Cloning and Nucleotide Sequence of Type 3 M Protein Gene (emm3) Consisting of an N-Terminal Variable Portion and C-Terminal Conserved C Repeat Regions: Relation to Other Genes of Streptococcus pyogenes

Streptococcal Diseases Study Group (Chief: Kiyoshi HOSHINA) Chihiro KATSUKAWA

Laboratory of Microbiology, Osaka Prefectural Institute of Public Health (Received: November 12, 1993) (Accepted: February 7, 1994)

Key words: Streptococcus pyogenes, M proteingenes, PCR, emm3, emm12

Abstract

The structural gene for type 3 M protein of *Streptococcus pyogenes*, which consists of an N-terminal variable portion and C-terminal conserved repeat regions, has been cloned by the polymerase chain reaction (PCR) with two primers (K-1 and K-2). They were selected from the best conserved region of the leader sequences and of the C-terminal portion near the Hexapeptide (LPSTGE) sequence found in different M proteins. From the nucleotide sequences of the product, 1645 base pairs were determined, including 32 amino acids of the leader sequences, the complete N-terminal variable region and the conserved C repeat regions. Analysis of the deduced amino acids of the sequence revealed the existence of two major repeat regions, the B and C repeat regions. Comparison of the C-repeat regions among M3 and other M proteins showed them to be more than 90% identical. The two B repeat blocks in M3 protein are also similar to those in M12 protein. Predictive secondary structure analysis of M3 protein reveals a strong alpha-helical potential. The algorithm also shows that the beta-sheet and turn potential for region 23–42 in M3 protein are similar to those for region 28–50 in M12 protein. The results indicate that M3 protein is closely related to M12 protein.

Introduction

Streptococcus pyogenes is responsible for a wide variety of human diseases, the most common of which are nasopharyngitis and impetigo¹⁾. Moreover, streptococcal pharyngeal infection in humans may develop into rheumatic fever or glomerulonephritis¹⁾. The principal virulence factor of *S. pyogenes* is a cell wall constituent known as M protein that gives the organism the ability to resist phagocytosis²⁾. This virulence factor displays antigenic diversity within its amino terminal region³⁾. The highly variable portions of M proteins form the basis of a serological typing scheme, and only antibodies directed to type-specific epitopes are capable of circumventing the antiphagocytic effect²⁾. M protein is thought to inhibit alternative C3 convertase formation to restrict deposition of C3 on the streptococci and also to inhibit the classical C5 convertase formation in order to interfere with efficient complement receptor-mediated phagocytosis⁴⁾. However, the relationship between M protein antigenic diversity and its antiphagocytic activity is not understood, and neither is the genetic basis for M protein antigenic diversity or the structural basis for

別刷請求先: (〒537)大阪市東成区中道1-3-69 大阪府立公衆衛生研究所 勝川 千尋

functions common to all M protein serotypes. In order to characterize them in more detail, the nucleotide sequences of genes encoding a number of different M protein serotypes need to be cloned, sequenced and compared. In this report, we describe the cloning and sequencing of emm3 from the *S. pyogenes* type 3 M protein gene which has not been known to date and compare this gene with others reported previously.

Materials and Methods

Bacterial strains, plasmid and media: *S. pyogenes*, the M⁺ (type 3) strain C203 (ATCC 12384) was used in this study. *Escherichia coli* SJ2, harboring plasmid pJRS42.50, consists of an Xba I-Pvu II fragment including the emm6 gene from *S. pyogenes* D471 cloned into pUC19; it was a gift from Dr. June R. Scott⁵⁾. *E. coli* JM109 was used as the recipient for plasmid transformation and for phage M13 propagation. The plasmid vector was pUC118 obtained from Takara Shuzo Co., Ltd, Kyoto, Japan. *S. pyogenes* was grown in Todd-Hewitt broth (Difco Laboratories, Detroit, MI, U.S.A.); *E. coli* strains were grown in LB broth.

Isolation of DNA: Chromosomal DNA from *S. pyogenes* strain C203 was prepared according to a procedure reported previously⁶⁾. Briefly, the cultured strain C203 (500 ml) was centrifuged at 8,000 rpm for 30 min and the resulting pellet was lysed in 25 ml of 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 (TE), 1.25 ml of 10% SDS and 0.125 ml of a 20 mg/ml solution of proteinase K. The resulting mixture was incubated at 37°C for 45 min. To the lysate, 4.74 ml of 5 M NaCl was added with thorough mixing, and then 4 ml of CTAB/NaCl solution (10% hexadecyl trimethyl ammonium bromide in 0.7 M NaCl) was added to the mixed lysate and this mixture was incubated at 65°C for 20 min. The CTAB-treated lysate was extracted with an equal volume of phenol/chloroform/isoamyl alcohol to remove CTAB-protein/polysaccharide complexes. The aqueous phase was transferred to a fresh tube and then 0.6 volume of isopropanol was added. The resulting precipitate was washed once with 70% ethanol. The washed DNA was suspended at 100 ng of genomic DNA per ml in TE.

E. coli SJ2 cultured in LB broth containing $50 \mu g$ of ampicillin per ml was collected by centrifugation at 8,000 rpm for 30 min, and the plasmid in the bacteria was isolated with a Qiagen column (QIAGEN-tip 100) (Diagen Inc. Chatsworth, CA, U.S.A.), according to the protocol of the manufacturer. The isolated plasmid was digested with Msp I and Pvu II restriction enzymes. After electrophoresis of the restricted plasmid in a 0.8% gel, the fragment, which contained only the region encoding the M6 protein and lacked 32 bases at the 5' end and 38 bases at the 3' end of the gene, was purfied with a Gene Clean Kit (Bio 101, Inc., La Jolla, CA, U.S.A.).

Oligonucleotides: Two oligonucleotides were synthesized with a 380 B automatic DNA synthesizer (Applied Biosystems Inc., Foster City, CA, U.S.A.) and used as specific primers for the required extension and amplification reactions. They were 31-mer (specific primer K-1; 5'-CCG<u>GGATCCTATTCGCTTA-GAAAATTAAAAA-3'</u>) and 30-mer (specific primer K-2; 5'-CCG<u>GTCGACAAGTTCTTCAGCTTGTT-TCGC-3'</u>). The K-1 and K-2 primers contained Bam HI and Sal I recognition sequences at the 5' end respectively. This made it easy to insert the amplified fragment into the vector.

Cloning of emm3 gene by the polymerase chain reaction: PCR was performed in a Hybaid Thermal Reactor (Hybaid Ltd., U.K.). The reaction mixture contained 10 μ l of 100 ng/ μ l genomic DNA, 25 pmol of each phosphorylated primer, 0.2 mM deoxynucleotide triphosphate mix, 0.01% (w/v) gelatin, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂ and 2 U of Taq polymerase (Takara) in a total volume of 100 μ l, and the reaction mixture was overlaid with 2 drops of mineral oil (Sigma Chemical Co., St. Louis, MO, U.S.A.). PCRs (30 cycles) were performed with each cycle consisting of denaturation (94°C, 1 min), annealing (55°C, 2 min) and extension (72°C, 2 min). The PCR product was precipitated with ethanol and digested with Bam HI and Sal I restriction enzymes. After electrophoresis, it was extracted from the preparative agarose gel and cloned into the cloning vector pUC118 which had been digested with Bam HI and Sal I.

Chihiro KATSUKAWA

DNA sequence: Plasmid pUC118 harboring the emm3 gene transformed in *E. coli* JM109 was grown in L-broth containing 50 μ g of ampicillin per ml and purified with Econopack Q (Pharmacia LKB Biotechnology Inc., Piscataway, N.J., U.S.A.). The purified emm3 gene inserted into the vector was digested with Bam HI, Sal I and Bgl II (fragments of about 440 and 1030 base pairs), with Bam HI, Sal I and Sca I (fragments of about 880 and 590 base pairs) and with Bam HI, Sal I and Stu I (fragments of about 1200 and 270 base pairs). Each fragment was purified with a Sephaglas® Band Prep Kit (Pharmacia) and inserted into M13 phage. Single-strand DNA was purified from the supernatant of the cultured phage with a Sephaglas® Phage Prep Kit (Pharmacia). The purified single-strand DNA was sequenced with an Auto Read® Sequencing Kit (Pharmacia) by A.L.F. DNA Sequencer (Pharmacia).

Hybridization: DNA samples were denatured by heating for 10 min at 95°C and then chilled rapidly on ice for dot blot hybridization. DNA samples also were denatured in NaOH for Southern hybridization before application to membrane filters (Schleicher and Schuell, Dassel, Germany) which had been soaked in distilled water and then in $20 \times SSC$ buffer ($1 \times SSC$ is 0.15 M sodium chloride, 0.15 M sodium citrate, pH 7.0). Hybridization of the restriction fragments was conducted by the method of Southern. The DNA samples were bound to a nitrocellulose membrane by baking for 2 hr in a vacuum at 80°C. The DNA bound to the filter was incubated for 2 hr at 68°C in prehybridization solution (0.25% powdered skim milk in $2 \times SSC$). Hybridization was carried out by overnight incubation at 65°C with labeled probe DNA diluted to a 5 ml final volume of prehybridization solution. The hybridized filters were washed twice in $2 \times SSC$, 0.1% SDS for 15 min at room temperature and twice at 68°C in 0.1 × SSC, 0.1% SDS for 15 min 68°C. Probes were labeled with digoxigenin-11-UTP (Boehringer Corp. Ltd., Sussex, U.K.) by random hexanucleotide primers according to the protocol supplied by the manufacturer. Digoxigenin-labeled probes were detected by using an anti-digoxigenin antibody alkaline phosphatase conjugate (Boehringer) and the substrates of BCIP (5-bromo-4-chloro-3-indolyl phosphate) and NBT (nitroblue tetrazolium).

Nucleotide sequence accession number. The 1465 base pair nucleotide sequence of emm3 gene is available from DDBJ, EMBL and GenBank Nucleotide Sequence Databases under accession number

Fig. 1 Analysis of the amplification product. (A) Ethidium bromide picture of the agarose gel (0.8%) of the amplified product. Lane 1, PCR product with K-1 and K-2 primers; lane 2, marker. (B) Dot blot hybridization analysis with the Msp I-Pvu II fragment of the emm6 gene probe. Serial five-fold dilutions of 6 ng of the Msp I-Pvu II fragment of the emm6 gene (lane a) and 5 ng of DNA amplified with K-1 and K-2 primers (lane b) were spotted onto nitrocellulose and probed with the digoxigenin-labeled Msp I-Pvu II fragment. (C) Southern blot analysis with the Msp I-Pvu II fragment of the emm6 probe. Lane 1, DNA amplified with K-1 and K-2 primers; lane 2, Msp I-Pvu II fragment of emm 6 gene. DNA hybridized with the probe is shown by the arrows.



700

D14415.

Results

Cloning and nucleotide sequence of the emm3 gene from S. *pyogenes* type 3 strain C203 with PCR: Much information is available on nucleotide sequences from various strains⁷)-12). We tried to obtain the

Fig. 2 Nucleotide and deduced amino acid sequences of the emm3 gene. The DNA strand is located at 5' to 3' and its nucleotides are numbered above each line. Amino acid residues are presented as single letters below each line. B repeat blocks and C repeat blocks are indicated by underlining.

