloned into ScaIdigested pSUP202, generating pKTN4.

E. coli SM10 and S17-1 were transformed with pKTN2 and pKTN4, respectively. The pKTN2 plasmid was subsequently mobilized to *R. sphaeroides* HR, and pKTN4 was mobilized to *R. sphaeroides* HRI. Exconjugants were selected by screening for streptomycin resistance (Sm^r) and Km^r (pKTN2) or rifampin resistance (Rif^r) and spectinomycin resistance (Sp^r) (pKTN4). Since pSUP202 cannot replicate in *R. sphaeroides*, Km^r or Sp^r can be obtained only via integration of the plasmid into the chromosome. Selection for replacement of the wild-type gene with the mutated gene via a double recombination event was achieved by screening exconjugants for tetracycline susceptibility (Tc^s). Southern hybridizations confirmed that a double recombination had taken place.

Nucleotide sequencing. The 2.6-kb HindIII-SalI fragment of pNIT122 was subcloned in both orientations in pTZ19U. A nested set of unidirectional deletions was subsequently created by digestion with exonuclease III and mung bean nuclease, essentially as described by Henikoff (21). Other subclones were obtained by digestion of pNIT122 with the appropriate restriction enzymes and subsequent ligation with pTZ18U and pTZ19U. Infection of E. coli JM107 containing the derivatives of pTZ18U and pTZ19U with the helper phage M13KO7 (59) and purification of singlestranded DNA were performed as previously described (51). Dideoxy sequencing reactions were performed with modified T7 DNA polymerase (Sequenase; U.S. Biochemical Corp., Cleveland, Ohio) and [35S]dATP as described by the manufacturer. In addition to the sequencing reactions using dGTP, at least one strand was also sequenced with dITP to eliminate compressions. The nucleotide sequence data were compiled and analyzed by using the programs supplied in the GCG sequence analysis software package (Genetics Computer Group, Madison, Wis.).

Nucleotide sequence accession number. The sequences reported in this article have been entered into GenBank under accession number M86823.

RESULTS

Cloning of the R. sphaeroides rpoN gene. R. capsulatus LJ1 carries a mutation in rpoN (31) and is unable to grow under nitrogen-fixing conditions (60). A genomic library of R. sphaeroides DNA previously used to isolate carbon dioxide fixation genes (62) was mobilized into R. capsulatus LJ1 via triparental mating as described in Materials and Methods. The mating mixture was plated on MSM plates lacking a nitrogen source and incubated under nitrogen-fixing growth conditions to select for complementation of R. capsulatus LJ1. Nitrogen-fixing colonies were purified on fructosetetracycline agar. Plasmid DNA was subsequently isolated and used to transform E. coli to Tcr. Two different plasmids which had retained the capability to complement R. capsulatus LJ1 could be distinguished (pNIT1 and pNIT2). Restriction mapping showed that these plasmids share a 9-kb HindIII fragment.

Localization of rpoN. The 9-kb HindIII fragment common to pNIT1 and pNIT2 was subcloned in pRK404, generating pNIT11 and pNIT12, respectively. The 9-kb HindIII fragment was found to complement the rpoN mutation of R. capsulatus LJ1. Subsequent subcloning of the 9-kb HindIII fragment reduced the complementing region to a 7.5-kb HindIII-EcoRI fragment, generating pNIT122. The ability to complement was not dependent on the orientation of the insert with respect to the lac promoter of pRK404. Plasmid pNIT122 was introduced into E. coli RM1628, and transposition of Tn5 from the chromosome to pNIT122 was allowed to occur. The resulting pNIT122:Tn5 plasmids were mapped by restriction analysis, and those plasmids containing Tn5 within the 7.5-kb HindIII-EcoRI insert were mobilized to R. capsulatus LJ1 via conjugation. Tcr exconjugants were tested for their ability to grow under nitrogen-fixing condi-



FIG. 1. Restriction map of the *nifUSVW-rpoN* cluster from *R. sphaeroides*. Closed triangles represent Tn5 insertions preventing the complementation of *R. capsulatus* LJ1. Open triangles represent Tn5 insertions that do not affect complementation. The positions and directions of transcription of the genes and ORF1 are indicated by arrows below the restriction map. The black bar shows where the heterologous *rpoN* probe hybridized. The replacement of the *Stul* fragment with the Ω element and the insertion of the kanamycin resistance (Km) cassette in *rpoN* are indicated with striped bars. The sequencing strategy is shown beneath the restriction map.

tions (Fig. 1). The Tn5 insertions delineated a 5.3- to 5.7-kb complementing region on pNIT122.

The position of rpoN was determined by Southern hybridization of restriction digests of pNIT122 using the rpoN gene from *R. capsulatus* as a heterologous probe. The probe hybridized to an *EcoRI-PstI* fragment (data not shown and Fig. 1). Combining the results from the Tn5 insertional analysis and the Southern hybridizations, we concluded that rpoN is located at the right end of the complementing region and is transcribed from left to right (Fig. 1). Since RpoN has a molecular weight of about 54,000 (37), the coding capacity of the region delineated by the Tn5 insertions is in excess of what is needed to accommodate rpoN. This suggested that transcription of rpoN was dependent upon transcription through upstream genes.

Nucleotide sequence of the rpoN gene and upstream DNA. In order to characterize *rpoN* and the upstream sequences, the nucleotide sequence of the complementing region was determined according to the sequencing strategy depicted in Fig. 1. Six ORFs, transcribed in the same direction and preceded by plausible ribosome binding sites, could be distinguished. An analysis of the codon usage showed that all six ORFs could have a coding function (data not shown). The ORFs may encode proteins with molecular weights of 10,780 (ORF1), 25,407 (ORF2), 41,225 (ORF3), 40,664 (ORF4), 11,905 (ORF5), and 47,870 (ORF6). The intergenic region between the first five ORFs is small, varying from 0 bp (ORF2-ORF3-ORF4) to 9 bp (ORF1-ORF2) and 10 bp (ORF4-ORF5). The intergenic region between ORF5 and ORF6 is larger (100 bp). A 33-bp stretch of only C and G nucleotides containing a 24-bp inverted repeat is present in the last intergenic region. A stretch of T's following the inverted repeat, characteristic of a rho-independent terminator, is not present (47). The nucleotide sequence and the predicted amino acid sequences are shown in Fig. 2.

Southern hybridization experiments strongly suggested that ORF6 might be an rpoN gene of *R. sphaeroides*. A search of the Swiss-Protein data base, using the deduced amino acid sequence of ORF6 as the query sequence,

J. BACTERIOL.

1	AGAGCACGCCAACCGAGACCGCGGCCCGCCTTCAAGGGCGCCGCCGCGGCGATCCGATGACGATCCCCGTCAGGATCGAGCGAG
101	AGGTCCGCGCGCGCGGGGAAAGCTCCGGTGCTTCCGGCCAGTGCCGGTCTGACCGTCTGCCGGGTCCGTGCGAGCGCGAGGACGGGGGGGG
201	CCTGTCCGTCTGCTGCCTCCCCCCCCCCCCCCCCCGACCCGCAACCCCGTCATTGTCGGATTCCCTCCGACGGGCCGGCAAGCAGGGTCCTCCGCAACTCCCT
301	GCAGTGCCTTGCTGCCGGAACTTGGCACGCGGCGTGCTTCTCCCTCAGCGGGAGCCCGCGACACGC <u>AGGAG</u> AGTTCGGATGATCCAGATTACCCCCGCCG M I Q I T P A A
401	CTCAAGCCGCCATCAAGGGAGCCATCGAGGGCGCCGGCCG
501	CATGTCGCTGGAGCTGACGGAGGCCCCCGACGATCTGGTGGTGGAGGCCGAGGGGCCTGCGGGTGCTAATCGATCCGCAGAGCGGCACCTATCTGAACGGC M S L E L T E A P D D L V V E A E G L R V L I D P Q S G T Y L N G
601	$ \begin{array}{llllllllllllllllllllllllllllllllllll$
701	GTGGTGCGATGCTCGACGAAACCGGAAAGGCGCTGGATCTCTTCTTCAACCCGCGGGAACGCGGGCCGCTCGAGGCGGCCGATGCGGTGGGCACGGCGGG M L D E T G K A L D L F F N P R N A G P L E A A D A V G T A G
801	CAGCCTCGAGGTAGGCGACGCGATCCGGCTGATGCTGCGGATCGAGGCGGGGCCGCGGGGCGCGGGGGCGCGGGGGCGCGGGGCCCATGCCATC S L E V G D A I R L M L R I E A G R V A E A R F L A F G G A H A I
901	GCCTGCGGCTCGGCGCTGACGGTGCTGGTGACGGGCCTCGATCTCGCCGCGCGCG
1001	CCGCACCGCGGCGGCCGGCCGGCCGGGCCTGGTCGGCGCCTCCAGATCGCGCCGGCCTACGAGGGCCGGACCTTCGTGGCGCCCTGAGCCCGCGCC A P R R P A A A R A W S A L Q I A L A A Y E G R T F V A P E P A P
1101	CGTTCCCGCGCCGGCCCGGGCCCGGTCAGGCTGCTCGCGCCGAAGCACGACTCGCAGCCCGGATCGTCCGCGATGTGCCGCCGCGGAGGAG V P A P A A A P V R L L A P K H D S Q P R I V R D V P L A P A E E
1201	GCGCGCCTGATCGCCGAGGTGATCGAGAGCGTGCGGCGGCGGCGGCGGACGGGGGGGG
1301	TCCATCTGACCGGGGCCTGCTCCGGCTGCCAGCTCGCCGGCCG
1401	CATCCCCGAGGAGAAGCGGCCGCTCGTCTCCATCGCGGGGGCGCGATGATGGAGCGGGTCTATCTCGACAACAATGCCACGACCCGCCCG
1501	CGCTTCAGGCCATGCTGCCCTTCCTGACCGAGGAGTTCGGCAATCCCTCGTCGCTGCACGGGCAGGGCGCGCGC
1601	GCGCGCGGTGCTGGAGCTGATCGGCGCGAGGGCCGACAGCGAGACCGAGGCCGACACGACG
1701	GCGGCGGATCCGTCGCGGCGGAGATCGTGACCTCGACGGTCGAACATGCGGCCGTCCTCGCGCCTCTGCGACCATCTGGAGCGGCAGGAGGGGTGACGG A A D P S R R E I V T S T V E H A A V L A L C D H L E R Q E G V T V
1801	TGCACCGCATCCCGGTGGACGGCGGACGGCCGGCTCGACATCGAGGCCTATCGCGCGGCCCTCTCGCCCCGGGTGGCGCTGGTCTCGCTCATGTGGGCCAA H R I P V D G D G R L D I E A Y R A A L S P R V A L V S L M W A N
1901	CAACGAGACCGGCACGGCATTTCCCGTCGAGGGGCTGGCCGAGGCGGCGCATCGGGCGGG
2001	GTGCCCATAGTGCTGCGGGGACCGAGATCGACATGCTGTCGCTCAGCGCGCGC
2101	GCGTGCCGTTCCAGCCCCTGATCCGCGGCGGCAGGCAGCAGCGCGGCGGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG
2201	GGAGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
2301	GATCCGCTCGACCGGCTGCCCAACACCTCCTGCGTGGCCTTCGACTTCGCCGAGGGCGAGGCGATCGTGATGCTTCTCGACCGGGCGGG
2401	CGTCGGGTGCGGCCTGCGCCTCGGGCGGCGATGGAGCCGAGCCATGTGGCGCGCCATGAGGTGCCCTTCACCGCCGCGCATGGCGCGCGATCCGCTTCTC S G A A C A S G A M E P S H V I R A M K V P F T A A H G A I R F S
2501	GCTCTCGCACTGGACGACGGCCGAGATCGACCGCCTCCGAGGTGCTGCCGCCCCATCGTCGACCAGCTGCGCCGCGCTCTCGGCGCCGAG L S H W T T A A E I D R L L E V L P P I V D Q L R A L S P F G A E
2601	GAGGTGAAGTGATGTCGCCCCAGCAGCGGCGGGGGGGGGG

FIG. 2. Nucleotide sequence of the DNA fragment encoding ORF1, *nifUSVW-rpoN*. Amino acids are represented below the nucleotide sequence by the single-letter code. Putative ribosome binding sites are underlined. Asterisks represent stop codons.

2701	GCAGACGGCGGCGTGGCCGTCACCCCGCGAGAAGCGGGCCATCGCCGAGGCGTGCGCGCGC
2801	ATGGGCGAGGAAGAGGGGGCGGACATCCGCGGGGGGGGGG
2901	CCGCGCAGCGCACGGGGCGCGGGGCGGGGCGGGGCGGG
3001	CAAGGTCGAGAAGCTCGTGCGCCGGCTCGCTGCGCCGGTCACAAGGTGTCGGTGGGGGCCGAGGATGCCTCGCGCCGATCCGTTCTTTCT
3101	ATCGCCCATGTCGCGGCCGAGGCGGGGCGCGATCCGCTTCCGGACCGCTGGGCGTGGCCGGGCCGGGCCGGGCGGG
3201	TCGTCACGCGCTGCCCGCTGCCGGTGGAGTTCCACGGCCACAACGATCTGGGCATGGCCACGGCCAACGGCCACGCGCGCG
3301	CCTGTCGGTCACGGTGAACGGGCTGGGCGAGCGGGCGAGGGGGGCGCGCGC
3401	GCGCTGGGCCAGCTCTGCGCCCTCTCGGAGCTGGTGGCCCGCGCCTCGGGGCGCGCCGCCGCCGCGCGCGCGCGCCGC
3501	ATGAATGCGGCATCCATGTCGACGGGCTGATGAAGGACCGCGCCACCTACGAGAGCGCGGGACCTGCGCCCCGAGCGGTTCGGCCGCAGCCGCACCGCATCGC E C G I H V D G L M K D R A T Y E S A D L R P E R F G R S H R I A
3601	CATCGGCAAGCATTCCTCGGCCGGGCTCGCCCGGGGCGCGGCGGGCG
3701	COCGACTGGGCGGCCATCACCAAGCGCGCCGCCGCCGCCGAGGATCTTGCGGCGCTCCTTGCCGCGCCAAACCGCAAACCGCACGGAGGAGAACCGAAGAGAGATG R D W A A I T K R A A A P E D L A A L L A A Q T E T A R * M
3801	ACCCCGGGAACCGCCGTGCTCGAGGAGCTGAAGCGACTGTCCTCTCCCCGAGGAGGTCTTCGACGCCCTCGACCATCCCTACCGGCCGG
3901	A R L H I M K R L G Q Y L A A V D F A T L D F A D A R A A A R D A
4001	L S R A Y T D F V D S S P L E Q K V F K V F A K P S R A F V P L S
4201	G L S V V E D *
4301	M D M M Q F Q R Q T T Q L A M T Q R M Q E S L R I L TGCAGATGAGCAACGCCGATCTTCGCGGATCTTCGAGCGCGCCCCGAGGGGGGCCCCGAGGGGGGCCCCGGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGGCCCCCGAGGGGCCCCCGAGGGGCCCCCGAGGGGCCCCCGGCGCCCCCC
4401	Q M S N A D L A D Y L T A Q A L E N P C L E V R V P E G A S V A P
4501	A L P S R G I Q A G L D R D A F A T V E G Q P P S L L A H V E A Q
4601	I D L A F F D P G D R R T A L A F A E A L E P S G W L G Q P V S E V TCGCCGCCGGGCCGAGGTGGAGGAGGAGGAGGAGGGGCGCTGGCCGGCGGTCGCCGGGCCGTTGGAGGCCGGCC
4701	A A A E V E E E A L V I L E R L Q A L E P A G L F A R S L A E ATGCCTCGCGCTGCAGCTCGAGGATCTGGGGGCTGCTGCGCGCGC
4801	C L A L Q L E D L G L L T W E L R T M L D H L P L L A E G R I A D CTCGCCCGCCGCCGCCGCGCGCGCGGCGGGCGAGGCCTGGGCGGGGGGGG
4901	LARRCDCEPEHIRENLALIRSLSPKPGEAFAADR GCACGCCGAATCCAGCCGCCGCAGGGGCGGGGGGGGGGG
5001	T P I Q P P D V R V L R G P E G W E V E L T R A Q L P R I R V S E GGCAGGGGACACCGGCGACGGCCGACGGCCGGCCGGCCG
5101	A G D T G D R Q A D A W L A R A R S Q A R W L E R A V E R R Q A T CTCCTGCGCACCGCCGTCTGCCTCGTGCGCCATCAGGCCGACTTTCTCGATCAGGGGCGCGCGC
5201	L L R T A V C L V R H Q A D F L D Q G P R A L R P L S M E E V A L E AACTCGATCTCCATCCCTCGACCATCAGCCGCCGCCACGGCCGACGAGACGCCGCGCGCG
5301	GTCTCCCCGAGGGCCCGAGGCGCGCAGGCAGGCGCGAGGACGCCGC
5401	GACGATGCGATCGTGAAGCAGGCGAAGCTCGCGGGGCGCGGGGTCCCGGCGCACCGTCACCAAATATCGCGAGACGCTGGGGATCCCCTCCTCTCACG D D A I V K Q A K L A G A V L A R R T V T K Y R R T L G T P S S V D
5501	ACCGCAAGCGCCGCCGCCGCGCGCCTGACGCAGGGCAGG

FIG. 2-Continued

revealed extensive similarities with other RpoN proteins. We therefore identified ORF6 as *rpoN*.

