JCI The Journal of Clinical Investigation

A mesangium-predominant gene, megsin, is a new serpin upregulated in IgA nephropathy.

T Miyata, ..., K Okubo, K Kurokawa

J Clin Invest. 1998;102(4):828-836. https://doi.org/10.1172/JCI2450.

Research Article

Mesangial cells play an important role in maintaining a structure and function of the glomerulus and in the pathogenesis of glomerular diseases. To identify a specific gene expressed in human mesangial cells, we used a rapid large-scale DNA sequencing and computerized data processing to compare the transcripts in cultured human mesangial cells with various different cells and organs. Using this novel approach, we discovered a new mesangium-predominant gene termed "megsin." We obtained a full-length cDNA clone of megsin, which coded for a novel 380-amino acid protein. Amino acid homology search revealed that megsin belonged to the serpin (serine protease inhibitor) superfamily. The amino acid sequences in the reactive loop site of megsin showed characteristic features of functional serpins. Northern blot and reverse-transcribed PCR analyses of various tissues and cells demonstrated that megsin was predominantly expressed in human mesangial cells. In situ hybridization studies showed the megsin expression in the mesangium of normal glomeruli, while it increased in the expanded mesangium of glomeruli from patients with IgA nephropathy with the degree of mesangial proliferation. Here we report a new human mesangium-predominant gene that may function as an inhibitory serpin in normal and abnormal biological processes of glomerulus.



Find the latest version:

https://jci.me/2450/pdf

A Mesangium-predominant Gene, Megsin, Is a New Serpin Upregulated in IgA Nephropathy

Toshio Miyata,* Masaomi Nangaku,* Daisuke Suzuki,* Reiko Inagi,* Ken-ichi Uragami,* Hideto Sakai,* Kousaku Okubo,[‡] and Kiyoshi Kurokawa*

*Molecular and Cellular Nephrology, Institute of Medical Sciences and Department of Medicine, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan; and [‡]Institute for Molecular and Cellular Biology, Osaka University, Suita, Osaka 565, Japan

Abstract

Mesangial cells play an important role in maintaining a structure and function of the glomerulus and in the pathogenesis of glomerular diseases. To identify a specific gene expressed in human mesangial cells, we used a rapid largescale DNA sequencing and computerized data processing to compare the transcripts in cultured human mesangial cells with various different cells and organs. Using this novel approach, we discovered a new mesangium-predominant gene termed "megsin." We obtained a full-length cDNA clone of megsin, which coded for a novel 380-amino acid protein. Amino acid homology search revealed that megsin belonged to the serpin (serine protease inhibitor) superfamily. The amino acid sequences in the reactive loop site of megsin showed characteristic features of functional serpins. Northern blot and reverse-transcribed PCR analyses of various tissues and cells demonstrated that megsin was predominantly expressed in human mesangial cells. In situ hybridization studies showed the megsin expression in the mesangium of normal glomeruli, while it increased in the expanded mesangium of glomeruli from patients with IgA nephropathy with the degree of mesangial proliferation. Here we report a new human mesangium-predominant gene that may function as an inhibitory serpin in normal and abnormal biological processes of glomerulus. (J. Clin. Invest. 1998. 102:828-836.) Key words: 3'-directed cDNA library • cell-specific protein • serpin • mesangial proliferation • glomerulonephritis

Introduction

Sixty trillion (6×10^{13}) cells in the human body share essentially identical genomic DNA. Nonetheless, for normal physiological function distinct for each cell type, expression of genes is tightly regulated by the cell lineage. It is, therefore, critical to uncover which genes are specifically expressed by individual cell types for further understanding of cell biology and body function.

Received for publication 9 December 1997 and accepted in revised form 10 June 1998.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc. 0021-9738/98/08/0828/09 \$2.00 Volume 120, Number 4, August 1998, 828–836 http://www.jci.org Mesangial cells play a central role in maintaining a structure and function of the glomerulus. In the pathophysiology of glomerular diseases, the proliferation of mesangial cells and the accumulation of extracellular mesangial matrix are primary events leading to the progression to glomerulosclerosis in patients with a variety of glomerular disorders, such as chronic glomerulonephritis and diabetic nephropathy, two major causes of end-stage renal failure.

To determine and characterize mesangium-specific genes is, therefore, essential to understand the pathogenesis of glomerular diseases. We used a novel strategy using a rapid largescale DNA sequencing and computerized data processing for this purpose. Here we report cloning and characterization of a new human mesangium-predominant gene, megsin, which was a new member of the serpin superfamily and was upregulated in IgA nephropathy, a form of mesangial proliferative glomerulonephritis.

Methods

Construction of the 3'-directed mesangium cDNA library. Details of construction of the cultured human mesangial cell library have been described elsewhere (1). Briefly, human glomeruli were isolated by differential sievings from a normal human kidney obtained from surgical specimens obtained during nephrectomy in a 58-yr-old male with kidney tumor. Isolated glomeruli were resuspended in medium 199 (GIBCO BRL, Gaithersburg, MD) containing 25 mM Hepes, 10% Nu-serum (Collaborative Biomedical Products, Bedford, MA), and antibiotics (10 mg/ml of penicillin, streptomycin, and fungizone), and were incubated in a 5% CO2 incubator. Mesangial cells at passage 3 were characterized by a series of criteria (1, 2): typical morphologic features, tolerance to trypsin, puromycin and D-valine, positive immunostaining for α -smooth muscle cell actin, and the absence of staining for von Willebrand factor, a marker for endothelial cells, and cytokeratine, a marker for parietal epithelial cells. Total RNA was isolated from human mesangial cells at passage 6 using the standard guanidine isothiocyanate method. PolyA⁺ RNA was isolated by oligo dT-cellulose column (Pharmacia, Tokyo, Japan) according to the manufacturer's protocol.

A vector primer is dam-methylated at the MboI sites (GATC) and has a 3' protrusion of homo dT stretch at one end. The doublestranded cDNA moiety was cleaved with MboI. The vector that was attached with the 3'-cDNA (from polyA to the nearest MboI site) was then circularized by ligation and was transformed into *Escherichia coli*. The resulting transformants were selected randomly and were lysed individually by brief boiling. The lysate was used as a template for PCR amplification. The resulting short double-stranded DNA was used for the cycle-sequencing reaction and was analyzed using an autosequencer.

Cloning of megsin. To identify the 5' end of megsin cDNA, the 5'-Full RACE Core Set (Takara, Tokyo, Japan) was used. The reverse transcriptase $(RT)^1$ reaction was performed with 1.0 µg of hu-

Address correspondence to Toshio Miyata, Molecular and Cellular Nephrology, Institute of Medical Sciences and Department of Medicine, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa 259-11, Japan. Phone: 81-463-93-1936; FAX: 81-463-93-1938; E-mail: t-miyata@is.icc.u-tokai.ac.jp

^{1.} *Abbreviations used in this paper:* IgA-N, IgA nephropathy; PAI, plasminogen activator inhibitor; RSL, reactive site loop; RT, reverse transcriptase; SCCA, squamous cell carcinoma antigen.

man mesangial cell polyA⁺ RNA with a gene-specific oligonucleotide, 5'-pTCAGAGAGGTCATTC. After degradation of the template mRNA with RNase H, the first strand cDNA was circulated by T4 RNA ligase and an aliquot was used as a template for the first PCR amplification with gene-specific primers, P1 (5'-AGATTCT-TGAGCTCAGAT) and P2 (5'-TCATTGATGGGTCCTCAA). The resulting PCR product was then used as a template for the second PCR amplification with nested gene specific primers, P3 (5'-AATG-GTGGCATAAACATG) and P4 (5'-ACAGACAAATTGAAC-TTC). The PCR products were cloned into the pCRII vector (Clontech, Palo Alto, CA) and DNA sequences were determined by dideoxy chain termination reaction using a DNA autosequencer.

Expression of megsin fusion protein. Expression of megsin fusion protein was performed as follows. The entire coding sequence of megsin cDNA was obtained by RT-PCR. Single-strand cDNA was prepared from human mesangial cell polyA⁺ RNA using oligo (dT) primer and RT. The resulting first-strand cDNA was amplified using specific primers from the 5' region and the 3' region of the coding sequence, P5 (5'-GTCGACTTATCAAGGGCAAGAAACTTTGCC) and P6 (5'-GAATTCCATGGCCTCCCTTGCTGCAGCAAA) containing EcoRI and SalI recognition sequences at their 5' end each. Amplified megsin DNA was cloned originally into pCRII vector (Clontech) for nucleotide sequencing, and then ligated into the maltose-binding fusion protein vector pMALc2 (New England Biolabs, Beverly, MA) using EcoRI and SalI site. Fusion protein was expressed in E. coli (DH5a), affinity purified according to the manufacturer's instructions (New England Biolabs), and dissolved in 50 mM Tris-HCl, pH 9.0.

The fusion protein was then reacted with Endo Lys-C (Acromobacter Protease I, Wako Pure Chemicals, Osaka, Japan) (enzyme/ substrate ratio = 1:50, W/W) at 37°C for 8 h. The digested peptides were separated by high performance liquid chromatography (model LC-10A; Shimadzu, Kyoto, Japan) on a C18 reverse-phase column (4.6 × 250 nm; Waters, Tokyo, Japan), detected by their absorbance at 214 nm, collected, and subjected to peptide sequencing on an automatic protein sequencer (Model 491 gas-phase sequencer; Perkin-Elmer, Chiba, Japan) or to SDS-PAGE using a 4–20% gradient acrylamide gel and Coomassie brilliant blue staining.

Analysis of the amino acid sequence. Motif search was performed using a Motif Finder program written by A. Ogiwara (Institute for Chemical Research, Kyoto University, Kyoto, Japan). This program searches PROSITE (3, 4) and MotifDic (5). To find a signal sequence, we applied PSORT (K. Nakai, National Institute for Basic Biology, Japan), a program to predict protein localization sites. Sequence homology search was performed using the FASTA program (6). Sequence alignment and pustell matrix comparison were performed according to the protein-scoring matrix pam250 with MacVector (International Biotechnologies, Inc., New Haven, CT). The Kyte-Doolittle scale (7) was used to investigate a hydrophilicity profile with MacVector. Secondary structure prediction was performed by PHD (Profile fed neural network system from Heidelberg) sec, according to the Rost and Sander secondary structure prediction method (8).

Northern blot analysis. RNA of cultured human mesangial cells and fibroblasts was separated on a 1% agarose gel containing 2.2 M formamide and transferred to a nitrocellulose filter. The filter was hybridized in a Rapid Hyb solution (Amersham Corp., Arlington Heights, IL). Blot was washed to a final stringency of $0.1 \times SSPE/$ 0.1% SDS at 60°C.

A human multiple tissue Northern blot and a human cancer cell line Northern blot were purchased from Clontech. A human multiple tissue northern blot contains 2 μ g of polyA⁺ RNA from heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. A human cancer cell line Northern blot has 2 μ g of polyA⁺ RNA from promyelocytic leukemia (HL-60), HeLa cell (S3), chronic myelogenous leukemia (K-562), lymphoblastic leukemia (MOLT-4), Burkitts' lymphoma (Raji), colorectal adenocarcinoma (SW480), lung carcinoma (A549), and melanoma (G361). Hybridization and washing was performed as described above. *RT-PCR.* Total RNA was isolated from monocytes, polymorphonuclear leukocytes, cultured human mesangial cells, and other human primary cultured cells (vascular smooth muscle cells, fibroblasts, endothelial cells, renal epithelial cells, and keratinocytes) purchased from Takara (Tokyo, Japan) by guanidinium thiocynate-phenol-chloroform extraction. The RNA was reverse transcribed using a T-Primed-First-Strand kit (Pharmacia) according to the manufacturer's instructions.

PCR amplification was performed in 25 μ l reaction (with 1 ml of the reverse-transcribed products) containing 0.625 U of ExTaq (Takara), 200 mM each of the four deoxynucleoside triphosphates, 20 pmol of sense primer (5' end), and 20 pmol of antisense primer (3' end). The oligonucleotide primers for megsin sequences were: sense, 5'-ATGATCTCAGCATTGTGAATG-3' and antisense, 5'-ACT-GAGGGAGTTGCTTTTCTAC-3'. The expected size of the amplified fragment was 773 bp. The β -actin served as an internal RNA control to allow comparison of RNA levels among different specimens. Specimens were amplified in a DNA Thermal Cycler (Perkin Elmer Cetus). Each cycle consisted of incubation for denaturation at 94°C for 1 min, annealing at 60°C for 2 min, and extension at 72°C for 2 min. The PCR products were resolved by electrophoresis in a 1% agarose gel and visualized by ethidium bromide staining.

In situ hybridization. In situ hybridization was performed as described previously (9). Nucleotide sequences from 391 to 428 of human megsin cDNA (5'-AGCGTGGACTGAGGGAGTTGCTTT-TCTACAATATTACT-3') were used as a probe. To evaluate the specificity of the signals, RNase pretreatment or a competitive study with 100-fold excess amount of homologous or unrelated, unlabeled oligonucleotides was performed. To quantitate the megsin mRNA expression, all nuclei and nuclei with surrounding positive cytoplasm in at least 10 randomly selected glomeruli (cross-section through their vascular poles) were blindly enumerated, and results were expressed as percentage of positive cells of total nuclei. Mann-Whitney U test was used for the statistical comparisons.

Results

Identification of a new human mesangium-specific gene. To identify specific gene(s) expressed in mesangial cells, we used a rapid large-scale DNA sequencing and computerized data processing to compare the transcripts in various different cells and organs (1, 10, 11). A large scale sequencing of a 3'-directed cDNA library of cultured human mesangial cells was performed to determine partial sequences of 1836 randomly picked clones. The sequence similarities of the clones were compared with each other and with the DNA databank Gen-Bank using the FASTA program (6). We selected six clones that had not been recorded in any other databases and were thought to be predominantly expressed in mesangial cells. Dot blot analysis of mRNAs from various organs and cells supported mesangium-predominant expression of these clones (data not shown). One clone was obtained, which was most abundant among those detected only in our human mesangial cell cDNA library. This clone consisted of 0.3% of all the mRNA population.

The new gene, megsin, is a new member of the serpin superfamily. Using PCR techniques, a full length cDNA of this gene was obtained (Fig. 1). The nucleotide sequencing analysis revealed an open reading frame of 1140 nucleotides coding for a predicted protein of 380 amino acids with four potential *N*-glycosylation sites (the deduced molecular weight, 43 kD). The sequence around the initiation codon (GCA<u>ATGG</u>) was consistent with Kozak's consensus sequence (A/GCC<u>ATGG</u>). Stop codons were in all three reading frames at the upstream of the initiation codon.

