Characterization of a Novel Chromosomal Virulence Locus Involved in Expression of a Major Surface Flagellar Sheath Antigen of the Fish Pathogen Vibrio anguillarum

ANDERS NORQVIST^{†*} AND HANS WOLF-WATZ

Department of Cell and Molecular Biology, University of Umeå, S-901 87 Umeå, Sweden

Received 9 November 1992/Accepted 11 March 1992

The fish pathogenic bacterium Vibrio anguillarum 775.17B was mutated by the use of transposon Tn5-132. Two hundred independent exconjugants were isolated and screened for a reduction of virulence in experimental infections of rainbow trout (Onchorhynchus mykiss). Two of these exconjugants, VAN20 and VAN70, showed a significant reduction in virulence after both intraperitoneal and immersion infections. The avirulent mutants showed no loss of any previously suggested virulence determinants of V. anguillarum. One of the mutants (VAN70) was further characterized. DNA sequence analysis revealed two open reading frames, the gene into which Tn5-132 had been inserted (virA) and a closely linked upstream gene (virB). A virB mutant of 775.17B, NQ706, was isolated and also shown to be avirulent. The deduced amino acid sequences of virA and virB correspond to proteins with molecular weights of 36,000 and 42,000, respectively. Insertional mutagenesis of the corresponding virA and virB genes of a clinical isolate of V. anguillarum, serotype O1, also resulted in avirulence. In immunoblot experiments, the total cell lysates of VAN70 (virA) and NQ706 (virB) did not respond to a rabbit polyclonal antiserum directed against whole cells of 775.17B (wild type). This suggests that virA and virB are involved in the biosynthesis of a major surface antigen important for the virulence of V. anguillarum. Immunogold electron microscopy showed that a constituent of the flagellar sheath was expressed by 775.17B (wild type) but not by VAN70 (virA) and NQ706 (virB), suggesting that the major surface antigen lacking in VAN70 and NQ706 is located on the outer sheath of the flagellum. Analysis of this major surface antigen revealed it likely to be lipopolysaccharide. Further analysis showed that the flagellum and the major surface antigen were expressed in vivo during fish infections.

The fish pathogen Vibrio anguillarum is the causative agent of vibriosis, a widespread bacterial disease among farmed fish. Although vibriosis has been the cause of a number of epizootics in fish farms, studies of the bacterial properties contributing to virulence have so far been limited. A major virulence factor is the plasmid-encoded, ironsequestering system which enables the bacterium to survive in the iron-deficient interior of the fish (40). The genetic organization and the regulation of this iron-sequestering system have been extensively investigated by Crosa and coworkers (1, 6, 30, 38). Loss of the plasmid or mutations within the iron-sequestering genes have been shown to render the bacterium avirulent (7, 33, 38). Another factor which has been indicated to be important for the invasion of V. anguillarum is the extracellular zinc metalloprotease. A regulatory mutation that resulted in a reduced extracellular proteolytic activity showed a concomitant reduction of virulence when rainbow trout (Onchorhynchus mykiss) were infected by immersion (25).

Properties such as serum resistance (36), possession of flagella (5), and production of extracellular cytotoxic products (17) and hemolysin (22) have also been suggested to be involved in the virulence, but so far only limited data have been shown. Thus, the role of these suggested determinants in virulence remains to be proven.

The aim of this study was to search for novel virulence factors of V. anguillarum by transposon mutagenesis followed by a screening for avirulent mutants by experimental

fish infections. Among 200 independently isolated mutants, 2 avirulent mutants were isolated. We describe the genetic characterization of one of these mutants, VAN70, and show that this mutant is affected in the expression of a major surface antigen which is located solely on the flagellar sheath and which is essential for the virulence of *V. anguillarum*.

MATERIALS AND METHODS

Bacterial strains and plasmids. Bacterial strains and plasmids used in this study are described in Table 1. Serotyping of the *V. anguillarum* strains is by the method of Sörensen and Larsen (34).

Conjugation and transposition experiments. Plasmid pRK2013, which contains the transposon Tn5-132 (tetracycline resistance), was transferred from *Escherichia coli* C600 to *V. anguillarum* 775.17B by conjugation on filters by the method of Norqvist et al. (24). Derivatives of the suicide vector pNQ705 were also transferred from *E. coli* SM10, which expresses the π protein necessary for the replication of the plasmids, to strains of *V. anguillarum* by filter conjugations. Earlier studies have shown that the isolation of spontaneous antibiotic-resistant strains of *V. anguillarum* might affect the virulence (24). Therefore, when insertional mutagenesis of *V. anguillarum* NB10 was performed, thiosulfate-citrate-bile-sucrose (Difco) agar plates were used as the counterselection, thereby avoiding the use of antibiotic-resistant strains as recipients.

Experimental fish infections. Experimental fish infections of rainbow trout (*O. mykiss*) were performed as described previously (24). The 50% lethal dose (LD_{50}) was calculated by the method of Reed and Muench (27).

^{*} Corresponding author.

[†] Present address: Department of NBC Defence, National Defence Research Establishment, S-901 82, Umeå, Sweden.

Strain or plasmid	Description	Reference or source
E. coli		<u>. 1</u>
C 600	pRK2013::Tn5-132, Tet ^r	9
S17-1	thi pro hsdR hsdM ⁺ recA RP4-2-Tc::Mu-Km::Tn7	32
SM10	thi thr leu tonA lacY supE recA RP4-2-Tc::Mu-Km	32
Y1088	supE supF metB trpR hsdR hsdM ⁺ tonA21 strA lacU169 (proC::Tn5 [pMC9])	42
Y1090	ΔlacU169 proÄ Δlon araD139 strA supF (trpC22::Tn10 [pMC9])	42
BL21	$F^- ompTr_B^-m_B^-$ pLysS	28
V. anguillarum		
775.17B	wt, serotype O1, Rif ^r Str ^r Amp ^r Kan ^r	24
VAN70	775.17B (virA::Tn5-132), Tet ^r	24
NB10	wt, serotype O1, Amp ^r Kan ^r	24
NQ104	NB10 (virA::pNQ7051), Cml ^r	This study
NQ105	NB10 (pNQ7051), Cml ^r	This study
NQ107	NB10 (<i>virB</i> ::pNQ7056), Cml ^r	This study
NQ706	775.17B (<i>virB</i> ::pNQ7056), Cml ^r	This study
6018/1	wt, serotype O1, Amp ^r Kan ^r	34
53-507	wt, serotype O1, Amp ^r Kan ^r	34
S. typhimurium LT2A	wt	B. N. Ames
Plasmid		
pUC19	Amp ^r	41
pUC4K	Kan ^r Amp ^r	37
pSUP202	Amp ^r Cml ^r Tet ^r	32
pSKT/pSKC + derivatives	Cml ^r /Tet ^r ; see text	This study
pGP704	Amp ^r	21
pNQ705	Cml ^r ; see text	This study
pNQ7051	Cml ^r ; see text	This study
pNQ7056	Cml ^r ; see text	This study
pET3a	Amp ^r	28
pAFO1	Amp ^r Cml ^r ; see text	This study
Bacteriophage λpir	pirRK6	18

TABLE 1. Bacterial strains and plasmids used in this study

Detection of proteolytic and hemolytic activities and of iron-sequestering capacity. Detection of protease activity was performed by patching the bacteria on Trypticase soy agar containing 1% skim milk (29). A transparent zone around the colony indicated proteolytic activity. Hemolytic activity was indicated by a zone of clearing when the bacteria were patched onto Trypticase soy agar plates supplemented with 5% horse blood. Bacteria were tested for the ability to sequester iron by patching them onto minimal salts (M9) agar plates supplemented with 2% NaCl and 25 μ M EDDA [ethylenediamine-di(*o*-hydroxyphenyl) acetic acid].

Construction of genomic libraries in $\lambda gt11$. Chromosomal DNA of the transposon mutant VAN70 was digested with *Eco*RI and ligated to *Eco*RI-digested dephosphorylated $\lambda gt11$ arms (Promega). The ligation mixture was then packaged into phage heads (Promega). Amplification of the recombinant phage library was done with *E. coli* Y1088 (42). A genomic library of strain 775.17B was also constructed. An adapter with *Eco*RI and *Bam*HI sticky ends was ligated to DNA fragments from a *Sau3A* partial digest of 775.17B chromosomal DNA, phosphorylated with T4 polynucleotide kinase, and then ligated to *Eco*RI-digested dephosphorylated $\lambda gt11$ arms. Packaging of the recombinant DNA into the λ phage heads and amplification of the genomic library were done as described above.

Screening of λ gt11 genomic libraries. Recombinant phages from the genomic libraries of strain 775.17B or VAN70 were plaqued on *E. coli* Y1090 (42). The phage DNAs of the

plaques were absorbed to nitrocellulose filters (BA85; Schleicher & Schuell), denatured with 0.5 M NaOH-1.5 M NaCl, neutralized with 0.5 M Tris-HCl-1.5 M NaCl, and baked under vacuum at 80°C for 2 h. Prehybridizations, hybridizations, and filter washings were performed by the methods of Maniatis et al. (20). Filters were then exposed to Dupont Cronex4 X-ray film with an intensifying screen. Positive plaques were picked and phage DNA was prepared by the methods of Maniatis et al. (20).

The probe used in screening the VAN70 genomic library was a $[\gamma^{-32}P]ATP$ 5'-end-labelled, 30-bp synthetic oligonucleotide, complementary to a part of the IS_L of Tn5-132. A 600-bp fragment of *Vibrio* DNA from a positive clone of the VAN70 genomic library was used as a probe when the λ gt11 genomic library of 775.17B was screened. The probe in this screening was radiolabeled with the Amersham Multiprime DNA Labelling System per the manufacturer's instructions.

Subcloning and DNA sequencing. Restriction fragments from a positive $\lambda gt11$ recombinant phage of 775.17B were subcloned into pUC19 (41). Dideoxy sequencing (31), with T7 DNA polymerase (Pharmacia), of plasmid DNA, isolated by the alkaline lysis method (20), was done by walking along each fragment with 17-bp oligonucleotide primers, which were synthesized on a Beckman SM automated DNA synthesizer. The sequencing data were analyzed with the computer program of the Genetics Computer Group (Madison, Wis.).

The virA and virB genes were expressed by the phage T7

expression system described by Rosenberg et al. (28), by cloning polymerase chain reaction (PCR)-generated fragments containing either *virA* or *virB* into the translation vector pET3a. The PCR primers used were 30-bp oligonucleotides complementary to bases (in Fig. 1) 372 (5') to 401 (3') (primer B1) and 1453 (3') to 1482 (5') (primer B2) for *virB* and bases 1470 (5') to 1499 (3') (primer A1) and 2376 (3') to 2405 (5') (primer A2) for *virA*.

Construction of recombinant plasmids used for transcomplementation experiments with VAN70 and NQ706. The mobilizable cloning vector pSUP202 (32) was shown to replicate in V. anguillarum and was therefore used as the vector in transcomplementation studies. From a positive $\lambda gt11$ recombinant phage of 775.17B, a 3.3-kb KpnI fragment containing virA and virB was subcloned into pSUP202 as follows. pSUP202 was digested with BamHI and SalI. The digested vector was purified from an agarose gel and ligated to the KpnI fragment by using two short synthetic oligonucleotides, which ligate to BamHI or SalI sticky ends and generate a KpnI sticky end. This recombinant plasmid was designated pSKT1.

pSKT1 was further manipulated (schematically illustrated in Fig. 2) for use in transcomplementation studies. (i) A 306-bp in-frame deletion (bases 609 to 914 in Fig. 1) was generated in virB by using a PCR method described by Ho et al. (15). The PCR product was then digested with Bg/II and SphI and ligated to Bg/II-SphI-digested pSKT1. In the generated PCR fragment of the resulting plasmid, designated pSKT4, BamHI and SpeI restriction sites were also inserted. (ii) A virA deletion was generated by digesting pSKT1 with SphI and BstEII, filling in the ends with T4 DNA polymerase, and then ligating (derivative designated pSKT9). (iii) An out-of-frame mutation in virB, pSKT42, was constructed by cloning a BamHI-digested kanamycin gene block from plasmid pUC4K (37) into virB of plasmid pSKT4, which had been digested with Bg/II and BamHI.

The 3.3-kb *Kpn*I fragment was also subcloned into the *Eco*RI site of pSUP202 by first digesting the vector with *Eco*RI and then ligating the digested pSUP202 with the *Kpn*I fragment by the use of a short synthetic oligonucleotide, which when ligated with the digested vector generated compatible ends with the *Kpn*I-digested DNA. The resulting plasmid was designated pSKC1. An in-frame deletion in *virB* of pSKC1, designated pSKC5, was constructed with the same strategy as for pSKT4.

For transcomplementation studies, the recombinant plasmids were mobilized from *E. coli* S17-1 (32) into *V. anguillarum* 775.17B (wild type [wt]), VAN70 (*virA*), and NQ706 (*virB*) by conjugation.

Construction of suicide vectors for insertional mutagenesis of virA and virB. Since the V. anguillarum strains used in this study were ampicillin resistant, the ampicillin resistance gene in the suicide plasmid pGP704 (21) was replaced by the chloramphenicol acetyltransferase (CAT) gene of pAF01. Plasmid pAF01 (kindly provided by Åke Forsberg, Umeå, Sweden) is a pBluescript derivative into which the CAT gene of pACYC184 has been subcloned. pACYC184 was digested with XhoII and AccI, generating a 1.4-kb fragment that harbored the CAT gene. This fragment was blunt ended and ligated into the EcoRV site of pBluescript, thereby generating plasmid pAF01. Plasmid pGP704 was digested with PstI (unique site in the ampicillin gene) and SalI (unique site in the multiple cloning cassette). The digested vector was gel purified and ligated to a PstI-SalI fragment from plasmid pAF01 which contained the CAT gene. The resulting plasmid, designated pNQ705, encoded chloramphenicol resistance instead of ampicillin resistance.

Plasmid pNQ705 can be mobilized into V. anguillarum by transfer functions provided by a derivative of RP4 integrated into the chromosome of SM10 (32), but it cannot replicate in V. anguillarum because the π protein, which is needed for plasmid replication, is lacking in V. anguillarum.

For insertional mutagenesis of V. anguillarum, recombinant plasmids of pNQ705 which contained internal fragments of virA and virB (underlined in Fig. 1) were constructed. The recombinant plasmids were transferred into V. anguillarum strains by conjugation as described above.

Preparations of total cell lysates and LPS. Total cell lysate preparations of *V. anguillarum* strains and proteinase K treatment of the total cell lysates were done by the methods of Hitchcock and Brown (14). When total cell lysates were treated with periodate, the cells were incubated for 1 h in the dark in 30 μ l of 10 mM periodic acid in 50 mM sodium acetate, pH 4.5. After incubation, 20 μ l of 4× sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer (19) was added, and the samples were incubated for 10 min at 95°C. Lipopolysaccharides (LPS) were prepared by the method of Westphal and Jann (39). When LPS were intrinsically radiolabelled with ³²P, cells were grown in Trypticase soy broth medium containing 10 μ Ci of NaH₂[³²P]O₄ per ml.

SDS-PAGE and immunoblotting. SDS-PAGE of total cell lysates and LPS preparations were performed by the methods of Laemmli and Favre (19). Immunoblotting (Western blotting) after SDS-PAGE was performed essentially as described by Swanson et al. (35). The antiserum used in the immunoblot experiments was either a rabbit antiserum raised against formalin-killed whole cells of *V. anguillarum* 775.17B (24) or a rabbit antiserum raised against LPS of *V. anguillarum* 40T (kindly provided by D. H. Shaw, Newfoundland, Canada).