RpoN. The *rpoN* gene has three possible initiation codons. Only the first initiation codon is preceded by a possible ribosome binding site, and it was chosen as the start of the rpoN coding sequence (Fig. 3). The RpoN proteins from the two purple nonsulfur bacteria are more closely related to each other than to any other RpoN protein (51% identity). The similarity with the other RpoN proteins in Fig. 3 did not exceed 35% identity. With the exception of the R. capsulatus protein, which has a pI of 6.8, the RpoN proteins all exhibit a pI of <5. This is also true for R. sphaeroides (pI 4.9). Merrick et al. (41) identified three regions of similarity in the RpoN sequences from several sources. Region I (residues 1 to 68) is characterized by a high Gln content; region III (residues 69 to 434) contains a helix-turn-helix motif (residues 315 to 340) and a stretch of highly conserved residues, the so called RpoN box (residues 408 to 417). The poorly conserved spacer region between regions I and III is virtually absent in the R. sphaeroides and R. capsulatus sequences.

ORF1, nifU, nifS, and nifW. Searches of the Swiss-Protein data base using the ORFs upstream from rpoN as query sequences led to the tentative identification of nifU (ORF2), nifS (ORF3), and nifW (ORF5). The alignment of the ORF1, NifU, NifS, and NifW proteins from R. sphaeroides with those from other sources is shown in Fig. 4. ORF1 displayed a high degree of similarity to ORF6 of the Azotobacter vinelandii nitrogen fixation cluster (7, 29). NifU proteins from other diazotrophic organisms have seven conserved Cys residues, which are present as Cys-X-X-Cys and Cys-X-Cys motifs (3, 7, 8, 11, 29, 44). The R. sphaeroides NifU protein was poorly conserved compared with the other NifU proteins (Fig. 4B). In fact, only three Cys residues, two of which are also conserved in the other NifU proteins, can be found in the R. sphaeroides NifU protein. In the NifU proteins from A. vinelandii and Klebsiella pneumoniae, a consensus sequence (Cys-X-Cys-His) typical of a hemebinding site of cytochrome c proteins is present (40). The heme-binding site is present in a region completely absent from the NifU protein of R. sphaeroides.

In contrast to the NifU proteins, the NifS proteins displayed a high degree of similarity throughout the amino acid sequence (Fig. 4C).

ORF1 upstream from nifU displayed a high degree of similarity with an ORF in the nitrogen fixation clusters from an *Anabaena* sp. (11, 42) and *A. vinelandii* (7, 29), especially at the carboxyl terminus (Fig. 4A). This ORF is found at the same position in *A. vinelandii*, upstream from *nifU*. Starting at residue 97 in the *R. sphaeroides* sequence, a motif (G-X-X-G-X-G-K-S) characteristic of a nucleotide-binding site (22) was present in a highly conserved region of the protein.

Analysis of NifV. The data base searches were used to identify ORF4 as *nifV*. Besides being similar to other NifV proteins, NifV is similar to isopropylmalate synthase (LeuA), as previously noted (14). Inspection of an amino acid sequence alignment between the NifV and LeuA proteins reveals two regions of high similarity (Fig. 5, regions 1 and 2). To determine whether these conserved regions are also present in other proteins, region 2 in the sequence alignment was used to construct a profile as described by Gribskov et al. (19a). The Swiss-Protein data base was searched in order to find proteins matching the region 2 profile. By using this technique, a data base can be searched with a sequence alignment, a procedure which is more

Pyruvate carboxylase and oxaloacetate decarboxylase were found to contain a sequence similar to region 2 of NifV. In addition, a sequence similar to region 1 was present at the proper spacing relative to region 2, compared with the sequence alignment in Fig. 5. The sequence between regions 1 and 2 was not highly conserved.

R. sphaeroides rpoN mutant. To determine the role of rpoN in the metabolism of *R.* sphaeroides, an rpoN mutant strain, EM1, was constructed. The rpoN mutation did not prevent diazotrophic growth of *R.* sphaeroides EM1. *R.* sphaeroides EM1 was not a glutamine auxotroph, nor was growth on malate impeded. These phenotypes are generally associated with a rpoN mutation. In the Southern hybridization used to verify the construction of the rpoN mutation (Fig. 1), a 4-kb band could be seen in addition to the 1.2-kb band representing rpoN (data not shown). The Southern hybridization was repeated at a lower stringency (60°C), using a 0.5-kb SacI-BamHI fragment internal to rpoN and lacking the variable 3' end of the rpoN gene as a probe. Again, 1.2- and 4.0-kb BamHI fragments hybridized to the rpoN probe (Fig. 6).

R. sphaeroides nifUS mutant. The functionality of the nifUSVW-rpoN operon was determined by deleting part of the nifUS genes and replacing them with an Ω element (Fig. 1). Since the Ω element is flanked by transcriptional terminators, it has a polar effect on the transcription of the genes downstream (48), nifVW-rpoN. Therefore, any phenotype resulting from this insertion may be caused by either the deletion of nifUS or the lack of transcription of downstream genes. R. sphaeroides EM11 and three other independently isolated mutants displayed strongly reduced diazotrophic growth. The R. sphaeroides EM11 colonies appearing after 1 month of incubation were very small and only slightly pigmented compared with the wild-type colonies. The growth of the mutant strain was indistinguishable from that of the parental strain on media containing NH₄⁺.

DISCUSSION

In contrast to the situation in other bacteria, where the rpoN gene is flanked by two conserved ORFs of unknown function, the rpoN gene of R. capsulatus is located between nif genes. The present study shows that this is also true for another purple nonsulfur bacterium, R. sphaeroides. In R. capsulatus, the nifAB genes are duplicated. The genetic organization of region A is nifUSVW-nifAB, and the organization of region B is nifHDK-nifU-rpoN-nifAB (1, 33, 34, 39). A duplication of a nifUSVW-rpoN-nifAB cluster and subsequent deletion of nifSVW-rpoN in region B and deletion of *rpoN* in region A could explain this organization in R. capsulatus. Furthermore, it would explain the unusual organization in region B, where a nifU gene is present in the absence of nifSV. Our finding that a nifUSVW-rpoN cluster actually exists in a bacterium belonging to the same genus lends credence to this hypothesis.

Insertions of Tn5 in the *R. sphaeroides nifUSVW* genes located on pNIT122 prevented complementation of the *rpoN* mutation of *R. capsulatus* (31, 60) by this plasmid. This strongly suggests that transcription of *rpoN* of *R. sphaeroides* is dependent on a promoter upstream from *nifUSVW*. The absence of a recognizable terminator structure between *nifW* and *rpoN* further supports this conclusion. The cotranscription of *rpoN* with upstream genes has not been found in other bacteria. The promoter of the *R. capsulatus rpoN* gene was mapped downstream from *nifHDK* via

	2000 2000 000 000 000 000 000 200 000 200 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 0	
Rsp	MDMMQFQRQTTQLAMTQRMQESLRILQMSNADLADYLTAQALENPCLE	62
Rca	MELAQTLSQRQTMQMAGQMLHSLAILGMSSQDLSEHLTEQATSNPFLTPAFI	56
Avi	MKPSLVLKMGQQLTMTPQLQQAIRLLQLSTLDLQQEIQEALDSNPMLERQEDAEDYDSPDMLGEHGDQSTLDTTPGSYQEGYESGAASEDGGILE	95
Ppu	MKPSLVLKMGQQLTMTPQLQQAIRLLQLSTLDLQQEIQEALESNPMLERQEDGEDFDNSDPMADNAENKPAAEVQDNSFQESTVSADNLE	90
Kpn	MKQGLQLRLSQQLAMTPQLQQAIRLLQLSTLELQQELQQALDSNPLLEQTDLHDEVETKEAEDRESLDTVDA	74
Tfe	MKQGLELKLGQHLAMTPQLQQAIRLLQLSTVDLQQEVQGMLESNPLLDEETGDEGGGGGPIPFIVELPSEERQLD	74
Rme	MALSASLHLRQSQSLVMTPQLMQSIQLLQMNHLELSHFTAQEVEKNPLLEVQPADEPTISDREDAGPHPAETGGETDEAAG-QSDL1DSAPSKS-G	94
Bjal	MALTORLEFROSOSLWISPOLMOAIKLLOLSNIDLMTFVEEELECNPLLERASDDAAGAEAPTE	100
Bja2	MALTQRLEFRQSQSLVMTPQLMQAIKLLQLSNLDLTTFVEEELEKNPLLEKANDEASGGEAPAEASQFSDSDGGINDEFGGGFGEAFEFGQLEWHSKDLG	100
Den		118
Rsp	SR	109
Avi	ALFOLDER TO SET PUT AWE DIVOTSASNI PST PERPENDENTPRST GESLOSHI WOLLNI. PMSDTDRI. LAVTL IDS SOU	177
Pnu	DEPUSER I DUEL PUT AWEDIVOTSASSI ESNDDEWDFTTRTSAGESI OSHI JUULNIA PASDTDRI I AVTLIDSINGOGY	172
Kon	OKEMPEEL PLDASWDEIYTAGTPSCNGVDYDDDEL PYYOGETTOSLODYLMOVELTPFTDTDRAIATSIVDAVDDTGY	153
Tfe	LAA-ENILPDELFVDSOWDDIFDMGTSGSGNGSDEDLFDFESRNSRTQSLQDYLRWQADMTHFTADERNMAELIIDAIDERGY	156
Rme	ERLSEGLDADFANVFPDDTAPORADAPELLGOWKSMPGAGDAEGYDLDDFVGGRKTLRETLAEQLPFALSAVSDRLIARYFIDQLDDAGY	184
Bja1	VASETTLSDHLAEQVRDARDGAMTTYTEWGGGGSGDEDYNLEAFVASETTLSDHLAEQLSVAFTAPAQRMIGQYLIDLVDEAGY	147
Bja2	$TRAE {\tt IEQTLDTGLDNVFSEE} PAEAAARNAQDAAPTTYTEWGGGASGDEDYNLEAFVAAEVTLGDHLAEQLSVAFTAPAQRMIGQYLIDLVDEAGY$	195
•		
Rsp	LGQPVSEVAAAAEVEEEEALVILERLQALEPAGLFARSLAECLALQLEDLGLLTWELRTMLDHLPLLAEGRIADLARRCDCEPEHIRENLA	209
Rca	LGQSLDSIALAAGVSLSRAESMLAVLQGFEPTGLFARDLSDCLILQAREADILTWEVETLIRNIRLIAENRLSDLADLCDCDIGDIPEIIK	200
Avi	LEAALEEILASLDPELGVELDEVEMVLRRIQQFEPAGIAARDLSESLLLQLRQLP-PDTPWLEEAKRLAKDYLDLLGARDFTQLMRRMKLKEEELRPVIE	276
Ppu	LEDTLEEISAGFDPELDIELDEVEAVLHRIQQFEPAGVGARNLGECLLLQLRQLP-ATTPMMTEAKRLVTDFIDLLGSRDYSQLMRRMKIKEDELRQVIE	271
Kpn	LTISVED-IVESIGDDEIGLEEVEAVLKRIQRFDPVGVAAKDLRDCLLVQLSQFA-KETPWIEEARLIISDHLDLLAHHDFRSLMRVTRLKEEVLKEAVN	251
Tfe	LADSLEDLAATMAVQEDALLAVLLRVQDFDPPGVGARNLSECLLLQLKQMVVEKDDAHVLLAQRIVKDDLQALGKHDIPRLCIVLGVDEAALKAATA	275
Rme	LHADLAETAETLGAAGEUVARVLHVLQVFDPPGVFARTLGEULAIQURARNKLDPAREALVARLELLARKDFASLAKIGVDEEDLIDTLA	273
Bjai Bio2	LPPDLGQAAERLGATQEDVEHVLAVLQEFDPFGVCARNLKGLLATQLKELDKIDPARQALVEHLDLLARNDIASLKALGADDEDIADHID I DED	230
bjaz	FEERANNERFORMAN AFAR AFAR AFAR AFAR AFAR AFAR AFAR AF	200
	0 11 11 11 11 11 11 11 11 11 11 11 11 11	
Rsp	LIRSISPEGEAFAADRTPTOPPDVRVI.RGPE-GWEVELTRAOLPETRVSEAGDTGDROADAWLARARSOARWLERAVERROATLLRTAVCLV	301
Rca	OIRHLNYKYGLAFDHOPTPYFPPDLIAVRGAE-GYTVELNRATSPTITYREDRFADGTADAKARAERRRRGRGPGAGEALERRRDTLLRTAAVLV	294
Avi	LIOSLNPRPGAOIESSEPEYVYPDVIVRKHNDR-WLVELNOEAVPRLRINPHY-AGFIRRADASADNT-FMRNOLOEARWFIKSLOSRNETLMKVSTQIV	373
Ppu	LVQSLNPRPGSQIESSEPEYVVPDVIVRKDSDR-WLVELNQEAIPRLRVNPQY-AGFVRRADTSADNT-FMRNQLQEARWFIKSLQSRNETLMKVATRIV	368
Kpn	LIQSLDPRPGQSIQTGEPEYVIPDVLVRKVNDR-WVVELNSDSLPRLKINQQY-AAMGNSTRNDADGQ-FIRSNLQEARWLIKSLESRNDTLLRVSRCIV	348
Tfe	LISALNPKPGEDVGTESTEYVIPDVIVRWAGSR-LRTDLNPEAMPKLRINRHY-ADMAGGKDAAHK-YIQDQLNEARWFIKSLQSRQDTILKVARAIV	347
Rme	EIRKLDPKPGTSFETGVFEAIIPDVVVRAAPDGGWLVELNPDALPRVLVNHDYFTEISRSSRKNSGEQAFLNECLQNANWLTRSLDQRARTIMKVASEIV	375
Bjal	ELRRLSPKFGMKFGSARLQTMVPDVYVRPAPDGGWHVELNSDTLPRVLVNQTYYSKLSKKIGKD-VDKSYFNDALQNATWLVRALDQRARTILKVATEIV	337
Bja2	EIRRLNPKPGMKFGAARLQTMVPDVYVRPGPDGGWHVELNSDTLPRVLVNQTYYSELSKKIGKD-GDKSYFTDALQNATWLVRALDQRARTILKVATEIV	385
_		
Ksp	RHQADFLDQGFRALRPLSMEEVALELDLDFSTISRATATRLIETFRGLIPLRAFFSRSVSSDGPEAPQSQDALMALVREIIARLDRTKFFSDDAIVKQAK	401
KCa Acci	ARQSAFLDKGPAHLVPLTLEDVASELGLHASTISKAVSGKMIQTQTKALPLKAFFSKAVSTQGGEAVSKD5L-DFVQKIWAAKIRQNPLSDDAIVILAE	393
AV1 Prov	αι τος του του του του το του το του του του τ	467
r pu Km	Engroup unduration for the average of the state of the st	407
Tfo	EQUALITE SUCCEPTION DUT DI LA ALVENIESTI DAVI TUNI LINSTATI ELSI TUTE VETTO SULVITE SOLVATORO SUL ALTERI ALVENITA MALTERIA ALVENITA	446
Rme	ENGLEFF HUGT SCHEMMENTEN IND VERTIGET VORVERGENTET FOR VERDELER FURDER VORDER SCHEMMENTEN VERDELER VORVERGEN VERDELER VERDELER VERDELER VORVERGEN VERDELER VORVERGEN VERDELER VORVERGEN VERDELER VERDEL	475
Ria1	ROONGPETI GVAHL PRIMI KAVAFA TOMESTVENT DINATALI TANGTELI VEFTAST PSADGGARSAFAVRHR IKOLI ESEEPSAVLSDDA IVERLR	437
Bia2	RODGEFTHGVAHLRPLNIKAVANA IOMESTVSEVTANKYMATNEGTFELKYFFTASIASADGGEABSAEAVRHHIKOLIDSEAPAALLSDDTIVERLR	485
-54-		
Rsp	LAGAVLARRTVTKYRETLGI PSSYDRKRAAAAA	434
Rca	RAGLRIARRTVAKYRSTLGLASSYERRRAAAAR	426
Avi	EQGIQVARRTVAKYRESLSIAPSSERKRLM	502
Ppu	AQGIQVARRTVAKYRESLGIAPSSERKRLM	497
Kpn	DQGIMVARRTVAKYRESLSIPPSNQRKQLV	477
Tfe	DQGIQIARRTVAKYREAANVPPASQRRRL	475
Rme	QAGVDIARRTVAKYREAMSIPSSVQRRREKRALPRPRDSERCRQAASA	523
Bja1	VSGIDIARRTVAKYREAMRIRSSVQRRRDNMWSTMINRASGGTGLDK	484
BjaZ	ASGIDIARKTVAKIREAMRIPSSVQRRDKQSALGNVLSTAMSDRSKNPEPA	557

