Panton-Valentine Leucocidin and Gamma-Hemolysin from Staphylococcus aureus ATCC 49775 Are Encoded by Distinct Genetic Loci and Have Different Biological Activities

G. PRÉVOST,¹ B. CRIBIER,² P. COUPPIÉ,² P. PETIAU,² G. SUPERSAC,¹ V. FINCK-BARBANÇON,²† H. MONTEIL,¹ AND Y. PIEMONT¹*

Institut de Bactériologie de la Faculté de Médecine, Université Louis-Pasteur,¹ and Clinique Dermatologique des Hôpitaux Universitaires de Strasbourg,² 67000 Strasbourg, France

Received 27 February 1995/Returned for modification 1 May 1995/Accepted 17 July 1995

Staphylococcus aureus ATCC 49775 produces three proteins recognized by affinity-purified antibodies against the S component of Panton-Valentine leucocidin (LukS-PV) and two proteins recognized by affinity-purified antibodies against the F component of this toxin (LukF-PV). Purification of these proteins and cloning of the corresponding genes provided evidence for the presence of two loci. The first one, encoding Panton-Valentine leucocidin, consisted of two cotranscribed open reading frames, *lukS-PV* and *lukF-PV*, coding the class S and the class F components, respectively. The second one coded for a gamma-hemolysin and consisted of two transcription units, the first one encoding an HlgA-like protein, a class S component, and the second one encoding two cotranscribed open reading frames identical to HlgC and HlgB, class S and class F components, respectively, from gamma-hemolysin from the reference strain Smith 5R. It appears that the Panton-Valentine leucocidin from *S. aureus* ATCC 49775 (V8 strain) should not be confused with leucocidin from ATCC 27733 (another isolate of V8 strain), which had 95% identity with HlgC and HlgB from gamma-hemolysin. The cosecretion of these five proteins led to six possible synergistic combinations between F and S components. Two of these combinations (LukS-PV-LukF-PV and HlgA-LukF-PV) had dermonecrotic activity on rabbit skin, and all six were leukocytolytic on glass-adsorbed leukocytes. Only three were hemolytic on rabbit erythrocytes, the two gamma-hemolysin combinations and the combination LukF-PV-HlgA.

Staphylococcus aureus is one of the most frequently isolated pathogens in hospital routine. It produces numerous virulence factors, including the bicomponent toxins gamma-hemolysin (2, 10, 21, 28, 33) and Panton-Valentine leucocidin (PVL) (8, 9, 20, 37–39). These toxins consist of two separately secreted and nonassociated proteins (class S and class F components) acting synergistically on human and rabbit polymorphonuclear cells, monocytes, and macrophages. In addition, gamma-hemolysin is able to lyse erythrocytes from a wide range of mammalian species (28). Gamma-hemolysin is produced by more than 99% of *S. aureus* strains, and 2% of these strains also produce PVL (22).

PVL has been purified (8, 39) from a V8 strain (ATCC 49775), and apparent molecular masses of 32 and 38 kDa were determined for the S and F components of PVL, respectively (termed LukS-PV and LukF-PV in this work). Its mode of action has been investigated (3, 7, 38, 40). By use of an immunoprecipitation test (8), it was found that PVL-producing strains were strongly associated with human primary necrotizing cutaneous infections such as furuncles (5, 6, 22).

The data concerning leucocidins appear to be ever more complex, since another leucocidin (17) having different apparent molecular masses for its two constituting components (31 and 32 kDa) was purified from a different V8 strain (ATCC 27733). The biological activity of the latter toxin has been studied by Noda and coworkers (13, 17–19). The genes encoding this toxin have never been characterized.

Furthermore, three other S. aureus bicomponent toxins which are leukotoxic, hemolytic, and structurally related have been reported. These toxins are (i) gamma-hemolysin (2, 10, 21, 33), which is encoded by the hlg locus of strain Smith 5R (4), (ii) bovine leucocidin (or leucocidin R) (29), which is encoded by the lukR locus (31) of strain P83, and (iii) a leucocidin produced by a methicillin-resistant S. aureus (MRSA) strain which is encoded by the luk locus (12, 23, 24). These three loci all consist of two consecutive transcription units in the same orientation, the first encoding a single open reading frame (ORF) and the second encoding two ORFs. For each of these three loci, the first two ORFs encode two class S components (HlgA and HlgC, respectively) and the third encodes a class F component (HlgB). In the three loci, each gene encoding a given ORF from gamma-hemolysin, bovine leucocidin, and leucocidin of an MRSA strain has at least 95% identity with its counterpart in the two other loci. Thus, all three toxins should be considered structural variants of the gamma-hemolysin and will be referred to in this work as the gamma-hemolysin group. Since each member of the gamma-hemolysin group is composed of two different (65% identity) class S proteins and one class F protein and biological activity results from the combination of one class S component with one class F component (3), each toxin in the gamma-hemolysin group could, in fact, give rise to two biologically active toxins.

On the basis of a few peptide sequences derived from PVL (31), evidence indicating that PVL and gamma-hemolysin are members of a single toxin family was obtained. We proposed the name synergohymenotropic (SHT) toxins for this toxin

^{*} Corresponding author. Mailing address: Institut de Bactériologie de la Faculté de Médecine, 3 rue Koeberlé, F-67000 Strasbourg, France. Phone: 33 88 21 19 70. Fax: 33 88 25 11 13.

[†] Present address: Medical College of Wisconsin, Department of Microbiology, Milwaukee, WI 53226.

family since these toxins act by the synergy of two proteins and are pore forming.

Here, we report the entire primary structure of LukS-PV and LukF-PV from *S. aureus* ATCC 49775 and characterize the other immunoreactive components purified from this strain. The purification of these different, but related, proteins offered the opportunity to test all of the possible synergistic combinations on human polymorphonuclear leukocytes (PMNs), on sheep, rabbit, and human erythrocytes, and on rabbit skin.

MATERIALS AND METHODS

Bacteria. The *S. aureus* V8 strain producing PVL was kindly provided by S. Thornley (Wellcome Laboratories, London, United Kingdom) and deposited in the American Type Culture Collection with the reference strain ATCC 49775 because of its high level of PVL production compared with that of another V8 strain (ATCC 27733). *S. aureus* P83 producing the so-called leucocidin R was a generous gift from N. L. Norcross (Cornell University, Ithaca, N.Y.). *Escherichia coli* NM522 [*supE thi hsdR* Δ (*lac-proAB*) F' (*proAB lacI*⁴ ZM15)] was used as the host strain of the recombinant pUC19 plasmid (41).

In this article, the two genes encoding S and F components from PVL are called *lukS-PV* and *lukF-PV*, and the three genes encoding members of the gamma-hemolysin group are called *hlgA*, *hlgC* (S-class genes), and *hlgB* (F-class gene), as described by Cooney et al. (4).

Purification of LukS-PV, LukF-PV, and other antigenically related components. The LukS-PV and LukF-PV antigenically related components as well as the two components of PVL were purified simultaneously from the culture supernatant fluid of *S. aureus* ATCC 49775. Fractions obtained at different elution steps were assayed by immunoprecipitation with polyclonal affinity-purified rabbit antibodies against LukS-PV and LukF-PV (8). Briefly, bacteria were grown to the stationary phase with vigorous shaking in 12 2-liter Erlenmeyer flasks, each containing 170 ml of modified CCY medium (8). Bacteria were separated from the culture supernatant by a 20-min centrifugation at $10,000 \times g$ at 4°C for 20 min, followed by filtration (Millipore, Paris, France; 0.45-µm pore diameter). Proteins were precipitated for 16 h at 4°C in 80% (wt/vol) saturated ammonium sulfate solution. The precipitated material was collected by centrifugation for 20 min at 20,000 \times g. Proteins were suspended in 100 ml of 0.2 M NaCl-buffer 1 (50 mM Na phosphate buffer [pH 7.0]) and chromatographed on a 40-ml Sepharose SP Fast Flow column (Pharmacia, Uppsala, Sweden) equilibrated in buffer 1. Fractions of the isocratic elution were stored for further purification. Proteins bound to the matrix of the column were eluted with 0.7 M NaCl-buffer 1. Fractions containing immunoreactive material were dialyzed against 0.2 M NaCl-buffer 1 and then subjected to cation-exchange Mono S fast performance liquid chromatography (FPLC; Pharmacia). Elution was performed with a linear gradient ranging from 0.2 to 0.56 M NaCl in buffer 1. After the addition of ammonium sulfate to 1.5 M, these protein fractions were further purified on an Alkyl-Superose FPLC (Pharmacia) with an ammonium sulfate linear gradient ranging from 1.5 to 1.0 M in buffer 1. Immunoreactive fractions were then reequilibrated in 0.4 M NaCl-buffer 1. Their concentration was adjusted to $1 A_{280}$ unit, and the fractions were stored frozen at -80° C

In a second step, the fractions of the isocratic elution mentioned above were equilibrated in buffer 2 (30 mM Na phosphate buffer [pH 6.5]). The immunoreactive components were purified by chromatography through a cation-exchange Mono S FPLC with a linear gradient ranging from 0 to 0.25 M NaCl in buffer 2. Each fraction containing immunoreactive protein was further purified to high homogeneity through an Alkyl-Superose FPLC with a linear gradient ranging from 1.5 to 0 M ammonium sulfate in buffer 2. The purified components were then dialyzed against 0.2 M NaCl-buffer 2. Their concentration was adjusted to $1A_{280}$ unit, and the components were stored frozen at -80° C.

Protein analysis was performed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (15) with the PHAST-System apparatus (Pharmacia). Immunoblotting was carried out with affinity-purified LukS-PV or LukF-PV rabbit polyclonal antibodies as described previously (35). Characterization of the purified proteins was performed by determining their N-terminal peptide sequences.

DNA cloning and sequencing methods. Restriction endonucleases and DNAmodifying enzymes were used as recommended by the manufacturers (Gibco-Bethesda Research Laboratories and New England Laboratories [Beverly, Mass.]). The DNA methods used were described previously (25). For the cloning of *luk-PV*, a partial DNA library was constructed by selecting *Eco*RV-digested fragments of *S. aureus* ATCC 49775 DNA, ranging from 4.2 to 6.8 kb, and inserting them in a *Hin*cII-linearized pUC19 plasmid. After transformation of *E. coli* NM522, recombinant clones were screened with a 5'-labeled degenerated oligonucleotide probe as described previously (31). This 27-mer oligonucleotide (5'-TATA/C/TAGC/TGAAGTAGAACAC/TCAA/GAAC-3') was deduced from the underlined portion of the N-terminal sequence of a trypsin-generated peptide of LukS-PV, i.e., TISYNQQN<u>YISEVEHONSK</u>.

The *hlg* locus of *S. aureus* ATCC 49775 was cloned with a partial DNA library of *ScaI* DNA fragments ranging from 3.5 to 5.6 kb. These DNA fragments were inserted into a *HincII*-linearized pUC19 plasmid. The screening of the partial

library was performed with a 5'-labeled oligonucleotide probe from nucleotide 1771 to 1753 (5'-TTGTTTTATCTTCTGTCCT-3') deduced from the *hlgC* sequence (EMBL-GenBank data bank library accession no. L01055) as described previously (31).

Double-stranded DNA sequencing was performed by the dideoxy chain termination method (26) and with T7 DNA polymerase (32) on double-stranded DNA. Initially, the universal primer and the cloning oligonucleotides were used; those deduced from previous sequencing steps were used subsequently. Analysis of the nucleotide and the peptide sequences was performed with DNAStar (London, United Kingdom) computer software. The nucleotide sequences of the *luk-PV* and the *hlg*-like loci from *S. aureus* ATCC 49775 are recorded in the EMBL-GenBank data bank, accession numbers X72700 and X81586, respectively.

Southern and Northern (RNA) blot analysis. The preparation of total DNA from *S. aureus* ATCC 49775 or P83 was described previously (31), and total RNA was isolated as described earlier by Kornblum et al. (14) and modified by Supersac et al. (31). DNA-DNA-specific hybridization (30) for *luk-PV* used a *lukF-PV*-directed oligonucleotide probe, namely, 5'-ATGACTCAGTAAACGT TGTAGAT-3' (nucleotides 2338 to 2360), deduced from the sequences of PVLencoding genes. Specific DNA hybridization for *hlg* genes from *S. aureus* Smith 5R (EMBL-GenBank data bank library accession no. L01055) used oligonucleotide probes specific for *hlgA*, 5'-TCTGCTTTAGAAATTTCTATA-3' (nucleo tides 207 to 197), and for *hlgC*, 5'-TTGTTTTATCTTCTGTCCT-3' (nucleo tides 1771 to 1753 (4).

RNA-DNA hybridizations (34) were performed with DNA fragments which were ³²P labeled by nick translation (Gibco-Bethesda Research Laboratories). Hybridization specific for *luk-PV* was performed with a cloned 2,750-bp *HincII-HindIII S. aureus* DNA fragment which includes the two *luk-PV* ORFs as the probe. Hybridization specific for *hlgA* used a cloned 1,000-bp *Eco*RV-*HindIII S. aureus* DNA fragment, which includes the whole gene, as the probe, and hybridization specific for *hlgC* and part of *hlgB*, as the probe. Hybridizations were performed for 15 h at 60°C in 6× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaH₂PO₄, and 1 mM EDTA [pH 7.7])-5× Denhardt's solution (25)–0.1% (wt/vol) SDS and autoradiographed.

Hemolytic activities of SHT toxins. Blood from sheep, rabbits, or humans was collected with heparin (4 IU/ml). Erythrocytes were sedimented by 5 min of centrifugation at 1,000 × g at 4°C and washed three times with 9.55‰ (wt/vol) phosphate-buffered saline (PBS)–Dulbecco minimal essential medium without Ca²⁺ or Mg²⁺ (Seromed, Berlin, Germany) plus 0.5% (wt/vol) gelatin (buffer 3). The erythrocytes were resuspended at a 1% (vol/vol) concentration in the same buffer. It has been reported that LukS-PV binds with a K_d of 6 nM (3) to the membranes of target cells prior to interaction with LukF-PV. Therefore, biological activity should be limited by the amount of class S components bound to the cells. For these experiments, limiting amounts of class S components (not exceeding 1 nM) were tested in the presence of an excess of class F components. Erythrocytes (100 µl at 10⁷ cells per ml) were mixed gently with an excess of one class F component (125 ng in 50 µl) and with twofold serial dilutions (50 µl) of the class S component to be assayed. Following a 1-h incubation at 37°C and a 5-min centrifugation at room temperature, hemolysis was determined by the A_{450} of the supernatant.

Each assay was performed four times and included a positive control (100% hemolysis with 1 mg of saponin per ml) and a blank (0% hemolysis with buffer only). One unit of hemolytic activity is defined as the smallest amount of assayed class S component inducing 50% hemolysis.

Leukotoxic activity of SHT toxins. The method used to determine the leukotoxic activity of SHT toxins was described previously by Finck-Barbançon et al. (8) and carried out with human glass-adherent leukocytes. For the same reasons as those described above, specific activities were determined only with variable concentrations of one class S component plus an excess of one class F component (50 ng) in 20 μ l of buffer 3.

Activity of SHT toxins in a rabbit skin model. Shaved New Zealand White rabbits were injected intradermally with 50 μ l of sterile 9.55‰ (wt/vol) PBS– Dulbecco minimal essential medium containing equal amounts (3, 30, 300, and 3,000 ng) of one class S- and one class F-related component. Similarly, formaldehyde-treated PVL toxoids obtained as described previously (6) were injected into rabbit skin. Assays were performed in duplicate. Macroscopic observations of the lesions were recorded after 0, 4, 24, and 48 h, and lesions were biopsied under general anesthesia induced by intramuscular injection of 250 μ l of reconstituted Zoletil-100 (Reading Laboratories, Carros, France). Skin biopsies were fixed in Bouin solution, paraffin embedded, cut in 4- μ m-thick sections, and stained with hematoxylin-eosin-safran-astra blue. The tissue sections were examined in a blind fashion.

RESULTS

Protein purification. Since the immunoprecipitation test directed against LukS-PV and LukF-PV had indicated that other antigenically related proteins were produced by *S. aureus* ATCC 49775, these compounds were purified by use of a 123456



FIG. 1. Electrophoresis under reducing conditions of purified protein components of SHT toxins from *S. aureus* ATCC 49775. The five protein components of SHT toxins (1.4 μ g of each loaded) produced by *S. aureus* ATCC 49775 have different apparent molecular masses, whereas their calculated molecular masses are very similar. Lanes: 1, purified HlgA; 2, purified HlgC; 3, purified LukS-PV; 4, protein molecular weight marker; 5, purified HlgB; 6, purified LukF-PV.

purification scheme analogous to that used for PVL. Two proteins which were antigenically related to LukS-PV were found to bind to the Sepharose SP Fast Flow column. The first was eluted from the cation-exchange Mono S FPLC column with 335 mM NaCl, and the second was eluted with 450 mM NaCl. In the final purification step (Alkyl-Superose FPLC), the first was eluted with 1.2 M ammonium sulfate and the second was eluted with 1.1 M.