Immunogold electron microscopy. Preparation of the samples for the immunogold electron microscopy was done as described by Nyberg et al. (26).

Recombinant DNA techniques. Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, T4 polynucleotide kinase, and the Klenow fragment of polymerase I were used as described by Maniatis et al. (20). Restriction fragments were isolated from agarose, using the GeneClean Glassmilk kit (Bio101 Inc.).

RESULTS

Isolation of avirulent transposon mutants of V. anguillarum. V. anguillarum 775.17B was mutagenized by transposition of Tn5-132. Southern hybridization, using EcoRI-digested chromosomal DNA of six randomly picked exconjugants and an oligonucleotide probe complementary to Tn5-132, showed that the transposon incorporated randomly into the chromosome (data not shown). Two hundred exconjugants were screened for their virulence properties by immersion infections of rainbow trout (O. mykiss) at 18°C for 30 min. The infection dose was 5×10^6 bacteria per ml, which was the LD_{50} of the parent strain 775.17B. Mortality was monitored daily for 7 days. Two avirulent exconjugants, VAN20 and VAN70, were isolated. To confirm the avirulent phenotypes of the mutants, LD_{50} s were determined. Results were similar for both mutants. The LD_{50} after intraperitoneal (i.p.) infection was 10⁷ bacteria; after infection by immersion, the LD_{50} exceeded 2 × 10⁸ bacteria per ml. The corresponding LD_{50} values for the parent strain 775.17B were 10^3 (i.p.) and 5×10^6 (immersion) per ml.

Characterization of the transposon mutants. The two avirulent mutants were tested for the ability to sequester iron by growth on minimal plates containing the iron chelator EDDA. They both grew, indicating that the transposon had not been incorporated into the 65-kb pJM1 plasmid (38) or into some chromosomally located genes involved in the iron uptake system of the bacterium. This was confirmed by digesting the 65-kb plasmids of the two mutants with *Bam*HI. No difference was observed when the digestion pattern of the plasmid from the parent strain 775.17B was compared with those of the plasmids from the mutants (data not shown).

Southern hybridization of EcoRI-digested chromosomal DNA from the two mutants revealed that two copies of the transposon had been incorporated into the chromosome of VAN20, while only one copy was inserted into the chromosome of VAN70 (data not shown). We therefore concentrated our studies on VAN70.

Two extracellular proteins have been suggested to be involved in the virulence of V. anguillarum, a metalloprotease (25) and a hemolysin (22). The ability of VAN70 to express proteolytic and hemolytic activities was detected by a zone of clearing on Trypticase soy agar plates containing either 1% skim milk or 5% horse blood. The parent strain 775.17B and the mutant VAN70 expressed comparable proteolytic as well as hemolytic activities (data not shown).

Growth rate and auxotrophy are two other factors that might affect the virulence of bacteria. To study whether the incorporation of the transposon changed the growth rate of VAN70 and/or whether the mutant had become auxotrophic, 775.17B and VAN70 were grown at 20°C in MOPS (morpholinepropanesulfonic acid) minimal medium containing only salts and glucose as the carbon source (23). Both strains grew well in the minimal medium, showing that the mutant had not become auxotrophic. Moreover, the generation time for both 775.17B and VAN70 was 210 min, indicating that the change in virulence was not due to a change in growth rate of the mutant.

Cloning of the transposon-inactivated gene and nucleic acid analysis. To clone the transposon-inactivated gene, a λ gt11 genomic library of *Eco*RI-digested chromosomal DNA of mutant VAN70 was constructed and screened by using a probe complementary to the transposon. Plaques containing part of the transposon-inactivated gene were isolated by plaque hybridization, and a 600-bp fragment of the target gene (isolated from a positive plaque of VAN70) was used as a probe in plaque hybridization of a λ gt11-775.17B (wt) genomic library. This resulted in the isolation of a recombinant phage containing a 6-kb fragment of 775.17B DNA. Restriction enzyme analysis showed that the transposon insertion site mapped in the middle of the 6-kb fragment.

A 3.3-kb KpnI fragment of the cloned 6-kb fragment was subcloned into pUC19 and sequenced (Fig. 1). The sequence revealed two open reading frames, virA and virB, which encode possible polypeptides of 304 and 362 amino acids, respectively. The insertion site of Tn5-132 was mapped to the 3' end of virA (base 2281 in Fig. 1). A potential promoter, homologous to *E. coli* consensus sequences, was found upstream of the putative transcriptional start of virB. In addition, a second possible promoter was found upstream of the suggested transcriptional start of virA and within virB. Potential terminator sequences were found in the 3' end of both virB and virA (Fig. 1). No other open reading frames could be found when the DNA sequences 380 bp upstream of *virB* and 800 bp downstream of *virA* were analyzed.

The virA and virB genes were subcloned under the control of the phage T7-inducible expression system (28). After induction of the virA and virB clones, protein profiles of total cell lysates were analyzed by SDS-PAGE. The virA-containing construct overexpressed a protein of 36 kDa, and the virB construct showed a predominant band of 39 kDa (data not shown). The molecular weights of these proteins corresponded to the estimated sizes calculated from the deduced amino acid sequences of virA and virB.

Insertional mutagenesis in virB. To investigate whether VirB was also an essential virulence determinant of V. anguillarum, an insertional mutation was made in virB. A 340-bp internal virB fragment (underlined in Fig. 1) was cloned into the suicide vector pNQ705 (Cml^r). The resulting plasmid (pNQ7056) was mobilized into the wt strain 775.17B by conjugation. Since pNQ705 derivatives cannot replicate in V. anguillarum, chloramphenicol-resistant exconjugants should contain pNQ7056 integrated into the chromosome within virB by a single homologous recombination step. The insertion of pNQ7056 into virB was confirmed by PCR, using a virB-specific primer and a pNQ705-specific primer (see Fig. 4, lanes I and J), and by Southern hybridization of XbaIdigested chromosomal DNA of the isolated exconjugants (data not shown). One exconjugant, designated NQ706 (virB), was used for i.p. infections of rainbow trout (O. mykiss). The LD₅₀ of NQ706 was 7×10^6 , compared with the parent strain 775.17B, which had an LD_{50} of 10³, indicating that virB was also important for the virulence of V. anguillarum.

Transcomplementation of VAN70 and NQ706. To establish whether *virB* and *virA* each were essential for the virulence of *V. anguillarum*, plasmid derivatives of the mobilizable cloning vector pSUP202 (Cml^r Tet^r Amp^r) containing *virA* or *virB* or both were constructed and used for transcomplementation of VAN70 (*virA*) and NQ706 (*virB*).

A 3.3-kb KpnI fragment, containing virA and virB, was subcloned into either the tetracycline gene (resulting plasmid pSKT1) or the CAT gene (resulting plasmid pSKC1) of pSUP202. Subcloning into the CAT gene was necessary for use of the plasmid derivatives in transcomplementation studies of NQ706 (virB), which was chloramphenicol resistant. The subcloned KpnI fragment was oriented in the opposite direction compared with the tetracycline and chloramphenicol genes to ensure that the expression of virA and virB was determined by their own promoters. pSKT1 and pSKC1 were further manipulated by creating an in-frame mutation in virB (pSKT4 and pSKC5), by deleting a part of virA (pSKT9), and by inserting a kanamycin gene block into virB (pSKT42). The constructs are illustrated in Fig. 2 (for detailed descriptions of the construction of the different plasmid derivatives, see Materials and Methods).

The various plasmid derivatives were transferred by conjugation to VAN70 (*virA*), NQ706 (*virB*), and also to the parent strain 775.17B. Experimental i.p. fish infections with the various exconjugants were performed. The LD₅₀s of 775.17B (wt), VAN70 (*virA*), NQ706 (*virB*), and the exconjugants are shown in Table 2. As shown, pSKT1 (*virA*⁺ *virB*⁺), pSKT4 (*virA*⁺), and pSKT42 (*virB*::kana *virA*⁺) restored the virulence of VAN70 (*virA*). On the other hand, the transcomplementation analysis of NQ706 (*virB*) was not as conclusive as that of VAN70 (*virA*). Neither pSKC1 (*virA*⁺ *virB*⁺) nor pSKC5 (*virA*⁺) could transcomplement NQ706 (*virB*). However, the negative results could be explained by poor growth of the strain carrying pSKC1 and