FIG. 3. Alignment of the R. sphaeroides (Rsp) RpoN amino acid sequence with those of other bacteria. Abbreviations: Rca, R. capsulatus (31); Avi, A. vinelandii (41); Ppu, Pseudomonas putida (27); Kpn, K. pneumoniae (42); Tfe, Thiobacillus ferrooxidans; Rme, Rhizobium meliloti; Bja1, B. japonicum RpoN1 (36); Bja2, B. japonicum RpoN2 (36). Symbols: \Box identical residue in at least eight sequences; \bigcirc , conservative substitution in at least eight sequences according to the scheme PAGST, QNED, ILVM, HKR, YFW, C. Numbers at the right refer to amino acid residues.

A)		
_	11 0 011 11 0 0 000 11 0000 11 0 0 0 0	
Rsp	M-IQITPAAQAAIKGAIEGAGQPVAGLRLMVQSGGCAGLKYGMSLELTEAPDDLVVEAEGLRVLIDPQSGTYLNGVTIDFVTSLEGTGFVFDNPNA	95
AV1	M-11LTESAKSAVTRFISSTGKFIAGLRRVEGGGCSGLKYSLKLEEAGAEDQLVDCDGITLLIDSASAPLLDGVTMDFVESMEGSGFTFVNPNA	95
мар	MIVILIERAEFRERAFERSAKDANEITKGIKVSVKDGGCSGYEYLMDVTSQPQPDDLVTQQGSVLVYVDAKSAPLLEGIVIDFVEGLVESGFKFTNPNA	100
_		
Rsp	KGGCGCGKSFC	106
Avi	TNSCGCGKSF-AC	107
Asp	TSTCGCGKSPKGG	113
B)		
Avi	MADYSEKVKEHFYNPKNAGAVEGANAIGDVGSLSCGDALRLTLKVDPETDVILDAGFQTFGCGSAIASSSALTEMVKGLTLDEALKISNQDIAD	94
Kpn	MANYSEKVKDHFFNFRNARVVDNANAVGDVGSLSCGDALRLMLRVDPQSEIIEEAGFQTFGCGSAIASSSALTELIIGHTLAEAGQITNQQIAD	94
Asp	MADYTOKVLELFYDPKNQGVIEENGEPGVKVATGEVQSIACGDALRLHIKVEVESDKIVDSRFQTFGCTSAIASSSALTEMIKGLTLDEALKVSNKDIAD	100
кар	MLDEIGKALDLFFNPRNAGPLEAADAVGTAGSLEVGDAIRIMLRIEAGRVAEARPLAFGGAHAIACGSALTVLVTGLDLAAARAVTPEEIEA	92
	○ 2222 ○ ○ 22 ○ ○ ○ ○ ○ ○ ○ 22 ○ ○	
Avi	YLDGLPPEKMHCSVMGREALQAAVANYRGETIEDDHEEGALICKCFAVDEVMVRDTIRANKLSTVEDVTNYTKAGGGCSACHEAIERVLTEELAA	189
Kpn	YLDGLPPEKMHCSVMGQEALRAAIANFRGESLEEEHDEGKLICKCFGVDEGHIRRAVQNNGLTTLAEVINYTKAGGGCTSCHEKIELALAE	185
Asp	YLGGLPEAKMECSVMGQEALEAAIYNYRGIPLAAHDEDDEGALVCTCFGVSENKVRRIVIENDLTDAEQVTNYIKAGGGCGSSCLAKIDDIIKD	193
Rsp	AVGGLPAFRRPAAARAWSALQIALAAYEGRTFVAPEPAPVPAPAAAPVRLLAPKHDSQPRIVRDVPLAPAEE	164
Avi	RGEVEVAAPIKAKKKKKVKVLAPEPAPAPVAEAPAAAPK LSNLORTERIETVL AATEPTLORDKGDVELIDVOGKNVVVLTGACTGCOMSMTLG GIO	286
Kpn	ILAOOPOTTP. AVASGK DPHNOSV, VDTL AFLICPHTOADGGMAILSVTNHOVTVSLSSSCSGAMTINTIA LIG	258
Asp	VKENKAATNI NNKGSKPTNI PNSGKRPI. TNVOKTAL I OKVI. DEEUR PVI. LADGGDVEL VDVGDI VKUVI. GACGSCSSSTATI KTATE	284
Rsp	ARLIAEVIESVRPRI.RADGEDVTI.VAVESKVRVHI.TGACSGCOLAALTIG GLO	218
-		
A1		
NV1 V		312
Age		274
Pen		300
veb	KRLAD I LOKF I KVI FEERREL VOI AGAR	24/
C)		
Rsp	MERVYLDNNATTRI.APEALOAMI.PFI.TEEFGNPSSI.HGOGRAPARAI MAARRAVI.EI.TGARADSETI.FTSGGTFADTTATRSAI.AADPSRRETVTSTVEH	
-		100
Avi	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVGMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH	100 100
Avi Asp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVQMALKKARQSVQKLLGAEHDSEILPTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH	100 100 99
Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH	100 100 99 97
Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH	100 100 99 97
Avi Asp Kpn Rsp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH AAVLALCDHLEROEGTYWEIPDGOGLDIEAYRAALSPRVALVSLMMANNETGTVEPVEGLAELAHRAGALFHTDAVOAVGKVPIVLRGTEIDMLSLS	100 100 99 97 200
Avi Asp Kpn Rsp Avi	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVQAALKKARQSVQKLLGAEHDEEILFTSCGTESDSTATLSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH COCCUPACIENT COCCUPACIENT COCUPACIENT COCUCUPACIENT COCUCUCUPACIENT COCUCUCUPACIENT COCUCUPACIENT COCUCUPACIENT COCUCUCUCUPACIENT COCUCUCUCUPACIENT COCUCUCUCUCUCUCUPACIENT COCUCUPACIENT COCUCUCUCUCUCUCUCUCUCUCUCUCUCUCUCUCUCUC	100 100 99 97 200 199
Avi Asp Kpn Rsp Avi Asp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLHSFGNQVQAALKKARQSVQKLLGAEHDEEILFTSCOTESDSTATLSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCOTESDSTATLSALKAQPERKTVITTVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH AAVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSPRVALVSLMWANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDIEHYASLLTDDVAVVSVMWANNETGTLFPIEEMARLADDAGIMFHTDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKOLETO.GYTVYLSVNSHGOLDLDELEASLTGNTALVYIMVANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVFIDLMMKTSTIDMLTIS	100 100 99 97 200 199 198
Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALSPRVALVSLMMANNETGTVPPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHRLFVDKKGRLDLEHTASLLTDDVAVVSVMAANNETGTVPPVEGLAELAHRAGALFHTDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVHXLSVNSHGQLDLLELASLTGNTALVTIMYANNETGTVPPIEEIGKRVKERGAIFHTDAVQAVGKVPIDLKNSSIHMLSCS PATLAACEHLERQ.GYTHILAVDSEGALDMAQFRAALSPRVALVSVMAANNETGTVFPIEBIGKRVKERGAIFHTDAVQAVGKVPIDLKNSSIHMLSCS	100 100 99 97 200 199 198 196
Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AAVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSPRVALVSLMWANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLFVDKKGRLDLEHYASLLTDDVAVVSVMAANNETGTVFPIEEIGKRVKERGAIFHTDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKIPLNMKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMANNETGVFPIEEIGKRAKERGAIFHVDAVQAVGKIPIAVGQTRIDMLSCS	100 100 99 97 200 199 198 196
Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAQPAKRHIITTQVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLFERREIITSVVEH AKVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSPRVALVSLMMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKIPVDGDGRLDIEAYRAALSPRVALVSLMMANNETGTVFPIEEIGKRVKERGAIFHTDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLEHXASLITGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSC PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMMANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSC PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMMANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDAVQQTRIDMLSCS	100 100 99 97 200 199 198 196
Avi Asp Kpn Rsp Avi Asp Kpn Rsp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLHSFGNQVGMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAPERKTVITTVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH AAVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSFRVALVSLMWANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKIPVDKKGRLDLEHYASLLTDDVAVVSVMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSC PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMANNETGVFPIEEIGKRVKERGAIFHCDAVQVVGKIPIAVGQTRIDMLSCS	100 100 99 97 200 199 198 196
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGGQLGKAVRTAREQVAALLGAEDSEILFTSCGTESDSTAILSALKAQPERKTVITTVVEH MKQVYLDNNATTKLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH B B B B O O DO B BDO DO BOB BOOKENEH ONDO AAVLALCDHLERQEGVTVHRIFVDGDGRLDIEAYRAALSPRVALVSLMWANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVVSVMANNETGTLFPIEEMARLADDAGIMFHTDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHQQLDLDLELASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMANNETGVLFPIEEIGKRVKERGAIFHVDAVQAVGKIPIAVGQTRIDMLSCS OGS BOOKED BOOKED B BOOKED B B DOOKED BOOKENEH ONDO B B DOOKED ON B B OOK ON B B OOK O AHKFHGPKGVGALWLRKGVPFQPLIRGGRQQRGHRAGTENIPGIVGLGRAAELALGGDHGAVRLLRDRLEQGILARVPKARVLGDPLDLPNTSCVA GHKLHAFKGVGVLYLLRGTRFRELLRGGHORGGRAGTENIASIIGLGVAAERALODMEHENTEVALIEDUT EAGILAVUPHARVTGPDDUD DHTANTA	100 100 99 97 200 199 198 196 297 299
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVEH MSVIYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWPRATPRHAAIALLPERREIITSVVEH AAVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSPRVALVSLMAANNETGTVPPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVVSVMAANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTYTYLSVNSHGQLDLDELEASLTGNTALVTIMTANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKIPLNMKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVLFPIGEMAELAHEQGALFHCDAVQVVGKIPIAVGQTRIDMLSCS OGD BOOGED BDOGOG B BEDGED BOOGED BOOGED B BOOGED OF BOOGED OF BOOGED OF BOOGED OF BOOGED OF A AHKFHGFKGVGALMIRKGVFPQLIRGGRQQRGRAGTENIFGIVGLGKAAELALGGDHGAVRLLDRLEQGILARVFKARVLGPPLDRLPHTSCTA GHKLHAPKGVGALVIRRGVFRFLLIGGGHQERGRRAGTENIFGIVGLGKAAELAL.HITETATKKTETIDNLEAGILARVPKARVLGPDLDRLPNTANIA GHKLHAPKGVGALVYRGVFRFLLIGGGHQERGRRAGTENIFGIVGLGKAAELALITHETATKKTETIDEIFOT LAKT PLOCKUNGDITTDI	100 100 99 97 200 199 198 196 297 299 298
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDEEILFTSCGTESDSTATLSALKAQPERKTVITTVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGQQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTESDSTATLSALKAQPERKTVITTVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH AAVLALCDHLERQEGVTVHRIFPDGOGRLDIEAYRAALSPRVALVSLMAANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLNVCKQLETQ.GYTYTYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTYTYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKIPLNMKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVLFPIGEMAELAHEQGALFHCDAVQVVGKIPIAVGQTRIDMLSCS OGE GOODE GOOGO GOODE GOODE GOODE GOODE GOODE GOODE OCOOD GOODE OCOO GOODE GOODE OCOOD GOODE	100 100 99 97 200 199 198 196 297 299 298 295
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSMHTFGQQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTESDSTATLSALKAQPERKTVITTVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMFFLTDFYGNDGORLDIEAYRAALSPRVALVSLMAANNETGTVPPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVVSVMAANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTYYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKIPLNMKTSTIDMLTIS PATLAACEHLERQ.GYTHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVLPPIEEIGKRVKERGAIFHVDAVQAVGKIPLNMKTSTIDMLTSS AHKFHGFKGVGALMLRKGVPFQELIRGGRQQRGHRAGTENIPGIVGLGRAAELALGGDHGAVRLLRDRLEQGILARVFKARVLGDPLDRLPNTSCVA GHKLHAPKGIGALVVRGVRFRPLLIGGHQRGRRAGTENNPGIVGLGRAAELALGGDHGAVRLLRDRLEQGILARVFKARVLGDPLDRLPNTSCVA GHKLHAPKGIGALVVRGVRFRPLLIGGHQRGRRAGTENVPGIVGIGVGLGKAAELLIHIETAIKKETRLRDRLEQTLLAKIPDCEVNGDITQRLPNTNIG AHKFHGFKGVGCLYLRRGTRFPLLIGGHQEKGRRAGTENVPGIVGIGVGAACELANIHL.PMTHIGQLRNRLEHRLLASVPSVMVMGGQPRVPGTVNLA	100 100 99 97 200 199 198 196 297 299 298 295
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBFFGNQVGMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBFGNQVGMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDPDVVEAIMPYLTDYYGNPSSHFTFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAPPAKRHIITTQVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLFERREIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLFERREIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLFERREIITSVVEH AAVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSPRVALVSLMMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHRIPVDGDGRLDIEAYRAALSPRVALVSLMMANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIVLRGTEIDMLSLS PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLEHASLITGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSCC PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLEASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSTHMLSCS PATLAACEHLERQ.GYRTHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSTHMLSCS PATLAACEHLERQ.GYRTHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKIPIAVGQTRIDMLSCS PATLAACEHLERQ.GYRTHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKIPIAVGQTRIDMLSCS AHKFHGPKGVGALMURKGVPFQPLIRGGRQQRGRRAGTENIFGIVGUGRAAELALGGDHGAVRLLRDRLEQGILARVFKARVLGDPLDRLPNTSCVA GHKLHAPKGVGVLYLRRGTRFPLLRGGHQERGRRAGTENIFGIVGUGVGAACELANIHL.GGDHGAVRLLRDRLEQGILARVFKARVLGDPLDRLPNTANIA GHKLHAPKGVGCLYLRRGTRFPLLRGGHQERGRRAGTENVFGIVGUGVGAACELANIHL.FGMTHIGQLRNRLEHRLLASVPSVMMGGGQPRVPGTVNLA	100 100 99 97 200 199 198 198 196 297 299 298 295
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Kpn Rsp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDPDVVEAIMPYLTDYYGNPSSHFFGGQLGKAVRTAREQVAALLGAD.ESEIVFTSCGTEGDNAAIRAALLAPPAKHIITTQVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH AAVLALCDHLERQEGVTVHRIPVDGDGRLDIEAYRAALSFRVALVSLMMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHXASLLTDDVAVVSVMANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVHYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVHLIPODKGRQLDDAELASITGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSCS OGG SOCIOL SOCIOL SOCIOLOGO SOCIONALSPRVALVSVMAANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKIPINAKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVFPIEEIGKRVKERGAIFHCDAVQVVGKIPIAVGQTRIDMLSCS OGG SOCIOL SOCIOL SOCIONALSPRVALSPRVALVSVMAANNETGVFPIEEIGKRVKERGAIFHCDAVQVGKIPIAVGQTRIDMLSCS GHKLHAPKGVGVLYLRRGTRFPLLRGGHQERGRAGTENIFGIVGLGRAAELALGGDHGAVRLLRDRLEQGILARVFKARVLGDPLDRLPNTSCVA GHKLHAPKGVGVLYLRRGTRFPLLRGGHQERGRAGTENIFGIVGGIVGGAAELALIGGDHGAVRLLRDRLEQGILARVFKARVLGDPLDRLPNTANIA GHKHAPKGJGALVYRGVGRFPLLIGGHQERGRRAGTENIFGIVGGIVGAGAACELANIHL.FCMTHIGQLRNRLEHRLLASVPSVMVMGGQPRVPGTVNLA G O EDDEDO DO ED EEGOODOCHDEGOED SOEIDOCHDEGOOD OGE OO OG DO EDOCHD FDFAEGEAIVMLLDRAGICVSSGAACASGAMEPSHVIRAMKVFFTAAHGAIRFSLSHVTAAEIDRLLEVLPPIVDQLRALSFFGAEE.VK	100 100 99 97 200 199 198 196 297 299 298 295 387
