Nucleotide Sequence of *pvdD*, a Pyoverdine Biosynthetic Gene from *Pseudomonas aeruginosa*: PvdD Has Similarity to Peptide Synthetases

TONY R. MERRIMAN, † MARILYN E. MERRIMAN, † AND IAIN L. LAMONT*

Department of Biochemistry and Centre for Gene Research, University of Otago, Dunedin, New Zealand

Received 8 August 1994/Accepted 9 October 1994

Pseudomonas aeruginosa secretes a fluorescent siderophore, pyoverdine, when grown under iron-deficient conditions. Pyoverdine consists of a chromophoric group bound to a partly cyclic octapeptide. As a step toward understanding the molecular events involved in pyoverdine synthesis, we have sequenced a gene, *pvdD*, required for this process. The gene encodes a 2,448-residue protein, PvdD, which has a predicted molecular mass of 273,061 Da and contains two highly similar domains of about 1,000 amino acids each. The protein is similar to peptide synthetases from a range of bacterial and fungal species, indicating that synthesis of the peptide moiety of pyoverdine proceeds by a nonribosomal mechanism. The *pvdD* gene is adjacent to a gene, *fpvA*, which encodes an outer membrane receptor protein required for uptake of ferripyoverdine.

Pseudomonas aeruginosa is an opportunistic pathogen which infects injured, immunodeficient, or otherwise compromised patients. The bacteria secrete a siderophore, pyoverdine, which is likely to play an important role in infection by competing with transferrin for iron in order to overcome the iron-withholding defense mechanism present in mammals (3, 11, 42). Almost all isolates of *P. aeruginosa* from infected patients secrete pyoverdine (9, 22, 29), and it has been shown that pyoverdine is present in the sputa of cystic fibrosis patients infected with *P. aeruginosa* (21). A mutant of *P. aeruginosa* which is unable to synthesize pyoverdine showed reduced virulence in an animal model of infection (24).

Pyoverdine from *P. aeruginosa* PAO is a water-soluble, yel-low-green fluorescent compound. It consists of a dihydroxyquinoline chromophoric group linked to an eight-residue partly cyclic peptide (D-Ser-L-Arg-D-Ser-LN⁵-OH-Orn-L-Lys- $L-N^5$ -OH-Orn-L-Thr-L-Thr) via the N-terminal serine, with a small dicarboxylic acid attached to the chromophore. Iron complexation is thought to occur through oxygen atoms present on the dihydroxyquinoline and two hydroxamate units supplied by the L-N⁵-OH-Orn residues. Several Pseudomonas species produce similar compounds, variously called pyoverdines or pseudobactins, all of which have the same dihydroxyquinoline group but differ in the nature of the attached peptide (reviewed in references 1 and 7), and it is likely that all of these are synthesized by similar mechanisms. Synthesis of the chromophoric group begins with condensation of D-tyrosine and L-2,4-diaminobutyric acid (7), with glutamic acid being the precursor of the amide-linked dicarboxylic acid (50). It has been suggested that biosynthesis of the peptide moiety of pyoverdine occurs by a nonribosomal mechanism (30).

Little is known about the molecular nature of enzymes involved in the biosynthesis of pyoverdine in *P. aeruginosa* or any other pseudomonad. Recently, *pvdA*, which encodes an enzyme (L-ornithine N^5 -oxygenase) responsible for catalyzing the hydroxylation of ornithine, an early step in pyoverdine biosynthesis in *P. aeruginosa*, has been characterized at the molecular level (55). The *pbsC* gene, which is involved in synthesis of pseudobactin by *Pseudomonas* sp. strain M114, has also been sequenced (2). To gain further insight into the molecular mechanism of pyoverdine biosynthesis, we have determined the nucleotide sequence of the *pvdD* (pyoverdine synthetase D) gene from *P. aeruginosa*.

Nucleotide sequence of pSOT1 and identification of pvdD. We have previously shown that restriction fragments D and E of pSOT1 (Fig. 1) are part of a locus required for synthesis of pyoverdine (34, 35). To gain insight into the nature of the corresponding pyoverdine biosynthetic gene(s), the DNA sequence of fragments B through F of pSOT1 was determined. DNA fragments to be sequenced were subcloned into M13mp18 or mp19 (60), using standard procedures (4), with Escherichia coli TG1 (16) used as the host strain. Doublestrande M13 DNA was prepared (25), and partial deletions of the cloned fragments were generated by using exonuclease III (Erase-a-Base system; Promega) or by using restriction enzymes which cleaved the DNA within the insert and within the M13 polylinker. Oligonucleotide primers were used to sequence regions not covered by these strategies; these were synthesized by using a model 180B DNA synthesizer (Applied Biosystems, Inc., Foster City, Calif.) or were purchased from Macromolecular Resources (Colorado State University). Single-stranded DNA of M13 subclones was prepared (5) and sequenced, using a Taq DNA polymerase sequencing kit (Amersham) or Sequenase Version 2.0 (United States Biochemical, Cleveland, Ohio), both with 7-deaza-dGTP. Sequence ambiguities were resolved by using the Sequenase Version 2.0 kit with dITP instead of 7-deaza-dGTP. A model 373A automated DNA sequencer (Applied Biosystems) was used to sequence across all of the cloning sites in pSOT1. Ninety-four percent of the sequence data were obtained from both strands. The remaining single-stranded data were obtained from at least two gels and were clear and unambiguous. The DNA sequence was analyzed on a MicroVaxII computer system by using various softwares: Ssedit (48), Vtutin (47), Codonuse (45), NLDNA (48), Map Zap (41), Diagon (44), Dbextract (46), and the

^{*} Corresponding author. Mailing address: Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand. Phone: 64-3-479-9869. Fax: 64-3-479-7866. Electronic mail address: ilamont@sanger.otago.ac.nz.

[†] Present address: Wellcome Centre for Human Genetics, Nuffield Orthopaedic Centre, Oxford OX3 7BN, United Kingdom.



FIG. 1. (A) Clone pSOT1. pSOT1 is the largest (~9.6-kb) SalI fragment of λ OT1 (35) cloned into pUC9 (54) and contains DNA required for pyoverdine synthesis (34, 35). Restriction fragments are labelled by using the same nomenclature as previously (36). The locations and orientations of two genes (*pvdD* and *fpvA*) sequenced in this study are shown. Ba, BamHI; P, PstI; S, SalI.

FoldRNA program from the version 7 UNIX of the Genetics Computer Group package (12).

The sequenced DNA contained a large open reading frame (ORF) of 7,374 bp (Fig. 2). Downstream from this ORF and oriented in the opposite direction was a partial ORF of 792 bp. Both ORFs had patterns of codon usage which are very similar to those of other genes from *P. aeruginosa* (58), indicating that they are likely to encode proteins.

Mutations mapping to restriction fragments D and E, which are entirely contained within the 7.4-kb ORF, prevent synthesis of pyoverdine by *P. aeruginosa* (34, 35). This finding shows that this gene is required for synthesis of pyoverdine, and it has been named *pvdD*, following the assignment of *pvdA*, *pvdB*, and *pvdC* to other genes involved in synthesis of pyoverdine (55). Analysis of codon usage shows that translation is likely to start at a GTG codon (Fig. 2), which nonetheless results in methionine being incorporated as the first amino acid in the protein (18). This is six bases downstream from a Shine-Dalgarno sequence which is complementary to seven of the eight bases at the 3' end of 16S rRNA. The *pvdD* gene encodes a protein of 2,448 residues with a predicted molecular mass of 273,061 Da.

Two transcriptional start sites have been identified upstream from pvdD (36) (Fig. 2). Two inverted repeat sequences downstream from pvdD could function as transcriptional terminators for the gene. Both are capable of forming stem-loop structures; one is present 21 bp downstream from the predicted TGA termination codon (Fig. 2) and comprises an 8-bp stem and 4-bp loop ($\Delta G = -14.8$ kJ/mol, as calculated by using the FoldRNA program), and the second is 122 bp downstream from the termination codon (Fig. 2) and consists of a 12-bp stem and 3-bp loop ($\Delta G = -19.1$ kJ/mol). Both inverted repeats are followed by multiple thymidine nucleotides; this is characteristic of transcription termination sequences which function independently of the Rho termination factor (59). Transcription of *pvdD* could be terminated at either of these sites, giving a transcript of about 7.5 kb. The size of the transcript corresponding to *pvdD* was previously estimated to be 5 kb following Northern (RNA) analysis (36), but the sequence of the gene indicates that this is an underestimate, presumably due to the high degree of instability of this transcript (36).

Direct repeats within *pvdD*. Diagon analysis indicated that the *pvdD* gene contains two direct repeats of about 3 kb each. The repeats span bases 139 to 3180 and 3331 to 6363 of the sequence presented in Fig. 2 and are 91.8% identical. The deduced amino acid sequences of the repeats are 90.7% identical and 94.7% similar (Fig. 3), with one three-amino-acid insertion in the first repeat 83 residues from the N-terminal end. The C-terminal regions of the repeats (amino acids 407 to 1040) were different at only seven positions, with four of the differences representing conserved changes. In contrast, there were tracts of relatively different sequence in the N-terminal parts of the repeats.

Identification of proteins similar to PvdD. The GenBank translated nucleic acid database (version 83) was searched by using FastA (32) for proteins showing sequence similarity to the first repeat of PvdD (1,013 amino acids). The protein showing greatest similarity was the product of the *pbsC* gene from Pseudomonas sp. strain M114, and this is discussed below. Significant similarity was also found to a family of enzymes including GrsB (53), TycA (57), SrfA (10), AcvA, PcbA, and PcbB (8, 27, 40), EntF (37), and AngR (14, 52). These enzymes are peptide synthetases, with the last two being involved in siderophore synthesis. After PbsC, the protein showing greatest similarity was GrsB, which is a well-characterized peptide synthetase from Bacillus brevis. This protein contains four domains, each responsible for the adenylation of a specific amino acid and its subsequent incorporation into the peptide antibiotic gramicidin (53). Each domain was similar to each of the repeats within PvdD, and an alignment of the PvdD repeats with the domains of GrsB is shown in Fig. 3.

The mechanism of nonribosomal peptide synthesis by GrsB and other peptide synthetases has been studied (reviewed in references 26 and 56). Each domain of a peptide synthetase recognizes a specific amino acid and activates this substrate by adenylation. It is thought that the amino acyladenylate is then covalently attached to the enzyme by a thioester bond between the carboxyl group of the amino acid and a thiol group, which is probably supplied by a 4'-phosphopantetheine cofactor. The amino acid is then incorporated into the peptide by the enzymatic formation of a peptide bond which joins it to the amino acid bound to the adjacent domain. This is coupled with scission of the thioester bond between the cofactor and the amino acid, regenerating the enzyme (26).