	10	30	50	70	90
GGTACCT	TTAGTCGAAGATGATAATG	CGCATGTATGGCACCTTTT'	TGTGGTAGTGGTCGAGGATAG	AGATAAATTAGTCGAACATC	TGAGTGAAAACCAA
ATTCAGT	CATTAATACATTACCCAAT	TCCACCACACAAGCAAGAG	I SU GCTTATATGGAGTGGAATAAC	AGTGCCTTTCCGTTATCAGA	GAAAATGCATAAGC
2	210	230	250	270	290
AAGTTTT	AAGTCTTCCACTAAGCTCT	GTTTTGGATAAAAGTGAAA'	TTGACAAAGTTATAGATGTAT	TGAACTCTTTTTCTTGGTTAA	TTAAGCCGTCATAC
8T8CTTT	510 FTCTCCBCCBTCBCTTTTT	(- <i>53)</i> CCAAAAACTCTTTACTATT	(-10)	(SD) start virB	390
AIAGIII		GOAAAAACIGIIIACIAI	AAAAIAATIGAGCITCCATTA	M H	L S D S K
4	410	430	450	470	490
GCCTCCT	TAGTAAGTTTATGTATAC	TAACTTTTTAATCATGAGAA	ATATATAACTAAAGCAATAAA	TAGTTGTTTGGCGCAATCTT	ATAGTAATATAGAG
	510 S L C I L	530 FNHER	IITKAIN 550	SCLAUSI 570	5 N I E 590
ATAATTA	ITGTAGACAATAATTCAAG	TGATGGAACTGTAAATAAG	ATTAGATCTGATTTTAAAAAT	GAGCTAGAAGTTGGGGAAAT	AAAGCTTTTTGATT
III	VDNNSS	DGTVNK	IRSDFKN	ELEVGEI	KLFDL
TGGAACA	DIU CAATACTTATCCTTCACAT	030 667777788677876686768	650 333367070336636337376	670 	690 CATACCCTTTCTT
ЕН	NTYPSH	GFNYALK	K S Q G E Y V	S L F S G D	DTLCL
	710	730	750	770	790
N K V	VEROIGT	M V K E G L	S N L F T W V	N I I N D K D	ATGAAATTATAAAG
	810	830	850	870	890
TGTGATT	ACTTAGAAAGTATTTTTAA	TCGCAATTATAATAGCCAA	CAAATAAAAGAGCATTTTATT	CATAGTGGGAATATGTTATC	CGCACTGTCCGTCA
	910 E S I F N	RNYNSQ (930	950 950	H S G N M L S	A L S V M
TGCTATC	TAGA GATGTATTTGATAGG	TATEGTCATTTCGACGAGC	GTCTTGTTCAGCTACAGGATT	TTGATTTTTGGCTGCGTATG	GCCTCTAATGATGA
LS	RDVFDR	YGHFDER	LVQLQDF	DFWLRM	ASNDD
11 TCTTAAT	010 ГТАТТСАСССАААААСТАТ	1030 CAAACTATAGATTGAGAGA	1050 TGATGGCGGAAATTTAAGTCT		1090 Эстальсьюетсья
LNI	LLTEKLS	N Y R L R D	D G G N L S L	A N H K S R O	L R T D
11	110	1130	1150	1170	1190
TTTGAAGA	AAGTTTATGTATATAGACA	I I N F D I	AAGACAATTCAAAGTGTTGTT	GGTATGCTGAATAAAGACCA	ATCTATCGCTATGG
12	210	1230	1250	1270 L N K D Q	1290 A M A
CTCTGCA	IGGCTATTATCATAATGAG	AATAAAATGAAGTTGGCCA	AAGGTTTTTTGTTGTCCATAT	ATGAAGAATTAGGGACAAAT	ATTGTTTTCCCTTC
LH	GYYHNE	N K M K L A K	GFLLSIY	EELGTN	IVFPS
ATCTCACT	TATAGCTACTTCTTTGACA	TATATTCTAAATGTGAATT	TTTCAACTCAGATGAGAATGA	TGAAGTATCAAAGCTGAAAG	1390 GAGAAAATACACGTT
SHY	Y S Y F F D I	YSKCEF	FNSDEND	EVSKLKE	скіну
(-35)1	410 (-10)	1430	1450 (SD)	1470 start virA	(← Terminator
TATGAAA	ATTCTAGGGCGAT <u>TAGATT</u>	TACTAGCTTAATCGTTTCT	TGGCTTGGT <u>AGGTT</u> AAAAAAA	TAAGTGAGAAAA <u>ATG</u> AAGTA	TCCACGAAGCTTAT
YEN	SRAIRF E10	TSLIVS	WLGRLKK	мку	PRSLS
				1 6 773	
CATGGGA	AAAAATTTCTAGCAGTACT	ATCTCTATTAGAGAGTTTA	ATCGTTTTGAGAATAAGAAAA	1570 AAGTAGCTATATTTTGTGGG	TATGTAAGGGATAA
CATGGGA	AAAAATTTCTAGCAGTACT K I S S S T	ATCTCTATTAGAGAGTTTA I S I R E F N	ATCGTTTTGAGAATAAGAAAA R F E N K K K	1570 AAAGTAGCTATATTTTGTGGT V A I F C G	TATGTAAGGGATAA Y V R D K
CATGGGA	AAAAATTTCTAGCAGTACT K I S S S T 510	ATCTCTATTAGAGAGTTTA I S I R E F N 1630	ATCGTTTTGAGAATAAGAAAA R F E N K K K 1650	1570 NAAGTAGCTATATTTTGTGGT V A I F C G 1670	TATGTAAGGGATAA Y V R D K 1690
CATGGGA W E 10 ATATGAAT Y E F	AAAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T	ATCTCTATTAGAGAGTTTA ISIREFN 1630 CTTATCAATGGTTAAGAAA YOWLRN	ATCGTTTTGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y Y V T L	1570 NAAGTAGCTATATTTTGTGGT (V A I F C G 1670 NAGTTATGCCTAAAAGTGGAA V M P K S G I	TAGGAAGGGATAA Y V R D K 1690 <u>ATTCTACTAGAATCT</u>
CA <u>TGG</u> GAJ W E 10 ATATGAAT Y E F	AAAAATTTCTAGCAGTACT K I S S S T 510 ITTAACTATCATGATAGAA F N Y H D R T 710	ATCTCTATTAGAGAGTTTA ISIREFN 1630 CTTATCAATGGTTAAGAAA YQWLRN 1730	ATCGTTTTGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y Y V I L 1750	1570 IAAGTAGCTATATTTTGTGGT (VAIFCG 1670 AGTTATGCCTAAAAGTG <u>GAA</u> VMPKSGI 1770	TATGTAAGGGATAA Y V R D K 1690 <u>NTTCTACTAGAATCT</u> I L L E S 1790
CATGGGA W E ATATGAAT Y E F 17 GATTTAT	AAAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T 710 CAAAGCTTGTGATGTTTT	ATCTCTATTAGAGAGTTTA ATCTCTATTAGAGAGTTA I S I R E F N 1630 CTTATCAATGGTTAAGAAA Y Q W L R N 1730 TATTGAAAGAGAAAACTTT	AIGGTITTGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GGCTACGATTTTGGATCTTAT	1570 IAAGTAGCTATATTTTGTGGT (VAIFCG 1670 AGTTATGCCTAAAAGTG <u>GAA</u> VMPKSGI 1770 ICCATGCGGATTACAATATGT	TTATGTAAGGGATAA Y V R D K 1690 <u>TTCTACTAGAATCT</u> I L L E S 1790 TTAACTTAATCCAGG
CATGGGA W E ATATGAAT Y E E 17 GATTTAT D L C	AAAAATTTCTAGCAGTACT K I S S S T 510 TITTAACTATCATGATAGAA F N Y H D R T 710 CAAAGCTTGTGATGTTTT K A C D V F 810	ATTICATTAGAGAGTTTA ISIREFN 1630 CITATCAATGGTTAAGAAA YQWLRN 1730 TATTGAAAGAGAAAACTTT IERENFC	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y Y V I L 1750 GOCTACGATTTTGGATCTTAT 5 Y D F G S Y 1850	1570 LASCTAGCTATATTTTGGGI V A I F C G 1670 V M P K S G I 1770 GCATGCGGATTACAATATGI A C G L Q Y V	TIATGTAAGGGATAA Y V R D K 1690 <u>NTTCTACTAGAATCT</u> I L L E S 1790 <u>TAACTTAATCCAGG</u> N L I E G
CATGGGAN W E 10 ATATGAAT Y E F 17 GATTTATC D L C 18 GCTCAGAC	ANAAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T 710 CAAAGCTTGTGATGTTTTT K A C D V F 310 SIGATTGATAGGTTACTT	ALSO ARCTCTATTAGAGAGTTTA I S I R E F N 1630 CTTATCAATGGTTAAGAAA Y Q W L R N 1730 TATTGAAAGAGAAACTTT I E R E N F C 1830 TYCGTGAAGAGATAGTTTAA	ADJO ATCGTTTTGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GGCTACGATTTTGGATCTTAT 5 Y D F G S Y 1850 TGGGCCTTTGGGATACTGCA	1570 LISTOR CTATATTTTGGGI $\langle V A I F C G$ 1670 CAGTATAGCCTAAAAGTG <u>GAA</u> V M P K S G I 1770 CGCATGCGGATTACAATAGGA A C G L Q Y V 1870 ATTAATAGAAGAATTCTAGC	TAJGTAAGGGATAA Y V R D K 1690 <u>VITCTACTAGAATCT</u> L L E S 1790 <u>TAACTTAATCCAGG</u> N L I E G 1890 <u>GAATTTGGGGAAA</u>
CATGGGAN W E ATATGAAT Y E F GATTTAN D L C GCTCAGAC S E	AAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T 710 SCAAAGCTTGTGATGTTTT K A C D V F 810 INGGATTGTTAGTAGGTTACTT R I D R L L	1330 ATCTCTATTAGAGAGATTTA ISIREFN 1630 CTTATCGATGGTTAAGAAA YQWLRN 1730 TATTGAAAGAGAAAACTTT IERRFC 1830 TCGTGAACGATAGTTTAT	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 LS70 X = X + X + X + X + X + X + X + X + X +	TAJGTAAGGGATAA Y V R D K 1690 <u>IL L E S</u> 1790 <u>IL L E S</u> 1790 N L I E G 1890 <u>GAATTTGGGGAAA</u> E F W G N
CATGGAA W E ATATGAAA Y E I GATTTAT D L C 16 GCTCAGAA S E 19 TACAGAT	MARATITCTAGCAGTACT K I S S S T 510 TITAACTATCATGATAGAA F N Y H D R T 710 SCAACGATGATAGGTTACTT K A C D V F 810 INGCATGATAGGTTACTT R I D R L L 910 900	1330 ACTOCTATTAGAGAGATTTA I S I R E F N 1630 CTTATCAATGGTTAAGAAA Y Q W L R N 1730 I E R E N F C 1830 TTCGTGAACGATAGCTTTTTAT F V N S F I 1930 O S F I 1930	ALLOSTITIGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GCTAGGATTTAGATCTTAG GCTAGGATTTAGATCTTAG 1850 TTGGCCCTTTGGCATATCGCA G P F G Y C N 1950	1570 LASGTAGCTATATTTTGGGT (V A I F C G 1670 1670 V M P K S G I 1770 CCATGCGGATTACAATAGT A C G L Q Y V 1870 ATTTAATAGAAGATTCTACC L I E D S 1970	1390 Y V R D K 1690 VTCTACTAGAATCT 1790 TTAGTTAATCCAGG N L I E G 1890 GAATTTTTGGGGAAA E F W G N 1990 ACCTTTTTTTTTTTTTTTCCC
CATGGAN W E 10 ATATGAAN Y E F GATTTAN D L C 18 GCTCAGAO S E 19 TACAGAT T D S	ARAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T 710 XAAAGCTTGTGATGATGATGATTT K A C D V F 810 INGGATTGATAGGTTACTT R I D R L L 910 GCAACCAAGTAAAATATC S N Q V K Y H	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 LISTOR A LAGTAGETATATTTTGGGT $\langle V A I F C G$ 1670 AGTTATGCCTAAAAGTG <u>GAA</u> V M F S G I 1770 A C G L Q Y V 1870 A C G L Q Y V 1870 A C G L Q Y V 1870 A C G L Q S 1970 ATTTAATAGAAGATTCTAGC AACTTAGATATTAAAAAA N L D I I N N	1330 TATGTAAGGGATAA Y V R D K 1690 <u>TTCTAACTAGAATCT</u> 1 L E S 1790 <u>TTAACTTAATCCAGG</u> N L I E G 1890 N L I E G 1890 R N C N 1990 <u>ACTTTTTTTTTTTACC</u> F F S
CATGGAN W E 10 ATATGAAT Y E F GATTTAN D L C 18 GCTCAGAC S E 19 TACAGAT T D S 20	AAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T 710 CCAAAGCTTGTGATGTTTT K A C D V F 810 INGGATGATAGGTTACTT R I D R L L 910 INGGATGGATAGGTAGATATC 5 N Q V K Y H 101	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccc} 13 & 3 & 3 \\ 15 & 3 & 7 & 7 & 7 & 7 & 7 & 7 & 7 & 7 & 7$
CATGGGAI W E 10 ATATGAAT Y E I 17 GATTTAT GATTTAT D L C 18 GCTCAGAC S E 19 TACAGAT T D S 20 CGAGGGAC	AAAATTTCTAGCAGTACT K I S S S T 510 TTTAACTATCATGATAGAA F N Y H D R T 710 CAAAGCTTGTGATGTTTT K A C D V F 810 INGGATTGTATGGTTACTT R I D R L L 910 GCCAACCAAGTAAAATATC S N Q V K Y H 110 ATATATACAGATGATAA	1350 TATCATTATAGAGAGATTTA I S I R E F N 1630 CTTATCAATGGTTAAGAAA Y Q W L R N 1730 TATTGAAAGAGAAAACTTT I E R E N F C 1830 TCGTGAACGATAGTTTTAT F V N S F I 1930 ACTATCAGAGGTTATCGATATT Y Q S Y L F 2030	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 LS70 X = X + X + X + X + X + X + X + X + X +	1350 TATGTAAGGGGATAA Y V R D K 1690 ITTCAACTAGAATCT I L L E S I790 TAACTTAATCCAGG N L I E G 1890 ACTTITTGGGGAAA E F W G N 1990 ACCTITTTTTTTTTAGC F F F S 2090 ACCATGTAGTGTAGTGTG
CATCGGAI W E 10 ATATGAAA Y E F 17 GATTTAT D L C 18 GCTCAGAA S E 19 TACAGATI T D 2 CCCAGGAG R G D 21 21	$\begin{array}{cccc} & \text{MAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{T} \\ & \textbf{510} \\ & \textbf{TTTAACTATCATGATGATAGAAA } \\ & \textbf{F} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ & \textbf{7} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ & \textbf{7} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{910} \\ & \textbf{300} & \textbf{300} & \textbf{300} & \textbf{300} & \textbf{300} \\ & \textbf{300} & \textbf{300} & \textbf{300} & \textbf{K} & \textbf{M} & \textbf{H} \\ & \textbf{300} & \textbf{V} & \textbf{K} & \textbf{M} & \textbf{H} \\ & \textbf{300} & \textbf{V} & \textbf{K} & \textbf{M} & \textbf{H} \\ & \textbf{300} & \textbf{V} & \textbf{K} & \textbf{M} & \textbf{H} \\ & \textbf{300} & \textbf{V} & \textbf{K} & \textbf{M} & \textbf{H} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{X} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ & \textbf{300} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{M} \\ \end{array} $	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{ccccc} 1390\\ Y & R & D & K\\ 1690\\ YTCTACTAGCAGATCT\\ T & L & L & S\\ 1790\\ TAACTTAATCCAGG\\ N & L & I & E & G\\ 1890\\ GAATTTTGGGGAAA\\ E & F & W & G & N\\ 1900\\ ACTTTTTTTTTTTTTTTGGGGAAA\\ C & F & F & S\\ 2090\\ AAGCATGTAGTGGTGT\\ R & C & S & L\\ 2190\\ \end{array}$
CATGGA, A W E 10 ATATGAAT Y E F 11 GATTTAT D L C 16 GCTCAGAN S E 10 TACAGAT T D S 20 CGAGGAG/ R G D 21 TACATCCC	$\begin{array}{cccc} & \text{MAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{S} \\ \text{510} \\ TTTAACTATCATGATAGAAAAAAAAAAAAAAAAAAAAAA$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{rcrcrc} 1350\\ 1690\\ 171CTAACTAGCAATCTI L L S 1790\\ 17AACTAATCCAGATCTI L L S 1790TAACTTAATCCAGGN L I E G 1890GAATTTGGGGAAAE F W G N 1990ACTTTTTTTTGGGGAAAE F W C N 1990ACATGTAGTGGTGTGT F S 2090AGAGTGTAGTGGTGTGT R C S V L 2190TTTTTATATAATGGT$
CATGGAIN W E 10 ATATGAAT Y E E 17 GATTTAT D L C 18 GCTCAGAN S E 19 TACAGATT T D S 20 CGAGGAG R G D 21 TACATCCC H P	$\begin{array}{ccccc} & \text{SIG} \\ \text{MARATTTCTAGCAGTACT} \\ \text{K I S S T} \\ & \text{SIG} \\ \text{TTTARCTATCATGATAGAAA} \\ \text{F N Y H D R T} \\ & \text{TOAAGCTTGTGATGTTTTT} \\ \text{K A C D V F} \\ \text{SIGGATGATAGGTTACTT} \\ \text{R I D R L L} \\ & \text{SIGGATGATAGGTTACTT} \\ \text{R I D R L L} \\ & \text{SIGGATGATAGATAGATAAATATC} \\ & \text{SIGAACCAAGTAAAATATC} \\ & \text{SIGAACCAAGTAAAATATCG} \\ \text{SIGAACCAAGTTTAAAAATACG} \\ \text{I S V L K S} \\ \end{array}$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