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDPDVVEAIMPYLTDYYGNPSSHHTFGQQLGKAAVRTAREQVAALLGAEJPSEIIFTSWPRATFRHAAIALLPERKEIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAALLGAEYPSEIIFTSWPRATFRHAAIALLPERKEIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAALLGAEYPSEIIFTSWPRATFRHAAIALLPERKEIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAALLGAEYPSEIIFTSWPRATFRHAAIALLPERKEIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAALLGAEYPSEIIFTSWPRATFRHAAIALLPERKEIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAALLGAEYPSEIIFTSWPRATFRHAAIALLPERKEIITSVVEH MKQVYLDNNATTRUDFMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAALLGAEYPSEIIFTSWPRATFRHAAIALLPERKEIITSVEH MKULLCHLERQEGVTVHRIFVDGGGRLDIEAYRAALSPRVALVSLMMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKIPVDKKGRLDLEHYASLITDDVAVVSVMANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIULKSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLDELEASITGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLDELEASITGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKIPINKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGVLFPIGEMAELAHEQGALFHCDAVQVVGKIPIAVGQTRIDMLSCS OGD BOOGOG BOOGOG BOOGOGO BOOGOGOGOGAGAENE OOGOGOG BOOGOGO OG BOOGOGO O AHKFHGPKGVGALWLKKGVPFQPIIRGGRQQGGRAGTENIPGIVGGLGKAAELALGGDHGAVRLLRDKLEAGILAVVPHARVKARVLGDPLDRLPHTANIA GHKHAPKGIGALVVRRGVFRPLLIGGHQERGRAGTENNPGIVGGLGKAAELALGGDHGAVRLLRDKLEAGILAVVPHARVKARVLGDPLDRLPHTANIA GHKHAPKGIGALVVRRGVFRPLLIGGHQERGRAGTENNASIIGUGVGGAACELANIHL.FCMTHIGQLRNRLEHRLLASVPSVMVMGGQGPRVPGTVNLA GHKHAPKGIGALVVRGVGRFRPLLIGGHQEYGRRAGTENICGIVGMGAACELANIHL.FCMTHIGQLRNRLEHRLLASVPSVMVMGGQPRVPGTVNLA GHKHAPKGIGALVVLLDRAGICVSSGAACASGAMEPSHVIRAMKVPFTAAHGAIRFSLSHVTTAAEIDRLLEVLPFIVQQLRALSPF.GAEE.VK FEYIEGGAILLLLNKUGIAASSGSACTSGSLEPSHVRAMDIPYTAAHGTVRFSLSRYTEEEIDRVIREV	100 100 99 97 200 199 198 196 297 299 298 295 387 398
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDEEILFTSCGTESDSTALLSALKAQPERKTVITTVVEH MSVIYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDEEILFTSCGTESDSTALLSALKAQPERKTVITTVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSLBFGNQVQMALKKARQSVQALLGAED.ESEIVFTSCGTEGDNAAIRAALLAPFARHIITTQVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSLBFGNQVQMALKKARQSVQAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSLBDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSLBDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDFMVLEAMMFFLTDFYGNPSSLBDFGIPAQAALERABQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH AAVLALCDHLERQEGVTVHRIFVDGDGRLDIEAYRAALSPRVALVSLMWANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVVSVMANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVVSVMANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVPIULKNSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVPIDLKNSSIHMLSLC PAVLNVCKQLETQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKIPIAVGQTRIDMLSCS CEE SOCEEDOS BOOGOG BLEDO BOOEDDE BOOEDDE BLEDOS BLEDOSED OED OED OE BLEDOSED OED OE BLEDO AHKFHGPKGVGALWLRKGVPFQPLIRGGRQRGRAGTENIFGIVGGGRAACELAGGDHGAVRLLRRLEQTLLRALPDEVTGDPDNRLPNTANIA GHKIHAPKGIGALVVRGVRFRPLLGGHQERGRAGTENASIIGLGVAARALQFMEHENTEVNALRDKLEAGILAVVPHAPVTGDPDNRLPNTANIA GHKIHAPKGIGGLYVRGVRFRPLLGGHQEYGRRAGTENICGIVGGGAACELANIHL.FGMTHIGQLRNRLEHRLLASVPSVMVMGGQPRVPGTVNLA B O DEDEBOO BO DE DEDOGEOODEEDOODE BOOD DEDOGEOD OG OED OO OE BO DEDOGEO FDFAEGRAIVMLLDRAGICVSSGAACASGAMEPSHVIRAMKVFTAHGAIRFSLSHVTTAAEIDRLLEVLPFIVQLRALSFF.GAEE.VK FEYIEGEAILLLINKVGIAASSGSACTSGSLEPSHVRAMDIPYTAAHGJFYTLIGSIRFSLCRYTTEAUIDRVIEVPFIVAQLRNVSPVSGNGP.VEDFGKAFA FKYIEGEAILLSLNKYGICASSGSACTSGSLEPSHVRAMGIFYTLHAGSLFYTLEGSLRFSLCRYTEAQIDRVIEVMPEIVERLRALSPFKNDEAGHLQAQBQTLA	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn	MADVYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVVEH MSVIYLDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFGNQVQMALKKARQSVQKLLGAEHDSEILFTSCGTESDSTATLSALKAQPERKTVITTVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRLDPMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAEQQAAALLGAEYPSEIIFTSWPRATFRHAAIALLPERREIITSVEH AAVLALCDHLERQEGVTVHRIFVDGGGGRLDIEAYSALSPRVALVSLMMANNETGTVFPVEGLAELAHRAGAALFHDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVVSVMAANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLPVDKKGRLDLEHYASLLTDDVAVSVMAANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVPIULKNSSIHMLSLC PAVLNVCKQLETQ.GYTVTYLSVNSHQQLDLDELEASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHDVAVQAVGKIPLNKKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKIPLNKKTSTIDMLTIS PATLAACEHLERQ.GYRIHRIAVDSEGALDMAQFRAALSPRVALVSVMAANNETGTVFPIEGHGAVLLRDLEQALLARPKARVLGDPLDRLPNTSCVA GHKHLAPKGVGALWLKGVGVFQPLIRGGRQGRGRAGTENIFGIVGGIGGAAAELALGGDHGAVRLLRDRLEQGILARVFKARUGDPDNRLPNTAIA GHKHLAPKGVGALVLRKGVPFQPLIRGGRQRGRGREGENAGSIIGLGVAGAACELANIHL.GGDHGAVRLLRDLEQULARVPKARVLGDPDNRLPNTAIA GHKHLAPKGIGGLVYRGVRFRPLLIGGHQERGRAGTENIGGIVGGIGAACELANIHL.FGMTHIGQLRNRLEHRLLASVPSVMVMGGGQPRVPGTVNLA GHKHLAPKGIGGS DO GENERGOOGE DOG GENEGOOGO OGO OGO OGO OGO OGO OGO OGO OGO O	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 398
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn	MADVILDNNATTRVDDE I VQAML PFFTEQFGNPSSLHSFGNQVGMALKKARQSVQKLLGAEHDSE ILFTSGGTESDSTALLSALKAQPERKTVI TTVVEH MSVIYLDNNATTKVDPDVVEAIMPYLTDYYGNPSSHHTFGQQLGKAVRTAREQVAALLGAD.ESEIVFTSGGTESDSTALLSALKAQPERKTVI TTVVEH MKQVYLDNNATTKVDPDVVEAIMPYLTDFYGNPSSHHTFGQQLGKAVRTAREQVAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTKVDPDVVEAIMPYLTDFYGNPSSHHTFGQQLGKAVRTAREQVAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTKVDPDVVEAIMPYLTDFYGNPSSHHTFGQQLGKAVRTAREQVAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH MKQVYLDNNATTRUDPMVLEAMPPLTDFYGNPSSHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATFRHAAIALLPERREIITSVVEH AAVLALCDHLERQESVTVHRIFVDGKGRLDIEAYAALSFRVALVSIMMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVFIVLRGTEIDMLSLS PAVLSLCDYLAS.EGYTVHKLFVDKKGRLDLEHYASLITDDVAVSVMAANNETGTVFPIEEIGKRVKERGAIFHDAVQAVGKVFIVLRGTEIDMLSLS PAVLSLCDYLAS.GYTVTKLSVNSHGQLDLEHYASLITDDVAVSVMAANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVFIULKSSHDULKSSHDALSC PAVLNVCKQLETQ.GYTVTILSVNSHGQLDDAQFRAALSFRVALVSVMAANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVFIUNGGTRIDMLSCS OCH BOODDO BOODO BODO BODO BOODO BODOD BODO OD OD OD OD OD OD OD O AHKFHGPKGVGALMLRKGVFPQFIRGGRQRGRRAGTENIGGIVGUGRAAELALGGDHGAVRLLRQLEQGILARVPKARVLGDPLDRLPNTSCVA GHKHAPKGIGALVVRGVFPFPLIIGGGQQRGRRAGTENIGGIVGUGRAAELALGGDHGAVRLLRQLEQGILARVPKARVLGDPLDRLPNTSCVA GHKHAPKGIGALVVRGVFPFPLLIGGHQERGRAGTENIGGIVGUGRAAELALGGDHGAVRLLRQLEQGILARVPKARVLGDPLDRLPNTSVAL AHKFHGPKGVGCLYLRRGTRFFPLLIGGHQERGRAGTENIGGIVGUGRAAELALGGDHGAVRLLRQLLEQTLLAKIPCEVMGGQPFVPGTVNLA BOODDO BO DO DO BOODO BODOD BODOD BODODO BODODO OD DO OO O DO DO DODOD FDFAGGAIVHLUDRAGICVSSGAACASGAMEPSHVIRAMKVPFTAAHGAIRFSLSRYTTEETIDRVIREVPFIVAQLRAUSPFWNGGGQPFVPGTVNLA FFYIEGEAIVHLLDRAGICASSGSACTSGSLEPSHVIRAMKVPFTAAHGAIRFSLSRYTTEEEIDRVIREVPFIVAGLRNSPYWNGGOP, VEDRGKAFA FKYIEGEAILLLINNVGIAASSGSACTSGSLEPSHVRAMINIPYTAAHGTIRFSLSRYTREKEIDYVATLPPIIDRLRALSPFKNDEAGWLQAQEQTLA FEFIEGEAILLLLNQAGIAASSGSACTSGSLEPSHVRAMINIPYTAAHGTIRFSLSRYTREKEIDYVATLPPIIDRLRALSPFWQNGKFRPADAVFT	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp	MADVILDNNATTRVDDEIVQAMLPFPTEQFGNPSSLHSFGNQVGMALKKARQSVQKLLGAEHDSELLPTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTRVDEIVQAMLPFPTEQFGNPSSLHSFGNQVGMALKKARQSVQKLLGAEHDSELLPTSCGTESDSTAILSALKAQPERKTVITTVVEH MKQVYLDNNATTRVDEIVQAMLPFPTEQFGNPSSLHDFGIPAQAALERAHQQAAALLGAEYPSEIIPTSWPRATPRHAALALLQPAKRHIITTQVEH MKQVYLDNNATTRVDEMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAALALLPERREIITSVVEH MKQVYLDNNATTRVDEMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAALALLPERREIITSVVEH MKQVYLDNNATTRVDEMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAALALLPERREIITSVVEH MKQVYLDNNATTRVDEMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAALALLPERREIITSVVEH MKQVYLDNNATTRVDEMVLEAMMPFLTDFYGNPSSIHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWPRATPRHAALALLPERREIITSVVEH MKQVYLDNNATTRVDEMSCHUPGKGRUDLEHYASLSTDDVAVVSWMANNETGTVFPVEGLAELAHRAGALFHTDAVQAVGKVPIVLRGTEIDMLSSS PAVLSCCQLEQ.GYTVTHSIVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIVLRGTEIDMLSSS PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLDELEASLTGNTALVSVMAANNETGVFPIEEIGKRVKERGAIFHVDAVQAVGKVPIVLRGTFINMKTSTIDMLTSS PATLAACEHLEQ.GYTVHRIAVDSEGALDMAQFRAALSFRVALVSVMAANNETGVFPIEGMAELAHEQGALFHCDAVQVVGKIPIANGKIPINTSTIDMLSSS PAVLNVCKQLEXGGYFPQPLIRGGQQRGRAGTENIPGIVQUGKAAELALGGDHGAVRLLRDRLEQGILARVPKARVLGDPLDRIPTSSVA GHKLHAPKGYGVLYLRGTFRFILLGGGQERGRRAGTENIPGIVQUGKAAELALGGDHGAVRLLRDRLEQGILARVPKARVLGDPLDRIPTSSVA GHKLHAPKGYGVLYLRGTFRFILLGGGQERGRRAGTENVGOIVGLGKAAELELIHIETATKKETRLRDRLEQTLLAKIPDCEVNGDI TQLPNTINIG AHKFHGFKGVGCLYLRRGTRFRFILRGGHQEYGRAGTENICGIVGMGAACELANIHL.POMTHIGQLRNRLEHRLLASVPSVMVMGGQPRVPGTVNL AHKFHGFKGVGCLYLRRGTRFRFILRGGHQEYGRAGTENICGIVGMGAACELANIHL.POMTHIGQLRNRLEHRLLASVPSVMVMGGQPRVPGTVNL FYYIEGEAIVLLDRAGICVSSGAACASGAMEPSHVIRAMKVPFTAAHGAIRFSLSRYTTEEEIDRVVEVIEVQLRALSPF.QEE.VK FEYIEGEAIVLLDRAGICASSGACTSGSLEPSHVIRAMKVPFTAAHGAIRFSLSRYTTEEEIDRVVEVIEVMEEVQEXRALSPFKNDEAGMQAQQETAA FFYIEGEAILLLLNQAGIAASSGSACTSGSLEPSHVRAMNIPYTAAHGTIRFSLSRYTTEEKIDYVVATLPPIIDRLRALSPFKNDEAGNAQAQQETAA FFYIE	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp	MADVYLDNNATTRVDDEIVQAMLPFPTEQFGNPSSLHSFGNQVAMALKKARQSVQKLLGAEHDSELLPTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTRVDEIVQAMLPFPTEQFGNPSSLHSFGNQVAMALKKARQSVQKLLGAEHDSELLPTSCGTESDSTAILSALKAQPERKTVITTVVEH MSVIYLDNNATTRVDEVVEAIMFYLTDYGNPSSHDFGGQLGKAWRTAREQVAALLGAD.SEETVFTSCGTESDSTAILSALKAQPERKTVITTVVEH MKQVYLDNNATTRVDEVVEAIMFYLTDYGNPSSHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRVDEVVEAIMFYLTDYGNPSSHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRVDEVVEAIMFYLTDYGNPSSHDFGIPAQAALERAHQQAAALLGAEYPSEIIFTSWFRATPRHAAIALLPERREIITSVVEH MKQVYLDNNATTRVDEVVEAIMFYLTDYGNPSSHDFGIPAQAALSFRVALVSIMMANNETGTVFPFEGLAELAHRAGALFHTDAVQAVGKVFIVURGTEIDMLSLS PAVLSLCDVLAS.EGYTVHKIPVDGKGRLDLEHYASLITDVAVVSVMAANNETGTVFPIEEIGKKVKERGAIFHTDAVQAVGKVFIVURGTEIDMLSLS PAVLSUCQLETQ.GYTVHKIPVDGKGRLDLEHYASLITDVAVVSVMAANNETGTVFPIEEIGKKVKERGAIFHTDAVQAVGKVFILMKKTSIIDMLISS PATLAACEHLERQ.GYTVHKIPVDGKGRQDDAGPRAALSFRVALVSVMAANNETGVLPPIEEIGKKVKERGAIFHTDAVQAVGKVFILMKKTSIIDMLISS PATLAACEHLERQ.GYTHRIJAVDSEGALDMAQFRAALSFRVALVSVMAANNETGVLPPIEEIGKKVKERGAIFHTDAVQAVGKVFILMKKTSIIDMLISS GHKIHAFKGYGALWLRKGVFPQLIRGGRQGRGRAGTENAGTENIPGIVGUGRAAELALGCDHGAVKLLRDRLEQGILAKVFKARVLGDPLDRLPNTSCVA GHKIHAFKGYGALWLRKGVFPQLIRGGRQGERGRAGTENAGTENIPGIVGUGRAAELALGCDHGAVKLLRDRLEQGILAKVFKARVLGDPDNRLPNTANG GHKIHAFKGYGQLYURGGYFRFLILGGRQGERGRAGTENIPGIVGUGRAAELALGCDHGAVKLLRDRLEQGILAKVFKARVLGDPDNRLPNTANG GHKIHAFKGYGQLYURGGVFFPLIRGGHQERGRAGTENVFOIVGUGGAACELANIHL.GOMTHIGQLRNRLEHLLASVPSVMVMGGQQPRVPGTVNLA GHKIHAFKGIGALVURGGUGSGGACASGAMEPSHVIRAMKVPFTAAHGAIRFSLSHVTAAEIDRULEVLPPIVQLRALSFF.GAELVK FYIEGGAILULLNKVGICASSGSACTSGSLEFSHVRAMD IYTAAHGTIRFSLSRYTTEEIDRVIREVPFIVAQLRNVSFYNGGNGP.VEDPGKAFA FYYIEGGAILLLLNKGICASSGSACTSGSLEFSHVRAMM IPYTAAHGTIRFSLSRYTREKEIDVVATLPPIIVGURALSFFWNDGKPRPADAVFT PVYG HR	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn	MADVILDNNATTRVDDEIVQAMLPFFTEQFONPSSLBSFONQVGALKKARQSVQKLLGAEBDSILFTSGGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDEIVQAMLPFFTEQFONPSSLBSFONQVGALKKARQSVQKLLGAEBDSILFTSGGTESDSTAILSALKAQPERKTVITTVEH MSVIYLDNNATTRVDEIVQAMLPFFTEQFONPSSLBJFGQQGKAVRTAREQVALLGAD.ESEIVFTSGGTESDSTAILSALKAQPERKTVITTVEH MKQVYLDNNATTRVDENVEAIMPFLTDYGNPSSHDFGIPAQAALEAAQVAALLGAD.SEIVFTSGGTESDSTAILSALKAQPERKTVITTVEH MKQVYLDNNATTRVDENVEAIMPFLTDYGNPSSHDFGIPAQAALEAAQVAALLGAD.SEIVFTSGGTESDSTAILSALKAQPERKTVITTVEH MKQVYLDNNATTRVDENVEAIMPFLTDYGNPSSHDFGIPAQAALEAAQQAAALLGADYPSEIIFTSWFRATFRHAAIALLPEREIITSVEH ANVALCDHLERQEGVTVHRIPVDGGRUDIEAYAALSERVALVSIMMANNETGTVPFVEGLAELAHRAGALFHTDAVQAVGKVPTVLRGTEIDMLSIS PAVLSLCDYLAS.EGYTVHKLIPVDKKGRUDIEHYASLLTDDVAVVSVMAANNETGTVPFVEGLAELAHRAGALFHTDAVQAVGKVPTULKRSTEIDMLSIS PAVLNVCKQLETQ.GYTVTYLSVNSHGQLDLDELEASLTGNTALVTIMYANNETGTVPFIEEIGKRVKERGAIFHVDAVQAVGKVPTULKNSSTBLSIC PAVLNVCKQLETQ.GYTHRIAVDSEGALDMAQFRAALSFRVALVSVMAANNETGTVFFIEEIGKRVKERGAIFHVDAVQAVGKVPTULKNSSTBLSIC PAVLNVCKQLETQ.GYTHRIAVDSEGALDMAQFRAALSFRVALVSVMAANNETGTVFFIEEIGKRVKERGAIFHVDAVQAVGKVPTULKNSSTBLSIC GAUNCKQLETQ.GYTHRIAVDSEGALDMAQFRAALSFRVALVSVMAANNETGTVFFIEEIGKRVKERGAIFHVDAVQAVGKVPTULKSTIDMLTSC GHKHAPKGVGVLYLRRGTRFFILRGGHQERGRAGTENVFOIVGLGKAAELALGGDHGAVKLLRDRLEQGILARVFKARVLGDFLDRLPNTSCVA GHKHAPKGVGVLYLRRGTRFFILRGGHQERGRAGTENVFOIVGLGKAAELALGGDHGAVRLLRDRLEQGILARVFKARVLGDFLDRLPNTSCVA GHKHAPKGVGVLYLRRGTRFFILRGGHQERGRAGTENVFOIVGLGKAAELALGGDHGAVRLLRDRLEQTLLAKIFDCEVNODIJQRLPNTNIG AHKFHGFKGVGCLYLRRGTRFFILRGGHQEYGRAGTENICGIVGMGAACELANIHL.FGMTHIGQLRNRLEHRLLASVFSVMVMGGQPRVPGTVNL AHKFHGFKGVGCLYLRRGTRFFILRGGHQEYGRAGTENICGIVGMGAACELANIHL.FGMTHIGQLRNLEHRLLASVFSVMVMGGQPRVPGTVNL FYYG FYYIGEAILLLINKVGIAASSGSACTSGSLEPSHVRAMKIPTTAAHGTIRFSLSRVTTEKEIDRVIREVPFIVAQLRNVSFWSONGG, ADAGAT FYYIG HR FYYIG	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn	MADVYLDNNATTRVDDE IVQAHLPFFTEQFQMPSSLBSFQMQVAALKKARSSVQKLGAEHDEE ILFTEGGTESDSTAILSALKAQPERKTVITTVVE MSVIYLDNNATTRVDDE IVQAHLPFFTEQFQMPSSLBSFQMQVAALKKARSSVQKLGAEHDEE ILFTEGGTESDSTAILSALKAQPERKTVITTVVE MSVIYLDNNATTRVDDE VVEALMEYLTDYYGMPSSLBSFGMQVAALKKARSSVQKLGAEHDEE ILFTEGGTESDSTAILSALKAQPERKTVITTVVEH MKQVYLDNNATTRVDDE VVEALMEYLTDYYGMPSSLBFFGQQLGKAARTAREQVAALLGAD. ESEIVFTSCGTESDNATRAALLQPAKHEITTTYVEH MKQVYLDNNATTRVDE VVEALMEYLTDYYGMPSSLBFFGQQLGKAARTAREQVAALLGAD. ESEIVFTSCGTESDNATRAALLAPERKTVITTVVEH MKQVYLDNNATTRVDE VLEAMPFLTDFYGMPSSLBFFGQQLGKAARTAREQVAALLGAD. ESEIVFTSCGTESDNATRAALLAPERKTVITTVVEH MKQVYLDNNATTRVDE VLEAMPFLTDFYGMPSSLBFFGQQLGKAARTAREQVAALLGAD. ESEIVFTSCGTESDNATRAALLAPERKTVITTVVEH MKQVYLDNNATTRVDE VLEAMPFLTDFYGMSSLBFSSLBFFGQQLGKAARTAREQVAALLGADEYSEIFTSWPRATPRHAALLAPERKTVITTVVEH MKQVYLDNNATTRVDE VLEAMPFLTDFYGMSSLBFFGQQLGKAALSERVALVSLMAANNETGTVFPTEGLAELAHAGALFHTAAGALFHTDAVQAVGKVPTULKGSTEIDLSS PAULACEULERQ. GYTVTVLSVNSHGQLDLDE LEASITGHTALVTIMVANNETGTVFPTEGLAELAHAGQALFHTDAVQAVGKVPTULKSSTEMI.SC PAULNCKQLETQ. GYTVTVLSVNSHGQLDLDE LEASITGHTALVTIMVANNETGTVFPTEGLAELAHEQGALFHCDAVQAVGKVPTDLVNSTKNSTIMLISS PAULACEULERQ. GYTVTVLSVNSHGQLDLDE LEASITGHTALVTIMVANNETGTVFPTEGEMAELAHEQGALFHCDAVQAVGKVPTDLVNSTMAKSSTEMI.SC PAULNCKQLETQ. GYTVTVLSVNSHGQLDLDE LEASITGHTALVTIMVANNETGTVFPTEGGAARLAHAGQALFHCDAVQAVGKVPTDLVNSTVMATSTIMLISS PAULACEULERQ. GYTVTVLSVNSHGQLDLDE LEASITGHTALVTIMVANNETGTVFPTEGGAARLAHEQGALFHCDAVQAVGKVPTDLVNSTVMATSTMSTK GHKHGPKGVGULVLRGTRFFFLLRGGHQERGRAGTENTOGTVGLGKAARLELL HIETATKKETRLRORLEQGLLANCVPAAVTGUPDIDUNTNT GHKHGPKGVGULVLRGTRFFFLLRGGHQERGRAGTENTOGTVGLGKAARLEL HIETATKKETRLRORLEQTLLASITTSCHTMGQQRCPRYFGTVNLA FYGGEATIVLLDRAGTCVSSGAACSGSACTSGSLEPSHVIRAMKVPTTAAHGTIRFSLSRVTTRELEIDVVATLPPTIDDLRALSFYNQNGKPR PADAVTT PVYG HR PVYG	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393 402 400 397