The three proteins which eluted in the isocratic elution of the Sepharose SP Fast Flow column were identified as LukF-PV, a LukF-PV antigenically related protein, and LukS-PV. They were eluted from the cation-exchange Mono S FPLC column with 60, 130, and 200 mM NaCl, respectively, and were purified to homogeneity after elution from the Alkyl-Superose FPLC column at 980, 870, and 375 mM ammonium sulfate, respectively.

There was no contamination of any of the five proteins by any other synergistic component or by another toxin (such as the alpha-toxin) with comparable physicochemical properties because each protein component alone had neither leukotoxic nor hemolytic activity. Sequencing of the 15 N-terminal amino acids of the five proteins purified from S. aureus ATCC 49775 revealed that the first eluted LukS-PV-related protein has the same sequence as that of HlgC (KGNDVEIIKRTEDKT), that the second eluted LukS-PV-related protein has the same sequence as that of HlgA (ENKIEDIGQGAEIIK), and that the LukF-PV-related protein has the same sequence as that of HlgB (AEGITPVSVKKVDDK). In this new purification scheme, LukS-PV and LukF-PV were identified by immunoprecipitation with LukS-PV and LukF-PV affinity-purified antibodies; these proteins are identical to the two PVL components purified previously in our laboratory (8). These LukS-PV and LukF-PV proteins have N-terminal sequences (DNNIENIG DGAEVVK and AQHITPVSEKKVDDK, respectively) that differ from those of the three other purified proteins but are identical to those described previously for these components (31). This new purification procedure led to the purification of 12 mg of LukS-PV, 10 mg of LukF-PV and HlgB-like protein, 5 mg of HlgC-like protein, and 3 mg of HlgA-like protein from 2 liters of culture supernatant (Fig. 1). These amounts are in accordance with the intensity of protein bands detected by

immunoblotting crude culture supernatant of *S. aureus* ATCC 49775.

Cloning and sequencing the genes encoding the SHT toxins from *S. aureus* **ATCC 49775.** The results of protein purification established that *S. aureus* **ATCC 49775** is able to produce a gamma-hemolysin or gamma-related hemolysin in addition to PVL. This was confirmed by hybridization experiments. As shown in Fig. 2A and B, oligonucleotide probes specific for *hlgA* and *hlgC* bound to restriction enzyme-digested DNA fragments of the same size for DNA isolated from *S. aureus* strains ATCC 49775 and P83, the latter producing a toxin of the gamma-hemolysin group.

(i) Genes coding for gamma-hemolysin. The three ORFs constituting the *hlg*-like locus of *S. aureus* ATCC 49775 were cloned in a 4.5-kb *ScaI-ScaI* DNA fragment into the pUC19 plasmid. This fragment hybridized with both *hlgA* and *hlgC* probes (data not shown). Crude extracts of positive recombinant *E. coli* NM522 were able to induce hemolysis of rabbit erythrocytes at dilutions up to 1/800 and were leukotoxic at dilutions up to 1/300, whereas crude extracts of *E. coli* transformed with pUC19 alone had no hemolytic or leukotoxic activity.

Of the cloned ScaI-ScaI DNA fragment, 3,763 bp were sequenced. Three ORFs, hlgA, hlgC, and hlgB, were identified with the same genetic organization as the hlg-related genes described previously (4, 12, 31). The two cotranscribed ORFs, the hlgC- and hlgB-like ORFs, encoded proteins identical to those encoded by lukR (31). Only two silent base substitutions (alanine substituted by glycine at position 430 [A430G] and T943C) differed in the genes. In the case of the hlgA-like ORF (3), of the five point mutations observed (T481C, C767A, G855A, T919A, A933C), only the third and the fifth were responsible for amino acid substitutions (R258K and H284P) as compared with those of lukR. The intergenic region between hlgA and hlgC of the hlg-like locus of S. aureus ATCC 49775 is 528 bp long, whereas that of the hlg locus (4) from S. aureus Smith 5R is 566 bp long. It is notable that a variable sequence exists within a region extending from 100 to 200 bp downstream of the *hlgA* ORF. This variable sequence is responsible for the size difference between the two intergenic sequences.



FIG. 2. Southern blot analysis showing that total DNA from *S. aureus* ATCC 49775 hybridizes with the *hlg*-specific probes and with the *luk-PV*-specific probe, whereas total DNA from *S. aureus* P83 hybridizes only with the *hlg*-specific probe; (A) Hybridization with the *hlg*A-specific probe; (B) hybridization with the *hlgC*-specific probe; (C) hybridization with the *luk-PV*-specific probe; (C) hybridization with the *luk-PV*-specific probe. Lanes: 1, 5'-labeled *Bst*EII-restricted DNA from bacteriophage lambda; 2 to 5, *Eco*RL-, *Eco*RV-, *Acc*I-, and *ScaI*-restricted DNA fragments from *S. aureus* ATCC 49775, respectively; 6 to 9, *Eco*RL-, *Eco*RV-, *Acc*I-, and *ScaI*-restricted DNA fragments from *S. aureus* P83, respectively.

INFECT. IMMUN.