Signal peptide →																										
m> 77		10			20			30			10	-	50			60	0			70			80			90
Y S	L	R	K	LI	члл/ (1	CAG	GAAC	.GGC A	STIC	V V	AGCGC	TTG A	L CTT	GACI T	V V	TTT/ T.	GGGG	ACA T	GGI	ACT L	GG1 V	AGC	G	GCA	GAC T	V
	M	atu	re p	rote	n -	÷ -										-		-		-				•		
	1	00		11	0		12	20		13	30		140			150	5		16	50			170			180
AAGG	CAGA	TGC	TAGG	AGTO	TTA	ATG	GAGA	\GTI	TCC	TAG	CATO	TTA	AATT	AAA/	AAT	IGA	AAT.	GAG	AAG	CTT	GTT	AGA	TCA	GGI	TAC	ACAA
K A	D	А	R	S \	/ N	G	Б	F	Р	R	н (ĸ	L	ĸ	N	Е	I	E	N	L	r	D	Q	v	т	Q
	1	90		20	0		21	10		22	20		230			240)		25	50			260			270
TTAT	атаа	TAA	асат	AAT	GTA	ATT	ACCA	ACA	ATA	TAG	GCAC	AAG	CTGG	CAGA	CT	IGAC	CT	AGA	CA	AAA	GGC	TGA	ATA	ICT	АЛА	AGGC
ΓY	N	ĸ	н	N S	: N	Y	Q	Q	Y	s	A C	A	G	R	r	D	L	R	Q	ĸ	A	Е	Y	г	к	G
	,			20			20						220			2.20							250			260
CTTA	ATGA	TTG	GGCI	GAGA	GGC	TGT	TACA	AGA	GTT	AAAT	GGAG	AAG	320 ATGT	****		AGTT	, rrr,	GGT	34 (AA	i U AGTA	GGC	TTT	TGA.		AGA	TGAT
LN	D	W	A	EF	L	L	Q	Е	L	N	GE	D	v	ĸ	ĸ	v	L	G	ĸ	v	A	F	Е	к	D	D
	3	70	~~~~	38	0		39	0		40	0		410			420)		43	30		_	440			450
LE	K AAAA	E	v	AAAG K F	AAC T	TTA	AGAA	IAAA K	TAAI	AGAC	X X X	AGG/ F	AAA W	GGAA	V	CAG	GAC	TTA	GAI	ĽAA.	AGA	TTT F	TGA	CTT L	GGC	CAAA
			•				~	ĸ	-	U	K F	5			1	¥	5	5	D	r	5		5	2	A	ĸ
	4	60		47	0		48	0		49	0		500			510)		52	20			530			540
CAGG	GTA	TGT	TTTA	TCAG	ATA	AAA	GACA	TCA	ACA	AGAA	CTAG	AAG/	AAA	AGAA	AAG	SAAA	GTI	ACA	GA/	AGC /	AAC	TGC	TAA	AGT	TGG	CCAG
Q G	Y	v	L	S C	ĸ	R	н	Q	Q	E	LE	E	ĸ	Е	ĸ	ĸ	v	т	Е	Α	т	А	к	v	G	Q
	5	50		56	0		57	0		58	0		590			600			61	0			620			630
ATTA	GCGA	AGA	GCTA	GAGA	CAG	TTA	AACA	- 	AGT	TGAA	AGTA	CTAT	GCA	AGAT	тта	ACT	GAA	АЛА	CAA	AA	rcg	IGT	TTC:	ICA(GTT	AGAG
I S	E	Е	L	е т	v	ĸ	Q	к	v	Е	s T	м	Q	D	L	т	E	ĸ	Q	N	R	v	s	Q	L	Е
		••			•						-															
CARC	0 10 mm	40	TA CT	65 80778	0		66	0	NCN:	67 2017	0		680			690			70	0			710	-		720
QE	L	A	T	T K	0	N	A	K	E	D	FE	T.	AGC	A	ття Т.	A	N	A	A SC-1	D	K	O O	K	этт. т.	E	A
					~										-						••	*		~	~	
		30		74	0		75	0		76	0		770			780			79	0		4	800			810
AAGA:	rtgC	30 CGA	TTTA	74 Зала 5 т	0 CAA	AAC	75 ГАЛА У	0 AGA	GGCI	76 AAAG	0 GAAG	ATTI	770 TGA	ACTA	GCA	780 GCA	TTA	GGT	79 CAC	0 CAP	ACA:	rger	BOO ICAT	'AA'	IGAC	810 Stat
AAGA: K I	rtgc A	30 CGA D	L L	74 Зааа Е Т	0 САА. К	AAC: L	75 ГААА К	0 AGA E	GGCI A	76 AAAG K	0 GAAG E D	ATTI F B	770 TGAI E 1-re	ACTA L Dae	GCA A t	780 GCA A	TTA L	GGT G	79 CAC H	0 CAP Q	ACA: H	IGC A	BOO ICAI H	n N	IGA(E	810 Stat Y
AAGA: K I	rtgc A 8	30 CGA D 20	L L	74 SAAA E T 83	0 САА. К	AAC: L	75 ГААА К 84	0 AGA E 0	GGCI A	76 AAAG K 85	0 GAAG E D 0	ATTI F B	770 TGA E 1-re 860	аста L рае	GCA A t	780 GCA A 870	TTA L	GGT G	79 CAC H 88	0 CA2 Q 0	ACA: H	IGC A	B00 ICA1 H 890	n N	IGA(E	810 STAT Y 900
AAGA K I CAAGO	RTGC A 8 CAAA	30 CGA D 20 ACT	ITTA L AGCA	74 3AAA E T 83 3AAA	0 CAA K 0 AAG	AAC: L ATGJ	75 ГАЛА К 84 АТСА	0 Aga E 0 Aat	GGCI A TAAJ	76 AAAG K 85 ACAA	0 GAAG E D 0 CTAG	ATTI F B	770 TGA E 1-re 860	L Dae	GCA A t <u>CAA</u>	780 GCA A 870 ATC	TTA L CTA	GGT G	79 CAC H 88 <u>3CT</u>	0 CAP Q 0 <u>AGC</u>	H H	IGC A	BOO ICAT H B90 AGGI	N N	IGAC E	810 STAT Y 900 AGA
AAGA K I CAAGO Q A	A A 8: CAAA K	30 CGA D 20 ACT L	L L AGCA	74 3AAA E T 83 3AAA E K	0 CAA. K 0 AAG. D	AAC: L ATGJ D	75 ГАЛА К 84 АТСА Q	0 AGA E 0 AAT I	GGCI A TAAJ K	76 AAAG K 85 ACAA Q	0 GAAG E D 0 CTAG L E	ATTI F B AA <u>GA</u> E	770 TGA E 1-re 860 GCA Q	L Dae AAAA K	GCA A t <u>CAA</u> Q	780 GCA A 870 <u>ATC</u> I	TTA L CTA L	GGT G GAT D	79 CAC H 88 <u>3CT</u> A	0 Q 0 AGC S	H H CG1 R	IGC A I I I I I K	800 ICA1 H 390 AGGI G	TAA	IGAO E A <u>GC</u> A	810 STAT Y 900 AAGA R
AAGA K I CAAGO Q A	A B CAAA K 9	30 CGA D 20 ACT L 10	ITTA L	74 3AAA E T 83 3AAA E K 92	0 CAA K 0 AAG. D	AAC: L ATGJ D	75 FAAA K 84 ATCA Q 93	0 AGA E 0 AAT I 0	GGCI A TAAJ K	76 AAAG K 85 ACAA Q 94	0 GAAG E D 0 CTAG L E 0	ATTI F B AA <u>GA</u> E	770 TGA E 1-re 860 GCA Q 950	L Dae AAAA K	GCA A t <u>CAA</u> Q	780 GCA A 870 <u>ATC</u> I 960	TTA L CTA	GGT G GAT D	79 CAC H 88 <u>3CT</u> A 97	0 Q 0 AGC 5	H H CGT R	IGC A EAAJ K	800 ICA H 390 AGGI G	TAA N TACI T B2	IGAC E A <u>GC</u> A -re	810 STAT Y 900 AGA R peat 990
AAGA K I CAAGO Q A <u>GACC</u>	A B B CAAA K 9 TTGA	30 CGA D 20 ACT L 10 AGC	ITTA L	74 3AAA E T 83 3AAA E K 92 <u>2GCC</u>	0 CAA. 0 AAG. D 0 <u>AAG:</u>	AAC: L ATGI D	75 ГАЛЛ К 84 АТСА Q 93 <u>АЛ</u> ЛЛ	0 AGA E 0 AAT I 0 AGC	GGCI A TAAJ K TACO	76 AAAG K 85 ACAA Q 94 3GAA	0 GAAG E D 0 CTAG L E 0 GCTG	ATTI F B AA <u>GA</u> E	770 TGAJ E 1-re 860 <u>GCAJ</u> Q 950 XAAA	L pae AAAA K	GCA A t <u>CAA</u> Q CTC	780 GCA A 870 <u>ATC</u> I 960 AAA	TTA L CTA L GCA	GGT G : GAT D : GAG	79 CAC H 88 <u>3CT</u> A 97 CTT	0 Q 0 <u>AGC</u> 5 0 GCA	H CGT R	IGC A I I I I I I I I I I I I I I I I I I	800 ICAT H 390 AGGI G 80 FACA	TAA N TACI T B2	IGAC E A <u>GC/</u> A -re	810 3TAT Y 900 <u>AGA</u> R peat 990 <u>AAA</u>
AAGA K I CAAGO Q A <u>GACC</u> D L	A B CAAA K 9 <u>TTGAI</u> E	30 CGA D 20 ACT L 10 <u>AGC</u> A	ITTA L AGCA A L I <u>GTIV</u>	74 3AAA E T 83 3AAA E K 92 <u>5GCC.</u> R Q	0 CAA. K 0 AAG. D 0 <u>AAG.</u> A	AAC L ATGJ D <u>CTAJ</u> K	75 FAAA K 84 ATCA Q 93 AAA K	0 AGA E 0 AAT I 0 AGC	GGCI A TAAJ K TACO T	76 AAAG K 85 ACAA Q 94 SGAA E	0 GAAG E D CTAG L E 0 GCTG. A E	ATTI F B AA <u>GA</u> E AATT	770 TGA E 1-re 860 <u>GCA</u> Q 950 XAAC N	L pae AAAA K N	GCA A t <u>CAA</u> Q CTC. L	780 GCA A 870 <u>ATC</u> I 960 AAA K	TTA L CTA L GCA	GGT G GAT D GAG E	79 CAC H 88 <u>3CT</u> A 97 CTT	0 Q 0 <u>AGC</u> S 0 GCA	ACA: H R R AAA/ K	IGC A I I I I I I I I I I I I I I I I I I	BOO ICAT H 390 AGGI G 80 FACA T	n N T B2 <u>GAJ</u> E	IGAC E AGC/ A - re ACA/ Q	810 STAT Y 900 MAGA R peat 990 MAAA K
AAGA K I CAAGO Q A <u>GACC</u> D L	7 TTGC A 8: ZAAA K 9: <u>TTGA</u> E	30 CGA D 20 ACT L 10 AGC A	ITTA L AGCA A CATN	74 3AAA E T 83 3AAA E K 92 <u>56CC</u> R Q	0 CAA K 0 AAG D 0 <u>AAG</u> A	AAC: L ATGI D <u>CTAJ</u> K	75 FAAA K 84 ATCA Q 93 <u>A</u> AA K	0 AGA E 0 AAT I 0 AGC A	GGCI A TAAJ K TACO T	76 AAAAG K 85 ACAA Q 94 GGAA E	0 GAAG E D 0 CTAG L E 0 GCTG A E	ATTI F B AA <u>GA</u> E AATT L	770 TGAI E 1-re 860 GCAI Q 950 AAAC N	L pae AAAA K CAAC N	GCA A t Q CTC L	780 A 870 A T 960 AAA K	CTA L GCA A	GGT G GAT GAG GAG E	79 CAC H 88 <u>3CT</u> A 97 CTT	0 Q 0 <u>AGC</u> S 0 GCA	ACA H R R AAAZ K	IGC: A <u>FAAI</u> K S GT: V	BOO ICAT H 390 AGGI G 80 FACA T	n N T B2 G <u>GA</u> E	E AGCI A -re A CAI	810 STAT Y 900 MAGA R peat 990 MAAA K