Several consensus motifs associated with these activities are present in different peptide synthetases (19), and all of these are present in PvdD (Fig. 3). From the extensive similarities found with peptide synthetases, including the presence of consensus sequence motifs (Fig. 3), it is very likely that each domain of PvdD recognizes, activates, and incorporates a specific amino acid into the peptide moiety of pyoverdine.

PvdD also contains a thioesterase active-site motif. The C-terminal segment of PvdD (336 amino acids), which is not part of either repeat sequence, was independently compared with sequences in the translated GenBank database. The only significant similarities were to EntF from E. coli (37) and the AcvA protein from Penicillium chrysogenum (40), both of which are peptide synthetases. In each case, similarity was centered around a thioesterase active-site motif (GXSXG [20], where X is any amino acid) which is present at residues 2240 to 2244 in PvdD. Similar sequence motifs are also present downstream from the domains of gramicidin S synthetase and surfactin synthetase which incorporate the C-terminal amino acids into the peptides (10, 53). Motifs of this sort are thought to be involved in termination of peptide synthesis by cleavage of the thioester bond between the 4'-phosphopantetheine cofactor of the final amino acid-activating domain and the last amino acid incorporated into the peptide (reference 10 and references therein). The presence of this motif indicates that PvdD is likely to have a thioesterase activity associated with release of the completed pyoverdine peptide, and this implies that the protein incorporates the two C-terminal threonines into the peptide. Further experiments will be required to investigate the possible thioesterase activity and to determine the amino acid substrates of PvdD.

Similarity of PvdD to PbsC. The protein found during database searching which was most similar to the first repeat of PvdD was the product of the *pbsC* gene of *Pseudomonas* sp. strain M114. This protein is predicted to contain 803 amino

-75 -51 1 --> 1 --> 5D -1 TGGCTANATCOCTGGA00000CTCAAAGTCTATCTGCCGATGAAAACTGATTGCCTAAGTCTCAACGC00CTAG GTGCAAGCACTCATAGAGAAGGTG M Q A L I E K V 1 50 75 Soctcoctttcococcagaaaggaaggcattggctgtcctgctcagcag S L S P Q E R K A L A V L L K Q GS Q G 26 CASGAGCGACAGTGGTTTCCTGGGCAACTGCAGGCGGAAAGGGGGGGCGACAGTGGGTTTCCTGGGCAACTGCAGGCGGAAAGGGGGGGCCTAACGAGTGGTGCTGCGGCGA Q E R Q W F L W Q L E P E S A X Y E I P S V L R L 51 R 76 CGTTTTCGCCTCGACGGCGACGAGGCGCGCCAGGAGATCGCCGCATCCATGGCATTGCCGTTGGATATCGTCGCG R L D G D E A R Q E I A A S M A L P R 101 F TTGGGGCCGCTGGAGGAGGAGGAGGAGGACGACGCGCGTTCGACCTGGAGCGGC L G P L E E G A L A R Q V E T T I A R P F D L E R 26 GGGCCGCTGCTGCGGGTGAGCCTGTGCGGGCTGGCCGAGGACGACCATGTGCTGGTGCTGGTGCAGCATCACATC G P L L R V S L L R L A E D D H V L V L V Q H H I ۍ 151 CGACGGTTGGTCGATGCAGGTATGGTCGAGGAACTGGTCCAGCGCTATAGTCGAGGGCT D G W S M Q V N V E E L V Q L Y A A Y S R G L GTGTC V S 176 E 201 775 CGGCCGCCCCGGTTCGCCAAAGCCAACTGGTGGCCCAGTTCATCCTGGATATTGATCTGTCCCAGGCG R P R P V R Q S E R G A Q F I L E L D I D L S Q A 51
276
925
950
975

CGCTACAGCGGGCAGGCGATATCOCGTGTCCGGGTGCGGATCGCCAATCGCAATCGCCAATCGCAATCGCAATCGCAATCGCAATCGCAATCGCCAATCGCAAT
351
1150
1175
1200

CCGGRAGECACHTCTTAGCCACACCECTETTCCAGGTGCTGTTCAACTACCAGAGCGAAGCCCGTGGCCAACGACCGAAGCCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGAGCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGTGGCAACGACGGAGCCGTGGCCGATGCCGAAGCCGAGCCGTGGCCGAGCCGTGGCCGACGAGCCGAGCCGCTGGCCGACGCCGTGGCCGACGACGAAGCCGAAGCCGAGCCGTGGCCGACGACGACGGAGCCGTGGCCGATGCCGAAGCCGAGCCGCAGGCCGTGGCCGACGACGAAGCCGAAGCCGAGCCGCTGGCCGACGACGAAGCCGAAGCCGAGCCGAGCCGAGGCCGTGGCCGACGAAGCCGAAGCCGAAGCCGAGGCCGTGGCCGAGCCGAGCGCAGGCCGCTGGCCGACGACGAAGCCGAAGCCGAAGCCGAGGCCGTGGCCGACGAAGCGAAGCCGAAGCGAAGCGAAGCCGAAGCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCCGAAGCGAAGCGAAGCGAAGCCGAAGCCGAAGCGAAGCGAAGCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCCGAAGCGAAGCCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCGAAGCAAG Q 401 GTGGAACGCCTGGCCGCCATGGCGCAACCTGTTGCGCGCATCGTCGCCAACCCACGACACGGCGCGGCGGCGGC V E R L A G H W R N L L R G I V A N P R Q R L G E TTGCCGCTGCTGGATGCGCCGGAGCGCCG L P L L D A P E R R 476 EGCEGGEAGACCETETECGAATGGAACCEGGECEAGEGEGAGTGEGGGTG R R Q T L S E M N P A Q R E C A V 1575 CAGGECACCTIGCAGCACGATTICGAGAAGAACAGGGGCAACGGCCACAGGGCAGGTGGGCGGTTGCGGCGGAGGAACAGGGA Q G T L Q Q R F E E Q A R Q R F Q A V A L I L D E 501 CANCEGTTGASCTACEGCEGAATCECCEGECCAATCECCEGECCCATCECCEGATCEGCCEGTEGEC Q R L S Y G E L N A R A N R L A H C L I A R G V G 526 <u>а</u> 551 GGCGGCGCCTACTGCCGTTGACCCGGCGGCGCAGAGGGCGCCTGGCGCATATCCTCGACGACAGTGGGGTA G G A Y L P L D P A A P E E R L A B I L D D S G V G 576 GGACTGGTGCGACGGCTACGCCGAGGCGATCCGCTCCCGACGCCTATCGGCGACAACCTGGCCTACGTGATC G L V L D G Y A E S D P L P T L S A D N L A Y V I G 626 GACTTCTACGGTCTGTGGGGGGAAGGGGTGACGGTGGCGTCCAACGAGGGGGTTCAAGGACGTGACGGCGTAAGCGACGTGACGGCGTGGGGGGTTCAAGGAACTGATG D F Y R L L C R E G V T V L N Q T P S A F K Q L H 26 D 726 A 751