1	CATGTTCTAG	AGATTCATTT	AGACGCAGCA	GGAGAAAGTG	CAAGTGGTGG	GCATGTTATT	ATTTCAAGTC	AATTCAATGC	80
81	AGATACTATT	GATAAAAGTA	TACAAGATGT	TATTAAAAAT	AACTTAGGAC	AAATAAGAGG	TGTAACACCT	CGTAATGATT	160
161	TACTGAACGT	TAATGTATCA	GCAGAAATAA	ATATCAATTA	TCGTTTATCT	GAATTAGGTT	TTATTACTAA	TGAAAAAAGA	240
241	TATGGATTGG	ATTAAGAAGA	ATTATGACTT	GTATTCTAAA	TTAATAGCTG	GTGCGATTCA	TGGTAAGCCT	ATAGGTGGTT	320
321	TGGTAGCTGG	TAATGTTAAA	CATCAGCTAA	АААССААААА	AATCCACCAG	TGCCAGCAGG	TTATACACTT	GATAAGAATA	400
401	ATGTGCCTTA	TAAAAAAAAAG	GCTGGTAATT	ACACAGTTGC	CAATGTTAAA	GGTAATAACG	TAAGGGACGG	CTATTCAACT	480
481	AATTCAAGAA	TTACTGGTGT	ATTACCTAAT	AACGCAACAA	TCAAATATGA	CGGCGCATAT	TGTATCAATG	GCTATAGATG	560
561	GATTACTTAT	ATCGCTAATA	GTGGACAACG	TCGTTATATC	CCCACACGAC	ACCTACATAA	ACCACCTART	ACCATABOTA	640
641	CTTTTCCTAA	CTTTACCOLC	OT COMON NEW	CENERATCHEC	A DECEMBER OF	AGGINGAIAA	GENERACOCE	AGGAIRAGIA	2040
201	STITIGGIAA	BCTTINCGCAG	DIGAIAATI	GIAIAIGAIG	AAICIIAGGC	AGGIACIICG	GIAIIIGCCI	ALIAITIAAA	720
001	TCCTATAAAC	AGIIAAIIII	IACATAAATA	TATTAAATTT	TAAAAAAAAAA	AACGITTTTA	GTATATAAAT	TATTTTGTGT	800
001	ATTATTTATA	ATTTCTTCC	TAAAAAGIIGI	CERTERIT	ANTANCGIGG	ANANANGITT	TIATCAATT	AAAATATAAA	880
001	AI INI I INI A	ATTIGITIGG	INAIGAACGG	GITTTTCGA	AATAATAGTA	AAAAAACACA	TITGIAGAIA	TITIMACIC	960
961	GGTAAATCTT	TTAATAAA <u>TA</u>	<u>TTTA</u> ATTTTA -10	ΤΤΑΑΑ ΑGTTA	AAAAGGTTTA	ATATAAAAAT	GTAATAAAAT	-35 TTATAAA <u>GAA</u>	1040
1041	AGGAAATGAT	TTTTATGGTC	AAAAAAAGAC	TATTAGCTGC	AACATTGTCG	TTAGGAATAA	TCACTCCTAT	TGCTACTTCG	1120
	IDS	<u>M_V_</u>	KKKL	<u> </u>	T_L_S		<u> </u>	<u>A T S</u>	
1121	TTTCATGAAT	CTAAAGCTGA	TAACAATATT	GAGAATATTG	GTGATGGCGC	TGAGGTAGTC	AAAAGAACAG	AAGATACAAG	1200
	<u>F H E Ş</u>	KA_D	NNI	ENIG	DGA	EVV	KRTE	DTS	
1201	TAGCGATAAG	TGGGGGGTCA	CACAAAATAT	TCAGTTTGAT	TTTGTTAAAG	ATAAAAAGTA	TAACAAAGAT	GCTTTGATTT	1280
	SDK	WGVT	QNI	QFD	FVKD	ккү	NKD	ALIL	
1281	TAAAAATGCA	AGGTTTTTATC	AATTCAAAGA	CTACTTATTA	CAATTACAAA	AACACAGATC	ATATAAAAGC	AATGAGGTGG	1360
	K M Q	G F I	NSKT	TYY	NYK	NTDH	IKA	MRW	
1361	COTTICCAAT	ACAATATTGG	TCTCAAAACA	AATGACCCCA	ATGTAGATTT	AATAAATTAT	CTACCTAAAA	ATAAAATAGA	1440
	PFQY	N I G	LKT	NDPN	VDL	INY	LPKN	KID	
1441	TTCAGTAAAT	GTTAGTCAAA	CATTAGGTTA	TAACATAGGT	GGTAATTTTA	ATAGTGGTCC	ATCAACAGGA	GGTAATGGTT	1520
	S V N	VSQT	LGY	NIG	GNFN	SGP	STG	GNGS	
1521	CATTTAATTA	TTCAAAAACA	ATTAGTTATA	АТСААСАААА	CTATATCAGT	GAAGTAGAAC	ATCAAAATTC	AAAAAGTGTT	1600
	FNY	SKT	ISYN	QQN	YIS	EVEH	QNS	K S V	
1601	CAATGGGGAA	TAAAAGCTAA	TTCATTTATC	ACATCATTAG	GTAAAATGTC	TGGACATGAT	CCAAATTTAT	TTGTTGGATA	1680
	QWGI	KAN	SFI	TSLG	KMS	GHD	PNLF	VGY	
1681	TAAACCATAT	AGTCAAAATC	CGAGAGACTA	TTTTGTGCCA	GACAATGAAT	TACCCCCATT	AGTACACAGT	GGTTTCAATC	1760
	КРҮ	SQNP	RDY	FVP	DNEL	P P L	VHS	GFNP	
1761	CTTCATTTAT	TGCTTCTGTT	TCTCATGAAA	AAGGCTCAGG	AGATACAAGT	GAATTTGAAA	TAACGTATGG	CAGAAATATG	1840
	SFI	ASV	з н е к	GSG	DTS	EFEI	ΤΥG	RNM	
1841	GATGTTACTC	ATGCTACTAG	AAGAACAACA	CACTATGGCA	ATAGTTATTT	AGAAGGATCT	AGAATACACA	AGGCATTTGT	1920
	D V T H	ATR	RŤT	HYGN	SYL	EGS	RIHK	AFV	
1021	ARACRCARAT.	TACACACTTA	A ATATCA ACT	GAACTGGAAA	ACTCATGAAA	TTAAAGTGAA	AGGACATAAT	TGATATGAAA	2000
1921	nnnononnni	INCICIOITA	minion						
1921	N R N	Y T V K	Y E V	NWK	тнеі	кvк	GHNS	top <u>M_K</u>	
1921	N R N	Y T V K	Y E V	N W K	тнеі	кvк	G H N s rbs	top <u>M_K</u>	
2001	N R N	Y T V K	Y E V AGTTACATCA	N W K	T H E I TTTTGCTATC	K V K	G H N S rbs GATGCAGCTC	top <u>MK</u> AACATATCAC	2080
2001	N R N AAAATAGTCA K I V K	Y T V K AATCTAGAGA S R E	Y E V AGTTACATCA VT S	N W K ATTGCATTGC I A L L	T H E I TTTTGCTATC L L S	K V K CAATACACTT N T L	G H N s rbs GATGCAGCTC D A A Q	top <u>M_K</u> AACATATCAC H I T	2080
2001	N R N AAAATAGTCA <u>K I V K</u> ACCTGTAAGT	Y T V K AATCTAGAGA S R E GAGAAAAAAGG	Y E V AGTTACATCA V. T S TTGATGATAA	N W K ATTGCATTGC I A L L AATTACTTTG	T H E I TTTTGCTATC L L S TACAAAACAA	K V K CAATACACTT <u>N T L</u> CTGCAACATC	G H N S CDS GATGCAGCTC D A A Q AGATTCCGAT	top <u>MK</u> AACATATCAC H I T AAGTTAAAAA	2080 2160
2001 2081	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S	Y T V K AATCTAGAGA S R E GAGAAAAAAGG E K K V	Y E V AGTTACATCA V. T S TTGATGATAA D D K	N W K ATTGCATTGC I A L L AATTACTTTG I T L	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T	K V K CAATACACTT N T L CTGCAACATC A T S	G H N S cbs GATGCAGCTC D A A Q AGATTCCGAT D S D	TOP <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I	2080 2160
2001 2081 2161	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT	Y T V K AATCTAGAGA S R E GAGAAAAAAGG E K K V TTTAACTTTT	Y E V AGTTACATCA V. T S TTGATGATAA D D K AATTTTATTA	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA	G H N S Cbs GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC	AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC	2080 2160 2240
2001 2081 2161	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F	Y E V AGTTACATCA V. T S TTGATGATAA D D K AATTTTATTA N F I K	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I	G H N S cbs GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N	2080 2160 2240
2001 2081 2161 2241	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG	Y T V K AATCTAGAGA <u>S R E</u> GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA	Y E V AGTTACATCA V. T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTAGTTCTCA	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG	G H N s cbs GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT	AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC	2080 2160 2240 2320
2001 2081 2161 2241	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTT L T F GCTATACAAA Y T K	Y E V AGTTACATCA V. T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA P N P	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTAGTTCTCA S S Q	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTATTGG F Y W	G H N s GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S	2080 2160 2240 2320
2001 2081 2161 2241 2321	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG	Y E V AGTACATCA V. T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA P N P ACTCAGTAAA	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTAGTTCTCA S S Q TATGCACCTA	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA	G H N s GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y TGAAGAATTT	top \underline{M} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC	2080 2160 2240 2320 2400
2001 2081 2161 2241 2321	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S	Y T V K AATCTAGAGA S R E GAGAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D	Y E V AGTTACATCA V. T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA P N P ACTCAGTAAA S V N	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT V V D	T H E I TTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTAGTTCTCA S S Q TATGCACCTA Y A P K	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTITATTGG F Y W AAAATCAAAA N Q N	G H N S GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y TGAAGAATT E E F	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC Q V Q Q	2080 2160 2240 2320 2400
2001 2081 2161 2241 2321 2401	N R N AAAATAGTCA K I V K ACCTGTAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGTAGG	X T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTTAT	Y E V AGTACATCA V.T S TICATGATAA D D K AATTITATTA N F I K GCCAAATCCA P N P ACTCAGTAAA S V N GGTGGAGATA	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTAGTTCTCA S S Q TATGCACCTA Y A P K TAACGGCTTG	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTITATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG	G H N s GATGCAGCTC D A A Q AGATTCCAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y TGAAGAATT E E F GTAATGGTTC	top MK AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q AAAATCTTTT	2080 2160 2240 2320 2400 2480
2001 2081 2161 2241 2321 2401	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGTAGG T V G	Y T V K AATCTAGAGA S R E. GAGAAAAAGG E K K V TTTAACTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTAT Y S Y	$\begin{array}{cccc} Y & E & V \\ A & A & A & C & A \\ V & T & S \\ T & T & G & T & A \\ D & D & K \\ A & A & T & T & T & A \\ T & A & T & T & A \\ N & F & I & K \\ G & C & A & A \\ S & V & N \\ G & G & G & C & A \\ G & G & G & D & I \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S	$\begin{array}{cccc} T & H & E & I \\ TTTTGGTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T & T \\ TATGATAAA \\ Y & D & K \\ TTAGTTGTA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ TAACGGCTTG \\ N & G & L \\ \end{array}$	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTITATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G	G H N s GATGCAGTC GATGCAGTC D A. A Q AGATTCCGAT D S D TATCCAAAGC L K A GGTTCTAAGC C S K Y TGAAGAATTT E E F GTAATGGTTC N G S	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAA A G N ACAACATTC N I S CAAGTACAAC Q V Q Q AAAATCTTT K S F	2080 2160 2240 2320 2400 2480
2001 2081 2161 2241 2321 2401 2481	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGTAGG T V G TCAGAGACAA	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TITAACTITT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTAATCTATA Y S Y TTAACTATAA	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N \\ ACTCAGTAAA \\ S & V \\ N \\ GGTGGGAGATA \\ G & G & D & I \\ ACCAGGAAATGC \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTCC N I S TATAGAACTA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTAGTTCTCA S S Q TATGCACCTA Y A P K TAAGGGCTTG N G L GCTTAGGATAA	K V K N T L CTGCAACATCA A T S GATACATTAA D T L I ATTTTATTGG F Y W ANAATCAANA N Q N TCAGGTGGAG S G G G AAGAACTAAT	G H N s cbs GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAG G S K Y TGAAGAATT E E F GTAATAGGTCC N G S TCCAAAAAAA	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q AAAATCTTTT K S F TIGGTTGGGA	2080 2160 2240 2320 2400 2480 2560
2001 2081 2161 2241 2321 2401 2481	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGTAGG T V G TCAGAGACAA S E T I	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTAT J T K GATTCTAATG D S N D S N J S N TAATCTTAT Y S Y TTAACTATAA N Y K	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & K \\ AATTTTATA \\ N & F & I & K \\ GCCAAATCCA \\ F & N & P \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & G & D & I \\ ACAAGAAAGC \\ Q & E & S \end{array}$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TATAGAACTA Y R T S	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACARAACAA \\ Y & K & T & T \\ TATGGTARAA \\ Y & D & K \\ TAGTTCTCA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ A & P & K \\ TAACGGCTTG \\ N & G & L \\ GCTTAGATAA \\ L & D & K \\ \end{array}$	K V K CAATACACTT N T L CTGCAACATC GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AAGAACTAAT R T N	G H N S GATGCAGCTC D A. A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y TGAAGAATTT E E F GTAATGGTTC N G S TTCAAAAAAA	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGGTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q V Q Q V Q Q V Q Q AAAATCTTTT K S F TTGGTTGGGAA G W D	2080 2160 2240 2320 2400 2480 2560
2001 2081 2161 2241 2321 2401 2481 2561	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I N S AATTAATTCA I N S AAACGGTAGG T V G TCAGAGACAA S E T I GTTGGAAGCA	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTAATG Y S Y TTAACTATAA N Y K CATAAAATTA	Y E V AGTTACATCA V.T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA P N P ACTCAGTAAA S V N GGTGGAGATAA G G D I ACAAGAAAGC Q E S TGAATAATGG	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAAT V V D TTAATATCTC N I S TATAGAACTA Y R T S TATGGGAGCACA	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T & T \\ TATGATAA \\ Y & D & K \\ TATGCACCTA \\ Y & A & P & K \\ TAACGGCTG \\ N & G & L \\ GCTTAGATAA \\ L & D & K \\ TATGCGACGAG \\ \end{array}$	K V K CAATACACTT N T L CTGCAACATCA A T S GATACATTAA D T L I ATTITATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AACAACTAAT R T N ATAGTTATCA	G H N S Cbs GATGCAGCTC D A. A Q AGATTCCGAT D S D TATCCAAAGC L K A GGTTCTAAGC L K A GGTS K Y TGAAGAATTT E E F GTAATGGTTC N G S TTCAAAAAAA F K K I I TTCAACTAT	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q AAAATCTTT K S F TTGCTGGGA G W D GGTAATGAAA	2080 2160 2240 2320 2400 2480 2560 2560
2001 2081 2161 2241 2321 2401 2481 2561	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCT ATTTATTCT I Y S G AATTAATTCA I N S AAACGGTAGG T CAGAGCAA S E T I TGTTGAAGCA V E A	AATCTAGAGA S.R.E. GAGAAAAAGG E.K.K.V TTTAACTTTT L.T.F. GCTATACAAA Y.K. GATTACTAATG D.S.N.D TTATTCTAT Y.S.Y TTAACTATAAA N.Y.K. CATAAAATTA H.K.I.M.	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCA \\ P & N & F & I \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & D & I \\ ACCAGGAAAGC \\ Q & E & S \\ TGAATAATCG \\ N & N & G \end{array}$	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAAGACACATA K D T I CGTTGTAGAT V V D TTAATATCTCC N I S TATAGAACATA Y R T S TTGGGGACCA W G P	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACARAACAA \\ Y & K & T \\ TTATGATAAA \\ Y & D & K \\ TAGTCTCCA \\ S & Q \\ TATGCACCTA \\ X & A & P \\ K \\ TAACGGCTTG \\ N & G & L \\ GCTTAGATAA \\ L & D & K \\ TATGGCAGAG \\ Y & G & R \\ D \\ G & R & D \end{array}$	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AAGAACTAAT R T N ATAGTTATCA S Y H	G H N S cbs GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAG G S K Y TGAAGGATTT E E F GTAATGGTTC N G S TTCAAAAAAA F K K I TTCAACTTAT S T Y	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q AAAATCTTTT K S F TTGCTTGGGA G W D GGTAATGAAA G N E M	2080 2160 2240 2320 2400 2480 2560 2640
2001 2081 2161 22241 2321 2401 2481 2561 2561	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGTTAGG T V G TCAGAGACAA S E T I TGTTGGAAGCA V E A TGTTTTAGG	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATCTTAT Y S Y TTAACTATAA N Y K CATAAAATTA H K I M CTCAAGACAA	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATATA \\ D & D & K \\ AATTTATATA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCCAGTAAA \\ G & C & D & I \\ ACAGGAAAGC \\ Q & E & S \\ TGAATAATCG \\ N & N & G \\ AGCAACTTAA \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAG D K S AAGACACTA K D T I CGTTGTACAT V V D TTAATATCTC N I S TATAGGACTA Y R T S TTGGGGACA W G P ATGCTGGGACA	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACARAACAA \\ Y & K & T & T \\ TTATGATAAA \\ Y & D & K \\ TAGTTCTCA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ A & P & K \\ CTTAGGACGTTG \\ N & G & L \\ GCTTAGGACAG \\ L & D & K \\ TATGGCACAG \\ Y & G & R & D \\ AAACTTCTTG \\ \end{array}$	$\begin{array}{c} \mathbf{K} \mathbf{V} \mathbf{K} \\ \mathbf{K} \mathbf{V} \mathbf{K} \\ \mathbf{C} \\ $	G H N S CATGCAGCTC D A A Q AGATTCCGAT D S D TATCCAAAGC L K A GGTTCTAAGT C S K Y TGAAGGATTT E E F GTAATGGTTC N G S TTCAAAAAAA TTCAACTTAT S T Y AAATGCCAGT	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACTAC Q V Q Q V Q Q V Q Q V Q Q V Q Q V Q Q GCTATCCAGA	2080 2160 2240 2320 2400 2480 2560 2640 2720
2001 2081 2161 2241 2321 2401 2481 2561 2641	$\label{eq:response} \left\{ \begin{array}{ccc} N & R & N \\ N & R & N \\ \end{array} \right.$	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TITAACTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTAATG Y S Y TTAACTATAA N Y K CATAAAATTA H K I M CTCAAGACAA S R Q	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & G & D & I \\ ACAAGAAAGC \\ Q & E & S \\ TGAATAATGG \\ N & N & G \\ AGCAACTTAA \\ S & N & L & N \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTTG I T L AAGATAAAAG D K S AAAGACACTA K D T I CGTTGTAGAAT V V D TTAATATCTC N I S TATGAACTA S TTGGGGACCA W G P ATGCTGGACA A G Q	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T & T \\ TTATGATAAA \\ Y & D & K \\ TTAGTCTCA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ TAACGGCTTG \\ N & G & L \\ GCTTAGATAA \\ L & D & K \\ TATGCACAGAG \\ Y & G & R & D \\ AAACTTCTTG \\ N & F & L \end{array}$	K V K CAATACACTT T L CTGCAACATC A T GATACATTAA D T L A T S G GATACATTAA D T L I T ATTTTATTGG F Y W ANAATCAANA N Q N TCAGGTGGAG G G AAACAAAA N Q N TCAGGTGAAG R T N AAACATATATCA ATAGTATATCACA S Y H GAATATCACA S Y H	G H N S cbs GATGCACGCTC D A A Q AGATCCGAT D S D TACTCAAAGC L K A GGTCTAAGT G S K Y TGAAGAATT CAAAAAAAA F K K I S TTCAAAAAAA F K K I AAATGCCAGT M P V	top <u>M_K</u> AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC Q V Q Q AAAATCTTT K S F TTGGTTGGGA G W D GGTAATGAAA G N E M GTTATCCAGA L S R	2080 2160 2240 2320 2400 2480 2560 2640 2720
2001 2081 2161 2241 2321 2401 2481 2561 2561 2641 2721	$\begin{array}{c c} N & R & N \\ N & R & N \\ \end{array} \\ \begin{array}{c c} A \\ A \\ A \\ A \\ C \\ C \\ C \\ C \\ C \\ C \\$	AATCTAGAGA S. R. E. GAGAAAAAGG E. K. K. V. TITAACTITT L. T. F. GCTATACAAAA Y. T. K. GATTCTAATG D. S. N. D. TITATTCTATT Y. S. Y. TTAACTATAAAATTA N. Y. K. CATAAAAATTA H. K. I. M. CTCAAGACAA S. R. Q. ATCCAAGAATT	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCGAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & G & D & I \\ ACCAAGAAAGC \\ Q & E & S \\ TGAATAATCG \\ N & N & G \\ AGCAACTTAA \\ S & N & L \\ N & L \\ N & L \\ TATTGGTGTC \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAAGACACATA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TATAGAACATA Y R T S TTGGGGACCA W G P ATGCTGGACA A G Q CTATCTGGAA	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACARAACAA \\ Y & K & T \\ TATGATAAA \\ Y & D & K \\ TAGTTCTCA \\ S & Q \\ TATGCACTA \\ Y & A & P & K \\ TAACGGCTTG \\ N & G & L \\ GCTTAGATAA \\ L & D & K \\ TATGGCAGAG \\ V & G & R & D \\ AAACTTCTTG \\ N & F & L \\ AACAAAACGC \\ \end{array}$	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AACAACTAAT R T N ATAGTTATCA S Y H GAATATCACA E Y H K	G H N S CATGCAGCTC GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y TGAAAGGATTT E E F GTAATGGTTC N G S TTCAAAAAAA F K K I TTCAACTTAT S T Y ANATGCCAGT M P V TCAAAAATA	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q V AAAATCTTTT K S F TTGGTTGGGA G W D GGTAATGAAA G N E M GTTATCCAGA L S R CTGTTACCAGA	2080 2160 2240 2320 2400 2480 2560 2560 2640 2720 2800
2001 2081 2161 2241 2321 2401 2561 2561 2561 2641 2721	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGTAGTAGG T V G TCAGAGCAA S E T I TGTTGTAGCAG F L G GGTAACTTCA G N F N	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TATTCTATT Y S Y TTAACTATAA N Y K CATAAAATTA H K I M K CATAAAATTA H K I M CTCAAGACAA S R Q ATCCAGGAATT P E F	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCCAGTAAA \\ C & C & D & I \\ ACAGGAAAGCA \\ C & E & S \\ TGAATAATCG \\ Q & E & S \\ TGAATAATCG \\ N & N & G \\ ACCAACATTAA \\ S & N & L & N \\ TATTGGTGTC \\ I & G & V \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAG D K S AAGACACTA K D T I CGTTGTACAT V V D TTAATATCTC N I S TTAGGGACCA W G P ATGCTGGGACA A G Q CTATCTCGAA L S R K	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACARAACAA \\ Y & K & T & T \\ TTATGATAAA \\ Y & D & K \\ TAGTCTCCA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ A & P & K \\ CTTAGGCACGA \\ K & G & L \\ GCTTAGCTAA \\ L & D & K \\ TATGGCACAG \\ Y & G & R \\ D & AACTTCTTG \\ N & F & L \\ AACARAACGC \\ Q & N & A \end{array}$	$\begin{array}{cccc} K & V & K \\ & & & & & \\ & & & & & \\ & & & & &$	G H N S CATGCAGCTC D A A Q AGATTCCGAT D S D TATCCAAAGC L K A GGTTCTAAGT E E F GTAATGGTTC N G S TTCAAAAAATA S T Y AAATGCCAGT M P V TCAAAAAATA	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACTACA Q V Q Q Q V Q Q Q V Q Q V Q Q V Q Q V Q Q AAAATCTTTT K S F TTGGTTGGGAA G W D GGTAATGAAA G N E M D GGTAATCCAGA L S R CTGTTACTATA V T Y	2080 2160 2240 2320 2400 2480 2560 2560 2640 2720 2800