CATGGAN W E 11 ATATGAAT Y E F 12 GATTTATC D L C 16 GCTCAGAN S 12 TACAGAT CGAGGAG 22 CGAGGAG R G D 22 TACATCCC H P 22 CATCATCCC	$\begin{array}{rcrcr} & \text{JAAAATTTCTAGCAGTACT} \\ & \text{K I S S T} \\ & \text{510} \\ & \text{TTTAACTATCATGATGATAGAA} \\ & \text{F N Y H D R T} \\ & JAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 1570 $X \vee A I F C G$ 1670 $X \vee A I F C G$ 1670 $X \oplus I F C G$ $Y \oplus F S G I$ 1770 $X \oplus C G L Q Y V$ 1870 ACC G L Q Y V 1870 ATTTAATAGAAGATTCTAGC L I E D S S 1970 AAACTTAGATATTATAATAA N L D I I N N 2070 TATTTAATGGGAAAGGCCT Y F N G K G L 2170 ATCTAACATCCAAAATTTC L T S K I F 2270 TTA	1330 TARGTAAGGGATAA Y V R D K 1690 ITCTACTAGCATGGATAA 1690 ITCTACTTAATCAGATCT I L L E S 1790 ITAACTTAATCAGGGAAA E F W G N 1990 ISACTTTTTTGGGGAAA E F W G N 1990 ACTTTTTTTGGGGAAA E F F S 2090 AGGATGATGGTGTGT R C S V L 2190 TTTTATATAATGGTGT TTTTATATAATGGTGT F Y I M V 2290
CATGGAIGAN W E 14 ATATGAAA Y E E GATTTATI D L C C CCCAGA R G D C CCCAGGAGA R G D C C CCCAGAGA R G D C C C CCCAGAGA R G D C C C C C C C C C C C C C C C C C C C	$ \begin{array}{cccc} 0.510 \\ MARATTTCTAGCAGTACT \\ K I S S T \\ 510 \\ TTTAACTATCATGATAGAAA \\ F N Y H D R T \\ 710 \\ TCAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT$	1330 TATCATTATAGAGAGATTTA I S I R E F N 1630 CTTATCCATGGTTAAGAAA Y Q W R N 1730 I E R E N F C 1830 TTCGTGGACGATAGTTTTAT I E R E N F C 1930 ACTATCAGAGGTTACCTGTTT Y Q S Y L F 2030 ACTATCAGTAGTTATCGAAAATT S L V I E N E 2130 GATTTTATTAAACAGACTTT D F I K Q T F 2230 ATCAACTATGGTTGCAACTT	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 1570 1570 V A I F C G 1670 AGTTATCCCTAAAGTG <u>GAA</u> V M P K S G I 1770 A C G L Q Y V 1870 A C G L Q Y V 1870 A C G L Q Y V 1870 A C G L Q Y V 1870 AACTTAGAGATATTACAAGATTCTAGC L I E D S S 1970 AAACTTAGATATTATAAATA N L D I I N N 2070 TATTATATGGGAAAGGCCTTTCG L T S K I F ATCCAAGACCCCAAAATTTCC L T S K I F ATCCAAGACCCTTAGAAGCTTTTCG 1 K K L L R	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
CATGGAIN W E 14 ATATGAAA Y E F GATTATGAA D L C 16 GCTCAGAA S E 10 TACAGAT T D 2 CCGACGAG R G D TACATCCC H P 22 GATTGCAC I A F 22 CGATTGCAC I A F 22 CGACGAC I A F 20 CCCAGAC I A F 20 CCCAGACAC I A F 20 CCCAGAC I A F 20 CCCAGACAC I A F 20 CCCAGACAC I A F 20 CCCAGACAC I A F 20 CCCAGACACAC I A F 20 CCCAGACACAC I A F 20 CCCAGCACACAC I A F 20 CCCAGCACACACACAC I A F 20 CCCAGCACACACAC I A F 20 CCCACGACACACAC I A F 20 CCCACGACACACAC I A F 20 CCCACGACACACACACAC I A F 20 CCCACGACACACAC I A F 20 CCCACGACACACAC I A F 20 CCCACGACACACACACAC I A F 20 CCCACGACACACACACACACACACACACACACACACACA	$\begin{array}{rrrr} 3.1.3 \\ MARATTTCTAGCAGTACT K I S S T \\ 510 \\ TTTAACTATCATGATAGATAGAT F N Y H D R T \\ 710 \\ \hline CAAAGCTTGTGATGGTAGTTTT K A C D V F \\ 710 \\ \hline K A C D V F \\ 710 \\ \hline CAAAGCTGTGTGATGGTTACTT R I D R L L \\ 710 \\ \hline S N Q V K Y H \\ \hline S N Q V K Y H \\ \hline S N Q V K Y \\ \hline S N Y \\ \hline S N Q V K Y \\ \hline S N Q V K Y \\ \hline S N Q V K Y \\ \hline S N Y \\ \hline S N Q Y \\ \hline S N Y \\ \hline$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
CATGGGAI W E 11 ATATGGAAT Y E F I GATTTATC D L C I GATTTATC D L C I GATTGGAA S E 1 TACAGATC T D S CCACGGAG R G D I TACATCC H P 22 GATTGCAAC I A F 23 GATTGCAAC	$\begin{array}{ccccc} & \text{MAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{T} \\ & \text{510} \\ & \text{TTTAACTATCATGATGATAGAAA} \\ \textbf{F} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ & \text{710} & SCAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 LISTOCTATATTTTGTGGT X V A I F C G 1670 AGTTATCCCTAAAAGTG <u>GAN</u> Y M F S G I 1770 R C G L Q Y V 1870 ATTTAATAGAAGAATTCTAGC L I E D S S 1970 AAACTAACATCCAAAATTTC L T N G K G L 2170 ATCTAACATCCAAAATTTC L Z K I F ATATAAGAAAGAACTTTGA I K E L L R 2370 AAACAAATCCTACAAAATCA	$\begin{array}{ccccc} 1390\\ Y & R & D & K\\ 1690\\ \hline YTCTACTAGCAGATAA\\ I & L & L & S\\ 1790\\ \hline TAACTTAATCCAGG\\ N & L & I & E & G\\ 1890\\ \hline GAATTTAATCCAGG\\ N & L & I & E & G\\ 1890\\ \hline GAATTTTGGGGAAA\\ E & F & W & G & N\\ 1900\\ \hline TATTATTAATAGTGGT\\ R & C & S & V & L\\ 2190\\ \hline TTTATATAATGGTGT\\ T & M & V\\ 2290\\ \hline GAGATAATCCGACA\\ 2390\\ T & T & TAGGACAATAG\\ 2390\\ \hline \end{array}$
CATGGOAL W E 14 ATATAGAAT Y E I GATTTAT D L C I GCTCAGAM S E I TACAGGATI T D S CCAGGAGGAT R G D I TACATCCO H P 22 CCAGGAGGAT C CCAGGAGGA R G D I CCAGGAGGAG R G D I CCAGGAGGAG R G D I CCAGGAGGAG R G D I CCAGGAGGAG R G D I CCAGGAGGAG R G D I CCAGGAGGAG R G D I CCAGGAGAG R G D I CCAGGAGAG R G D I CCAGGAGAG R D I C CCAGGAGAG R D I C CCAGGAG R D I C C CCAGGAGAG R D I C C CCAGGAGAG R D I C C CCAGGAG R D I C C CCAGGAGAG R D I C C C CCAGGAG C C C C C C C C C C C C C	$\begin{array}{cccc} \text{MAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{T} \\ \text{510} \\ \textbf{TTTAACTATCATGATAGAAA } \\ \textbf{F} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ \textbf{710} \\ \textbf{TTAAAGCTTGTGATGATGATGATTTT } \\ \textbf{K} & \textbf{A} & \textbf{C} & \textbf{D} & \textbf{V} & \textbf{F} \\ \textbf{810} \\ \textbf{MGGATTGATAGGTTACTTT } \\ \textbf{K} & \textbf{R} & \textbf{L} & \textbf{D} & \textbf{R} & \textbf{L} \\ \textbf{910} \\ \textbf{MGGATTGATAGGTAAAATATC } \\ \textbf{S} & \textbf{N} & \textbf{Q} & \textbf{K} & \textbf{M} \\ \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{910} \\ \textbf{GCAACCAAGTAAAATATC } \\ \textbf{S} & \textbf{N} & \textbf{Q} & \textbf{K} & \textbf{K} \\ \textbf{10} \\ \textbf{CATAGTGTTTATATACAGATGATAAATCG } \\ \textbf{I} & \textbf{S} & \textbf{V} & \textbf{K} & \textbf{K} \\ \textbf{10} \\ \textbf{CATAAGTGTTTTAAAATCG } \\ \textbf{SCACAGCGTGAATCCAACTCC } \\ \textbf{R} & \textbf{D} & \textbf{V} & \textbf{P} & \textbf{T} & \textbf{H} \\ \textbf{310} \\ \textbf{SGGACCTTACAAAAAATC } \\ \textbf{SGGACCTTACAAAAAACTCAACTC } \\ \textbf{SGGACCTTACAAAAAACTCAACTCC } \\ \textbf{SU} & \textbf{V} & \textbf{K} & \textbf{V} \\ \textbf{310} \end{array} $	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GCTATGGATTTTGGATCTTAT G P F G Y C N 1850 CGGATTCAGATTACTCGAAAAAG G P F G Y C N 1950 TGGATTCATTATTCATATATGAA F E L S L Y E 1000 TTCACTTTATTCATATATGAA F F I S Y P Y 2250 TTCACTTTATTCATATCCT H F I S Y P Y 2250 TTCACTTTATTCATATCCT	1570 1570 1570 $X \vee A I F C G$ 1670 AGTTATACCTANAGTG <u>GAN</u> Y M P K S G I 1770 A C G L Q Y V 1870 ATTAATAGAAGATCTAGC L I E D S S AACCTAGATATTATAATAG ATTAATAGAAGATCTAGC L I E D S S AACCTAGATATTATAATAGAAGGCCT Y F N G K G L 2170 ATCTAACATCCAAAATTTC L T S K I F 2270 TATAAGAAAGAAGGCCTTTGA I K E L L R 2370 I F K N H 2470	1330 TATGTAACAGGGATAA Y V R D K 1690 ITICTAACTAGAATCT I L E S 1790 ITAACTTAATCCAGG N L I E G 1890 ACTTITTGGGGAAA E F W G N 1990 ACTTITTTTTTTTTGG F F S 2090 AGCATTAGTGGGTGT R C S V L 2190 TTTTATATAATGGT F Y I M V GAGATAATCCGACA C N P T 2390 ITTTATGAGCAATAG L 2400
CATEGEAN W E 14 ATATAGAAT Y E I GATTATAT D L C 16 GCTCAGAA S E 16 TACAGATI T D S 20 CCAGAGAG R G D 21 TACATCCC H P 22 GATTACATCC GG Y P 24 CG Y P 24 CG Y P 24 CCTCAGAG	$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GGCTACGATTITGGATCTTAT 5 Y D F G S Y 1850 TGGGCCTTCGGATACTGCA G P F G Y C N 1950 CG F N L E K V 2050 C F N L E K V 2050 TTGAGCCTTCATTATATGAAA F L S L Y E 2150 TTCACTTTATTCCATATATGAA F E L S L Y E 2150 ATCAAAAGAGGCTTCCTTT F K R G F P F 2250 ATCAAAAGAGGCTTCCTTT F K R G F P F 2350 ATCAAAAGAGGCTTCCTTT F K R G F P F 2350 ATCAAAAGAGGGCTTCCTTT F K R G F P F 2350 ATCAAAAGAGGGCTTCCTTT F N G E Y 77777777777777777777777777777777777	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
CATEGOA W E ATATGAAT Y E I GATTTAT D L C I GATTATAT D L C I GCTCAGAA R G D CCGAGGAGY R C D CCGAGGAGY R C D CCGAGGAGY R C D CCGAGGAGY R C D CCGAGGAGY R C D C C C C C C C C C C C C C C C C C C	TARAATTTCTAGCAGTACT K I S S T 510 TTTAACTATCATGATAGAAA F N Y H D R T 710 XAAAGCTTGTGATGATGATGTTT K A C D V F 810 INGCATGATGATAGGTTACTT K I D R L L 910 GCAACCAAGTAAAAATATC 95 N Q V K Y H 101 CATAAGTGTTTTAAAATCG I S V L K S 210 CATAAGGGTTTACAAAAATCG I S V L K S 210 CAGGACCTGGAATCCAACTCC R D V N P T H 310 CAGGACCTTACAAAAAAGT E L Y K K V 410 CAGAGCTTTGGTATTCGTATTCG	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 LTARAGTAGCTATATTTTGGGT X V A I F C G 1670 AGTTATCCCTAAAAGTG <u>CAA</u> V M F K S G I 1770 X G L Q Y V 1870 ATTTAATAGAAGATTCTAGC 1870 ATTTAATAGAAGAATCTTGA N L D I I N N 2070 ATCTAACATCCAAAATTTC L T S K I F 2270 TTTTAACGGAAAAGCTTTGA I K K L L R 2370 AAACAAATCTTCAAAAATCT K K C I F K N H 2470 CCTCTCCCCACAGATTTTTT 2570	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
CATEGOAL WE 11 TYEI GATTTATAT DLC GATTCAT DLC CCACGAGA RGD 22 CCACGAGA CCACGAGA RGD 22 CCACGAGA RGD 22 CCACGAGA RGD 22 CCACGAGA CCACGA CCACGAGA CCACGAGA CCACGAGA CCACGAGA CCACGAGA CCACGA CCACGAGA CCACGA CCACGAGA CCACGA CCACGAGA CCACGA CCACGAGA CCACGA CCACGAGA CCACGACAC CCACGAC CCACGAC CCACGAC CCACGAC CCACGA CCACGA CCACGAC CCACGAC CCACGAC CCACGAC CCACGAC CCACGAC CCAC	$\begin{array}{cccc} \text{MAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{T} \\ \text{510} \\ \textbf{TTTAACTATCATGATGATGATAGAA } \\ \textbf{F} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ \textbf{10} \\ \textbf{TAAAGCTTGTGATGATGATGATTTT } \\ \textbf{K} & \textbf{A} & \textbf{C} & \textbf{D} & \textbf{V} & \textbf{F} \\ \textbf{310} \\ \textbf{300} \\ 3$	1330 TATCATATTAGAGAGATTTA ISI R E F N 1630 CTTATCCAATGGTTAAGAAA Y Q W R N 1730 IE R E N F C 1830 TTCGTGAACGATAGTTTTAT I E R E N F C 1930 ACTATCAGAGGTTACCTGTTT Y Q S Y L F 2030 ACTATCAGAGTTATCGAACTT S L V I E N F 2030 ATCTATCAGGTTACCGGTTGCAACTT Q L W L Q L 2300 ATCACTATGGTTGCAACTT Q L W L Q L 2300 ATCACTATGGTTGCAACTT	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 1570 $X \vee A I F C G$ 1670 AGTTATACCTAAAAGTG <u>GAA</u> Y M F S G I 1770 A C G L Q Y V 1870 A C G L Q Y V 1870 A C G L Q Y V 1870 AACTTAGAAGAATCTAGC I I F D S AACTTAGATATTATAAAAA N L D I I N N 2070 TATTTAATGGGAAAGGCCTT Y F N G K G L 2170 ATCTAACATCCAAAATTTTC L T S K I F 2270 TATAAAGAAAGAGCTTTGAAAATCTA K & L L R 2370 AAACAAATCTCAAAAATCTA K Q I F K N H 2070 CCTGGCCCACAGATTTTTT 250 TTCATCGCGAATTACACTT	1330 TATGTAAGAGGATAA Y V R D K 1690 ITCTAACTAGCATGGATAT I L E S 1790 ITAACTTAATCCAGG N L I E G 1890 CATTTTTGGGGAAA E F W G N 1990 ACTTTTTTTTTTTGGGGAAA E F W G N 1990 ACTTTTTTTTTGGGGAAA C F F S 2090 AGGATTATGGGGATATCGGACA C V L 2190 TTTTATGAGCATAGG L 2390 TTTTATGAGCAATAG L 2390 TTTATGAGCAATAG L 2590 TTCAACCGGATCTA 2590
CATGGGAI W E 14 ATATAGAAT Y E H GATTTATC D L C GCTCAGAM S E 17 CAGAGAT R G D CAGAGAC R G D T CAGAGAC R G D T A F I A F I A F I A 22 GGATACCC G Y P CCAGAGC 22 TTCTAGAAG 22 CGACACT 22 CGACACT 22 CGACACT 22 CGACACT 22 CGACACT 22 CGACACT 22 CCACCACT 22 CCACACT 22 CCACCACT 22 CCACCACT 22 CCACCACT 22 CCACCACT 22 CCACCACT 22 CCACCACT 22 CCACCACCT 22 CCACCACT 22 CCACCACT 22 CCACCACCACC 22 CCACCACCT 22 CCACCACCACC 22 CCACCACCCCC 22 CCACCACCACCCCCCCC	$\begin{array}{rrrr} \textbf{J}_{10} \textbf{J}_{$	$\begin{array}{cccccc} 1.5.5\\ 1.5.$	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GCTATGGATTTTGGATCTTAG GCTATGGATTTTGGATCTTAG I1850 TIGGCTCTTCGGATACTGCA G F F G Y C N 1950 CG F N L E K V 2050 TIGGATCTTCATTATATGAA F E L S L Y E 2150 TIGCACTTATTTCATATCCAT H F I S Y P Y 2250 ATCCATTTATTCATATCCAT H F I S Y P Y 2250 ATCCATTTAATGGAGGGTAT T F K R G F P F 2350 AATGATTTAATGGAGGGTAT N D F N G E Y <i>rminator</i>) TICCACTATTATATGGAGGGTAT TAGGGAGTATTTTATGGTAG	1570 1570 1570 X = X = X = C = G 1670 X = X = X = C = G 1670 X = X = X = C = G 1670 X = X = X = C = X = C = X = G 1770 X = C = C = Q = Y = V 1870 AC G L Q Y V AC G L Q Y V	$\begin{array}{ccccccc} 113 \\ 1590 \\ 1590 \\ 171 \\ 1590 \\ 171 \\ 1590 \\ 171 \\ 1590 \\ 1790 \\ 171 \\ 1590 \\ 171 $
