Table 1. Formation (affinity) constants.

Ion	Bi		
	$\log K_{1\mathrm{B}}$	$\log K_{2B}$	1ES, $\log K_{1B}$
H ⁺	8.39 ± 0.01		7.60 ± 0.01
Cu(II)	8.07 ± 0.01	$5.40 \pm 0.02^{*}$	$3.90 \pm 0.02^{\dagger}$
Co(II)	5.30 ± 0.01	3.38 ± 0.02	2.07 ± 0.02
Zn(II)	5.37 ± 0.01	2.67 ± 0.03	2.08 ± 0.02
Mn(II)	3.02 ± 0.01		‡
Ca(II)	2.66 ± 0.01		~ 0

*See Eqs. 10 and 11. complex formation. †See Eqs. 12 and 13. ‡Hydrolysis and precipitation are concomitant with any

The values for K_{1a} and K_{OH} are $10^{-12.36 \pm 0.02}$ and $10^{-10.42 \pm 0.02}$, respectively. On the other hand, TES forms much weaker metal complexes with metal ions (no complex formation with Ca^{2+}) than bicine. However, the 1:1 Cu(II)-TES chelate undergoes hydrolysis (Eq. 12) and subsequent dimerization (K_D) (Eq. 13) in the neutral pH range (5)

 $Cu(TES)^+ \leftarrow Cu(TES)OH + H^+(12)$ 2Cu(TES)OH $\leftarrow \frac{K_{\rm R}}{\leftarrow}$ Cu₂(TES)₂(OH)₂ (13)

The values for K_{OH} and K_D are $10^{-6.62 \pm 0.02}$ and $10^{+2.30 \pm 0.04}$, respectively. It is obvious from the above reactions that not only do "Good's" buffers complex metal ions but also that the metal complexes formed have a rich solution chemistry. These latter reactions further buffer H^+ concentration (Eqs. 10) through 12) and strengthen the metal complexes formed (Eq. 13).

Although the constants listed in Table 1 are on the whole small compared to those of metal ions for active sites of apoenzymes $(10^{10} \text{ to } 10^{20})$, they are of sufficient strength to lead workers to erroneous conclusions in systems of low metal-binding capabilities, for example, fluorescence quenchers, spin labels, and paramagnetic shift reagents. Furthermore, the "Good's" buffer ADA forms a very stable chelate with Cu(II) $(K_1 = 10^{10.57})^2$, rivaling that of apoenzymes, and Ca(II) $(K_1 = 10^{4.01})^2$. The latter constant is high enough to cause the extraction of labile Ca(II) from enzymes or other biochemicals, thereby altering the structure of the protein with concomitant loss of activity.

The users of "Good's" buffers should no longer assume, as is generally believed, that they do not complex metal ions below or at physiological pH values. Studies on the following "Good's" buffers, ACES {2-[(2-amino-2-oxoethyl)amino]ethanesulfonic acid}, ADA, BES, bicine, and TES indicate that all complex metal ions. The general tendency to bind M(II) is in the order ADA > bicine > $ACES > BES \sim TES$. The last three of these buffers, however, do not bind Ca(II). Those doing studies in the presence of these buffers and metal ions should take into account the metal complexes formed and their possible deleterious effects on the interpretation of data collected in their presence. Routinely used buffers other than those developed by Good *et al.* (1) have also been shown to complex metal ions, most notable tris, [tris(hydroxymethyl)methylamine] (6).

ROBERT NAKON

C. R. KRISHNAMOORTHY Department of Chemistry, West Virginia University, Morgantown 26506

References and Notes

- N. E. Good et al., Biochemistry 5, 467 (1966).
 R. Nakon, Anal. Biochem. 95, 527 (1979); J. M. Pope, P. R. Stevens, M. T. Angotti, R. Nakon, *ibid.* 103, 214 (1980).
- 3. Potentiometric formation curves were obtained at 25.00° \pm 0.05°C and 0.1*M* (KNO₃) ionic strength on a Corning Digital 130 research mod-el *p*H meter. Metal ion concentrations were ~ 0.025*M*.
- 4. Visible spectra (not shown) of 2:1 bicine-Cu(II) solutions from a = 0 to a = 2.0 (a is the number of moles of base per mole of complexing agent) evince a monotonic shift in maximum wave-breacher (i) and the solution of the sol length (λ_{max}) and extinction coefficient (ϵ_{max}) from **a** = 0.0 to 1.0, indicating the formation of [Cu(bicine)₂]. The λ_{max} value at **a** = 1.0 is 610 nm, in good accord with the value for a variety of other big(amisocaidta) correct(U) abelates Init, in good accord with the value for a value of other bis(aminoacidate)copper(II) chelates ($\lambda_{max} \approx 605 \text{ nm}$ [R. Nakon, E. Beadle, Jr., R. J. Angelici, J. Am. Chem. Soc. **96**, 719 (1974)]. From **a** = 1.0 to 2.0, λ_{max} shifts monotonically to lower energy ($\lambda_{max} = 724 \text{ nm}$ at **a** = 2.0); this From $\mathbf{a} = 1.0$ to 2.0, λ_{max} shifts monotonically to lower energy ($\lambda_{max} = 724$ nm at $\mathbf{a} = 2.0$); this behavior indicates the loss of an aminoacidate ligand (Eq. 10). The λ_{max} and ϵ_{max} values of 1:1 and 2:1 bicine–Cu(II) solutions at $\mathbf{a} = 3.0$ and 2.0, respectively, are 724 nm and 64 M^{-1} cm⁻¹, an indication that both solutions contain the same species, [Cu(H₋₁ bicine)OH⁻]. The elec-tron spin resonance spectra of 1:1 and 2:1 tron spin resonance spectra of 1:1 and 2: bicine—Cu(II) solutions at a = 3.0 and 2.0, re bicine—Cu(II) solutions at $\mathbf{a} = 3.0$ and 2.0, respectively, are identical; similarly, at $\mathbf{a} = 1.0$ and 0.5, respectively, identical spectra were observed. However, the 2:1 spectrum at $\mathbf{a} = 1.0$ was never seen in any 1:1 spectrum from $\mathbf{a} = 0.0$ to 3.0. The species at $\mathbf{a} = 3.0$ (2.0) is [Cu(H₋₁ bicine)OH⁻], at $\mathbf{a} = 1.0$ (0.5) is [Cu(bicine)⁺], but at $\mathbf{a} = 1.0$ (2:1 system) is [Cu(bicine)⁺],
- Visible spectral data of 1:1 TES-Cu(II) solutions show no large blue shift until pH > 11, an indication that there are no alcohol deprotonation reactions in the acidic or neutral pH regions
- J. L. Hall, J. A. Swisher, D. G. Brannon, T. M. Linden, *Inorg. Chem.* 1, 409 (1962); K. S. Bai and A. E. Martell, *J. Inorg. Nucl. Chem.* 31, 1697 (1969).

2 March 1983; revised 9 May 1983

Nucleotide Sequence of a Light Chain Gene of the Mouse I-A Subregion: Aβ^d

Abstract. Ia (I region-associated) antigens are cell-surface glycoproteins involved in the regulation of immune responsiveness. They are composed of one heavy (α) and one light (β) polypeptide chain. We have sequenced the gene encoding the $A\beta^d$ chain of the BALB/c mouse. The presence of six exons is predicted by comparison with the complementary DNA sequences of human β chains and with partial protein sequence data for the $A\beta^d$ polypeptide. Sequence comparisons have been made to other proteins involved in immune responses and the consequent implications for the evolutionary relationships of these genes are discussed.

The major histocompatibility complex (MHC) of the mouse is a cluster of genes encoding at least three different classes of proteins involved in immune responses (1). Class I molecules, the classic transplantation antigens, and class II molecules are cell-surface, membranebound glycoproteins. Class III molecules are serum protein components of the complement pathway. While class I proteins are found on essentially all cells, class II representatives appear limited mainly to the surface of B cells and to antigen-presenting cells such as macrophages. Class II proteins are required for antigen presentation and lymphocyte interactions involved in the activation and differentiation of antibody-producing

cells or B cells (2). Murine class II genes are located in the I (immune response) region of the MHC and are consequently referred to as I region-associated or Ia antigens (3).

Two types of Ia antigen have been defined in mice, I-A and I-E. Both are composed of two noncovalently linked polypeptides, a heavy chain (α) of about 34,000 daltons and a light chain (β) of about 28,000 daltons (molecular size). In the past 2 years, advances in protein microsequencing and recombinant DNA technologies have led to a definition of the primary structure of various human and murine class II molecules (4-14). Both α and β polypeptides may be divided into two external domains of approximately 90 residues each (termed α_1 and α_2 for the α chain or β_1 and β_2 for the β chain), a transmembrane region about 30 residues long, and a cytoplasmic tail of about 15 residues.

We have determined the complete nucleotide sequence of the class II $A\beta^d$ gene. This structure, together with class I and class II genes that have been analyzed, permit us to discuss several interesting evolutionary features about the supergene family including the MHC and antibody genes.

An A β gene was isolated from a cosmid library constructed from BALB/c (d haplotype) sperm DNA by screening with a complementary DNA (cDNA) probe of a human I-A region homolog, DC- β (15). We have sequenced 5.8 kilobases (kb) of cosmid clone 34.2 overlapping the region of DC- β hybridization. The sequencing strategy and gene organization are shown in Fig. 1 and the DNA sequence in Fig. 2. The predicted amino-terminal protein sequence is identical to that previously defined for the BALB/c A β^d protein (4, 5).

The coding sequence of $A\beta^d$ is distributed among six exons. All exons have the appropriate donor and acceptor RNA splice sites. The first exon encodes a 27residue-long putative signal peptide with 25 consecutive hydrophobic or neutral amino acids. It is identifiable because it begins with a methionine and the first four amino acids of the mature protein. The length of the $A\beta^d$ messenger RNA, 1.4 kb, and the lack of hybridization of probe 1 (Fig. 1) with BALB/c

 $poly(A)^+$ (polyadenylated) messenger RNA (mRNA) (see below) limits the extent of the 5' untranslated region to less than 150 bp upstream of the initiation codon. Consistent with this, the canonical promoter-associated sequence CCAAT (C, cytosine; A, adenine; T, thymine) is found about 135 bp 5' to the methionine codon. However, no recognizable candidate for the Hogness-Goldberg TATA box can be found within the 200-bp stretch 5' of this codon. The $E\alpha^k$ gene also lacks a canonical TATA box (13), although it does have an AT-rich area missing from $A\beta^d$. It remains to be seen whether the partial or complete absence of such sequences will be a common feature of murine class II genes and, if so, whether this will have any significance for their regulation.