2001 2081 2161 2241 2321 2401 2481 2561 2561 2561 2721 2801	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGTAGG T V G TCAGAGGACAA S E T I GTTGAACACAA S E T I GGTAACTCA G N F N	TAACTAGAGA AATCTAGAGA S. R. E. GAGAAAAAGG E. K. K. V. TITAACTITT L. T. F. GCTATACAAA Y. T. K. GATTCTAATG D. S. N. D. TTAACTATAA N. Y. K. GATTCAATAANATAA N. Y. K. CATAAAAATAA H. K. I. M. CTCAAGACAA S. R. Q. ATCCAGAATT P. E. ATGGATAGAT	$\begin{array}{c} Y \ E \ V \\ AGTTACATCA \\ V \ T \ S \\ TTGATGATAA \\ D \ D \ K \\ AATTTTATTA \\ D \ D \ K \\ AATTTTATTA \\ N \ F \ I \ K \\ GCCARATCCA \\ P \ N \ P \\ ACTCGAGTAAA \\ S \ V \ N \\ GGTGGAGATA \\ G \ G \ D \ I \ I \\ ACAGGAAACC \\ Q \ E \ S \\ TGAATAATCG \\ N \ G \\ AGCAACTTAA \\ S \ N \ L \ N \\ TATTGGGTGTC \\ I \ G \ V \\ ATACAAACTT \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAGGACACTA K D T I CGTTGTAGAT V V D TTAATATCCC N I S TATAGGAGCAC X R T S TATGGGGACCA A G Q CTATCTGGAA L S R K TTGGATCAAC	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T \\ TTATGATAAA \\ Y & D & K \\ TTAGTTCTCA \\ S & Q \\ TATGCACCTA \\ X & A & P \\ K \\ TAAGGGCTTG \\ N & G \\ L \\ GCTTAGGATAA \\ L & D \\ K \\ TATGGCACAA \\ L & D \\ K \\ TATGGCACAA \\ R & D \\ AAACTTCTTG \\ N & F \\ L \\ AACAAAACGC \\ Q & N \\ TTCGACTGGA \\ \end{array}$	K V K K CATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I A T S GATACATTAA D T L I A ATTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S AAGAACTAAT R T N AAGAACTAAT N AAGAACTAAT K T N AAGAACTAAT N Q N TAGGTAATAAA K K TAGGTAATAAA K K TAGGTAATAAA K K TAGGTAAAAAAA K K TAGGTAATAAA T N T	G H N S cbs GATGCACGTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTCTAAGT G S K Y TGAAGAATT E E F GTAATAGGTTCC N G S TTCCAAAAAAA F K K I TCAAAAATTA S K I TATAAAGAT TATAAAGAT	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC G V Q Q AAAATCTTT K S P TIGGTTGGGA G W D GGTAATGAAA G N E M GTTATCCAGA L S R CTGTTACTTA V T Y CACATAACAG	2080 2160 2240 2320 2400 2480 2560 2640 2720 2800 2880
2001 2081 2161 2241 2321 2401 2481 2561 2561 2641 2721 2801	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTATT Y S Y TTAACTATAAA N Y K CATAAAATTA H K I M CTCAAGACAA S R Q ATCCAGAATT P E F ATGGATAGAT	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & G & D & I \\ ACCAGAAAAGC \\ Q & E & S \\ TGAATAATGG \\ N & N & G \\ AGCAACTTAA \\ S & N & L & N \\ TATTGGTETC \\ I & G & V \\ ATACAAACTT \\ T & N & F \end{array}$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAG D K S AAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TTGGGGCCA W G P ATGCTGGACA A G Q CATCCTGGAC L S R K TTGGATCAAL N S R K TTGGATCAAL	$\begin{array}{ccccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T \\ TATGATAAA \\ Y & D & K \\ TTATGATAAA \\ Y & D & K \\ TATGCACTA \\ S & Q \\ TATGCACTA \\ Y & A & P \\ K \\ TATGCACTA \\ L & D \\ K \\ TATGCCAGA \\ V & G & L \\ GCTTAGATAA \\ L & D \\ K \\ TATGCCAGA \\ N & F & L \\ AACAAAACTCT \\ Q & N \\ TCAACTGGA \\ C & N \\ F & N \\ TCAACTGGA \\ K \\ N & F \\ N & K \\ T \\ TCAACTGGA \\ N & K \\ T \\ N & M \\ M \\ T \\ N & M \\ M$	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AAGAACTAAT R T N ATAGTTATCA S Y H GAATATCACA E Y H K TGCAAAAAAA A K K TAGCTAATAA G N N	G H N S CATGCAGCTC GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT G S K Y TGAAAGGATTT E E F GTAATGGTTC N G S TTCAAAAAAA F K K I TTCAAAAAAT S T Y AAATGCCAGT M P V TCAAAAATTA S K I T TTATAAAGAT Y K D	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q V Q AAAATCTTTT K S F TTGGTTGGGA G W D GGTAATCAAA G W D GGTAATCAAA L S R CTGTTATCCAGA L S R CTGTTATCCAGA L S R CTGTTATCTAA	2080 2160 2240 2400 2480 2560 2560 2720 2800 2880
2001 2081 2161 2241 2321 2401 2481 2561 2561 2641 2721 2801 2881	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TITCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGTAGG T V G TCAGAGACAA V E A TGTTTTAGG F L G GGTAACTTCA Q N F N TCAAAGGAA	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TITAACTITT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTAATCTTAAT Y S Y TTAACTATAA N Y K CATAAAATTA H K I M CTCAAGACAA S R Q ATCCAGAATT P E F ATGGATAGAT	$\begin{array}{cccc} Y & E & V \\ & & & & & & \\ & & & & & \\ & & & &$	N W K ATTGCATTGC I A. L L AAATTACTTG I T L AAGATAAAAAG D K S AAAGACAATA K D T I CGTTGTACAT V V D TTAATATCTC N I S TATAGAACTA W G P ATGCTGGGACA A G Q CTATCTCGAA A G Q CTATCTCGAA L S R K TTGGATCAAC W I N GGGAAAAATCA	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T & T \\ TTATGATAAA \\ Y & D & K \\ TAGTCTCAA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ TAACGCCTTG \\ N & G & L \\ GCTTAGATAA \\ L & D & K \\ TATGGCAGAG \\ Y & G & R \\ D & AACTTCTTG \\ N & F & L \\ AACAAAACGC \\ Q & N & A \\ TTCAACTGA \\ F & W & I \\ TACGGTAAA \\ \end{array}$	K V K CAATACACTT N T L CTGCAACACTC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCACGTGGAG S G G G AAGAACTAAT R T N ATAGTTATCA E Y H K TGCAAAAAAA A K K TAGCTAATAA G N N TAATAGATA	G H N S CATGCAGCTC D A. A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAGT C S K Y TGAAGAAATTT E E F GTAATGGTTC N G S TTCAAAAAATA S T Y AAATGCCAGT M P V TCAAAAAATTA AAATGCCAGT M P V TCAAAAAATTA S K I T TTATAAAAATTA Y K D CTCAATCTAA	top \underline{M} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACTAC Q V Q Q Q V Q Q AAAATCTTTT K S F TTGGTTGGGA G W D GGTAATCAAA G N E M GTTATCCAGA L S R CTGTTACTAA V T Y CACATAAGAG H I R A GGAAAAAAAT	2080 2160 2240 2320 2400 2480 2560 2640 2720 2800 2880 2880 2960
2001 2081 2161 2241 2321 2481 2561 2561 2721 2801 2881	$\begin{array}{c c} N & R & N \\ N & R & N \\ \end{array} \\ \begin{array}{c c} A \\ A \\ A \\ C \\ T \\ C \\ T \\ C \\ C \\ C \\ C \\ C \\ C$	$ \begin{array}{cccc} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{Y} \mathbf{T} \mathbf{V} \mathbf{K} \\ \mathbf{A} \mathbf{A} \mathbf{X} \mathbf{C} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{G} \mathbf{A} \\ \mathbf{S} \mathbf{R} \mathbf{E} \\ \mathbf{G} \mathbf{A} \mathbf{G} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{G} \\ \mathbf{E} \mathbf{K} \mathbf{K} \mathbf{V} \\ \mathbf{T} \mathbf{T} \mathbf{A} \mathbf{A} \mathbf{C} \mathbf{T} \mathbf{T} \mathbf{T} \\ \mathbf{L} \mathbf{T} \mathbf{F} \\ \mathbf{G} \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \mathbf{C} \mathbf{A} \mathbf{A} \\ \mathbf{G} \mathbf{X} \mathbf{T} \mathbf{C} \mathbf{A} \mathbf{A} \\ \mathbf{G} \mathbf{A} \mathbf{T} \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \\ \mathbf{T} \mathbf{T} \mathbf{X} \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \\ \mathbf{N} \mathbf{Y} \mathbf{K} \\ \mathbf{G} \mathbf{A} \mathbf{T} \mathbf{C} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \\ \mathbf{N} \mathbf{X} \mathbf{K} \\ \mathbf{G} \mathbf{A} \mathbf{T} \mathbf{C} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \\ \mathbf{K} \mathbf{I} \mathbf{M} \\ \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \\ \mathbf{K} \mathbf{I} \mathbf{M} \\ \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \\ \mathbf{K} \mathbf{I} \mathbf{M} \\ \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \\ \mathbf{K} \mathbf{I} \mathbf{M} \\ \mathbf{C} \mathbf{T} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \mathbf{A} \\ \mathbf{K} \mathbf{I} \mathbf{M} \\ \mathbf{K} \mathbf{I} \mathbf{K} \\ \mathbf{K} \mathbf{K} \mathbf{K} \\ \mathbf{K} \mathbf{K} \mathbf{K} \\ \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \\ \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \\ \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \\ \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K}$	$\begin{array}{c} Y \ E \ V \\ AGTTACATCA \\ V \ T \ S \\ TTGATGATAA \\ D \ D \ K \\ AATTTTATTA \\ D \ D \ K \\ AATTTTATTA \\ N \ F \ I \ K \\ GCCAAATCCA \\ P \ N \ P \\ ACTCAGTAAA \\ S \ V \ N \\ GGTGGAGATA \\ G \ C \ D \ I \\ ACAAGAAAGC \\ Q \ E \ S \\ TGAATAATCG \\ N \ N \ G \\ AGCAACTTAA \\ S \ N \ L \ N \\ TATTGGTGTC \\ I \ G \ V \\ ATACAAACTT \\ T \ N \ F \\ GAAGTGATTGAT \\ T \ N \ F \\ GAAGTGATTGAT \\ T \ N \ F \\ GAAGTGGTGATT \\ T \ N \ F \\ GAAGTGGTGGTG \\ T \ N \ F \\ GAAGTGGTGGTG \\ T \ N \ F \\ GAAGTGGTGGTG \\ T \ N \ F \\ GAAGTGGTGATT \\ T \ N \ F \\ GAAGTGGTGGTG \\ T \ N \ F \\ GAAGTGGTGATT \\ T \ N \ F \\ GAAGTGGTGGTG \\ T \ N \ F \\ GAGTGGTGGTG \\ T \ N \ F \\ T \ N \ F \\ T \ N \ F \\ T \ T \ T \ T \ T \ T \ T \ T \ T \ T$	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAGGACACTA K D T I CGTTGTAGAT V V D TTAATATCCC N I S TATAGAACTA Y R T S TATGGGGACAA W G P ATGCTGGACA A G Q CTATCTGGAA L S R K W I N GGGAAAATCA E N H	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T \\ TTATGATAAA \\ Y & D & K \\ TAGTCTCAA \\ S & Q \\ TATGCACTA \\ S & Q \\ TATGCACTA \\ C & C \\ CTTAGGATAA \\ L & D & K \\ CTTAGGCACGA \\ G & C \\ G & C \\ CTTAGGCACAA \\ L & D & K \\ C & C \\ C & $	K V K CATACACTT T L CTGCAACATC A T A T L I CTGCAACATC A T S GATACATTAA D T L I A ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGAGG S G G G G AAGAACTAAT R T N AAGAACTAAT S Y H GAAGAACTAAT R CAAGAACAAAAAA S Y H GAAGAACTAAT G N A TAGGTAAAAAAAA A K K TAGGTAATAAA G N N TAAATAGAATA L I D T	$ \begin{array}{cccc} G & H & N & s \\ cbs \\ cbs$	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC G V Q Q AAAATCTTT K S F TTGGTTGGGA G W D GGTAATGGAAA G N E M GTTATCCAGA L S R CTGTTACTTA V T Y CACATAACAG H I R A GGAAAAAAAT E K N	2080 2160 2240 2400 2400 2560 2560 2640 2720 2800 2880 2960
2001 2081 2161 2241 2321 2401 2481 2561 2561 2721 2881 2881 2961	$\begin{array}{c c} N & R & N \\ \hline N & R & N \\ \hline A \\ A \\ A \\ A \\ C \\ C \\ T \\ T \\ C \\ T \\ C \\ C \\ C \\ C$	AAACAGATAG Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAAA Y T K GATTCTAATG D S N D TTATTCTATT Y S Y TTAACTAGAAATTA N K CATAAAAATTA N K CATAAAAATTA N K CATAAAAATTA N K CATAAAAATTA N CATCAAGAAATG ATCCAGAATG M D R Y AAACAGATAGATAG	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & G & D & I \\ ACCAGAAAAGC \\ Q & E & S \\ TGAATAATCA \\ S & N & L & N \\ GCCAAACTTA \\ S & N & L & N \\ TATTGGTETC \\ I & G & V \\ ATACAAACTT \\ T & N & F \\ GAAGTTGATT \\ E & V & D & W \\ ATAATCAAAA \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAAG D K S AAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TTGGGGCCA A G Q ATGCTGGACA L S R K TTGGATCAAL S R N H N GGGAAAATCA W I N	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TATGATAAA Y D K TTATGATAAA Y D K TAACTGCACTA S S Q TATGCACCTA X A P K TAACGGCTTG N G L GCTTAGATAA L D K TATGCAGAA N F L AACAAAACGC Q N A TTCAACTGAA T V K ATGTTAAAAT	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G MAGAACTAAT R T N ATAGTTATCA S Y H GAATATCACA E Y H K TGCAAAAAAA A K K TAGCAAACAA G N N TTACAAACAC	$ \begin{array}{cccc} G & H & N & s \\ \mbox{cbs} \\ \mbox{gatgcagetc} \\ $	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q V Q AAAATCTTTT K S F TTGGTTGGGA G W D GGTAATCAAA G W D GGTAATCAAA G N E M GTTATCCAGA L S R CTGTTATCCAGA L S R CTGTTATCCAGA L S R CTGTTATCCAGA L S R CACATAAGAG K H I R A GGAAAAAAAT E K N TATTAGGGTA	2080 2160 2240 2320 2480 2560 2640 2720 2880 2880 2880 2960 3040
2001 2081 2161 2241 2321 2401 2481 2561 2641 2721 2801 2881 2961	N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGTAGTAGG T V G TCAGAGACAA V E A TGTTTTAGG F L G GGTAACTTCA Q N F N TCAAAGTGAA Q S E CAACTCATAC T H T CCTATGAGCCT P M S st	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTAATG D S N D TTATTCTTAT Y S Y TTAACTATAA N Y K CATAAAATTA H K I M CTCAAGACAA S R Q ATCCAGAATTAT P E F ATGGATAGAT M D R Y AACAGATAG S I Y AAACAGATAG Op	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCAGTAAA \\ S & V \\ N & GGTGGAGATA \\ G & C & D & I \\ ACAAGAAAGC \\ Q & E & S \\ TGAATAATCAAAAT \\ S & N & L \\ N & TATTGGTGTC \\ I & G & V \\ ATACAAACTT \\ T & N & F \\ GAAGTTGATT \\ E & V & D & W \\ ATAATCAAAA \\ \end{array}$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAAG D K S AAAGACACAA K D T I CGTTGTACAT V V D TTAATATCTC N I S TATAGGACTA W G P ATGCTGGACA A G Q CTATCTGGACAA A G Q CTATCTGGACAA L S R K TTGGACACAAC W I N GGGAAAAATCA E N H AATCTTAAAT	$\begin{array}{cccc} T & H & E & I \\ TTTTGCTATC \\ L & L & S \\ TACAAAACAA \\ Y & K & T & T \\ TATGATCAA \\ Y & D & K \\ TAGTCTCA \\ S & S & Q \\ TATGCACCTA \\ Y & A & P & K \\ A & C & C \\ CTTAGCACCTA \\ L & D & K \\ TATGGCACAG \\ Y & G & R & D \\ AAACTTCTTG \\ N & F & L \\ AACAAAACG \\ Q & N & A \\ TTCAACTGAAAA \\ F & N & W & I \\ TACAGTTAAAAT \\ T & V & K \\ ATGTTAAAAT \\ \end{array}$	K V K CAATACACTT N T L CTGCAACATCC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCACGTGGAG S G G G AAGAACTAAT R T N ATAGTTATCA E Y H K TGCAAAAAAA A K K TAGCTAATAA A K K TAGCTAATAA L I D T TTACAAAACAC	G H N S cbs GATCCAACTC DA A Q AGATCCGAT D S D TACTCAAAGC L K A GGTCTAAGT G S K Y TGAAGAATT E E F GTAATGGTTC N G S TTCAAAAAAA F K K I TTCAACATAT S T Y AAATGCCAGT M P V TCAAAAATTA S K Y TCAAAAATTA S K Y TCAAAAATTA Q S K TTICTTCTA	top \underline{M} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q AAAATCTTT K S P TTGGTTGGGA G W D GGTAATCAAA G N E M GTTATCCAGA L S R CTGTTACTAA V T Y CACATAAGAG H I R A GGAAAAAACA	2080 2160 2240 2320 2400 2480 2560 2640 2720 2800 2800 2800 2960 3040
2001 2081 2161 2241 2321 2401 2481 2561 2641 2721 2801 2881 2961 3041	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGCTAGG F L G GGTAACGCAGCA S E T I TGTTGAAGCA V E A TGTTTTTAGG F L G GGTAACTCA G N F N CCAACTGCAAC Q S E CAACTCATAC C T H T CCTATGACCT T H T CCTATGACCT	AATCAGAAAAG S. R. E. GAGAAAAAG E. K. K. V. TITAACTITT L. T. F. GATTACAAA Y. T. K. GATTCTAATG D. S. N. D. TIAATCTAATA N. Y. K. CATAAAATTA N. Y. K. CATAAAAATTA N. Y. K. CATAAAATTA N. Y. K. Y.	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCA \\ P & N & F \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & C & D & I \\ ACCAGGAAAGC \\ Q & E & S \\ TGAATAATGG \\ N & N & G \\ AGCAACTTAA \\ S & N & L \\ N & TATTGGTGTC \\ I & G & V \\ ATACAAACTT \\ T & N & F \\ GAAGTGAATTTTTT \\ T & M \\ TAATCAAAACTT \\ T & M \\ TTAATCAAAACTT \\ T & M \\ TTAATCAAAAACTT \\ T & M \\ TTAATCAAAAACTT \\ T & M \\ TTATTTTTT \\ T & M \\ TTATTTTTT \\ T & M \\ TTATTTTTT \\ T & M \\ TTATTTTTTTT \\ T & M \\ T $	N W K ATTGCATTGC I A L L AATTACTTGG I T L AAGATAAAAG D K S AAGGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TATGGAGACTA Y R T S TATGGGACAA M G P ATGCTGGACA A G Q CTATCTGGAA L S R K TTGGAGCAAAATCA B N H AATCTTAAAT CAGGGCAAAA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TAGTTCTCA S S Q TATGCACCTA TAGGCACTG N G L GCTTAGGCAGGA CAGACAAACGC Q N A TAGGCAGAA F N W I TACAGGTAAAA F N W I TACAGGTAAAA T V K ATGTTAAAAA	K V K K X K CTGCAACATC A T L T CTGCAACATC A T L I ATTTTATTGG F Y W AATATTATTGG F Y W AAAAATCAAAA N Q N TCAGGTGGAG G G G S G G G AAGAACTAAT R T N AAGAACTAAT S Y H GAATAACTAAAA A K K TGCAAAAAAAA A K K AGGTAATAAA G N N TTAACAAAAAAA L I D T TTACAAACAC TTATTTAAAT T K	G H N S cbs GATGCAGGCTC D A A Q AGATTCCGAT D S D TACTCAAAGC L K A GGTTCTAAG G S K Y TGAAGAATT E E F GTAATAGGTCC N G S TTCAAAAAAAA F K K I TTCAACTAT M P V TCAAAAATTA S T Y AAATGCCAGT M P V TCAAAAATTA S K I T TATAAAGT Y K D CTCAAAAATTA S K I T TATAAAGAT Y K D CTCAAAAATTA S K I T TATAAAGAT	top \underline{M} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC G V Q Q AAAATCTTTT K S P TIGGTTGGGA G W D GGTAATGAAA G N E M GTTATCCAGA L S R CTGTTACTA V T Y CACATAACAG H I R A GGAAAAAA T ATTAGGGTA ACTTGTGGAA	2080 2160 2240 2320 2480 2560 2640 2720 2800 2800 2960 3040 3120
2001 2081 2161 2241 2321 2401 2481 2561 2721 2881 2881 2961 3041 3121	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TITCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I Y S G AATTAATTCA I Y S G AATTAATTCA S E T I TGTTGTAGCA G N F N TCACAGACAA S E T I TGTTGTAGCA G N F N TCACAGTCTTCA G N F N TCACAGTCATT H T CCTATGAGCCA	AAACASATA Y T V K AAACTAGAGA S R E GAGAAAAAGG E K K V TTTAACTATT I T F GCTATACAAAA Y T K GATTCTAATG D S N U S N TAACTATAAAATTA N Y K CATAAAAATTA N Y K CATAAAAATTA N Y K CATAAAAATTA N Y K CATAAAAATTA N Y K CATAAAAATTA S R Q AATCCAGATAGATA S R Q AAACAGAATAG OD AATCGACGTG GTTAAAAAATA	$\begin{array}{cccc} Y & E & V \\ A & GTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ A & ATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ F & N & P \\ ACTCAGTAAA \\ S & V & N \\ GGTGGAGATA \\ G & G & D & I \\ C & ACAAGAAAGC \\ Q & E & S \\ TGAATAATGGTC \\ I & G & V \\ ATACAGAACTT \\ T & N & F \\ GAAGTTGATT \\ T & N & F \\ GAAGTTGATT \\ T & N & F \\ ATAATCAAAA \\ TATAAGAACT \\ T & ATAAACAAACT \\ T & T & T & T \\ T & T & T & T \\ T & T &$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAAG D K S AAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TTGGGGACCA W G P ATGCTGGACA A G Q CTATCTGGACAA L S R K TTGGGACAAA C N N GGGAAAATCA E N H AATCTTAAAT	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TATGGATAAA Y D K TTATGGATGAA S S Q TATGCACCTA S S Q TATGCACCTA Y A P K TAACGGCTTG N G L GCTTAGATAA L D K TATGGCAGAG V G R D AAACTTCTTG N F L ACCAAAACGC Q N A TTCAACTGGA F N W I TACAGTTAAAAT AAAGGGCGGA AAAGGGCGGA	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G ANGAACTAAT R T N ATAGTTATCA S Y H GAATATCACA S Y H GAATATCACA S Y H GAATATCACA S Y H CAATATCACA S N N TTAATAGATA L I D T TTACAAACAC TTATTTAAAT	G H N s GATGCAGCTC D A A Q AGATTCCGAT D T D T C D T A Q A AGATTCCGAT D T	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q Q Q Q V Q V Q AAAATCTTTT K S F TTGGTTGGGA G N E M GGTAATCAAA L S R CTGTTATCCAGA L S R CTGTTATCCAGA L S R CTGTTATCTAGA H I R A GGAAAAAAAT E K N TATTAGGGTA	2080 2160 2240 2320 2400 2400 2560 2560 2640 2720 2800 2800 2800 2960 3040 3120