61 TTOTCAGTGAGAAGATCACCATAGGAAACCAATTCCCATCATTCTTCTGGCATCAGTGAGCGCATCACTCCTCTCTTTTTTTT	1	GTAC	AGGG	AAGCT	CTCC	ГТСАТ	CACCI	TCCT	AAGTO	CATG	GGGGZ	AAAT	ACCTA	GGGC	FCAAC	AGTC	TGAG	AAGTO	TGGA	AACAI	TTTC	80
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$																						180
by D N O G N G N G N V F F S S L S L F A L L 241 GAT GAC AAT CAA GGA AAT GGA AAT GGA TT TC TT TCC TCT CG AG CTC TTC GCT GCC CTG A L V R L G A Q D S L S Q I D K L L H 301 GC CTG GTC GGC TTG GGC GCT CAA GAT GAC TCC CTC TCT CAG AAT CAG TTA GCC TC AT 360 V N T A S G V G N ⁴ S S N S Q S G L Q S Q 1 C R R V F S D I N ⁴ A S H K D Y D L S I V 421 CTG AAA AGA GTT TT TCT GAT ATA AAT GGA AAC TCT TCT AAT AGT CAG GCA GG CTC CAG TT CAA 420 L R R V F S D I N ⁴ A S H K D Y D L S I V 421 CTG AAA AGA GTT TT TCT GAT ATA AAT GGC TAT CAT AAG GAC TCC AGA GT GAT CAC TC AGC ATT GTG AAC ACT GCC TAG GAA AAA GTG TAT GGA TAT GGC TTT CAT AAG GAC TAC ATT GGA TAT CTT AGA ATT GGA GAC ACT AGA N G L F A E K V Y G F H K D Y I E C A E 1 AAA TTA TAC GAT GCC AAA GTG GAG CG GAT GAC ACT TA CAA AAT CAT TTA GAA GAC ACT AGA K L Y D A K V E R R V D F T N H L E D T R 600 TAT ATTA ATA GAT GCC TAT GGT AAT GAA TGA ATT GAA ATC CAT TTA GGA GAC ACT AGA 601 CT ATA TAT ATA TAA TGG GT GAG GAT GAC CAT AGA AATC CAT TTA GGA GAC ACT AGA 661 GAA GGT GGC AT AGC CTCA CTC AGT GAT ATG GAT GCT GTG GTG ATA CAT TTA GAA GCC GTA TAG 70 L TA TAT AAC GG TG GCT CAA AT GAA ACC ATA AAT TAC AAT TAC GAT GAC CAT AGA 71 AGG GG CAA TAC GCC CTC ACC AAG AGC GAA ACC ATA AAT TCC AGA ATC TC CAA AGC 72 K W Q S A F T K S S S A V M V L V N A V Y F K G 721 AAG GGG GCA TA GC CC TCC ACC AGA AGC GAA ACC ATA AAT TGC CAT TTC CAA ATC TCC CAA 780 C S G K A V A M M H Q C R K F M L S V I 741 TGC TCT GGG AAG GCA GTC GCC TTC AGA AAC CATA AAT TGC CAT TTA CAA TGT CCC GAT 780 751 GGA GAC CCA ACA GCC TTC ACC AGA GAC CATA AAT GGA GCA ACT AAA GCT TA CG TT 760 761 GAA GGA GAC ATT GAC GCC ATG GAC CTC TAG GA AAC CATA AAT CGA CTC ACA GG TGA CTA ACG TT 760 961 GAA TGG ACC CAT ACG AGG CGA ATG ACC TCT TGGA ATA CAA TT TT TTT CCT CAG AT TCT TAC GA AT CCA ATG GAT TT TTT CTT CCT CAG AT ACC 76 M M H K K S Y I L E V T E N K Y V E V F F P Q F 961 GAA TGG ACC AAT GCA AT GAA CTC TTG GGA TAC CAT TTT TTT CCT CAG ATA TTC 771 TT TCT TGA AAT GAA GAA ATG AC CTC TGT GAA AAC AATA CTC TTT TTT CTT CAG GAA			M	А	S	_L	A	A	A	N	A	E	F		F	N	L	F	R	Е	М	
241 GAT GAA GAA GAA AT GGA AAT GGA AAT GGA TT TGT TT TCC TTT TCC TCT GAC CTC TTT GCC GCC CTG GCC CTG A L V R L G A Q D S L S Q I D K L L H 300 GCC CTG GTC CGT GT GG CT CAA GAT GAC TCC CTC TCT CAG ATT GAT AMG TTG CTT CAT V N T A S G Y G N S S N S Q S G L Q S Q 161 GT AAC ACT GCC TCA GGA TAT GGA AAC TCT TCT AAT AGT CAT CAA GTG CC CGA TTC CAA 172 V N T A S G Y G N S S N S Q S G L Q S Q 161 GT AAC ACT GCC TCA GGA TAT GGA AAC TCT TCT AAT AGT CAC CAG GC CC GA TC CAA 162 V N T A S G Y G N S S N S Q S G L Q S Q 163 GT AAC ACT GCT CAA GAA AAC TCT TCT AAT AGT CAC TAG AT GAC TCA GC ATT GGA TTA GA CAT GAC ATT GGA ATT AT GAA CAT GAC ATT GGA ATT AT GAA CAT GCA TTT TG CT GAA AAA ATT GAA ATT GAA CAT CAA TT GAG GT GT GCC GA 1641 AAT GGG CTT TT GCT GAA AAA ATT GGA TTT ACCA AAT GAA CAT CAA TT GAG GT GCC GAA 1641 AAT TAA C GAT GCC AAA GTG GAG GGA GTT GAC TTT ACG AAT CAT TAG GA CAC TAG GAG AC ACT AG 1641 AAT TAA C GAT GCC AAA GTG GAG GGA GTT GAC TTT AGA AAT CAT TTA GAA GAC ACT AGA 1641 AAT TAA TAAC GAT GCC AAA GTG GAG GTG GAC CAT AAAT GAA ATC CAT TAG AAG AAC GTG ATT GGT 1660 C CT AAT ATT AAT AAG TGG GTT GAA ATT GAA ATC GAT TTA GGA AAT CAT TAG AAG AAC GTA ATT GGT 1660 C CT AAT ATT AAT AAG TGG GTT GAA ATT GAA ATC GAT CAT GAG AAT CAT TAG AAG AAC GTA ATT GGT 1660 C C TAAT ATT AAT AAG TGG CTC ACT CAG AAG GTG AAC CAT AAAT TCC CAT TAA GAA GAC ATT AGA 1711 AAG GGC CAT AGC CA AC CAT CAG CAT GAA ACC AAT AAAT TCC CAT TAC AAA CAT GAT CTC AAA GC 1720 C S G C A TA AGC CAT CGC GCC ATG ATG CTC TAG AAT GTG GGG CAT AAC ATT TAT CAA ATC GT 720 1731 TGC TCT AGA AAT GGC CTT CAC AAG GAG GAC ACC ATA AAAT TCC ATT TCC AAA AC ATT TAT 1741 AAG GGA CCA TA AGC CAT CG CC CTG AGA GTA CCA TA AAA ATG GAC CTT TCC AAA CAC ATT ATT 1741 AAG GGA CCA AA GC ACTC CTC TGT GAA TT TT TCC AAAT GAG TT CTT CAA ATG 1741 AAG GGA CCA AA CCA CTC CTC TGT GAA TAT AAA CAG ATT TT TCC AAAT AC ATT TAT 1741 AAG GGA CCA CA AG ATT CTT GAA ATT ATT AAA CAA TTT TTT CCA TAT GAG ATT ATT CAA 1741 AAG GGA CCA CA AAG ATT CTT GAA ATC CTT TGAA AAC GGA GCC TTT CAAAA AAC ATG TAC	181	GCA	ATG	GCC	TCC	CTT	GCT	GCA	GCA	AAT	GCA	GAG	TTT	TGC	TTC	AAC	CTG	TTC	AGA	GAG	ATG	240
241 GAT GAC GAT CA GA AT GA AAT GA AAT GA AAT GA GAT GT TT TC TT TC TCT CTG AG CTC TT CG CT GCC CTG 300 A L V R L G A Q D S S L S Q I D K L L H 360 V N T A S G Y G A^{*} G C CC CT TC CA A AT GA AAG TG CT CAT 360 V N T A S G Y G A^{*} S S N S Q S Q L Q S Q L K R V F S D I A^{*} A S S N S Q S Q L S Q I Q L S I V L K R V F S D I N^{*} A S H K D Y D L S I V V N G L F A E K V Y G F H K D Y D L S I VN G L F A E K V Y G F H K D Y I E C A A G G A A A A A G A		D	D	N	0	G	N	G	N	v	F	F	S	S	L	S	L	F	A	A	L	
301 GCC CTG GTC CGC TTG GCC GCT CAA GAT GAC TCC CTC TCT CAG ATT GAT AAG TTG CTT CAT 360 V N T A S G Y G N [*] S S N S Q S G L Q S Q L N R V F S D I N [*] A S N K D Y D L S I V 421 CTG AAA GAG GTT TTT TCT GAT ATA ATT GA CAC CAC TAG GAT TGA GTG CTC CAG TCT CAA N G L F A E K V Y G F H K D Y I L S I V 421 CTG GG CTT TTT GCT GAT ATA ATA TGA CAC CCAC AAG GAT TAT GAT CTC AGC ATT GTG 480 N G L F A E K V Y G F H K D Y I E C A E S 481 AAT GGG CTT TTT GCT GAA ATA GAA TGA GTG GAC TAC ATT GAG TGT GCC GAA K L Y D A K V E R V D F T N H L E D T R 541 AAA TTA TAC GAT GCC AAA GTG GAG CGA GTT GAC TTT ACG AAT CAT TA GAA GCA CA AGA F N I N K W V E N P T H G K I K N V I G 661 CGT AAT ATT AAT AAA GTG GG TT GAA ATT GAA AAT GAA ACC ATA GGC AAA ATC GAG TGT GAT ATA GAG CAC AAG K W Q S A F T K S S A V M V L V N A A V Y F K G 661 GAA GGT GGC ATA AGC TCA TCT GCT GTA ATA GAA CAC ATA GGT GAA GTG GTG ATA GCT CAT AAG ACC TCT GGG AAT ACG CCA TC TC GCT GTA ATA GAG CAA AAT GGA CAT AAA TTGC CAT TTC AAA ACC GAT GGT ATT 780 C S G K A V A M M M H Q E R K F N [*] L S V I F T M G GG CAT TCA CAC GCC TTC ACC AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA ATC GCT TAT 840 C S G K A A V A M M M H Q E R K K F N [*] L S V I F D F S M K T L E E L R Y N G G I N M M Y V 841 GAG GAC CCT ACT CAC AGG CCT CT CT GAA ATT GAA CAT ATA AAC ATT ATT GT C TTT ATT 840 1001 CTG CTG CGA GAA GAT CTC TCT GAA ATT GAA CAC ATA AAT TGT GAA GTA ACT TAC GTT ATT 840 1011 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC AAA TGT AAAC ATT TTT CTC CC GAT TAT 840 1021 AAC ATA GAA CAA ATT GAA ATG GAA CAA TAT TTT GAA GAC ATT ATT 1020 AAG ATA GAC AAT CCC AAG CAC CTC TCT GAA ATT GAA AAC AAA CTG ACC TTT C GAA ATC AAT ATG 1020 AAG ATA GAC AAT CCC AAT CCA ATG CAT CTA TAT GAA AAC AAA ATA TTT TTT CCT CAG GAT ACT ATG 1020 1021 L P E N D L S E I I E N K L T F Q N L M 1022 ACC GAA GAT CCA ATA GAA ATT GAA ATG GAA CAA TAT TTT GAA GAC ATA ACC ATT ATT TAT CAA GAA ATA ATT TAT GAA AAC AAA TAT TTT GAA GAC ATA ATT TAT CAA GAA ATA ATT TAT GAA AAC AAT ATT TAT GAA ATT GAA AAG CAAT ATT TTT AC	241	GAT	GAC		CAA	GGA	AAT	GGA	AAT	GTG	TTC	TTT	TCC	TCT	CTG	AGC	CTC	TTC	GCT	GCC	CTG	300
301 GCC CTG GTC CGC TTG GCC CT CA GAT GAT GAC TCC CT TT CAG ATT GAT AAG TTG CTT CAT 360 V N T A S G Y G N ⁴ S S N S Q S G L Q S Q ATT GA AT GA AC CT TCT TA AT AGT CAG TCA GG GC CC CAG TCT CAA L K R V F S D I N ⁴ A S H K D Y L L S I V ATT GG CTT TTT GCT GA AAA GT GG CT CC CA GG GAT TAG GAT CAT GTG GTG GT N G L F A E K V Y G F H K D Y I L C A TT GAT GC GC GAA ATT GG CTT TTT GCT GA AAA GT GG GT GA CTT AT GAT CAT AGG GT GCC GAA K L Y D A K V E R V J G F H K D Y I E C A E 481 AAT GG CTT TTT GC GA AAA GT GG GAT GAA CTT TA CAT AAG GAC TAC ATT GAG GT GCC GAA R N I N K W V E R V D F T N H L E D T R 541 AAA TTA TAC GAT GCC AAA GTG GAG GAT GAA AAT GAA AAC CAT GGC AAA ACC ATA AGG AAC GTG ATT GGT GCC R W Q S A F T K S E T I N C H F K S P K 721 AAG GG CAT ACA GC TCA CT TC GCT GTA AAT GAA ACC CAT GGC AAA ATT GAA GAA CAC TA AGG AG GC AT ACA GC TCA CT GCT GTA ATG GTG CTG GTG GTG AAT GGT GTA CTTC AAA GAC (720 K W Q S A F T K S E T I N C H F K S P K 721 AAG GG CAT CAA GC CCA TC ACG GC GAA ACC ATA AAT TGA AAC CAT AAA TTTC CAT AGG AAC GTG ATT GTT GTT ATT 840 641 GAA GAC GCC TCA ACG GCC TTC ACC AGG AAA CC ATA AAT TGC CAT TTC AAA ACC ATA GGT C S G K A V A M M M H Q E R K F K L S V I F 1 TGC TCT GGA GAA GAT CTC ACC AGG GAA ACC AAT AAT TGG TG GCA TTA ATT 840 641 GAA GAC CCA CAT CAA GAA GAT CTT GAG AAT CAA GAA CC ATA AAT TGG CA GAT AAC ATT TAC GAT TAC F D F S M K T L E E L R Y N G G I N M M Y V 841 GAA GAC ACC ATC AATG AGA TC TT GAA AAT GAA ACC ATA AAT TGG GA GAA AAC ATT ATC F D C S C G C AAT GAC CTC TCT GAA ATT GAA AAC AAT ATT TG CTG TTA TAT 840 901 CTG CTG CCA GAA ATT GAA ATT GAA ATT GAA AAC AAA ATT TTT CCT CAG AAT CTA ATG 900 12 L P E N N D L S E I I E N K L T F Q N L M 901 CTG CTG CCA AAT GAC ATT CTT AGA CAA CTA TAT TTT GAA GAC ATT ATA TAA TAA 1140 7 M M H K S Y I E V T E K V D E V F F P Q F 900 12 L L P E N N D L S G I A S G G R L Y I S 1061 TTT GAT GAA ATT GAA AAT GAA CTC TTA GAA AAT GTT GAG GAA GAC ATT ATT TAA CAA ITT TTT CCT CAG GAA CTA ATT GTT GAA AAT ATT TAG AAA AG AT ATT TTT GAA GAA		Δ	т.	v	R	т.	G	A	0	п	D	s	т.	s	0	Ι	D	к	L	L	н	
V N N T À S G Y G GA NĂ GA NĂ S S N S Q G G L Q S G L Q S Q 361 GT AAC ACT GCC TCA GGA TAT GGA AAC TCT TCT AAT AGT CAG TCA GGG CTC CAG TCT CAA L K R V F S D I N À A S H K D Y D L S V 421 CTG AAA AGA GTT TTT TCT GAT ATA AAT GCA TCC CAC AAG GAT CAT GTG GCC GAT GTG N G L F A E K V Y G F H K D Y I L S C A E 481 AAT GGG CTT TTT GCT GAA AAA GTG TAT GGC TTT CAT AAG GAC TAC ATT GAG GTG GCC GA E K L Y D A K V E R V D F T N H L E D T R 540 K L Y D A K V E R V D F T N H L E D T R 600 CT ATA TAT AAG GGT GAA GTG GAG CGA GTT GAC TTT AGA AAC CAT GAA GAC ACT AGA 601 CTT AAT AAT AAG GGT GAA AAA GTG GAG CGA GTT GAC TTT AGA AAC CAT GTG GAA GAC CAT AGA 601 CTT AAT AAT AAA GTG GGC GTG GAA AAT GAA AC CAT GGC GAA ATC AAG AAC CAT GGC 601 CTT AAT AAT AAG GG CTC ACT TC GCT OTA ATG OTG CTG GTG GAT GCT GTG TAT CTC AAA GGC 720 K W Q S A F T K S S A V M V L V N A V Y F K G 661 GAA GTG GCA ATA AGC TCA TCT GCT OTA ATG OTG CTG GTG GAT GCT GTG TAT TTC AAA GGC 720 K W Q S A F T K S E T I N C C H F K S F K 721 AAG TGG CAA TCA GCC TAC CA GC AAG ACC AAT AAT TGC CAT TTA GAT GTT CAAA GGC 720 K W Q S A F T K S E T I N C C H F K S F K 731 TGC TCT GGG AAG GCA GTC GCC ATG ATG GTT GCT GTG AAT GCT GTG TAT TTC AAA GGC 740 K W Q S A F T K S E T I N G G I N M Y V 841 GAG GAC CCA TCA ACG AAG ATT CTG AAG GCC ATA AAC CAT AGC ATA AAT TGC CAT TTA GAT GTT GTT ATT 840 E D P S M K I L E L R Y N G G I N M Y V 841 GAG GAC CCA TCA ATG AAG AAT ATT GAA ACC AAA ACT AAC TG ACC TTT CCA GAT TTA CAG TT 900 TT GT GCT GCG GAA GCCA CTC CTC TG GAA ATG CATA CTA ATG GT GCC ATA AAC ATG TTA CTT ATT 840 F W T N P R R R M T S K Y V E V F F P Q F 961 GAA TGG ACC ATA CAA AG CAA ATT TG AAA ACA AAA CTA ACC TA ACC TT AGC TTA CTA ATG 960 T C GTG CT GCA GAA ATC CAA AG CAA CCC TC TCT GAG ATT GAT GTT GTT GTT ATT ATA CAA 1141 AGG ATG AAT CCA AG GCA ATG CTC TCT GGG ATT GCT TG GT GC TG TGT ATT ATA CAA 1142 AGG ATG ACT AAA GAA AGA AAA CAA TAT TTG AGA GCC CAA GAC GTA CTT ATT CCA GAT ACC 7 F L F V T R K N D L S G I A S G G R C C C GG GG ATA ACC AAA ACT A CT C GG GG CT A	301						-			-	-		_				-				CAT	360