AAGA K I CAAGC Q A <u>GACC</u> D L CAAAD	A B CAAAA K 9 TTGAA E 10	30 CGA D 20 ACT L 10 AGC A A 00 AGA	ITTA L AGCA A CA V	74 3AAA E T 83 3AAA E K 92 <u>5GCC</u> R Q 101 AGTC	0 CAA. K 0 AAG. D 0 AAG. A 3 TAJ	AAC' L ATGJ D <u>CTAJ</u> K	75 TAAA K 84 ATCA Q 93 AAA K 102 STACE	0 AGA E 0 AAAT I 0 AGC' A AGC	GGCI A TAAI K TACO T	76 K 85 ACAA Q 94 3GAA E 103	0 GAAG E D 0 CTAG L E 0 GCTG. A E 0 CTTG.	ATTI F B AA <u>GA</u> E AATT L 1	770 TGA E 1-re 860 <u>GCA</u> Q 950 AAAC N 040 AGT	L pae AAAA K XAAC N	GCA A t CAA CTC L 1 CAA	780 GCA A 870 <u>ATC</u> I 960 AAA K 050 GCA	TTA L CTA L GCA A	GGT G : GAT D : GAG GCAC	79 CAC H 88 <u>3CT</u> A 97 CTT L 106 CAA	0 Q 0 <u>AGC</u> S 0 GCA A 0 GTI	ACA: H R K K	IGC A EAAI K S AGT V 1(AGC	800 H 390 AGGT G 880 FACA T 070	N N T B2 GAJ E	IGAC E AGC/ A -re ACCA/ Q 1	810 STAT Y 900 AAGA R peat 990 AAAA K 080 CAA
AAGA K I CAAGC Q A <u>GACC</u> D L <u>CAAAT</u> Q I	A 8: CAAA K 9: TTGAI E 10: CTTI	30 CGA D 20 ACT L 10 AGC A A O 0 D	ITTA L AGCA A CA N IGTT V I IGCT/ A	74 3AAA E T 83 3AAA E K 92 <u>CGCC</u> R Q 101 AGTC 3 R	0 CAA. K 0 AAG. D 0 <u>AAG.</u> A 3 <u>TAJ</u> K	AAC L ATGJ D C <u>TAJ</u> K AAGC	75 FAAA K 84 ATCA Q 93 AAAA K 102 STACI T	0 AGA E 0 AAAT I 0 AGC A AGC	GGCI A TAAJ K TACO T R R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D	0 GAAG E D 0 CTAG L E 0 GCTG. A E 0 CTTG. L E	ATTI F B AAA <u>GA</u> E AATTI L 1 AAAGC A	770 TGAJ E 1-re 860 <u>GCAJ</u> Q 950 N AAAA N 040 <u>AGTT</u> V	ACTA L pae AAAAA K CAAC N N	GCA A t CAA Q CTC L 1 CAA	780 GCA A 870 <u>AATC</u> I 960 AAA K 050 <u>GCA</u>	TTA L CTA L GCA A A AAA K	GGT G GAT D GAG GAG GCAG	79 CAC H 88 <u>3CT</u> A 97 CTT L 06 CAA	0 CAF Q 0 <u>AGC</u> S 0 GCA A 0 GCA V	ACA: H R R AAA/ K CGA/ E	A A K S A GTT V 1(A GCT A	BOO H 3900 G 9800 FACA T 0700 NGCT A	N N E B2 GAJ E	IGA(E A <u>GC/</u> A -re <u>ACA/</u> Q 1 1 CAA/ K	810 STAT Y 900 AGA R peat 990 AAAA K 080 ACAA Q
AAGA K I CAAGC Q A <u>GACC</u> D L <u>CAAAT</u> Q I	A 8: CAAA K 9 <u>9</u> TTGAA E 100 CTTTA	30 CGA' D 20 ACT. L 10 AGC' A 00 AGA' D	L : AGCAI A : <u>IGTIN</u> V :	74 33AAA E T 83 33AAA E K 92 <u>CGCC</u> R 92 <u>CGCC</u> R 92 101 101 AGTC 3 R	0 CAA K 0 AAAG D 0 AAAG A A STAJ K	AAC L D CTAJ K G	75 FAAA K 84 ATCA Q 93 <u>AAAA</u> K 102 <u>STAC:</u> T	0 AGA E 0 AAAT I 0 AGC A A C A A	GGCI A TAAJ K TACC T R R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D	0 GAAG E D CTAG L E GCTG A E 0 CTTG L E C1-r	ATTI F B AAA <u>GA</u> E L I I A <u>AAGC</u> A epe	770 TGAJ E 1-re 860 <u>GCAJ</u> Q 950 NAAAC N 040 AGT V at	ACTA L pae AAAAA K CAAC N R	GCA A t CAA CTC L 1 CAA	780 GCA A 870 <u>ATC</u> I 960 AAA K 050 <u>GCA</u> A	TTA L CTA L GCA A A AAA K	GGT G GAT D GAG GCAG A (79 CAC H 88 <u>3CT</u> A 97 CTT L 06 CAA	0 Q 0 AGC S 0 GCA S 0 GCA A 0 GTI V	ACA: H R AAA/ K SGA/ E	A A FAAA K S GTT V I (AGCT A	BOO ICAT H 390 AGGI G 70 FACA T 70 FGCT A	N N E CAC T B 2 GAA E CTX L	IGAO E AGC/ A -re Q 1 I XAAA K	810 STAT Y 900 <u>AAGA</u> R peat 990 <u>NAAA</u> K 080 <u>CAA</u> Q
AAGA K I CAAGC Q A GACC D L CAAAT Q I) TTGC: A 8. CAAA K 9. TTGAA E 10. CTTM L	30 CGA' D 20 ACT L 10 AGC A A GO AGA' D 90	ITTA L : AGCA A : IGTTM V) IGCTM A :	74 3AAA E T 83 3AAA S K 92 CGCC R Q 101 AGTC G R 110	0 CAA. K 0 AAG. D 0 AAGC A 3 TAJ K 0	AAC: L ATGJ D <u>CTAJ</u> K <u>AAGC</u> G	75 TAAA K 84 ATCA Q 93 AAAA K 102 STAC: T 1111	0 AAGA E 0 AAAT I 0 AGCC A C A A O	GGC/ A TAA/ K TACC T R AAG/ R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D 112	0 GAAG E D 0 CTAG CTAG 0 GCTG. A E 0 CTTG. L E C1-r 0	ATTI F B AA <u>GA</u> E AATTI L 1 AAGC A epe 1	770 TGAJ E 1-re 860 <u>GCAJ</u> Q 950 N 040 N 040 N 040 N 040 N 040 N 130	ACTA L pae AAAA K CAAC N N CGCC R	GCA A t CAA Q CTC. L 1 CAA Q	780 GCA 870 A 870 I 960 AAAA K 050 GCA A 140	TTA L GCA A AAA K	GGT G : GAT D : GAG GCAG A (1	79 CAC H 88 3 <u>CT</u> A 97 CTT L 106 CAA 2	O CAP Q O AGC S O GCA A O GTI V O	ACA: H R AAA/ K CGA/ E	1000 A (1000 A (1000 A A (1000 A A (1000 A A (1000) A A (1000) A A (1000) A A (1000) A A (1000) A A (1000) A A (1000) A A (1000) A A (1000) (1000) (100) (1000) (1000) (100) (100) (1000) (1000) (100) (100) (1000) (100) (100) (1000) (100)	B00 TCAT H 390 AGGT G 7 070 TACA T 070 NGCT A	TAA N B2 GAN E	IGAC E AGC/ A -re Q 1 XAAA K 1	810 3TAT Y 900 AAGA R peat 990 AAAA K 080 AAAA Q 170
AAGA K I CAAGC Q A <u>GACC</u> D L <u>CAAAA</u> Q I CTTG2 L E	7 TTGC: A 8: ZAAAI K 9: <u>TTGAI</u> E 100: <u>CCTTII</u> L 100: AGGAI E	30 CGA D 20 ACT: L 10 AGC: A D 90 ACAJ O	ITTA L AGCA A C IGTI V V I IGCTI A AAACI N	74 33AAA E T 83 33AAA 5 K 92 <u>2GCC</u> 2 <u>2GCC</u> 3 Q 101 <u>3 GTC</u> 3 R 110 <u>3 GGA</u> 3 I	0 CAA. K 0 AAG. D 0 AAG. A 3 <u>TAI</u> K 0 TTTX S	AAC L ATGJ D <u>CTAJ</u> K G G CAGA	75 TAAA K 84 AATCA Q 93 AAAA K 102 TAC: T 1111 AGCZ A	0 AGA E 0 AAAT I 0 AAGC A A C AAGC S	GGCJ A TAAJ K TACO T R R R CCGJ R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D 112 TAAG K	0 GAAG E D 0 CTAG L E 0 GCTG. A E 0 CTTG. L E C1-r 0 CTTG. GGTC'	ATTI F B AA <u>GA</u> L I <u>AAGC</u> A epe 1 I <u>TTCG</u> R	770 TTGAJ E 1-re 860 <u>GCAJ</u> Q 950 N AAAC N 040 <u>AGTT</u> V at 130 <u>CCGT</u> R	ACTA L pae AAAA K CAAC N CCGC R R	GCA A t CAA Q CTC L 1 CAA Q 1 TTG L	780 GCA 870 A 870 I 960 AAA K 050 GCA 140 GAC	L CTA L GCA A A AAA K	GGTM G : GAGC GAGC E 1 J GCAC A (J I TCAC	79 CAC H 888 <u>3CT</u> A 97 CTT L 106 CAA 2 L 115 C <u>GT</u>	Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	ACA H R AAA/ K E GGA/ E	A A K K A G C A A G C C A A C C A A C C C A A C C C A A C	B00 TCAN H 390 AGGT G 980 TACA T 070 TGCT A 160 SAAG K	N N E CAC E CAC CAC CAC	IGAC E AGC/ A -re Q 1 I XAGC/ K I I XAGTI V	810 3TAT Y 900 <u>AAGA</u> R peat 990 <u>AAAA</u> K 080 <u>AAAA</u> Q 170 <u>K</u>
AAGA K I CAAGO Q A <u>GACCT</u> D L <u>CAAAT</u> Q I CTTG2 L E) TTGC: A 8: ZAAAJ K 9 9 <u>TTGAA</u> E 100 <u>CCTTM</u> L 100 L 100 CCTTM L	30 CGA' D 20 ACT: L 10 AGC: A D 90 ACAJ Q	ITTA L AGCA A S C <u>IGTTY</u> V I <u>IGCT</u> A A AAACI N I	744 SAAA E T 83 SAAA 92 CCCC Q 92 CCCC Q 101 AGTC S R 110 AGGA X I	0 CAA K 0 AAG D 0 AAG A A S TAI K 0 FTTC S	AAC' L ATGJ D CTAJ K G G CAGJ E	75 FAAA K 84 ATCA Q 93 <u>AAAA</u> K 1022 <u>5TAC</u> T 1111 AAGCJ A	0 AGA E 0 AAAT I 0 AAGC A A 0 AAGC S	GGCJ A TAAJ K TACO T R R R CCGJ R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D 112 TAAG K	0 GAAG E D 0 CTAG CTAG 0 CTTG 0 CTTG L E C1-r 0 GGTC G L	ATTI F B AA <u>GA</u> L I L I L E E E R R	770 TGAJ E 1-re 860 <u>GCAJ</u> Q 950 N 040 N 040 N 040 N 040 N 040 R R	ACTA L pae AAAA K CAAC N N <u>CGCC</u> R <u>CGAC</u>	GCA A t CAA Q CTC. L 1 CAA 1 TTG L	780 GCA 870 I 960 AAA K 050 GCA 140 GAC D	TTA L CTA L GCA A A AAA K GCA A	GGT G : GAT D : GAG GCA GCA A (I I TCA S . I	79 CAC H 88 3CT 97 CTT L 106 CAA 2 L 15 CGT R	Q Q AGC S Q GCA A Q GCA A Q GAA E	ACA H R AAAA K SGAA E	A FIGC' A FIAAI K C2 FIGC' A C2	800 ICAN H 390 G 980 ICACA T 070 ICACA T 070 ICACA C C C C C C C C C C C C C	TAA' N T B2 GAU E C C C C C C C C C C C C C C C C C C	IGAC E AGC/ A -re Q Q 1 1 CAA/ K 1 L CAA/ K 1 V V	810 STAT Y 900 MAGA R peat 990 MAAA K 080 CAAA Q 170 CGAA E