2001 2081 2161 2241 2321 2481 2561 2641 2721 2801 2881 2961 3121 3121	N R N N R N AAAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCA I N S AAACGGTAGG T V G TCAGAGGACAA S E T I GTTGAACGTAGG F L G GGTAACTCA G N F N CCAACTCATCA Q S E CAACTCATCATCA Q S E CAACTCATCATCA C N S S S CAACTCATCATCATCA ATTTAAAAG	AATCTAGAGA Y T V K AATCTAGAGA S. R. E GAGAAAAAG E K K V TITAACTITT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATCTATA Y T K GATTCTATA N Y K CATAAAATA H K I M CTCAAGACAA S R Q ATCCAGAATT P E F ATGGATAGAT M D R Y ATCAACAATAG OP AATT <u>GACGTG</u> GTTAAAAATA	$\begin{array}{c} Y \ \in \ V \\ AGTTACATCA \\ Y \ E \ V \\ AGTTACATCA \\ V \ T \ S \\ TTGATGATAA \\ D \ D \ K \\ AATTTTATTA \\ D \ D \ K \\ AATTTTATTA \\ N \ F \ I \ K \\ AATTTATTA \\ N \ F \ I \ K \\ S \ V \ N \\ GCTGAGTAAATCA \\ S \ V \ N \\ GGTGGAGATA \\ G \ G \ D \ I \ N \\ GATGAGAAGC \\ Q \ E \ S \\ TGAATAATCG \\ N \ G \\ AGCAAACTTA \\ S \ N \ L \ N \\ TATTGGTGTC \\ I \ G \ V \\ ATACAAACTT \\ T \ N \ F \\ GATGTGATTTTT \\ ATAAAGAACT \\ ATGTTATTTTT \\ ATGATAAGAACT \\ ATGTTATTTTT \\ ATGTTATTTT \\ ATGTTATTTT \\ ATGTTATTTT \\ TATTAGGAGATA \\ ATGTTATTTTT \\ ATGTTATTTTTT \\ ATGTTATTTTT \\ ATGTTATTTTTT \\ ATGTTATTTTTT \\ ATGTTATTTTT \\ ATGTTATTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTTT \\ ATGTTATTTTTTT \\ ATGTTATTTTTTTTTT$	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAAG D K S AAAGACACTA K D T I CGTTGTAGAT V V D TTAATATATCC N I S TATAGAACTA Y R T S TATAGAACTA Y R T S TATGGGACAA A G Q CTATCTGGACA L S R K M G P ATGCTGGACAA L S R K M I N GGGAAAATCA GGGAAAAATCA CAGGGCAAAA TGGTATAACA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTATGTTCTCA S S Q TATGCACCTA X A P K TAAGCACCTA C	K V K K V K CTGCAACATC T L CTGCAACATC A T D T L I ATTTATATCG F Y W ATATTATTGG F Y W AAATCAAAA N Q N TCACGTGAGG G G AGAACTAAT R T N AAGAACTAAT S G G G AAGAACTAAT R T N AAGAACTAAT K K TGCAAAAAAA A K K TAGGTAATAA G N N TTAACAAACAC TTATAAAGAAT T D T TTACAAACAC TTATTTAATTGGT T T	$ \begin{array}{cccc} G & H & N & s \\ cbs \\ cbs$	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC Q V Q Q AAAATCTTT K S F TTGGTTGGGA G W D GGTAATGAA G N E M GTTATCAGA L S R CTGTTACTAA L S R CTGTTACTAA V T Y CACATAACAAG H I R A GGAAAAAAA TATTAGGGTA AATGTCAAGA	2080 2160 2240 2320 2400 2480 2560 2640 2720 2800 2880 2960 3040 3120 3220
2001 2081 2161 2241 2321 2401 2481 2561 2721 2801 2881 2961 3041 3121 3201	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TITCTCAGAT S Q I ATTTATTCA I Y S G AATTAATTCA I N S AAACGGTAGG T V G TCACAGGCAACAA S E T I TGTTGAAGCA V E A TGTTTTTAGG F L G GGTAACTTCA C N F N TCAAAGTCAA Q S E CAACTCATAC T H T CCAAGGCAA C N F N CAAAGTCAA C N S S C ACCACGTCTT AATTTAAAAG TACCAACTCTT	AATCCAGATA S. R. E. GAGAAAAAGG E. K. K. V. TITAACTITT L. T. F. GCITATACAAA Y. T. K. GATICTAATG D. S. N. D. TITATTCTATT Y. S. Y. TITAACTATAAAATTA N. Y. K. CATAAAAATTA N. Y. K. CATAAAAATTA H. K. I. M. CICAAGACAA X. TGAATGATAGATA J. ATGGATGATA GITAAAAATA TIAACGATAGA GITAAAAATA TIAATGATGAT TIATGGATGACAA	$\begin{array}{cccc} Y & E & V \\ AGTTACATCA \\ V & T & S \\ TTGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N \\ GGTGGAGATCA \\ P & N \\ GGTGGAGATCA \\ S & V & N \\ GGTGGAGATCA \\ S & V & N \\ GGTGGAGATCA \\ S & V & N \\ GGTGGAGATCA \\ S & N \\ C & S \\ TGAATAATCG \\ N & N \\ S \\ MCAACAACATTAA \\ S & N \\ CTAATCAAACTT \\ T & N \\ F \\ GAAGTTCATTTTT \\ ATAAAGAACT \\ ATTAATCAAAA \\ CTTATTTATTT \\ ATAAAGAACT \\ ATTAATCAAAACT \\ ATTAATCAAACACT \\ ATTAATCAAACT \\ ATTAATCAAACACT \\ ATTAATCAAAACT \\ ATTAATCAAACACT \\ ATTAATCAAACACACACACACACACACACACACACACAC$	N W K ATTGCATTGC I A L L AATTACTTG I T L AAGATAAAAG D K S AAAGACACATA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TATAGAACATA Y R T S TTGGGGACCA A G Q CTATCTGGAAC X G P ATGCTGGACA L S R K TGGATCAAC W I N GGGAAAATCA CAGGGCAAAA TGGTATAACA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T TTATGATAAA Y D K TAGTTCTCA S S Q TATGCACCTA X A P K TAACGGCTG N G L GCTTAGCACTA CGTTAGCAGTA AACAAAACGC Q N A TCAACTGAT F N W I TACAGGTGAAA T V K TGTTAAAAT AAAGAATGA	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G MAGAACTAAT R T N ATAGTTATCACA E Y H K TAGATAATACAAA A K K TAGGTAATAAA G N N TTAATAGAAACAC TTATTTAAAT ACATTTGCGT CGTCTTGATC	G H N S CATGCAGCTC GATGCAGCTC D A A Q AGATTCCGAT D S D TACTCAAGC G S K Y TGAAGGATTT E E F GTAATGGTTC N G S TTCAAAAAAA F K K I TTCAACTTAT M P V TCAAAAATTA S K I TTAAAAAATA S K I TTAAAAAATA C S AACTATCTA AAAGGCAAACA AAAAGGCAAAA AAAAGGCAAATA	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC G V Q Q AAAATCTTTT K S F TTGGTTGGGA G N E M GGTAATGAGA G N E M GGTAATGAGA A CTGTTACCAGA L S R CTGTTACCAGA H I R A GGAAAAAAA N TATTAGGGTA ACTTGTGGAA AAAGTCAATT ACGTTCAAACT	2080 2160 2240 2320 2400 2560 2560 2640 2720 2880 2880 2960 3040 3120 3280 3280
2001 2081 2161 2241 2321 2401 2481 2561 2641 2721 2881 2961 3121 3121 3121 3281 3361	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTCTG ATTTATTCTG I Y S G AATTAATTCA I Y S G AATTAATTCA S AAACGGTAGG T V G TCGGAGCACA S E T I TGTTGTAGCA G N F N TCAAAGTCAAC G N F N TCAAAGTCAAC Q S E CAACTCATAC G N F N TCAAAGTCAAC Q N F N TCAAAGTCAAC Q S E CAACTCATAC T H T CCTATGAGCT P M S St ACCCATCTTC TTGTTTGTAAA	AAACAGATA Y T V K AAACTAGAGA S R E GAGAAAAAGG E K K V TTTAACTATT I T F GCTATACAAA Y T K GATTCTAATG D S N D S N D S N D S N D S N D S N D TAATCTAATA N Y K CATAAAAATA N Y K CATAAAAATA H K I M CTCAAGACAA S R Q ATCCAGAATA S R Q ATCCAGAATA S I Y AAACAGATAGAT GTTAAAAATA TTATGATGAT GGTTAAAAATA	$\begin{array}{cccc} Y & E & V \\ A & T & S \\ T T \\ T \\ T \\ T \\ T \\ S \\ T \\ T \\ T \\$	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAAG D K S AAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTCC N I S TTGGGGACAA W G P ATGCTGGACAA X G Q ATGCTGGACAA L S R K TTGGGACAAA CAGGGAAAATCA E N H AATCTTAAAT CAGGGCAAAA TAAACACTCA GGCAATAAC	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TATGGATGAAA Y D K TAGGTCTCA S S Q TATGCACCTA Y A P K TAACGGCTGA N G L GCTTACGACAA A ACTTCTTG N F L AACACTACTAG N F L AACACAACGG Y A P K AACCTCTTG N F L AACACAACGGA T V K ATTCAACTGAAA T V K ATGTTAAAAT AAAGGCCGGA AAACAATGGA	K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AAGAACTAAT R T N ATAGTTATCA S Y H GAATATCACAA S Y H GAATATCACAA A K K TAGGTAATAA A K K TTAATAGATA L I D T TTACAAACAC TTATTTAAATAGATA	G H N s GATGCAGCTC D A A D A. A Q AGATCCGAT D S D TACTCAAAGC L K A GGTCTAAGC C K A GGTCTAAGC C K Y TGAAGGAATT F K K TTCAAAAAAAA F K K TTCAAAATAAA F X K AATGCCAGT M P V TCCAAAATAA S K I TTAAAAAAAAA S K I TTAAAAAAAAAA S K I TTCAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q Q V Q Q Q V Q V Q Q Q V Q V Q Q GTAATACAAA G N E M GTTATCCAGA L S R CTGTTACCAGA L S R CTGTTGGAA GTTCAGAATTA	2080 2160 2240 2320 2400 2400 2560 2640 2720 2800 2800 2800 3040 3120 3220 3320 3360 3360
2001 2081 2161 2241 2321 2401 2481 2561 2561 2721 2881 2961 3121 3120 3201 3201 3201 33441	N R N N R N AAAATAGTCA K I V K ACCTGTAAGTC V S TTTCTCAGAT S Q I ATTTATTCTG I Y S G AATTAATTCG I Y S G AATTAATTCA I N S T V G TCAGAGGCAGG F L G GGTAACGCTAGG F L G GGTAACTCA C N F N CCAACTCATCA C N F N CCAACTCATCA C N S ST AACCACTCTTT TACATCTGACTA AATTCAACTT	TRACEASE IN AATCTAGAGA S. R. E. GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTACTATTA Y S Y TTAACTATAA N Y K CATAAAAATAA N Y K CATAAAAATAA N Y K CATAAAAATAA N Y K CATAACAGATAGA ATCGACGGTG GTTAAAAATAA AATTGACGGTG GTTAAAAATAA TAGGTTGAACGATAA	Y E V AGTTACATCA V.T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATTTA F N P ACTCAGTAAA S V N GGTGGAGATA G G D I ACAAGAAAACC Q E S TGAATAATCG N N G AGCAACTTAA S N L N TATTGGTGTC I G V ATACAAAACTT T N F GAAGTGAATTTTTT ATAAAGAACT ATACTAAAACTT ATAATAAGAACT ATACTAAAAGACT	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAAG D K S AAAGACACATA K D T I CGTTGTAGAT V V D TTAATATCC N I S TATAGAACATA Y R T S TATAGAACATA Y R T S TATGGGGACAA W G P ATGCTGGACA K S R K I S G CTATCTGGAA L S R K I S G GGAAAATCA E N H AATCTTAAAT CAGGGCAAAA TGATATAACA TAAACACCTA AGCAGTATAC	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TAGTTCTCA S S Q TATGCACTA A P K TAGGCACTA GCTTACGATAA L D K ATGGCAGAG Y G R D AAACTTCTTG N F L AACAAACGC Q N A T TCAACTGGA F N W I TACAGTGAA A AGGGCAGAA AGGGTTTATA AACAATGGA	K V K K V K CTGCAACATC T L CTGCAACATC T K A T L I A T L I ATTTTATTGG F Y W ATATTATTGG F Y W ANAAACAAAA N Q N TCAGGTGAGG G G A AAGAACTAAT R T N AAGAACTAAT K K TGGCAAAAAAA A K K TGGCAAAAAAA A K K TAGGTAATAA G N N TTAACAAACAC TTATTAGAAT A K K ACATTGCATTAAACACA T T T TAAGAAAAAAA K K K CATTTAAAACACA T T T TAAGAAAAAAAAAAAAAA T T T CATTTACAAAAACAAAAAAAAAA<	G H N S cbs GATGCAGGTC D A A Q AGATCCGAT D S D TACTCAAAGC L K A GGTTCTAAG GSTCTAAG GSTCTAAG GS K Y TGAAGAATT E E F GTAATAGGTCC N G S TTCAAAAAAA F K K I TTCAACTAT M P V TCAAAAATA S K I T TATAAAGAT S K I T TATAAAGAT Y K D CTCAAAAATAG S K I T TATAAAGAT X K D CTCAAAAAATA S K I T TATAAAGAT Y K D CTCAAAAAATA S K I T TATAAAGAT X K D CTCAAAAAATA S K I T TATAAAGAT	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC G V Q Q AAAATCTTT K S F TTGGTTGGGA G W D GGTAATGGAA G N E M GTTATCCAGA L S R CTGTTACTTA L S R CTGTTACTTA A ATGTCAACA H I R A GGAAAAAAAT TATTAGGGTA AATGTCAACAT TTGGAAATA	2080 2160 2240 2320 2400 2480 2560 2640 2720 2880 2880 2960 3040 3120 3200 3280 3360 3440
2001 2081 2161 2241 2321 2401 2481 2561 2721 2861 2961 3041 3121 3281 33281 33281 33361 3341	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTTATTGTG I Y S G AATTAATTCA I N S AAACGGTAGG T V G CAAGGGTAGG T V G TCACAGGACAA S E T I TGTTGAAGCA V E A TGTTTTTAGG F L G GGTAACTTCA G N F N TCAAAGGCAT T CTAAAGGCAT T CTAAGGCTTT P M S SI ACCTACGTCTT ACCTACGTCTTAAAAG TACCAACGTTTAAAAG TACCAACGTTTAAAAG TACCAACGTTTAAAAG TACCAACGTTTAAAAG	Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTITT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTATT Y S Y TTAACTATAAAATA N Y K CATAAAAATA N Y K CATAAAAATA H K I M CTCAAGACAA S R Q ATCCAGAATT P E F ATCGATAGATAT TATGGATAGATA GTTAAAAATA TTAACGATAA GTTAAAAATA TTAACGATAA	Y E V AGTTACATCA V.T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA P N P ACTCCAGTAAA S V N GGTGGAGATA G C D I ACAAGAAAACC Q E S TGAATAATGG N N G AGCAACATTA T N F GAAGTGATCAT T N F GAAGTGATCATT T N F GAAGTGATCATTATT ATTAGGAGACT ATTATTAGTACAACAT T ATTAGAAAACT ATTATTAGAAAACT AGTTATTACAAC	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAAGACACATA K D T I CGTTGTAGAT V V D TTAATATCTCC N I S TTAGAGACTA Y R T S TTGGGGACCA W G P ATGCTGGACA A G Q CTATCTGGAA L S R K TTGGATCAAC W I N GGGAAATCA M ATCTTAAAT CAGGGCAAAA TGGTATAACA TAGACATAAA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T TATGTATAAA Y K T TAGTATCAA S S Q TATGCACCTA S S Q TATGCACCTA S S Q TATGCACCTA CACTAGCAGAA L D K TAACGGCAGAA N G L GCTTAGCAGAA CTTCAACAAAACGC Q N A TACACAAAACGC Q N A TCAACTGCA AACAAAACGGCGGA AAGGGTCTTAT AAGGGCGGGA TAATATTTTA	K V K CAATACACT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N O N TCAGGTGGAG S G G G M C G AAGAACTAAT R T N ATAGTTATCA S Y H GAATATACACA A K K TAGCAAAAAA A K K TAGCAAAAAA A K K TAGCAAAAAA C N N TTAATAGATA L I D N TTAACAACAC CGTCTTGATC CGATCTCCTT TTAAAACCAA	$ \begin{array}{cccc} G & H & N & s \\ cbs \\ catcological construction \\ catcological construction$	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTTC N I S CAAGTACAAC Q V Q V Q AAAATCTTTT K S F TTGGTTGGCA G W D GGTAATGAGA G N E M GGTAATGAGA G N E M GTTATCCAGA L S R CTGTTACCAGA L S R CTGTTACCAGA H I R A GGAAAAAT E K N TATTAGGGTA AATGTCAATT ACGTTTGAGAATAA TTAATGTGAATAA	2080 2160 2240 2320 2400 2560 2560 2640 2720 2880 2880 2880 2960 3040 3120 3220 3280 3360 3340 3520
2001 2081 2161 2241 2321 2401 2401 2401 2561 2721 2801 2881 2961 3121 3281 3281 3281 3281 3281 3321 3341 3321	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TTTCTCAGAT S Q I ATTATATTCTG I Y S G AATTAATTCA I N S AAACGTTAGG T V G GTAACTAGGACAA V E A TGTTTTAGG F L G GGTAACTCAA Q S E CAACTCATAC T H T CCTATGAGCT P M S St ACCTATCAAAGTTA TACCATCTTC	AAACAAAAAAA Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTITT L T F GCTATACAAA Y T K GATTCTAATG D S N D S N D S N D TTATCTITAT Y S Y TTAACTATAA H K I M CATAAAAATTA H K I M D R V ATCCAGACAA S R Q ATCCAGAATT P E F ATGGATACAAT S I Y AAACAGATAA S R Q ATCCAGAATTAT S I Y AAACAGATAA TGATTTGACGTG CATAGACAATA CTGCTATCAG	$\begin{array}{cccc} Y & E & V \\ A & GTTACATCA \\ V & T & S \\ TTGATGATGATAA \\ D & D & K \\ AATTTTATTA \\ N & F & I & K \\ GCCAAATCCA \\ P & N & P \\ ACTCAGTAAA \\ F & N & P \\ ACTCAGTAAA \\ G & C & D & I \\ ACAAGAAAGC \\ Q & E & S \\ TGAATAATCA \\ G & G & D & I \\ ACAAGAAAGC \\ Q & E & S \\ TGAATAATCAAACTT \\ T & N & F \\ GAAGCTGATTA \\ S & N & L & N \\ TATTGGGTGC \\ I & G & V \\ ATACAAACTT \\ T & N & F \\ GAAGTTGATT \\ T & N & F \\ GAAGTTGATTTTT \\ T & N & F \\ GAAGTTGATTTTTT \\ TATAATCAAAACTT \\ ATAAATCAAAA \\ ATGTTTTATTT \\ TAGAAAAACT \\ AGTGTTCAAC \\ AGTGTTCAAC \\ GATTTACAA \\ GATGTTCAAC \\ CABAGCTTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCACC \\ CABAGCTTCCAC \\ CABAGCTTCACC \\ CABAGCTTCACC \\ CABAGCTTCACC \\ CABAGCTTCACC \\ CABAGCTTCACC \\ CABAGCTTCCAC \\ CABAGCTTCACC \\ CABAGCTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCCAC \\ CABAGCTTCACC \\ CABAGCTCACC \\ CABACC \\ CABACCACC \\ CABACCACC \\ CAB$	N W K ATTGCATTGC I A L L AATTACTTGC I A L L AATTACTTGC I T L AAGACACTA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TATAGAACTA Y R T S TATAGAACTA Y R T S TATGGGACAA A G Q CTATCTGGAA L S R K TGGGACAAATCA CAGGCAAAATCA CAGGCAAAAAT TGGGACAAAAT TGACACAATAAC	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TTATGATAAA Y D K TTATGATAAA Y D K TAGTCTCTCA S S Q TATGCACCTA X A P K TAACGCACTA C R D CTTAGATAA L D K GCTTAGATAA L D K TATGCACAGA G R D AAACTTCTTG A AACATCGA F N W I TCCAACTGGA F N W I ACGCTATAAAAT AAGGCGTTAA AAGGCGTTAA TAATATTATTA AAGGCAGTGAA AAACAATGGA	K V K K V K CAATACACTT N T L CTGCAACATC A T S GATACATTAA D T L I ATTITATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G AAGAACTAAT R T N ATAGTTAATCACAA C N N TCAGTAATAACAAA A K TGCAATAACAAA C N N TTAATAGATA L I D T TTACAAACACA CATTTCGTCT TTAAAACTAA	G H N S cbs GATGCACGTC DA A Q AGATCCGACTC DS D TACTCAAAGC L K A GGTCTAAGT G S K Y TGAAGAATT G S K Y TGAAGAATT F K K I TTCCAAAAAAA F K K I TTCAACATAT S T Y AAATGCCACT M P V TCAAAAAATA S K I TTATAAAGAT Y K D CTCAATCTAT AAAGGCCAAA ACCACGACA ACAACGACAA AAAGGCCAAAA ACTGTTTCA	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTCC N I S CAAGTACAAC Q V Q Q AAAATCTTT K S F TTGGTTGGGA G W D GGTAATGAAA G N E M GTTATCAGA L S R CTGTTACTAA L S R CTGTTACTAA L S R CTGTTACTAA C W T Y CACATAAGAG H I R A GGAAAAAGAA GGAAAAAAGAA AATGTCAAACT TGAGAATAA	2080 2160 2240 2400 2400 2400 2640 2640 2720 2800 2800 2960 3040 3120 3200 3200 3320 3340 3352 33600
2001 2081 2161 2241 2321 2401 2481 2561 2641 2721 2881 2961 33201 33201 33201 33201 33201 33201 33201 33201 33201 33201	N R N N R N AAAATAGTCA K I V K ACCTGGTAAGTC V S TTTCTCCAGAT S Q I ATTTATTCA I Y S G AATTAATTCA I Y S G AATTAATTCA I Y S G AATGGTGAAGCA S E T I TGTTGAAGCA C S E CTACAGGCAGCA C S E CAACTCATCA C N F N A TGTTTTAGGA C N F N CCAACGCATCATCA C N F N N S ST ACCACCTCTTCAA AATTCCAGTT TACCATCTTTT TACCACCTTT TACCACCTTTTT TACCACCTTCAA	AATCAGAGAA Y T V K AATCTAGAGA S R E GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTTAT Y S Y TTAACTATAA N Y K CATAAAAATTA N Y K CATAAAAATTA N Y K CATAAGACAAT P E F ATGGATAGATTAY AATGACGATA S I Y AATCAGATAGATA TATGACGGTG GTTAAAAATA TATGACGGTG GTTAAAAATA TATGACGGTG GTTAAAAATA TATGACGGTG GTTAAAAATA TATGACGTACAA CGCATCACAA CATACTACACAA	Y E V AGTTACATCA V.T.S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCA P N P ACTCCAGTAAA S V N GGTGGAGATA G C D I ACCAGAAAACT Q E S TGAATAATGG N N G AGCAACTTAA S N L N TATTGGTGTC I G V ATACAAAACTT T N F GAAGTGAATTTTTT ATAAAGAACTT ATTAATCAAAA GTTATTTTTT ATAAAGAACTT ATGATATTACAAG GAAATCGAA	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAAGACACATA K D T I CGTTGTAGAT V V D TTAATATCTC N I S TATAGAGACTA Y R T S TATGGGGACAA Y R T S TATGGGACAA K G Q CTATCTGGAA A G Q CTATCTGGAA CAGGCAAAATCA E N H AATCTTAAAT CAGGGCAAAATCA CAGGGCAAAATCA ACCACTATACA AACCACTACC	T H E I TTTTGCTATC L L S TACAAAACAA Y K T TTATGATAAA Y K T TTATGATAAA Y D K TAGTCTCTCA S S Q TATGCACCTA TAGGCACCTG N G L GCTTAGCACTA CTAGGCAGGA N F L AACAATCTCT Q N A TAGGCAGAA F N W I TACAGGTTAAAA AGGGCTTTATA AGAGCAGGAA AGGGTTATATTA	K V K K X K CTGCAACATC A T L CTGCAACATC A D T L I A T L I A T L I A T L I F Y W AAAAATCAAAA N O N CCAGGTGGAG S G G A ACAACTAAT R T N AAGAACTAAT R T N AAGAACTAAT E Y H K TGCAAAAAAAA A K K T T TAGCTAATAAA K K K T TAGCAAAAAAAA K K T T TAGCAAAAAAAA K K T T TACAAAAAAAAA K K T T TATATTAAAAA A <	G H N S cbs GATGCAGGTC D A A Q AGATTCCAT D S D TACTCAAAGC L K A GGTTCTAAG G S K Y TGAAGAATT E E F GTAATAGGTC N G S TTCAAAAAAA F K K I TTCAACTAT M P V TCAAAAATA S T Y AAATGGCAGT K K I TTATAAAGAT Y K D CTCAAAAATTA S K I T TATAAAGAT X K D CTCAAAAATTA S K I T TATAAAGAT X K D CTCAAAAATTA S K I TTATAAAGAT X K D CTCAAAAATTA S K I TTATAAAGAT X K D CTCAAAAATTA S K I TTATAAAGATA X AAGGCAAAA AAATGGCACA AAATGGCACAA AAATGGCACAA	top M_{-K} AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAAAC A G N ACAACATTC N I S CAAGTACAAC G V Q Q AAAATCTTT K S P TIGGTTGGGA G W D GGTAATGAAA G N E M GGTAATGAAA G N E M GTTATCCAGA L S R CTGTTACTTA V T Y CACATAACAAC H I R A GGAAAAACAT AAGGCAAATTACAGGAA AATGTCAATT ACGTTGTGGAA AATGTCAATT TTGAGAATAA TTAATCTTAA GGAAAAATGGCAATTA	2080 2160 2240 2320 2400 2480 2560 2640 2720 2880 2880 2880 2960 3040 3120 3200 3280 3360 3360 3440
2001 2081 2161 2241 2321 2401 2401 2401 2561 2721 2881 2961 3201 3201 33281 33281 33281 33281 33281 33281 33281 33281 33281 33281 33281 33281 33281	N R N N R N AAAATAGTCA K I V K ACCTGTAAGT P V S TITCTCAGAT S Q I ATTATTCTG ATTATTCTG I Y S G AATTAATTCA I Y S G AATTAATTCA I Y S G AATGGTAGGCA S E T I TGTGGAGCA S E T I TGTGTGAGCA C N F N TCACAGGCACTA S E T I TGTGTGAGCA G N F N TCACAGCACTTCA G N F N TCACAGCTCTTA AATTCAAAGTT GTACCTCTTT AGCTACGTTCTAA	Y T V K AATCTAGAGA S. R. E. GAGAAAAAGG E K K V TTTAACTTTT L T F GCTATACAAA Y T K GATTCTAATG D S N D TTATTCTATT Y S Y TTAACTATAAA N Y K CATAAAAATTA N Y K CATAAAAATA N D R Y AACAAGAATAG OP AATCGGTAGGAT TGGTTTGAACA CAGACAATAAAATA TAAGAACAATAAAATA	Y E V AGTTACATCA V.T S TTGATGATAA D D K AATTTTATTA N F I K GCCAAATCCA P N P ACTCAGTAAA S V N GGTGGAGATA G G D I ACAAGAAAACT C E S TGAATAATGG N N G AGCAACTTAA S N L N FATAGGTGTC I G V ATAACAAACTT T N F GAAGTTGATT ATAAGAAACT ATAATCAAAA GTTATTTATTA ATAATGAAGAC AGTTATTTACAA	N W K ATTGCATTGC I A L L AATTACTTGC I T L AAGATAAAAG D K S AAGGCACATA K D T I CGTTGTAGAT V V D TTAATATCTCC N I S TTGGGGCCA Y R T S TTGGGGCCA Y R T S TTGGGGCCA W I N GGGAAAATCA W G P ATGCTGGACA L S R K TTGGATCGAA L S R K TTGGATCAAA CAACGCAAAA CGCACTATAAAT CAGGGCAAAAAT CAACAACTAA	T H E I TTTTGCTATC L L S TACAAAACAA Y K T T TATGATAAA Y K T T TATGATAAA Y D K TAACGACTAA S Q TATGCACCTA CTTAGGCAGGA N G L GCTTAGGCAGGA N G L GCTTAGGAGAG N F L AACAAAACGC Q N A TTCAACTGGT AACAAAACGG AACATAAAAT AAGGGCGGAA AAGGGTTTTAT AGGACAAGGAA CAATCTAATA	K V K CAATACACT N T L CTGCAACATC A T S GATACATTAA D T L I ATTTTATTGG F Y W AAAATCAAAA N Q N TCAGGTGGAG S G G G MACAACTAAT R T N ATAGTTATCA S Y H GAATATCACA E Y H K TAGCAAAAAA A K K TAGCTAATAA G N N TTAACAACAT TTACAAACAC TTATTTAAAT ACATTTGCTT AAGAACTAAT AAGAACAAA A K K TAGCTAATAACTAA ACATTTGCGT CGTCTTGATC CGTCTTGATC CGATCACCTAA AAGAACAAA	$ \begin{array}{cccc} G & H & N & s \\ cbs \\ GATGCAGCTC \\ D & A & A & Q \\ AGATTCCGAT \\ D & S & D \\ TACTCAAAGC \\ L & K & A \\ GGTTCTAAGT \\ G & S & K & Y \\ TGAAGGAATT \\ G & S & K & Y \\ TGAAGGATT \\ F & K & I \\ TTCAAAAAAA \\ F & K & K & I \\ TTCAAAAAAAA \\ F & K & K & I \\ TTCAAAAAAAA \\ F & K & I \\ TTCAAAAAT \\ S & K & I \\ TTCAAAACAACAAC \\ AAATGCAAAC \\ AAAGGCAAAA \\ ATCTGTAACA \\ ACGAATGTTC \\ ACGAATGTTC \\ \end{array} $	top M_K AACATATCAC H I T AAGTTAAAAA K L K I TGCTGGAACA A G N ACAACATTTC N I S CAAGTACAACQ Q V Q Q AAAATCTTTT K S F TTGGTTGGAA G W D GGTAATGAAA G W D GGTAATGAAA G N E M GGTAATGAAA L S R CTGTTACCAGA L S R CTGTTACCAGA CTGTTACAGA CTGTCAGAAAAAT CGGTTCAAACTT TGGGAAAAAAAGAACAGAAATGG CGATAAAGAACAGAAC	2080 2160 2240 2320 2400 2560 2560 2560 2880 2880 2960 3040 3120 3280 3280 3360 3340 3520 3600 3760