ACCENERCEACEGTECACE	2425 TAACCTACCGTCCGGTGAGCGAGG	2450 CCGACCTGGAAGGTGCCCTGCTCACTCO	2475
ТЕТТVНV 801	TYRPVSEA	DLEGGLVSP	I
GGCGGGACCATCCCGGACC	2500 TGTCCTGGTACATCCTCGACCGTG	2525 ACCTGRACCCGGTGCCGCGCGCGCGGGGG	2550 GGGC
826 826	2575	2600 · · · · · · · · · · · · · · · · · ·	G 2625
GAGCIGIACATCGGICGCGG E L Y I G R A	CCGGGCTGGCGCGCGCGCTACCTGA G L A R G Y L R	GCCGCCCCGGGTTGAGTGCCACCCCCTT	CGTG
851	2650	2675	2700
CCGAACCCGTTCCCCGGCGG	GCGCCGGCGAGCGGCTGTACCGTA A G E R L Y R T	CCGGCGACCTGGCACGGTTCCAGGCGGA	TGGC G
876	2725	2750	2725
N I E Y I G R	I D E Q V K V R	GEGGETTEEGTATEGAACTGGGEGAGAT	E
GCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		2825 TGGCCCATGACGGAGTCGGCGGCACGCA	2850 ACTG
AALAGLA 926	GVRDAVVL	AEDGVGGTQ	L
GTGGGATACGTGGTGGCGG	2875 ACTCGGCGGAGGATGCCGAGCGTC	2 90 0 TGCGGGAGTCGCTGCGGGAGTCGCTGAA	2925 GCGG
951 951	S A E D A E R L	RESLRESLK	R 2000
CACCTGCCGGACTACATGG	TECCOECECACCTEATECTECTOE	AGCGGATGCCGCTGACGGTCAATGGCAA	GCTC
976	3025	3050	3075
GACCGGCAGGCGTTGCCGCI D R Q A L P Q	AACCOGATGCGAGCCTGTCGCAAC P D A S L S Q Q	AGGCCTATCGAGCGCCGGGTAGCGAGCT	GGAG E
1001	3100	3125	3150
Q R I A A I W	SEILG VER	CGGTCGGCCTGGACGACAACTTCTTCGA VGLDDNFFE	ACTG L
GOCOGTCATTCCTTCCTCC	3175 CTACCCGGGTGATTTCTCGGGTTTC	3200	3225 GANG
G G H S L L A	TRVISRVR	QEQQLDASL	K
GCGTTGTTCGAGCGGCCGG	3250 TTCTGGAAGCGTTCGCCCAGGGAT	3275 TGGAACGCACGACGGATGCGGTCTCGAC	3300 GAT7
ALFERPV 1076	LEAFAQGE	ERTTDAVST	I
COCCTTCCCCATCCCCACC	AACCETTGGCACTGTCCTTCGCTC	AGGAGCGTCAGTGGTTCCTCTGGCAACT	3375 GGA0
1101	3400	3425	2450
$\begin{array}{cccc} cccccccccccccccccccccccccccccccc$	ATATTCCGAGTGCCTTGCGCCTAC	GCGGGCGGCTGGACGTGGATGCCTTGCA	ACGC R
1126	3475	3500	3525
AGCTTCGACAGCCTGGTCGA S F D S L V A	CGCGGCATGAAACCTTGCGTACCC R H E T L R T R	CCTTCCCGCTGGAGGGGGGGGGCGTTCGTA	CC AC Q
1151	3550	3575	3600
Q V Q P A V S	V S I E R E Q F	GEEGLIERI	Q
GOCATOGTTGTGCAGOCAT	3625 TCGACCTGGAACGGGGGCCGCTGC	3650 TGCGGGTGAACCTGTTGCAACTGGCCGA	3675 GGAG
			n
A I V V Q P F 1201	DLERGPLI	RVNLLQLAE	
A I V V Q P F 1201 GACCATGTACTGGTGCTGG	D L E R G P L L 3700 TCCASCACCACATCGTGTCCGATG	R V N L L Q L A E 3725 STTGGTCGATGCAGGTGATGGTCGAGGA	3750 ACTO
A I V V Q P F 1201 GACCATGTACTGGTGCTGG D H V L V L V 1226	D L E R G P L I 3700 TCCACCACCACATCGTGTCCCATG Q H H I V S D G 2775	R V N L L Q L A E 3725 SCTTOGTICGATCCACGGTCGATCCGACCGA C N S M Q V M V E E 3800	3750 ACTO L
A I V V Q P F 1201 GACCATGTACTGGTGGTGGT D H V L V L V 1226 GTCCAGCTCTATGCCCCCTT V O L Y A A Y	D L E R G P L I 3700 TCCAOCACCACATCGTGTCCGATG Q H H I V S D G 3775 ATACCCAAGGGCTCGACGTGGTGT S O G L D V V I	R V N L L Q L A E 3725 IGTGGTCGATGCAGGTGATGGTGGAGGA I N S M Q V M V E E 3800 TGCCAGCCTCCCGATCCAGTACGCGGGA . P A L P I O Y A D	3750 ACTO L 3825 CTAC
A I V V Q P F 1201 GACCATCTACTOGTCGTCGT D E V L V L V 1226 GTCCAGCTCTATGCCGCCT. V Q L Y A A Y 1251	D L E R G P L I 3700 TCCACCACACATCGTGTGCCAATG Q H H I V S D G 3775 ATAGCGAAGGGCTCGACGTGGTGT S Q G L D V V I 3850	, R V N L L Q L A E 3725 STTGGTCGATGCAGGTGATGGTCGAGGA S M S M Q V M V E E 3800 TGCCAGGCCTGCCGATCCAGTAGGCGGA . P A L P I Q Y A D 3875	3750 ACTO L 3825 CTAC Y 3900
A I V V Q P F 1201 GACCATCTACTOCTGCTGCT D E V L V L V 1226 GTCCAGCTCTATGCCGCCT. V Q L Y A A Y 1251 CCCCTGTGCGAGCGCAGCTT A L M Q R S M	D L E R G P L I 3700 TCCACCACCACATCGTGTCCGATG Q E E I V S D G 3775 S Q G L D V V I 3850 GATGGAGGCGGGGGGAAAAGAGGAG M E A G E K E B	, R V N L L Q L A E 3725 STTGGTCGATGCAGGATGGTCGAGGA S M S M Q V M V E E 1900 TGGCAGCCTGCCGATCCAGTAGGGGGA . P A L P I Q Y A D 3875 SGCCAGTTGGCGTACTGGACCGGCCTGCT L Q L A Y M T G L L	3750 ACTO L 3825 CTAC Y 3900 GGGC G
A I V V Q P F 1201 GACCATGTACTGGTGGTGGT D H V L V L V 1226 GTCCAGCTCTATGGCGCCT V Q L Y A A Y 1251 GCCCTGTGCCAGCGCAGCT A L M Q R S M 1276	D L E R G P L I 3700 TCCASCACCACATCGTGTCCGATG Q H H I V S D G 3775 S Q G L D V V I 3850 OGATGGAGCCCGGGGAAAAGCACG M E A G E K E R 3925	, R V N L L Q L A E 3725 STTGGTCGATCCAGGTGATGGTCGAGGA S M S M Q V M V E E 3800 TGGCCAGCCTAGCCGATCCAGTACGGGGA P A L P I Q Y A D 1875 SGCCAGTTGGCGTACTGGACCGGCCTGCT K Q L A Y M T G L L 3850	3750 ACTO L 3825 CTAC Y 3900 GGGC G GGGC G 3975
A I V V Q P F 1201 CACCATETACTOCTOCTO D H V L V L V 1226 STCCASCTCTATGCCOCCT. V Q L Y A A Y 1251 CCCCTCTGCCASCGCASCT A L W Q R S W 1276 CCCCCTCTGCCASCGCASCT C C Q P V L E 1301	D L E R G P L I 3700 TCCACCACCACATCGTGTCCGATG Q H H I V S D G 3775 S Q G L D V V I 3850 GCATGCACCGCGGGGAAAAGACG M E A G E K E R 3925 AGTTCCCGTTCGATCGACGCGCGCGT	, R V N L L Q L A E 3725 STTGSTCGATCATGCAGGATGGTCGAGGA S M S M Q V M V E E 3800 1875 3875 3875 3875 3876 3875 3876 3950 3050	3750 ACTO L 3825 CTAC 3825 CTAC 3900 GGGC GGGC GGGC GGGC GGGC GGGC L
A I V V Q P F 1201 GACCATGTACTGGTGCTGGT D H V L V L V 1226 GTCCAGCTCTATGGCGCCT. V Q L Y A A Y 1251 GCCCTGTGGCAGCGCAGCTT A L M Q R S M 1276 GCGAGCAGCGGTGCTGGG G E Q P V L E 1301 GGTTTGAGCTATGGGGGG	D L E R G P L I 3700 TCCACCACCACATCGTGTCCGATG Q H H I V S D G 3775 TAGCCAAGGGCTCGACGTGGTGT S Q G L D V V I 3850 GGATGGAGGCGGGGGAAAAGGACG M E A G E K E R 3925 AGTTGCGATGGAGGCGGTGAGAGCCT L P F D R P R P 4000 AACTGGTCGAGGCCGTGAGAGCCT	, R V N L L Q L A E 3725 STTGGTCGATGCATGCTGATGGTCGAGGA ; M S M Q V M V E E 3800 1860 1875 3875 3875 3875 3875 3875 3950 3876 4 Q L A Y M T G L L 3950 39	3750 ACTO L 3825 CTAC Y 3900 G G G G G G G G TTO L 4050 CATO
A I V V Q P F 1201 GACCATGTACTGGTGCTGGT D E V L V L V 1226 GTCCAGCTCTATGGCGCCT. V L V L V 1226 GTCCAGCTCTATGGCGCCCT. Y A A Y 1251 GCCCTGTGGCAGCGCAGCTT A L N Q R S M 1276 E Q P Y L E 1301 GCTTCGAGCTATGCGAGGG G P Q L E 1301 GCTTTCGAGCTATGCGAGGG F E S R E 1326	D L E R G P L I 3700 G H H I V S D G 3775 X G G L D V V I 3850 GGATGGAGGCGGGGAAAAGAGC M E A G E K E R 3925 3925 CTTGCGTTGGATCGGCGGCGTC L P F D R P R P 4000 ACTGGTCGAGGCCGTGAAAGCCT L V E A V R A I	, R V N L L Q L A E 3725 3726 37607 37607 37607 37607 37607 3770	3750 ACTC L 3825 CTAC S 3900 GGGC G GGC G G 3975 L 4050 CATC M
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	D L E R G P L I 3700 G H H I V S D G 3775 NASCCARGUECACATCGATGTCCGATG G H H I V S D G 3775 S Q G L D V V I 3850 GGATGGAGGCGGGGAAAAGAGC M E A G E K E R 3925 ACTTOCCGTTCGATCGGCGGCGTC L P F D R P R P ACTTOCCGTTCGATCGGCGGCGACACCCT L V E A V R A I 4075 AGGCGCTGTTGTATCGCTACACCG	, R V N L L Q L A E 3725 STTGGTCGATCCAGGATGCAGGATGGTCGAGGA ; M S M Q V M V E E 3800 3875 SGCCAGTGGCGTACCGGATCCAGGCCTGGT L Q L A Y M T G L L 3950 CAGCCGCGGAAGACCATCGTGCGCCCCA · A R Q S B R G A Q 4025 TGGCCCGGTGAAGCGCGAAGACGCCACTAGTTT . A Q R E G A S S F 4100 SGGCAGGCGATATCCGTGTCGCCCTGCTCCCC	33750 ACTC 1 38255 CTAC 33900 GGGC G GGC G GGC CATC M 41255 GATC M
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	D L E R G P L L 3700 G R R I V S D G 3775 S Q G L D V V I 3850 GATGGCGGGGGGAAAAGAGG M E A G E K E R 3925 GATGGCGGGGGGGGAAAGGAG L P F D R P R P 4000 ANCTGGTCGAGGCCGTGAGAGGCCT L V E A V R A I 4075 AGGGGCTGTTGATCGCTACAGCC A L L Y R Y S C 4150	, R V N L L Q L A E 3725 3726 3727 3800 3800 3875 38	3750 ACTO L 3825 CTAC Y 3900 GGGG G G 3975 CTAC L 4050 CATO M 4125 GATO I 4200
A I V V Q P F 1201 GACCATGTACTGGTGCTGGT D E V L V L V 1226 GTCCAGCTCTATGCCGCCT. V Q L Y A A Y 1251 GCCCTGTGCCAGCGCAGCTA A L W Q R S W 1276 GCCGAGCAGCCGGTGCTGG G E Q P V L E 1301 GGTTTGCAGCTATGCGGG G F E L S R E 1326 CTGTTGCTGGCCTGGTGCTGG G C TGTTGCTGGCCTGGTGCC L L L A S F Q 1351 GCCAATGCCAGCCGCGCGCGCGCTGCTGC L L A S F Q 1351	D L E R G P L I 3700 G R E I V S D G 3775 TRACCARGECCACATCGTGTCCGATG G R E I V S D G 3775 TRACCARGECCGCGGCGAAAAGGACC M E A G E R E R 3925 ACTGCCGTCGATCGATCGGCCGCGTC L P F D R P R P 3927 ACTGCGTCGATCGATCGGCCGCGTC L V E A V R A I 4000 AACTGGTCGAGCCGCGTAGAACCCT L V E A V R A I 4075 AGCGCGTGTTGTATCGGCTACAGCC A L L Y S S C 4150 ACACGGAGCGCGCTGATCGGCTTCG T E R L I G F F	, R V N L L Q L A E 3725 3726 3727 3800 3800 3875 38	3750 ACTO L 3825 CTAC GGGC G GGCC G GGCC G GGTC L 4050 CATO M 4125 GATO L L 42000 L
$\begin{array}{c cccc} A & I & V & V & Q & P & F \\ \hline 1201 & \\ GACCATGTACTGGTGCTGGT \\ D & E & V & L & V & L & V \\ 1226 & \\ GTCCAGCTCTATGCCGCCCCT \\ V & Q & L & Y & A & A & Y \\ 1231 & \\ GCCCTGTGGCAGCAGCGCAGCTT \\ A & L & W & Q & R & S & W \\ 1276 & \\ GGCGAGCAGCCGGTGCTGGC \\ G & E & Q & P & V & L & E \\ 1301 & \\ GGTTTGGAGCTATCGCGGGC \\ G & F & E & L & S & R & E \\ 1326 & \\ GTTTGCTGGCCTGTGGCCTGGTCGC \\ L & L & A & S & F & Q \\ 1351 & L & L & A & S & F & Q \\ 1351 & GCCATCGCAACCGCGTGGG \\ A & N & R & N & R & V & E \\ 1376 & \\ \end{array}$	$ \begin{array}{cccccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline & & & & \\ 3700 \\ \hline & & & & \\ 1700 \\ \hline & & & & \\ 3775 \\ \hline & & & \\ 3850 \\ \hline & & \\ 3950 \\ \hline $, R V N L L Q L A E 3725 3726 3727 3800 3800 3800 3870 3870 3875 3870 3875 38	3750 ACTX L 3825 CTA 3900 GGGC G GGC G GGC C A 1 4050 C CATX M 4125 GATX I 4200 C CCTX L 4275
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{ccccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline & 3700 \\ \hline & CCACCACACCACATCGTGTCCGATG \\ Q & E & I & V & S & D & G \\ \hline & 3775 \\ \hline & 3775 \\ \hline & 3775 \\ \hline & 3850 \\ \hline & 3850$, R V N L L Q L A E 3725 3726 3727 3800 3800 3800 3800 3875 38	3750 ACTX 1 3825 CTAX 3900 GGGX G 3975 G GTT L 4050 CATX M 4125 GGATX 1 42200 CCTX L 1 42200 CCTX L 1 42275 GGAX L
$\begin{array}{c} \mathbf{A} \mathbf{I} \mathbf{V} \mathbf{V} \mathbf{Q} \mathbf{P} \mathbf{F} \\ 1201 \\ \mathbf{GACCATGTACTGGTGCTGGT} \\ \mathbf{D} \mathbf{E} \mathbf{V} \mathbf{L} \mathbf{V} \mathbf{L} \mathbf{V} \\ 1226 \\ \mathbf{GTCCAGCTCTATGCCGCCT} \\ \mathbf{V} \mathbf{Q} \mathbf{L} \mathbf{Y} \mathbf{A} \mathbf{A} \mathbf{Y} \\ 1251 \\ \mathbf{GCCCTGTGGCACCGCAGCTGGT} \\ \mathbf{GCCCTGTGGCACCGCGCTGGT} \\ \mathbf{GCCCTGTGCCACCGCGCTGGT} \\ \mathbf{GCCCTGTGCCGCCTGGTGGGGT} \\ \mathbf{G} \mathbf{F} \mathbf{L} \mathbf{L} \mathbf{A} \mathbf{S} \mathbf{F} \mathbf{G} \\ 1301 \\ \mathbf{GCTTTGCAGCTATGCCGGGTGGG} \\ \mathbf{G} \mathbf{F} \mathbf{L} \mathbf{L} \mathbf{S} \mathbf{F} \mathbf{G} \\ 1326 \\ \mathbf{GCCATGCCACCGCTGGTGGGTGGG} \\ \mathbf{A} \mathbf{N} \mathbf{N} \mathbf{N} \mathbf{V} \mathbf{E} \\ 1376 \\ \mathbf{GCCGATGGCTGGGATGGGCTTGGG} \\ \mathbf{D} \mathbf{G} \mathbf{R} \mathbf{M} \mathbf{G} \mathbf{F} \mathbf{D} \\ 1401 \\ \end{array}$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$, R V N L L Q L A E 3725 3726 3727 3800 TROCCARCCAGATGATGATGATGAGAGA FROCAGCCTGCCGATCCAGATAGGCGGA P A L P I Q Y A D 3875 CCGCGGCAGAGCGTATGAGCGGCCAGCC Q L A Y M T G L L 3850 CCGCCGGCAGAGCCATCGTGCGCCCAGC A R Q S B R G A Q 4025 CGCCCGGCAGCGTGAAGCGGCCAAGGCCCAGCGC Q A D I R V G V P 4100 CGCCAGGCGCAGGCCCAGGCCCAAGGCCCACCA Q X L L A V K A D CAGCGGGCGCTGAAGGCCCAAGGCCCACCAC	3750 ACTX 1 38255 CTAC Y 3900 GGGC G GGC GGC Y 4050 CCTX H 4125 GATC L 4200 CCTX L 4275 GGATC L 4350 CCTX CCTX CCTX CCTX CCTX CCTX CCTX CCT