The other five exons were defined on the basis of similarity with two human cDNA sequences: DC- β (11) and DR- β (12). The two external domains are encoded by exons 2 and 3. Each contains a pair of cysteines, presumably involved in intradomain disulfide bridges. The predicted β_1 domain amino acid sequence contains one potential glycosylation site (Asn-Glu-Thr) beginning at residue 19. This site is also conserved in both human sequences. The fourth exon encodes three different structural elements: a hydrophilic-connecting peptide (amino acids 190 to 199), a hydrophobic or uncharged transmembrane segment (200 to 220), and a hydrophilic cytoplasmic tail (221 to 226). The transmembrane element is apparently anchored in the cell membrane by three basic residues (Arg-His-Arg) on the cytoplasmic side. Exon 6 encodes the last four amino acids of the cytoplasmic tail. A poly(A) addition signal (AATAAA) is found 280 nucleotides 3' to the termination codon. The predicted 3' untranslated sequence is found by computer analysis to be weakly, but recognizably similar to that of the DC- β cDNA.

The cytoplasmic region of the DR-B cDNA is eight codons longer than that of the DC- β cDNA. Alignment of all three β sequences suggests that the size difference is internal and not simply a 3'truncation of the DC- β coding sequence relative to that of the DR-B cDNA. This relative insertion is located between exon 4 and exon 6 and led us to suspect there may be an additional cytoplasmic exon in the intervening sequence. A search of this intron revealed only one 24-base sequence bounded by the appropriate RNA donor and acceptor splice signals with an open reading frame in phase with the adjacent exons that shares similarity with the extra sequence of the DR-B clone. A restriction fragment including this sequence (probe 3, Fig. 1) was hybridized to BALB/c spleen $poly(A)^+$ mRNA on a nitrocellulose filter (Northern blot) (16) yielding a band at 1.4 kb, which is the same size as the $A\beta^{d}$ mRNA. The same mRNA did hybridize to another coding sequence (probe 2) but did not hybridize to a noncoding region (probe 1). We therefore include exon 5 as part of the expressed mouse gene sequence and assume that the DC- β gene



Fig. 1. The organization and sequencing strategy for the $A\beta^d$ gene. (A) Partial restriction map of the cosmid clone 34.2. Filled areas indicate exons. The arrow above shows the direction of transcription. Restriction enzyme names have been abbreviated as *R* (Eco RI) and *B* (Bam HI). (B) Structure and sequence strategy for the $A\beta^d$ gene. Arrows beneath indicate the sequencing strategy. Straight arrows give the sequence obtained by the dideoxy chain termination method (20). The two Hind III fragments were subcloned into the Sma I site of M13mp8 (21) in both orientations. Each subclone was used to generate a series of overlapping subclones by a deoxyribonuclease I deletion technique (22). Wavy arrows indicate the direction and length of sequences obtained by the chemical degradation method (23). *SP* refers to the signal peptide, β_1 and β_2 to the two external domains, *TM* the transmembrane portion of the protein, *CY* the cytoplasmic part, and *3'UT* the 3' untranslated region of the mRNA. The region between the two dotted lines shown in the gene map has not been sequenced. Abbreviations for restriction enzymes not listed in (A) are *H* (Hind III), *S* (Sac I), and *C* (Sac II). (C) Different probes used in the Northern blot analysis (16).