FIG. 3. Nucleotide sequence of *lukS-PV* (positions 1050 to 1993) followed by that of *lukF-PV* (positions 1995 to 2972) and deduced corresponding peptide sequences from *S. aureus* ATCC 49775. The stop codon of *lukS-PV* is located 1 base upstream from the ATG codon of *lukF-PV*. Signal peptides of the two peptide sequences are underlined. Putative consensus sequences (i.e., ribosome binding site [rbs] and promoter sequences) are also underlined. An inverted repeat sequence (positions 3039 to 3064) is indicated by arrows and is probably used as a transcription terminator for *luk-PV*. Other ORFs are underlined, and their orientations are indicated.

(ii) Genes coding for PVL. Although the N-terminal peptide sequences of LukS-PV and LukF-PV were determined, they could not be used to clone the corresponding genes, first, because a reasonably low number of oligonucleotides could not be deduced from the N-terminal extremity of LukS-PV and, second, because high sequence homologies existed between the N-terminal extremities of LukF-PV and HlgB. Therefore, an oligonucleotide probe was deduced from the N-terminal peptide sequence of a trypsin-generated fragment of LukS-PV. Hybridization of total DNA from *S. aureus* ATCC 49775 with this specific oligonucleotide probe yielded DNA fragments (Fig. 2C) different from those obtained by hybridization with *hlg* probes (Fig. 2A and B). The two ORFs constituting the *luk-PV* locus of *S. aureus* ATCC 49775 were cloned into the pUC19 plasmid as an *Eco*RV-*Eco*RV DNA fragment of 6.5 kb. The crude extract of *E. coli* carrying the recombinant plasmid