AAGA K I CAAGO Q A <u>GACCT</u> D L <u>CAAAT</u> Q I CTTG2 L E) TTGC A 8. 8. 7 8. 7 9. 7 7 7 6 100 2 7 7 6 100 2 7 7 6 100 2 7 7 6 100 2 7 7 100 2 8 8 8 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7	30 CGA D 20 ACT L 10 AGC A A CA D 90 ACA Q 80	ITTA L : AGCA A : I <u>GTTN</u> V) I <u>GCTJ</u> A : N 1	744 SAAA E T 83 SAAA 5 K 92 CGCC CGCC 101 101 AGTC 3 R 110 AGGA 3 I 119	0 CAA K 0 D 0 0 <u>AAAG</u> A 0 <u>3TAI</u> K 0 TTTC S 0	AAC' L ATGJ D CTAJ K AAGC G G CAGJ E	75 TAAA K 84 ATCA Q 93 <u>AAAA</u> K 102 <u>STAC</u> T 1111 NAGCI A 120	0 AGA E 0 AAT I 0 AGC A A C A A C S 0 0	GGCI A TAAJ K TACO T R R R CCGI R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D 112 TAAG K 121	0 GAAG E D 0 CTAG L E 0 GCTG CA E C 1-r 0 GGTC G G C 1-r 0 0 GGTC 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	ATTI F B AA <u>GA</u> L I I A <u>AGC</u> A epe 1 TCG R	770 TGAJ E 1-re 860 Q 950 N 040 AGT V at 130 CCGT R 220	ACTA L pae K CAAC N N <u>CGCC</u> N N <u>CGCC</u> D	GCA A t CAA Q CTC. L 1 CAA 1 TTG L	780 GCA 870 A 7 960 AAA K 050 GCA 140 GAC D 230	TTA L CTA L GCA A A A A K GCA A	GGT G : GAT D : GAG GCA GCA A (1 I TCA S . I I	79 CAC H 88 <u>3CT</u> A 97 CTT L 106 CAA 2 L 15 CGT R	0 CAF Q 0 AGC S 0 GCA A 0 GTI V 0 GAA E 0	ACA: H R AAAZ K SGAZ E SGAZ E	A FAAI K S AGT V 10 AGC A 11 TAAC K C2 12	B00 ICA H 390 AGGT G C B80 IACA T D70 IGCI A L60 SAAG K -re 250	N N T B2 GAU E C C C C C C C C C C C C C C C C C C	IGAC E AGC/ A -re Q 1 I CAAA K 1 I V V I 1 I I I I I I	810 STAT Y 900 MAGA R peat 990 MAAA K 080 MCAA Q 170 CGAA E 260
AAGA: K I CAAGG Q A GACC: D L CAAAA Q I CTTGJ L E AAAAGJ	7 TTGC/ A 8. 2. 2. 2. 3. 3. 4. 100 2. 100 100 2. 100 2. 100 100 100 100 100 100 100 10	30 CGA D 20 ACT: L 10 AGC: D 90 ACAJ Q 80 ACAJ	ITTA L AGCA A C I <u>IGTT</u> V V I I <u>IGCT</u> A A AAACI N I	744 SAAA E T 83 SAAA S K 92 CGCC CGCC R 92 101 101 AGCC S R 1100 AGCA X I 119 FTGAA	0 CAA.K 0 AAG. D 0 AAGC A 3TAJ K 0 STAJ S 0 CTGC	AAC: L ATGJ D CTAJ K G G CAGJ E	755 TAAA K 84 ATCA Q 93 <u>AAAA</u> K 102 <u>STAC</u> T 1111 AGCJ A 1200	0 AGA E 0 AAT I 0 AAGC A A 0 AAAG S 0 CAA	GGCJ A TAAJ K TACC T R R CCGI R R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D 1122 TAAG K 1211 CAT	0 GAAG E D 0 CTAG CTAG GGCTG 0 CTTG 0 CTTG 0 CCTTG 0 CCTTG 0 CCTTG 0 CCTTG 0 CCTTG 0 CCTTG 0 CCTTG 0 CCTAG CCTAG CCC	ATTT F B AAA <u>GA</u> E AAATT L 1 1 AAAGC A epe 1 TTCG R I 1 AAAGA	770 TGAJ E 1-re 860 Q 950 AAAA N 040 AGT V 130 CCGJ R 220 AAAA	ACTA L pae AAAA K CAAC N N <u>CGCC</u> D	GCA t CAA Q CTC. L 1 CAA Q 1 TTG L 1 1 TTG L 1 1 TTG	780 GCA A 870 ATC I 960 AAA K 050 GCA A 140 GAC D 230 TCA	TTA L CTA L GCA A A A A A GCA GCA GCA	GGT G GAT D GAG GCA A () J TCA S SCAJ	79 CAC H 88 3CT A 97 7 CTT L 06 CAA 2 L 106 CAA 2 L 115 CGT 2 L 124 4 4 CC	0 CAF Q 0 CAF S 0 CGTI V 0 CGTI 0 CGTI	ACA H R AAAA K GAA K GAA A A A A A A A A A A A	A TAAN K S A TAAN V 10 A C2 12 A C2 12 A C2	800 1CA1 H 390 G 6 80 1CA1 6 980 1CA1 6 980 1CA1 1 90 1 1 1 1 1 1 1 1 1 1 1 1 1	TAA' N B2 GAV E CTC L CAV Q Pea	IGAC E AGC/ A -re Q I I SAAA K I I AGTJ V V I I I I I I I I I I I I I I I I I	810 STAT Y 900 AGA R peat 990 AAAA K 080 CAA Q 170 NGAA E 260 NGAC
AAGA K I CAAGG Q A GACCT D L CTTGJ L E AAAGSJ K D	A R R SAAAI K 9 TTGAI E 100 NCTTI L 100 AGAI E 110 TTTTI L	30 CGA D 20 ACT L 10 AGC A D 90 ACAJ Q 80 AGCJ A CA	ITTA L AGCA A I I I I I I I I I I I I I I I I I I	74 33AAA E T 83 33AAA E K 92 <u>CGCC</u> R 92 <u>CGCC</u> R 92 <u>CGCC</u> R 92 <u>CGCC</u> R 92 <u>CGCC</u> R 92 <u>CGCC</u> R 101 119 119 TTGAA T	0 CAA. K 0 AAG 0 0 AAG A A 3 TAI 5 0 CTGC A	AAC: L ATGJ D CTAJ K G G CAGJ E CTGJ E	75 TAAA K 84 ATCA Q 93 AAAA K 102 STAC: T 1111 AGC: A 1200 L	0 AGAA E 0 AAAT I 0 AAGC A A C AAGC S 0 AAGC S 0 C C C C C C C C C C C C C C C C C	GGCJ A TAAJ K TACC T R R CCGJ R R CCGJ R R	76 AAAAG K 85 ACAA Q 94 SGAA E 103 AGAT D 112 TAAG K 121 SGTT. V	0 GAAG E D 0 CTAG 1 E 0 CTTG 0 <u>CTTG</u> 1 E C1-r 0 <u>GGTC</u> 3 L 0 AAAG K E	ATTT F B AAA <u>GA</u> L 1 AAGC A P P P C C G R 1 AAGA F	770 TGAJ E 860 Q 950 AAAA N 040 AGT V 040 AGT V 220 R 220 K	ACTA L pae AAAA K CAAC N CGGC D CAAA Q	GCA A t CAA Q CTC. L 1 CAAA Q 1 TTG L 1 : ATC'	780 GCA A 870 AAC I 960 AAA K 050 GCA A 140 GAC D 230 TCA S	TTA L GCA A A A A A K GCA A GAC D	GGT G GAT D GAC E C GCAC A C C CAC S S C A S C A S C A	79 CAC H 888 GCT A 97 CTT L 06 CAA 2 L 15 CGT L 124 AGC 3	0 CAP Q 0 CAP CAP CONT	ACA H R AAA K K SGAA E SGAA A CAA Q	IGC" A K <u>FAAA</u> K V 10 AGC" A C2 I1 XAAC K C2 I2 AGGT G G	800 1CA1 H 390 G 980 TACA T 070 NGCI A 60 SAAG K -re 250 CCTI L	TAA N T B2 GAU E C L C C C C R	AGC/ A A -re Q 1 2AAA K 1 1 XGT7 V at 1 2CG7 R	810 STAT Y 900 <u>AAGA</u> R peat 990 <u>AAAA</u> K 080 <u>CAA</u> Q 170 <u>COAA</u> 2 260 <u>COAC</u> D
AAGAA K I CAAGGG Q A GACCT D L CTTGJ L E AAAGAA K D	10° 10° 10° 10° 10° 10° 10° 10°	30 CGA D 20 ACT: 10 AGC: A 00 AGC: A 00 AGC: A 0 0 ACAJ 0 80 AGC: A 70	ITTA L AGCA A C IGCTI V V V V V V V V V V V V V V V V V V V	74 3AAA E T 83 3AAA S K 92 2 <u>CGCC</u> 2 101 <u>AGCC</u> 3 R 110 <u>AGCC</u> 3 R 110 <u>AGCC</u> 3 R 110 <u>AGCC</u> 3 R 110 <u>AGCC</u> 3 R 112 119 119 128 128	0 CAA K 0 D 0 AAGC A A 0 STAJ K 0 STAJ S 0 CTGC A 0	AAC L ATGJ D CTAJ K G G CAGJ E E CTGJ E	75 FAAA K 84 ATCA Q 93 AAAA K 102 STAC: T 1111 AGC: A 1200 L 1299	0 AGA E 0 AAAT I 0 AGC A A C C C D 0 C C C 0 0 C C 0 0 C C 0 0 C C 0 0 C C 0 0 C	GGCI A TAAJ K TACC T R R R R R R R R R R R R R R R R R R	76 AAAAG K 85 ACAA Q 94 GGAA E 103 AGAT D 112 TAAG K 121 GGTT. V 130	0 GAAG E D 0 CTAG L E 0 CTTG 0 CTTG 1 E C1-r 0 0 GGTC ² 0 AAAG K E 0 0 0 0 0 0 0 0 0 0 0 0 0	ATTI F B AAA <u>GA</u> L 1 <u>AAGC</u> A epe 1 <u>TTCG</u> R 1 AAGA E 1	770 TGAJ E 860 Q 950 XAAAC N 040 AGT V 130 CCGT R 220 K 310	ACTA L pae AAAA K CAAC N <u>CGGC</u> R D	GCA A t CAA Q CTC. L 1 CAA Q 1 TTG L 1 : ATC' I	780 GCA 870 A 870 A A 050 GCA A 140 GCA D 230 C CA S S 320	TTA L CTA L GCA A A A A K GCA A GAC	GGTY G : D : GAGC E 1 ; GCAG A (1 ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;	79 CAC H 88 3CT A 97 CTT L 06 CAA 2 L 15 CGT 2 CTT L 124 AGC 3 3 3	Q Q AGC S O GCA A O GTI V O GAA E O CGT R O CGT	ACA H R AAAZ K GAA E GAA A Q	IGC A K S G G G G G G G G G G G G G G G G G G	800 ICAT H 390 AGGT G 980 FACA T 070 IGCT A 160 SAAG K -re 250 K -re 250 L	RAA N T B2 GAU E CTC L CTC L CTC R	AGC/ E AGC/ A -re Q 1 CAAA K 1 CAAA K 1 CAAA K 1 CAAA K 1 CAAA R R	810 STAT Y 900 AAGA R peat 990 AAAA K 0800 ACAA Q 170 NGAA E 260 NGAC D 350