1	GTTTGGAAAGAGGCTCAGC	GACTCCCCAA	GAACTCTGCTI	TCTCACTGAG	GCACCCCAG	AGGGTCAGATO		TGATTTCTC	GTCTCTTAA	TACAACGTCACACTCATTTC	T 120
121	CCCACTACAACCTCAAGTG	TAAGAACAGT	AGATTTGAAAT	AATTATCCCA	AAGTTGTCT	ATACTTCACG	ACTTTTCTCT	TTAAACCCCC		I CAGTCTGCAGTTTGGAAAAG	T 240
241	TTACTCTGTAACTGAGAGG	AGCATTCACC	TCAAACGTTC	TCCAGTGGGG	CTCATGAAG.	AAACAGACCAT	GCCTTGCATA	GAGAGCCTTT	GTAAACAAA	, STTTACCCAGAGACAGACGA	C 360
361	AAGCTTCAGCTCCAATGCT	GATTGGTTCC	TCACTTGGGA	CAACCCTGAC	CACTCTGGGA	I TTTCAGATCAC	TCTAGGCTAC		TTTCTGAAG	I GGGGCACAGCAGGTGTGAGT	C 480
	I	I	1	. I		1 1					
-27 481	CTGGTGACTGGCATTACCT	GTGCCTTAGA	M A L GATGGCTCTGC	AGATCCCCA	S L L L SCCTCCTCCT	S A A CTCAGCTGCTG	уууу Каралан Улавтартаст	GATGGTGCT	S S P Agcagccca	R T E G∳VG N S Aggactgaggggggaaactc	3 C 600
	1	1	1 1	· I						1 1	
4 601	E GAAA <u>GT</u> AAGTGCCGGGGGC	GGGCTCTCCA	GAGCCAGCAC1	CTGGGGGTCG	TGTTCTGAG	GGACCCTGGGG	GCCCTGGGAT	GTGGGAGGG	GTGGCAAAA	TCCCCATTTTACTATAGATT	4 T 720
721	CCTTTCTCTGAAAAGAGCG	I TGAGACTCCA	I TTCTCACTTCI	GTGTTCTAAT	 ATTGAGTTA	I I GGATGTGCCGC	I CGCACGGGCC	AGCCTGACA		 	A 840
841	I GTTTAAATTGACCACCAAA	GATAAGGGTT	 G	· · · · · · · · · · · · · · · · · · ·	l	I I			GGAGAAAGC	I I Agggcttacaataccagtgc	000
901		AGGCTTACAA	I I TACCAGTGCCC	I TCACCCCACC	I CGTACATTG	I I TCTTGAGGAA#	I I ATGCCAGATTO	TGAAGTGGG		I I AGCGATGCTGGCCTCCCGAT	3 1020
1021		GTGGAGTGGA		GCGAACTCT	GGCGTCTAA			AATAAACCT	ATTTTTTG	I I STCAAATGGAATCCAGGACT	C 1140
1141		I	AGAAAACACT	GGCCCAGAAG	GGACATATG	I I TGATACATACC	CTGGGGTGAG	CGGAGTGTC		I CGGAGACCAGAGAGTAAATA	A 1260
1261			GGTGTCTTCAC	TCTCGCAGT	ACCCTTAAG	I I ATGCCTGGAGT		GCAGGCAGT	CTCAGAGGT	I IGCCTGGATTCCCATTTGCT	3 1380
1381			AGTGCCCTCTC		GCCTTTTGC		SAGTCCCACAG	GGTGGGGGT	GGGTAGCAG		T 1500
1501			GCTGAGAATTI		TGATCATAT			CCCAGGCCCA	тессессес		T 1620
1621	GICCCTTAGITITATIATI				TAGCGTTTC			ACGAGCTCG	TCGCCCCAT		1740
1741					6666666666			000000000000000000000000000000000000000	CGGTTCCCA		3 1860
	I	I	1	I I		1	I I			1 1	
5 1864	GAAGCGCCGGCGCGGGGTG	6660060666	GAGCCGTGGT(тесетеетсс	ссостотос	CCCGGGCTGAC	сесетссето		F V V TTCGTGGTC		18
			1	I							
17 1981						R V Y					56 2100
1001		1	1	Ĩ							- 2100
57						J A T		E G P		T S L R R L	94 7 2220
2224								TCCTTCCTC			1 2340
2244								GTGACCTCAG	COTOTOTOTO		2460
2341								TTTOTOCOCC			5 2400
2401											2360
2704								AATTAAACCO	CTOCCCTAC		2700
2024								TCCTCACAC			2040
2021								TOTTALACTAL			A 2940
2941									GIAIGUIGI		a 3000
3061								CACAGICIG	TTACATTA		
3181											A 3300
3301		AAATACACAG									3420
3421								CACALICIIC	STITICCIA6/		3 3540
96										E	96
3541			I					CICALATII	ACICALIGI	I I I	3638
97		S L S R				V C S	V T D F	Y P A	K I K		136
3008							I GALAGAIII	LIALLLAGU	AAGAILAAAU	SIGLGLIGGIILAGGAAIGG	. 3//8
137	QEETVG									1 1	
3//9	CACCACACACACTOCCC	V S S T						M T P	H Q G		176
	CAGGAGGAGACAGTGGGGG	V S S T TCTCATCCAC	Q L I ACAGCTTATTA	R N G C NGGAATGGGGA) W T F ACTGGACCTT(I	Q V L CCAGGTCCTGG I I	V M L E STCATGCTGGA	M T P GATGACCCC1	H Q G CATCAGGGAG	I I I E V Y T C H V GAGGTCTACACCTGCCATGT I I	176 3 3898
177	CAGGAGGAGACAGTGGGGG	V S S T STCTCATCCAC	Q L I ACAGCTTATTA	R N G L) W T F Actggacctt(I	CCAGGTCCTGG	V M L E Stcatgctgga	M T P	H Q G CATCAGGGAG	E V Y T C H V GAGGTCTACACCTGCCATGT	176 3898 188
177 3899	CAGGAGGAGACAGTGGGGG I E H P S L K GAGCATCCCAGCCTGAAGA	V S S T STCTCATCCAC I S P I T GCCCCATCAC	Q L I ACAGCTTATTA V E W TGTGGAGTGGA	R N G L GGAATGGGGA I I <u>GT</u> AAGGGAAT	O W T F ACTGGACCTTO I ITTCTTTTTC	Q V L CCAGGTCCTGG I ITTCAACTGTG	V M L E STCATGCTGGA I GGGCCCCACAT	M T P GATGACCCC1 GACATGGGA1	H Q G CATCAGGGAC	E V Y T C H V SAGGTCTACACCTGCCATGT	176 3898 188 4018
177 3899 4019	CÂGGĂGGĂGAĊAGTGGĜGG I E H P S L K GAGCATCCCAGCCTGAAGA I TCACCCACCCATCACTTG I	V S S T STCTCATCCAC I GCCCCATCAC I STCCTATATGA	Q L I ACAGCTTATTA V E W TGTGGAGTGGA I TCTGGTTCCAC I	R N G [AGGAATGGGGA A <u>GT</u> AAGGGAAT TACTGAGCTG) W T F ACTGGACCTT(I ITTCTTTTTC BAGACCTACA	Q V L CCAGGTCCTGG I I TTTCAACTGTG GGAAATCATAT I I	V M L E STCATGCTGGA	GACATGGGAT GACATGGGAT	H Q G CATCAGGGAG	I I C H V SAGGTCTACACCTGCCATGT I ITATTATCCCATCCCTCCAAT I G AGCCCTGACCTATCTTCTCA I I	176 3898 188 4018 6 4138
177 3899 4019 190	CÁGGÃGGÃGACAGTGGĜGG E H P S L K GAGCATCCCAGCCTGAAGA I TCACCCACCCCATCACTTG	V S S T STCTCATCCAC I S P I T GCCCCATCAC I TCCTATATGA I	Q L I ACAGCTTATTA I V E W I GTGGGAGTGGA I TCTGGTTCCAG	R N G [AGGAATGGGGA) W T F ACTGGACCTT(I ITTCTTTTTC BAGACCTACA	Q V L CCAGGTCCTGG I I TTTCCAACTGTG GGAAATCATAT I I	V M L E STCATGCTGGA	M T P GATGACCCCT GACATGGGA1 GACAATCAGGG	H Q G CATCAGGGAG TTTTAGGTGTT CATCAGGAA	E V Y T C H V GAGGTCTACACCTGCCATGT I IATTATCCCATCCCTCCAAT AGCCCTGACCTATCTTCTCA I B A Q S E	176 3898 188 4018 4138
177 3899 4019 190 4139	CÁGGÃGGÃGAGACAGTGGĜGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCATCACTTG I ACAGCAGTTCTGGAGATCA	V S S T STCTCATCCAC S P I T GCCCCCATCAC I ICCTATATGA CTACATACAC	Q L I ACAGCTTATT, Y E W TGTGGAGTGG, TCTGGTTCCAC	R N G L AGGAATGGGGA I I TACTGAGCTG I I AGGTTTTTCCC	D W T F ACTGGACCTT TTTCTTTTTC GAGACCTACA CTAATATCCA	Q V L CCAGGTCCTGG I I TTTCCAACTGTG GGAAATCATAT I GAGGAATTGGC I I	V M L E STCATGCTGGA GGGCCCCACAT I CTCTCTCACCTC	GACATGGGA Gacatggga Gacatcagg Gacatcagg Gacatcagg	H Q G CATCAGGGA(TTTTAGGTGTT CATCAGGAA)	E V Y T C H V SAGGTCTACACCTGCCATGT I I AGCCCTGACCTATCTTCTCA I I AGCCCTGACCTATCTTCTCA I I TCCCCC <u>AG</u> GGGCCCGTCCGA I	176 3898 188 4018 4138 4138 4138 4138
177 3899 4019 190 4139 195	CÁGGAGGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA I TCACCCACCCCATCACTTG I ACAGCAGTTCTGGAGATCA I S A R S K M	V S S T STCTCATCCAC S P I T GGCCCCATCAC I TCCTATATGA I CTACATACAC I L S G I	Q L I ACAGCTTATT/ I TGTGGAGTGG/ I TGTGGAGTGG/ I TGGGGGCCCTG/ I G G C C	R N G L AGGAATGGGGA) W T F CTGGACCTTI I GAGACCTACAI CTAATATCCAI	Q V L CCAGGICCTGE I ITTTCAACTGIE I GAGGAATCATAT GAGGAATTGGC I GAGGAATTGGC	V M L E STCATGCTGGA	GATGACCCCI GACATGGGAI ACAATCAGGG TGTAGACACI	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAA/		176 3898 188 4018 4018 4138 194 4258 2255
177 3899 4019 190 4139 195 4259	CAGGAGGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGCAAGATGT	V S S T STCTCATCCAC S P I T GCCCCATCAC STCCTATATGA CTACATACAC L S G I TGAGCGGCAT	C L I ACAGCITATI	R N G L AGGAATGGGGJ V STACTGAGGTAA V AGGTTTTTCCCC V V GGTTTGGGGT) W T F CTGGACCTTI GAGACCTACA CTAATATCCA I I GATCTTCCT	Q V L CCAGGICCTGE I I GGAAATCATAT GAGGAATTGGC I GAGGGATTGGC CGGGCTCGGCC	V M L E	GATGACCCCT GACATGGGAT ACAATCAGGG TGTAGACACT H R S TCACAGGAG	H 0 G CATCAGGGA(TTTAGGTGTT CAGGAAGG <u>GT</u> CAGGAAAG <u>GT</u>	E V Y T C H V SAGGTCTACACCTGCCATGT I AGCCCTGACATCCCTCCAAT I AGCCCTGACCTATCTTCTCA I TCCCCAGGGGCACAGTCCGA I GAGGAGCTCTGGAGAATTGG	176 3898 188 4018 4018 4138 194 34258 225 4378
177 3899 4019 190 4139 195 4259 4379	CAGGAGGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCATCACTTG ACAGCAGTTCTGGAGATCA I S A R S K M TCTGCCCGGAGCAAGATGT GGTGGGTGGGCCGTGGCCACA	V S S T STCTCATCCAC S P I T GCCCCATCAC I CTACATACAC L S G I TGAACGGCAT I GGAAGGAGCCC	CTGGGGGCCCTGA	R N G L AGGAATGGGGA TACTGAGCTG TACTGAGCTG AGGTTTTTCCC C L G V TGCTTGGGGAAA GGGAGGAAA	D W T F CTGGACCTTI GAGACCTACA CTAATATCCA J F L GGACGTCCC	Q V L CCAGGTCCTGG I I GGAAATCATATI GGAAATCATATI G G G G C G GGGCTCGGCC I AGGAGGAGGACAT	SGGCCCCACAT GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC L F I F CTTTTCATCCG	GATGACCCCI GACATGGGAI GACAATGGGAI CACAATCAGGG CTGTAGACACI H R S ITCACAGGAGG ITTTTGCTGG	H 0 G CATCAGGGA(TTTTAGGTGTT CATCAGGAAA TAGCTCTATT 0 K CAGGAAAG <u>GT(</u> TATGTTACT	E V Y T C H V SAGGTCTACACCTGCCATGT I I I I I I I I I I I I I I I I I I	176 3898 4018 4018 4138 194 4258 225 4378 4498
177 3899 4019 190 4139 195 4259 4379 4499	CAGGAGGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGCAAGATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTCTGTG	V S S T STCTCATCCAC S P I T GCCCCATCAC I I GCTACATACAC L S G I TGAGCCGGCAT I GGAAGGAGCC C CCTGTCCACC I CCTGTCCACC	C L I ACAGCTTATT/ TGTGGAGTGG/ TCTGGTTCCAC CGGGGGCCCCTG/ CGGGGGCTGCAG GGGCTGCAGGG ACCCTCACTGI	R N G L AGGAATGGGGA TACTGAGCTA TACTGAGCTA GGGTTTTTCCC C GGGAGGAAAA CCTCTTTCCCA	D W T F CTGGACCTTO GAGACCTACA CTAATATCCA GATCTTCCT GAAGGTCCC AGAAGCTTCC	Q V L CCAGGTCCTGE I I GGAAATCATAT GGAAATCATAT I GGAAATCATAT GGGGCTCGGCC I AGGAGGGGCAT I GGCAATGAAAA	SGGCCCCACAT GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC L F I F TTTTCCATCCG TTGGGATCTGA	GATGACCCCI GACATGGGAI GACAATGGGAI CACAATCAGGG CTTCACAGGAG CTTCCCCGGC CTTCCCTGGTG	H 0 G CATCAGGGA(TTTTAGGTGTT CAGCATCAGGAA/ TAGCTCCTAT CAGGATAGG <u>GT</u> CAGTTCATC	E V Y T C H V AGGTCTACACCTGCCATGT I I I I I I I I I I I I I	176 3898 188 4018 4018 4138 34138 194 4258 225 34378 24498 4498
177 3899 4019 190 4139 195 4259 4379 4499 4619	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA I TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA I ACAGCAGTTCTGGAGATCA I S A R S K M TCTGCCCGGAGCAAGATGT GGTGGTGGGCCTGTGCCACA I TCATTCTGTGACTTCTGTC TAGGGTCTTAGTGAATACA	V S S T STCTCATCCAC S P I T IGCCCCATCAC I ICTACATACAC L S G I TGAGCCGCAT I IGGAAGGAGGACC I ICCTGTCCACC I GTTGGTGAGAA I	ACAGCTTATT	R N G L AGGAATGGGGA AGTAAGGGAA1 TTACTGAGCTC TACTGAGCTC GGGAGGAAA1 CTCTTTTCCC CCCTTTTCCC GCTAAGCTGG	D W T F CTGGACCTTC GAGACCTACA TGATATATCCA TGATCTTCCT GAGAGGTCCC CAGAAGCTTCC	0 V L CCAGGTCCTGC 1 TTTCAACTGTG 1 GGAAATCGTAT 1 GGAAATCGTAT 1 1 1 1 1 1 1 1 1 1 1 1 1	V M L E STCATGCTGGA GGCCCCCACAT CTCTCACCTC CTGAAGTAGAC CTGGAAGTAGAC TGGGATCTG TGGGATCTG TGGGATCTGGAC	M T P GATGACCCCI GACATGGGAI GACAATGGGAA TTGTAGACACCI H R S TCACAGGAG TTTTGCTGG TTCCCTGTGT(GAGGGCCCCT)	H 0 G CATCAGGGA(TTTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAG <u>GT</u> TATGTTACT(CAGTTCATC(TTTTTTTATA)	E V Y T C H V AGGETCTACACCTGCCATGT TATTATCCCATCCCTCCCAAT AGCCCTGACCTACTTCTCCA TCCCCCAGGGGGGCACCAGTCGGA GAGGAGCTCTGGAGAATTGG GCCACAGAGCCATGGTGGAG ACTTGGGGGGGGGCACCAC ACAACACCTTTCTCTATCTG I	176 3898 188 4018 4018 4138 194 4258 225 4378 4498 4498 44738