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn D)	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLEBFGNQVGMALKKARQSVQKLLGAREDSEILFTSCATESDSTATISALKAQPERKTVITTVEE MSVIILDNNATTRVDEVVEALMFFLTDFYGNPSSLEBFGNQVGMALKKARQSVQKLLGAREDSEILFTSCATESDSTATISALKAQPERKTVITTVEE MSVIILDNNATTRVDEVVEALMFFLTDFYGNPSSLEDGGQCKKARTARQQAAALLGAPYSEIIFTSWFRATFRHAAIALLQPAKREIITTQVEE MKQVYLDNNATTRUDEMVEAMFFLTDFYGNPSSLEDGGQCKKARTARQQAAALLGAPYSEIIFTSWFRATFRHAAIALLQPAKREIITTQVEE MKQVYLDNNATTRUDEMVEAMFFLTDFYGNPSSLEDGGQCKKARTARQQAAALLGAPYSEIIFTSWFRATFRHAAIALLQPAKREIITTQVEE MKQVYLDNNATTRUDEMVEAMFFLTDFYGNPSSLEDGGQCKKARTARQQAAALLGAPYSEIIFTSWFRATFRHAAIALLQPAKREIITTQVEE MKQVYLDNNATTRUDEMVEAMFFLTDFYGNPSSLEDGGQCKKARAINESTGTVEPSEGAELAHRAGALFRHDAVQAVKVPIVLNGETEIMISS SAUSLCDYLAS, EGYTVEKLFVDKKGRLDLEHYSLITDDVAVVSVMANNETGTVFPIEEIGKRKKERGAIFHVDAVQAVGKVPIVLNGTEIMISS PAULAVCKQLETQ, GYTYTYLSVNSHQQLDLDELASI.TGHTALVTIMVANNETGTVPPIEEIGKRKKERGAIFHVDAVQAVGKVPIDLKNSSIEMISG AALKACEHLERQ, GYRIHRIAVDSGGALDMAQFRAALSPRVALVSVMANNETGTVPPIEEIGKRKKERGAIFHVDAVQAVGKVPIDLKNSSIEMISG AALKACEHLERQ, GYRIHRIAVDSGGALDMAQFRAALSPRVALVSVMANNETGTVPPIEEIGKRVKERGAIFHVDAVQAVGKVPIDLKNSSIEMISG AALKHERGYGQCULYLRGTRFFFLUGGGQCRGRAGTENNGJUNGGGACHAALSPNEHENTIVNALRDKLEAGILAVVPKARVLGDPLDRLPNTSN GHKLBAPKGJVQULYLRGTRFFFLUGGGQCRGRAGTENNGJUNGGACHLANIL, GODHGAVRLLRDLEQGILARVPKARVLGDPLDRLPNTANG GHKLBAPKGJVQULYLRGTRFFFLUGGGQCRGRAGTENNGJUNGGACCHANAIL, PAMTHIQQLRNLEHRLLASVPSVMVMOGQQPRVGTVNL AAKHENGGACIVYLLDRAGTRFFFLUGGAQCGCRGAGTENNGJEG DOOL DOOL OOL OOL OOL OOL OOL GFDFAEGGAIVMLLDRAGICVSSGAACASGAMEPSHVIRAMKVPFTAABGAIRFSLSHVTTAAEIDRULEVPFIVQLRAVSPYNSGNOP, VEDFGKAFA FYYIEGGAILLLINKVGIAASSGSACTSGSLEPSHVWRAMNIPYTAABGTIRFSLSHVTTAEIDRULEVPFIVQLRVSPYNSGNOP, VEDFGKAFA FYYIEGGAILLLINKVGIAASSGSACTSGSLEPSHVWRAMNIPYTAABGTIRFSLSRVTREKEIDVVATLPPIIDRLRALSPYNQNGKFRPADAVT	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393 402 400 397
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn D)	<form></form>	100 100 99 97 200 199 198 196 297 299 298 295 387 398 393 393 402 400 397
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn Avi Asp Kpn D) Rsp	MADVYLDNNATTRVDDE I VAAHLPF JTE OF ON PSSLHSFON (VAALKKAR SVOKLLGAE EDSELL FTS OG TE SDSTALLSALKAOPERKTVITT VER SVJILDNNATTRUD PUVLEAM PFILT DYYGN PSSHH FGO QLGKAVRTAR QVAALLGA JSE IVFTS OG TE SDSTALLSALKAOPERKTVITT VER MKQVYLDNNATTRUD PUVLEAM PFILT DYYGN PSSHH FGO QLGKAVRTAR QVAALLGA JSE IVFTS OG TE SDSTALLSALKAOPERKTVITT VER MKQVYLDNNATTRUD PUVLEAM PFILT DYYGN PSSHH FGO QLGKAVRTAR QVAALLGA JSE IVFTS OG TE SDSTALLSALKAOPERKTVITT VER MKQVYLDNNATTRUD PUVLEAM PFILT DYYGN PSSHH FGO QLGKAVRTAR QVAALLGA JSE IVFTS OG TE SDSTALLSALALLOPERKTITT VER MKQVYLDNNATTRUD PUVLEAM PFILT DYYGN PSSHH FGO QLGKAVRTAR QVAALLGA JSE IVFTS OF TE SDSTALLSALALLOPERKTTT VER MKQVYLDNNATTRUD PUVLEAM PFILT DYYGN PSSHH FGO QLGKAVRTAR QVAALGKAVPI UND QVA VGKVPI ULKNSSHH JSE DATLAD CHILER QCGYT VERH FYD GGGRUD LEATRAL STRUT JSHAM NETGT VPPIEELGKAVRERGALFH VDA VQAVGKVPI ULKNSSHH JSE DATLAD CHILER QCGYT VERH FYD GGGRUD LEATRAL STRUT JSHAM NETGT VPPIEELGKAVRERGALFH ODA VQAVGKI PLAVKTSTI IMLT IS DATLAD CHILER QCGYT VERH FYD JSK GGU DDALLEASL T ONTAL VTI MANNETGT VPPIEELGKAVRERGALFH ODA VQAVGKI PLAVKTSTI IMLT IS DATLAD CHILER QCGYT VERH FYD JSK GGU DDAL OF DOO DOO OF OO OF OO OF OO OF OO OF OO OF OF O	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 402 400 397
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn D) Rsp Aca	MADVYLDNNATTRVDDE IVQAALPFF TEOFGNPSSLBSGNQVGAALKKARQSVQKLLGAEBDSE LLFTSCGTESDSTALLSALKAOPERKTVITTVUE MSVJILDNNATTRLOPDVCALH4FYLDJYGNPSSBHEFGOQLGKAVRTARQVAALLGAD.SESTVFTSCGTESDSTALLSALKAOPERKTVITTVUE MSVJILDNNATTRLOPMVLAAMPFLTDYGNPSSBHEFGOQLGKAVRTARQVAALLGAPSEB IFFSWPRATPRHAALALLPERKEIITSVUE MSVJILDNNATTRLOPMVLAAMPFLTDYGNPSSBHEFGOQLGKAVRTARQVAALLGAPSEIIFTSWPRATPRHAALALLPERKEIITSVUE AAVLALCDHLEQGEVJTWEIFJVDGGGLDIEATRAALSERVALVSLMAANEETGTVFPIEEIGKVKERGAIFHDAVQAVGKVFIVLRGTEIMLSLS PAVLSLCDYLAS. EGYTVHEIFJVDGGGLDIEATRAALSERVALVSLMAANEETGTVFPIEEIGKVKERGAIFHDAVQAVGKVFIVLRGTEIMLSLS PAVLSLCDYLAS. EGYTVHEIFJVDGGGLDIEATRAALSERVALVSLMAANEETGTVFPIEEIGKVKERGAIFHDAVQAVGKVFIVLRGTEIMLSLS PAVLSLCDYLAS. EGYTVHEIFJVDGGGGLDIEATRAALSERVALVSLMAANEETGTVFPIEEIGKVKERGAIFHDAVQAVGKVFIVLRGTEIMLSLS PAVLSLCDYLAS. EGYTVHEIFJVDGGGGLDIEATRAALSERVALVSUMAANEETGTVFPIEEIGKVKERGAIFHDAVQAVGKVFIVLRGTEIMLSLS GAUNVCKQLEFQ.GYTVHISJVNSHGQLDIDELEASLTGNTALVTIMVANNETGTVFPIEEIGKVKERGAIFHODAVQAVGKVFIVLRGTEIMLSLS PAVLSUCHENDO BOODO BOODO BOODO BOODO BOODO BOODO BOODO CON CONCOLONG OF DOODO CON CONCOLONG OF DOODO CONCOLONG O	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 402 400 397
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn D) Rsp Aca Avi	MADVYLDNNATTRVDDEIVQAMLPFFTEQFGNPSSLBSFGNQVGMALKKARQSVQKLLGAEHDSEILFTSCGTEEDSTAILSALKAQFERKTVITTVEE MSVIJLDNNATTRVDEVVEAIMPYLTDYYGNPSSNHTFGGQLGKAVRTARQVAALGAD.ESEIVFTSCGTEEDSTAILSALKAQFERKTVITTVEE MSVIJLDNNATTRVDEVVEAIMPYLTDYYGNPSSNHTFGGQLGKAVRTARQVAALGAD.ESEIVFTSCGTEEDSTAILSALKAQFERKTVITTVEE MSVIJLDNNATTRUDEVVEAIMPYLTDYYGNPSSNHTFGGULGKAALGAEVAETAGVAALGAE SEIVFTSCGTEEDSTAILSALKAQFERKTVITTVEE MSVIJLDNNATTRUDEVVEAIMPYLTDYYGNPSSNHTFGGULGKAALGAEVAETAGVAAULGAEVAETA AAVLALCDELERQGEVVENEIFVGGDCRLDIEATAALSERVALVSIJMAANNETGTLPPIEEMARLADDAGINFHTDAVQAVCKYPTULGSTEIDMLSLS PAVLSLCDVLAS, BOYTVELYDVKKGRLDLEHYASLLTDDVAVVSVMAANNETGTLPPIEEMARLADDAGINFHTDAVQAVCKYPTULGSTEIDMLSLS PAVLSLCDVLAS, BOYTVELYDVKKGRLDLEHYASLLTDDVAVVSVMAANNETGTLPPIEEMARLADDAGINFHTDAVQAVCKYPTULGSTEIDMLSLS PAVLSLCDVLAS, BOYTVELYDVKKGRLDLEHYASLLTDDVAVVSVMAANNETGTLPPIEEMARLADDAGINFHTDAVQAVCKYPTULGSTEIDMLSLS PAVLSLCDVLAS, BOYTVELYDVKGRUDEGALDMAQFRAALSERVALVSVMAANNETGVLPPIGEMAELAHEQGALFFDCAVQVVCKIPIAKSTIDMLSCS PAVLSLCDVLAS, BOYTVELGUNGGEGGUGGRAAGTENINGTULSVMAANNETGVLPPIGEMAELAHEQGALFFDCAVQVVCKIPIAKGTIDMLSCS PAVLSLCDVLAS, BOYTVELGUNGGUGGGRAAGTENINGTULSVMAANNETGVLPPIGEMAELAHEQGALFFDCAVQVVCKIPIAVGGTIDMLSCS COLOCION ON DE DE DE DOOD BOOD BOOD BOOD DE DE DOOD ON DE DE DOOD ON GRIEDEKSVGGLALVERGTFFFLLGGGQEGGRAAGTENNIGTVGIGGAAAELLAILGOBBGAVLLEAGILAVVFLAVVLGPIDULGLENFISCA GRIEDEKSVGGLALVERGTFFFLLGGGQEGGRAAGTENNIGTVGIGGAAAELLAINHEL FOUNDELEAGULAVFLAVVLGPIDULGULPNTNIG AHKFHGFKSVGCLILRGGTFFFLLGGGQEGGRAAGTENNIGTVGIGGAAAELLAINHEL, FOUTHLGQLARREEHELLASVFSVMVMGGGQERVFOTVNLA ON DE DOOD DE DE DE DOOD DE DOOD DE DOOD ON DOOD DE DOOD FDAGEGAILLLANGGTAASSGSACTSGSLEPSHVRAMDIPTTAAHGTIRFSLSKYTTEAGIDRVIEWFFIVERLAALSFFLNDEAGALQAQEQTA FEYIEGAILLLINKVGIAASSGSACTSGSLEPSHVRAMDIPTTAAHGTIRFSLSKYTTEAGIDRVIEWFFIVERLAALSFFLNDEAGALQAQEQTA FEYIEGAILLLINKUGIAASSGSACTSGSLEPSHVVRAMDIPTTAAHGTIRFSLSKYTTEAGIDRVIEWFFIVERLAALSFFLNDEAGALQAQEQTA FEYIEGAILLLINAQGIAASSGSACTSGSLEPSHVVRAMDIPTTAAHGTIRFSLSKYTTEAGIDRVIEWFFIVERLAALSFFLOGAGLAGXGKQKFYKY MTYQFPSEDSLTLLLEKNLSSAEEJFALLEVDUPQVVVALHILMKRYVVXLHILMKKYVKVKU	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 402 400 397 89 89 97
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn D) Rsp Aca Avi Kpn	MADVYLDNNATTRVDEDEIVQAMLPFFTEQFQNPSSLBSFQNQVQMALKKARQSVQKLLGAEHDSEILFTSCOTESDSTAILSALKAQFERKTVITTVVEH MKQVYLDNNATTRUDEMVLEAMHPFLTDYGNPSSHBFGGNQAGALKARQSVQKLLGAEHDSEILFTSCOTESDSTAILSALKAQFERKTVITTVVEH MKQVYLDNNATTRUDEMVLEAMHPFLTDYGNPSSHBFGGNQAGALKARQSVQKLLGAEHDSEILFTSCOTESDSTAILSALKAQFERKTVITTVVEH MKQVYLDNNATTRUDEMVLEAMHPFLTDYGNPSSHBFGGIPGAALKARQSVAKLLGAEHDSEILFTSCOTESDSTAILSALKAQFERKTVITTVVEH MKQVYLDNNATTRUDEMVLEAMHPFLTDYGNGSSHBFGGIPSSLBAFGNAALLAQFARRHLITQVEH MKQVYLDNNATTRUDEMVLEAMHPFLTDYGNGORLDIEKASL AAVLALCDELERQSVTVERIFVDGORLDIEKASLSDTVALVSLMAANNETGTVFPTEGIALAHRAGALFHTDAVQAVGKYPTVLGGTEIDMLSLS FAVLSLCDVLAS. GSYTVERLFVDKKGRLDLEHXSLLTDDVAVVSVMANNETGTVFPTEGIGRVKEGAIFFVDAVQAVGKYPTVLGGTEIDMLSLS FAVLSLCDVLAS. GSYTVERLFVDKKGRLDLEHXSLLTDDVAVSVMANNETGTVFPTEGIGRVKEGAIFFVDAVQAVGKYPTULKSSIBMLSLC AAVLALCDELERQ.GYRIHRIAVDSEGALDMAQFRAALSFRVALVSVMANNETGVLFPIGEMAELAHEQGALFHCDAVQVVGKIPIAVGQTRIDMLSCS OM BOGGDOD BOOGOOL DE	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393 402 400 397 89 89 89 89
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn D) Rsp Aca Avi Kpn Rsp	MADVYLDNNATTREVDDEIVQAHLPFFTEQFGNPSSLESFGNQVGMALKKARGSVGKILGAEHDSEILFTSGGTESDSTAILSALKAQPERKTVITTVVEH MKQVILDNNATTREVDEDVVALNHYLTDYVGNFSSNHTFGGLGKAVRTAREQVAALLGAD. ESELVFTSGCTESDSTAILSALKAQPERKTVITTVVEH MKQVILDNNATTREVDEMVFELIGYVGNPSSNHTFGGIGKAVRTAREQVAALLGAD. ESELVFTSGCTESDSTAILSALKAQPERKTVITTVVEH MKQVILDNNATTREVDEMVFELIGYVGNSSNHTFGGIGKAVRTAREQVAALLGAD. ESELVFTSGCTESDSTAILSALKAQPERKTVITTVVEH MKQVILDNNATTREVDEMVFELIGYVGNSSNHTFGGIGKAVRTAREQVAALLGAD. ESELVFTSGCTESDSTAILSALKAQPERKTVITTVEH MKQVILDNNATTREVDEMVFELIGYVGNSSNHTFGGIFGSLENGEN AAVLALCHLERQEGVTVHRIFYDGDGRLDIELASLIDDVAVVSVMANNETGTVFFVEGLAELAHBAGALPHTDAVQAVGKVFIVLRGTEIDMLSLS PAVLNCQLTQ, GYRTHRLFYDKKGALDLERYASLIDDVAVVSVMANNETGTVFFVEGLAELAHBAGALPHTDAVQAVGKVFIULKSSIBHLSLC PAVLNVCQLTQ, GYRTHRLFYDKKGALDLERYASLIDDVAVVSVMANNETGTVFFIEEIGRRVKERGGALPHDAVQAVGKVFIULKSSIBHLSLC PAVLNVCQLTQ, GYRTHRLFYDKGALDLEDKASLISTONTAUTNATEGTVFFIEEIGRRVKERGALHPDAVQAVGKVFIULKSSIBHLSLC PAVLNVCQLTQ, GYRTHRLFOGGGCGRRAAGTENIFOLYGUGGRAAELAL,, GGDBGAVLLADDAGLPHTAVQAVGKVFIULKSSIBHLSLC PAVLNVCQLTQ, GYRTHRLFOGGGCGRRAAGTENIFOLYGUGGRAAELAL,, GGDBGAVLLADDAGLPHTAVQAVGKVFIULKSSIBHLSLC GHKLHAFKSVGVJLWRGTFFULLGGGRQGRGRAAGTENIFOLYGUGGRAAELAL,, GGDBGAVLLADDLEGGILARVFKARVLGDPLDRIFTSVA GHKLHAFKSVGVJLWRGTFFULLGGGRQGRGRAAGTENIFOLYGUGGRAAELAL INTEATKICHTALDDLEGTLAXVFHAFVLGDPDNRLPHTNIG AHKPGFKSVGALWLKKSVFFQULGGRGGRGRAAGTENIFOLYGUGGRAAELAL INTEATKICHTALDDLEGTLLAVVFHAFVLGDPDNRLPHTNIG AHKPGFKSVGALWLKKSVFFQULGGRGGRGRAAGTENIGGI VGMGAACELANIHL, PGMTHIGQLRNRLEHRLLSVFSVMMGGQQPRVPGTVHLA GHKILAPKSIGUALVSKGSGSLEPSSWVFAMADI FYTAHGTIRFYLSSISMTTAAETDRIFTEEDIRVTFVQUGGALSPFLAQUNGKFR., PADAVT FYTIEGEAILLLINGGGASGSAACSGSLEPSSWVFAMADI FYTAHGTIRFYLSSISMTTAEETDRVFFVPTVQUGALSPFLAQUNGKFR., PADAVT FYTIEGEAILLLINGGGASGSAACSGSLEPSSWVFAMADI FYTAHGTIRFSLSSNTTAEETDRVFFVPTVAQLANSPFLAQUNGKFR., PADAVT MYG HR FYTIEGEAILLLINGGGASGSAACSGSLEPSSWVFAMADI FYTAHGTIRFSLSSNTTAEETDRVFFVPTVALLSPFVQAGGASGLAGAGQQGTAS MATAGGILDQLNKASSASGLEPSFVVFVYVYNNHLHIMFXLGQULA. AVDFA., TLDPADARAAABABASSSQLQAADSSFLEGKVFKV MYQCFSDSDLILDAADELVSASDFJEFYGYPTQVQVHV	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393 402 400 397 89 89 97 86
Avi Asp Kpn Rsp Avi Asp Kpn Rsp Avi Asp Kpn Avi Asp Kpn D) Rsp Aca Avi Kpn D) Rsp Aca	MADVYLDNNATTREVDDE IVQAHLPFFTEQFGNPSSLESPGNQVGAALKKARGSVQKILGAEHDSE ILFTSOGTESDSTAILSALKAQPERKTVITTVVER MKQVYLDNNATTREVDE MVKALHPYLTDYYGNPSSLEDFGNQVGAALKKARGSVQKILGAEHDSE ILFTSOGTESDSTAILSALKAQPERKTVITTVVER MKQVYLDNNATTREVDE MVFYLTDYYGNPSSLEDFGIPAQAALERARQAAALGAD.ESE VFTSOGTESDSTAILSALKAQPERKTVITTVVER MKQVILDNNATTREVDE MVFYLTDYYGNPSSLEDFGIPAQAALERARQAAALGAD.ESE VFTSOGTESDSTAILSALKAQPERKTVITTVVER MKQVILDNNATTREVDE MVFYLTDYYGNPSSLEDFGIPAQAALERARQAAALGAD ESE VFTSOGTESDSTAILSALKAQPERKTVITTVVER MKQVILDNNATTREVDE MVFLTDYVDGOGRLD LEAYRALSPRVALVSIMAANETGTVFPVEGLAELARBAQALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS. BGYTVERLFYDKKGLDLEELYASLLDDVAVVSWAANETGTVFPVEGLAELARBAQALFHTDAVQAVGKVPIVLRGTEIDMLSLS PAVLSLCDYLAS. BGYTVERLFYDKKGLDLEELYASLLDDVAVVSWAANETGTVFPIEEIGRKVKKGALDADGIPHTDAVQAVGKVPIDLKSSIBHLSLC PAVLSKCQLTQL, GYRIHRIAVDSEGALDMAQFRAALSPRVALVSWAANETGTVFPIEEIGRKVKKAGLADDAGUPHTAVQAVGKVPIDLKSSIBHLSLC GALGHOG BOOGG BOO BOO BOOG BOOGGE BOOGG	100 100 99 97 200 199 198 196 297 299 298 295 387 398 398 393 393 402 400 397 89 89 89 89 89 89 89