CATEGOAL W E ATATGAAT Y E I GATTTAT D L C I GATTTAT D L C I GCTCAGAA S E I CGAGGAGY R G D CCGAGGAGY R C C CCGAGGAGY R C C CCGAGGAGY R C C CCGAGGAGY R C C CCGAGGAGY R C C CCGAGGAGY R C C C CCGAGGAGY R C C C CCGAGGAGY R C C C C CCGAGGAGY R C C C C C C C C C C C C C C C C C C C	$\begin{array}{ccccc} & \text{JAAATTTCTAGCAGTACT} \\ & \text{K I S S T} \\ & \text{510} \\ & \text{TTTAACTATCATGATAGAAA} \\ & \text{F N Y H D R T} \\ & \text{YAAAGCTTGTGATGATGATGATGATTTTT K A C D V F \\ & \text{510} \\ & \text{TAGAAGCTTGTGATGATGATTGTTAGTTTGATAGGTTTGATAGGTTTTAAAATATCAGATGATAAAATATC \\ & \text{F N Y T D D K } \\ & \text{510} \\ & \text{CACACGTGATACAAAAATATC } \\ & \text{J S N Q V K Y H } \\ & \text{J 10} \\ & \text{CACAGTGATATATACAGATGATAAA } \\ & \text{I Y T D D K } \\ & \text{J 10} \\ & \text{CACAGTGATATAAAAATCC } \\ & \text{J V N P T H } \\ & \text{J 10} \\ & \text{CACGGACCTGGAATCCAACTCC } \\ & \text{R U V N P T H } \\ & \text{J 10} \\ & \text{CAGAGCTTTACAAAAAATC } \\ & \text{J L Y K V V } \\ & \text{J 10} \\ & \text{CACAGCTTTGGATATGGATATCT } \\ & \text{J 10} \\ \end{array}$	$\begin{array}{cccccc} 1333\\ 1330\\ 1330\\ CTTATCCATGGTTAAGAAAY Q W L R N1730\\ 1300\\ 1000\\ 1300\\ 1300\\ 1300\\ 1300\\ 1300\\ 1300\\ 1300$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 1570 $X \vee X I F C G$ 1670 AGTTATCCCTAAAAGTG <u>CA</u> Y M F K G I 1770 $X \oplus K S G I$ 1770 1770 1770 $X \oplus C G L Q Y V$ 1870 ATTTAATAGAAGATTCTAGC 1970 AAACTTAGATATTATAATAA N L D I I N N 2070 ATCTAACATCCAAAATTTC L T S K I F 2270 TATTAACAAGGGATAAATTTC 2370 AAACAAATCTTCAAAAATCT GCTCTGCCCACAGATTTTT 2570 TCATCCGTAACATGCGGAAGG	$\begin{array}{cccc} 13 & 3 \\ 13 & 3 \\ 15$
CATEGOAL WE 11 ATATAGAAT YE II GATTTAT D L C GATTCAGAT S L CCARGAGI R G D CCARGAGI R G D CCARGAGI R G D CCARGAGI R G D CCARGAGI CCARGAGI R G D CCARGAGI CCARGAGI R G D CCARGAGI CCARGAG	$\begin{array}{cccc} & \text{JAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{T} \\ \hline \textbf{50} \\ \textbf{TTTAACTATCATGATGATGATAGAA } \\ \textbf{F} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ \hline \textbf{70} & \textbf{70} & \textbf{70} & \textbf{70} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{31} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{310} & \textbf{30} & \textbf{30} \\ \hline \textbf{30} & \textbf{30} & \textbf{30} & \textbf{30} \\ \hline \textbf{310} & \textbf{310} & \textbf{310} \\ \hline \textbf{310} & \textbf{310} & \textbf{310} \\ \hline \textbf{310} & \textbf{310} & \textbf{310} \\ \hline \textbf{310} & \textbf{310} & 310$	1330 TCTCTATTAGAGAGATTTA ISI R E F N 1330 CTTATCCAATGGTTAAGAAA Y Q W R N 1730 IE R E N F C TATTCGAAAGGAAAGATTTTA IE R E N F C 1930 TCGTGAACGATAGCTTTAT Y Q S Y L F 2030 ACTATCAGAGTTACCTGTTT Y Q S Y L F 2030 ACTATCAGAGTTATCGAAAAATT S L V I E N F 2030 ATCTACTAGGTTATCGACTTC Q L W L Q L 2330 TCTACACTATGGTTGGAACTT Q L W L Q L 2330 CT F V M G S T 2330 CT F V M G S T 2330 CT CTCCCCCCCCCAAAGT2 2530 TCTTCCCCCCCCCCAAAGT2 2530 TCTTCCCCCCCCCCAAAGT2 2730 ACTACCTTGGAAAACT	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GCTAGGATTT <u>GGATACTGCA</u> G F N G S Y 1850 TGGCCCTTCGGATACTGCA G F N G S Y 1950 CGATTTAATCGAAAAAG G F N L E K V 2050 TCGACTTCATTATATATATA F E L S L Y E 2150 TCGACTTCATTATTCATATATGAA F E L S L Y E 2150 TCGACTTCATTATTCATATATGAA F F I S Y P Y 2250 ANGGACTATTATTCATATTCTTC T K R G F P F 2350 TAGGCATATTATATTTTAGT 2550 TAGGCATATTATTTTATGGAGAGATAT N D F N G E Y TCGACCTGTTGTTATTGCAT AGGCATATTTTTTGTCGAT 2550 TAGGCATATTTTTTTTGCATATATTTTGC 2550 TAGGCAGTATTTTTTTGCGAT 2550 TAGGCAGTATTTTTTGCGAT 2550 TAGGCAGTATTTTTTGCATA 2550 TAGGCAGTATTTTTTGCATA	1570 1570 X = X + X + X + X + X + X + X + X + X +	1330 TATGTAAGAGGATAA Y V R D K 1690 ITCTAACTAGCATGCA I L E S 1790 ITAACTTAATCCAGG N L I E G 1890 GAATTTTGGGGAAA E F W G N 1990 ACTTTTTTTTTTTAGC AGGATGTAGTGGTGT R C S V L 2190 TTTTATGTAGTGGTGAT C S V L 2190 TTTTATGTAGCGACA C D N F T 2390 ATATAATATTTTA 2390 ATATAAATATTTTA 2390 ATATAAATATTTTA 2590 ATATAAATATTTTA 2590 ATATAAATATTTTA 2590 ATATAAATATTTTA 2590 ATATAAATATTTTA 2590 ATATAAATATTTTA 2590 ATATAAATATTTTA 2590
CATEGRAF W E 14 ATATAGAAT Y E I GATTTATAT D L C GCTCAGAAT S E 17 ACAGGAT R G D 20 CCAGGAGA R G D 21 TACATCCC H P 22 CCAGGAGA 22 GATGCACA 22 GATGCACA 22 GATGCACA 22 GGTACCCC 22 GGTACCCC 22 GGTACCCC 22 GGTACCCC 22 GGTCGCT 21 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGATATT 22 CCGCGT 22 CCGCGT 22 CCGATATT 22 CCGATATT 22 CCGCGT 22 CCGATATT 22 CCGATATT 22 CCGCCGT 22 CCGCCGCT 22 CCGCCGT 22 CCGCCGT 22 CCGCCGCCCCCCCCCC	$\begin{array}{cccc} \text{MAAATTTCTAGCAGTACT} \\ \textbf{K} & \textbf{I} & \textbf{S} & \textbf{S} & \textbf{T} \\ \text{510} \\ \textbf{TTTAACTATCATGATAGAAAA \\ \textbf{F} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ \textbf{7} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ \textbf{7} & \textbf{N} & \textbf{M} & \textbf{C} & \textbf{C} & \textbf{N} \\ \textbf{7} & \textbf{N} & \textbf{H} & \textbf{D} & \textbf{R} & \textbf{T} \\ \textbf{7} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{2} \\ \textbf{8} & \textbf{A} & \textbf{C} & \textbf{D} & \textbf{V} & \textbf{F} \\ \textbf{8} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{2} \\ \textbf{9} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{9} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{9} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{9} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{9} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{9} & \textbf{10} & \textbf{R} & \textbf{L} & \textbf{L} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{10} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf{R} & \textbf{R} & \textbf{R} \\ \textbf{R} & \textbf$	$\begin{array}{cccc} 1330\\ 13$	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCATT N D Y V I L 1750 GCTATGGATTTIGGATCTTAT G P F G Y C N 1850 CGGATTCAGATTACTCGAAAAAG G P F G Y C N 1950 TCGACTTCATTATATGAAAAG F E L S L Y E 100 TTCACTTTATTCATATATGAA F E L S L Y E 2150 TCCACTTTATTCATATATGAA F F I S Y P Y 2250 TTCACATTTAATGGAGGAGTAT N D F N G E Y rminator) TCCACTATTATATGGAGGATAT N D F N G E Y rminator) TCCACTATTATATGTTATATTTTAGT 2550 AGGCTGTGTTGTTATTGATATAT 2550 TAGGAGATTTTATGGCAGGATAT N D F N G E Y rminator) TCCACTATTATATTTTTTGTCGAT 2550 AGGCTGTGTTGTTATTGATATAT 2550 TAGGAGATTTTATGCCGCACAAAA 2850	1570 1570	$\begin{array}{cccccc} 13703 \\ 1490 \\ 1400 \\ 140$
CATEGOAL W E ATATGAAT Y E I GATTTATUT D L C I GATTTATUT D L C I GCTCAGAA S E I CCGAGGAGY R G 21 TACACATCC H P CCGAGGAGY R G 21 TACACATCC H P CCGAGGAGY I A I S GGTTACCAC G Y P CCGAGGAGY I CCGACATCC CCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{ccccc} & \text{JAAATTTCTAGCAGTACT} \\ & \text{K I S S T} \\ & \text{50} \\ & \text{TTTAACTATCATGATAGAAA} \\ & \text{F N Y H D R T} \\ & \text{YAAAGCTTGTGATGATGATGATGATTTTT \\ & \text{K A C D V F} \\ & \text{510} \\ & \text{TTTACTGTGATGATGATGATGATTTTT \\ & \text{K A C D V F} \\ & \text{510} \\ & \text{TT D R L L} \\ & \text{510} \\ & \text{TT D R L L} \\ & \text{510} \\ & \text{TT D R L L} \\ & \text{510} \\ & \text{TT D R L L} \\ & \text{510} \\ & \text{TT D R L L} \\ & \text{510} \\ & \text{TT C ATATGATGATGATATTCT} \\ & \text{T D R L L} \\ & \text{510} \\ & \text{TT TATATATACAGATGATATAT \\ & \text{1 Y T D D K} \\ & \text{510} \\ & \text{CATAGTGTTTTAAAAATCG} \\ & \text{S V V N P T H} \\ & \text{510} \\ & \text{CAGCACCTGGAATCCAACTCC} \\ & \text{R D V N P T H} \\ & \text{510} \\ & \text{CAGCACCTTGGATATGGTATTTGGTATTGGTATTCT} \\ & \text{510} \\ & \text{CTCCCACGTTGATATGGTATTCTGC} \\ & \text{710} \\ & \text{TTTTGATTGGTTATGGTATGCTA} \\ & \text{510} \\ & \text{TTTTTTTTTTTGATTAGCTAGGTATATTCGGTATTCTGGTATGGTATTCGGAGATGTTAAGCATGTTAAGCATGTTAAGCATGTTAAGCTA} \\ & \text{510} \\ & \text{TTTTTTTTTTTGATTAGATCAGGTATATTCTGGTATTCTGATATATCGGTATTCTGGTATGGTATTCGGTATTCGGTATTCGCATGTTAAGCATGTTAAGCTA} \\ & \text{510} \\ & \text{TTTTTTTTTTTTTTTTTGATCAGGGTATTTTTTTTTGATCAGGGATGTTAAGCATGTTAAGCATGTTAAGCATGTTAAGCATGTTAAGCTA} \\ & \text{510} \\ & TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT$	$\begin{array}{cccccc} 1333\\ 1330\\ $	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGACTATACT N D Y V I L 1750 GCTACGATTTAGGATCTTAG GCTACGATTTAGGATCTTAG 1850 ITGGCTATCGGATACTGCA G P F G S Y 1850 CGCATCGATTATGGATACTGCA G N L E K V 2050 CGCATCATTATTCGATAAAGAT G N L E K V 2050 TTCGAGCTTTCATTATATGAAA F E L S L Y E 2150 TTCCACTTTATTCATTCATTATATGAA F E L S L Y E 2250 AATGATTTAATGGAGGGCTTTCCTTT F K R G F P F 2350 AATGATTTAATGGAGGGGTAT T F N G E Y TTCCACCTATTATATTTTAGT 2550 TTCGGGGTATTGTTATTGTCGAT 2550 TTGGGGGTATCGCGCGCCCAAAA 2750 TTGGGGTATCGCGCCGCACAAA 2850 TTGGGCTATCGCGCCGCACAAA	1570 1570 (V A I F C G) 1670 AGTAATGCCTAAAAGTGGAA V M P K S G I 1770 170	$\begin{array}{ccccc} 1390\\ Y & R & D & K\\ 1690\\ YTCTACTAGCAGAATAA\\ Y & V & R & D & K\\ 1690\\ YTCTACTAGCAGAATACACAGAATACACAGAATAATACAGAATATTTGGGGGAATAATAATAGTGTGTGGGGAATAAT$
CATEGOAL WE 11 ATATAGAAT YE I GATTTAT D L C GATTAT C C GATTAGAT T D C CCAAGAAT T D C CCAAGAAT C CCAAGAAT C CCAAGAAT C CCAAGAAT C CCAAGAAT C C CCAAGAAT C C CCAAGAAT C C CCAAGAAT C C CCAAGAAT C C C CCAAGAAT C C C CCAAGAAT C C C C C C C C C C C C C C C C C C	$\begin{array}{cccc} & \text{MAAATTTCTAGCAGTACT} \\ & \text{K I S S T} \\ & \text{50} \\ & \text{TTTAACTATCATGATGATAGAA} \\ & \text{F N Y H D R T} \\ & \text{7 N Y H D R T} \\ & 7 N Y H D R T T T T T T T T T T T T T T T T T T$	$\begin{array}{cccccc} 1333\\ 1330\\ $	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCTT N D Y V I L 1750 GCTAAGACTATTACGACTTATG GCTAAGATTTGGACTCTTAT 3 Y D F G S Y 1850 TIGGCCCTTCCGATAACTGCA G F N L E K V 2050 TCGACTTCATTATATGAAAAAGA F E L S L Y E 150 TCGACTTCATTATATGAAAAGAGT H F I S Y P Y 2250 ATCCAAAAGAGGCCTTCCTTT F K R G F P F 2350 TAGGAGGATATTATATGAGAGAGTAT N D F N G E Y 2550 TAGGAGGATATTATATGTCAGT AGGGGTATCGCGTCGGCACAAA 2850 TATGAGCTGTGTTATTGGATAAT 2750 TGGCTATCGCGCACGGCCGAACT 2850	1570 1570 1570 X = X = X = C = G 1670 X = X = X = C = G 1670 X = X = X = C = G 1770 X = X = C = G 1770 X = C = G = Q = Y = V 1770 X = C = G = Q = Y = V 1870 ANTITATAGAGAATATATAAAAA A C G L Q Y V 1870 ANTITATAGAGAATATATAAAAAA A C G L Q Y V 1870 AACTAGATATATAGAAGAATCTAGC L I E D S S 1970 AACTAGATATATATAGAAAAGAGCTTTGA ATTTAAAGAAAAGAGCTTTGA TATATAAAGAAAAGAGCTTTGA TATAAAGAAAAGAGCTTTGA 1 K K E L L R AAACAAAATCTCAAAAATCA K Q I F K N H 2370 GCTCTGCCCACAGATTATCAT 2570 TCCACGGAATAACCGAAAAACCT 2570 TCCACGGAATAAGCGAAAGACCA 2770 ACATAGCAATAACCGGATGGAAA 2970 ACATAGCAATAACCGGTTGGAGA 2970	1330 TATGTAAGAGGATAA Y V R D K 1690 ITCTAACTAGCATAGATGA I L E S 1790 ITAACTTAATCCAGG N L I E G 1890 GAATTTTGGGGAAA E F W G N 1990 ACTTTTTTTTTAGC AGGATGTAGTGGTGT R C S V L 2190 TTTTATAAATGGTG TTTTATATAATGGT F Y I M V 2290 GAGATAATCCGACA C D N F T 2390 CTACCCGATCAT 2590 ACTACCGACATAATGAGA 2590 ACTACCGGATATTC 2790 AATGGTAATGAGAAATGAGA 2890 ACTACCGGATATTGGTCGA 2900 ACTACCGGATATTGGTCGA 2900 ACTACCGGATATTGGTCGA 2900 ACTACCGGATATTGGTCGA 2900 CTACCGGATATTGGTCGA 2900 CTACCGGATATTGGTCGA 2900 CTACCGGATATTGGTCGA 2900