361 GTT ARC ACT GCC TCA GGA TAT GGA ARC TCT TCT ART AGT CRG GGG CTC CRG TCT CRA 420 L K R V F S D I N ⁴ A S H K D Y D L S I V 480 N G L F A E K V Y G F H K D Y D L S I V 480 N G L F A E K V Y G F H K D Y I E C A E 541 ARA GGC TTT TGT GAT ARA ART GGA TCC CAC ARG GAT TAT GAT CTC AGC ART GTG K L Y D A K V E R V D F T N K L E D T R 541 ARA TTA TAC GAT GCC ARA GTG GGA GGA GGT GAT GAC TTA AGA ACC ACT AGA F N I N K W V E N E T H G K I K N V I G 601 CGT ART ATT ART AGC GGC TTC GCC ARA GGA GGA GTG GGC GA ATC AGA ACC ACT AGA 600 C A A GG GGC ATA AGC GGA GTG GAC GGA GTT GAA ART CAA GAC CTA GAT GGT 601 CGT ART ATT ART AGG GG GT TG ARA ART GAA ACC CAT GGC ARA ACC ACT AGA 600 C A A GG GGC ATA AGC GGA GTG GAA ACC CAT GGT GGT GAT AGC TCC ATA GAT GGT 601 CGT ART ATT ART AGG GG GTT GAA ART GAA ACC CAT GGC GG GAA ATC ARG ARC CAC AGA 602 K W Q S A F T K S S E T I N C H F K S P K 721 ARG TGG CAA TCA GCC TTC ACC ARG AGC GAA ACC ATA ART TGC CAT TCC ARA TCT CCA ARG 720 C S G K A V A M M H Q Q E R K F N ⁴ L S V I 731 TGC CTC GGG ARA GAT CCC CCC ATC AGT GAT CAG GAA ACC ARA ART GGC GCA TTA AAC ATG CAC ARG 740 C S G K A V A M M H Q Q E R K F N ⁴ L S V I 741 AGG CCC TCA ATG AGA CTC CCC ATG AGT CAG AGA CAC ARA GGT GGC ATT AAC ATG CC GTT ATG 750 C TCT GGG AA GAG CCC CTC CAC ATG AGT CAC AGA ACC AATA AAT TGC CAT TA AAC ATG CT GTG TAC 760 C TG GC CAA TCA AGA ATC TCA GC CC TCA GAG CAC AAGA TGC TT C ARA TTG TC GTT ATT 761 TGC CTG AGA ART GAC CTC CTC AGA AGA CT CAG ATA CAT AGG GG C TTT C ARA ATC AGG 960 C L L P E N N D L S E T E N K L T F Q N L M 901 CTG CTG CC AATG GAA ACC CTC TGA ATT AGA TGC CTT GAG CTC AGA ACC ATA ACC TT CAG ATT AAA AC TTA ACA 700 C L L P E N N D L S E T E N K L L T F Q N L M 901 CTG CTG CC AATG GAA ACC CTC TC GAA ATA TTG GAA GC AGG CT CTT AGA GCT CTG GG GG CT CTT CAG ATA TTA TTA ATA TCA AAA ATG AGA GAA CCC TTT CAA ATA TTG AAA ACA TTT TAT TTT TTT CCC CAG GT CT TT TTT TTT CCC CAG GT CTT TA CAA ATA TTG GAA ACC AATA CCC AGG GCT CTT CTA CAA ATA CAA TT TGA GAA GCC CT TAGA GCT CAA TT TGA GAA CCC CCC AT	201							++-			-											••••
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		-		-								-									-	420
421 CTG AAA AGA GTT TTT TCT GAT ATA AAT GCA TCC CAC AAG GAT TAT GAT CTC AGC ATT GTG 480 N G L F A E K V Y G F H K D Y I E C A E AT GGC CTT TG GT GAA AAA GTG TAT GGC TTT CAT AAG GAC TAC ATT GAG GTG GCC GAA K L Y D A K V E R V D F T N H L E D T R 541 AAA TTA TAC GAT GCC AAA GTG GAG CGA GTG GCC GTT CAT AAG GAC TAC ATT GAG GTG CC GAA F N I N K W V E N E T H G K I K N V I G F N I N K W V E N E T H G K I K N V I G E G G I S S S A V M V L V N A V Y F K G 661 GAA GGT GGC ATA AGC TCT GTG AT GAG GTG GTG GTG GTG GTG GTG GTG	301						·															-20
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$					-	_		-	_			-			-	_	-	_	-	-	-	400
481 AAT GGG CTT TTT GCT GAA AXA GTG TAT GGC TTT CAT AAG GAC TAC ATT GAG TGT GCC GAA K L Y D A K V E R V D F T N H L L E D T R 540 K L Y D A K V E R V D F T N H L L E D T R 540 K L Y D A K V E R F V D F T N H L L E D T R 540 R N I N K W V E N [*] E T H G K I K N V I G 601 CGT AAT ATT AAT AAG TGG GTT GAA ATT GAA CAC CAT GGC AAA ATC GAT TTG GAA GAC ACT AGA 602 G G G I S S S A V M V L V N A V Y F K G 661 GAA GGT GGC ATA AGC TCA TCT GCT GTA ATG GTG GTG AAT GCT GTG TAC TTC AAA GGC 720 K W Q S A F T K S S L I N C V N A V Y F K G 721 AAG TGG CAA TCA GCC TTC ACC AAG ACC GAA ACC ATA AAT TGC GTG GTA ATT GTC TAC TTC AAA GGC 720 K W Q S A F T K S E T I N C H F K S F K 721 AAG TGG CAA TCA GCC TTC ACC AAG ACC GAA ACC ATA AAT TGC GTG TAC TTC AAA GGC 720 C S G K A V A M M H Q E R K F N [*] L S V I 840 E D F S M K I L E L R Y N G G I N M Y V 841 GAG GAC CCA TCA ATG AAG CTC TCT GGA ATT GAA CTA CAG GAA CCA TA AAT TGC GAT AAC TAC TAC TAT T 840 901 CTG CTG GTG GAAT GAA GAC CCC TTC GAA ATT GAA AAT CAA CAA CAA CAA TAT TTF CCT GAT ATT 900 E W T N F R R M T S E I E N K L T F Q N L M 901 CTG CTG CTG AGA AAT GAA ATG GAA CTC TTG AGA ATT GAT GAT GAT GAT CTA ATG 960 F D E S K A Q Y E M K Q Y L R A L G L K D I 1021 AAG ATA GGA AGA GAA TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTA TAA ATG TAA TTC 1040 F D E S K A D L S G I A S G G R R L Y I S 1081 TT GAT GAA CCC AAT GCA GGA GAA TCC TCT GGA ATT GTG GGG CTG TCT ATA TTC 1040 R M M H K S Y I E W T <u>E G G T E A T A A</u> 1141 AGG ATG CGC AAA TTG TAT AAG GAA CCC CTC CG GG GGT CCC GAG GGCT AAA GAT ATC 1260 F L F L F V T R R K M D L S G I A S G G R C L Y I S 1081 TTT GAT GAA ATC GAA ATG GAA CCC TCT GG GG GGT CCT GGG CTAT ATA TTCA 1140 R M M H K S Y I E V T <u>E E G T E A T A A</u> 1141 AGG ATG ATG CCA ATT TTG TA ATA GAA AGA AAA CAA ATT TTG AGG GGC ACC GAG GCT ACT GCC GCC CCC 1260 CT TCA ATT GTT ATA CAA AGG GAA CTC CCC CCC CAG GGC CCC CCC AGG CCC CCC CC	421	CTG	AAA	AGA	GTT	TTT	TCT	GAT	ATA	AAT	GCA	TCC		AAG	GAT							480
K L Y D A K V E R V D F T M GO GAT GAG CGA AT GAT GAT CAT TA GAA GAC ACT AGA K L Y D A K V E R E R V D F T N H L E D T R 541 AAA TTA TAC GAT GCC AAA GTG GAG CGA GT GAC TTT ACG AAT CAT TTA GAA GAC ACT AGA R N I N K W V E N E N E T H G K I K N V I G 601 CGT AAT ATT AAT AAG TGG GT GAA AAT GAA ACA CAT GGC AAA ATC AAG GAC GTG ATT GGT 601 CGT AAT ATT AAT AAG TGG GT GAA AAT GAA ACA CAT GGC AAA ATC AAG AAC GTG ATT GGT 601 CGT AAT ATT AAT AAG TGG GTT GAA AAT GAA ACA CAT GGC AAA ATC AAG AAC GTG ATT GGT 601 CGT AAT ATT AAT AAG TGG GTT GAA AAT GAA ACA CAT GGC AAA ATC AAG AAC GTG ATT GGT 601 CGT AAT ATT AAT AAG GCC GT GTA AATG GTG GTG GTG AAT GCT GTG TAC TTC AAA GGC 720 K W Q S A F T K S S E T I N C H F K S P K 721 AAG TGG GAA TCA GCC TTC ACC AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA TCT CCC AAG C S G K A V A M M H Q E R K F N L S V I 781 TOC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CAG AAG TTC AAT TTG TCT GTT ATT 840 E D P S M K I L E L R Y N G G I N M Y V 841 GAG GAC CCA TCA ATG AAG ATT CTT GAA CTC AGA TAC AAT GGT GGC ATA AAC ATG TAC GTT 900 L L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CCC AGA AGC CAT AAG CCC TTA AAG TAG ACC TTT CAA ACG ATG CTA ATG 960 C GAT GGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GAT GAG CTA TTT TTT CCC CAG TTC 1020 K I E K N Y E M K T S K Y V E V F F F Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG AAC CAT TAT TTG AGA GCC TTT CAG AT CTT ATG 1020 K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAA AAT CA AAG CAA TAT TTG AGA GCC CTA GGG CTG TAT ATA TCC 1140 R M M H K K S Y I E V T E E G T E A T A A 1141 AGG AGA ATC CAAA GCA GAT CTT CTT CTT GG GAT CAC CGC CG GGG CTCT GTAT ATA TCC 1260 P S LO P 1201 CT GA AAATCCAATGTTACTTAAGA AGG CAC CCC CC GGG GCT CGT CTG TAT ATA TCC 1260 P S LO P 1201 ACA GGA AGA TAT TG GAA AGG CAA CCC CCC CCAG GCC GAG GCC ACC GCC G		N	G	L	F	A	Е	ĸ	v	Y	G	F	н	ĸ	D	Y	I	Е	С		_	
541 AAA TTA TAC GAT GCC AAA GTG GAG CGA GTT GAC TTT ACG AAT CAT TTA GAA GAC ACT AGA R N I N K W V E N E T H G K I K N V I G 601 CGT AAT ATT AAT AAG CG AT CGA AAT (GAA AAC CAT GGC AAA ATC AAG AAC CGT AGA E G G I S S S A V M V L V N A V Y F K G 661 GAA GGT GGC ATA AGC TCA TCT GCT GTA ATG GTG GTG GTG AAT GTT GTG GTG TAT ATT GGC K W Q S A F T K S E T I N C H F K S S P K 721 AAG TGG CAA TCA GCC TTC ACC AAG AGC GAA ACC ATA AAT GGT GTG TAC TTC AAA TCT CCC AAG C S G K A V A M M H Q E R K F N L S V I 781 TGC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CCG ATA TTC CAA AT TTC CCC AAG 780 C S G K A V A M M H Q E R K F N L S V I 781 TGC TCT GGG AAG GCA GCC GCC ATG ATG CAT CAG GAA CCG ATA AAT TTG CCT GTT ATT 840 E D F S M K I L E E L R Y N G G I N M Y V 841 GAG GAC CCA TCA AAG AAT CTT GAC CTC TGT GAA ATG GGC CAT AAC ATG TAC GTT 900 C TG CTG CCT GAG AAT GAC CTC TCT GAA ATG GAA CCA ATA GTG GCC ATA AAC ATG TAC GT 900 C TG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC GAA CTG AC ATG TAC CT ATG 840 F D F S M K I L E E N K L T F P Q N L M 901 CTG CTG CCC ATG ATG ACC TCT TG GAA ATG GAA CCT TTA GAA CTG ATA CTA ATG 960 F W T N P R R M T S K Y V E V F F F P Q F 961 GAA TGG ACC ATA CAA GGC GA ATG ACC TCT TAG GAA ATG TAT CA AAT GT 1020 AAG ATA GAG AAT CCA AG GCG ATG AAC CTC TAT AGT ATA TTG GAG GCT ATT TTT CCT CAG TTC 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG TAT ATA CA 1140 R M M H K S Y I E V T E E V T E E A T A A 1141 AGG ATG ATG CAA ATG TTA CATA GAG GAT CCTC CG GGG GGG CGC GCC GCC GCC CTC GGC C 1200 F D E S K A D L S G I A S G G R E Y I S 1201 ACA GGA AGT ATA TT GTA GAA AAG CAA CTC CCT CAG GGG ACC GAC GAC GAC GCC ACT GCC CCC 1200 P Stoop 1325 CCT TCC ATT CTT ATA CAG GAA GGC ACC CCC CCT GAG GGG ACC GAC GAC GAC CAC CAC 1326 1326 AAAATCCAAT ATT GTA CAG AAA GCA CAC CCC CCT CAG GCA CACC CACC	481	AAT	GGG	CTT	TTT	GCT	GAA	ААА	GTG	TAT	GGC	TTT	CAT	AAG	GAC	TAC	ATT	GAG	TGT	GCC	GAA	540
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		к	L	Y	D	A	ĸ	v	E	R	v	D	F	т	N	н	L	Е	D	т	R	
601 CGT AAT ATT AAT AAG TGG GTT GAA AAT GAA ACA CAT GGC AAA ATC AAG AAC GTG ATT GGT 660 E G G I S S S A V M V L V N A V Y F K G 661 GAA GGT GGC ATA AGC TCA TCT GGT GTA ATG GTG CTG GTG GAA ATCC AGA GGC 720 K W Q S A F T K S E T I N C H F K S P K 721 AAG TGG CAA TCA GCC TTC ACC AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA GCC 720 C S G K A V A M M H Q E R K F N ⁶ L S V I 781 TGC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CGG AAG TTC AAT TTG TCT GTT ATT 840 E D P S M K I L E L R Y N G G I N M Y V 841 GAG GAC CAT CA CAA AAT GAC TCT GAG CAA ATT GAA ATG GGC ATA AAC ATG ACC ATT 840 C G G CA CTCA ATG AAG ATT CTT GAG CTC AGG GAA ATC AAA TGG ACC TTT CAG AAT CAC ATG 840 GAG GAC CCA CCA ATG AAG ATT GTA GAG ACC TCT AAG GAA CTG ACC TTT CAG AAT CAC ATG 900 L L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC CTA ACC ATT CAG AAT CTA ATG 960 E W T N P R R R M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GTT GAG GTA TTT TTT CCG GAG ATA CCAA ATG R M M H K S Y I E M K Q Y L R A L G L K K D I 1021 AAG ATA GAA GAA GAT TAT GAA AG AA ATG GAAT CAT GAT GCT TGG GG GG CGT CAT ATA AAA GAA ATA T 1040 R M M H K S Y I E V T I E E G T E A T A A 1141 AGG ATG GAC CAC AAA TCT TAC ATA GAA AAG CAC ACT GCT CGG GG GGT CCT CTG TAT ATA TACA 1142 1141 AGG ATG GAG AAT ATC TAAA AAG CAA CTC CTC CAG GGA GGC CCC GAG GCT ACT GCC GCC 1200 1325 CCT TGA AAA TCT TAC AAA AAG CA ACC ACT CAC CAG GGA GGC CAC CAG GCC ACC GAC GCC CAC 1260 1326 1325 CCT TGA AAATCCAATGGTTTAGTAGAAA AAG CA ACC ACT CAC TTA TAT ACA CAC 1260 1325 CCT TGA AAATCCAATGGTTTAGTAGCAATCCACTCAAGTAGACCACCACACACA	541	AAA	TTA	TAC	GAT	GCC	AAA	GTG	GAG	CGA	GTT	GAC	TTT	ACG	AAT	CAT	TTA	GAA	GAC	ACT	AGA	600
601 CGT AAT ATT AAT AAG TGG GTT GAA AAT GAA ACA CAT GGC AAA ATC AAG AAC GTG ATT GGT 660 E G G I S S S A V M V L V N A V Y F K G 661 GAA GGT GGC ATA AGC TCA TCT GGT GTA ATG GTG CTG GTG GAA ATCC AGA GGC 720 K W Q S A F T K S E T I N C H F K S P K 721 AAG TGG CAA TCA GCC TTC ACC AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA GCC 720 C S G K A V A M M H Q E R K F N ⁶ L S V I 781 TGC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CGG AAG TTC AAT TTG TCT GTT ATT 840 E D P S M K I L E L R Y N G G I N M Y V 841 GAG GAC CAT CA CAA AAT GAC TCT GAG CAA ATT GAA ATG GGC ATA AAC ATG ACC ATT 840 C G G CA CTCA ATG AAG ATT CTT GAG CTC AGG GAA ATC AAA TGG ACC TTT CAG AAT CAC ATG 840 GAG GAC CCA CCA ATG AAG ATT GTA GAG ACC TCT AAG GAA CTG ACC TTT CAG AAT CAC ATG 900 L L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC CTA ACC ATT CAG AAT CTA ATG 960 E W T N P R R R M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GTT GAG GTA TTT TTT CCG GAG ATA CCAA ATG R M M H K S Y I E M K Q Y L R A L G L K K D I 1021 AAG ATA GAA GAA GAT TAT GAA AG AA ATG GAAT CAT GAT GCT TGG GG GG CGT CAT ATA AAA GAA ATA T 1040 R M M H K S Y I E V T I E E G T E A T A A 1141 AGG ATG GAC CAC AAA TCT TAC ATA GAA AAG CAC ACT GCT CGG GG GGT CCT CTG TAT ATA TACA 1142 1141 AGG ATG GAG AAT ATC TAAA AAG CAA CTC CTC CAG GGA GGC CCC GAG GCT ACT GCC GCC 1200 1325 CCT TGA AAA TCT TAC AAA AAG CA ACC ACT CAC CAG GGA GGC CAC CAG GCC ACC GAC GCC CAC 1260 1326 1325 CCT TGA AAATCCAATGGTTTAGTAGAAA AAG CA ACC ACT CAC TTA TAT ACA CAC 1260 1325 CCT TGA AAATCCAATGGTTTAGTAGCAATCCACTCAAGTAGACCACCACACACA		R	N	т	N	ĸ	w	v	E	N*	E	т	н	G	к	I	к	N	v	I	G	
E G G G I S S A V M V L V N A V Y F K G 661 GAA GGT GGC ATA AGC TCA TCT GCT GTA ATG GTG CTG GTG AAT GCT GTG TAC TTC AAA GGC 720 K W Q S A F T K S E T I N C H F K S P K 721 AAG TGG CCA TCA CCC ACA AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA TCT CCC AAG C S G K A V A M M H Q E R K F N ⁴ L S V I 781 TGC TCT GGG AAG GCT GCC ATG ATG CAT CAG GAA CGG AAG TTC AAT TTG TCT GTT ATT 840 E D P S M K I L E L R Y N G G I N M Y V 841 GAG GAC CCA TCA ATG AAG ATT CTT GAG CTC AGA TAC AAT GGT GGC ATA AAC ATG TAC GTT ATT 900 L L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CTG CAG ATG GAC TCT CTG TA ATG AAT GAA AAC AAA CTG ACC TTA CAAT GA F D P S K A V L A M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG GAA TGG ATA CTA TTG AGG TAT TTT TTT CCT CAG ATT CTA ATG 960 F D E S K A D L S E I E N K L T F Q N L M 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGG GTA TTT TTT TCC CAG TC 1020 K I E K N Y E M K Q Y L R A L G G L K D I 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGG GTG CTG TA ATA TC 1080 F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCG AAA GCA GAT CTC TCT GG AAT GCT TTG GGG GGC CTG TAAT ATA TCA 1140 R M M H K K S Y I E V T E G G T E A T A A 1141 AGG ATG ATG CAC AAA TCTA ATA GAA GCA CCC CCT CAG GGG GGC CCT AGG CTA CTG CC 1200 T G S N I V E K O L F O S T L F R A D H 1201 ACC GGA AGT AAT TT GTA GAA AG CAA CCC CCT CAG TTC TAGA GCT GAC ACC 1260 1261 CCA TTC CTA TTT GTT ATC AGA AG CAA CCC CCT CAG TCC AGG CGA AGT TTC TGC 1320 CCT TGA AAATCCAATTGGTTCTTGTTATAGCAGTCCCACCACACACA	601			_				-	_		_	_		-		_			GTG			660