FIG. 4. Northern blot analysis of expression products from *hlg* and *luk-PV* loci among the total RNAs isolated from *S. aureus* ATCC 49775. Lane 1, photograph of electrophoresed total RNAs from *S. aureus* stained with ethidium bromide. Lanes 2 to 4, Northern blot autoradiograph: 2, total RNAs hybridized with a ³²P-labeled 1,000-bp *Eco*RV-*Hind*III DNA fragment containing the structural gene *hlgA*; 3, total RNAs hybridized with a ³²P-labeled 1,150-bp *Hind*III-*PsI*I DNA fragment containing *hlgC* and part of *hlgB*; 4, total RNAs hybridized with a ³²P-labeled 2,750-bp *Hinc*II-*Hind*III DNA fragment containing *luk-PV*.

was able to induce leukotoxicity (8) at dilutions up to 1/1,000 but did not exhibit any hemolytic activity, whereas that of *E. coli* transformed with pUC19 alone did not show leukotoxicity.

The nucleotide sequence of 3,819 bp from the cloned DNA fragment containing both lukS-PV and lukF-PV was determined (Fig. 3). Like the cotranscribed genes encoding other SHT toxins, *lukS-PV* and *lukF-PV* ORFs were separated by only one thymine. The lukS-PV ORF was 939 nucleotides long (from positions 1055 to 1993), whereas the lukF-PV ORF was 978 nucleotides long (from positions 1995 to 2972). The G+C contents of lukS-PV and lukF-PV were 31.5 and 30.3%, respectively. Two identical putative ribosome binding sites, GAAAG GA from positions 1039 to 1045 and the same from positions 1978 to 1984, were found 10 bases upstream from the ATG codons of lukS-PV and lukF-PV, respectively. The putative ribosome binding site of lukF-PV is thus included within the lukS-PV ORF. Two putative transcription signals are proposed to be located in the 80 bases upstream from *lukS-PV*. A 10-bp inverted repeat was found 67 nucleotides (from positions 3039 to 3064) downstream from lukF-PV and may represent a transcription terminator for the operon.

Whereas all previously sequenced *hlg*-like loci showed a third ORF located approximately 550 bp upstream from *hlgC*-like genes, no similar ORF existed in the 1,000 bp upstream

from *lukS-PV*. In this area, a small ORF without similarity to any previously cloned gene was identified (from positions 231 to 680). Another small ORF without any similarity to any previously known sequence was found 500 bp downstream from *lukF-PV*.

Therefore, it appears that *S. aureus* ATCC 49775 encodes PVL as well as a toxin of the gamma-hemolysin group. This strain, able to produce three class S components and two class F components, could generate six possible combinations of class S and class F components. The precise biological activities of each of these combinations are investigated herein.

Transcription of the loci of SHT toxins from *S. aureus* **ATCC 49775.** Northern blot analysis confirmed that the *hlg*-like locus consists of only two transcriptional units (Fig. 4, lanes 2 and 3): the *hlgA* mRNA is detected as a 1,000-base-long nucleic acid (lane 2), while *hlgC-hlgB* mRNA is 2,100 bases long (lane 3), corresponding to the transcription of *hlgC*-like and *hlgB*-like genes as a single mRNA. Similarly, *lukS-PV* and *lukF-PV* are cotranscribed (Fig. 4, lane 4) as a unique 2,100-base-long mRNA detected by the *Hinc*II-*Hind*III probe overlapping the two genes. For the cotranscribed genes, no additional transcription product which could correspond to the transcription of *hlgB* or *lukF-PV* alone was detected.

Peptide sequence of PVL. As observed for the other bicistronic operons of SHT toxins where the class S encoding genes are transcribed first, *lukS-PV* is the first transcribed gene of the operon. It encodes a 312-amino-acid protein with a calculated molecular mass of 35,297 Da and a calculated pI of 9.1. This protein is probably secreted after cleavage which occurs between A-28 and D-29, as determined by the N-terminal sequencing of the secreted protein. Therefore, the mature polypeptide of 284 amino acids has a calculated molecular mass of 32,374 Da and a calculated pI of 8.95.

The second gene of the operon, *lukF-PV*, encodes a 325amino-acid protein with a calculated molecular mass of 37,012 Da and a calculated pI of 9.17. The N-terminal peptide sequence of the mature protein begins at A-25, giving a 301amino-acid polypeptide having a calculated molecular mass of 34,336 Da and a calculated pI of 9.02.

Therefore, both LukS-PV and LukF-PV are translated as precursors and have signal peptides of 28 and 24 amino acids, respectively, containing sequences typical of procaryote signal peptides (36).

Protein alignments. As shown in Table 1 and 2, the peptide sequences of all the proteins of the gamma-hemolysin group described herein are very similar. Peptide sequences of the HlgA, HlgC, and HlgB proteins had 96 to 99%, 97 to 100%, and 95 to 98% identity, respectively.

TABLE 1. Identity scores of the different class S components of SHT toxins

	% Identity with								
Class S component	LukS-PV	HlgC-ATCC 49775	HlgC-5R	HlgC-P83	HlgC-MRSA	HlgA-ATCC 49775	HlgA-5R	HlgA-P83	HlgA-MRSA
HlgA-MRSA	62.8	65.4	66.0	65.4	65.4	99.0	99.6	99.0	100.0
HlgA-P83	62.8	65.1	65.8	65.8	65.2	100.0	99.0	100.0	
HlgA-5R	62.8	65.4	66.0	65.4	65.4	99.0	100.0		
HlgA-ATCC 49775	62.6	65.2	65.8	65.2	65.2	100.0			
HlgC-MRSA	75.1	97.1	99.4	97.8	100.0				
HlgC-P83	74.8	100.0	97.5	100.0					
HlgC-5R	75.1	96.8	100.0						
HlgC-ATCC 49775	74.8	100.0							
LukS-PV	100.0								

Like HIgA proteins, which had only 66% identity with HIgC proteins, LukS-PV differed markedly from the other class S components, since it had nearly 75% identity with all HIgC-like proteins and 62% identity with all HIgA-like proteins. The difference between LukS-PV and the other S-related proteins was significant since, of the 74 amino acid residues distinguishing LukS-PV and HIgC of strain Smith 5R (HIgC-5R), only 31 (42%) were conservative modifications; similarly, of the 113 amino acid residues distinguishing LukS-PV and HIgA-5R, only 44 (39%) were conservative modifications.

In the case of LukF-PV, this protein is approximately 68 to 71% identical to the other HlgB-like components. In fact, LukF-PV and HlgB from strain 5R sequences differ by 97 residues, of which only 30 (31%) are conservative modifications. These differing residues are distributed throughout the sequences (Fig. 5), and no putative domain could account for the additional hemolytic activity of gamma-hemolysin compared with that of PVL.

Hemolytic activity. As mentioned above, *S. aureus* ATCC 49775 could generate six pairs of SHT toxins when the five protein components were secreted from the bacterium. The biological activity on erythrocytes of each combination was investigated. When assayed alone, none of the five proteins induced detectable hemolysis, indicating that none of the proteins was contaminated by any toxin having a lytic activity on rabbit, human, or sheep erythrocytes.

As shown in Table 3, of the different combinations of class S and class F components secreted by *S. aureus* ATCC 49775, all those including either LukS-PV or LukF-PV had very low hemolytic specific activities on rabbit erythrocytes except the combination HlgA–LukF-PV, which exhibited significant hemolytic activity. In contrast, the SHT toxin HlgA-HlgB had the most potent specific hemolytic activity, being approximately 30 times more effective than the other combination, HlgC-HlgB, constituting gamma-hemolysin.

When tested on sheep erythrocytes, only HlgA-HlgB had a significant lytic activity $[(1.6 \times 10^6) \pm (3 \times 10^5) \text{ U} \text{ of each component per mg}]$. Results were similar with human erythrocytes, but these cells appeared less sensitive to HlgA-HlgB.

Leukotoxic activity. The six possible combinations of class S and class F components (Table 4) were all able to promote lysis of human glass-adherent leukocytes. The morphological changes of PMN cells (8, 9) observed by phase-contrast microscopy were identical for all combinations tested and consisted of a decrease in cell refringency, cell swelling, appearance of rounded nuclei, and occasional bud-like formations.

When assayed alone, none of the five proteins induced detectable leukotoxicity on glass-adherent leukocytes, indicating that none of the proteins used were contaminated by any toxin having lytic activity on human leukocytes.

PVL appeared to be the most leukocytolytic toxin, and all combinations including either LukS-PV or LukF-PV had high leukotoxic specific activities, ranging from 2×10^7 to 7×10^7 U/mg. The two SHT toxins resulting from the gamma-hemolysin locus (HlgC-HlgB and HlgA-HlgB) appeared 30 and 100 times less active, respectively, than PVL.

Activity on rabbit skin of SHT toxins. None of the five components alone had a macroscopic effect when injected in rabbit skin. Only a high amount of protein (30,000 ng) induced transient inflammatory lesions, which disappeared within 72 h. Histological observation provided evidence that this dose (30,000 ng) induced edema, dermal infiltration of PMN cells, and caryorrhexis 4 and 24 h after the injections. These features disappeared within 72 h.

Four combinations of the SHT toxins in equimolar concentrations, HlgC-LukF-PV, LukS-PV-HlgB, HlgC-HlgB, and

INFECT. IMMUN.

TABLE 2. Identity scores of the different class F components of SHT toxins

	% Identity with:						
Class F component	LukF- PV	HlgB-ATCC 49775	HlgB- 5R	HlgB- MRSA	HlgB- P83		
HlgB-P83	71.2	100.0	97.2	94.7	100.0		
HlgB-MRSA	68.4	94.7	97.2	100.0			
HlgB-5R	70.2	97.2	100.0				
HlgB-ATCC 49775	71.2	100.0					
LukF-PV	100.0						

HlgA-HlgB, induced inflammatory lesions alone if 300 ng or more of each component was injected. Intradermal injection of 30,000 ng of these combinations generated papules within 72 h, but no skin necrosis was observed macroscopically or on the histological sections. In contrast, HlgA–LukF-PV and LukS-PV–LukF-PV (i.e., PVL) were the only two SHT toxins able to produce localized acute inflammation associated with skin necrosis in rabbits. However, PVL had the most potent dermonecrotic activity, since the association of as little as 300 ng of each constitutive component was sufficient to induce skin necrosis within 24 h as described previously (5). The HlgA– LukF-PV combination was less potent since similar injuries were obtained only within 72 h and required injection of at least 3,000 ng of each of these components.

Injections in rabbits previously immunized with PVL toxoid (6) showed that protection against skin necrosis was conferred, since up to 30,000 ng of PVL was necessary to induce skin necrosis, which was only observable 72 h after injection. Under the same conditions, the inflammatory lesions induced by HlgA–LukF-PV never produced necrosis, although amounts as high as 30,000 ng of each component were injected. Therefore, immunization of rabbits with PVL prevented necrosis by amounts of toxin 100-fold greater than those necessary to induce equivalent lesions in a nonimmunized animal.

DISCUSSION

The first synergistic bicomponent toxin to be reported in the literature was PVL. Woodin (39) purified this toxin, which was known to be produced by certain S. aureus strains (20). This toxin, having a cytolytic effect on only human and rabbit PMNs (8, 9), has been described recently as a marker of virulence, since PVL-producing strains were strongly associated (5, 6, 22)with necrotizing primary cutaneous infections, especially furuncles, and vice versa. A second toxin, gamma-hemolysin, was later identified as a bicomponent toxin (2, 10, 21, 33) having lytic activity on erythrocytes from a wide range of animal species. Its leukotoxic capacity (16) was not generally considered. A third toxin, bovine leucocidin, which was considered to be active essentially on bovine PMNs, was also characterized (29). A fourth toxin has been studied more extensively by Noda et al. (13, 17–19). This toxin was isolated apparently from the same V8 strain as the toxin studied by Woodin and as such was referred to as PVL. However, we find that the V8 strain (ATCC 27733) from which Noda's toxin was isolated is devoid of any of the genes encoding PVL (Fig. 6, lane 2), whereas another V8 strain (provided by Wellcome Laboratories) produces the toxin and was secondarily deposited as the ATCC 49775 strain. Furthermore, S. aureus ATCC 27733 possesses the genes encoding a member of the gamma-hemolysin group. One could imagine that the V8 strain recorded as ATCC 27733 was contaminated by another S. aureus strain. The sequence of

A		
LukS-PV	(49775)	MVKKRLLAATLSLGIITPIATSFHE-SKADNNIENIGDGAEVVKRTEDTSSDKWGVTQNIQFDFVKDK 67
HlgC	(P83)	*L*NKI**T***VSLLA*L*NPLL*NA**A*DT*D**K*NDV*II*****KT*N***************************
HlgC	(5R)	*L*NKI*ST***VSLLA*L*NPLL*NA**A*DT*D**K*NDV*II*****KT*N***************************
HlgC	(MRSA)	*L*NKI**T***VSLLA*L*NPLL*NA**A*DT*D**K*NDV*II*****KT*N***************************
HlgA	(5R)	*I*NKI*T***AV*L*A*L*NP*I*I***E*K**D**Q*****II***Q*IT*KRLAI**********************70
HlgA	(49775)	*I*NKI*T***AV*L*A*L*NP*I*I***E*K**D**Q*****II***Q*IT*KRLAI**********************
HlgA	(MRSA)	*I*NKI*T***AV*L*A*L*NP*I*I****E*K**D**Q*****II***Q*IT*KRLAI********************** 70
LukS-PV	(49775)	KYNKDALILKMQGFINSKTTYYNYKNTDHIKAMRWPFQYNIGLKTNDPNVDLINYLPKNKIDSVNVSQTL 137
HlgC	(P83)	**************************************
HlgC	(5R)	**************************************
HlgC	(MRSA)	**************************************
HlgA	(5R)	******VV*****S*R***SDL*KYPI**R*I*****S***K*S**********************
HlgA	(49775)	*******VV******S*R***SDL*KYPI**R*I******S***K*S*********************
HIGA	(MRSA)	*******VV******S*R***SDL*KYPI**R*I******S***K*S****K*S**************
LukS-PV	(49775)	GYNIGGNFNSGPSTGGNGSFNYSKTISYNQQNYISEVEHQNSKSVQWGIKANSFITSLGKMSGHDPNLFV 207
HIgC	(P83)	******Q*A**L***************************
HIGC	(5R)	******Q*A**L***************************
HIGC	(MRSA)	********Q*A**L*************************
HIGA	(5R)	*********Q*A**I**S**********************
HIGA	(49/75)	210
HIGA	(MRSA)	**************************************
LukS-PV	(49775)	GYKPYSQNPRDYFVPDNELPPLVHSGFNPSFIASVSHEKGSGDTSEFEITYGRNMDVTHATRRTTHYG 277
HlgC	(P83)	****H*KD*******\$\$*****Q*****Q******T****S**********
HlgC	(5R)	****H*KD*******************************
HlgC	(MRSA)	****H*KD*******\$\$*****Q*****Q******T*****S**********
HlgA	(5R)	Q-D*TGPAA******Q****IQ****IQ****TTL***R*K**K************************
HlgA	(49775)	Q-D*TGPAA******Q****IQ*****IQ*****TTL*****K**K**********************
HlgA	(MRSA)	Q-N*TGPAA******Q****IQ*****IQ*****TTL***R*K**K***********A*Y*YV**HR 279
LukS-PV	(49775)	NSYLEGSRIHKAFVNRNYTVKYEVNWKTHEIKVKGHN 312
HlgC	(P83)	****D*H*V*N**K*******************************
HlgC	(5R)	****D*H*V*N*****************************
HlgC	(MRSA)	****D*H*V*N*****************************
HlgA	(5R)	*AVD*K*D**K***V*************************
HlgA	(49775)	*AVD*K*D**K***V*************************
HlgA	(MRSA)	*AVD*K*D**K***V*************************
_		
в		
LukF-PV	(49775)	MKKIVKSREVTSIALLLLSNTLDAAQHITPVSEKKVDDKITLYKTTATSDSDKLKISQILTFNFIKDK 68
HlgB	(P83)	*NMN*L***SVA**M*******AN*EGK*****V*****V*****A***A***A***F**********
HlgB	(5R)	**MN*L***SVA**M******G*AN*EGK*****V*****V*****A****A***************
HlgB	(MRSA)	**MN"L***SVA**M*****G*AN*EGK****V*****V*****A***A***F****F**********
LukF-PV	(49775)	SYDKDTLILKAAGNIYSGYTKPNPKDTISSQFYWGSKYNISINSDSNDSVNVVDYAPKNQNEEFQVQQTV 138
HlgB	(P83)	******V******N***ER*****YDF*KI*** A ***V**S*Q******************************
HlgB	(5R)	******V***T***N**FV****N*YDF*KI***A***V**S*Q***************************
HlgB	(MRSA)	******V***T***N**FV****N*YDF*KI***A***V**S*Q*****A***A*****************
LukF-PV	(49775)	GYSYGGDINISNGLSGGGNGSKSFSETINYKQESYRTSLDKRTNFKKIGWDVEAHKIMNNGWGPYGRDSY 208
HlgB	(P83)	**TF****S******************************
HlgB	(5R)	**TF****S*******L**NTA***************************
HlgB	(MRSA)	**TF****S******L**NTA****************************
LukF-PV	(49775)	HSTYGNEMFLGSROSNLNAGONFLEYHKMPVLSRGNFNPEFIGVLSRKONAAKKSKITVTYOSEMDRYTN 278
HlqB	(P83)	*P****L**AG***SAY*****IAO*O**L***S*****LS***HR*DG********************************
HlgB	(5R)	*P****L**AG***SAY*****IAO*O**L***S*****LS***HR*DG********************************
HlgB	(MRSA)	*P*****L**AG***SAY*****IAQ*Q**L***S***R*LS***HR*D********************************
LukF-PV	(49775)	FWINFNWIGNNYKDHIRATHTSIYEVDWENHTVKLIDTQSKEKNPMS 325
HlgB	(P83)	R*NG*Y*A*A***NFKTR*FK*T**I****K***L**KET*N*K 325
HlgB	(5R)	R*NG*Y*A*A***NFKTR*FK*T**I****K***L** KET*N*K 325
HlgB	(MRSA)	R*NG*Y*A*A***NFKTR*FK*T**I****K***L**KET*N*K 323

FIG. 5. Alignment of peptide sequences of the known class S (A) and class F (B) components of SHT toxins from S. aureus. Asterisks indicate identical amino acids. Horizontal dashes indicate the absence of amino acids.

the toxin studied by Noda et al. has never been reported, so the data published on the mode of action of this staphylococcal leucocidin cannot be assigned to one of the well-characterized toxins; however, it is likely that this toxin is a member of the gamma-hemolysin group. The work of Kamio et al. (12, 23, 24) reported the cloning and sequencing of the genes encoding yet another leucocidin from an MRSA strain. This new toxin has more than 95% identity with the gamma-hemolysin (4). No comparative study of the biological properties of this leucocidin from the MRSA strain (12) with that from the S. aureus ATCC 27733 strain (17) or with another SHT toxin was ever reported. Here, we report a hemolytic activity on rabbit erythrocytes for the two combinations derived from the gammahemolysin entity from strain ATCC 49775, but it was not the same (12) for the corresponding entity from the MRSA strain. This discrepancy might be the result of divergent evolution of genes or of point mutations. Nevertheless, both toxins generated by gamma-hemolysin are hemolysins and leucocidins.