FIG. 4. Alignment of ORF1 (A), NifU (B), NifS (C), and NifW (D) deduced amino acid sequences of R. sphaeroides (Rsp) with those of other bacteria. Abbreviations: Kpn, K. pneumoniae (3, 7, 8); Avi, A. vinelandii (7, 29); Aca, Azotobacter caulinodans (2); Asp, Anabaena sp. (11, 44). Symbols: \square , identical residue; \bigcirc , conservative substitution according to the scheme PAGST, QNED, ILVM, HKR, YFW, C. Numbers at the right refer to amino acid residues.

-		58
Ami	MASTITUTTI REFERENCE AND A CONTRACT	43
Kon	MERVLINDTTLRDGEQSPGVAFRTSEKVAIAEALYAAGITAME	43
StyLeuA	MSQQVIIFDTTLRDGEQALQASLSAKEKLQIALALERMGVDVME	44
SceLeuA	${\tt MVKESIIALAEHAASRASRVIPFVKLAYKNMLKDPSSKYKPFNAPKLSNRKWPDNRITRAPRWLSTDLRDGNQSLPDPMSVEQKKEYFHKLVNIGFKEIE$	100
Scepca	TLLMDTTWRDAHQSLLATRV	
Kpnoxd	VAITDVVLRDAHQSLFATRL	
	Region 1	
	ଷା ଅନେ ତ ତ ତ ତ ତ ତ ତ ତ ତ ପଥିଚ ସେଥି ତ ତ ତ ପଥିଚ ତ	
Rsp	VGVPAMGEEERADIR.AVAAVLKTAAPVVWCRLRAEDLAAAQRTGVVRLHIGVPVSERQISAKLGKDAAWVRDKVEKLVRAASWAGHK	145
Avi	IGIPSMGEEEREVMH.AIAGLGLSSRL.LAWCRLCDVDLAAAR.STGVTMVDLSLPVSDLMLHHKLNRDRDWALREVARLVGEARMAGLE	130
Kpn Charlen	VGTPAMSDEETARIQ.LVRRQLPDATL.MTWCRWHALEIRQSA.DLGIDWVDISTPASDKLRQYKLREPLAVLLERLAMFIHLAHILGLK	130
SceleuA	VSFPSASQTDFDSTRYAVENAPDDVSIQCLVQSREHLIKRTVEALTGAKKATIHTYLATSDMFREIVFNMSREEAISKAVEATKLVRKLTKDDPSQQATR	200
_		
Rsp	. VSVGAEDASRADPFFLAEIAHVAAEAGAIRFRISDTLGVLDFFAHELVGRVVIRCPLFVEFHGHNDLGFAIANSLAAARAGASDLSV	233
AV1 Var	VCLGCEDASRADLEFYVQVSEVAQAAGARFLKFADIVSVMET GTLDKFKFLSKKLDRELEVTAANDU GLAIARILLANTKSGAIIIMI	218
StylenA	VCIGCEDARAASQILARAIAEVANAA	227
SceLeuA	WY EFSPECTS DT PGEFAVE ICEAVKKAWETTEEN FI IFNL PATVEVAS PAVYADO IEVFATH I TEREKVCI STHCHNDRGCGVAATELGMLAGADRVEG	300
Scepca	GTHILGIKDMAGTMKPAAAKLLIGSLRAKYPDLPIHVHTHDSAGTRVASMTACALAGADVVDV	
Kpnoxd	GVDSVAIKDMSGILTPHAAFELVSEIKKRYDVTLHLHCHATTGMAEMALLKAIEAGVDGVDT	
	Region 2	
Rsp	TYNGLGERAGN. AALEEVAAALEA.AGRATGVALGOLCALSELVARASGRPLSPOKPIVGEGYFTHEGGIHVDGLMKDRATYES	315
Avi	TVNGLGERAGN. AALEECVLALKNLHGIDTGIDTGIDTGIPAISALVERASGRQVAWQKSVVGAGVFTHEAGIHVDGLLKHRRNYE	300
Kpn	TVLGLGERAGNAAAWKPSALGLERCLGVETGVHFSALPALCQRVAEAAQRAIDPQQPLVGELVFTHESGVHVAALLRDSESYQSQS.	302
StyLeuA	AMNGIGERAGN.CALEEVIMAIKVRKDIMNVHTNINHHETGAPARPSVQICNIADPSQQSDCRQRRFRHSSGIHQOSVLKIRENYEIMITPES	318
SceLeuA	CLFGNGERTGN.VDLVTVAMNMYTQGVSPNLDFSDLTSVLDVVERCNKIFVSQRAPYGGDLVVCAFSGSHQDAIKKGFNLQNKKRAQGETQWRIFY	392
Scepca	AINSMSGLTSQPS	
Kpnoxd	AISSMSATYGHPA	
Rsp	ADLRPERFGRSHRIAIGKHSSAAGLARALAEAGLPADAATLAALMPALRDWAAITK.RAAAPEDLAALLAAQTETAR	392
Avi	GLNPDELGRSHSLVLGKHSGAHMVRNTYRDLGIELADWOSOALLGRIRAFSTRTKRRSPQPAELQDFYRQLCEQGNP	377
Kon	IAPSLMGRSVR. LVLGKHSGROAVNGVFDO. MGYHLNAAOINOLLPAIRP. AENWK.RSPKDYELVAIYDELCGESAL	377
StylenA	IGSEPDTAEPDLP LWPCRRETSHGRDGLOGHRLOHGPPVRRVPEAGDKKGOVPDVDLEAL. AFINKOOEEPEHFRLDYFSVOSGSSDI	405
SceleuA	LPL DEVIGED VERVIEW SOSGEGGAAWUILESLGIDIPENNOTEPSSAVODHADSLGRELKSDEISKLEKEAUNUNDEOVOAISLVNYNVEKFGTER.	494
2002041		
	0	
Avi	ELAAGGMA	285
Kpn	RARG	381
StyLeuA	ATASVKLACGEEIKAEAANGNGPVDAIYQAINRITGYDVELVKYDLNAKGRARRAGSGRYRREPSWSPLPRRGLATDIVESSAKAMVHVLNNI	498
SceLeuA	RVFTGQVKVGDQIVDIEGTGNGPISSLVDALSNLLNVRFAVANYTEHSLGSGSSTQAASYIHLSYRRNADNEKAYKWGVGVSEDVGDSSVRAIFATINNI	594
StyLeuA	WRAAEVEKELQRKAQNKENNKETV	522
SceLeuA	IHSGDVSIPSLAEVEGKNAAASGSA	619