661 GAA GGT GGC ATA AGC TCA TCT GCT GTA ATG GTG CTG GTG GAT GCT GTG TAC TTC AAA GGC 720 K W Q S A F T K S E T I N C H F K S F K 721 AAG TGG CAA TCA GCC TTC ACC AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA CTC CCC AAG C S G K A V A M M H Q E R R K F N L S V I 781 TGC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CGG AAG TTC AAA TCT CCC AAG 780 C S G K A V A M M H Q E R R K F N L S V I 781 TGC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CGG AAG TTC AAA TCT GTC GTT ATT 840 841 GAG GAC CCA TCA ATG AAG ATT CTT GAG CTC AGA TAC AAT GGT GGC ATA AAC ATG TAC GTT 900 1 L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC AAT CTG ACC TTT CAG AAT CTA ATG 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC ATT TTG AAC CTA CTA TTG CT CAG GAT CTA ATG 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT TAAT TTO AAG ATC TA CTA ATG 961 GAA TGG ACC AAT CCA AGG CGA ATG AAC CAT ATT TTO AGA GCC CTA GGG CTG AAA GAT ATC 1020 AAG ATA GAG AAG AAG TAT TAT GAA AAT GAA ACCA TAT TTO AGA GCC CTA GGG CTG AAA GAT ATTC 1080 F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GTC GAA GAT ATA TCA 1140 AGG ATG ATG CAC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CTG TAT ATA TCA 1141 AGG ATG GAT ATG CAA AAG CAA ACC TC CT GG GAT GCT CTG GAG GTA ATA TCA 1141 AGG ATG ATG CAC AAA TCT TAC ATA, GAG GAT GAC CT CT CAG GGG GG CCC CAG GC CACT GCC CCC 1261 CCA TTC CTA TTT GTT ATC AGA AAG CAA CCC CCT AGG GAG GCC ACC GAG CAC CAC 1260 1325 CCT TGA AAATCCAATTGGTTTATGGATCTGTATGGCCCACACACATGGTGAATAGATTGGATTTATGGATGGATGAT	001																	v			G	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	c c 1	_	-	-						-							-		-		-	720
721 AAG TGG CAA TCA GCC TTC ACC AAG AGC GAA ACC ATA AAT TGC CAT TTC AAA TCT CCC AAG C S G K A V A M M H H Q E R K F N ⁶ L S V I TGC TCT GGG AAG GCC GC ATG ATG CAT CAG GAA CGC AAT GT TG CAT TTC AAA TCT CCC AAG E D P S M K T L E L R Y N G G I N M Y V 841 GAG GAC CCA TCA ATG AAG ATT CTT GAG CTC AGA TAC AAT GGT GGC ATA AAC ATG GTA CGTT 900 L L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CCT GCT AGA ATG GAC CTC TCT GAA ATT GAA AAC AAA CTG ACC TTT CAG AAT CTA ATG E W T N P R R M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AGA ATT GAA ACC AAA CTG ACC TTT CCAG AAT CTA ATG F D E S K A D L S G I A S G G R L Y I S 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC 1020 K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAA CAA A CG AAC CAT CTC TG GG GT GGT CTG TAT ATA TCA 1141 AGG ATG GCA CAAA TCT TAC ATA GAG GTC ACT GAG GAT GCT CTG GGG GGT CGT CTG TAT ATA TCA 1141 AGG ATG CAC AAA CTC TAC ATA GAG GAT CTC TCT GG AGG GGC ACC GAG GCT ACT GCT GCC 1200 $T G S N I V E K V L E V T E G G T E A T A A 1141 AGG ATG CAC AAA TCT TAC ATA GAG GAT CTC CCT CAG GGG GGC CGT CTG TAT ATA TCA 1141 AGG AGG AAG AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG GCG CGG CTG CTA GT GCC CAC 1200 \overline{T G S N I V E K V L E V T E G G T E A T A A 1141 AGG AGG AGG AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG GCG CTG TAT ATA TCA 1140 \overline{T G S N I V E K V L E V T E G G T C C G G G G C C C AG G CT ACT G C C C C C C C C C C C C C C C C C C$	001																					/20
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$								-			_	_	_	-						-		700
781 TGC TCT GGG AAG GCA GTC GCC ATG ATG CAT CAG GAA CGG AAG TTC ART TTG TCT GTT ATT E D F S M K I L E L R Y N G G G I N M Y V 841 GAG GAC CCA TCA ATG AAG ATT CTT GAG CTC AGA TAC AAT GGT GGC ATA AAC ATG TAC GTT L D F E N D L S E I E N K L T F Q N L M 901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAT GAA CTG ACC TTT CAG AAT CTA ATG 960 E W T N F R R M T S K Y V E V F F F Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AGA TAT GTA GAG GTA TTT TTT CCT CAG TTC 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CCT TAT GTA GAG GTA TTT TTT CCT CAG TTC 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC 1021 AAG ATA GAA AGC AAT CTA TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC 1021 AAG ATA GAA AAG CAA TCT TAC ATA GAG GTC CTC TGGG ATT GCT GGG GGG CTG TAT ATA TCA 1141 AGG ATG CAC AAA TCT TAC ATA GAG GTC CTC TGGG ATT GCT GGG GGC ACC GAG GCT ACT GCC GCC 1200 $T $	721	AAG	TGG	CAA	TCA	GCC	TTC	ACC	AAG	AGC	GAA	ACC		AAT								/80
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		С	S	G	ĸ	A	v	A	м	м	н	Q	Е	R		_				-		
841 GAG GAC CCA TCA ATG AAG ATT CTT GAG CTC AGA TAC AAT GGT GGC ATA AAC ATG TAC GTT 900 L L P E N D L S E I E N K L T F Q N L M 901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC AAA CTG ACC TTT CAG AAT CTA ATG P61 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GAT AAAC CTG ACC TTT CAG AAT CTA ATG 960 K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAA CAA AGC AAA CTT TT TTT CCT CAG TTC 1020 K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAA CAA AGC AAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCT 1080 F D E S K A D L S G I A S G G R L Y I S 1081 TT GAT GAA CCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA 1141 AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GGT GGC ACC GAG GCT ACT GGT GCC 1200 ACC GGA AGT AAT ATT GTA GAA AAG CAA CCT CCT CAG GGA GGC CACC GAG GCT ACT GGT GCC 1201 ACC GGA AGT AAT ATT GTA GAA AAG CAA CCT CCT CAG TCC AGT GCC AAA GGT GAC CAC 1201 ACC GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC AGT GGC AAA GTT TCT TGC 1201 ACC GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC AGT GGC AAA GTT TCT TGC 1201 ACC GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC AGT GGC AAA GTT TCT TGC 1201 ACC GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC AGT GGC AAA GTT TCT TGC 1201 ACC GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC AGT GGC AAA GTT TCT TGC 1201 ACC GGA AGT AAT ATT GTA AAA GAA GAA GAC ATC ATC TTA TTA CAFT GGC AAA GTT TCT TGC 1200 P Stop 1225 CCT TGA MAATCCAATTGGTTTCTGTTATAGCAGTCCCCACCACAACATCAAAGGACCCACCACAAGTCAATAGATYTGGGTTAATTGGA 1335 AAAATGTGGTGTTTCCTTGGTTATAGCAGTCCCCACCACAGTCGAAGAGCCCACCAGAGTCAATAGATYTGGGTTAATTGGAATTT 1376 1386 AAAATGTGGTGTTCCTTGGCATTCTACCACCATTGGTAGCCACCACGAGAGTGAACCACGCTCTTTTGAAAACCACCGCCACTTTCTAGAAATATTTAGGAGTTAATCTAGGAATTCTAGCACTCAGGAAATAGGCCTTCCTT	781	TGC	TCT	GGG	AAG	GCA	GTC	GCC	ATG	ATG	CAT	CAG	GAA	CGG	AAG	TTC	AAT	TTG	TCT	GTT	ATT	840
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		Е	D	P	s	м	к	I	L	E	L	R	Y	N	G	G	I	N	М	Y	v	
901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC AAA CTG ACC TTT CAG AAT CTA ATG E W T N P R R M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GTT GAG GTA TTT TTT CCT CAG TTC 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGG GCC CTA GGG CTG AAA GAT ATC 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGG GCC CTA GGG CTG AAA GAT ATC 1021 AAG ATA GAA ATG CAA GAA CAA TAT TTG AGG GCC CTA GGG CTG TAT ATA TCA 1021 AAG ATA GAA ATG CAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA 1041 AGG ATG GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA 1140 R M M H K S Y I E V T E E G T E A T A A 1141 AGG ATG CAC AAA TCT TAC ATA, GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCC 1200 120 ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC 1200 \overline{P} F L F V I R K D D I I L L F S G K V S C 1261 CCA TTC CTTA ATT GTA AGA GAA GGAT GAC ATC ATC TTA TTC AGT GGC AAAG GTT TCT TGC 1320 \overline{P} S top 1325 CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCCACAACATCAAAGGACCACCACAAGTCAATAGATYTGGGTTTAATTGCTACATAGGAGTGCCCCACACATGGTAATAGCACCTCGTTGCATTGAACACCACCACATGGTACACACAGGGAACCTTAGACCCTCCAAGTCAATAGGATTTAAATCTAGATTTAGAATTTGGCATTTCAATTGGCCTTCATAGAAAATTTCATTGGCAATTCAAGTGGTAATAAGTCAACAAAATTTCATTGGCAATTCAAGTGGTAAAAAGTAAATTTCATTGGTAAATTTCATTGGCAATTCAAGTGGTAAAAAGTGGAAATTTCATTGGCAATTCTACGATGAATATTCAATTGGCAATTCTACGATGTAATTTCAATTGGCAATTCTACGATGAATATTCAATTGGCAATTCTACGTGAATATTCAATTGGCAATTCTAACATAGGTAGAAACTAAGGAAACCTAAGAAAAAAGCCAACCTAGGTAAAAAGTGGAAAAATATTTCATTGAATTTCAATTGGCAATTCTAAGAAAAATTTAAATTTCAATATTCAATATTCAATAGGTATTAAATTCAATTGGCAATTCTAACAATTGGTAAAAATGTAAAATTTCAATTGAAAAAAAGCCAACCAA	841	GAG	GAC	CCA	TCA	ATG	AAG	ATT	CTT	GAG	CTC	AGA	TAC	AAT	GGT	GGC	ATA	AAC	ATG	TAC	GTT	900
901 CTG CTG CCT GAG AAT GAC CTC TCT GAA ATT GAA AAC AAA CTG ACC TTT CAG AAT CTA ATG E W T N P R R M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GTT GAG GTA TTT TTT CCT CAG TTC K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC AGG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC 1080 F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA 1140 R M M H K S S Y I E V T E E G T E A T A A 1141 AGG ATG CAC AAA TCT TAC ATA GAG GAT GCT ACT GAG GAG GGC ACC GAG GCT ACT GCT GCC 1200 T G S N I V E K O L P Q S T L F R A D H 1201 ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC P F L F V I R K D D I I I L F S G K V S C 1261 CCA TTC CTA TTG GTT ATA GAG AAG GAT GAC ATC ATC TTA TTC AGG GGC AAA GTT TCT TGC 1320 P Stop 1325 CCT TGA AAATCCAATGGTTCTGTTATAGCAGTCCCACACATCAAAGGACCACCACAGTCAATAGATYTGGGTTTAATTGGAA 1386 AAAATCCAATTGGTTCTGTTATATGCAGGCCCCACACATCAAAGGAACCACCGACACTGGTGACTTGACCCTTCTAGAAATATTGGAATTATGGAGAATTATGGTATAATCTAGGTGTATATTCCATAGGAGAATCAACTGGTAGAACCACCGACACATGGAAATAAGGCCTTGACCCTCGTGGATTCTTTATATTCTCCAACATGGTCACCACCACACACA		T.	T.	P	Е	N	D	L	S	Е	I	Е	N	к	L	т	F	ο	N	L	м	
E W T N P R R M T S K Y V E V F F P Q F 961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GTT GAG GTA TTT TTT CCT CAG TTC K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CTG TAT ATA TCA R M M H K S Y I E V T <u>E E G T E A T A A</u> 1141 AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GTC ACC GAG GCT ACT GCT GCC T G S N I V E K Q L P Q S T L F R A D H 1201 ACA GGA AGT ATT GTA GAA AAG CAA CTC TCC TCG TAG TAT ATA TCA 1200 <u>T F L F V I R K D D I I L F S G K V S C</u> 1201 CCT TCG TGA TATT GTT ATC AGG AAG GAT GAC CTC CT CAG TCA CAG GTG GAAT GAC CAC CAC 1260 P F L F V I R K D D I I L F S G K V S C 12161 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC CAC CACAAATAGATYTGGGTTAATGGA 1325 TCTCCCCAATTGGTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGGGTTAATTGGA 1326 AAAATGTGGTGTTCCTTTGGGTATCAGCATCAACATTGGTAGACACTGGTGACTTGCCCTCCAGGTCATTCTATCATCA 1376 TCTCCCCAATTGGCTTCTGTATAGCAGTCCCCACAACAAAAGRACCACCACAAAGTCAATAGATYTGGGTTAATTGGA 1376 TCCTCGATGACAATTGGATTCTAGCATTCAACATTGGTCACCCACATTCTAATTCATAGCAGTCAATAGATYTGGGTTAATTGGA 1376 TACAGTTAGCGCACTTTGAGATTATGGAGGTGCCCAACAACTAGAATAGGATAAATATTCATGGAACTCGGGACCTAGTCTTTCTT	901	_		_	_			_		_		_			_	-	-	~		CTA	ATG	960
961 GAA TGG ACC AAT CCA AGG CGA ATG ACC TCT AAG TAT GTT GAG GTA TTT TTT CCT CAG TTC K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA R M M H K S Y I E V T E E G T E A T A A 1141 AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCT GCC I C S N I V E K V O L P Q S T L F R A D H 1201 ACA GGA AGT ATT GTT ATC AGA GAG GAT CAC TC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC P F L F V I R K D D I I L F S G K V S C 1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ACT CTC TCT TAT TC AGT GGC AAA GTT TCT TGC P Stop 1325 CCT TGA AAATCCAATGGTTTATTCTATAGCAGTCCCCCAACAACACAACAACAAGAAGTAGCCCTCCAAGGGATCAATTGGAATAATT 1365 TACAGTTATCCACGAGAAATTATGGAGGTGCCCAACACGAAGAAGAAGTAGCCCTCCAAGGGAATAATTGGAAATAATT 1376 CCTATGGCCCGTCTTGGAATTATGGAGGTGCCCAACACGAAGAAGAAGTAGCCCTCCAAGGAATAAGTTGAAAAATTT 1365 CAAATATCAAATGGAGTTATTCAACATTGGAAATTATGGAAATAAGCAATTTGAATTATAGAAATTTGGACTTAATTAGAATTTTGAATTTAAAGAATTTGACTTATAGAAGTTTTTAACAGATTTTAACAATTTGAATTTAAAGAATTTGACTTTAATTAA																	F	F	Ð	0	F	