177 3899 4019 190 4139 195 4259 4379 4499 4619 227	CAGGAGGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA I TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA I ACAGCAGTTCTGGAGATCA I S A R S K M TCTGCCCGGAGCAAGATGT GGTGGTGGGCTGTGCCACA I TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA I G P	V S S T STCTCATCCAC S P I T IGCCCCATCAC I ITCCTATATGA L S G I TGAGCGGCAT I GGAAGGAGGAGCC I CCTGTCCACC GTTGGTGAGA H R G P P	C L I ACAGCITATI) TGTGGAGTGG/ TCTGGTTCCA(CGGGGGCCCTG/ G G C CGGGGGGCTGCA GGGCTGCAGGT ACCCTCACTGT GAACAATTCC/ J	R N G L AGGAATGGGGA AGTAAGGGAA1 TTACTGAGCTC AGGTTTTTCCCC STGCTTGGGGG CTCCTTTTCCC CCCTATTTCCC I	D W T F CTGGACCTTC GAGACCTACA CTAATATCCA GATCTTCCT GAAGGTCCC GGAAGGTCCC CAGGACTTCC	0 V L CCAGGTCCTGE 1 TTTCAACTGTE 1 GGAAATCATAT 1 GAGGGAATTGGC 1 GGGCAATGAAAA 1 GGCCATGAAAA 1 GGCCATGAAAA	SOCOLOGICACA SOCOCCCCACAA CTCTCACCCTC CTGAAGTAGAA CTGGAAGTAGAA TGGGATCTGA ATGCATAGGCC STCGTTGTGAA	M T P GATGACCCCT GACATGGGAT GACATGGGAGG TGTAGACACC H R S TCACAGGAG TTTTGCTGG TTCCCTGTGT GAGGGCCCCT	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAA/ TAGCTCTAT CAGAAAG <u>GT</u> TATGTTACT(AGTTTCATC(TTTTTTTATA/	E V Y T C H V AGGETCTACACCTGCCATGT I TATTATCCCATCCCTCCCAAT AGCCCTGACCTATCTTCTCA R A Q S E TCCCCCAGGGGGGCACCAG GAGGAGCTCTGGAGAATTGG GCCACAGAGCCATGGTGGAG ACCATGGGGGGGGGCACCAC ACAACACCTTTCTCTATCTG	176 3 3898 4018 4018 4138 194 4258 225 3 4378 24498 1 4618 4 4738 2359 2350
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739	CAGGAGGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA I TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGGAAGAATGT GGTGGTGGGCGGGCTGTGCCACA I TCATTCTGTGACTTCTGTG I TAGGGTCTTAGTGAATACA CTTCTTTTTGCAGGACCTC	V S S T STCTCATCCAC S P I T IGCCCCATCAC I ICTACATACAC L S G I TGAGCGGCAT IGGAGGGAGGGCAT IGGAGGGAGGAGAC I IGTTGGTGAGAGA R G P P IGAGCCCTCC I IGAGCCCTCCC	ACAGCTTATT/	R N G L AGGAATGGGGA AGTAAGGGAA1 TTACTGAGGTT AGGTTTTTCCCC I V L G V TGCGGAGGAAA1 TCTCTTTTCCAGC I AGGTAAGCTGG	CTAATATCCA GAGACCTACA GAGACCTACA GAAGGTCCC GAAGGTCCC GGAAGGTTCC GGAAGCTTCC CTCTATCTGC	Q V L CCAGGTCCTGE I GGAAATCATAT GGAAATCATAT I GGAGGAATTGGC CGGGCTCGGCC AGGAGAGACAT I AGGAGAGAGACAT I TGCCCCTTGGG I AGGGTGGTGGTGCA	V M L E STCATGCTGGA SGGCCCCCACAT ICTCTCACCCC CTGAAGTAGAG CTGGAAGTAGAG TTGGGATCTGA STCGTTGGGATCTGA STCGTTGTGGAGGG	M T P GATGACCCCI GACATGGGAI ACAATCAGGA TGTAGACACCI H R S TCACAGGAG TTTTGCTGGG AGGGCCCCT/	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAG <u>GT</u> TATGTTACT(AGTTTCATCA TTTTTTTATAA	E V Y T C H V AGGETTACACCTGCCATGT TATTATCCCATCCTGCCATGT AGCCCTGACCTATCTTCTCA TCCCCCAGGGGGCACAGTCCGA AGGGAGCTCTGGGAGAATTGG CCACAGAGCCATGGTGGAG ACTTGGGGGGGGGGCACCAC ACAACACCTTTCTCTATCTG CAAGCTATGCTTGGCACAGT	176 3898 3898 188 4018 4018 4138 194 4258 225 3498 225 4378 2498 4498 4738 4458 4858 4858 4858
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGGCAAGATGT GGTGGGGGGGGCGTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGC <u>AG</u> GACTTCT CTGGTTCCCAATCTGTCTC	V S S T STCTCATCCAC S P I T GCCCCATCAC I SCTACATACAC L S G I TGAGCGGCAT GGAAGGAGGACC I CCTGTCCACC CCTGTCCACC I GGAGGCCCTCC I TGCTGTGAGAGA I TGCTGTGAGAGA I TGCTGTGAGAGA	CAGCAGGET	R N G (AGEAATGGGGAA TACTGAGGCTG TACTGAGCTG V L G V TGGGAGGAAAA TGGGAGGAAAA TGGGAGGAAAA TGGGAGGAAAA TGGCTAAGCTGCTA AATATTTCAGG CATAGCTGCTA	W T F CTGGACCTTO CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTCTATCTGC CTCTATCTGC	Q V L CCAGGTCCTGE I I GGAAATCATAT GGAAATCATAT GGAGGAATTGGC CGGGCTCGGCC I AGGAGAGACAT I GGCCATGAAAA I TGCCCCTTGGG I AGGGTGGTGCCATA	V M L E STCATGCTGGA GGGCCCCACAT TCTCTCACCTC CTGAAGTAGAG L F I F TTGGGATCTGA TTGGGATCTGA GGGAGGTGAGGG	M T P GATGACCCCI GACATGGGAI ACAATCAGGA TGTAGACACI H R S TCACAGGAG TTTTGCTGGG TTCCTGTGTG AGGGCCCCT/ GGGGTTGTC	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAA/ TAGCTCTAT CAGAAAG <u>GT</u> TAGTTCATC TTTTTTTATA/ AGGGTGCGAG(CCAGCAGGAGA	E V Y T C H V SAGGTCTACACCTGCCATGT	176 3898 4018 4018 4018 4138 3194 4258 4378 44258 4378 4498 4458 4458 4458 4458 4458
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979	CAGGAGGAGAGACAGTGGGGG GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA SARSKM TCTGCCCGGAGGCAGGATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGCAGGACTCG CTGGTTCCCAATCTGTCT GGTGTCTCGGTTGGAGCC	V S S T STCTCATCCAC S P I T IGCCCCATCACA I ICCTACATACAC L S G I TGAGEGGGCAT I IGGAAGGAGGACC I ICCTGTCCACC I IGCTGGTGGAGA I TGAGECCCTCC I TGCAGECCAGTT I I I I I I I I I I I I I	C L I ACAGCTTATT/ TGTGGAGTGG/ TCTGGTTCCAC TCTGGTTCCAC CGGGGGCCCCCG/ GGCCGCGCCCCG/ ACCCTCACCGCI GACAATTCC/ TCCAGCAGGCI/ TTGACAGGGCI/ TTGACAGGTGCC	R N G L AGGAATGGGGAA1	W T F CTGGACCTTIT AGGACCTACA CTAATATCCA V I F L GGATCTTCCT CTAATATCCCA CTGCACGGCCCC CAGGACCTTCC CTCTATCTGC CTCCTATCTGC	Q V L CCAGGTCCTGE I I GGAAATCATAT GGAAATCATAT I GGGGGAATTGGCC GGGGCTCGGCC I GGCAATGAAAA I GGCCACTGGGGCCA I CGGGGTGCCATAT I CGCCCCCACACT	V M L E STCATGCTGGA GGGCCCCACAT CTCTCACCCTC CTGAAGTAGAG TTGGGAAGTAGAG TTGGGATCTGG GGGAGTGAGGC GGGAGGTGAGGC GGGAAGCTCCC	GATGACCCCI GACATGGGAI CACAATCAGGA CTGTAGACACCI CACAGGAGA CTTTTGCTGGG CTCCTGTGTG CACGGGCCCCT/ CGGGGTTGCCC GCCGTGGAC	H 0 G CATCAGGGA(TTTAGGTGT) CATCAGGAAA TAGCTCTAT CAGAAAAG <u>GT</u> TAGTTCATCT AGGTTCATCT TTTTTTTATA/	E V Y T C H V AGGETCTACACCTGCCATGT TATTATCCCATCCCTCCAAT AGCCCTGACCTATCTTCTCA TCCCCAGGGGGCACAGTCCGA TCCCCCAGGGGGGCACAGTCCGA CACTGGGGGGGGGGGCACCAC ACAACACCTTTCTCTATCTG CAGGCAAGAGGCCACGACGTC CAGGCAAGAGGGCCACGAGGTC CAGGCAAGAGGGCCACGAGGTC	176 3898 188 4018 4018 4138 194 2498 4258 4378 4498 4498 4498 44738 4858 4978 5998 2008 2009
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099	CAGGAGGAGAGACAGTGGGGG GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA SARSK TCTGCCCGGAGGCAAGATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGCAGGACCTCG CTGGTTCCCAATCTGTCT GGTGTCTCGAGTTGGAGCC AGCAGAGGATATGGATCCG	V S S T STCTCATCCAC S P I T ISCCCCATCAC I ICCTATATGA I CTACATACAC I L S G I I TGAGECGGCAT I I GGAAGGAGGAGCC I CCTGTCCACC I I GGAGGCCCTCC I TGCAGCCCCTCC I TGCAGCCCCTCC I TGCAGCCCCTCC I TGCAGCCCCTCC I TGCAGCCCCTCC I TGCAGCCCCTCC I TGCAGCCCCTCC I I TGCAGCCCCTCC I I CGAGCCCCTCC I I CGAGCCCCTCC I I CGAGCCCCTCC I I CGAGCCCCTCC I I CCTGTCCACC I I I I I I I I I I I I I	CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	R N G L AGGAATGGGGAA AGTAAGGGAAA TTACTGAGCTG B TGCGGAGGAAAA ICTCTTTGGGG CTCTTTTCCA AGGAAGGAGAAAA ICTCTTTTCCA AGGAAGCTGCC ATAGCTGCTA AGGAGGTGCCA	W T F CTGGACCTTO AGGACCTACA CTAATATCCA V I F L GGATCTTCCT CAGGACTTCC CAGGACTTCC CAGGACTTCC CAGGACTTCC CTCTATCTGC	0 V L CCAGGTCCTGE I I GGAAATCATAT GGAAATCATAT GGGAGAATCATAT GGGAGAGGAGGACAT I GGCAATGAAAAA I GGCCACTGGGGGCA I AGGGTGGGTGCATA I GGCCCCCACACT I AGGATGGGTGGC	V M L E STCATGCTGGA GGGCCCCCACAT CTCTCACCTC CTGAAGTAGAC TTGGAAGTAGAC TTGGGAACTAGGC GGGGAGGCGAGGC	M T P GATGACCCCI GACATGGGAI CACAATCAGGA CTGTAGACACCI CACAGGAG CTCCCGGAG CTCCCGGGAC CCCCCCCACAC	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAG <u>GT</u> TAGCTTCATC(AGTTTCATC) AGGGTGCGAGG CCCAGCAGGAGGAG TTGACAGTG(GTTTGACAGTG(E V Y T C H V AGGTCTACACCTGCCATGT I TATTATCCCATCCCTCCCAAT AGCCCTGACCTATCTTCTCA C TCCCCCAGGGGGGCACAGTCCGA C CCACGAGGCCTGGAGAATTGG CAGCTATGCTGGGGGGGCACCAC C CAGCAACACCTTTCTCTATCTG CAGCCATGCTGGCACAGT CAGCCATGCTGGCACAGT CCAGGAAGAGGCTGAAGGTC CTGTGAAGTTCTGTGGGTTTA	176 3898 3898 188 4018 4018 4138 194 4258 225 4378 2498 4498 4618 4738 233 4858 4978 5 5218 4958 5218
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099 5219	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA CTCCCGCGGAGCAAGATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGCGGACTCTG GGTGTTCCCAATCTGCTCC GGTGTTCTGGATGGATCCC AGCAGAGGATATGGATCCC J	V S S T STCTCATCCAC S P I T ISCCCCATCAC STCCTATATGA I CCTACATACAC I L S G I I GGAAGGAGAGCC I SCTGGTGGGGAGA A B GGAGGCCCTCC I TGAGCCAGTT I TTGAACTTAC I TTGAACTTAC I TTGCACCAGG I	CAATGGCAAGGCAG	R N G L AGGAATGGGGAA L L L AGTAAGGGAA L L L TACTGAGCTC L L L AGGTATGGGGAA L L L AGGTATGGGGAA L L L AGGTAGGGAA L L L AGGTAGGGAAA L L L AGGTAGGGGAAA L L L AGGTAGGGGAAA L L L AGGTAGGGGAAA L L L AGGTAGGGGGAAA	CTAATATCCA GAGACCTACA GAGACCTACA GAGACCTACA GAAGGTCCC GAAGGTCCC CAGGAAGCTTCC CTCTATCTGC GCCTATCCGC GGCCATAACA	0 V L CCAGGTCCTGG 1 TTTCAACTGTG GGAAATCATAT GGAAATCATAT GGAAATCATAT GGAAATCATAT GGAAATCATAT GGAATCATAT AGGAGAGAGACAT 1 CGCCCCCTGGCTG 1 AGGAGTGGTGGCGAT 1 CAGGTGTCCCATA 1 CCAGCTGTCCGAT 1 CCAGCTGTCTG 1 1 CAGCGGTGGTGGCG 1 1 1 1 1 1 1 1 1 1 1 1 1	V M L E STCATGCTGGA GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC TGGGAAGTAGGC TTGGGATCTGA AGGGGTGAGGC GGGAGAGCTCCA CTGGTAGCCCACTCCA CTCCTTTCCACCCC CTCCTTTCCACCCCACTCC	GATGACCCCI GACATGGGAI GACAATGGGAI CACAATCAGGG CTTCACAGGAG CTTCCTGTGTG GAGGGCCCCCI GAGGGCCCCCI GGCGTGGGAC CATCCCACA(CTATGAGATT(H 0 G CATCAGGGAG TTTAGGTGTT CATCAGGAAA TAGGTCCTAT CAGAAAAG <u>GT</u> TATGTTACTC CAGATCAGAGAG TTTTTTTTATAJ	E V Y T C H V AGGTCTACACCTGCCATGT I TATTATCCCATCCCTCCCAAT AGCCCTGACCTATCTTCTCA T TCCCCCAGGGGGCACAGTCCGA C CCCCCAGGGGGGGGCACCAC C CCCCCAGGGGGGGGGG	176 3898 188 4018 4018 194 194 24138 225 4258 225 4378 2498 4498 4458 4858 4858 4858 4978 55098 5218 5338
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099 5219 2255	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGCAAGATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTGCCACA CTTCTTTTGCAGGACTCTGTC GGTGTTCCCAATCTGTCTC GGTGTCTTGAGTTGGAGCC AGCAGAGGATATGGATCCC GGGTTTAGCTGCTTTCACC	V S S T STCTCATCCAC S P I T GCCCCATCAC CTACATATGA CTACATACAC L S G I TGAGCGGCAT CGTGGTGGGGAGA CCTGCTGTCACC I CCTGTGTGGTGAGA C CGTGGTGGTGAGG C CGTGGTGGAGGA C CGTGGTGGAGGA C CGTGGCCACTAC C CTTCCACAGG C C C C C C C C C C C C C	CAATGGCAGTGGC	R N G L AGEAATGGGGAA1 L L L CTACTGAGCTC L L L CTGCTTGGGGAAA1 L L L V L G L L STGCTTGGGGAAAA L L L L V L G L L L STGCTTGGGGAAAA L <td>CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTCTATCTCC CAGAAGCTTCC CAGAAGCTTCC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCCTATCTGC CTCCTTGCTCTTG CTCCTTGCTCCTTG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTTG</td> <td>Q V L CCAGGTCCTGG I TTTCAACTGTG GGAAATCATAT GGAAATCATAT GGAAATCATAT I GGGAATCATAT GGCGGCCGGCC I GGCAATGGAGACAT I GGCCATGGTGCCATA I CAGGTGTCCCATAT I GGCCACCGTCCCC I CAGGTGTGGTGCC I CCAGCTGTCTGCCG I CCAGCTGTCTGCCG</td> <td>V M L E STCATGCTGGA GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC TGGGAAGTAGGC GGGGAGCTGAG GGGGAGGCTGAGGC CTCCTTGCACCCA GGGAGGCTGAGGC CTCCTTTCCA</td> <td>GATGACCCCI GACATGGGAI GACAATCAGGA TGTAGACACCI TTTTGCTGGG TTTTTGCTGGG GAGGGCCCCT/ GAGGGCCCCT/ GGGGGTTGTCC GCCGTGGGAC CATCCCACAC</td> <td>H 0 G CATCAGGGAG TTTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAG<u>GT</u> CAGAAAG<u>GT</u> TATGTTACT(CAGTTCATCA CCCAGCAGGAG CCCAGCAGGAG CTTGACAGTG(CAGAGCCAGGAG CTTGACAGTGC</td> <td>E V Y T C H V AGGTCTACACCTGCCATGT I TATTATCCCATCCCTCCCAAT AGCCCTGACCTATCTTCTCA C C C C C C C C C C C C C</td> <td>176 3898 4018 4018 4138 194 4258 225 4378 2498 4498 4458 4498 4498 4498 4858 4978 5598 5218 5338 228</td>	CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTCTATCTCC CAGAAGCTTCC CAGAAGCTTCC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCCTATCTGC CTCCTTGCTCTTG CTCCTTGCTCCTTG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTCCCTTGG CTCCTTGCCTTG	Q V L CCAGGTCCTGG I TTTCAACTGTG GGAAATCATAT GGAAATCATAT GGAAATCATAT I GGGAATCATAT GGCGGCCGGCC I GGCAATGGAGACAT I GGCCATGGTGCCATA I CAGGTGTCCCATAT I GGCCACCGTCCCC I CAGGTGTGGTGCC I CCAGCTGTCTGCCG I CCAGCTGTCTGCCG	V M L E STCATGCTGGA GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC TGGGAAGTAGGC GGGGAGCTGAG GGGGAGGCTGAGGC CTCCTTGCACCCA GGGAGGCTGAGGC CTCCTTTCCA	GATGACCCCI GACATGGGAI GACAATCAGGA TGTAGACACCI TTTTGCTGGG TTTTTGCTGGG GAGGGCCCCT/ GAGGGCCCCT/ GGGGGTTGTCC GCCGTGGGAC CATCCCACAC	H 0 G CATCAGGGAG TTTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAG <u>GT</u> CAGAAAG <u>GT</u> TATGTTACT(CAGTTCATCA CCCAGCAGGAG CCCAGCAGGAG CTTGACAGTG(CAGAGCCAGGAG CTTGACAGTGC	E V Y T C H V AGGTCTACACCTGCCATGT I TATTATCCCATCCCTCCCAAT AGCCCTGACCTATCTTCTCA C C C C C C C C C C C C C	176 3898 4018 4018 4138 194 4258 225 4378 2498 4498 4458 4498 4498 4498 4858 4978 5598 5218 5338 228