$\begin{array}{cccc} A & I & V & V & Q & P & F \\ 1201 & \\ GACCATGTACTGGTGCTGGTD & E & V & L & V & L & V \\ 1226 & \\ & & & & & & & & & & & & & & & & & $	$ \begin{array}{ccccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline & 3700 \\ \hline & CCACCACACCACATCGTGTCCGATG \\ Q & E & I & V & S & D & G \\ \hline & 3775 \\ \hline & 3775$, R V N L L Q L A E 3725 3726 3727 3800 TROCCARCCACAGAGAGATACTOCAGAGA FRACCARCTCACTACACAGAGACACAGA P A L P I Q Y A D 3875 CCCACTCACACACTACTOCACCACTACT Q L A Y N T C L L 3950 CCACCTCACACACCACTACTCCCCCCCAC P A R Q S B R G A Q 4025 CACCCACACTCAACACCACTACTCCCCCCACTACT 4 S R V V L R C A D 4 225 CACCCACCACACCCACACCACACCACCACCACCACCACC	33750 ACTX 1 38255 CTX 3900 G G G G G G G G G G C C C C C C C C C
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$, R V N L L Q L A E 3725 3725 3725 3725 3725 3800 FROCCAGATOCAGATOCAGATOCTOCAGAGA FROM CONCENTRONGGATOCTOCAGAGA CAGE AND A CONCENTRONGGATOCTOCCOCCAGA 2875 COCCAGATOCCAGATOCTOCCOCCAGA CAGE A Y N T G L L 3950 COGCCOGCAGACCAGACCACTOCTOCCOCCAGA 4 A Y N T G L L 3950 COGCCOGCAGACTOCAGACCACTATOTTT A Q R E G A S S F FROM COCAGCACACCACTATOCTCCAGATOCTCCAGATOCTCAGACCACACCAC	3750 ACTCL 3825 CTAC Y 3900 GGGCG G 3975 GGTCL 4050 CATC M 4125 GGTCL 42275 GGTCL 42350 CATCL 44425 GGCCL
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{ccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline & 3700 \\ \hline & 775 \\ \hline & 755 \\ \hline & $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	3750 ACTX I 3825 CTAC GGCG G 3975 GCTAC GGCG G 4050 CCAN M 4125 GGATX I 4225 GGATX I 4350 CCAN L 4350 CCAN L 4350 CCAN L 4350 CCAN CCAN C CCAN C CCAN C CCAN C C C C
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{ccccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline 3800 \\ \hline 38$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	3750 ACTA 3825 CTAA 3825 CTAA 3825 CTAA 33905 GGGGG GGGGG GGGGG CATA 4050 CATA 4125 GGAT L 4275 GGAT L 4275 GGGCC A 43505 CTAA CATA A 4275 CTAA A 43505 CTAA A 4275 CTAA A 43505 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 43505 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA A 4275 CTAA CTAA A 4275 CTAA CTAA A 4275 CTAA CTAA A 4275 CTAA
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{ccccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline & 3700 \\ \hline & CCACCACACCACATCGTGTCCGATG \\ Q & E & I & V & S & D & G \\ \hline & 3775 \\ \hline & 7750 \\ \hline & 7750 \\ \hline & 7760 \\ \hline & 7760$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	33750 33750 23750 23750 23825 CTAX 33900 GGC GGC CAX M 41257 GGAT L 4200 CCAX L 4275 GGC A 4350 GGC CAX L 4425 CCAX L 4425 CCAX L 4425 CCAX CCAX CCAX CCAX CCAX CCAX CCAX CCA
$\begin{array}{cccc} A & I & V & V & Q & P & P \\ \hline A & I & V & V & Q & P & P \\ \hline 1201 & GACCATGTACTGGTGCTGGT \\ D & H & V & L & V & L & V \\ \hline 1226 & GTCCAGCTCTATGCCGCCT. \\ V & Q & L & Y & A & A & Y \\ \hline 1231 & GCCGTGTGCCAGCCAGCTGGCTGGCCAGCCAGCTGGCCAGCAGCAGCGGGGGGGG$	$ \begin{array}{ccccccc} D & L & E & R & G & P & L & L \\ 3700 \\ \hline & 3700 \\ \hline & 3700 \\ \hline & CARCERCACATCCCACCTCGATCG \\ Q & E & I & V & S & D & G \\ \hline & 3775 \\ \hline & 3850 \\ \hline & GATOGAGCCCCCCCCCCACCTCGATCT \\ S & Q & C & L & D & V & V & I \\ \hline & 3850 \\ \hline & GATOGAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c cccc} , \ \mbox{R} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	3750 3750 3750 23750 23750 39750 3900 GGCC 3900 GGCC 4050 GGCC 4050 GGCC 4050 GGCC 4050 GGCC 4125 GGCC 4125 GGCC 4275 GGCC 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 44255 GCCT 445555 GCCT 4455555 GCCT 4455555 GCCT 445555555 GCCT 445555555555555555555555555555555555
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	3750 3750 3750 3750 3750 3750 3750 3750
$\begin{array}{cccc} A & I & V & V & Q & P & P \\ \hline & & & & & \\ \hline & & & & & \\ \hline & & & &$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	3750 3750 ACTX L 38255 CTAX SCTAC J 3900 GGGGG G 3975 CTAX J 3900 GGGGG G 3975 CTAX L 3900 GGGGG G 3975 L 4050 G GGGCT L 4275 GGCTX L 4257 CTAX L 2000 C CTAX L 200 CTAX L 2000 C CTAX L 2000 C CTAX L 2000 CTAX L 2000 CTAX L 20 C CTAX L 2000 C C C CTAX L C C C C C C C C C C C C C C C C C C
$\begin{array}{c} \mathbf{A} \mathbf{I} \mathbf{V} \mathbf{V} \mathbf{Q} \mathbf{P} \mathbf{F} \\ \mathbf{I201} \\ \mathbf{GACCATGTACTGGTGCTGGT} \\ \mathbf{D} \mathbf{E} \mathbf{V} \mathbf{L} \mathbf{V} \mathbf{L} \mathbf{V} \\ \mathbf{I226} \\ \mathbf{GTCCAGCTCTATGCCGCCT} \\ \mathbf{GTCCAGCTCTATGCCGCCT} \\ \mathbf{G} \mathbf{I217} \\ GTCCAGCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	3750 3750 3750 3750 3750 3750 3750 3750
$\begin{array}{cccc} \mathbf{A} & \mathbf{I} & \mathbf{V} & \mathbf{V} & \mathbf{Q} & \mathbf{P} & \mathbf{F} \\ 1201 & & \\ \mathbf{GACCATGTACTGGTGCTGGT} \\ \mathbf{D} & \mathbf{E} & \mathbf{V} & \mathbf{L} & \mathbf{V} & \mathbf{L} & \mathbf{V} \\ 1226 & & \\ \mathbf{GTCCAGCTCTATGGCGGCGCGT} \\ \mathbf{V} & \mathbf{Q} & \mathbf{L} & \mathbf{Y} & \mathbf{A} & \mathbf{A} & \mathbf{Y} \\ 1251 & & \\ GCCCTGTGGCAGCGGGGCGGGGGGGGGGGGGGGGGGGGG$	D L E R G P L I 3700 G R E I V S D G 3715 ATAGOCAMOGOCTOCACONGTOT S Q G L D V V I 3850 GOATOGAGOCGOGGAAAAGGAO M E A G E K E R 3925 ACTIOCCATCGATOGACOGOCGAAAGGAO M E A G E K E R 4000 AACTGATCGAGOCGTGAGOGOCACO L P F D R P R E 4000 AACTGATCGAGOCGTGAGOGOCACACO C L V E A V R A I 4015 ACCCACOGOCGAGOCCCAGOCCCCC E L L Y R Y S C 4150 ACCCACTGCTGCAGOCCCAGOCCCCCC E L L A Q A R C 4150 ACCCACTGCTGCAGOCCCAGOCCCCCC E L L A Q A R C 4150 ACCCACTGCTGCCAGOCCCGAGOCCCCC E L L A Q A R C 4150 ACCCACTGCTGCCGAGOCCCGAGOCCCCCC E L L A Q A R C 4150 ACCCCCCCCCCCCGAGOCCCGAGOCCCCCC C L L C R L C C 4150 ACCCCCCCCCCCCGAGOCCCGAGOCCCCCC C L L A Q A R C 4150 ACCCCCCCCCCCCGAGOCCCGAGOCCCCCC C L L C R C C 4150 ACCCCCCCCCCCCCGAGOCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	9750 3750 3750 38255 CTA 38255 CTA 3900 GGGC GGCC M 4050 GGCC CAT M 41255 GAT CAT M 41255 GAT CAT CAT CAT CAT CAT CAT CAT CAT CAT C
$\begin{array}{c} \mathbf{A} \mathbf{I} \mathbf{V} \mathbf{V} \mathbf{Q} \mathbf{P} \mathbf{F} \\ \mathbf{I201} \\ \mathbf{GACCATGTACTGGTGCTGGT} \\ \mathbf{D} \mathbf{E} \mathbf{V} \mathbf{L} \mathbf{V} \mathbf{U} \mathbf{V} \\ \mathbf{I226} \\ \mathbf{GTCCAGCTCTATGGCGGCCAGCT} \\ \mathbf{V} \mathbf{Q} \mathbf{L} \mathbf{Y} \mathbf{A} \mathbf{A} \mathbf{Y} \\ \mathbf{I231} \\ \mathbf{GCCCTGTGGCAGCGGCGGCTCTGG} \\ \mathbf{G} \mathbf{C} \mathbf{Q} \mathbf{V} \mathbf{L} \mathbf{E} \\ \mathbf{I301} \\ \mathbf{GGCTGTGGAGCAGCGGTGCTGGG} \\ \mathbf{G} \mathbf{E} \mathbf{Q} \mathbf{P} \mathbf{V} \mathbf{L} \mathbf{E} \\ \mathbf{I301} \\ GGCTTTGGAGCTATGGGGGGGGGGGGGGGGGGGGGGGGG$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccc} , \ R \ V \ N \ L \ L \ Q \ L \ A \ E \\ 3725 \\ 3775 \\ 3870 \\ 3870 \\ 3870 \\ 3870 \\ 3875 $	3750 3750 ACTOL 1 38235 3750 38257 33900 33975 33900 33975 339075 339075 339075 339075 33975 339075 40500 COATOL 1 42050 COATOL 1 40500 COATOL 1 4000 COATOL 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
$\begin{array}{cccc} A & I & V & V & Q & P & F \\ 1201 & GACCATGTACTGGTGCTGGT \\ D & E & V & L & V & L & V \\ 1226 & GTCCAGCTCTATGCCGCCTGT \\ V & Q & L & Y & A & A & Y \\ 1231 & GCCCTGTGGCAGCCAGCTGT \\ A & L & W & Q & R & S & M \\ 1276 & GGCGAGCAGCCGGTCTGG \\ G & E & Q & P & V & L & E \\ 1301 & GGTTTGGAGCTATCGCGGGTGGGT \\ G & F & E & L & S & R & E \\ 1302 & GGTTTGGAGCTATCGCGGGTGGGT \\ G & F & E & L & S & F & Q \\ 1351 & GGCAATCGCAACCGCTGTGGG \\ G & R & M & G & F & D \\ 1401 & G & R & M & G & F & D \\ 1401 & G & R & M & G & F & D \\ 1401 & GGCAATCGCAACCGCGTGGG \\ L & P & F & Q & L & V \\ 1426 & F & D & L & F \\ 1376 & GACGGTCGGACCAGCAGCGGGGGGGGGGGGGGGGGGGGG$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccc} , \ \mbox{R} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	3750 3750 ACTOL 1 38255 3900 CTAL 33905 CTAL 33905 CTAL 33905 CTAL 33905 CAN A 4050 CCAN A 4275 GGCK CCAN A 4275 GGCK CCAN A 4275 GGCK CCAN A 4255 CCAN A 4555 CCAN A 4555 CCAN A 4555 CCAN A 455 CCAN A 4555 CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN A CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN C CCAN CCAN CCAN C CCAN CCAN C CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN CCAN C C C C
$\begin{array}{c} \mathbf{A} & \mathbf{I} & \mathbf{V} & \mathbf{V} & \mathbf{Q} & \mathbf{P} & \mathbf{F} \\ \mathbf{I201} \\ \mathbf{GACCATGTACTGGTGCTGGT} \\ \mathbf{D} & \mathbf{E} & \mathbf{V} & \mathbf{L} & \mathbf{V} & \mathbf{L} & \mathbf{V} \\ \mathbf{I226} \\ \mathbf{GTCCAGCTCTATGCCGCCCT} \\ \mathbf{V} & \mathbf{Q} & \mathbf{L} & \mathbf{Y} & \mathbf{A} & \mathbf{A} & \mathbf{Y} \\ \mathbf{I231} \\ \mathbf{GCCCTGTGCCACCGCAGCTATGC} \\ \mathbf{G} & \mathbf{C} & \mathbf{Q} & \mathbf{P} & \mathbf{V} & \mathbf{L} & \mathbf{E} \\ \mathbf{I301} \\ \mathbf{GGCTTGCACCTATCCGAGCTGTGCT} \\ \mathbf{G} & \mathbf{G} & \mathbf{C} & \mathbf{Q} & \mathbf{P} & \mathbf{V} & \mathbf{L} & \mathbf{E} \\ \mathbf{I301} \\ \mathbf{GGCTTGCACCTATCCGAGCTGTGCT} \\ \mathbf{G} & \mathbf{G} & \mathbf{F} & \mathbf{E} & \mathbf{L} & \mathbf{S} & \mathbf{R} & \mathbf{E} \\ \mathbf{I301} \\ \mathbf{GGCTTGCACCTATCCGAGCTGTGCT} \\ \mathbf{G} & \mathbf{F} & \mathbf{E} & \mathbf{L} & \mathbf{S} & \mathbf{F} & \mathbf{Q} \\ \mathbf{I326} \\ \mathbf{G} & \mathbf{F} & \mathbf{E} & \mathbf{L} & \mathbf{S} & \mathbf{F} & \mathbf{Q} \\ \mathbf{I326} \\ \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{K} & \mathbf{N} & \mathbf{R} & \mathbf{V} & \mathbf{E} \\ \mathbf{I376} & \mathbf{R} & \mathbf{N} & \mathbf{R} & \mathbf{V} & \mathbf{E} \\ \mathbf{I376} & \mathbf{R} & \mathbf{N} & \mathbf{R} & \mathbf{V} & \mathbf{E} \\ \mathbf{I376} & \mathbf{R} & \mathbf{N} & \mathbf{R} & \mathbf{V} & \mathbf{E} \\ \mathbf{I376} \\ \mathbf{G} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{C} \\ \mathbf{G} & \mathbf{C} & \mathbf{G} & \mathbf{C} & \mathbf{G} & \mathbf{C} & \mathbf{C} \\ \mathbf{G} & \mathbf{R} & \mathbf{N} & \mathbf{R} & \mathbf{R} & \mathbf{V} & \mathbf{E} \\ \mathbf{I376} \\ \mathbf{G} & \mathbf{R} & \mathbf{N} & \mathbf{R} & \mathbf{R} & \mathbf{R} & \mathbf{L} \\ \mathbf{I} & \mathbf{I} & \mathbf{R} & \mathbf{R} & \mathbf{R} & \mathbf{R} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{R} & \mathbf{R} & \mathbf{R} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{G} & \mathbf{G} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} &$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$, R V N L L Q L A E 3725 3726 3727 3800 3800 3800 3875 38	3750 3750 3825 CTA Y 3900 G G G G G G G G G G G G G G G G G G