K I E K N Y E M K Q Y L R A L G L K D I 1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC 1080 F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA R M M H K S S Y I E V T E E G T E A T A A 1141 AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCC GCC T G S N I V E K O L P O S T L F R A D H 1201 ACA GGA AGT ATT GTT GAT AGA AAG CAA CTC CCT CAG TCC CAG TCT AGA GCT GAC CAC P F L F V I R K D D I I L F S G K V S C 1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC 1325 ANAATGTGGTGTTCCTTTGGGTTTATTCTTCCTAACATTGGTCACCACAACACCAAGGACAATAGAATYGGGATCAATAGGATTATATGGA 1385 ANAATGTGGTGTTCCTTGGGAAATTATGGGAGGGCCCCACAACACCAACATCAAAGGACCAACGAAGAGAAGAGGACCAACGGGACCCTAGGGACCTTTTGAAACTC 1595 TTCTCCCCCAGGACCGTCTGGAAATTATGGAGGGGCCCCAACATCGGTAAAGGAACGAAGAGGACCTAAGGAATAAGGGTAAAAATGTGGAAATAT 1775 TCTTGTGTGACCATAGGACTTCTACCACCATCTATTGAATTGGAAATTAGGCAATTTGGCATTTTTAAATCAGAATTTTCAATGGCATTTTTAAATCAGATTTTAAATCAGAATTTTGATTTTCATAGGAAATAAGGTGAAAAATGTGAAAATATTTAAGCAGTTTAAGGAATAATGGTAAAATATGGAAATAATGTGAAGAA	961	_		-		-				-						-	-		-			1020
1021 AAG ATA GAG AAG AAT TAT GAA ATG AAA CAA TAT TTG AGA GCC CTA GGG CTG AAA GAT ATC F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA R M M H K S Y I E V T E E G T E A T A A 1141 AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCT GCC T G S N I V E K V O L P O S T L F R A D H 1201 ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC P F L F V I R K D D I I L F S G K V S C 1210 CCT TGA AAATCCAATTGGTTATGCAGTGCCCCACAACATCAAAGRACCACCACACACAAGAGTAAAGATYTGGGTTAATTGGA 1385 AAAATGTGGGTGTTTCCTTTGAGATTATTCTCCCAACATTGGTCCAACACTGAGAGAAGACCTGGTGACTTGACCCTTCCTAGAACATCTATGGATTATTGGAAGTATTGGAATTATGGAAGTATATGGAAGTATATGGAATTATGGAAGTATATGGAATTATGGAAGTATATGGAAGTATATGGAATTATGGAATTATGGAATTATGGAATTATGGAATTATGGAATTATGGAATTATAGGAGTAAACGCAACTGGTAAAGAACCCTGGTGTAAAGAAATGTGGAAAATATTAGGAGTTAACCTCTAAGAATATAAGGATTTTGAATTAGAAGTATATGGAATTATGGAATTATGGAATTATAGGAGTAAACATTGTGATTTAAATCAGGTGTAAAAGACAATATGTAAAGAAAATGTGAAATATTTAAATCGGAATTATGGAATTATGGAATTATGGAATTATGGAATTATGGAAGTATATGGAATTATGGAATTATAGGAATTATAGGAATTATAGGAATTATAGGAATTATAGGAATTATAGGAATTATAGGATTTTAAATCAGTGTTAAATATAAAGACAATATGTAAATATAAATTTAAATGTGGAGTTAATTCAATAGTGTTAAATGAAATATGGAAATATGGAAATTATGGAATTTTGAATTATAGGATTTTAAATCAGATTTTGAATATAAGACAATATGTAAATATAAATTTAAATCAGTGTTAAAGACAATATGTAAAATATTAAATTTCATTAAATCAGTGTAAAAAATGTGGAAATATGGAAATATTTCAATTGGAATTTTGAATTTAAATCAGTGTTAAAAAATGTGGAAATATGGAAATATTTCAATTGAATTTCAATTGGAATTTTGAATTTAAATCAGTGTAAAAAATGTGGAAATATGGAAATTTTGGAATTTTGAATTATAGGAATCTTCAATTGAAATATTAGAATTTAAATCAGTGTTAAATTAAAGACAAATATTAAATTTAAATCAGTGTAAAAAATGTGGAAATATGGAAATATTAAATCAGTGTAAAAAATATTAAATTTAAATTAAATTTAAATCAGTGTAAGAAAATATTAAATTAAATTTAAATTAAATTAAATTAAATTTAAATCAGATGTTAAATATAAGACAAATATGTAAATATTAAATTAAATTAAATTAAATTTAAATTAAATTTAAATTAAATTAAATTTAAA	901						-															
F D E S K A D L S G I A S G G R L Y I S 1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA R M M H K S Y I E V T E E G T E A T A A 1141 AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCT GCC T G S N I V E K Q L P Q S T L F R A D H 1201 ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC P F L F V I R K D D I I L L F S G K V S C 1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC P Stop 1325 CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTTAATTGGA 1385 AAAATGTGGTGTTTCCTTTGGAGTTTATTCTCCCAACATTGGTCAACCACGAGAAGACACTGGTGAACTTGACCCTTCTTTGAACACCTGGTTGA 1595 TTGTCCCCCAGACACTCAACATTGGAATTATGGCAGTCCCACACGATGAACACTGGTGAACTGGGACTTGACCCTTCTTTGAACATTCAATTGGAATTAT 166 1775 TCTTGGTGACCCTTGGAAATTATGCAGTTCACTTCAATTGGAATTATAGCAGTTTCAATTGGAAATTA 1865 GGGATTAGGAGCCTGCCTACCAACATCAAAGATCTAAAGACCATCTTTTTGAAAAAAATGTGGAAATTT 1866 GGGATTAGGGACCTACCAAGATTTTAAGAGTTTTCATTATGCAATTGGAATTTAAAGCATTTTAAAGGTTTTCATTAAAGTTTTGAAATTT 1946 1955 CAAAATT <u>TAAGAGATCTTTTAACTGTGGCAGTGCTTCCAACGAAATAAGATTTTGGCTTTCTTT</u>	4004		_										_			_				-		1090
1081 TTT GAT GAA TCC AAA GCA GAT CTC TCT GGG ATT GCT TCG GGG GGT CGT CTG TAT ATA TCA R M M H K S Y I E V T E E G T E A T A A 1141 AGG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCT GCC T G S N I V E K Q L P Q S T L F R A D H 1201 ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC P F L F V I R K D D I I L F S G K V S C 1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC 1385 AAAATGTGGTGGTTTCCTTTGGATTTATGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTTAATTGGA 1385 TTGTCCCGATCCCTGGCATTGTTATGGAGTGCTCCACCACATGGTAAGCACTGGTGACCTTGCCTACTGGACCTGGTTGA 1595 TTCCCCCCATGACCCGGCTGTTGTATGGAGTGTCCACCACTGGTAAGGACACTGGTAGCCTTCCTAGGAACCTGGTTGA 1595 TTCCCCCCATGACCCTTCGAATTATGGAGTGGTGCTCAACTGGTAAGGACACTGGTAGGCCTTGCTAGGGATCTTTTGAAACTC 1685 TACAGTTATCGCAGATATTATGGAGTGTGCCCACACTGGTAAGGACACTGGTAGAGCCTAGGGACCTTGGGAATTATAGGATTTTAAGGATTTTAAGGATTTAAGGATTTTAATGCAATTTGGATATTTCATATGGAAATAAGCCTGCTGCTGCTGCTCACCGGTGATCCTTTTTGAAACTC 1685 TACAGTTATCGCAGAATATTCTAGGATTTAAGGATTTAAGGATTTAAAGCCATCTAGGAAATATGGTAAAATATTTCAATAGGGTATTTAAAGGATTTTAAAGGATTTTAAATGCTTCAATTGGAAATTTTGGCCTTTCTTGGAAGGAA	1021																					1000
RMMHKSYIEVTEEGTEATAA1141AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCT GCCI200TGSNIVEKOLPQSTIFRADH1201ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC1260 \overline{P} FLFVIRKDDIILFSGKVSC1261CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC1320PStop1325CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAACATCGACACTGGTGACTTGACCATCAATAGATYTGRGTTAATTGGA13761385AAAATGTGGGTGTTCCCTTGGGATATTATGCCACTCAACATCGGTCACACACGGTGACACTGGTAACTTGACACCCGGCTCATTCTAACACACCGGTGGAA14961505TTCTCCCCCATGACCCGTCTTAGCATTCTACCACCATGGTCACACTGGTAAAGGACAACATGGGAACCTTGGGAACCTTTGAAACTC16761685TACAGTTATCGCAGAATTATGGAGTTAACCTCACTGGTAAATGGGAAACTAGGGAACGAAGAGGATAAAATT17661685TACAGTTATCGCAGAATTATGGAGTTTAACTTCAATGCTTCATAGGAAATTTAGGCCTTGCTGCTGCTTCTTGAAAGACAATATGTAAAGACAATATGTAAATGTAAAGACAATATGTACATGTTTTAAATGACTTTCAAATGACTTTCAAATGACATGTTATAAGCAGTATATTGACATGTTAAATTAGAGTTTTAAATGACTTTCAAATGACATGATTATTGGCCTTCTTTGAAAGAACAATATGTACATGTTTTT19461955CAAATAT <u>TAAAGAATCTTTTAAACTGTTGGCAGTTGTTATATCAAAGAACAATATGCTTTCATATTTCATATGCTTTCATATGCTTTCATATGCTTTATAAGGTTTTTTT1946</u>		-	-	_				_										_		_	-	
1141AGG ATG ATG CAC AAA TCT TAC ATA GAG GTC ACT GAG GAG GGC ACC GAG GCT ACT GCT GCC1200 $T \ G \ S \ N \ I \ V \ E \ K \ Q \ L \ P \ Q \ S \ T \ L \ F \ R \ A \ D \ H$ 1201ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC1260 $P \ F \ L \ F \ V \ I \ R \ K \ D \ D \ I \ I \ L \ F \ S \ G \ K \ V \ S \ C$ 12611261CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC1320 $P \ Stop$ 1325CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTAATTGGA13761385AAAATGTGGGTGTTTCCTTTGAGTTTATTCTTCCTAACATTGGTCCACCAGAAGTCAACATGGTGACTTGACCCTCCTAGACACCTGGTGA14961505TTGTCCCGCATCCTGGGAAATTATGGAGTGTCCCACCACTGGTGAACACGGGAGACGTAGAAGTAGACTCTTTTGAAACTC16761685TACAGTTATCGCAGAATTTCTAGCTTCATTGTAAGGAAGTCTAGGAAATAAGGCCTGCTGCTAGAAAGAGGAAAATATTTTTGAAAGGAGAAAATATTTAGAGTTTTAACTGTTGAATTCAAAGGACTATTTGAAGAGAATAATGTGAAATATGTGAAATATGTGAAATAGGGAAATTATGACATGTTAAATTATGACATTTTTAAATCTAGAGTTTAAAGACAATATGTAAAGACAATATGTACATGTTTTT19461955CAAAATAT <u>TAAAGATCTTTTAAACTGTTGGCAGTGTTAACACACAGAATATTTTCATATGCTGTGAGAGAAGAATAAGGTTTAAAGTTTTTCCTCTATTT1946</u>	1081	TTT	GAT	GAA	TCC	AAA	GCA	GAT	CTC	TCT	GGG	ATT	GCT	TCG	GGG	GGT	CGT	CTG				1140
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		R	М	М	H	ĸ	S	Y	I	E	v	т	E			T		A				
1201 ACA GGA AGT AAT ATT GTA GAA AAG CAA CTC CCT CAG TCC ACG CTG TTT AGA GCT GAC CAC 1260 P F L F V I R K D D I L F S G K V S C 1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC 1320 P Stop 1325 CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTAATTGGA 1376 1385 AAAATGTGGGTGTTTCCTTTGAGTTTATTCTTCCTAACATTGGTCCACCAGAGAGAG	1141	AGG	ATG	ATG	CAC	AAA	TCT	TAC	ATA	GAG	GTC	ACT	GAG	GAG	GGC	ACC	GAG	GCT	ACT	GCT	GCC	1200
P F L F V I R K D I I L F S G K V S C 1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC TTA TTC AGG AAG GGT TTC TTC<		Т	G	S	N	I	v	E	к	Q	L	P	Q	S	т	L	F	R	A	D	H	
1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC 1320 p Stop 1325 CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTTAATTGGA 1376 1385 AAAATGTGGGTGTTTCCTTTGAGTTTATTTCTTCCTAACATTGGTCAGCAGATGACACTGGTGACTTGACCTTCCTAGACACCTGGTTGA 1496 1505 TTGTCCCGACTCTTAGCATTCTACCACCACTGGTGTCTCACCCATTTCTAATTGCCTTCCTCCCCAGGCTCATTCTATTCAATTGGAGTGTCTACCTCCACGGTAGAAGTAGGAGACCTGGGGACCTATTCTAGAATTAGGAGTCTTTTGAAACTC 1566 1685 TACAGTTAGCACGATATTTCAATGCTTCATCTCACATGGAAATAAGGAGAAGTAGAAGTGGTAAAAGTGGTAAAATGTGAAATT 1766 1865 GGGATTAGGGACCTACCAAAATATTCATTGATTTCAATGCTTTCAATTTGACATTTTTGATAAGGAAATATGTGAAATT 1946 1955 CAAATAT <u>TAAAGATCTTTTAAACTGTTGGCAGTTGTTAACCACGAAATATTTCATATGCTGTGTAGAAGTAGAGAATATAGGTTTATAAGGTTTTTCCTCATTT 2036 </u>	1201	ACA	GGA	AGT	AAT	ATT	GTA	GAA	AAG	CAA	CTC	CCT	CAG	TCC	ACG	CTG	TTT	AGA	GCT	GAC	CAC	1260
1261 CCA TTC CTA TTT GTT ATC AGG AAG GAT GAC ATC ATC TTA TTC AGT GGC AAA GTT TCT TGC 1320 P Stop 1325 CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTTAATTGGA 1376 1385 AAAATGTGGGTGTTTCCTTTGAGTTTATTCTTCCTAACATTGGTCAGCAGAGAGACACTGGTGACTTGACCCTCCTAGACACCTGGTTGA 1496 1505 TTGTCCCGATCCTGGCAGTCTACCACCACCACGTGGTCCACCCAC		P	F	L	F	v	Ī	R	к	D	D	I	I	L	F	S	G	ĸ	v	S	С	
P Stop 1325 CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTTAATTGGA 1376 1385 AAAATGTGGGTGTTTCCTTTGAGTTTATTCTTCCTAACATTGGTCAGCAGATGACACTGGTGACTTGACCTTCCTAGACACCTGGTTGA 1496 1505 TTGTCCTGATCCCTGCTTAGCATTCTACCACCACTGTGTCTCCACCATTTCTAATTGCTTTCTTCCCCACGCTCCATTTCTATCA 1586 1595 TTCTCCCCCCATGACCCGCTCGGAAATTATGGAGRGTGCTCAACTGGTAAGGAGAAGTAGAGAGAGAGGGGACCTTTTTGAAACTC 1676 1685 TACAGTTACCGCAGATATTCTAGCATTCTACTTCATAGAAATAAGCCCTGCGGCTTTCTAGAAATAAAGTGTGAAAATT 1766 1775 TCTTTGTTGACCTACCAAAATTTCATTGACTTCCATAGTTTGAATTTAAACGGTGTATAATCTAGAAGAGACAATATGTGAAATT 1856 1865 GGGATTAGGGACCTACCAAAATATTCATTAATGCTTTCAATTGACAAATTTGGCCTTTCTTT	1261	_	TTC		-	GTT	ATC	AGG		GAT	GAC	ATC	ATC		TTC	AGT	GGC	ААА	GTT	TCT	TGC	1320
1325CCT TGA AAATCCAATTGGTTTCTGTTATAGCAGTCCCCACAACATCAAAGRACCACCACAAGTCAATAGATYTGRGTTTAATTGGA13761385ANAATGTGGTGTTTCCTTTGAGTTTATTTCTTCCTAACATTGGTCAGCAGATGACACTGGTGACTTGACCTTCCTAGACACCTGGTTGA14961505TTGTCCCGATCCCTGGTCTTAGCATTCTACCACCATGTGTCTCCACCCATTTCTAATTGCATTGCCTTCCTCCCCAGGCTCATTCTATCA15861595TTCTCCCCCATGACCCGGCTGGAAATTATGGAGRGTGCTCAACTGGAAATAAGGAGAGCGTGGTAGAAGTGGAAAGTGGAAAATATTTGATTGCTTTCTACGAAATAAGTGGAAAATT17661685TACAGTTACGCAGATATTTAGAGTTTAACTTCATTGACATTCCATAGGAAATTAAGGAGTGTATAATCTAGAAGTGGAAAATATGTGAAATT18561865GGGATTAGGGACCTACCAAAATATTCATTAATGCTTTCAATTGACAAATTTTGGCCTTCTTTGAAAAGACAATATGTACATGTTTTT19461955CAAATAT <u>TAAAGATCTTTTAAACTGTTGGCAGTTGTTAACCACAGAAATATTCCATATGCCGTGGTAGAGTTATAAGGTTTATAAGGTTTTTCCCTCTATTT</u> 2036	1001																					
1385AAAATGTGGGTGTTTCCTTTGAGTTTATTTCTTCCTAACATTGGTCAGCAGATGACACTGGTGACTTGACCCTTCCTAGACACCTGGTTGA14961505TTGTCCTGATCCCTGGTCTTAGCATTCTACCACCATGTGTCTCACCCATTTCTAATTTCATTGTCTTTCTT	1325				CC22'	րանշատ	ምምሮምሶ	<u>ሙ የ</u>	AGCA	TCCC	CACAA	CATC		ACCA	CACA	AGTC	ATAG	ATYTO	RGTT	TAATT	GGA	1376
1505TTGTCCTGATCCCTGGCTCTTAGCATTCTACCACCATGTGTCTCACCCATTTCTAATTTCATTGTCTTTCTT																						1496
1595TTCTCCCCCCATGACCCGCTGGAAATTATGGAGRGTGCTCAACTGGTAAGGAGAACGTAGAAGTAGCCCTAGGGATCCTTTTTGAAACTC16761685TACAGTTATCGCAGATATTCTAGCTTCATTGTAAGCAACTGGGAAATAAGCCCTGCTGCTGCTTTCTAGAAATAAGTGTGAAGGATAAATTT17661775TCTTTGTTGACCCTATGAAGATTTTAGAGGTTACCTTCATATGTTTGAATTTAAATCAGTGTAAAGATCAAATCTAGATGGTAAAAAATGTGAAATT18561865GGGATTAGGGACCTACCAAAATATTTCATTAATGCTTTCAATTGACAATTTTGGCCTTTCTTT																						1586
1685TACAGTTATCCGAGATATTCTAGCTTCATTGTAAGCAATCTAGGAAATAAGCCCTGCTGCTTTCTAGAAATAAGTGTGAAGGATAAATTT17661775TCTTTGTTGACCTATGAAGATTTTAGAGGTTACCTTCATATGTTTGATTTTAAATCAGTGTATAAATCTAGATGGAAGAAAAATGTGAAATAT18561865GGGATTAGGGACCTACCAAAATATTTCATTAATGCTTTCAATTGACAAATTTTGGCCTTTCTTT																						1676
1775 TCTTTGTTGACCTATGAAGATTTTAGAGTTTACCTTCATATGTTTGATTTTAAATCAGTGTATAAATCTAGATGGTAAAAAATGTGAAATT 1856 1865 GGGATTAGGGACCTACCAAAATATTTCATTAATGCTTTCAATTGACAAATTTTGGCCTTTCTTT																						1766
1865 GGGATTAGGGACCTACCAAAATATTTCATTAATGCTTTCAATTGACAAATTTTGGCCTTTCTTT																						1856
1955 CAAATAT <u>TAAAGATCTTTTTAACTGTTGGCAGTTGTTATCTACAGAATCATATTTCATATGCTGTGTAGTTTATAAGTTTTTCCTCTATTT</u> 2036																						1946
																						2036
			_				•															