CATEGEAL WE 14 ATATAGAAT YE I GATTTAT D L C GATTTAT D L C GCTCAGAM S I I TACAGATI T D S CCACGAGATI R G D I CCACGAGAT CCACGAGA C GTTACATCCC H P 22 GGATTACCC G Y P 21 CCACGAGAG 22 GGTTACCCC G Y P 21 CCACGAGAG 22 GTCCACATCCC G Y C CCACGACCC 23 GTCCACATCCC 23 GTCCACATCCC 23 GTCCACATCCCC 23 GTCCACATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{cccc} & \text{MAAATTTCTAGCAGTACT} \\ & \text{K I S S T} \\ & \text{50} \\ & \text{TTTAACTATCATGATAGAAA} \\ & \text{F N Y H D R T} \\ & \text{TAAAACTTGTGATGATGATGATGATTTT } \\ & \text{K A C D V F} \\ & \text{SAAAGCTTGTGATGATGATGATTGATAGGTTACTT} \\ & \text{K A C D V F} \\ & \text{SAGATTGATAGGTTACTT} \\ & \text{R I D R L L} \\ & \text{S10} \\ & \text{GCAACCAGTGAAAAATATC} \\ & \text{S N Q V K Y H} \\ & \text{S10} \\ & \text{GCAACCAGTGAAAAATATC} \\ & \text{S N Q V K Y H} \\ & \text{S10} \\ & \text{CACAGCGAGTGAAACAATAAAAGT } \\ & \text{S V L K S} \\ & \text{S10} \\ & \text{CGCAGCGTGAATCCAAAAAAGT } \\ & \text{E L Y K V V} \\ & \text{S10} \\ & \text{CACAGCATTTCAAAAAAGT C } \\ & \text{S10} \\ & \text{CACAGCGTTACCAAAAAAGT } \\ & \text{S10} \\ & \text{CACAGCCATTTCGAATCCAAACACT C } \\ & \text{S210} \\ & \text{CGCGAGCTTTACAAAAAAGT } \\ & \text{S10} \\ & \text{CACAGCATTTGATTGGTATTCT } \\ & \text{S10} \\ & \text{CTGCAGCATTTGATTGGTATTCGC } \\ & \text{S10} \\ & \text{CTGCCACGATTACTATATGACTA } \\ & \text{S10} \\ & \text{CTCCCCAGTTACTATTGGCTA } \\ & \text{S10} \\ & \text{CTCCCCAGTTACTATTCGACTA } \\ & \text{S10} \\ & \text{TTTTGAGCCAAGGTTTACAAGGTA } \\ & \text{S10} \\ & \text{CTCCCCAGTTACTATTCGACTA } \\ & \text{S10} \\ & \text{CTCCCCCAGTTACTATTCGACTA } \\ & \text{S10} \\ & \text{CTCCCCCAGTTACTAGGATTACAGG } \\ & \text{S10} \\ & \text{CTCCCCCAGTTACAAGGTTTACAAGGTT } \\ & \text{S10} \\ & \text{CTCCCCCAGTTACAAGGTTTACAAGGT } \\ & \text{S10} \\ & \text{CTCCCCCCCCCCCCCCCATTACGAAGTTTACAGGG } \\ & \text{S10} \\ & CTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{cccc} 1330\\ 13$	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCATT N D Y V I L 1750 GCTACGATTTGGATCTTAG GCTACGATTTGGATCTTAG GCTACGATTTGGATCTTAG ISS0 CGTATTGGACCTTCGGATACTGCA G F N G S Y 1850 G F N L E K V 2050 TGGATTTAATCTCGAAAAAGT F E L S L Y E 17CAGCTTCATTATATATAA F E L S L Y E 17CAGCTTCATTATATATATATA H F I S Y P Y 2250 MTCAAAAGAGGCTTCCTTT H F I S Y P Y 2250 TTCAACTTTATTCATATCTCAT H F I S Y P Y 2250 TTCAACTTTATTCATATTCTCAT H F I S Y P Y 2250 TTCGACTATTATATGAGAGGTAT N D F N G E Y TTGGGCTATTATATTTTAGT 250 TTCGGCTATTATATTTTAGT 250 TTCGGCTATCGCGTCGGCACAAA 2850 STATTGGAGAGGTATATAAAGTTT 3050	1570 1570 1570 X = X = X = C = G 1670 X = X = X = C = G 1670 X = X = X = C = G 1670 X = X = C = X = C = X = G X = X = X = G = G = X = G 1770 1770 AC G L Q Y V 1870 AC G L Q Y V AC C AC AC G C C C C C C C C C C C C C C	1330 14300 Y R 1690 ITTATGATAGGATAA Y R 1690 IL L TACTTACTAGAATCE N L E F F F 2090 AAGATGTAGTGGTGT R C Y I 2090 GAGATATCCGACA D P 2190 TTTATATAATGGGTAT TTTTATGTAGGAATAATCCGACA D D P 2390 GAGATAATCCGACA C N 2390 TATATAAATATTATA 2430 ATATAAATATTATA 2430 ATATAAATATTATA 2430 AATATGAACCGAATAT 2430 AATATAAATATTATA 260 AATATAAATATTATA 2800 GTGCATCAAGGAAAT 2830 GTGCATTTTGGCCAA 290 CAGAATATCCCATATATCCTA 3090 SAGGATCAATCACCCGATCA
CATEGOAL W E 1 ATATGAAT Y E 1 GATTTATUT D L C 1 GATTTATUT D L C 1 CGACGAGAG S E 1 CCACAGAGA T D 2 CCACGAGAG T CCACA T CCACAGAG T CCACAGAG T CCACAGAG T CCACAGAG T CCACAGAG T CCACAGAGAG T CCACAGAGAGAG T CCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	$\begin{array}{ccccc} & & & & & & & & & & & & & & & & &$	13.30 TRCTCTATTAGAGAGATTTA I S I R E F N 130 CTTATCAATGGTTAAGAAA Y Q W L R N 1730 I E R E N F C 1830 TRCGRGAACGATAGTTTTAT F V N D S F I 1930 CTTCCTAGGATTACTGATTACTGTT S L V I E N F 2030 ACTACCAGACTTACTGATAGATAC S L V I E N F 2130 ACTACCAGCATTATTGAAAATT S L V I E N F 2230 ACTCCTAGTTATGGTGCAACTT Q L W L Q L 2300 ACGAGGAGATTATGGTGCAACTT C L W L Q L 2300 ACGAGGAGATTATGGTGCAACTT C L W L Q L 2300 ACGAGGAGTTATCCGCTGGAAACT 2300 ACTCCTCTGGTGCAAACT3 2300 ACTCCTCTGGTGCAAACT3 2300 ACTCCTCTGGTGCAAACT3 2300 ACTCCTCTGGTGCAAACT3 2300 ACTCCTCTGGGTTTCCAGGTCC 2300 ACTCCTCTGGGGTTCCAAGTCC 3030 TCTTGTATTACCTCCGGGTTCCCAGTCC	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 1570	1330 TATGTAAGAGGATAA Y V R D K 1690 ITTCTACTAGAATCT I L E S 1790 ITAACTTAATCAGAATCT I L E S 1790 GAATTATAATCAGGAA R C I E G 1890 ACTITITTGGGGAAA C F M G N 1990 ACTITITTTTTTTACC C S V L 2090 AGAGATGTAGTGTGTGT R C S V L 2190 TTTTATATAATGGTGT TTTTAGAGCAATAG C S V L 2190 TTTTAGAGCAATATTATA 2390 ACAATATAAATATTATA 2490 ATATAAAATATTATAT 2590 TCGACCCGATCTAT 2790 ACTACAGGATATTGGTGA 2890 GTGATTTTGGTGAAGA
CATEGOAL WE II ATATGAAT YE II GATTTATUT DL C GATTTATUT DL C CGACGAT SE SE SE SE SE SE SE SE SE SE SE SE SE	$\begin{array}{cccc} & \text{JAAATTTCTAGCAGTACT} \\ & \text{K I S S T} \\ & \text{510} \\ & \text{TTTAACTATCATGATGATAGAA} \\ & \text{F N Y H D R T} \\ & \text{JAAAGCTTGTGATGATGATGATGATTTT } \\ & \text{K A C D V F} \\ & \text{J10} \\ & \text{SCAAAGCTTGTGATGATGATGATTTGTA } \\ & \text{J10} \\ & \text{R L D R L L} \\ & \text{J10} \\ & \text{J10} \\ & \text{R L L} \\ & \text{J10} \\ & \text{SCAAAGCTAGTAAAATATC} \\ & \text{GCAACCAAGTAAAAATATC} \\ & \text{GCAACCAAGTAAAAATATC} \\ & \text{GCAACCAAGTAAAAATATC} \\ & \text{J210} \\ & \text{GCAACCAAGTAAAAATATC} \\ & \text{J210} \\ & \text{CACAAGTAAAAAATATC} \\ & \text{J210} \\ & \text{CACAAGTAAAAAATATC} \\ & \text{J210} \\ & \text{GCGGACGTGAATCCAACTC} \\ & \text{K D V N P T H} \\ & \text{J10} \\ & \text{CACAAGTTTTACAAAAAAGT } \\ & \text{L Y K K V} \\ & \text{J10} \\ & \text{GCGGACGTTAACAAAAAAGT } \\ & \text{CAGGCTTTACAAAAAAAGT } \\ & \text{CAGGCTTTACAAAAAAAGT } \\ & \text{CAGGCTTTACTAATACGAATATCGAATACCGAATTCGACATCCAAGTATATTGGATATCT } \\ & \text{J10} \\ & \text{TCTCCACGCATTTACATATCGCTA} \\ & \text{J10} \\ & \text{TTTTAGGGCCAAGGTTTTAA} \\ & \text{J10} \\ & \text{TCTCCCCAGTTACTATCGCT} \\ & \text{J10} \\ & \text{TCTGCCAACCCGATAGAAGCTA} \\ & \text{J10} \\ & \text{TCTACCCGATAGAAACAGAGT } \\ & \text{TCTACCCGATACCAAGAAGCT } \\ & \text{J10} \\ & \text{TCTACCCGATAGAAACAGAGGTTTAACAGAGACT } \\ & \text{J10} \\ & \text{TCTACCCGATATGTAGAAACAGAGCT } \\ & \text{J10} \\ & \text{TCTACCCGATAGAAACCAAGCAAGAGCT } \\ & \text{J10} \\ & \text{TCTACCCGATAGAAACAGAGCT } \\ & \text{J10} \\ & \text{TCTACCCGATAGAAACAGAGCT } \\ & \text{TTTATTGTATTGTAGTAGAAACAGAGCT } \\ & \text{J10} \\ & \text{TCTACCCGATACCAAGAAGCAAGACT } \\ & \text{TCTACCCGATACCAAGAAGCACCAACCCAAGAAGCAACCAAGACT } \\ & TCTACCCCAACTACCAAGAAGCACCAACCCAACCCAACC$	$\begin{array}{cccccc} 1330\\ $	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	1570 1570 1570 X = X = Y = C = G 1670 X = Y = X = C = G 1770 Y = Y = X = G = G 1770 Y = Y = X = G = G 1770 X = C = G = Q = Y = Y 1870 A = C = G = Q = Y = Y 1870 A = C = G = Q = Y = Y 1870 A = C = G = Q = Y = Y 1870 A = C = G = Q = Y = Y 1970 ADACTTAGATATTATAAAAAAAAAAAAAAAAAAAAAAAAA	1330 TATGTAAGAGGATAA Y V R D K 1690 ITCTAACTAGCATGCA I L E S 1790 ITAACTTAATCCAGG N L I E G 1890 GAATTTTGGGGAAA E F W G N 1990 ACTTTTTTTTAGC F F F S 2090 ACTATTTTGGGGAAA E F W G N 1990 ACTATTTTAGCA AGCATGTAGTGGTGT R C S V L 2190 TTTTATAAATGGTGT TTTATGACGACATAG L 2390 ACGATAATCCGACA C N P T 2390 ACGATAATCCGACA L 2490 ATATAAATATTTAT 2590 TTCCAACGGATATCT 2790 CAGCATAGGATATCCGACA 290 CAGCATATTGGTCAAGGATAGGCA 290 CAGCATATTGGTCAAGGATAGCA 290 CAGCATATATCCTA 3090 CAGCATGACTCGTT 3190
CATEGEAL WE 11 ATATAGAAT YE II GATTTATC D L C GATTACAGATI T D C CCACGAGATI T D C CCACGAGAT R G D CCACGAGAT R G D CCACGAGAT R G D 22 CCACGAGAC R G D 22 CCACGAGAC R C D 22 CCACGACAC C Y P CCACGACAC C Y P CCACGACAC C Y C CCACGACAC C C CCACGACAC C C CCACGACAC C C CCACGAC C C C C C C C C C C C C C C C C C C C	ARAATTTCTAGCAGTACT K I S S T 50 TTTAACTATCATGATAGAA F N Y H D R T 710 SCAAAGCTTGTGATGATGATTT K A C D V F 310 SCAAAGCTAGTAAGGTTACTT R I D R L L 310 SCAACCAAGTAAAAATATC S N Q V K Y H 100 SCAACCAAGTAAAAATATC S N Q V K Y H 110 SCAACGAGTAAAAAATATC 210 SCAACGAGTAATACAGATGATAA I Y T D D K 10 SCAACGAGTAATACAGATGATAA I Y T D D K 10 SCGACGTGAATCCAACTC R D V N P T H 310 SCGACGTTACAAAAAAATT E L Y K K V 410 ITAATTTGATTTGGTATCTATTCGC 10 SCGACCTTACAAAAAAAGT E L Y K K V 410 ITAATTGATTTGGTATCTATTCGC 10 SCGCACCCAGTTACAAAAAAGTT 310 ITTTAGGCATTCGATATCGCATTTTAA 310 ITTTTAGGGCCAAGGTTTTAAAAGCTA 310 ITTTAGGGCCAAGGTTTTA 310 ITTTGGTCCCGATACGAGAGCT 110 310 ITTTAGGGCCAAGGTTTTTAAAAGCAG 310 ITTTAGGGCCAAGGTTTTTAAAAGCAGT 110 310 ITTTAGGGCCAAGCTTTTTAAAAGCAGT 110	1330 TATCATATTAGAGAGATTTA ISI R E F N 1330 CTTATCCATGGTTAGGAAA Y Q W R N 1730 IE R E N F C 1830 TATTGAAAGAGAAAGATTTTA IE R E N F C 1830 ACTATCAGAGGTTACCTGTTT Y Q S V F 2030 ACTATCAGAGTTACCTGTTT Y Q S V F 2030 ACTATCAGAGTTACCGGTTGCAACTT Q L V I E N F 2130 GATTTTATTAAACAGACTTTGAAAATT S L V I E N F 2230 ATCACTAGGTTGCAACTT Q L V Q T F 2230 ACTACTGGTTGGAACAT Q L V Q C S 230 ACTACTCAGGTTAGGGTGGCAACTT Q L V Q C S 230 ACTACCCTGGTTGGAACAT Q L V G S T 230 ACTACCCTGGTTGCAACTTGGAAACT 250 TCTCCCCCCCCCCAAGGTJ 250 CTTCCCTTGGATGAAACT 230 ACTCATCCTTGGATATCCTTGGAAACT 230 ACTCATCCTTGGATTACCCAGTCC 3130 GGGGTTTTAGCAGCTAGGATAC	RIGHTITGAGAATAAGAAAA R F E N K K K 1650 CAATGACTATTACGTCATCATT N D Y V I L 1750 GCTAGGATTTGGATCTTAG GCTAGGATTTGGATCTTAG GCTAGGATTTGGATACTGCA G F N G S Y 1850 G F N G Y C N 1950 GGATTAATCTCGAAAAAGAA F E L S L Y E 1950 TCGACTTCATTATATATAGAA F E L S L Y E 2050 TCCACTTATTCATATATATAG F N L E K V 2050 TCCACTTATTCATATATATATA F I S Y P Y 2550 ATAGGAGATTATATTTTTGG T K R G F P F 2350 ATAGGAGATTATATTTTTGG TCGACTATTATATGGAGAGATAT N D F N G E Y 2550 ATAGGAGATTATATATTTTAGT 2550 ATAGGAGATCTGGTATAGAAAAATA 2850 ATAGGAGATCTGGTAAAAAGTT 3050 SAATGGCAGATATGAAAAAATA 3150 TTGGGGAATCTGATAAAGTT 3250	1570 1570 V A I F C G 1670 AGTTATGCCTATAAGTGGAA V M F S G I 1770 A C G L Q Y V 1870 A C G L Q Y V 1870 AACTTAGATATTACAAGATCTAGC L I E D S S 1970 AACTTAGATATTATAGAGATCTAGC I L I E D S S 1970 AACTTAGATATTATAGAGATCTAGC I L I E D S S 1970 AACATTAGGATATTATAAATA N L D I I N N 2070 TATTATAAGGAAAGGGCCTT Y F N G K G L 2170 ATCTAACATCCAAAATTCTC L T S K I F 2370 AAACAAATCTTCAAAATCTCAAAATTCT 2570 GCTCTGCCCACAGATTTTTT 2670 GCTACTGCCCACAGATTATCACTT 2670 GCTACTGCCCACAGTATTATCACTT 2670 17CATCAGCAATTAGCAAATCCG 2870 17CGTATAACCAATGCGGAAGA 2970 3070 17CTTGATCCGTCAGTGGTGGAGA 3070 17CTTGATCCGTCAGTAGTAGTAGTGG 3070 17CTTGATCCGTATTAGTGATTAGGAGAACTGCGAACACCGGTGGTGGTGG 3270 170 170 170 170 170 170 170 1	1330 14300 1690 V R D K 1690 VICTACTAGCAGATAA 1690 IL L E S 1790 TAACTTAATCEAGG N L I E G 1890 GAATTATCGGGGAAA E F W G N 1990 ACTTTTTTTTTTAGC C F F S 2090 AAGATGTAGTGGTA C S V L 2190 TTTTATATAATGGTAT C S V L 2190 GAGATTATCGGACAA D N F T 2390 TTTAACTAAGGATAATCGGACAA L S NO L S NO AGGATAATCCGACAA L AGGATAATCCGACAA L AGGATAATCGGATAATATATATA 2590 TTCAACCCGATCTA 2590 ACCAACAGGATAATCCGACAA 290 CAGAATAATCAGTAATAAGAA 290 CAGAATAATCCGATATATATATA 100 290 CAGAATAATCCGATATATCCTAT