Q 1.65 L T G \$ 2 7 9 À 1801 185 ñ 1876 \$750 5725 G R 5825 G 192 1 95 L 6050 200 L TG 2026 6200 ٥ G 207 L 2101 6400 6425 GGC L 2126 215 L 2176 6625 6650 G GA D E L D 2201 6775 6800 227 692 2000 7029 ATGAGCGACTGGGCGGAAG E 232 L 2351 7150 7175 7200 GCGG λ 7250 Ē 7325 7300 7350 SCGTGCCGGGCGCTGATCT ٥ ٨ L



FIG. 2. Nucleotide sequences of the *pvdD* gene and part of the *fpvA* gene. The predicted amino acid sequences of the proteins are also shown. Nucleotide numbering is relative to the likely translational start site of *pvdD*. Transcription start sites for *pvdD* ($|^{->}$), a Shine-Dalgarno sequence (SD), and inverted repeats downstream from *pvdD* (IR1 and IR2) are shown. *Bam*HI, *Ps*I, and *SaI* restriction sites are indicated. Part of the sequence (nucleotides -75 to 390) has been published elsewhere (36) and is included here for completeness.

acids and to have a molecular mass of 89 kDa (2). However, when the translated *pbsC* sequence was aligned with that of PvdD by using the GAP program (12), similarity extended over the first 988 amino acids of PvdD, with the start of PbsC predicted by Adams et al. (2) corresponding to amino acid 185 of PvdD. It seems likely that the GTG codon in the *pbsC* sequence which corresponds to the proposed GTG start site in *pvdD* is the start of the *pbsC* gene.