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099 5219 235 5339	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGCAAGATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGCAGTGGACTCC GGTGTCTTGAGTTGGAGCC AGCAGAGGATATGGATCCC GGGTTTAGCTGCTTCACC	V S S T STCTCATCCAC S P I T GCCCCATCAC I TCACATACAC I CTACATACAC I CTACATACAC I CTACATACAC I CCTACATACACAC I CCTACATACACAC I CCTACATACACAC I CCTACATACACACACACACACACACACACACACACACAC	G L I ACAGCTTATT/ TGTGGAGTGG/ TCTGGTTCCAC GGGGGGGCTGCC GGGGGGGCTGCC GGGCTGCAGGC ACCCTCACTGI GTTCCAGCAGGC TTGACAGGGCAGGC CAAAAGATCGC CAAAAGATCGC CAATGGCGAAG	R N G L AGGAATGGGGAA1 TTACTGAGGTTTTTCCCC I STGCTTGGGGGAGAAA1 I GGGGAGGAAAA1 I GGGGAGGAAAA1 I GGGAAGTGGCAAG I SAGAAGTTGCCI SAGAGGGGCAACI I SAGGAGGGGCAACI I GCTCCCCAC <u>AGG</u> I CTCCCCAC <u>AGG</u>	CTAATATCCA GAGACCTACA CTAATATCCA CTAATATCCA CTAATATCCA CAGAGGTCCC CAGAGGTCCC CAGGACTTGT CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCCTTGCTGCAGG CTCCTGCTGCAG CCCCCTGCAG	Q V L CCAGGTCCTGE I GGAAATCATAT GGAAATCATAT GGAAATCATAT GGAGGGAATCGGCC I GGGCAATGGAGGACAT GGCCACTGGCCC GGCAATGGAGACAT I CAGGGGGGGGGCCA I AGGATGGTGCCACT I AGGATGGTGGCGC I TGACTCAGAGT	V M L E STCATGCTGGA GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC TGGGAGTCTGG GTTTTCATCCC GTCGTTGGGATCTGA GGGGGGGGGG	M T P GATGACCCCI GACATGGGAI GACAATGGGAA TGTAGACACCI TCCCAGGGAG TTTTGCTGG GAGGGCCCCT/ GTGTGTGGGGAC GCCGTGGGAC CATCCCACA(TATGAGATT(CAGTTGACTG	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAA) TAGCTCTAT CAGAAG <u>GT</u> CAGAAG <u>GT</u> CAGATAGCAGAG CCAGCAGGAG CCAGCAGGAGGAG CTTGACAGTGC CAAGAGGTCAG	E V Y T C H V AGGETCTACACCTGCCATGT I I I I I I I I I I I I I	176 3898 188 4018 194 225 194 194 194 194 195 194 1978
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099 5219 235 5339 5459	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGGAGAAGATGT GGTGGTGGGCGGCTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGCAGGACTCTC GGTGTCTTGAGTTGGAGCC AGCAGAGGATATGGATCCC GGGTTTAGCTGCTTCACC CTCTGCTCCCCCCACTCCCT TCGAATTCCCGACTGCCA	V S S T STCTCATCCAC S P I T GCCCCATCAC I CTACATACAC I CTACATACACA I CTACATACACA I CTACATACACA I CTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACA I CCTACATACACACA I CCTACATACACACA I CCTACACACACACACA I CCTACACACACACACACACACACACACACACACACACAC	G L I ACAGCTTATT/ TGTGGAGTGG/ TCTGGTTCCAC TCTGGTTCCAC G G C CGGGGGGCTGCC GGGGGGCTGCC GACCACTCACTGC CACCCCAGCAGGT/ CAAAAGATCGC GAAAAAGATCGC CAAAGACGCGAGC TTTCCTTCCC	R N G (AGGAATGGGGAA1 AGGATATGGGGAA1 TACTGAGCTC AGGTTTTTCCCC V L G V TGGGAGGGAAA1 IGGGAGGGAAA1 IGGGAGGGAAA1 IGGGAGGGAAA1 IGGGAGGGAAA1 IGGGAGGGGAAA1 IGGGAGGGGAAA1 IGGGAGGGGAAA1 IGGGAGGGGCAAC IGGGAGGGGCAAC IGGGGGGGGGCAAC IGGGGCGGCAAC IGGGGGCAAG IGGGGGCAAG	CTAATATCCA GAGAGGTCCC GAGAGGTCCC GAGAGGTCCC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCCTTGCTCCTTG CTCCTTGCTCCTTG CTCCCTTGG CTCCCCCGCGG CATCGCCCCA	Q V L CCAGGTCCTGE I GGAAATCATAT GGAAATCATAT I GGAAATCATAT I GGCCCCGGCC I AGGAGAGAGACAT I GGCCATGAAAA I CAGGAGGAGGACAT I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGGTGGTGCCATA I CAGGTGTCCACATI I CAGGTGTCCACATI I CAGGTGTCCACATI I CAGGTGTCCCACATI I CAGGTGTCCCACATI I CAGGTGTCCCACATI I CAGGTGTCCCACATI	V M L E STCATGCTGGA GGGCCCCACAT CTCTCACCTC CTGAAGTAGAC TGGGAGTCTG GTGCTTAGGCC GGGGGTGAGGC GGGGGGGGAGGC GGGAGACTGTG CTTCTTTTCCA TGGAGACCTGTC TGGTTTTGACTC	M T P GATGACCCCT GACATGGGAT GACAATGGGAT TGTAGACACCT TTTTGCTGG GAGGGCCCCT/ GGGGGTGGGAC CATCCCACAG TATGAGATT(CAGTTGACTGT	H 0 6 CATCAGGGA(TTTAGGTGTT CATCAGGAAA CAGCACCAGAAAGG <u>C</u> TAGCTCCTAT CAGAAAG <u>GC</u> TTATGTTACT(CAGAAAG <u>GC</u> CCAGCAGGAG CCCAGCAGGAG CTTGACAGGGC CGTTGTGTCCC CAAGAGTCAG CCCCCCCC	E V Y T C H V AGGETCTACACCTGCCATGT I I I I I I I I I I I I I	176 3898 188 4018 4138 194 24138 225 3478 2498 4498 44738 233 4858 4978 5598 35218 15338 12338 2335 24978 2598 35218 15338 12578 238 12578
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099 5219 235 5339 5459	CAGGAGGAGAGACAGTGGGGG E H P S L K GAGCATCCCAGCCTGAAGAA TCACCCACCCCATCACTTG ACAGCAGTTCTGGAGATCA S A R S K M TCTGCCCGGAGGAAGAATGT GGTGGTGGGCTGTGCCACA TCATTCTGTGACTTCTGTG TAGGGTCTTAGTGAATACA CTTCTTTTTGCAGGAGCACATCTGTCT GGTGTCTCGATCTGGATCCC GGTGTCTGGGGGTTTCACC CTCGCTCCCCCCCCCCCC TCGATTCCCGACTGCCAC	V S S T STCTCATCCAC S P I T GCCCCATCAC I CTACATACAC I CTACATACAC I CTACATACAC I CTACATACAC I CTACATACAC I CTACATACAC I CCCTGTCCACC I CCCTTCACCAC I CCTTTCACCACCAC I CCTTTCACCAC I CCTTTCACCACCAC I CCTTTCACCACCAC I CCTTTCACCACCAC I CCTTTCACCACCAC I CCTTTCACCACCAC I CCTTTCACCACACACACACACACACACACACACACACAC	CAAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	R N G (AGEAATGGGGAA1 TACTGAGGTTTTTCCC AGGTTTTTCCGA TAGGGAGGAAA1 TGGGGAGGGAAA1 TGGGGAGGGAAA1 TGGCTAAGCTGCG AATATTTCAGC CATAGCTGCTA AGGTGGCAAGCTGC AATACTTCCAA AGGTGGCATG CCCACAGGCTTA	CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CAGAAGCTTCC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCCATCACAGCA CTCCCCGCAGA CATCGCACTA	Q V L CCAGGTCCTGE I GAAGGAATTGGC GGAAATTGGC GGAAATTGGC GGGATTGGCGC AGGGGCGGGCGGCC I GGCAATGAAAA I GGCCATGATGAAAA I GGCCATGGTGGCCATA I TGCCCCCTTGGG I AGGGTGGTGGTGCC I CCAGCTGCCCATA I TGCCCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCAGCTGCCCACTI I CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	V M L E STCATGCTGGA SGGCCCCCACAT CTGAAGTAGAG CTGAAGTAGAGA CTGGAAGTAGGAG TGGGATGTGGA STGCTTGGGAGCGAG STGCTTGGGAGACGTGG CTGCTTGGGAGACGTGG CTGTTTTGACTG CACCCAGCTG ACCCAATGTA	M T P GATGACCCCT GACATGGGAT GACAATGGGAA TGTAGACACC TGTAGACACCT TTCCCGGGAG GGGGCTGGGAC GGGGGTGGGAC CATCCCACAG CAGTGACTG CTGTGAGATT(GGTTATCCC	H 0 G CATCAGGGA(TTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAG <u>GT</u> CAGAAAG <u>GT</u> CAGATCATCA CAGATCATCA CCAGAAAGG <u>T</u> CCAGAAGGGAGA CCTAGAGAGGAG CCTCAGACTGC CAAGAGTCCCCC CACCAGATTA	E V Y T C H V AGGETCTACACCTGCCATGT I I I I I I I I I I I I I	176 3898 188 4018 4138 194 4258 225 4378 2498 4498 4428 4498 44738 233 24958 5598 35218 75388 238 238 25578 5578 5578
177 3899 4019 190 4139 195 4259 4379 4499 4619 227 4739 4859 4979 5099 5219 235 5339 5459 5579 5699	CAGGAGGAGAGAGAGGGGGG E H P S L K GAGCATCCCAGCCTGAAGAA TCACCCACCCCATCACTTG ACAGGCAGTTCTGGAGGATCA S A R S K M TCTGCCCGGAGGAGAAGATGT GGTGGTGGGCTGTGGCACA TCATTCTGTGACTTCTGTC TAGGGTCTTAGTGAATACA CTGGTTCCCAATCTGTCTC GGTGTCTGGAGGACCCC AGCAGAGGGATATGGATCCCC GGGTTTAGCTGCTTCCACCCC TCGATTCCCTGACTGCCA CTCTGCTCCCCCACTCCCCT TCGAATCCCTGACTGCCA CTGGTCTTGGTGGTGCTACCAGAGAC CTGGTCTTGGTGGTGCTACCAGAGAC	V S S T STCTCATCCAC S P I T GCCCCATCAC I CTACATACAC I CTACATACAC I CTACATACAC I CTACATACAC I CCCTACATACAC I CCCTGTCCACC I CCCTGTCCACCAGG I CCCTTTCACTGC I CCCTTTCACCAGG I CCCTTTCACCAGG I CCCTTTCACCTGC I CCCCTCCCTGC I CCCCTCCCTGC I CCCCTCCCTGC I CCCCTCCCTGC I CCCCTCCCTGC I CCCCTCCCTGC I CCCCTCCCTGC I CCCCCTCCCTGC I CCCCCTCCCTGC I CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CAACAGCTTATTC CGGGGGCCCTGC CGGGGGCCCTGC CGGGGGCCCTGCC CGGGGGCCCCCGCC CGGGGGCCCCCGCC CGGGGGCCCCCGCC CGGGGGCCCCCGCCCC CGACCACCACTCCCC CAACAGCAGCAGCCC CAACAGCAGCAGCCC CAACAGCAGCAGCCC CAACGCCGCCCCGCC	R N G L AGETAAGGGAA1 AGETAAGGGAA1 L L TACTGAGCTG L L L L V L G L <	TITCTTTTC AGAACCTACA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATATCCA CTAATCTCCT CAGAACCTTCC CTCTATCTGC CTCTATCTGC CTCTATCTGC CTCTATCTGCA CTCCTCCAGA CATCGCACTA CTCGCACTACTACTACTACTACTACTACTACTACTACTACTACT	O V L CCAGGTCCTGE I GAAGGAATTGGC GGAAATTGGC I GGGGATTGGC GGGGCTCGGCC I AGGGAGAGACAT I GGCCATGAAAA I GGCCATGAAGAG I CCAGTGGTGCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCATA I CCAGCTGCCCACAT I CCAGCTGCCCACAT I CCAGCTGCCCACAT I CCAGCTGCCCACAT I CCAGCTGCCCACAT I CCAGCTGCCCACAT I CCCGCCCCACACAT I CCCGCCCCACACAT I CCCGCCCCACACAT I CCCGCCCCCCCCACACAT I CCCGCCCCCCCCCCCCACACA I CCCGCCCCCCCCCC	V M L E STCATGCTGGA SGGCCCCACAT TCTCTCACCTC CTGAAGTAGAGA TGGCATGTGGAGACTGA TGGGAGAGCTGAGGA STCGTTGTGAGAGCA STGCTTAACTCA TGTTTTGACTACTCA TGTTTTGACTACTCA TGTTTTGACTACTCA TGTTTGACTACTCA TGTGTTTGACTACTCA TGTGTTGGAGAGCTGA TGTGGTTGCAAGTTACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA TGTGGTTGCTAACTCA	M T P GATGACCCCI GACATGGGAI ACAATCAGGA TGTAGACACI TGTAGACACCI TTTTGCTGG TTCCTGTGTG GGGGCTCGCCCCI GGCGTGGGAC CATCCCACAC CAGTGACTG TTGTGAGATTC CAGTGACTG CCTCCCCCAC	H 0 6 CATCAGGGA(TTTAGGTGTT CATCAGGAAA TAGCTCTAT CAGAAAGG <u>GT</u> CAGAAAG <u>GT</u> TATGTTACT(AGGTTCATC(CAGATCAGAC TTTATTTTATAA AGGGTGCGAGAG CTTGACAGGGG CAGAGAGTCAGA CAGAGTCCCCC CAGAGACCCCCA CACCGATTAA	E V Y T C H V AGGETTACACCTGCCATGT I TATTATCCCATCCTGCCATGT AGCCCTGACCTATCTTCTCA I CCCCCAGGGGGGCCACAGTCTGG AGGGAGCTCTGGAGAATTGG CCACAGAGCCATGGTGGAGA ACATCGGGGGGGGGCACCAC ACAACACCTTTCTGGCACAGT CCAGGAAGAGGGCTGAAGGTC I CCAGGAAGAGGGCTGAAGGTC I CTGTGAAGTCTGTGGGCACAGT I TAAGGCCTGCATGTCTGTGG ATAACTGTCTGTATCTCTG I TAAGGCCAGGAACCCAGAC I TAAGGCCTGCATGTCTGTG CAACCACCTTCTGTAGCCCTG I TAAGGCCTGCATGTCTGTG ATAACTGTCTGTATCTCTTG I TAACGCCTGCATCTCTGTATCTCTTG I TAACGCCTGCATCTCTGTATCTCTTG I TAACGCCTGCATCTCTCTG I TAACGCCTGCATCCTCTGTATCTCTTG I TAACGCCTGCATCTCTCTG I TAACGCCTGCATCTCTCTG I TAACGCCTGCATCTCTCTG I TAACGCCTGCATCTCTGTATCTCTTG I TAACCGTCTGTATCCTCTG	176 3898 4018 4018 4138 194 24138 225 4378 2498 4428 4428 4428 4498 4498 4498 4498 4498 4978 5098 5218 5338 2338 25578 5578 5818