AAGAA K I CAAGGG Q A GACCT D L CTTGJ L E AAAGJA K D	100 CTTA 88: CAAAA K 90 CTTA E 100 CTTA L 100 CTTA L 110 L	30 CGA D 20 ACT L 10 AGC D 90 ACA D 90 ACA Q 80 ACA Q 80 ACA A C A TO A TO	ITTA L : AGCA A COLL A AGCA M I A AAAC N I AAAC N I AAAC	74 3AAA E T 83 3AAA S K 92 20 20 20 20 20 20 20 20 20 2	0 CAA. K 0 D 0 AAGC A A 0 STAI K 0 STAI CTGC A 0 CTGC A 0 CTGZ	AAC L D CTAJ K G G CAGJ E CAGJ E	755 FAAA K 84 ATCA Q 933 AAAA K 102 STAC: T 1111 LAGCI A A L 1200 L L 1290 ACAJ	0 AGAN E 0 AAAT I 0 AAGC A A A C A A A C A A A C A A C A A C A C A C A C A C A C A C A C A C A	GGCI A TAAJ K TACO T R R CCGI R R R R R R R R R R R R R R R R R R R	766 AAAAG K 855 ACAA Q 944 SGAA E 103 AGAT 121 SGTT. V 130 AAAA	0 GAAG E D 0 CTAG CTAG 0 GCTG L E C 1-r 0 GGTC G L 0 AAAG K E 0 GGTC C 1 0 GGTC C 1 0 0 GGTC C 1 0 0 C 1 0 0 C 1 0 0 C 1 0 0 0 C 1 0 0 0 C 1 0 0 0 C 1 0 0 0 C 1 0 0 0 C 1 0 0 0 C 1 0 0 C 1 0 C 1 0 C 1 0 C 1 0 C 1 C 1 0 C 1 C 1 0 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1	ATTI F B AAA <u>GA</u> E AAATT L I I AAAGA R I I AAGA E I I AAGA	770 TGAJ E 860 Q 950 N 040 AAAAC N 040 AGTT V at 130 CCGT R 220 K 310 K 310	ACTA L pae AAAA K XAAC N XCGC N XCGC R D	GCA A t CTC L 1 CAAC Q 1 TTG L 1 1 AAC	780 GCA 870 A 870 A A C C C C C C C C C C C C C C C C C	TTA L CTA L GCA A A A A A A A A A A A A A	GGTY G : GAGY GAGY GGCAY A (J J GGCAY A (J J GCCAY A (J J CTCAY A (J J CTCAY A (J J CTCAY A (J J) J J) J J J J J J J J J J J J J J	79 CAC H 88 3CT A 97 CTT L 06 CAA 97 CTT L 106 CAA 2 2 115 CCT 124 A GCC 3 3 3 CT	Q Q Q AGCA S O GCA A O GCA C GTI V O GAA C GTI R O C GTI R O C GTI R	ACA H R AAA K SGA2 E A CAA Q CTT	IGC: A K S G TAAC V 10 A G C 2 A C 2 C 2 C C 2 C C C C C C C C C C	8000 ICAT H 3990 AGGT G 980 TACA T 980 ICCT I L 60 SAAG K -re 250 K -re 250 K 1 L	TAA' N B2 B2 GAU E CTC L CTC R CCTC R	IGAO E AGCJ A A -re Q 1 1 CAAA K 1 1 1 CAAA R 1 1 CAAA R 1 1 CAAAA	810 STAT Y 900 AAGA R peat 990 AAAA K 080 ACAA Q 170 R 260 CAAA D 350 CAAA
AAGA K I CAAGG Q A GACCT D L CAAAAT Q I CAAAAT Q I CAAAAT K D TTGGAA	A R R R SAAAA K 9 <u>TTGAA</u> E 100 <u>CTTTA</u> L 110 <u>CTTTA</u> L 12 <u>CCGCA</u>	30 CGA D 20 ACT L 10 AGC A D 90 ACA D 90 ACA Q 80 ACA X Q 70 ATC S	ITTA L AGCA A IGTTM V I IGTTM V I I AAAC N I AAAC N I AAAC R I I I I I I I I I I I I I	74 33AAA E T 83 33AAA 5 K 92 CGCC 7 Q 101 101 107 3 R 110 110 119 TTGAA 5 T 128 3AAG 2 T 128 3AAA	0 CAA K 0 D 0 AAGC A 3 TAI K 0 CTGC A CTGC A CTGC A CTGC A CTGC A	AAC L D CTAJ K AAGC G CAGJ E CAGJ E K G K	75 FAAA K 84 ATCA Q 93 <u>AAAA</u> K 102 <u>5TAC</u> T 1111 AGCJ A 120 L 129 Q	0 AGAA E 0 AAAT I 0 AAAC S 0 AAAGC D 0 AAGC D 0 AAGC V	GGCI A TAAJ K TACO T R R CCGI R R R R R R R R R R R R R R R R R R R	766 AAAAG K 85 ACAA Q 944 GGAA E 103 AGAT 121 SGTT. V 130 AAAA K	0 GAAG E D 0 CTAG CTAG 0 GCTG L E C 1-r 0 GGTC G L 0 GGTC G L 0 0 AAAG K E 0 0 GCTT A L	ATTTI F B E AAA <u>GA</u> E I I I I I I I I I I I I I I I I I I	770 TGA E 1-re 860 Q 950 XAAA N 040 N 040 N AGTT R 220 R 220 K 310 AGAA K 310 AGAA E	ACTA L pae AAAA K CAAAC N CAAAC N CAAAC CAAA Q Q GCAA	GCA A t CTC. L 1 CAAA Q 1 TTG L 1 1 1 AACC N	780 GCA 8700 ATC 960 AAA K 050 GCA 1400 GAC D 2300 TCA S 1320 AGC S	CTA L GCA A A A A A A A A A A A A A A A A A A	GGTN G : GATC D : GAC E 1 J GCAC A (J ICAC S .1 J GCAC A (J ICAC S .1 J ICAC C J J CAC C J C L J	79 CAC H 888 3CT A 97 CTT L 106 CAA 2 L 15 CGT L 124 A GCC 3 3 3 CTC 4 3 3 CTC 4 4 3 3 CTC 4 1 2 4 1 2 4 3 3 CTC 4 1 1 5 CTT 4 1 1 5 CTT 4 1 1 5 CTT 4 1 1 5 CTT 4 1 1 5 CTT 4 1 1 5 CTT 4 1 5 C CTT 4 1 5 C CTT 4 1 5 C C C C C C C C C C C C C C C C C C	0 CAP Q 0 CAP CAP 0 CAP CAP CAP CAP CAP CAP CAP CAP	ACA H R AAA K K SGAA E SGAA A CAA Q CTTI L	A FAAA K S AGT V 10 AGC X A C22 12 AGG G 13 SGAA E	800 TCAN H 390 390 390 390 390 390 390 390	TAA' N T B2 B2 GAU E C C C C C C C C C C C C C C C C C C	IGAO E AGCA A A CAAA V I I CAAA R I I CCGJ R I I CCGJ R I I CCGJ R I I CCGJ R I I I CCGJ R I I I I I I I I I I I I I I I I I I	810 STAT Y 900 <u>AAGA</u> R peat 990 <u>LAAA</u> K 0.080 <u>CAAA</u> Q 1.70 <u>CAAA</u> 2.260 <u>CAAC</u> D 3.500 <u>CAAA</u> X
AAGAX K I CAAGG Q A GACCT D L CAAAA Q I CAAAA Q I CAAAA C CAAAA C CAAAA C C CAAAA C C CAAAG X C CAAGC C C C CAAGG C C C C C C C C C C	A 88. CAAAA K 99. TTGAA E 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L 100. CTTM L L L L L L L L L L L L L	30 CGA D 20 ACT 10 AGC A D 00 AGA D 90 ACAJ Q 80 AGC X C A C C A C C C A C C C A C C A C C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C A C C C A C C C A C C C A C C C A C C C C A C C C C C A C	ITTA L AGCA A COTT N L COTT R I L L L L L L L L L L L L L	74 33AAA E T 83 33AAA E K 92 CGCC R Q 101 101 AGTC: 5 R 1100 AGGA: 119 1199 TTGAA C T 128 SAAGC: A	0 CAA K 0 AAGC D 0 3 TAA K 0 7 TTX S 0 7 TTX S 0 7 TTX A 0 7 TTX A	AACC L D CTAJ K AAGC G CAGJ E CAGJ E K K	75 TAAA K 84 ATCA Q 93 <u>JAAA</u> K 1022 <u>TT</u> 1111 JAGCJ A L 1200 <u>JACAJ</u> Q	0 AGA E 0 AAAT I 0 AAGC A A C 0 AAGC D 0 AAGT V 0	GGCJ A TAAJ K TACC T R R R R R R R R R R R R R R R R R R	766 AAAAG K 85 ACAA Q 944 GGAA E 103 AGAT 121 SGTT. V 130 AAAA K	0 GAAG E D 0 CTAG CTAG CTAG CTAG CTAG C CTAG C C C C C C C C C C C C C C C C C C C	ATTTI F B AAA <u>GA</u> E I I AAGC A epe 1 I MAGC R I I AAGA E I I CAGA	770 TGAN E 1-ree 860 Q 950 XAAA N 040 AGTT R 040 X 220 X AAAA K 310 AGAA E	ACTA L pae AAAAA K CAAC N CGCC R D CGAC Q Q Q Q Q Q Q Q	GCA A t CAA Q CTC. L 1 CAAA Q 1 TTG L 1 1 AATC' I I 1 1 AATC' N	780 GCA 8700 ATC 960 AAA K 0500 GCA 1400 GCA 1400 GAC 5 1 3200 AGC 5 1 4100 1 1 1 1 1 1 1 1 1 1 1 1 1	CTA L GCA A A AAA K GCA A A A A A A A A A A A A A A A A A A	GGTY G : GAT(D) GAC(E) J GCA(A (J J ICA(S .] J SCA(A S I ICA(S .] J ICA(L) J	79 CAC H 888 3CT A 97 CTT L 06 CAA 2 106 CAA 2 115 CCAA 2 124 L 06 CCAA 3 CTT L 06 CCAC 4 CCAC CCAC CCCAC CCAC CCAC CCAC CCAC CCAC CCAC CCCAC CCCAC C	0 CAP Q 0 CAP CAP 0 CAP CAP CAP CAP CAP CAP CAP CAP	ACA H R AAAA K SGAA E SGAA A Q Q CTTI L	A TAA K S A G TAA K C22 12 G G 13 S CAA E	800 ICAN H 390 AGGT G 980 ICAN T 070 ICAN T 160 ICAN T 160 ICAN T 160 ICAN T 160 ICAN T 160 ICAN K ICAN	RAA N B2 GAA E CTV L CTV L CTV L CTV L	IGAC E A C-re Q I I XAAA K I I XAAA R I I ZCCJI R I I ZCCJI R I I ZCCJI R	810 3TAT Y 9000 <u>NAGA</u> R peat 990 <u>NAR</u> 0080 <u>CCAA</u> Q 170 <u>CCAA</u> Q 170 <u>CCAA</u> Q 170 <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> A <u>CCAA</u> <u>CCAA</u> A <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAA</u> <u>CCAAA</u> <u>CCAAA</u> <u>CCAAA</u> <u>CCAAA</u> <u>CCAAAA</u> <u>CCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA</u>