Mutations of *pbsC* in *Pseudomonas* sp. strain M114 prevent the synthesis of pseudobactin, and it has been suggested that PbsC may be a member of the peptide synthetase family which is involved in the biosynthesis of M114 pseudobactin (2). PbsC contains most of the sequence motifs present in peptide synthetases, but motifs VI and X (Fig. 3) are missing. The exact relationship of PbsC with members of the peptide synthetase family remains to be determined, but its striking similarity to PvdD indicates that the two proteins are likely to have some functions in common in siderophore synthesis. DNA hybridization experiments have shown that genes similar to *pvdD* are also present in siderophore synthesis DNA from other fluorescent pseudomonads (34, 35), indicating that proteins similar to PvdD are also present in these species.

pvdD is adjacent to a gene encoding a ferripyoverdine uptake protein. DNA sequencing also revealed a partial ORF of 792 bp downstream from *pvdD* and orientated in the opposite direction (Fig. 2). The 264-residue predicted protein product of the partial ORF downstream from *pvdD* was 99.2% identical to amino acids 552 to 814 of FpvA, the outer membrane receptor for ferripyoverdine (33). There was a two-amino-acid insertion and one amino acid difference when the translated sequence of this ORF was compared with that of Poole et al. (33). Further sequencing and genetic analysis has shown that this ORF does indeed encode the ferripyoverdine outer membrane receptor (51). The differences between our *fpvA* se-

J. BACTERIOL.

PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	1 11 LSYAQEROWFLWQLE F USPWQEGMLFHALLD VSSVQKRMFILNEFD VSSAQKRMFILNEFD VSSAQKRMFIVNQFV -SSAQ-RML-Q-E	21 3 PESAAYH IPSVLRLR A KDKNAHLVOMS IA IE RSGTAYNLPGVMFLD GAGITYNVPNVMF IE GVGI SYNMPS IMLIE -EG-AYN-PSVME	1 41 GRIDIDALQRSFDSLV V U GIVDVELLSESLNILJ GKLNYRQLEAAVKKLN GKLDYQRFEYAIKSLV GKLERTRLESAFKRJJ GKLD-D-LER-FKSLV	51 VARHETLRTRF-F IDRYDVFRTTFLI VERHEALRTSF-F IERHESLRTSF-F /-RHE-LRTSFL	61 RLDGDEARQE I E GRSY QV HEKIKQPLQVV HSINGEPVQRV SLINGEPVQRV EI INGKPVQKI INGEPVQRV	71 AASMALPLDIV QPAVSVSIERR LKERPVQLQFF HQNVELQIAYS HQNVELQIAYS HEEVDFNMSYQ HQ-VELQIAYS	81 VALGPLEEGA CQF EG KDISSLDEEK SESTED SEAKED 2VASNE SE-G-LAEED	91 LARQVETTIARP IERIQAIVVQ REQAIEQYKYQD QVERIAEFMQP EIEQIVESFVQP QVEKINDEFIQP E-I-AEFVQP
PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus)1 111 FDLERGPLLRV GETVFDLTRDPLMRV FALEVRPLLRV FDLEIAPLLRV GETVFDLER-PLLRV GETVFDLER-PLLRV	121 13 SLIRLAEDDHVLVLV N Q AIFQTGKVNYQMIWS GLVKLEAERHLFIMD GLVKLASDRYLFIMD GLLKLAEDRHVLIFD GLLKLAEDRHVLIFD	HHIVSDG SMOVMVE HHIIMDG CFNIFY HHIISDG SMOIMIC HHIISDG SMOIMIC HHIISDG SMILMF HHIISDG SMOIM- Accer motif	151 CELVQLYAAYSRO Q NDLFNIYLSLKEP 2EIADLYK (ELGELYQ -ELLY-AYS-(161 GLEVALPALPI D V KKPLQLEAVQ- EKELPTLGI GKELAELHI GNALPELRI GLLPALPI	171 QYADYALWORS PYKQFIKW QYKDFTVWHNI QYKDFAVWQNI QYKDFAVWQNI QYKDFAVWQN-	181 SWMEAGEKER -LEKQDKQE RLLQSDVIEK SWFQSDALEK EWFQSEAFKK -WSKEK	191 QLAYWTGLLGGE ALRYWKEHLMNY QEASLAERICRR QKTYWLNTFAED QEEYWVNVFADE QLAYWT-LL-GE
20 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	D1 211 QPVIELPFDRPRPVR A DQSVTLPKKKAAINN DSSIESTDRLPKY IPVILSTDYPRPTI RPILDIPTDYPRPMQ -PVIELP-D-PRP-R	221 QSHRGAQFI-LELDI LG F SR TTYEPAQFR-FAFDK QPFKALMVKDLHSVQ QSFEGDIVT-FSACK QSFDGAQLT-FGTGK QSF-GAQ-TDFEL-K	DLSQALRRVAQ E VE V AL F VLTQQLLRIAN ESSLWMIYTRWQQKQ ESSLWL-ALREAR QLAEELKRLAAR QLMDGLYRVAT -SSLWLALRR-AQE	251 2EGATAFALLLAS 8 SS M 2SQVTLNIVFQT 2QHYIWFYLLRIN 2TGTTLYMLLLA 2-G-TLFMLLLA-	261 SFQALLYRYSG MFFLSKYSG AYNVLLHKYSG AYNVLLSKYSG -FLL-KYSG	271 QADIRVGVPI TNHVVYGSVV QDDIVVGTPI QEEIVVGTPI QEDIVVGTPI Q-DIVVGTPI	281 ANRNRVET SGRP SE I SG I AGR SHADV AGR SHADV VGR SHTDL AGR SHE I-D-	291 ERLIGFFVNTQV EKMVGLFINTLP ENMLGMFVNTLA ENIVGMFVNTLA ENIVGMFVNTLA EN-VGMFVNTLA
30 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus)1 311 LKADLDGRMGFDELL LR IQTQKDQSF IELV IRSRLNNEDTFKDFL LKNTP IAVRTFHEFL MRNKPEGEKTFKAFV LDLDG-MTF-E-L	321 33 AQARQRA LEAQAHQD KTVHQNVLFSQQHEX ANVKQTALHAYENPD LEVKQNA LEAFENQD SE IKQNA LEAFENQD A-VKQNA LEAQE-QD	1 341 LPFEQLVEALQPERSI NF FPLYEIQNHTELKQNI YPFDLVEKLGIQRDI YPFENLIEKLQVRRDI YPFELIEKLEIQRDI YPFE-LVEKLQRDI	351 LSHNPLFQVLFN A LIDH IMV I ENYPI LSRNPLFDTMFVI LSRNPLFDTMFSI LSRNPLFDTLFSI LSRNPLFDTLF-I	361 YQSEARGNGQA H I SVTPE LVEELQKNS LQNTDRKS LSNIDEQVE LQNIGEES LQNEDRK	371 FRFDELQMESY VQLED RL GI - IMQKVGFTVI FEVEQITITP IGIEGLNFSP FELAELTCKPI FELAELTCKPI	381 VQFDSRTAQ- LAW GQ RDVK-MFEPT YVPNSRHS YEMQYWIA FDLVSKLEHA YD-DSRTA	391 -FDLTLDLTDEE IQED NYDMTVMVLPRD KFDLTLEVSEEQ KFDLSTDILEKQ KFDLSLVAVVFE KFDLTLD-LEEE
4(PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	1 411 QRFCAVFDYATDLFD NGIW S E-ISVRLDYNAAVYD NEILLCLEYCTKLFT DDIQFYFNYCTNLFK EEIAFGLQYCTKLYK -EIDYCT-LFD	421 43 ASTVERLAGHWRNLI DIFIKKIEGHMKEVA DKTVERMAGHFLQII KETIERLATHFMHI EKTVEQLAQHFIQIV A-TVERLAGHFR-II	1 441 RGIVANPRQRLGELPI LCVANNPHVLVQDVPI HAIVGNPTIISEIE QEIVINPEIKLCEIN KAIVENPDVKLSDID IVANPRL-EIPI	451 LLDAPERRQTLSH LLTKQEKQHLLVI ILSEEEKQHILFI MLSEEEQQRVLJ MLSEEEKKQIMLJ LLSEEEKQQ-LSI	461 EWNPAQRECAV ELHDS ITEYP- EFNDTKTTYPL DFNGTDATYAT EFNDTKIQYTC EFNDTEYAV	471 QGTLQQRFEE(CKQFKDYLRN NKIFHELFEE(NQTIQELFEE(~KT-QQLFEE(481 QARQRPQAVA QVEKTPEHVA RWRRRADHVA QVEKTPDHIA QVKKTPEHIA QVRK-P-HVA	491 LILDEQRLSYGE VVFEDEKVTYRE VGWKDQTLTYRE VIDEREKLSYQE IVWEGQALIYHE VI-E-Q-LSY-E
50 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	DI 511 INARANRLAHCLIAF IHERSNQLARFLREK INERANQVARVLRQK INAKANQLARVLRQK INIKANQLARVLREK INARANQLARVLREK	521 53 GVGADVPVGLALERS GVQPDNIVGLLVERS GVQPDNIVGLLVERS GVQPNHPVAIMTERS GVPPD-PVG-M-ERS	1 541 LDN LVGLLAILKAGG VEN IVGILGILKAGG PELLVGIMGILKAGG LDN IVGMLGVLKAGG LEN IVGIFSILKAGG L VGILGILKAGG	551 AYLP LDP AAPEEI AFVP IDPEYP KEI AYLP LDPEYP ADI AYVP IDI DI DYPQEI AYVP IDPAYPQEI AY-P-DPAYP-EI	561 RIAH I LDDSGV RIGYMLDSV RISYM IQDCGV RISYMMEDSGA RIQYLLEDSGA RIQYLLEDSGA	571 RLLLTQGHLLI RLVLTQRHLKI RIMLTQQHLL: ALLLTQQKLTQ ALLLTQSHVLI RLLLTQ-HLLI	581 ERLPRQAGVE DKFAFTKETI SLVHDEFDCV QQIAFSGDIL NKLPVDIEWL E-LP-QA-V-	591 VLAIDGLVLDGY VIEDPSISHELT ILDEDSLYKGDS YLDQEEWLHEEA DLTDEQNYVEDG VLD-LE-Y
60 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	01 611 AESDPLPTLSADNLA EEIDYINESEDLF SNLAPVNQAGDLA SNLEPIARPHYIA TNLEFMNQSTDLA LDP-NTL-ADDLA	621 63 YVITTSGSTGKPKGV YIITTSGSTGKPKGV YIITTSGSTGKPKGV YIITTSGTTGKPKGV YIITTSGTTGKPKGV YIITTSG-TGKPKGV	1 641 VI LLTHRNALRLFSATEJ MLEHKNIVNLL MUEHRNVIRLV MIEHQSYNVAM MIEHQSINCIA	651 AWFGFDERJ KNTNYVQV AW-KDAYRLDTFI QWRKEEYEFO AW-KGFDE-1	661 DVWTL-FHS REDDRIIQT PVRLLQMAS GPGDTALQV DVNFWDLQS	671 YAFDFSVWI TNAVLTCVTKI GAIGFDALTFI FAFAFDVSAGI FSFAFDGFVA -AF-FDVWT-I	681 EIFGALLYGG KFFSTLLSGG EVFGSLLHGA DFARALLTGG SIFAPILAGA E-FGALLYGG	691 CLVIVPQWVSRS R QLYLIRKETQRD ELYPVTKDVLLD QLIVCPNEVKMD TSVLPKEEEAKD QLV-VP-EVSRD
76 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	01 711 PEDFYRLLCREGVTV VEQLFDLVKRENIEV AEKLHKFLQANQITI PASLYAIIKKYDITI PVALKKLIASEEITF PEDLY-LL-REGITV	1 721 73 VLNQTPSAFKQLMAVA VLSFPVAFLKFIFN MWLTSPLFNQLSQGT FEATPALVIPLMEYI VYGVPSLFSAILDVS VLNQTPSLFKQLMAVA	1 741 CSADMATQQPA-LRY1 EEREFINRFPTCVKH: EEMFAG-LRSI YEQKLDISQ-LQII SSKDLQN-LRC1 C-AD-ATQ-PACLRY	751 VIFGGEALDLQSJ IITAGEQLVVN LIVGGDALSPKH LIVGSDSCSMEDJ VTLGGEKLPAQIV VI-GGEAL-LQSJ	761 LRPWFQRFGDR -NEFKRYLHEH INNVKRK-C-P FKTLVSRFG-S VKKIKEK-N-K L-PWK-RFGDF	771 OPQLVNMKGI NVHLHNHKGP NLTMWNGKGP TIRIVNSKGV EIEVNNEKGP QLVNMKGP	781 TETTVHVTYR SETHVVTTYT TENTTFSTCF TEAALILAIM TENSVVTTIM TETVTY-	791 PVSEADLEGGLV K INPEAE IPEL LIDKEYDDN NQPLSSLHVTGT RDIQVEQE PV-EADLE-GLV
86 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	01 811 SPIGGTIPDLSWYII IPIGKPISNTWIYII IPIGKAISNSTVYIN VPIGKPYANMKMYIN ITIGRPLSNVDVYIV -PIGKPISNLS-YII	821 DRDLNPVPRGAVGET DQEQQLQPQGIVGET DRYGQLQPVGVGET INQYLQIQPVGVGET NCNHQLQPVGVGET DR-LQLQPVGVGET	11 841 YIGRAGLARGY RRP YISGANVGRGY INNO YISGANVGRGY INNO CVGGDGVARGY INRP CUGGOGLARGY INRP -IGGOGLARGY INRP	851 GLSATRFVPNPF ELTAEKFFADPF ALTEEKFVPNPF DLTAEKFVPNPF ELTADKFVVNPF -LTAEKFVPNPF	861 PGGAGER LYR APNER MYR APGER MYR VPGEK LYR VPGER MYR -GGPGER -YR	A71 TGDLA FQADG TGDLA WLPDG TGDLA WLPDG TGDLA WRSDG TGDLA W-PDG	NIEYLERID NIEFLERAD TIEYLERIX NVEFLERIM NIEYLERID	IQVKVRGFRIELG IXVKIRGHRIELG XVKIRGYRIEPG IQVKIRGIRIELG XVKVRGYRIELG IQVK-RGFRIELG
PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	911 Elenalaglagvrda Elenqlinckgvkea Elenqlinckgvkea Elenqlrkhdsikea Elenaleyekikea Elenalgvkea	921 VVUAHDGVGGTQ-V VVIDKADDKGGKY VIMVVEDNNGQKA VTVIAREDHMKEKY VVVSEHTASEQM VVV-AHEDVGG-Y	941 GY VVADSAEDAERLRI AY VVMEVEVND: AY VVPEEVVT: AY MVTEGEVNV: AY IVGEEDVLTI AY VVAESAED-EVL	951 ESIRESIKRHLPI SEIREYLGKALPI SEIREYIAKELPI AEIRAYLANDRA LDIRSYLAKHLPI -EIREYLAKHLPI	961 DYMVPAHIMLI DYMIPSFFVPI VYMVPAYFVQI A-MIPSYFVSI SYMIPNYFIQI DYM-PAYFV-I	EFMPLTVNGKI DIVRLHLNGK ECMPLTONGKU EFMPLTANGK DSIPLTPNGK EFMPLTVNGK	981 LDROALPOPD IDRKSLPNLE WRSALPKPD IDKRSLPEPD VDRKALPEPQ -DR-ALP-PD	991 ASLS-QQAYRAP GEFGTATEYVAP GSISIGTEYDRP TIGIMAREYVRP GSLSTEYVAP
100 PvdDrep1 PvdDrep2 GrsBdom1 GrsBdom2 GrsBdom3 GrsBdom4 Consensus	01 1011 GSELEQRIAAIWSEI A TNELEEKLAKIWEEV SSDIEMKLAEIWHNV RTMLEGKLEEIWKDV RNEIEAQLVLIWGEV -SELEQKLA-IW-EV	1021 LGVERVGLDONFFEJ LGISQIGIODNFFEJ LGVRKIGVIDNFFEJ LGLORVGIHDDFFT LGIELIGITDNFFEJ LGVER-GIDDNFFEJ	CGHSLLATRV LLML GGHSLKATTL CGHSLKAMTM IGGHSLKAMAV GGHSLKATLL GGHSLKAL X			•.	-	