has lost this exon subsequent to the divergence of humans and mice (17).

Similarity comparisons were made between defined regions of mouse and human Ia polypeptides. The greater similarity of $A\beta^d$ to DC- β (76 percent) than to DR- β (65 to 68 percent) supports the model that DC- β is the human homolog of the mouse I-A β molecule (18). The relative similarities between the equivalent domains of $A\beta^d$ and DC- β indicate that the first domain is the more divergent of the two external domains (66 and 86 percent, respectively). The high degree of similarity between the connecting peptides (90 percent) and transmembrane regions (86 percent) of human and mouse sequences is rather surprising in view of their presumably general functions. Conservation between similar regions of I-E-like α chains, class I molecules, and immunoglobulins is much less. This level of conservation leads us to suspect that the connecting and transmembrane regions may be involved in the interaction with the α chain or some other membrane protein rather than serving just as hydrophobic anchorage. The size difference between the $A\beta^d$ and DC- β cytoplasmic regions implies that significant variation of this region may have less severe selective consequences than it would for other regions of the gene.

Computer analysis reveals no significant similarity between the β_1 and β_2 domains of $A\beta^d$ nor between β_1 and any domain of any other gene involved in immune responses. However, the β_2 domain is significantly similar to the α_2 domains of class II proteins, the third external domain (α_3) of class I heavy chains, B2-microglobulin, and immunoglobulin constant-region domains. A computer-aided alignment was made between most of the known representatives of these "Ig-like" domains. The alignment (data not shown) shows a striking