AAGA'K I CAAGG Q A GACCT D L CAAAT Q I CAAAT Q I CAAAT CAAAT C AAAGJ K D TTIGGA GAGCT	A 8: CAAAA 8: CAAAA 8: CAAAA 8: CAAAA 8: CAAAA 8: CAAAA 10: CCTTA L 10: CCCTTA L 10: CCCTTA L 10: CCCTTA L 10: CCCCTA A L 10: CCCCCA A A A A A A A A A A A A A	30 CGA D 20 ACT L 10 AGA ACT A 10 ACA D 90 ACA D 90 ACA A S 5 50 AGA 2	ITTA	74 33AAA E T 83 33AAA S K 92 CGCC CGCC 2 101 AGCC 3 R 110 AGCA 3 119 FTGAA 4 2 2 3 119 FTGAA 5 119 128 3AAG 4 3 4 3 4 3 4 4 5 5 8 4 3 3 4 4 5 8 5 8 3 3 3 4 4 5 8 8 3 3 3 4 4 8 5 8 8 3 3 4 4 8 9 2 2 CGCC 3 8 8 8 8 8 8 8 8 8 8 8 8 8	0 CAA K 0 AAGC D 0 AAGC A C C C C C C C C C C C C C C C C	AAC L D CTAJ K AAG G CAGJ E CTGJ E K K AAG	755 TAAA K 84 ATCA Q 93 ATAAA K 1022 STACC T 1111 AGCJ ACAJ Q 1380	0 AGA E 0 AAAT I 0 AAGC A AGC A AGC V 0 AAAZ	GGCJ A TAAJ K TACO T R R CCGI R R R R R R R R R R R R R R R R R R R	76 AAAG K 85 ACAA Q 94 GGAA E 103 3 GGAT 121 SGTT. V 130 AAAA K 139	0 GAAG E D 0 CTAG CTAG GCTG. A E 0 CTTG. L E C1-r 0 CTTG. C1-r 0 C CTTG. C C C C C C C C C C C C C C C C C C C	ATTTI F B AAA <u>GA</u> E I I AAAGA A epe 1 I TTCG R I I AAGA E I I SAGA	770 TGAJ E 1-rea 860 Q 950 AAAA N 0400 AGTT V 2200 AAAAA K 310 AGAA E 4000	ACTA L pae k AAAA K XAAC N CGCC R CGAC CAA Q CGAA	GCA A t CAA Q CTC L 1 CAA Q 1 TTG L 1 1 AACC N 1 AACC	780 GCA 8700 ATC 1 9600 AAA K 0500 GCA 1400 GCA 1400 GCA 3200 AGC S 1 4100 CTTT	CTA L GCA A A AAA K GAC D AAAA K	GGTY G : GAC D : GAC E : I GCA S . I I GCA S . I I GCA L I I GCA C L I I GCA	79 CAC H 888 3CT 4 97 CTT 106 CAA 97 CTT 106 CAA 106 CAA 106 CAA 106 CAA 106 CAA 106 CAA 106 CAA 106 CAC 100 CAC 10 C CAC 10 C CAC 10 C CAC 10 C C CAC 10 C CAC 10 C CAC 10 C CAC 10 C	0 CAP Q 0 CAP CAP CAP CAP 0 CAP CAP CAP CAP CAP CAP CAP CAP	ACA H R K K GAM E SGAM C CAM Q CTTI L	A RGC ¹ A R CAAN K C C C C C C C C C C C C C	8000 TCAN H 3900 AGGT G 9800 TA 0700 TCAN T 0700 TCAN 1600 SAAG CTT L 1600 SAAG CTT L 1600 SAAG CTT L 1600 SAAG CTT SAAG SAAAG SAAAG SAAAG SAAAG SAAAG SAAAG SAAG SAAAG SAAAG SAAG SAAG	TAA' N B2 GAN E CTC Q Q Pea R CTT L	IGAC E AGC// A -re Q I I CAAM K I I CAAM K I I CCGJ R I I CAAC N I I I CAAM K I I I I I I I I I I I I I I I I I I	810 3TAT Y 9000 <u>JAGA</u> R peat 9900 <u>JAGA</u> K 0080
AAGA'K I CAAGG Q A GACCT D L CTTGJ L E AAAGJ K D TTGGA L D GAGCT E L	, rrdsca A 8: CAAAA K 9 <u>9</u> TTGAA E 100 CTTT L 100 CTTTAA E 110 CTTTAA L 122 CCCCA A A 130 CTTGAA E	30 CGA D 20 ACT L 10 AGA D 90 AGA Q 80 AGA A CA S 50 AGA A E	ITTA	74 3AAA 5 T 83 3AAA 5 K 92 2 2 2 2 2 101 AGGC 3 R 110 AGGC 3 R 110 AGGC 3 R 110 AGGC 3 R 110 AGGC 5 R 100 100 100 100 100 100 100 10	0 CAAL K 0 AAAG. D 0 0 AAAG. A 3 TAY S 0 CTGC A 0 CTGC A 0 CTGC A 1 CTGC A 1 CTGC A 1 CTGC A 1 CTGC A 1 C C AAG. C	AAC L D CTAJ K AAGC G CAGA E CAGA E CAGA E CAGA K K K AAC T	755 TAAAA K 844 ATCAA Q 93 ATCAA Q 93 ATCAA Q 102 STACC T 1111 AGCCI A Q 1386 C AGAAA E	0 AAGA E 0 AAAT I 0 AAGC A A C D 0 AAGC V 0 AAAA V 0 AAAA X	GGCJ A TAAJ K TACC T R R R R R R R R R R R R R R R R R R	76 AAAG K 85 ACAA Q 94 SGAA E 103 AGAT 121 CAAG K 121 SGTT. V 130 AAAA K 139 AAAA K	0 GAAG GAAG E D 0 CTAG GCTG. L E 0 CTTG. L E C1-r 0 GGTC' 0 GGTC' N E 0 0 GGCTGI A L 0 0 GGCTGI A E	ATTI F B AA <u>GA</u> E I I I I I I I I I I I I I I I I I I	770 TGAJ E 1-re 860 Q 950 AGTT V 0400 AGTT V 2200 AAAA K 310 AGAA K 310 AGAA K 4000 ACAA Q	ACTA L Dpae K CAAAC N CCGC N CCGC D CCAA Q Q GCAA A 1 GCAA	GCA A Q CTC. L 1 CAAU Q 1 TTG CAAU Q 1 1 TTG 1 1 AAC N 1 AAAC X AAAC	780 GCA 870 A 870 AAC 1 960 AAA K 050 GCA A 140 GAC 5 1 320 AGC 5 1 410 CTTX L 1	CTTA L CTA L GCA A A A A A A A A A A A A A A A A A A	GGTM GGTM D GGAGC GGAGC C GGCAC A SCAC A SCAC L J GGCAC A L J GGCAC A C C C C C C C C C C C C C C C C	79 CAC H 888 3CT A 97 CTT L 106 CAA 2 115 CGT 124 AGC 3 3 3 SCT 4 2 4 4 2 2 4 4 3 3 3 3 SCT 4 4 2 4 4 3 5 3 3 3 3 SCT 4 4 5 5 7 5 7 5 7 7 5 7 7 7 7 7 7 7 7 7	0 CAF Q 0 AGC S 0 GCA A 0 GCA 0 CGTI 0 CGTI 0 CGTI A 0 CGTA 0 CGTA A 0 CGTA A 0 CGTA A 0 CGCA A CGCA CGCA	ACA H R K K GAJ E SGAJ E CAA Q CTT L L AAA K	INCOM A K S S S S S S S S S S S S S S S S S S	B000 ICA) H 3900 G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) G 9800 ICA) ICA) G 9800 ICA) ICA) G 9800 ICA) ICA) G 9800 ICA)	TAA' N T B2 GAU E C C C C C C C C C C C C C C C C C C	IGAC E AGC/ A -re Q 1 CAAA K 1 CAAA R 1 CAAA N 1 1 GAAA E	810 3774 37 9000 <u>ANGA</u> R peat 9904 K 0800 CCAA Q 1700 CCAA Q 3500 CCAA K 4400 CCAA Q
AAGA'K I CAAGC Q A GACCT D L CAAAAA Q I CAAAAA C I CAAAAA Q I CTTGSA K D GAGCT E L	A R R SAAAA K 9 STGAA E 100 STTAA E 110 STTAA E 110 STTAA E 110 STTAA E 110 STTAA E 100 STTAA E STTAA E STTAA E STTAA E STTAA E STTAA E STTAA E STTAA E STTAA E STTAA E STTAA STTAA E STTAA E STTAA E STTAA STTAA STTAA E STTAA ST STTAA STTAA STTAA STTAA STTAA STTAA STTAA STTAA STTAA ST STTAA	30 CGA D 20 ACT L 10 AGC AGC D 00 AGA A C A C C C C C C C C C C C C C C C	ITTA L AGCA A CALL A IGTTM V I IGTTM V I IGTTM V I IGTTM V I I I A I I I I I I I I I I I I I	74 3AAA E T 83 3AAA 5 K 92 2 2 2 2 2 2 101 AGGC 3 R 110 AGGC 3 R 110 AGGC 3 R 110 AGGC 5 R 110 AGGC 5 K 101 AGGC 5 K 102 AGGC 5 K 103 AGGC 5 K 103 AGGC 5 K 103 AGGC 5 K 103 AGGC 5 K AGGC AGGC 5 K AGGC AGGC 5 K AGGCC AGGCC AGC	0 CAAL K 0 AAAG. D 0 AAAG. A 3 TAI S 0 CTGC A CTGC A CTGC A CTGC A L	AAC L D CTAJ K AAC G CAGJ E CAGJ E CAGJ K K K CAAC T	755 TAAA K 844 ATCA Q 93 MAAA K 102 STAC: T 1111 AGCAI Q 1290 ACAI Q 1380 E	0 AAGA E 0 AAAT I 0 AAGC A A 0 AAAG V 0 AAAA K	GGCJ A TAAJ K TACC T R R CCGJ R R R R R R R R R R R R R R R R R R R	76 AAAG K 85 ACAA Q 94 3GAA Q 94 3GAA D 103 103 1121 121 3GTT V 121 3GTT V 130 AAAA K 139 K	0 GAAG GAAG CTA	ATTI F B AA <u>GA</u> E I I I AAGA E I I AAGA E I I CAGA I I I CAGA I I I I CAGA I I I I I I I I I I I I I I I I I I	7700 7TGAI E 1-rea 8600 Q 9500 N 0400 AGTH V at 1300 CCGT R 2200 AAAAC R 3100 AGAAA E 4000 ACAAAC Q	ACTA L Dpae K CAAA R Q CCAA Q GCAA A L GCAA	GCA A Q CTC. L 1 CAAC Q 1 TTG- CAAC I I I I I I I I I I I I I I I I I	780 GCA 870 A 870 AAC 1 960 AAA K 050 GCA A 140 GAC 5 1 3200 AGC 5 1 410 CTTX L 1	CTTA L GCA A AAAA K GAC D AAAA K	GGTM G : GATM D : GGAGC E 1 1 GGCAG A (1 CCAC S .1 1 SCCAC L 1 1 GCCAC A F	79 CAC H 888 3CT A 97 CTT L 106 CAA 2 115 CCAA 2 115 CCAA 3 CTA 4 2 CAA 4 2 CTT L 106 CAA 4 2 CTT L 106 CAA 4 CTT L 106 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CAA 4 CTT L 106 CTT L 107 CTT L 106 CAA 4 CTT L 106 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTT L 107 CTTT L 107 CTTT L 107 CTTT L 107 CTTTT L 107 CTTTT L 107 CTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	0 CAZ Q 0 CAZ CAZ CAZ 0 CAZ CAZ CAZ CAZ CAZ CAZ CAZ CAZ	ACA: H R K K GAJ E CAJ A Q CCTT L AAAA K	TRECT A K S S S S S S S S S S S S S S S S S S	B000 ICA) H 3900 G B000 FACA T 0700 FACA T 1000 FACA T 1000 FACA T 1000 FACA T 1000 FACA T 1000 FACA T 1000 FACA T 1000 FACA T 1000 FACA T 1000 FACA FAC	TAA' N B2 GAU E NCTX L CCCX L C C C C	IGAO E AGC/J A CAA Q I I CAAA K I I CAAA R I I CAAA R I I CAAA R I I CAAA R I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I CAAA K I I I I CAAA K I I I I CAAA K I I I I CAAA I I I I I I I I I I I I I	810 3TAT Y 9000 <u>NAGA</u> R peat 9900 <u>NAGA</u> R 0800 <u>NAGA</u> K 0800 <u>NAGA</u> K 0800 <u>NGA</u> <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2600 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 2000 <u>C</u> 200