FIG. 5. Alignment of the *R. sphaeroides* (Rsp) NifV amino acid sequence with NifV from *K. pneumoniae* (Kpn) (3, 7) and *A. vinelandii* (Avi) (7) and with LeuA from Salmonella typhimurium (StyLeuA) (49) and Saccharomyces cerevisiae (SceLeuA) (5). For oxaloacetate decarboxylase from *K. pneumoniae* (Kpnoxd) (residues 2 to 21 and 165 to 239) and pyruvate carboxylase from *S. cerevisiae* (Scepca) (residues 557 to 576 and 727 to 802), only the parts similar to regions 1 and 2 are shown (38, 53). Symbols: \Box , identical residue in at least four sequences; \bigcirc , conservative substitution in at least four sequences according to the scheme PAGST, QNED, ILVM, HKR, YFW, C. Numbers at the right refer to amino acid residues.



FIG. 6. Hybridization of a 0.5-kb SacI-BamHI fragment internal to *rpoN* to BamHI-digested chromosomal DNA from *R. sphaeroi-* des. Sizes are indicated in kilobases.

mini-Mu insertions (4). However, the exact transcription initiation site is unknown, and therefore it may be possible that the *rpoN* promoter is situated upstream from the recently discovered *nifU* gene (33), as is the case in *R. sphaeroides*. The *nifUSVWZM* gene cluster from *A. vine-landii* is also transcribed from a promoter upstream from *nifU* (30).

The organization of the *nifUSVW* genes in *R. sphaeroides* is similar to those in *R. capsulatus* and *K. pneumoniae* (3, 33); however, *A. vinelandii* has two ORFs between *nifV* and *nifW* (29). The cotranscription of the *nifUSVW* genes of *R. sphaeroides*, as indicated by the Tn5 insertional analysis, is underscored by the overlapping stop and start codons of the *nifU*, *nifS*, and *nifV* genes, which are indicative of translational coupling.

The complementation of the R. capsulatus rpoN mutant with the genomic library of R. sphaeroides resulted in the isolation of only one rpoN gene. Much to our surprise, the insertion of a Km^r gene in rpoN had no phenotypic effect. Since rpoN is essential for diazotrophic growth in all organisms examined thus far, the possibility of the existence of a second rpoN gene was examined. Using the R. sphaeroides rpoN gene as a probe in Southern blots revealed a second hybridizing fragment. Duplication of nitrogen fixation genes is not without precedent; as previously described, R. capsulatus contains a duplicated set of nif genes in regions A and B (34). In addition, Bradyrhizobium japonicum was recently found to possess two functional copies of rpoN; one gene is expressed constitutively, while the expression of the second rpoN gene is controlled by oxygen (36). In R. sphaeroides, there are two operons encoding duplicate copies of many of the Calvin cycle CO₂ fixation structural genes (16-19, 56). Interestingly, these operons appear to be located on different chromosomes (55). A comparable situation could exist for the rpoN genes.

Detailed comparisons of deduced amino acid sequences indicated that the RpoN proteins from *R. sphaeroides* and *R. capsulatus* belong to a different group from other RpoN proteins. The most striking difference between the two RpoN families is the virtual absence of region II in the *R. sphaeroides* and *R. capsulatus* RpoN proteins. Since it has been proposed that region II is involved in melting of the promoter (52), the absence of this domain in the *R. capsulatus* and *R. sphaeroides* RpoN proteins suggests that, at least in these proteins, region II is not critical. Despite the low overall similarity of the RpoN proteins of the purple nonsulfur bacteria and those of other organisms, most residues thought to be functionally important are conserved.

In the NifU amino acid sequences of A. vinelandii and K. pneumoniae, a possible heme-binding site was identified (40). The His residue in the consensus sequence is the ligand of the heme group and is not conserved in the Anabaena sp. sequence (11), where it is replaced by a Leu residue. In R. sphaeroides, the heme-binding site is completely absent. Since the heme-binding site is conserved in only two NifU proteins, we conclude that NifU is not a heme-binding protein. The ORF upstream from nifU contains a consensus ATP phosphate-binding site (22), suggesting that the gene product of this ORF may bind ATP or other nucleotides. Although no function could be assigned to this ORF, the fact that it is highly conserved among unrelated organisms suggests that it is a functionally important protein.

The product of the reaction catalyzed by NifV has been identified as homocitrate (25). Substrates of the reaction catalyzed by NifV leading to the formation of homocitrate have not been identified. However, as has been noted by Evans et al. (14), the NifV primary structure is similar to that of isopropylmalate synthase, an enzyme which catalyzes the formation of isopropylmalate from acetyl coenzyme A (acetyl-CoA) and α -ketovalerate. Homocitrate synthase present in the lysine biosynthetic pathway catalyzes the formation of homocitrate from acetyl-CoA and α -ketoglutarate, a similar reaction. It is therefore reasonable to assume that the substrates for NifV are acetyl-CoA and α -ketoglutarate.