FIG. 1. Nucleotide sequence of the coding region for virB and virA. The sequence shown includes a 3.3-kb KpnI fragment. The deduced amino acid sequences of virB and virA correspond to proteins with molecular weights of 42,000 and 36,000, respectively. The possible translational starts of virB and virA, the potential promotor sequences (-35 and -10), the suggested ribosome-binding sites (Shine-Dalgarno [SD]), and the terminators are indicated. The DNA fragments that were cloned into vector pNQ705 for generating the insertion mutants NQ104 (virA), NQ706 (virB), and NQ107 (virB) are underlined. The nucleotide sequence of the 3.3-kb fragment has been submitted to GenBank and assigned accession number L08012.

pSKC5 and by uncertainties of plasmid stability within NQ706 (*virB*) during the fish infections. The presence of plasmid constructs in the wt 775.17B slightly reduced the virulence. Thus, these results suggest that both *virA* and *virB* appear to be essential for the virulence of *V. anguillarum*.

Insertional mutagenesis of a clinical isolate of *V. anguillarum.* To investigate whether the *virB-virA* region of the wt strain 775.17B could be found in other isolates of *V. anguillarum*, a PCR analysis was performed with *virB-* and *virA-*

specific primers. The strains tested were all of serotype O1. Fragments, which corresponded to the expected sizes of virA and virB, were found in all tested strains (Fig. 3).

The observation that the *virB-virA* region seemed to be conserved in strains of serotype O1 led to the question of whether *virA* and *virB* were important for virulence in those strains as well. We therefore studied a clinical isolate of serotype O1, NB10. For insertional mutagenesis, a 350-bp internal *virA* fragment of 775.17B (underlined in Fig. 1) was



FIG. 2. Schematic illustration of a 2.4-kb *KpnI-XbaI* fragment harboring *virB* and *virA*. Important restriction sites and insertion sites of the recombinant plasmids pNQ7051 and pNQ7056 and of the transposon Tn5-132 are marked. Plasmid derivatives of pSKT1 and pSKC1 are also shown. Deletions in pSKT4/pSKC5 and pSKT9 are indicated by shaded boxes. Note that the 1.4-kb kanamycin gene block in pSKT42 is out of scale.

subcloned into suicide plasmid pNQ705. The resulting plasmid, designated pNQ7051, and also pNQ7056 (described above) were mobilized into the wt strain NB10 by conjugation. Insertion of the plasmids into the putative virA and virB genes was confirmed by PCR (Fig. 4). Insertion mutants in virA (named NQ104) and virB (named NQ107) which had the mobilized suicide plasmids incorporated into virA and virB on the chromosome of NB10 were isolated. The insertion site of pNQ7056 in NQ107 was identical to the insertion in NQ706 (Fig. 4). The results were also confirmed by Southern hybridization of XbaI-digested chromosomal DNA of the isolated mutants (data not shown). A third exconjugant (NQ105) with an insertion site of pNQ7051 outside of virA and virB, by nonhomologous recombination, was also isolated for use as a control. The three strains, together with the wt strain NB10, were used to infect rainbow trout by i.p. infections. Both NQ104 (virA) and NQ107 (virB) showed considerably higher LD₅₀s compared with the wild-type NB10 and NQ105, showing that these two mutants were

avirulent (Table 3). This strongly suggests that virA and virB have identical chromosomal locations within NB10, as for 775.17B, and that virA and virB are essential for the virulence of both V. anguillarum strains NB10 and 775.17B.

Phenotypic characterization of the virA and virB mutations. To characterize the functional aspects of the virA and virB mutations, a rabbit polyclonal antiserum, raised against whole cells of strain 775.17B, was used. Total cell lysates of 775.17B (wt), VAN70 (virA), and NQ706 (virB) were prepared and separated by SDS-PAGE. The separated lysates were blotted to nitrocellulose filters, and Western analysis was done, using the whole-cell antiserum, to detect potential differences in the antigenic patterns of 775.17B, VAN70, and NQ706. Interestingly, the lysates of VAN70 (virA) and NQ706 (virB) did not respond to the antiserum (Fig. 5).

Total cell lysates of 775.17B were then treated with either proteinase K or periodic acid, and these lysates were separated on SDS-PAGE and blotted to nitrocellulose filters. Proteinase K treatment did not affect the antigenic pattern,

TABLE 2. LD₅₀ values after i.p. infection of rainbow trout of the parent strain 775.17B, VAN70 (virA), and NQ706 (virB) and transcomplemented derivatives of these strains

Strain	Genotype	Transcomplementing plasmid	Genotype of plasmid	LD ₅₀ (no. of bacteria)
775.17B	wt			1×10^{3}
775.17B	wt	pSKT4	virB virA ⁺	7×10^{3}
775.17B	wt	pSKT9	virB ⁺ virA	7×10^{3}
VAN70	<i>virA</i> ::Tn5-132	1		1×10^{7}
VAN70	<i>virA</i> ::Tn5-132	pSKT1	$virB^+$ $virA^+$	2×10^4
VAN70	<i>virA</i> ::Tn5-132	pSKT4	virB virA+	2×10^5
VAN70	<i>virA</i> ::Tn5-132	pSKT9	virB ⁺ virA	1×10^{7}
VAN70	<i>virA</i> ::Tn5-132	pSKT42	virB virA ⁺	5×10^{3}
NO706	virB::pNO7056	1		7×10^{6}
NO706	virB::pNO7056	pSKC1	virB ⁺ virA ⁺	2×10^{6}
NQ706	virB::pNQ7056	pSKC5	virB virA+	$>1 \times 10^{7}$



FIG. 3. PCR-generated fragments of different strains of V. anguillarum, serotype O1, using virA (primers A1 and A2; see Materials and Methods) (lanes A to D) and virB (primers B1 and B2; see Materials and Methods) (lanes F to I)-specific primers. The strains tested were 775.17B (lanes A and F), NB10 (B and G), 53-507 (C and H), and 6018/1 (D and I). Lane E is a 1-kb DNA ladder (Bethesda Research Laboratories).

while periodate treatment almost completely abolished the response, indicating that the antigen recognized by the antiserum was a carbohydrate structure, presumably LPS (Fig. 5). LPS was therefore prepared from 775.17B (wt) and VAN70 (virA) by the method of Westphal and Jann (39). Also, the LPS of Salmonella typhimurium LT2A was prepared to compare the prepared LPS of V. anguillarum with a classical LPS pattern of a smooth S. typhimurium strain. Both the water and phenol phases were analyzed for the presence of LPS. Intrinsically, ³²P-radiolabelled LPS preparations of the three strains were also prepared in parallel. The preparations were separated by SDS-PAGE and silver stained (data not shown) or autoradiographed (Fig. 6). The water-phase preparations of the V. anguillarum strains showed no obvious difference when 775.17B (wt) and



FIG. 4. PCR-generated fragments of V. anguillarum 775.17B and NB10 and avirulent mutants generated from these strains. Lane A, 775.17B with virA-specific primers (primers A1 and A2; see Mate-rials and Methods); lane B, VAN70 with a virA-specific primer (A1) and a Tn5-132-specific primer; lane C, VAN70 with a virA-specific primer (A2) and a Tn5-132-specific primer; lane D, NB10 with virA-specific primers (A1 and A2); lane E, NQ104 with a virAspecific primer (A1) and a pNQ705-specific primer; lane F, NQ104 with a virA-specific primer (A2) and a pNQ705-specific primer; lane H, 775.17B with virB-specific primers (primers B1 and B2; see Materials and Methods); lane I, NQ706 with a virB-specific primer (B1) and a pNQ705-specific primer; lane J, NQ706 with a virBspecific primer (B2) and a pNQ705-specific primer; lane K, NB10 with virB-specific primers (B1 and B2); lane L, NQ107 with a virB-specific primer (B1) and a pNQ705-specific primer; lane M, NQ107 with a virB-specific primer (B2) and a pNQ705-specific primer. Lane G is a 1-kb DNA ladder (Bethesda Research Laboratories).

TABLE 3. LD_{50} values after i.p. infection of rainbow trout with the wt strain NB10, NQ104 (*virA*), NQ107 (*virB*), and NQ105 (*virB*⁺ *virA*⁺)

Strain	Genotype	LD ₅₀ (no. of bacteria)
NB10	wt	20
NQ104	virA::pNQ7051	2×10^{7}
NO107	virB::pNO7056	2×10^{7}
NQ105	virB ⁺ virA ⁺ , pNQ7051	30

VAN70 (virA) were compared. Both strains showed a rough LPS pattern containing a predominant core similar to the core of S. typhimurium (Fig. 6, lanes A to C). In contrast, the phenol-phase preparations showed a clear difference between 775.17B (wt) and VAN70 (virA). The wt strain showed four high-molecular-weight bands which were lacking in the mutant (Fig. 6, lanes D and E). The silver-stained separated preparations, although the O-polysaccharide side chain ladder of S. typhimurium was more pronounced in the silver-stained gel (data not shown).

The LPS preparations of 775.17B (wt) and VAN70 (virA) were also blotted to nitrocellulose filters after SDS-PAGE. The whole-cell antiserum described above and, in addition, an antiserum raised against LPS from V. anguillarum 40T (a phenol-phase preparation) were used in the blottings. The sera gave identical results. In the water-phase preparations of 775.17B (wt) or VAN70 (virA), no antigenic reaction was observed, while a pattern identical to the one observed for the total cell lysate of 775.17B (Fig. 5, lane A) was found in the phenol-phase preparation of 775.17B (wt) but not for VAN70 (virA) (data not shown). Thus, these results suggest that the major surface antigen is LPS in nature.

The total cell lysates of the transcomplemented exconjugants of VAN70 (virA) and NQ706 (virB) were also tested by Western blotting (Fig. 7), and a perfect correlation was found between the expression of the major surface antigen and virulence for the transcomplemented VAN70 excon-



FIG. 5. Immunoblotting after SDS-PAGE of total cell lysates, using rabbit antiserum directed against whole cells of V. anguillarum 775.17B. Lane A, untreated lysate of 775.17B; lane B, proteinase K-treated lysate of 775.17B; lane C, periodate-treated lysate of 775.17B; lane D, untreated lysate of VAN70; lane E, untreated lysate of NQ706; lane F, prestained molecular weight (10³) standards (Bio-Rad).



FIG. 6. Autoradiogram after SDS-PAGE of intrinsically ³²Pradiolabelled LPS preparations isolated from *V. anguillarum* and *S. typhimurium*. Lanes A to C are water phases after extractions as described by Westphal and Jann (39); lanes D to F are the phenol phases. Lanes A and D, *V. anguillarum* 775.17B; lanes B and E, *V. anguillarum* VAN70; lanes C and F, *S. typhimurium* LT2A. The apparent molecular weights (10³) from the Electrophoresis Calibration Kit (Pharmacia) are indicated.

jugants. Transcomplementation of NQ706 with pSKC1 ($virA^+ virB^+$) resulted in a response with the whole-cell 775.17B antiserum, while transcomplementation with pSKC5 ($virA^+ virB$) did not. This established that an insertion into virB results in avirulence due to the absence of virB per se and not because of a downstream polar effect on virA.

When total cell lysates of the clinical isolate NB10 and its isogenic mutants NQ104 (*virA*) and NQ107 (*virB*) were analyzed by Western blottings, using the whole-cell 775.17B antiserum, the same result was achieved. The antigenic pattern obtained for 775.17B was also seen in the wt NB10 but not in the avirulent mutants NQ104 and NQ107.

The major surface antigen is a constituent of the outer sheath of the flagellum. The phenotypic characterization of the avirulent mutants suggested that the expression of a



FIG. 7. Western blotting after SDS-PAGE of total cell lysates, using rabbit antiserum directed against whole cells of the wt 775.17B. Lane A, prestained molecular weight (10^3) standards (Bio-Rad); lane B, lysate from 775.17B; lane C, lysate from VAN70 (*virA*); lane D, lysate from VAN70/pSKT1 (*virA*⁺); lane E, lysate from VAN70/pSKT4 (*virA*⁺); lane G, lysate from VAN70/pSKT42 (*virA*⁺); lane H, lysate from NQ706 (*virB*); lane I, lysate from NQ706/pSKC1 (*virB*⁺); lane J, lysate from NQ706/pSKC1 (

major surface LPS antigen was lost in the avirulent mutants. Immunogold electron microscopy, using the whole-cell antiserum described above, showed that the antibody only recognized the outer sheath of the polar flagellum of the wt strain 775.17B, not the sheath of the avirulent mutant VAN70 (Fig. 8A and B). The flagellum of V. anguillarum had been shown earlier to be composed of a flagellin core structure surrounded by an outer sheath (5). Our studies confirmed this observation and demonstrated that the antibody bound solely to the outer sheath and not to the flagellin filament (Fig. 8C). Interestingly, no binding of antibody to other surface-located structures was found, although the antiserum used was raised against whole cells. When the avirulent mutant VAN70 (virA) was transcomplemented with plasmid pSKT1 (vir A^+), the binding of the antibody to the flagellum could be restored (Fig. 8D). Moreover, immunogold electron microscopy of the avirulent insertion mutant NQ706 (virB) showed no binding of the antibody, while the transcomplemented NQ706/pSKC1 (virB⁺) recovered binding of the antibody to the outer sheath of the flagellum (data not shown).