FIG. 3. Alignment of the predicted peptide synthetase domains of PvdD with those of GrsB. The two repeats of PvdD (PvdDrep1 and PvdDrep2) were aligned with the four domains of gramicidin S synthetase B (GrsBdom1 to GrsBdom4) (53), using Homed (49) and CLUSTAL V (23). The two repeats of PvdD were identical except at sites where a residue is shown for PvdDrep2. Boxes denote consensus sequence motifs which have been identified for peptide synthetase-type enzymes; the consensus sequences are listed in reference 19. Motifs I to VI are conserved among the superfamily of adenylate-forming enzymes, including nonpeptide synthetases. Motifs VII to X are conserved among a subfamily of the adenylate-forming superfamily which includes all known peptide synthetases and also EntF and AngR. Motifs I and III have been associated with ATP-binding and hydrolysis activities, respectively (15, 17, 38). Motif X contains a conserved serine (position 1034 in the alignment) to which a 4'-phosphopantetheine cofactor is very probably attached (13, 39). A so-called spacer motif present at the N-terminal end of peptide synthetase domains involved in elongation, but not initiation, of peptide synthetase (15) is also indicated. It has been suggested that amino acids between 627 and 829 in the alignment are involved in substrate recognition by peptide synthetase domains (10).

quence and that of Poole et al. (33) probably reflect minor differences between the *P. aeruginosa* strains used in the respective studies. The presence of *fpvA* downstream from *pvdD* indicates that ferripyoverdine synthesis and uptake genes are clustered, as is the case in other fluorescent pseudomonads (6, 28, 31). In addition, this allows the *fpvA* gene to be located at 47 min on the genetic map of *P. aeruginosa*, as we have previously shown that clone pSOT1, which includes this DNA, is derived from that part of the chromosome (35).

Nucleotide sequence accession number. The sequence shown in Fig. 2 has been assigned accession number U07359 in the GenBank and EMBL libraries.

We are very grateful to Irene Rombel for providing some of the plasmid constructs used in this study and to Anthony Smith for critical reading of the manuscript.

T.R.M. was supported by a postgraduate studentship from the Health Research Council of New Zealand. M.E.M. was supported by the Health Research Council of New Zealand.