By using a profile search, a sequence similar to region 2 of the NifV-LeuA sequence alignment was identified in pyruvate carboxylase and oxaloacetate decarboxylase. Upon inspection of the amino acid sequence of these two proteins, a sequence similar to region 1 was also identified at the proper spacing relative to region 2. The conservation of these regions suggests that they play an important role in catalysis, substrate binding, or both. It has been suggested that the fragment of pyruvate carboxylase containing regions 1 and 2 (residues 550 to 900) is involved in the binding of the substrate, pyruvate (38). Pyruvate has a keto acid group in common with oxaloacetate, α -ketoglutarate, and α -ketovalerate, the substrates of the other enzymes used in the sequence alignment. Therefore, on the basis of the common chemical nature of the substrates of these enzymes, we propose that region 1 and region 2 are involved in the binding of the keto acid group of the respective substrates. Obviously, subsequent enzymological investigations are required to confirm these hypotheses.

ACKNOWLEDGMENTS

We thank Judy Wall and Robert Haselkorn for strains and plasmids used in this study.

This work was supported by USDA grants 89-37262-4566 and 91-37306-6325 from the National Research Initiative Competitive Grants Program.

REFERENCES

1. Ahombo, G., J. C. Willison, and P. M. Vignais. 1986. The *nifHDK* genes are contiguous with a *nifA*-like regulatory gene in *Rhodobacter capsulatus*. Mol. Gen. Genet. 205:442-445.

- Arigoni, F., P. A. Kaminski, H. Hennecke, and C. Elmerich. 1991. Nucleotide sequence of the *fixABC* region of *Azorhizobium caulinodans* ORS571: similarity of the *fixB* product with eukaryotic flavoproteins, characterization of *fixX*, and identification of *nifW*. Mol. Gen. Genet. 225:514–520.
- 3. Arnold, W., A. Rump, W. Klipp, U. B. Priefer, and A. Pühler. 1988. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. J. Mol. Biol. 203:715–738.
- Avtges, P., R. G. Kranz, and R. Haselkorn. 1985. Isolation and organization of genes for nitrogen fixation in *Rhodopseudomo*nas capsulata. Mol. Gen. Genet. 201:363–369.
- Beltzer, J. P., L.-F. Chang, A. E. Hinkkanen, and G. B. Kohlhaw. 1986. Structure of yeast LEU4. J. Biol. Chem. 261:5160-5167.
- 6. Berger, D. K., D. R. Woods, and D. E. Rawlings. 1990. Complementation of *Escherichia coli* σ^{54} (NtrA)-dependent formate hydrogenlyase activity by a cloned *Thiobacillus ferroaxidans ntrA* gene. J. Bacteriol. 172:4399–4406.
- Beynon, J., A. Ally, M. Cannon, F. Cannon, M. Jacobson, V. Cash, and D. Dean. 1987. Comparative organization of nitrogen fixation-specific genes from *Azotobacter vinelandii* and *Klebsiella pneumoniae*: DNA sequence of the *nifUSV* genes. J. Bacteriol. 169:4024–4029.
- Beynon, J., M. Cannon, V. Buchanan-Wollaston, A. Ally, R. Setterquist, D. Dean, and F. Cannon. 1988. The nucleotide sequence of the *nifT*, *nifY*, *nifX* and *nifW* genes of K. pneumoniae. Nucleic Acids Res. 16:9860.
- 9. Beynon, J., M. Cannon, V. Buchanan-Wollaston, and F. Cannon. 1983. The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. Cell **34**:665–671.
- Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening of recombinant plasmid DNA. Nucleic Acids Res. 7:1513–1523.
- Borthakur, D., M. Basche, W. J. Buikema, P. B. Borthakur, and R. Haselkorn. 1990. Expression, nucleotide sequence and mutational analysis of two open reading frames in the *nif* gene region of *Anabaena* sp. strain PCC 7120. Mol. Gen. Genet. 221:227-234.
- 12. de Bruijn, F. J., and F. M. Ausubel. 1983. The cloning and characterization of the glnF (ntrA) gene of Klebsiella pneumoniae: role of glnF (ntrA) in the regulation of nitrogen fixation (nif) and other nitrogen assimilation genes. Mol. Gen. Genet. 192:342-353.
- Ditta, G., T. Schmidhauser, E. Yakobson, P. Lu, X.-W. Liang, D. R. Finlay, D. Guiney, and D. R. Helinski. 1985. Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. Plasmid 13:149– 153.
- Evans, D. J., R. Jones, P. R. Woodley, J. R. Wilborn, and R. L. Robson. 1991. Nucleotide sequence and genetic analysis of the Azotobacter chroococcum nifUSVWZM gene cluster, including a new gene (nifP) which encodes a serine acetyltransferase. J. Bacteriol. 173:5457-5469.
- Figurski, D., R. Meyer, D. S. Miller, and D. R. Helinski. 1976. Generation in vitro of deletions in the broad host range plasmid RK2 using phage mu insertions and a restriction endonuclease. Gene 1:107-119.
- Gibson, J. L., J.-H. Chen, P. A. Tower, and F. R. Tabita. 1990. The form II fructose 1,6-bisphosphatase and phosphoribulokinase genes form part of a large operon in *Rhodobacter sphaeroides*: primary structure and insertional mutagenesis analysis. Biochemistry 29:8085–8093.
- Gibson, J. L., D. L. Falcone, and F. R. Tabita. 1991. Nucleotide sequence, transcriptional analysis, and expression of genes encoded within the form I CO₂ fixation operon of *Rhodobacter* sphaeroides. J. Biol. Chem. 266:14646–14653.
- Gibson, J. L., and F. R. Tabita. 1987. Organization of phosphoribulokinase and ribulose bisphosphate carboxylase/oxygenase genes in *Rhodopseudomonas* (*Rhodobacter*) sphaeroides. J. Bacteriol. 169:3685-3690.
- Gibson, J. L., and F. R. Tabita. 1988. Localization and mapping of CO₂ fixation genes within two gene clusters in *Rhodobacter*

sphaeroides. J. Bacteriol. 170:2153-2158.

- 19a. Gribskov, M., M. McLachlan, and D. Eisenberg. 1987. Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA 84:4355–4358.
- Helmann, J. D., and M. J. Chamberlin. 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. 57: 839–872.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351–359.
- Higgins, C. F., I. D. Hiles, G. P. C. Salmond, D. R. Gill, J. A. Downie, I. J. Evans, I. B. Holland, L. Gray, S. D. Buckel, A. W. Bell, and M. A. Hermodson. 1986. A family of related ATP-binding subunits coupled to many distinct processes in bacteria. Nature (London) 323:448-450.
- 23. Hirschman, J., P.-K. Wong, K. Sei, J. Keener, and S. Kustu. 1985. Products of nitrogen regulatory genes ntrA and ntrC of enteric bacteria activate glnA transcription in vitro: evidence that the ntrA product is a σ factor. Proc. Natl. Acad. Sci. USA 82:7525-7529.
- Hoover, T. R., J. Imperial, P. W. Ludden, and V. K. Shah. 1989. Homocitrate is a component of the iron-molybdenum cofactor of nitrogenase. Biochemistry 28:2768–2771.
- Hoover, T. R., A. D. Robertson, R. L. Cerny, R. N. Hayes, J. Imperial, V. K. Shah, and P. W. Ludden. 1987. Identification of the V-factor needed for synthesis of the iron-molybdenum cofactor of nitrogenase as homocitrate. Nature (London) 329: 855-857.
- Hoover, T. R., V. K. Shah, G. P. Roberts, and P. W. Ludden. 1986. *nifV*-dependent, low-molecular-weight factor required for in vitro synthesis of iron-molybdenum cofactor of nitrogenase. J. Bacteriol. 167:999–1003.
- Inouye, S., M. Yamada, A. Nakazawa, and T. Nakazawa. 1989. Cloning and sequence analysis of the *ntrA* (*rpoN*) gene of *Pseudomonas putida*. Gene 85:145–152.
- Ishimoto, K. S., and S. Lory. 1989. Formation of pilin in *Pseudomonas aeruginosa* requires the alternative σ factor (*rpoN*) of RNA polymerase. Proc. Natl. Acad. Sci. USA 86:1954-1957.
- Jacobson, M. R., K. E. Brigle, L. T. Bennett, R. A. Setterquist, M. S. Wilson, V. L. Cash, J. Beynon, W. E. Newton, and D. R. Dean. 1989. Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. J. Bacteriol. 171:1017– 1027.
- Jacobson, M. R., V. L. Cash, M. C. Weiss, N. F. Laird, W. E. Newton, and D. R. Dean. 1989. Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii*. Mol. Gen. Genet. 219:49–57.
- 31. Jones, R., and R. Haselkorn. 1989. The DNA sequence of the *Rhodobacter capsulatus ntrA*, *ntrB* and *ntrC* gene analogues required for nitrogen fixation. Mol. Gen. Genet. 215:507-516.
- 32. Kay, R., and J. McPherson. 1987. Hybrid pUC vectors for addition of new restriction enzyme sites to the ends of DNA fragments. Nucleic Acids Res. 15:2778.
- 33. Klipp, W. 1990. Organization and regulation of nitrogen fixation genes in *Rhodobacter capsulatus*, p. 467–474. *In* P. M. Gresshoff, L. E. Roth, G. Stacey, and W. E. Newton (ed.), Nitrogen fixation: achievements and objectives. Chapman & Hall, New York.
- Klipp, W., B. Masepohl, and A. Pühler. 1988. Identification and mapping of nitrogen fixation genes of *Rhodobacter capsulatus*: duplication of a *nifA-nifB* region. J. Bacteriol. 170:693-699.
- 35. Köhler, T., S. Harayama, J.-L. Ramos, and K. N. Timmis. 1989. Involvement of *Pseudomonas putida* RpoN σ factor in regulation of various metabolic functions. J. Bacteriol. 171:4326-4333.
- Kullik, I., S. Fritsche, H. Knobel, J. Sanjuan, H. Hennecke, and H.-M. Fischer. 1991. *Bradyrhizobium japonicum* has two differentially regulated, functional homologs of the σ⁵⁴ gene (*rpoN*). J. Bacteriol. 173:1125–1138.
- 37. Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss. 1989. Expression of σ^{54} (*ntrA*)-dependent genes is probably united by a common mechanism. Microbiol. Rev. 53:367–376.
- 38. Lim, F., C. P. Morris, F. Occhiodoro, and J. C. Wallace. 1988.

Sequence and domain structure of yeast pyruvate carboxylase. J. Biol. Chem. **263**:11493–11497.

- Masepohl, B., W. Klipp, and A. Pühler. 1988. Genetic characterization and sequence analysis of the duplicated *nifA/nifB* gene region of *Rhodobacter capsulatus*. Mol. Gen. Genet. 212:27-37.
- 40. Mathews, F. S. 1985. The structure, function and evolution of cytochromes. Prog. Biophys. Mol. Biol. 45:1-56.
- Merrick, M., J. Gibbins, and A. Toukdarian. 1987. The nucleotide sequence of the sigma factor gene ntrA (rpoN) of Azotobacter vinelandii: analysis of conserved sequences in NtrA proteins. Mol. Gen. Genet. 210:323–330.
- 42. Merrick, M. J., and J. R. Gibbins. 1985. The nucleotide sequence of the nitrogen-regulation gene ntrA of Klebsiella pneumoniae and comparison with conserved features in bacterial RNA polymerase sigma factors. Nucleic Acids Res. 13:7607-7620.
- Merrick, M. J., and W. D. P. Stewart. 1985. Studies on the regulation and function of the *Klebsiella pneumoniae ntrA* gene. Gene 35:297-303.
- Mulligan, M. E., and R. Haselkorn. 1989. Nitrogen fixation (nif) genes of the cyanobacterium Anabaena species strain PCC 7120. J. Biol. Chem. 264:19200-19207.
- 45. Ormerod, J. G., K. D. Ormerod, and H. Gest. 1961. Light dependent utilization of organic compounds and photoproduction of hydrogen by photosynthetic bacteria: relationship with nitrogen metabolism. Arch. Biochem. Biophys. 94:449–463.
- Paul, W., and M. Merrick. 1989. The roles of the nifW, nifZ and nifM genes of Klebsiella pneumoniae nitrogenase biosynthesis. Eur. J. Biochem. 178:675-682.
- 47. Platt, T. 1986. Transcriptional termination and the regulation of gene expression. Annu. Rev. Biochem. 55:339-372.
- Prentki, P., and H. M. Krisch. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene 29:303–313.
- Ricca, E., and J. M. Calvo. 1990. The nucleotide sequence of leuA from Salmonella typhimurium. Nucleic Acids Res. 18: 1290.
- 49a. Römermann, D., J. Warrelmann, R. A. Bender, and B. Friedrich. 1989. An *rpoN*-like gene of *Alcaligenes eutrophus* and *Pseudomonas facilis* controls expression of diverse metabolic pathways, including hydrogen oxidation. J. Bacteriol. 171:1093-1099.
- 50. Ronson, C. W., B. T. Nixon, L. M. Albright, and F. M. Ausubel.

1987. Rhizobium meliloti ntrA (rpoN) gene is required for diverse metabolic functions. J. Bacteriol. 169:2424–2431.

- 51. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 52. Sasse-Dwight, S., and J. D. Gralla. 1990. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor σ^{54} . Cell 62:945–954.
- Schwartz, E., D. Oesterhelt, H. Reinke, K. Beyreuther, and P. Dimroth. 1988. The sodium ion translocating oxaloacetate decarboxylase of *Klebsiella pneumoniae*. J. Biol. Chem. 263: 9640-9645.
- Simon, R., U. Priefer, and A. Pühler. 1983. A broad host range mobilization system for in vitro engineering: transposon mutagenesis in Gram-negative bacteria. Bio/Technology 1:784–790.
- 55. Suwanto, A., and S. Kaplan. 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. J. Bacteriol. 171:5850-5859.
- Tabita, F. R. 1988. Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. Microbiol. Rev. 52:155-189.
- 57. Tabita, F. R., J. L. Gibson, D. L. Falcone, B. Lee, and J.-H. Chen. 1990. Recent studies on the molecular biology and biochemistry of CO₂ fixation in phototrophic bacteria. FEMS Microbiol. Rev. 87:437-444.
- van Slooten, J. C., E. Cervantes, W. J. Broughton, C. H. Wong, and J. Stanley. 1990. Sequence and analysis of the *rpoN* sigma factor gene of *Rhizobium* sp. strain NGR234, a primary coregulator of symbiosis. J. Bacteriol. 172:5563-5574.
- 59. Vieira, J., and J. Messing. 1987. Production of single stranded DNA. Methods Enzymol. 153:3-11.
- Wall, J. D., J. Love, and S. P. Quinn. 1984. Spontaneous Nifmutants of *Rhodopseudomonas capsulata*. J. Bacteriol. 159: 652–657.
- 61. Weaver, K. E., and F. R. Tabita. 1983. Isolation and partial characterization of *Rhodopseudomonas sphaeroides* mutants defective in the regulation of ribulose bisphosphate carboxylase/ oxygenase. J. Bacteriol. 156:507-515.
- 62. Weaver, K. E., and F. R. Tabita. 1985. Complementation of a *Rhodopseudomonas sphaeroides* ribulose bisphosphate carboxylase-oxygenase regulatory mutant from a genomic library. J. Bacteriol. 164:147–154.