Figure 1. Nucleotide sequence of the megsin cDNA and its deduced amino acid sequence. The sequencing of cDNA revealed an open reading frame of 1140 nucleotides coding for a 380-amino acid polypeptide displayed over the nucleotide sequence in one-letter code. A double overline indicates a serpin signature. Pn-numbered portion enclosed by a box indicates the RSL and the arrow indicates the putative reactive site. Two putative hydrophobic regions are indicated by single underlines. Four potential *N*-glycosylation sites are indicated by asterisks. Confirmed peptide sequences of the fusion protein produced by *E. coli* expression system are indicated by single overlines. Initial nucleotide sequence obtained by a rapid large-scale DNA sequencing of a 3'-directed cDNA library of human cultured mesangial cells is indicated by a double underline. The accession number for the nucleotide sequence of megsin is AF027866 (DNA Data Bank of Japan, European Molecular Biology Laboratory, and GenBank).

The size of the fusion protein produced by *E. coli* expression system was ~ 85 kD (Fig. 2), which was equivalent to the calculated molecular weight of the deduced amino acid sequence of the fusion protein with the maltose-binding protein (42 kD). The amino acid sequencing analysis of two Endo Lys-C-digested peptides of the fusion protein confirmed that this open reading frame was correct (single overlines in Fig. 1).