The flagellum and the major surface antigen are expressed during fish infections. Since it had been shown that VAN70 (virA) was avirulent after i.p. infection, we asked whether the flagellum and the major surface antigen were expressed during fish infections. Therefore, the expression of these structures in the fish was analyzed by immunogold electron microscopy of bacteria isolated directly from the kidneys of infected moribund rainbow trout (Fig. 8E and F). During fish infections, it was found that the flagellar outer sheath was indeed expressed by both 775.17B (wt) and VAN70 (virA); however, the major surface antigen was expressed only by the wt 775.17B.

DISCUSSION

In this report, we have identified a novel chromosomal locus required for virulence of the fish pathogen *V. anguillarum*. This was achieved by screening transposon mutants of *V. anguillarum* 775.17B for the loss of virulence in fish. Of 200 tested exconjugants, 2 were found to be avirulent after both i.p. and immersion infections of rainbow trout, *O. mykiss*. One of these mutants, VAN20, had two copies of Tn5-132 inserted in the chromosome, while the other, VAN70, had only one. VAN70 was, therefore, chosen for further studies.

The plasmid-encoded iron-sequestering system and the secreted metalloprotease of the bacterium are, so far, the only well-described virulence determinants of *V. anguillarum*. No difference could be observed when the mutant, VAN70, and its parent strain, 775.17B, were compared for these factors. Moreover, no difference was seen for either strain when another putative virulence factor, hemolysin, was tested for activity. This suggests that VAN70 is defective in a novel virulence determinant not earlier described in *V. anguillarum*.

The mutation in VAN70 was cloned and mapped to a region on the chromosome which encompassed two proteins, VirB and VirA. The order of transcription of the genes was *virB-virA*, and the insertion site of transposon Tn5-132 was mapped to the 3' end of *virA*. By using the information gained from nucleic acid sequence analysis, an insertion mutant of *virB* was isolated (NQ706). This mutant was also avirulent when injected into rainbow trout, indicating that the gene products of both *virA* and *virB* are essential for virulence in *V. anguillarum*.



FIG. 8. Immunogold electron microscopy of the flagellum of V. anguillarum. (A) 775.17B (wt); (B) VAN70 (virA); (C) enlargement of a flagellum of strain 775.17B where the sheath is partly removed (arrow); (D) VAN70/pSKT1 (vir A^+); (E) 775.17B (wt) isolated from the kidneys of moribund fish; (F) VAN70 (virA) isolated from the kidneys of a moribund fish. The antiserum used was a rabbit antiserum directed against whole cells of 775.17B (24). Magnifications: ×28,480 in panels A, B, and D to F; ×70,310 in panel C.

Because of the close proximity of virA and virB, it was not obvious whether these genes were part of an operon or were separate transcriptional units. The nucleic acid sequence suggested potential promoters upstream of both virB and virA that showed homology to the E. coli -35 and -10consensus sequences. However, the -10 and the -35 boxes were not optimally spaced when compared with E. coli promoters. Also, potential ribosome-binding sites were found for both genes, but the Shine-Dalgarno site suggested for virA was not optimally positioned. There is a possibility, however, that the transcriptional start point could be at the GTG triplet situated 9 bp upstream of the indicated ATG transcriptional start point of virA, which would reduce the spacing between the Shine-Dalgarno site and the potential transcriptional start point of virA. Nucleic acid sequence analyses of other Vibrio genes have also displayed promoter sequences homologous to the consensus sequences of E. coli (1-3, 10), and in some of these cases the spacing of the different promoter sequences has not been optimal. Potential terminator structures could also be found downstream of the 3' end of virB and virA. These observations suggested that the two genes were contained within different transcriptional units. The transcomplementation studies, in which different plasmid derivatives were conjugated into VAN70 (virA) and NQ706 (virB) (Fig. 7; Table 2), showed that both virA and virB are essential for virulence. However, it was not conclusively shown whether virA and virB indeed were contained within different transcriptional units.

Computer analyses of the *virB* and *virA* sequences revealed that no signal sequences or membrane-spanning regions were present in *virA* or *virB*. When computer homology searches were performed with the search profiles of the Genetics Computer Group (Madison, Wis.) with the programs TFASTA and FASTA, no significant homologies could be found with any published nucleic acid or amino acid sequence. This indicates that these genes are part of a novel system important for bacterial virulence.

Western blotting experiments, in which a polyclonal rabbit antiserum directed against whole cells of the wt strain 775.17B was used, showed a difference in the expression of a major surface antigen between the wt strains and the avirulent mutants tested in this study. No expression of the major surface antigen was observed in Western blottings when total cell lysates of VAN70 (virA), NQ706 (virB), NQ104 (virA), and NQ107 (virB) were analyzed, suggesting that these mutants lacked the antigen. Transcomplementation analysis of virA and virB mutants by fish infections and Western blottings showed a perfect correlation between virulence and the presence of the major surface antigen, suggesting that the expression of the major surface antigen is a prerequisite for virulence. Interestingly, when a total cell lysate of the other isolated avirulent transposon mutant, VAN20, was analyzed by Western blotting, no reaction could be seen with the whole-cell 775.17B antiserum, suggesting that VAN20 did not express the major surface antigen (data not shown). In addition, Southern hybridization showed that the insertion sites of the transposon in this mutant differed from that of VAN70 since neither virA nor virB was affected in VAN20 (data not shown). This would suggest that the expression of the major antigen is dependent on several genes located at different loci of the chromosome of V. anguillarum. Western blottings, using total cell lysates from strains of V. anguillarum serotype O2 and the wholecell 775.17B antiserum, showed that the major surface antigen was not expressed in V. anguillarum serotype O2, indicating that the antigen is serotype specific (data not shown). PCR experiments, in which virA- and virB-specific primers designed for strain 775.17B were used on different strains of V. anguillarum serotype O1, supported this theory. The PCRs resulted in products with sizes corresponding to those of virA and virB, indicating that the virB-virA region is highly conserved, at least among strains of serotype O1. The general importance of virA and virB in the virulence of V. anguillarum was also shown by insertional mutagenesis of virA and virB of a clinical isolate of V. anguillarum serotype O1.

The major surface antigen was resistant to proteinase K treatment and sensitive to periodate treatment, indicating that the antigen was LPS. When water-phase LPS preparations were analyzed, no obvious difference in the LPS pattern could be detected between 775.17 (wt) and VAN70 (virA). However, in phenol-phase LPS preparations, discrete high-molecular-weight bands were found in 775.17B (wt) but not in VAN70 (virA). There are different views about whether the LPS of V. anguillarum can be isolated from the water phase or not. Banoub and Shaw (4) claimed that no LPS could be isolated from the water phase, while Eguchi et al. (8) used a water-phase preparation for chemical analysis of the LPS of V. anguillarum. We found that the major antigen described herein could be isolated from the phenol phase and that this antigen reacted with an LPSspecific antibody. Since we also isolated LPS from the water phase, this indicates that there are two different types of LPS in V. anguillarum, one of which is very immunogenic and located on the outer sheath of the bacterial flagellum. The question regarding the chemical differences of these different LPS structures is currently under investigation. At least in one other bacterial genus, Pectinatus, a similar observation with two distinct LPS has been reported (13).

Immunogold electron microscopy revealed that the major surface antigen of V. anguillarum was located on the outer sheath of the flagellum. In most described bacterial species, the flagellar filament has been found to consist mainly of flagellin subunits, but the flagella of some bacteria, including the human pathogens Helicobacter pylori (12), Vibrio cholerae (16), and other vibrios (11), have been found to have an outer sheath surrounding the flagellin filament. The studies of this sheath have been limited, and no function has been attributed to this structure. We demonstrate the novel finding that the outer sheath of the flagellum is essential for the virulence of V. anguillarum. We know from earlier studies that the avirulent mutant VAN70 can invade fish and trigger an immune response (24). This suggests that the sheath structure is not needed in the initial stages of invasion, but rather in the generation of a fulminant systemic infection in the fish. We could also demonstrate that the flagellar sheath indeed was expressed in vivo during fish infections. Thus, the flagellum of V. anguillarum is a virulence organelle expressed during the infection of fish and consists of a structure on the outer sheath that is highly antigenic and essential for virulence.

ACKNOWLEDGMENTS

We gratefully acknowledge the skillful technical assistance of Stina Bäckman, Ulla Eriksson, and Mats Jakobsson. The skillful technical assistance of Lenore Johansson in performing the electron microscopy is highly appreciated. We also thank Debra Milton for critical reading of the manuscript. We are very grateful for the gift of the LPS-antiserum from Derek H. Shaw, Newfoundland, Canada.

This work was supported by the Swedish Council for Forestry and Agricultural Research, 0905/89 V 79:1, and the Swedish Board for Technical Development, 86-03199P.

REFERENCES

- Actis, L. A., M. E. Tolmasky, D. H. Farrell, and J. H. Crosa. 1988. Genetic and molecular characterization of essential components of the *Vibrio anguillarum* plasmid-mediated iron-transport system. J. Biol. Chem. 263:2853–2860.
- 2. Alm, R. A., and P. A. Manning. 1990. Characterization of the hlyB gene and its role in the production of the El Tor haemolysin of *Vibrio cholerae* O1. Mol. Microbiol. 4:413–425.
- 3. Alm, R. A., U. H. Strocher, and P. A. Manning. 1988. Extracellular proteins of *Vibrio cholerae*: nucleotide sequence of the structural gene (hlyA) for the haemolysin of the haemolytic El Tor strain O17 and characterization of the hlyA mutation in the non-haemolytic classical strain 569B. Mol. Microbiol. 2:481– 488.
- Banoub, J. H., and D. H. Shaw. 1981. Isolation and characterization of 3-acetamido-3,6-dideoxy-L-glucose from the core oligosaccharides obtained from the aquatic gram-negative bacteria *Aeromonas hydrophila* and *Vibrio anguillarum*. Can. J. Biochem. 59:877-879.
- Chart, H. 1983. Multiflagellate variants of Vibrio anguillarum. J. Gen. Microbiol. 129:2193–2197.
- Crosa, J. H. 1980. A plasmid associated with virulence in the marine fish pathogen *Vibrio anguillarum* specifies an ironsequestering system. Nature (London) 284:566-568.
- Crosa, J. H., L. L. Hodges, and M. H. Schiewe. 1980. Curing of a plasmid is correlated with an attenuation of virulence in the marine fish pathogen *Vibrio anguillarum*. Infect. Immun. 27: 897-902.
- Eguchi, H., S. Kaya, and Y. Araki. 1992. Occurrence of 2,4dihydroxy-3,3,4-trimethylpyroglutamic acid as an N-acyl substituent in the O-polysaccharide chain of the lipopolysaccharide of *Vibrio anguillarum* V-123. Carbohydr. Res. 231:147–158.
- Ely, B. 1985. Vectors for transposon mutagenesis of non-enteric bacteria. Mol. Gen. Genet. 200:302-304.
- 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.
- 11. Follett, E. A. C., and J. Gordon. 1963. An electron microscope study of Vibrio flagella. J. Gen. Microbiol. 32:235–239.
- Geis, G., H. Leying, S. Suerbaum, U. Mai, and W. Opperkuch. 1989. Ultrastructure and chemical analysis of *Campylobacter pylori* flagella. J. Clin. Microbiol. 27:436-441.
- Helander, I. M., R. Hurme, A. Haikara, and A. P. Moran. 1992. Separation and characterization of two chemically distinct lipopolysaccharides in two *Pectinatus* species. J. Bacteriol. 174: 3348-3354.
- Hitchcock, P. J., and T. M. Brown. 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J. Bacteriol. 154:269–277.
- Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51-59.
- Hranitzky, K. W., A. Mulholland, A. D. Larson, E. R. Eubanks, and L. T. Hart. 1980. Characterization of a flagellar sheath protein of Vibrio cholerae. Infect. Immun. 27:597-603.
- Kodama, H., M. Moustafa, T. Mikami, and H. Izawa. 1985. Characterization of extracellular substance of *Vibrio anguillarum* toxic for rainbow trout and mice. Microbiol. Immunol. 29:909–920.
- Kolter, R., M. Inuzuka, and D. R. Helinski. 1978. Transcomplementation-dependent replication of a low molecular weight origin fragment from plasmid RK6. Cell 15:1199–1208.
- Laemmli, U. K., and M. Favre. 1973. Maturation of the head of bacteriophage T4. J. Mol. Biol. 80:575-599.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

- 21. Miller, V. L., and J. J. Mekalanos. 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregula-
- and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires taxR. J. Bacteriol. 170:2575-2583.
- Munn, C. B. 1978. Haemolysin production by Vibrio anguillarum. FEMS Microbiol. Lett. 3:265-268.
- 23. Neidhardt, F. C., P. L. Bloch, and D. F. Smith. 1974. Culture medium for enterobacteria. J. Bacteriol. 119:736-747.
- Norqvist, A., Å. Hagström, and H. Wolf-Watz. 1989. Protection of rainbow trout against vibriosis and furunculosis by the use of attenuated strains of *Vibrio anguillarum*. Appl. Environ. Microbiol. 55:1400-1405.
- Norqvist, A., B. Norrman, and H. Wolf-Watz. 1990. Identification and characterization of a zinc metalloprotease associated with invasion by the fish pathogen *Vibrio anguillarum*. Infect. Immun. 58:3731-3736.
- Nyberg, G., N. Strömberg, A. Jonsson, K.-A. Karlsson, and S. Normark. 1990. Erythrocyte gangliosides act as receptors for *Neisseria subflava*: identification of the Sia-1 adhesin. Infect. Immun. 58:2555-2563.
- 27. Reed, L. J., and H. Muench. 1938. A simple method of estimating fifty per cent endpoints. Am. J. Hyg. 27:493-497.
- Rosenberg, A. H., B. N. Lade, D.-S. Chui, S.-W. Lin, J. J. Dunn, and F. W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135.
- Sakai, D. K. 1985. Loss of virulence in a protease-deficient mutant of Aeromonas salmonicida. Infect. Immun. 48:146-152.
- Salinas, P. C., M. E. Tolmasky, and J. H. Crosa. 1989. Regulation of the iron uptake system in *Vibrio anguillarum*: evidence for a cooperative effect between two transcriptional activators. Proc. Natl. Acad. Sci. USA 86:3529–3533.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Simon, R., U. Priefer, and A. Pühler. 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Biotechnology 1:784– 791.
- 33. Singer, J. T., K. A. Schmidt, and P. W. Reno. 1991. Polypeptides p40, pOM2, and pAngR are required for iron uptake and for virulence of the marine fish pathogen *Vibrio anguillarum* 775. J. Bacteriol. 173:1347–1352.
- Sörensen, U., and J. L. Larsen. 1986. Serotyping of Vibrio anguillarum. Appl. Environ. Microbiol. 51:593-597.
- 35. Swanson, J., L. W. Mayer, and M. R. Tam. 1982. Antigenicity of Neisseria gonorrhoeae outer membrane protein(s) III detected by immunoprecipitation and Western blot transfer with a monoclonal antibody. Infect. Immun. 38:668–672.
- Trust, T. J., I. D. Courtice, A. G. Khouri, J. H. Crosa, and M. H. Schiewe. 1981. Serum resistance and haemagglutination ability of marine vibrios pathogenic to fish. Infect. Immun. 34:702-707.
- 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.
- Walter, M. A., S. A. Potter, and J. H. Crosa. 1983. Iron uptake system mediated by *Vibrio anguillarum* plasmid pJM1. J. Bacteriol. 156:880-887.
- Westphal, O., and K. Jann. 1965. Bacterial lipopolysaccharides. Methods Carbohydr. Chem. 5:83-91.
- Wolf, M. K., and J. H. Crosa. 1986. Evidence for the role of a siderophore in promoting *Vibrio anguillarum* infections. J. Gen. Microbiol. 132:2949-2952.
- 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.
- Young, R. A., and R. W. Davis. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science 222:778–782.