1450 1460 1465 TTAGCGAAACAAGCTGAAGAACTTG L A K Q A E E L

平成6年5月20日

Fig. 3 Comparison of the leader sequences or of the deduced signal peptide of M types 3 and various M genes. (A) Leader sequences are subdivided according to structural domains of the resulting leader peptide. (B) Signal peptides are divided into basic (B), hydrophobic (H) and cleavage (C) regions. Unknown sequences and amino acids are indicated by asterisks.

(A)

Basic	region		Primer	
M3	*****	***	***TATTCGCTTAGAAAAT1	AAAAACAGGAACG
M12	ATGGCTAA	AAATACCACGAATAGA	CACTATTCGCTTAGAAAATT	AAAAACAGGAACG
M1	ATGGCTAA	AAATAACACGAATAGA	CACTATTCGCTTAGAAAATT	TAAAAACAGGAACG
M6	ATGGCTAA	AAATAACACGAATAGA	CACTATTCGCTTAGAAAAT	TAAAAAAAGGTACT
M5	ATGGCTAG	AGAAAATACCAATAAG	CATTATTGGCTTAGAAAAT	TAAAAAAAGGCACT
M24	ATGACTAA	AAACAACACGAATAGA	CACTATTCGCTTAGAAAAT	TAAAAACGGGAACG
M49	ATGGCTAG	AAAAGATACGAATAAA	CAGTATTCGCTTAGAAAAT	AAAAACAGGTACA
M2	ATGGCTAG	AAAAGATACGAATAAA	CAGTATICGCITAGAAAAI	AAAAACAGGTACA
Hvdro	phobic re	aion		
M3	GCTTCAGT	AGCGGTTGCTTTGACA	GTTTTAGGGACAGGACTGG	FAGCA
M12	GCTTCAGT	AGCGGTTGCTTTAACA	GTCGTAGGAGCAGGGTTAG	FAGCA
M1	GCTTCAGT	AGCGGTAGCTTTGACT	GTTTTAGGGGCAGGTTTTG	CGAAT
M6	GCATCAGT	AGCAGTGGCTTTGAGT	GTAATAGGGGGCAGGATTAG	TTGTC
M5	GCATCAGT	AGCAGTAGCTTTGAGT	GTCTTAGGAGCAGGATTAG	TTGTC
M24	GCTTCAGT	AGCGGTAGCTTTGACA	GTTTTAGGGGCAGGATTAG	FTGTC
M49	GCATCCGT	AGCGGTCGCTGTGGCT	GTTTTAGGAGCAGGCTTTG	CAAAC
M2	GCATCCGT	AGCAGTCGCTGTGGCT	GTTTTAGGAGCAGGCTTTG	CAAAC
Cleava	age regio	n		
M3	GGGCAGAC	AGTAAAGGCA		
M12	GGGCAGAC	AGTAAGAGCA		
M1	CAAACAGA	GGTTAAGGCT		
M6	AATACTAA	TGAAGTTAGTGCA		
M5	AATACTA	TGAAGTTAGTGCA		
₩24	AATACTAA	TGAAGTTAGTGCA		
M49	CAAACAGA	AGTTAAGGCT		
M2	CAAACAAC	AGTTAAGGCG		
(B)	_			-
	в		н	C
M3	: ******	**YSLRKLKTGT	ASVAVALTVLGTGLVA	GQTVKA
M12	: MAKNTTN	RHYSLRKLKTGT	ASVAVALTVVGAGLVA	GQTVRA
M1	: MAKNNTN	RHYSLRKLKTGT	ASVAVALTVLGAGFAN	QIEVKA
M6	: MAKNNTN	RHYSLRKLKKGT	ASVAVALSVIGAGLVV	NTNEVSA
M5	: MARENTN	KHYWLRKLKKGT	ASVAVALSVLGAGLVV	NINEVSA
M24	: MTKNNTN	KHYSLKKLKIGT	ASVAVALIVLGAGLVV	NINEVSA
M49	MARKDTM	IKQYSLKKLKTGT	ASVAVAVAVLGAGFAN	QIEVKA
M2.	: MARKDIN	IKUYSLKKLKIGI	ASVAVAVAVLGAGFAN	Q I I VKA

Fig. 4 Comparison of C-repeat regions found among M3, M6, M5, M24 and M2 proteins. Identified amino acid residues with deduced M3 protein are indicated by colons.

C1-repeat regions

313	NET SEASBAGE BEDE DASBEAKKOVEKDE ANT TAFE DKVKE
nn J	
M6	NKVSEASRKGLRRDLDASREAKKQVEKDLANLTAELDKVKE
	: :: ::::::::::::::::::::::::::::::::::
M5	NKISDASRKGLRRDLDASREAKKQLEAEHQKLEEQ
	: :::::: ::::::::::::::::::::::::::::::
M24	NKISEASRQSLRRDLDASREAKKQLE
	:::::::::::::::::::::::::::::::::::::::
M2	Q I SEASRKSLRRDLEASRAAKKD
C2-re	epeat regions
MЗ	EKQISDASRQGLRRDLDASREAKKQVEK

- M6 EKQISDASRQGLRRDLDASREAKKQVEKALEE
- M5 NKISEASRKGLRRDLDASREAKKQLEAEQQKLEEQ
- M24 NKISEASRQSLRRDLDASREAKKQVEKALEE
- M2 QISEASRKSLRRDLEASRAAKKD

emm3 gene from the C203 strain by PCR. We selected a pair of forward and reverse primers, one (K-1; 31 mer) from the best conserved portion in leader sequences of seven different strains, and the other (K-2; 30 mer) from the C-terminal conserved portion^{5,7)~14}). PCR was performed with these two primers. The amplified product showed a single band approximately 1.4–1.5 kilobases (Kb) in length in agarose gel electrophoresis (Fig. 1-A) and was hybridized with the digoxigenin-labeled Msp I-Pvu II fragment of the emm6 gene (Fig. 1-B and C).

The DNA sequence of 1465 Kb of the amplified product was determined with a DNA sequencer and is shown in Fig. 2. The sequence indicated that the oligonucleotide sequences of the two primers existed in the 5' and 3' end (forward primer, 5' (CCGGGATCC) TATTCGCTTAGAAAATTAAAAA 3'; reverse primer, 5' (CCGGTCGA) CAAGTTCTTCAGCTTGTTTCGC 3'), and also encoded 488 amino acids which contained a leader sequence (1-32; defect first 9 amino acids), B repeat region (B1, 286-309; B2, 328-351) and C repeat region (C1, 374-391; C2, 416-433).

Comparison of the amplified product with other M protein genes: When the amplified product was compared with previously reported M protein genes^{7)~13)}, the leader sequence or C repeat regions showed

702

			B1 rep	repeat (100%)									
	261	271	281	291	301								
MЗ	AKEDFELAAL	GHQHAHNEYQ	AKLAEKDDQI	KQLE <u>EQKQIL</u>	DASRKGTARD								
	:: ::::::												
	265	275	285	295	305								
M12	AKKDFELAAL	GHQHAHNEYQ	AKLAEKDGQI	KQLEEQKQIL	DASRKGTARD								
		B2 repeat (91.7%)											
	311	321	331	341	351								
MЗ	LEAVRQAKKA	TEAELNNLKA	ELAKVTEQKQ	ILDASRKGTA	RDLEAVRQAK								
	315	325	335	345	355								
M12	LEAVRQAKKA	TEAELNNLKA	ELAKVTEQKQ	ILDASRKGTA	RDLEAVRKSK								
			C1 repea	t (94.5%)									
	361	371	381	391	401								
MЗ	AQVEAALKQL	EEQNRISEAS	RKGLRRDLDA	SREAKKQVEK	DLANLTAELD								
			::::::::	::::::::::									
	366	376	386	396	406								
M12	QQVEAALKOL	EEQNKISEAS	RKGLRRDLDT	SREAKKOVEK	DLANLTAELD								
		C2 rep	eat (100%)										
	411	421	431	441	451								
MЗ	KVKEEKQISD	ASRQGLRRDL	DASREAKKQV	EKALEEANSK	LAALEKLNKE								
	416	426	436	446	456								
M12	KVKEEKQISD	ASRQGLRRDL	DASREAKKQV	EKALEEANSK	LAALEKLNKD								
	461	471	481										
МЗ -	LEESKKLTEK	EKAELQAKLE	AEAKALKEQL	AKQAEEL									
	466	476	486										
M12	LEESKKLTEK	EKAELQAKLE	AEAKALKEQL	AKQAEEL									

Fig. 5 Comparison of homologous regions in M3 and M12 proteins, and relationship between B and C repeats in the predicted amino acid sequences of M3 and M12 proteins. The residues from each protein being compared are indicated by the numbering system of Robbins et al.¹²⁾ for M12 protein. B repeat blocks and C repeat blocks are indicated by underlining, and identified amino acid residues are indicated by colons. The numbers enclosed in parentheses indicate % homology.

high homology with the other emm genes (Fig. 3 and 4). While the N-terminal amino acid portion of the amplified product was variable, it was found to be identical to 96 of 98 nucleotides downstream of the leader peptide sequence of another emm3 gene¹⁵⁾. Therefore, the amplified product was characterized as the emm3 gene.