Amino acid homology search by the FASTA program throughout the SwissProt database revealed that this gene was highly homologous to members of the serpin superfamily (12–16), which is a group of structurally related proteins and generally serves as extracellular, irreversible serine protease inhibitors. Squamous cell carcinoma antigen 1 (SCCA1) showed the highest homology to megsin (41.2% identity, see also Fig. 3*A*),

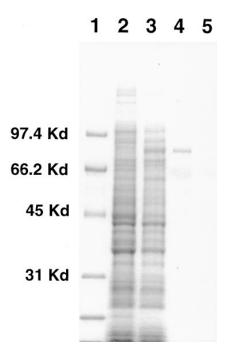


Figure 2. SDS-PAGE analysis of the fusion protein (maltose-binding protein megsin) expressed in *E. coli*. After SDS-PAGE on a 4–20% gradient acrylamide gel, the gel was stained with Coomassie brilliant blue. Lane *1*, molecular weight markers; lane *2*, *E. coli* lysate with pMalc2 (5 μ g); lane *3*, *E. coli* lysate with pMalc2 containing megsin cDNA (5 μ g); lane *4*, purified fusion protein (maltose-binding protein megsin) (1 μ g); lane *1*, Endo Lys-C-digested peptides of the fusion protein (5 μ g).

followed by other members of the serpin superfamily, including SCCA2 (40.6%), leukocyte elastase inhibitor (37.5%), and plasminogen activator inhibitor-2 (PAI-2) (35.2%). Therefore, we termed this gene megsin after <u>me</u>sangial cell-predominant gene with a homology to serpin.

Motif search of the amino acid sequence of megsin confirmed the serpin signature at the COOH terminus. Comparison with the other members of the serpin superfamily suggested that amino acids 331-352 corresponded to the reactive site loop (RSL) (P17 to P5') (17–20) (a box in Fig. 1 and Fig. 3 *B*).

Northern blot analysis and RT-PCR confirmed that megsin was predominantly expressed in human mesangial cells. By Northern blot analysis with a megsin cDNA probe, a single transcript of 3.0 kb was detected in cultured mesangial cells, but not in any other organs or cancer cell lines tested (Fig. 4 A). Although megsin transcript was detected in primary cultured human fibroblasts, the amount of the transcript was apparently much less than in mesangial cells (Fig. 4 B). The megsin transcripts were detected only in cultured human mesangial cells by RT-PCR at 20 cycles (Fig. 5A). When the samples were amplified up to 40 cycles, faint bands corresponding to megsin could be observed in kidneys from two different subjects, some other primary cultured cells, and circulating cells (Fig. 5 B). Without reverse transcription, no bands could be detected (data not shown). These results demonstrated that megsin was predominantly expressed in human mesangial cells.

In situ hybridization studies demonstrated that expression of megsin was upregulated in the proliferating mesangium. The megsin mRNA expression was evaluated on human renal tissues from 18 patients with IgA nephropathy (IgA-N) and three normal kidneys by in situ hybridization. Patients with IgA-N were divided into two subgroups: those with mesangial cell proliferation with minimal glomerular sclerosis (proliferative stage, n = 9) and those with glomerular sclerosis in > 30% of glomeruli (sclerotic stage, n = 9). Megsin mRNAs were demonstrated exclusively in the glomeruli of both normal subjects (Fig. 6 A) and IgA-N (Fig. 6 B). Within the glomeruli, megsin transcripts were detected in location consistent with mesangial cells (Fig. 6 C). Pretreatment of tissues with RNase before hybridization eliminated most of the signal detected with the megsin probe, and the megsin signals disappeared with a large amount of unlabeled homologous oligonucleotide competitors but not with nonhomologous oligonucleotides (data not shown). The number of positive cells for megsin in IgA-N of the proliferative stage was significantly greater than those in normal kidneys (Table I).

Discussion

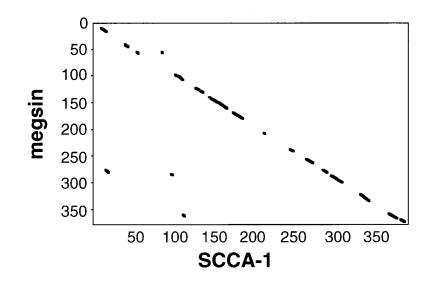
Using a novel approach entailing a rapid large-scale DNA sequencing and computerized data processing, we discovered a new human mesangium-predominant gene "megsin." Megsin was composed of 380 amino acids and was a new member of the serpin superfamily. Although some serpins have functions distinct from protease inhibition, three evidences suggest that megsin is an inhibitory serpin. First, the Peven residues of the megsin RSL were uncharged, small, and nonpolar, which is a characteristic feature of the inhibitory serpins. Second, inhibitory serpins have the sequence of Ala-Ala/Thr-Ala-Ala at the NH₂-terminal region of the RSL (P12 to P9), called the hinge region. P12 to P9 of the RSL of megsin were ATAA. The P17-P8 sequence of megsin (EEGTEATAAT) was indeed consistent with the consensus sequence among inhibitory serpins (EEGTEAAAAT) (Fig. 3). Third, it is known that the stretch of β -sheet immediately precedes the NH₂-terminal side of the RSL, which is a prerequisite for proteinase inhibition, and limits the sizes and charges of amino acids in the hinge region to achieve appropriate conformational change. Megsin preserved this β -sheet stretch.

Residues flanking the putative scissile bonds within the RSL (P1 and P1'), which are thought to be important in determining the substrate specificity (21–23), were Lys-Gln. No other inhibitory serpins have related sequences at this site. Serpins such as the Kunitz type bovine basic protease inhibitor

Table I. Quantitative Analysis of Megsin mRNA from Normal and IgA-N Kidneys by in situ Hybridization

	Percentage of positive cells
IgA-N/Proliferative stage $(n = 9)$	25.21±3.91*
IgA-N/Sclerotic stage $(n = 9)$	18.39 ± 4.91
Normal kidney ($n = 3$)	14.47 ± 1.96

To quantitate the megsin mRNA expression by in situ hybridization, all nuclei and nuclei-surrounding positive cytoplasm in at least 10 randomly selected glomeruli were blindly enumerated, and the results were expressed as percentage of positive cells of total nuclei. Data are mean \pm SD. **P* < 0.01 versus normal kidney or IgA-N/Sclerotic by Mann-Whitney U test.



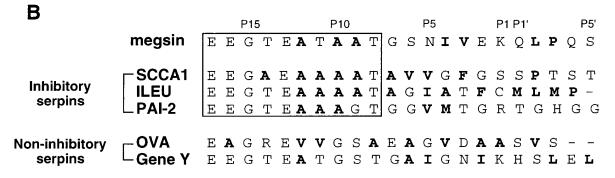


Figure 3. Comparison of the amino acid sequences of megsin with other members of the serpin superfamily. (*A*) A comparison of the protein sequences between megsin and SCCA1. Matching regions are shown as diagonal lines. These sequences were aligned according to the proteinscoring matrix pam 250 with window size 8, minimum percent score 60, and hash value 2. (*B*) Comparison of serpin RSLs. Inhibitory serpin RSLs from P17–P5' (Schechter and Berger numbering [37]) were aligned. Nonpolar residues are shown in bold. P17–P8, indicated by a box, corresponds to the consensus sequence for inhibitory serpins (*EEGTEAAAAT*) (38). *ILEU*, leukocyte elastase inhibitor; *OVA*, ovalbumin.

are known to have Lys at P1 and inhibit trypsin strongly. The target serine protease of megsin might be, therefore, a lysine-cleaving protease.

Hydrophilicity profile showed two hydrophobic regions in α helix A (amino acids 1–16) and B (amino acids 22–44) of megsin (single underlines in Fig. 1), whereas we could find no apparent NH₂-terminal signal peptide sequence. Several serpins appear to exist as dualistic molecules that can be either secreted or cytoplasmic, due to these noncleavable internal signal sequence in a helix A and B (24–28). It is possible that these α helix regions play important roles for translocation of megsin as dualistic molecules.

Northern blot analysis revealed the predominant expression of megsin in cultured human mesangial cells. Megsin transcript was not detected in poly A⁺ RNA from other cancer cells or organs tested by Northern blotting. Localization of megsin-producing cells in our in situ hybridization studies of human renal tissues from normal subjects and patients with IgA-N were consistent with mesangial cells. Furthermore, we cloned a megsin homologue from cultured rat mesangial cells, which were stained positive for Thy-1 antigen, a marker of rat mesangial cells (29) (our manuscript in preparation). These findings strongly suggest the predominant expression of megsin in mesangial cells.

Northern blot analysis did not detect megsin transcripts in human kidney samples, although RT-PCR analysis of human kidneys and some other primary cultured or circulating cells showed corresponding bands after the samples were amplified up to 40 cycles. One possible explanation for this discrepancy is due to the small number of mesangial cells in the kidney, which is also composed of endothelial cells, epithelial cells, and other various cells. Alternatively, cultured mesangial cells are known to acquire new phenotypes, and become more prone to proliferate or become activated (30, 31). Our in situ hybridization studies demonstrated that the expression of megsin was upregulated in IgA-N of the proliferative stage. Therefore, it is more likely that expression of megsin is upregulated only when mesangial cells are activated.