REFERENCES

- Abdallah, M. A. 1991. Pyoverdins and pseudobactins, p. 139–154. *In G.* Winkelmann (ed.), CRC handbook of microbial iron chelates. CRC Press, Fla.
- Adams, C., D. N. Dowling, D. J. O'Sullivan, and F. O'Gara. 1994. Isolation of a gene (*pbsC*) required for siderophore biosynthesis in fluorescent *Pseudomonas* sp. strain M114. Mol. Gen. Genet. 243:515–524.
- Ankenbauer, R., S. Sriyosachati, and C. D. Cox. 1985. Effects of siderophores on the growth of *Pseudomonas aeruginosa* in human serum and transferrin. Infect. Immun. 49:132–140.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience, New York.
- Bankier, A. T., K. M. Weston, and B. G. Barrell. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. Methods Enzymol. 155:51–93.
- Bitter, W., J. D. Marugg, L. A. de Weger, J. Tommassen, and P. J. Weisbeek. 1991. The ferric-pseudobactin receptor PupA of *Pseudomonas putida* WCS358: homology to TonB-dependent *Escherichia coli* receptors and specificity of the protein. Mol. Microbiol. 5:647–655.
- Budzikiewicz, H. 1993. Secondary metabolites from fluorescent pseudomonads. FEMS Microbiol. Rev. 104:209–228.
- Coque, J. J. R., J. F. Martín, J. G. Calzada, and P. Liras. 1991. The cephamycin biosynthetic genes *pcbAB*, encoding a large multidomain peptide synthetase, and *pcbC* of *Nocardia lactamdurans* are clustered together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. Mol. Microbiol. 5:1125–1133.
- Cornelis, P., D. Hohnadel, and J.-M. Meyer. 1989. Evidence for different pyoverdine-mediated iron uptake systems among *Pseudomonas aeruginosa* strains. Infect. Immun. 57:3491–3497.
- Cosmina, P., F. Rodrigeuz, F. de Ferra, G. Grandi, M. Perego, G. Venema, and D. van Sinderen. 1993. Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol. Microbiol. 8:821– 831.
- Cox, C. D., and P. Adams. 1985. Siderophore activity of pyoverdin for Pseudomonas aeruginosa. Infect. Immun. 48:130–138.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395.
- D'Souza, C., M. N. Makano, N. Corbell, and P. Zuber. 1993. Amino-acylation site mutations in amino acid-activating domains of surfactin synthetase: effects on surfactin production and competence development in *Bacillus* subtilis. J. Bacteriol. 175:3502–3510.
- Farrell, D. H., P. Mikesell, L. A. Actis, and J. H. Crosa. 1990. A regulatory gene, *angR*, of the iron uptake system of *Vibrio anguillarum*: similarity with phage P22 *cro* and regulation by iron. Gene 86:45–51.

- Fuma, S., Y. Fujishima, N. Corbell, C. D'Souza, M. M. Nakano, P. Zuber, and K. Yamane. 1993. Nucleotide sequence of 5' portion of *srfA* that contains the region required for competence establishment in *Bacillus subtilis*. Nucleic Acids Res. 21:93–97.
- Gibson, T. J. 1984. Ph.D. thesis. University of Cambridge, Cambridge, England.
- Gocht, M., and M. A. Marahiel. 1994. Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthesis TycA by site-directed mutagenesis. J. Bacteriol. 176:2654–2662.
- Gold, L., and G. Stormo. 1987. Translational initiation, p. 1302–1307. In J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
- Guilvout, I., O. Mercereau-Puijalon, S. Bonnefoy, A. P. Pugsley, and E. Carniel. 1993. High-molecular-weight protein 2 of *Yersinia enterocolitica* is homologous to AngR of *Vibrio anguillarum* and belongs to a family of proteins involved in nonribosomal peptide synthesis. J. Bacteriol. 175:5488– 5504.
- 20. Gutiérrez, S., B. Díez, E. Montenegro, and J. F. Martin. 1991. Characterization of the *Cephalosporium acremonium pcbAB* gene encoding α-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the *pcbC* gene as a cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains. J. Bacteriol. 173:2354– 2365.
- Haas, B., J. Kraut, J. Marks, S. C. Zanker, and D. Castignetti. 1991. Siderophore presence in sputa of cystic fibrosis patients. Infect. Immun. 59: 3997–4000.
- Haas, B., E. Murphy, and D. Castignetti. 1991. Siderophore synthesis by mucoid *Pseudomonas aeruginosa* strains isolated from cystic fibrosis patients. Can. J. Microbiol. 37:654–657.
- Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: improved software for multiple sequence alignment. Computer Appl. Biosci. 8:189– 191.
- Holder, I. A., Z. Akchich, and J.-M. Meyer. 1994. Pyoverdin iron binding and virulence of *Pseudomonas aeruginosa*, abstr. B-49, p. 37. *In* Abstracts of the 94th General Meeting of the American Society for Microbiology 1994. American Society for Microbiology, Washington, D.C.
- Ish-Horowicz, D., and J. F. Burke. 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. 9:2989–2998.
- Kleinkauf, H., and H. von Döhren. 1990. Nonribosomal biosynthesis of peptide antibiotics. Eur. J. Biochem. 192:1–15.
- MacCabe, A. P., H. van Liempt, H. Palissa, S. E. Unkles, M. B. R. Riach, E. Pfeifer, H. von Döhren, and J. R. Kinghorn. 1991. δ-(L-α-Aminoadipyl)-Lcysteinyl-D-valine synthetase from Aspergillus nidulans. J. Biol. Chem. 266: 12646–12654.
- Magazin, M. D., J. C. Moores, and J. Leong. 1986. Cloning of the gene coding for the outer membrane receptor protein for ferric pseudobactin, a siderophore from a plant growth-promoting *Pseudomonas* strain. J. Biol. Chem. 261:795–799.
- McMorran, B., and I. Lamont. 1992. Production of pyoverdine by isolates of *Pseudomonas aeruginosa* from patients with cystic fibrosis. Proc. Univ. Otago Med. School 70:11–12.
- Moores, J. C., M. Magazin, G. S. Ditta, and J. Leong. 1984. Cloning of genes involved in the biosynthesis of pseudobactin, a high-affinity iron transport agent of a plant growth-promoting *Pseudomonas* strain. J. Bacteriol. 157:53– 58.
- Morris, J., D. J. O'Sullivan, M. Koster, J. Leong, P. J. Weisbeek, and F. O'Gara. 1992. Characterization of fluorescent siderophore-mediated iron uptake in *Pseudomonas* sp. strain M114: evidence for the existence of an additional ferric siderophore receptor. Appl. Environ. Microbiol. 58:630–635.
- Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85:2444–2448.
- Poole, K., S. Neshat, K. Krebes, and D. E. Heinrichs. 1993. Cloning and nucleotide analysis of the ferripyoverdine receptor gene *fpvA* of *Pseudomonas aeruginosa*. J. Bacteriol. 175:4597–4604.
- 34. Rombel, I. T. 1992. Ph.D. thesis. University of Otago, Dunedin, New Zealand.
- 35. Rombel, I. T., and I. L. Lamont. 1992. DNA homology between siderophore

genes from fluorescent pseudomonads. J. Gen. Microbiol. 138:181-187.

- Rombel, I. T., B. J. McMorran, and I. L. Lamont. Identification of a DNA sequence motif required for expression of iron-regulated genes in pseudomonads. Mol. Gen. Genet., in press.
- Rusnak, F., M. Sakaitani, D. Drueckhammer, J. Reichert, and C. T. Walsh. 1991. Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *entF* gene, expression and purification of EntF, and analysis of covalent phosphopantetheine. Biochemistry 30:2916–2927.
- Saraste, M., P. R. Sibbald, and A. Wittinghofer. 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15:430–434.
- 39. Schlumbohm, W., T. Stein, C. Ullrich, J. Vater, M. Krause, M. A. Marahiel, V. Kruft, and B. Wittmann-Liebold. 1991. An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. J. Biol. Chem. 266:23135–23141.
- 40. Smith, D. J., A. J. Earl, and G. Turner. 1990. The multifunctional peptide synthetase performing the first step of penicillin synthesis in *Penicillium chrysogenum* is a 421 073 dalton protein similar to *Bacillus brevis* peptide antibiotic synthetases. EMBO J. 9:2743–2750.
- Smith, R. 1988. A finite state machine algorithm for finding restriction sites and other pattern matching applications. Computer Appl. Biosci. 4:459–465.
- Sriyosachati, S., and C. D. Cox. 1986. Siderophore-mediated iron acquisition from transferrin by *Pseudomonas aeruginosa*. Infect. Immun. 52:885–891.
- Staab, J. F., M. F. Elkins, and C. F. Earhart. 1989. Nucleotide sequence of the *Escherichia coli entE* gene. FEMS Microbiol. Lett. 59:15–20.
- Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res. 10:2951–2961.
- Staden, R., and A. D. McLachlan. 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. 10:141–156.
- 46. Stockwell, P. A. (University of Otago, Dunedin, New Zealand). 1988. Personal communication.
- Stockwell, P. A. 1985. VTUTIN: a full screen gel management editor. Computer Appl. Biosci. 1:253–259.
- 48. Stockwell, P. A. 1987. DNA sequence analysis software, p. 19-45. In M. J.

Bishop and C. J. Rawlings (ed.), Nucleic acid and protein sequence analysis. IRL Press, Oxford.

- Stockwell, P. A. 1988. HOMED: a homologous sequence editor. Trends Biochem. Sci. 13:322–324.
- Taraz, K., R. Tappe, H. Schröder, U. Hohlneicher, I. Gwose, H. Budzikiewicz, G. Mohn, and J. F. Lefèvre. 1991. Ferribactins—the biogenetic precursors of pyoverdins [1]. Z. Naturforsch. 46c:527–533.
- 51. Thorn, R. L., M. E. Merriman, and I. L. Lamont. Unpublished data.
- Tolmasky, M. E., L. A. Actis, and J. H. Crosa. 1993. A single amino acid change in AngR, a protein encoded by pJM1-like virulence plasmids, results in hyperproduction of anguibactin. Infect. Immun. 61:3228–3233.
- Turgay, K., M. Krause, and M. A. Marahiel. 1992. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. Mol. Microbiol. 6:529–546.
- Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.
- 55. Visca, P., A. Ciervo, and N. Orsi. 1994. Cloning and nucleotide sequence of the *pvdA* gene encoding the pyoverdin biosynthetic enzyme L-ornithine N⁵oxygenase in *Pseudomonas aeruginosa*. J. Bacteriol. **176**:1128–1140.
- Von Döhren, H. 1993. The organization of multifunctional peptide and depsipeptide synthetases. Biochem. Soc. Trans. 21:214–217.
- 57. Weckermann, R., R. Fürbaβ, and M. A. Marahiel. 1988. Complete nucleotide sequence of the tycA gene encoding the tyrocidine synthetase I from *Bacillus brevis*. Nucleic Acids Res. 16:11841.
- West, S. E. H., and B. H. Iglewski. 1988. Codon usage in *Pseudomonas* aeruginosa. Nucleic Acids Res. 16:9323–9335.
- 59. Yager, T. D., and P. H. von Hippel. 1987. Transcript elongation and termination in *Escherichia coli*, p. 1241–1275. *In* J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.