Comparison of homologous regions in M3 and M12 proteins: When the region of 252–488 deduced amino acids in M3 protein was compared with the region of 256–355 and 357–493 deduced amino acids in M12 protein, 96.6% homology was found between them (Fig. 5). Interestingly, the B repeat region showed high homology with only that of M12 protein (91.7 and 100%, respectively). However, the A repeat region in M12 protein was not present in the M3 protein. Predictive secondary structure analysis of M3 protein: From analysis of the predictive secondary structure of the amplified product by the algorithm of Robson¹⁶⁾, most of the product was found to exhibit strong alpha-helical potential. In addition, the beta-sheet and turn potential seen for region 23 to 42 in the M protein was similar to that seen for region 28 to 50 in M12 protein. The results suggest that M3 protein may be closely related to M12 protein.

Discussion

DNA sequence analysis has made it clear that all M proteins studied to date^{7)~12)} are structurally related and are therefore encoded by a family of genes. The regions of amino acid sequence homology in the protein include the signal sequences, the C repeat region in the central part of the protein chain and the carboxyl-terminal part. On the basis of available genetic information, we cloned the emm3 gene by using PCR and sequenced its DNA. The primers prepared originated from the best conserved leader sequence or the C-terminal portion of the emm genes^{7)~12)}. The amplified product was hybridized with an emm6 gene probe (Fig. 1). Sequence analysis of the amplified product identified sequences complementary to both oligonucleotide primers (Fig. 2). In addition, the product had both B and C repeat regions. By comparing the amplified product with known emm genes, we found not only that the N-terminal portion is very variable, but also that the C-terminal region is conserved^{7)~12),14)}.

Several streptococcal immunoglobulin-binding proteins have also been characterized as members of the M protein family and as having C repeats^{17,18}. Our selected primers have similar regions which can be

Chihiro KATSUKAWA

screened. However, our product amplified by PCR from the type 3 strain C203 of *S. pyogenes* was a single DNA and had no homology with amino terminal regions of the streptococcal immunoglobulin-binding proteins. Furthermore, 96 of 98 nucleotides downstream of the leader peptide sequence of the amplified product were found to be identical to the corresponding sequence of another emm3 gene of type 3 M strain 3-3/317¹⁵). The evidence shows that the amplified product is the emm3 gene.

B and C repeat blocks that exist in M3 protein are similar to those in M12 protein (Fig. 5). Furthermore, predictive secondary structure analysis of M3 protein revealed that the majority of the products exhibit strong alpha-helical potential as found with other M protein structures¹⁹. The algorithm also showed that region 23-42 exhibits beta-sheet and turn potential with a pattern similar to that for region 28-50 found by predictive secondary analysis of M12 protein.

S. pyogenes can be divided into two major classes on the basis of their immune reactivity with monoclonal antibodies (mAbs) directed against epitopes which lie within the conserved half of M proteins²⁰⁾. Class I serotype are defined as those which bind their mAbs, whereas class II isolates do not. Mainly the class I-specific mAb binding sites map to a region of C repeats within M proteins. Inasmuch as the C repeat region of our emm 3 gene represents more than 90% homology with the known emm genes, it belongs to the class I serotype. This agrees with the report of Bessen et al. who decided that M type 3 *S. pyogenes* had a class I protein²⁰⁾. Furthermore, we found similarity between emm 3 and emm 12 genes in their B repeat regions and predictive secondary structure. Thus, there may exist a subclass of class I M proteins. Bessen et al.²⁰⁾ discriminated between serotypes sharing both B and C repeat region epitopes and those sharing only C repeat region epitopes by using only antibody probes directed to antigenic sites within the B and C repeat regions of the M protein molecules in class I serotype. We suggest that M3 and M12 proteins belong to a subclass of class I M proteins.

Acknowledgements

I am grateful to Dr. Kyonsgu Hong of the Department of Bacteriology, Osaka University Medical School, Dr. Ikuya Yano of the Department of Bacteriology, Osaka City University Medical School, Dr. Masanao Makino and Dr. Yasuhiko Suzuki of Osaka Prefectural Institute of Public Health for technical advice and helpful discussions. In addition, I thank Dr. June R. Scott (Emory University) for providing the pUC19: M6 plasmid and Dr. Tatsuya Tanaka of the Central Laboratoey for Research and Education, Osaka University, Faculty of Medicine, for synthesizing the oligonucleotides. This work was supported by a grant-in-aid from the Ministry of Education, Science and Culture of Japan (C63570192).

References

- 1) Bisno, A.: Group A streptococcal infections and acute rheumatic fever. N. Engl. J. Med. 325: 783-793, 1991.
- Lancefield, R.C.: Current knowledge of the type-specific M antigens of group A streptococci. J. Immunol. 89: 307-313, 1962.
- 3) Beachey, E.H., Seyer, J.M., Dale, J.B., Simpson, W.A. & Kang, A.H.: Type-specific protective immunity evoked by synthetic peptide of *Streptococcus pyogenes* M protein. Nature (London) 292: 457–459, 1981.
- 4) Hong, K., Kinoshita, T., Takeda, J., Kozono, H., Pramoonjago, P., Kim, Y.U. & Inoue, K.: Inhibition of the alternative C3 convertase and classical C5 convertase of complement by group A streptococcal M protein. Infect. Immun. 58: 2535-2541, 1990.
- 5) Scott, J.R., Pulliam, W.M., Hollingshead, S.K. & Fischetti, V.A.: Relationship of M protein genes in group A streptococci. Proc. Natl. Acad. Sci. USA 82: 1822–1826, 1985.
- Murry, M.G. & Thompson, W.F.: Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Res. 8: 4321– 4325, 1988.
- 7) Bessen, D.E. & Fischetti, V.A.: Nucleotide sequence of two adjacent M or M-like protein genes of group A streptococci: Different RNA transcript levels and identification of a unique immunoglobulin A-binding protein. Infect. Immun. 60: 124-135, 1992.
- 8) Haanes, E.J. & Cleary, P.P.: Identification of a divergent M protein gene and an M protein-related gene family in

Streptococcus pyogenes serotype 49. J. Bacteriol. 171: 6397-6408, 1989.

- Hollingshead, S.K., Fischetti, V.A. & Scott, J.R.: Complete nucleotide sequence of type 6 M protein of the group A streptococcus: Repetitive structure and membrane anchor. J. Biol. Chem. 261: 1677–1686, 1986.
- 10) Miller, L., Gray, L., Beachey, E. & Kehoe, M.: Antigenic variation among group A streptococcal M proteins: Nucleotide sequence of the serotype 5 M protein gene and its relationship with genes encoding type 6 and 24 M proteins. J. Biol. Chem. 263: 5668-5673, 1988.
- Mouw, A.R., Beachey, E.H. & Burdett, V.: Molecular evolution of streptococcal M protein: Cloning and nucleotide sequence of type 24 M protein gene and relation to other genes of *Streptococcus pyogenes*. J. Bacteriol. 170: 676–684, 1988.
- 12) Robbins, J.C., Spanier, J.G., Jones, S.J., Simpson, W.J. & Cleary, P.P.: *Streptococcus pyogenes* type 12 M protein gene regulation by upstream sequences, J. Bacteriol. 169: 5633-5640, 1987.
- 13) Haanes-Fritz, E., Kraus, W., Burdett, V., Dale, J.B. & Beachey, E.H.: Comparison of the leader sequences of four group A streptococcal M protein genes. Nucleic Acids Res. 16: 4667-4677, 1988.
- 14) Hollingshead, S.K., Fischetti, V.A. & Scott, J.R.: A highly conserved region present in transcripts encoding heterologous M proteins of group A streptococci. Infect. Immun. 55: 3237-3239, 1987.
- 15) Podbielski, A., Baird, R. & Kaufhold, A.: The group A streptococcal M-type 3 protein gene exhibits a C terminus typical for class I M proteins. Med. Microbiol. Immunol. 181: 209–213, 1992.
- 16) Garnier, J., Osguthorpe, D.J. & Robson, B.: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97–120, 1978.
- 17) Gomi, H., Hozumi, T., Hattori, S., Tagawa, C., Kishimoto, F. & Björck, L.: The gene sequence and some properties of protein H. J. Immunol. 144: 4046-4052, 1990.
- 18) Heath, D.G. & Cleary, P.P.: Fc-receptor and M-protein genes of group A streptococci are products of gene duplication. Proc. Natl. Acad. Sci. USA 86: 4741-4745, 1989.
- 19) Phillips, G.N., Flicker, P.F., Cohen, C., Manjula, B.N. & Fischetti, V.A.: Streptococcal M protein: α-helical coiled-coil structure and arrangement on the cell surface. Proc. Natl. Acad. Sci. USA 78: 4689–4693, 1981.
- 20) Bessen, D., Jones, K.F. & Fischetti, V.A.: Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. J. Exp. Med. 169: 269-283, 1989.

PCR 法を用いてクローニングした A 群溶血レンサ球菌

M3蛋白遺伝子の解析および他菌型との比較

(レンサ球菌感染症研究会:会長 保科 清)大阪府立公衆衛生研究所

勝川千尋

(平成5年11月12日受付)(平成6年2月7日受理)

要 旨

A 群溶血レンサ球菌 M3蛋白遺伝子の N 末端 の高度型特異領域から C 未端の保存領域までの 部分を PCR 法を用いてクローニングを行った.

各菌型に共通なN末端のリーダーシクエンス 部分とC末端の保存領域部分をプライマーとし て用い,1465bpの遺伝子配列を決定し,他のM蛋 白遺伝子と比較検討した。その結果,N末端側の 100塩基から750塩基の範囲にM3型特異的な領域 を見いだすことができた。また、アミノ酸配列を 検討したところ, 2つの繰り返し配列を見いだし た(BリピートおよびCリピート).Cリピートは 現在知られている他の M 蛋白遺伝子の塩基配列 と非常に高い相同性を示した.これに対して,Bリ ピートは M12蛋白遺伝子の Bリピート配列との み高い相同性を示し,また二次構造の解析結果で もこの2菌型は構造が類似していた.これらの結 果より M3蛋白と M12蛋白は遺伝学的に非常に近 い蛋白であることが示唆された.

平成6年5月20日