Fig. 2. The DNA sequence of 5.8 kb of the cosmid clone 34.2 containing the A β gene. The RNA splicing donor and acceptor sites (GT-AG) (G, guanine) are underlined as is the probable 3' untranslated region. A possible CCAAT box and poly(A) addition signal sequence are boxed. The gap shown between bases 870 and 871 represents less than 200 bases of undetermined sequence. The predicted amino acid sequences encoded by the exons are shown above the DNA. Singleletter amino acid codes are placed above the middle bases of the translated codons. The single-letter amino acid code is abbreviated as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; and W, tryptophan.



Fig. 3. Comparison of the intron-exon organization of class I. class II. and B2microglobulin (24). The $A\beta^d$ is representative of a class II β chain, the E α^k (13) of a class II α chain. and L^d of a transplantation antigen heavy chain (25). Solid, filled areas encode residues present in the mature proteins, crossed areas show the portions en-

coding the signal peptides, and hatched boxes represent the 3' untranslated sequences. The abbreviations are the same as in Fig. 1; β_1 and β_2 and α_1 and α_3 refer to exons encoding the external domains of the protein.

conservation of functionally similar residues responsible in immunoglobulins for the β -pleated sheet regions of a tertiary structure referred to as the antibody fold (19). The alignment shows that the class I, class II, and β_2 -microglobulin sequences are roughly comparable in their similarities to each other as well as comparably less similar to immunoglobulin sequences, establishing two evolutionary subgroups.

The organizations of the $A\beta^d$ gene, a class II α gene, a class I heavy chain, and β_2 -microglobulin are shown in Fig. 3. The most striking comparison to be made from Fig. 3 is between class II α genes and β_2 -microglobulin. Note particularly (i) the distances between the signal peptide and the Ig-like domain, (ii) that the first few amino acids of each mature protein are encoded with the signal peptide by the first exon, and (iii) the rare occurrence of an intervening sequence in the 3' nontranslated region of both genes. The $A\beta^d$ gene shares the first two features but not the third one. Also, like the class I genes and not the class II α genes, the cytoplasmic coding sequence of $A\beta^{d}$ occurs on multiple exons.

These organizational comparisons support the suggestion (14) that class II α chains and β_2 -microglobulin are more directly related evolutionarily than are class II α and β chains. It also was suggested that class I and class II genes might not be paralogous evolutionarily, but are each the result of de novo gene construction through the process of "exon shuffling." However, the structure of the $A\beta^d$ gene as presented here is not as dissimilar to class I genes as is the $E\alpha$ gene. Moreover, though not significantly similar over their entire length, the first external domains of class I heavy chains (α_1) and the β_1 domains of class II B chains do share a highly conserved peptide sequence about residue 40 (Val-Arg-Phe-Asp-Ser-Asp). Aβ^d differs by one conservative substitution (Phe to Tyr). The shorter α_1 domains of class II α chains possess a similar and completely conserved sequence, Phe-Asp-Gly-Asp, about residue 30, but have no better than random similarity beyond this with the α_1 domains of class I or β_1 domains of class II. The conservation of sequence across all examples of class I and class II molecules and the structural similarities of the genes support a direct evolutionary connection between class I and class II genes, particularly those of class II β chains.