Α

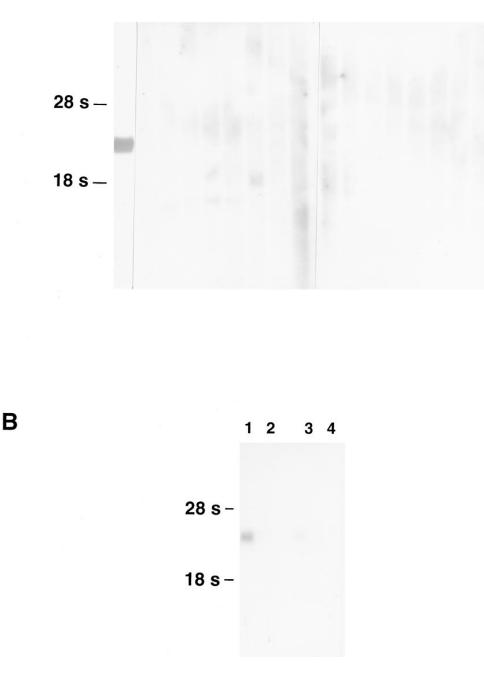
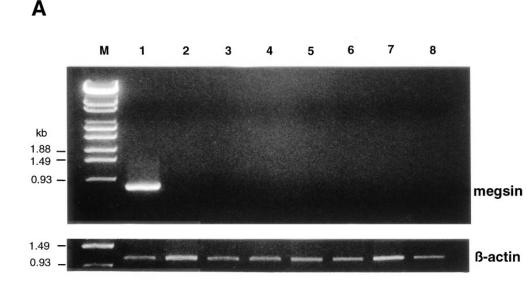


Figure 4. Detection of megsin transcripts. (A) Expression of megsin in cultured mesangial cells, various human organs, and cancer cell lines were examined by Northern blot. Each lane contained 2 μ g of poly A⁺ RNA. The expression of megsin was observed in mesangial cells. Lane 1, cultured mesangial cells; lane 2, pancreas; lane 3, kidney; lane 4, skeletal muscle; lane 5, liver; lane 6, lung; lane 7, placenta; lane 8, brain; lane 9, heart; lane 10, melanoma cells (G361); lane 11, lung carcinoma cells (A549); lane 12, colorectal adenocarcinoma cells (SW480); lane 13, Burkitts' lymphoma cells (Raji); lane 14, lymphoblastic leukemia cells (MOLT-4); lane 15, chronic myelogenous leukemia cells (K-562); lane 16, HeLa cells (S3); lane 17, promyelocytic leukemia cells (HL-60). (B) Northern blot analysis of primary cultured mesangial cells and fibroblasts. The expression levels of megsin in mesangial cells were much higher than in fibroblasts. Lane 1, 20 µg of total RNA from mesangial cells; lane 2, 20 µg of total RNA from fibroblasts; lane 3, 4 µg of total RNA from mesangial cells; lane 4, 4 µg of total RNA from fibroblasts. Horizontal bars on the right correspond to the sizes of 28S and 18S ribosomal RNA (5.1 kb and 1.9 kb, respectively).

The mechanism of mesangium-predominant expression of megsin is another important subject to be pursued. Further studies of the promoter and transcriptional factor of megsin will give an important insight to a mechanism of cell type– dependent gene expression. This information can also be useful in targeting specific cells when molecular genetics and gene transfer techniques are applied.

Furthermore, mesangium-predominant genes can also serve as biological markers for mesangial cells. Although Thy-1 antigen has been used to identify rat mesangial cells, rat Thy-1 antigen is also expressed by other cell lineages such as thymocytes and, more importantly, human mesangial cells lack this molecule (32). Human mesangium-specific markers provide great tools to investigate glomerular diseases, which renders identification of megsin even more importance.

Sequence homology and similar biological activity of megsin with PAI-2 and other inhibitory serpins gives us a clue for its putative function. Serpins play a central role in the regulation of a wide variety of physiologic and pathological processes including coagulation, fibrinolysis, matrix metabolism,



В

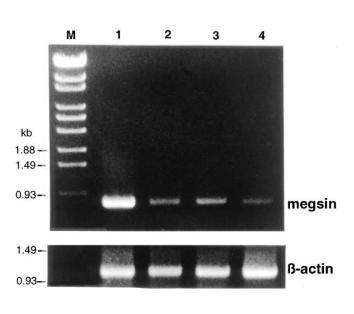


Figure 5. PCR detection of megsin transcripts. (A) PCR analysis at 20 cycles confirmed predominant expression of megsin in mesangial cells (top). The bottom shows PCR products of β-actin as an internal control. Lane 1, mesangial cells; lane 2. smooth muscle cells: lane 3. fibroblasts; lane 4, endothelial cells; lane 5, renal epithelial cells; lane 6, keratinocytes; lane 7, monocytes; lane 8, polymorphonuclear leukocytes. (B) PCR demonstrated bands corresponding to megsin in kidneys when the samples were amplified up to 40 cycles. Lane 1, cultured mesangial cells; lane 2, kidney (subject A); lane 3, kidney (subject B); lane 4, fibroblasts.

development, blood pressure, malignancy, and inflammation (14, 33). Glomerulosclerosis results from accumulation of extracellular matrix proteins as a consequence of their increased synthesis and relative inhibition of degradation by proteases. The enzymatic activity of a protease is regulated by a balance among proteinases and their inhibitors. Recent studies demonstrated that transgenic mice that overexpress the PAI-1 showed more marked lung fibrosis in response to bleomycin than wild-type mice, whereas PAI-1 knockout mice are relatively protected (34), emphasizing the role of serpin in extracellular matrix accumulation. It is also becoming clear that serpins inhibit critical cellular processes such as adhesion, migration, free radical production, and even apoptosis (35, 36). Detailed functional analysis of megsin is now under investigation.

In conclusion, megsin is a new serpin protein expressed predominantly in mesangial cells. The expression levels of megsin paralleled mesangial proliferation in vivo. Megsin could be the first biological marker for injured mesangial cells, which renders this molecule a high potential from the diagnostic perspective of glomerular diseases. Furthermore, marked upregulation of megsin in the injured mesangium suggests that megsin may play some roles in the pathogenesis of glomerulosclerosis. Studies of megsin might thus open up the way for future therapeutic approaches to treat various glomerular disorders.

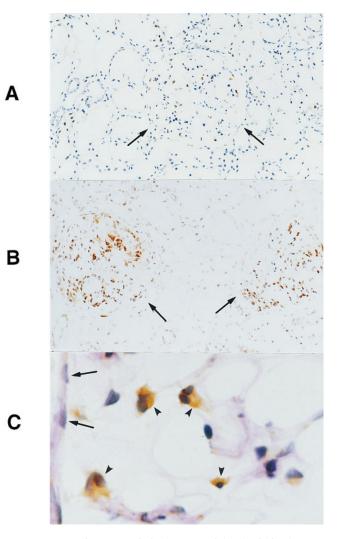


Figure 6. Expression of megsin in the mesangial cells within glomeruli from normal subjects and patients with IgA nephropathy. (A) The signals for megsin are observed in a glomerulus (arrows) of normal subjects. There are no signals for megsin in tubulointerstitial area. Original magnification, ×40. (B) The increased signals for megsin are observed in two glomeruli (arrows) of patients with IgA-N. Original magnification ×40. (C) A high magnification view shows megsin-producing cells in mesangial area (arrowheads), whereas cells of Bowman's capsule (arrows) were negative for megsin. Original magnification, ×200.

Acknowledgments

This study was supported by a grant from Research for the Future Program of the Japan Society for the Promotion of Science (96L00303 to T. Miyata).

References

1. Yasuda, Y., T. Miyata, M. Nangaku, Y. Iida, K. Maeda, K. Kurokawa, and K. Okubo. 1998. Functional quantitative analysis of the genome in cultured human mesangial cells. *Kidney Int.* 53:154–158.

2. Striker, G.E., and L.J. Striker. 1985. Glomerular cell culture. *Lab. Invest.* 53:122–132.

3. Bairoch, A. 1992. PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res.* 20:2013–2018.

4. Bairoch, A., P. Bucher, and K. Hofmann. 1996. The PROSITE database: its status in 1995. *Nucleic Acids Res.* 24:189–196.

5. Ogiwara, A., I. Uchiyama, Y. Seto, and M. Kanehisa. 1992. Construction of a dictionary of sequence motifs that characterize groups of related proteins. *Protein Eng.* 5:479–488.

6. Pearson, W.R., and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*. 85:2444–2448.

7. Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105–132.

8. Rost, B., and C. Sander. 1993. Prediction of protein structure at better than 70% accuracy. J. Mol. Biol. 232:584–599.

9. Suzuki, D., M. Miyazaki, K. Jinde, T. Koji, M. Yagame, M. Endoh, Y. Nomoto, and H. Sakai. 1997. In situ hybridization studies of matrix metalloproteinase-3, tissue inhibitor of metalloproteinase-1 and type IV collagen in diabetic nephropathy. *Kidney Int.* 52:111–119.

10. Matsubara, K., and K. Okubo. 1993. cDNA analyses in the human genome project. *Gene*. 135:265–274.

11. Okubo, K., N. Hori, R. Matoba, T. Niiyama, A. Fukushima, Y. Kojima, and K. Matsubara. 1992. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* 2:173–179.

12. Carrell, R., and J. Travis. 1985. α1-Antitrypsin and the serpins: variation and countervariation. *Trends Biochem. Sci.* 10:20–24.

13. Carrell, R., P.A. Pemberton, and D.R. Boswell. 1987. The serpins: evolution and adaptation in a family of protease inhibitors. *Cold Spring Harbor Symp. Quant. Biol.* 52:527–535.

14. Kruithof, E.K.O., M.S. Baker, and C.L. Bunn. 1995. Biological and clinical aspects of plasminogen activator inhibitor type 2. *Blood*. 86:4007–4024.

15. Potempa, J., E. Korzus, and J. Travis. 1994. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* 269: 15957–15960.

16. Remold-O'Donnell, E. 1993. The ovalbumin family of serpin proteins. *FEBS Lett.* 315:105–108.

17. Aertgeerts, K., H.L. De Bondt, C.J. De Ranter, and P.J. Declerck. 1995. Mechanisms contributing to the conformational and functional flexibility of plasminogen activator inhibitor-1. *Nat. Struct. Biol.* 2:891–897.

18. Hopkins, P.C., and J. Whisstock. 1994. Function of mapsin. *Science*. 265: 1893–1894.

19. Patston, P.A., and P.G.W. Gettins. 1996. Significance of secondary structure predictions on the reactive center loop region of serpins: a model for the folding of serpins into a metastable state. *FEBS Lett.* 383:87–92.

20. Wright, H.T. 1996. The structural puzzle of how serpin serine protease inhibitors work. *Bioessays.* 18:453–464.

21. Stein, P.E., and R.W. Carrell. 1995. What do dysfunctional serpins tell us about molecular mobility and disease? *Struct. Biol.* 2:96–113.

22. Gettins, P., P.A. Patston, and M. Schapira. 1993. The role of conformational change in serpin structure and function. *Bioessays.* 15:461–467.

23. Creighton, T.E., and I.G. Charles. 1987. Sequences of the genes and polypeptide precursors for two bovine protease inhibitors. *J. Mol. Biol.* 194:11–22.

24. Ye, R.D., T.-C. Wun, and J.E. Sadler. 1988. Mammalian protein secretion without signal peptide removal: biosynthesis of plasminogen activator inhibitor-2 in U-937 cells. *J. Biol. Chem.* 263:4869–4875.

25. Wohlwend, A., D. Belin, and J.D. Vassalli. 1987. Plasminogen activatorspecific inhibitors in mouse macrophages: in vivo and in vitro modulation of their synthesis and secretion. *J. Immunol.* 139:1278–1284.

26. Wohlwend, A., D. Belin, and J.D. Vassalli. 1987. Plasminogen activatorspecific inhibitors produced by human monocytes/macrophages. *J. Exp. Med.* 165:320–339.

27. Mikus, P., T. Urano, P. Liljestroem, and T. Ny. 1993. Plasminogen-activator inhibitor type 2 (PAI-2) is a spontaneously polymerising SERPIN: biochemical characterization of the recombinant intracellular and extracellular forms. *Eur. J. Biochem.* 218:1071–1082.

28. Genton, C., E.K.O. Kruithof, and W.D. Schleuning. 1987. Phorbol ester induces the biosynthesis of glycosylated and non-glycosylated plasminogen activator inhibitor 2 in high excess over urokinase-type plasminogen activator in human U-937 lymphoma cells. *J. Cell Biol.* 104:705–712.

29. Miyata, T., K. Isobe, R. Inagi, R. Taguchi, I. Takai, Y. Fujita, T. Iwamoto, T. Hasegawa, K. Maeda, K. Yamada, and I. Nakashima. 1989. Rat mesangial cells actively produce phosphatidylinositol-anchored Thy-1. *Immunology*. 67:531–533.

30. Johnson, R.J., J. Floege, A. Yoshimura, H. Iida, W.G. Couser, and C.E. Alpers. 1992. The activated mesangial cells: a glomerular "myofibroblast?" *J. Am. Soc. Nephrol.* 2:S190–S197.

31. Floege, J., H.R. Radeke, and R.J. Johnson. 1994. Glomerular cells in vitro versus the glomerulus in vivo. *Kidney Int.* 45:360–368.

32. Miyata, T., K. Isobe, R. Dawson, M.A. Ritter, R. Inagi, O. Oda, R. Taguchi, H. Ikezawa, I. Inoue, H. Seo, M. Hasegawa, S. Kobayashi, K. Maeda, K. Yamada, and I. Nakashima. 1990. Determination of the molecular nature and cellular localization of Thy-1 in human renal tissue. *Immunology*. 69:391–395.

33. Bachmann, F. 1995. The enigma PAI-2: Gene expression, evolutionary and functional aspects. *Thromb. Haemostasis.* 74:172–179.

34. Eitzman, D.T., R.D. McCoy, X. Zheng, W.P. Fay, T. Shen, D. Ginsburg, and R.H. Simon. 1996. Bleomycin-induced pulmonary fibrosis in transgenic

mice that either lack or overexpress the murine plasminogen activator inhibitor-1 gene. J. Clin. Invest. 97:232–237.

35. Dickinson, J.L., E.J. Bates, A. Ferrante, and T.M. Antalis. 1995. Plasminogen activator inhibitor type 2 inhibits tumor necrosis factor alpha-induced apoptosis. *J. Biol. Chem.* 270:27894–27904.

36. Kumar, S., and C. Baglioni. 1991. Protection from tumor necrosis factormediated cytolysis by overexpression of plasminogen activator inhibitor type-2. J. Biol. Chem. 266:20960–20964.

37. Schechter, I., and A. Berger. 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27:157–162.

38. Schick, C., Y. Kamachi, A.J. Bartuski, S. Cataltepe, N.M. Schechter, P.A. Pemberton, and G.A. Silverman. 1997. Squamous cell carcinoma antigen 2 is a novel serpin that inhibits the chyotrypsin-like proteases cathepsin G and mast cell chymase. *J. Biol. Chem.* 272:1849–1855.