The data presented in this report allow identification of five different SHT proteins purified from the culture supernatant of the reference strain S. aureus ATCC 49775 producing PVL. Thus, six possible synergistic combinations with different bio-

TABLE 3. Hemolytic specific activities on rabbit erythrocytes of each possible pair of protein components constituting the SHT toxins from *S. aureus* ATCC 49775

Component	Sp act $(U/mg)^b$ when paired with:				
in excess ^a	LukS-PV ^c	HlgC ^c	$HlgA^{c}$		
LukF-PV HlgB	$<\!$	$ \substack{ < 2.10^2 \\ 1.6 \times 10^6 (\pm 3 \times 10^5) } $	$\begin{array}{c} 2.2 \times 10^4 (\pm 8 \times 10^2) \\ 5.2 \times 10^7 (\pm 2 \times 10^6) \end{array}$		

^a Amount of component in excess in the assay, 125 ng.

^b Values indicate the means of quadruplicate assays (\pm standard deviations). One unit was defined as the limiting amount of class S component in the presence of an excess of a class F component contained in dilutions producing 50% hemolysis.

^c Limiting component.

logical activities are thought to occur naturally. The genes encoding each purified protein were cloned, sequenced, and found to be organized as two distinct loci. The first locus, containing two transcription units and three ORFs, encodes a gamma-hemolysin-like toxin. The second locus encodes the PVL and is organized as two tandemly cotranscribed ORFs. On these loci, the genes encoding a class F component are always located 1 base downstream from a gene encoding a class S component, and these genes have no evident promoter, as shown by the transcript analysis. As described in previous incomplete data (31), the sequences of LukS-PV and LukF-PV indicated that PVL and gamma-hemolysin are encoded by different loci but belong to a single family of toxins. The sequencing data obtained in this work confirm that this toxin belongs to the SHT family. Our results suggest a partitioning of this family of toxins, since PVL was divergent from the gammahemolysin group. Partitioning was also indicated by the distribution of the corresponding genes among S. aureus strains (22) and by the specific clinical association of PVL-producing strains (5, 6, 8) with primary cutaneous lesions, whereas the role of gamma-hemolysin in pathogenicity has never been studied.

Study of the biological activities of the six combinations of the SHT proteins also showed that each of the two combinations from the gamma-hemolysin genes induced significant hemolysis of rabbit erythrocytes. In contrast, none of the combinations which included LukS-PV was hemolytic. Study of the leukotoxic activities of the toxins indicated that combinations including HIgA had the lowest activity on human glass-adherent leukocytes. To further understand the role of the different toxins, the binding of each component to target membranes and kinetic studies of their biological activities are currently being investigated. To determine whether the combination of the five components produced by *S. aureus* ATCC 49775 led to a further synergy, we tested amounts of proteins equal to the

TABLE 4. Leukotoxic specific activities on human PMN cells of each combination of protein components of the SHT toxins

Component	Sp act $(U/mg)^b$ when paired with:					
in excess ^a	LukS-PV ^c	$HlgC^{c}$	$HlgA^{c}$			
LukF-PV HlgB	$\begin{array}{c} 7.10^7 (\pm 10^6) \\ 2.10^7 (\pm 10^6) \end{array}$	$\begin{array}{c} 2 \times 10^7 (\pm 10^6) \\ 2 \times 10^6 (\pm 10^5) \end{array}$	$\begin{array}{c} 3 \times 10^7 (\pm 10^6) \\ 8 \times 10^5 (\pm 5.10^4) \end{array}$			

^a Amount of component in excess in the assay, 50 ng.

^b Values indicate the means of quadruplicate assays (\pm standard deviations). One unit is defined as the limiting amount of class S component in the presence of an excess of class F component contained in dilutions producing 100% of the morphological changes in 10⁵ glass-adsorbed PMN cells.

^c Limiting component.



FIG. 6. Southern blot analysis showing that total DNA from *S. aureus* ATCC 27733 does not hybridize with the *luk-PV*-specific probes but does hybridize with the *hlg*-specific probe, whereas total DNA from *S. aureus* ATCC 49775 hybridizes with both probes. Lanes: 1 and 4, 5'-labeled *Bs*tell-restricted DNA from bacteriophage lambda; 2, hybridization with the *luk-PV*-specific probe of the *Acc*1-restricted DNA of ATCC 27733; 3, hybridization with the *luk-PV*-specific probe of the *Acc*1-restricted DNA of ATCC 49775; 5, hybridization with the *luk*C-specific probe of the *Acc*1-restricted DNA of ATCC 27733; 6, hybridization with the *hlg*C-specific probe of the *Acc*1-restricted DNA of ATCC 2775.

various minimal doses necessary to obtain biological activity in both in vitro and in vivo assays. The combination of the five components led to a leukotoxic activity in the range of that obtained with a single two-component combination. In hemolytic and dermonecrotic assays, the same combination had biological activity identical to that obtained with the most potent pair, i.e., HlgA-HlgB and LukS-PV–LukF-PV, respectively. These observations reflect the specificity of SHT and may indicate the absence of mobility of components after interacting with membranes.

PVL (LukS-PV–LukF-PV) was found to be highly dermonecrotic for rabbit skin (5). However, of the heterologous pairs naturally obtained from strains producing PVL and gammahemolysin, HlgA–LukF-PV seems to induce lesions similar to those induced by PVL but only with injection of higher amounts.

Therefore, it appears that all of the SHT toxins have distinct biological effects. Furthermore, all of the SHT toxins seem to have a potent inflammatory role since they induce leukocyte infiltration into tissues as observed on histological sections. At the molecular level, sublytic amounts of PVL (LukS-PV–LukF-PV) have been demonstrated to induce granule secretion (3) and to lead to the release of leukotriene B_4 (11) and interleukin 8 (27) from human PMNs. This may explain the chemotaxism of PMNs and the inflammatory lesions observed in the animal model when these toxins were tested in rabbit skin. The activity of other SHT combinations at the molecular level is currently being investigated.

As observed recently (22), all PVL-producing *S. aureus* strains tested to date produce, in addition, the three components of the gamma-hemolysin group, indicating that the generation of the six SHT pairs is not restricted to strain ATCC 49775 but could involve all of the PVL-producing strains (or most of them). The ancient terms gamma-hemolysin and leucocidin reflect the action of these toxins on selected models in

vitro but not their role in the virulence of *S. aureus*. Moreover, different names were given to almost identical toxins (gammahemolysin, leucocidin, and leucocidin R) and were sources of confusion. Therefore, we proposed (31) to call these related toxins by their common properties, i.e., SHT toxins, since they act by two nonassociated protein components upon target cell membranes. In addition, all 30 *Staphylococcus intermedius* strains tested to date in our laboratory produce a new SHT toxin (21a). Recently, a new protein that had a primary structure like a class S component (1) but was able to induce biological activity when combined with another class S component and not a class F component was described. The latter data may provoke questions regarding the structural role in pore formation of such a protein that remain to be studied.

ACKNOWLEDGMENTS

The skillful technical assistance of C. Fritsch and D. Keller was greatly appreciated. We thank G. Green for editing the English.

This work was supported in part by grants from the Institut National pour la Santé et la Recherche Médicale (INSERM 900316) and from the Fondation pour la Recherche Médicale.

REFERENCES

- Choorit, W., J. Kaneko, K. Muramoto, and Y. Kamio. 1995. Existence of a new protein component with the same function as the LukF component of leukocidin or g-hemolysin and its gene in *Staphylococcus aureus* P83. FEBS Lett. 357:260–264.
- Cline, M., T. H. Birbeck, and J. P. Arbuthnott. 1992. Characterization of staphylococcal gamma-lysin. J. Gen. Microbiol. 138:923–930.
- Colin, D. A., I. Mazurier, S. Sire, and V. Finck-Barbançon. 1994. Interaction
 of the two components of leukocidin from *Staphylococcus aureus* with human
 polymorphonuclear leukocyte membranes: sequential binding and subsequent activation. Infect. Immun. 62:3184–3188.
- Cooney, J., Z. Kienle, T. J. Foster, and P. W. O'Toole. 1993. The gammahemolysin locus of *Staphylococcus aureus* comprises three linked genes, two of which are identical to the genes for F and S components of leukocidin. Infect. Immun. 61:768–771.
- Couppié, P., B. Cribier, G. Prévost, E. Grosshans, and Y. Piémont. 1994. Leucocidin from *Staphylococcus aureus* and cutaneous infections: an epidemiological study. Arch. Dermatol. 130:1208–1209.
- Cribier, B., G. Prévost, P. Couppié, V. Finck-Barbançon, E. Grosshans, and Y. Piémont. 1992. *Staphylococcus aureus* leukocidin: a new virulence factor in cutaneous infections. Dermatology 185:175–180.
- Finck-Barbançon, V., G. Duportail, O. Meunier, and D. A. Colin. 1993. Pore formation by a two-component leukocidin from *Staphylococcus aureus* within the membrane of human polymorphonuclear leukocytes. Biochim. Biophys. Acta 1182:275–282.
- Finck-Barbançon, V., G. Prévost, and Y. Piémont. 1991. Improved purification of leukocidin from *Staphylococcus aureus* and toxin distribution among hospital strains. Res. Microbiol. 142:75–85.
- Gladstone, G. P., and W. F. Van Heyningen. 1957. Staphylococcal leucocidins. Br. J. Exp. Pathol. 38:123–137.
- Guyonnet, F., and M. Plommet. 1970. Hémolysine gamma de Staphylococcus aureus: purification et propriétés. Ann. Inst. Pasteur 118:19–33.
- Hensler, T., B. König, G. Prévost, Y. Piémont, and W. König. 1994. Leukotriene B₄ generation and DNA fragmentation induced by leukocidin from *Staphylococcus aureus*: protective role of granulocyte-macrophage colonystimulating factor (GM-CSF) and G-CSF for human neutrophils. Infect. Immun. 62:2529–2535.
- Kamio, Y., A. Rahman, H. Nariya, T. Ozawa, and K. Izaki. 1993. The two staphylococcal bi-component toxins, leukocidin and gamma-hemolysin, share one component in common. FEBS Lett. 321:15–18.
- Kato, I., and M. Noda. 1989. ADP-ribosylation of cell membrane proteins by staphylococcal α-toxin and leucocidin in rabbit erythrocytes and polymorphonuclear leucocytes. FEBS Lett. 255:59–62.
- Kornblum, J., S. J. Projan, S. L. Moghazeh, and R. P. Nowick. 1988. A rapid method to quantitate non-labeled RNA species in bacterial cells. Gene 63: 75–85.
- 15. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of

the head of bacteriophage T4. Nature (London) 227:680-685.

- Möllby, R. 1983. Isolation and properties of membrane damaging toxins, p. 619–669. *In* C. S. F. Easmon and C. Adlam (ed.), Staphylococci and staphylococcal infections. Academic Press Ltd., London.
- Noda, M., T. Hirayama, I. Kato, and F. Matsuda. 1980. Crystallization and properties of staphylococcal leukocidin. Biochim. Biophys. Acta 633:33–44.
- Noda, M., I. Kato, T. Hirayama, and F. Matsuda. 1980. Fixation and inactivation of staphylococcal leukocidin by phosphatidylcholine and ganglioside GM1 in rabbit polymorphonuclear leukocytes. Infect. Immun. 29:678–684.
- Noda, M., I. Kato, F. Matsuda, and T. Hirayama. 1981. Mode of action of staphylococcal leukocidin: relationship between binding of ¹²⁵I-labeled S and F components of leukocidin to rabbit polymorphonuclear leukocytes and leukocidin activity. Infect. Immun. 34:362–367.
- Panton, P. N., and F. C. O. Valentine. 1932. Staphylococcal toxin. Lancet 222:506–508.
- Plommet, M. 1988. Purification of staphylococcal gamma-hemolysin. Methods Enzymol. 165:8–16.
- 21a. Prévost, G. Unpublished data.
- Prévost, G., P. Couppié, P. Prévost, S. Gayet, P. Petiau, B. Cribier, H. Monteil, and Y. Piémont. 1995. Epidemiological data on *Staphylococcus aureus* strains producing synergohymenotropic toxins. J. Med. Microbiol. 42:237-245.
- Rahman, A., K. Izaki, and Y. Kamio. 1993. Gamma-hemolysin genes in the same family with *lukF* and *lukS* genes in methicillin resistant *Staphylococcus aureus*. Biosci. Biochem. 57:1234–1236.
- Rahman, A., H. Nariya, K. Isaki, I. Kato, and Y. Kamio. 1992. Molecular cloning and nucleotide sequence of leucocidin F-component gene (*lukF*) from methicillin-resistant *Staphylococcus aureus*. Biochem. Biophys. Res. Commun. 184:640–646.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
- Schreiner, A., B. König, M. Köller, G. Prévost, Y. Piémont, J. E. Alouf, and W. König. 1994. Activation of human effector cells by different bacterial toxins (leukocidin, alveolysin, and erythrogenic toxin A): generation of interleukin-8. Infect. Immun. 62:4831–4837.
- Smith, M. L., and S. A. Price. 1938. Staphylococcus γ-haemolysin. J. Pathol. Bacteriol. 47:379–393.
- Soboll, H., A. Ito, W. Schaeg, and H. Blobel. 1973. Leukocidin from staphylococci of different origin. Zentralbl. Bakteriol. Hyg. I Abt. Orig. A 224: 184–193.
- Southern, E. 1975. Detection of specific sequence among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
- Supersac, G., G. Prévost, and Y. Piémont. 1993. Sequencing of leucocidin R from *Staphylococcus aureus* P83 suggests that staphylococcal leucocidins and gamma-hemolysin are members of a single, two-component family of toxins. Infect. Immun. 61:580–587.
- Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 polymerase. Proc. Natl. Acad. Sci. USA 84:4761– 4771.
- Taylor, A. G., and A. W. Bernheimer. 1974. Further characterization of staphylococcal gamma-hemolysin. Infect. Immun. 10:54–59.
- Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77:5201– 5205.
- Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA 76:4350–4354.
- Von Heijne, G. 1985. Signal sequences, the limits of variations. J. Mol. Biol. 184:99–105.
- Ward, P. D., and W. H. Turner. 1980. Identification of Panton-Valentine leukocidin as a potent dermonecrotic toxin. Infect. Immun. 28:393–397.
- Woodin, A. M. 1972. The staphylococcal leukocidin, p. 281–289. In J. O. Cohen (ed.), The staphylococci. Wiley Interscience, New York.
- Woodin, A. M. 1960. Purification of the two components of leucocidin from Staphylococcus aureus. Biochem. J. 75:158–165.
- Woodin, A. M., and A. A. Wienecke. 1968. Role of leucocidin and triphosphoinositide in the control of potassium permeability. Nature (London) 220:283–286.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strain: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.