Antigen triggering of cellular and humoral immune responses requires the recognition by different T-cell populations of antigens presented in the context of class I or class II molecules. The charged nature of the conserved peptide probably ensures its residence on the exterior of the molecule. As a conserved, exposed determinant, this peptide sequence may play a homologous role in T cell recognition of both class I and class II antigens. It is possible then that not only are class I, class II, and β_2 microglobulin related by distant ancestry to immunoglobulin, but that class I and class II antigens are derived from the same ancestral two-chain polypeptide involved in prototypical lymphocyte interactions.

Note added in proof: Comparison of the $A\beta^d$ sequence with that of the bhaplotype allele (26) indicates a relative deletion of an Alu-like (β_1) repeat sequence in the second intron of $A\beta^d$. Computer analysis reveals a repeat-like element only vaguely similar to β_1 about 300 bp 5' of the β_2 domains of both alleles.

> MARIE MALISSEN TIM HUNKAPILLER LEROY HOOD

Division of Biology, California Institute of Technology, Pasadena 91125

References and Notes

- 1. J. Klein, A. Juretic, C. Baxevanis, Z. Nagy,
- J. Klein, A. Juretic, C. Baxevanis, Z. Pagy, Nature (London) 291, 455 (1981).
 A. Singer, Y. Asano, M. Shigeta, K. S. Hath-cock, A. Ahmed, C. G. Fathman, R. J. Hodes, *Immunol. Rev.* 64, 137 (1982).
 H. O. McDevitt, J. Immunogenet. 8, 287 (1981).

- 4. J. M. Cecka, M. McMillan, D. B. Murphy, H. O. McDevitt, L. Hood, Eur. J. Immunol. 9, 955 (1979)
- 5. E. Sung, M. W. Hunkapiller, L. E. Hood, P. P. Jones, in preparation. 6.
- H. Kratzin et al., Hoppe-Seyler's Z. Physiol. Chem. 362, 1665 (1981). 7
- L. Auffray, A. J. Korman, M. Roux-Dosseto, R. Bono, J. L. Strominger, *Proc. Natl. Acad. Sci.* Bono, J. U.S.A. 79, 6337 (1982) 8.
- C. O. Benoist, D. J. Mathis, M. R. Kanter, V. E. Williams II, H. O. McDevitt, *ibid.* 80, 534 E. Williams II, H. O. McDevitt, *ibia.* 60, 557 (1983).
 9. A. J. Korman, C. Auffray, A. Schamboeck, J. L. Strominger, *ibid.* 79, 6013 (1982).
 10. J. S. Lee *et al.*, *Nature (London)* 299, 750 (1992).
- (1982)11. D. Larhammar et al., Proc. Natl. Acad. Sci.
- D. Latnamma et al., Arost et al., Mach.
 U.S.A. 79, 3687 (1982).
 E. O. Long, C. T. Wake, J. Gorski, B. Mach, EMBO J. 2, 389 (1983). 12.
- D. J. Mathis, C. O. Benoist, V. G. Williams II, 13.
- M. R. Kanter, H. O. McDevitt, Cell 32, 745 (1983)
- J. McNicholas, M. Steinmetz, T. Hunkapiller, P. Jones, L. Hood, *Science* **218**, 1229 (1982). 14.
- M. Steinmetz et al., Nature (London) 300, 35 15.
- 16. M. Kronenberg, M. M. Davis, P. W. Early, L.

E. Hood, J. D. Watson, J. Exp. Med. 152, 1745 (1980)

- 17. The fifth exon sequence has been demonstrated in the cDNA sequence of the d-haplotype allele of A β [E. Choi, K. McIntyre, R. N. Germain, J. G. Seidman, *Science* 221, 283 (1983)]. 18. J. R. Bono and J. L. Strominger, *Nature (Lon-*
- don) 299, 836 (1982). 19. Feinstein, Quant. Rev.
- D. Beale and A. Biophys. 9, 135 (1976) 20. Sanger, A. R. Coulson, B. G. Barrel, A. J. H.
- F. Sanger, A. K. Colusin, B. O. Barlet, A. J. H.
 Smith, B. A. Roc, J. Mol. Biol. 143, 161 (1980).
 J. Messing and J. Vieira, Gene 19, 269 (1982).
 S. Anderson, Nucleic Acids Res. 9, 3015 (1981).
 A. M. Maxam and W. Gilbert, Methods Enzymol. 65, 499 (1980).
- 22
- 23.
- 24. R. Parnes and J. G. Seidman, Cell 29, 661
- (1982)K. W. Moore, B. T. Sher, Y. H. Sun, K. A.
 Eakle, L. Hood, *Science* 215, 679 (1982).
 D. Larhammar *et al.*, *Cell*, in press.
 We thank M. Steinmetz for his help during the 25.
- 27 initiation of this work, K. Minard for prep tion of the sequencing gels, and B. Larsh and S Olive for preparation of the manuscript. Supported in part by NIH grants (L.H.) and by the Fondation pour la Recherche Médicale (M.M.).

16 June 1983

Pancreatic Islet Allograft Prolongation by Donor-Specific **Blood Transfusions Treated with Ultraviolet Irradiation**

Abstract. Survival of allografts of islets of Langerhans in nonimmunosuppressed adult rats was prolonged by transfusions of donor blood irradiated with ultraviolet light before transplantation across a major histocompatibility barrier. This treatment is donor blood-specific and has produced greater than 160-day survival of transplanted islets without the administration of immunosuppressive agents.

The successful transplantation of allogeneic pancreatic islets in mice depleted of cells bearing I region-associated antigens (Ia) by antiserum to Ia and the enhancement of skin allografts with this antiserum (1) suggest that allografted tissue depleted of Ia-bearing cells is accepted without being recognized as foreign. The type of Ia-bearing cell eliminated in such experiments is not known, but appears to be the dendritic cell; Ia-bearing dendritic cells are present in frozen tissue sections of islets and in the parenchyma of human kidneys, hearts, thyroid glands, and skin (2). Such a wide distribution suggests that the depletion of Iabearing cells from organ allografts may have clinical applicability not only to pancreatic islet transplantation but to transplantation of other organs as well.

Table 1. Effect of UV irradiation on the stimulatory activity of Lewis rat peripheral blood lymphocytes (PBL) in MLR's. Values are means \pm standard deviations.

Re- spon- der	Stimulator	[³ H]thymidine incorporation (count/min)		
ACI	ACI	465 ± 153		
ACI	Lewis PBL	5371 ± 543		
ACI	Lewis PBL and	772 ± 102		
	UV irradiation			
	for 20 minutes			

Although abolition of the initial recognition of a foreign allograft by the host is critical to successful allografting without further immunosuppression, maintenance of a functioning allograft may depend on the initiation of donor-specific suppressor T lymphocytes in the host (3). Such a state of unresponsiveness to allogeneic tissue is seen when Ia-negative platelets and red blood cells are unable to provoke a primary immune response and attenuate the subsequent challenge with Ia-bearing cells (3).

This idea was further supported when treatment of diabetic mice with donor blood depleted of Ia-bearing cells allowed successful transplantation of fresh, untreated allogeneic islets of the blood donor strain (4). It appears, therefore, that immunization with Ia-negative donor blood cells induces immunological unresponsiveness to the donor strain in recipients by the stimulation of specific suppressor cells.

Our demonstration of a rapid and simple method of inducing donor-specific immunological unresponsiveness in adult animals that allows long-term survival of islet allografts is consistent with recent clinical studies in which donorspecific transfusions led to 1-year survival of kidney allografts in more than 90 percent of mismatched donor-recipient pairs of one haplotype (5). Since ultraviolet (UV) irradiation of the stimulating cell population in a primary mixed-lymphocyte reaction (MLR) leads to little or no proliferative response (6), we hypothesized that Ia-bearing cells may not need to be eliminated from blood before its use for immunization, but may need to be inactivated with UV light, leading to abrogation of the stimulating allogeneic signal while leaving major histocompatibility complex antigens intact for the induction of donor-specific immunological unresponsiveness.

Rats of strain ACI (RTl^a) were made diabetic with intravenous streptozotocin (60 mg/kg). A rat was used as a recipient of blood and islets only if its blood glucose concentration exceeded 300 mg/ dl for more than 3 weeks. Islet allografts were considered to have been rejected when plasma glucose was greater than 200 mg/dl on two successive daily measurements.

Whole blood was obtained from normal Lewis rats (RTl¹) by intracardiac puncture. The blood was diluted 1:50 in phosphate-buffered saline (PBS), placed with a magnetic stirring bar into 250-ml petri dishes, and irradiated for 20 minutes with two Sylvania FS-20 lamps located 10 cm from the dishes. The blood cells were then centrifuged and the resulting pellet was resuspended in PBS to 50 percent packed cell volume. Each diabetic ACI rat received 1 ml of UVirradiated blood or 1 ml of identically treated nonirradiated blood adjusted to 50 percent packed cell volume through the penile vein 3 weeks, 2 weeks, and 1 week before islet transplantation. One group of diabetic ACI rats received islets without previous transfusions.

Pancreatic islets were harvested from Lewis (RTl¹) and Wistar Furth (WF) (RTl^u) rats by the collagenase technique (7) and Ficoll gradient separation (8), with subsequent handpicking under a dissecting microscope. Some 1200 to 1500 freshly prepared allogeneic islets

Table 2. Effect of UV light on the serological reactivity of Lewis rat PBL surface antigens. Values are mean counts (\pm standard deviations) of ¹²⁵I-labeled staphylococcal protein A bound per assay (background, 200 count/ min).

Antigen	PBL	PBL and UV irradiation (20 minutes)
Rabbit antise- rum to rat lymphocytes	2996 ± 172	3315 ± 434
Monoclonal antibodies to rat Ia (MRC/OX4)	2050 ± 421	1963 ± 268