Isolation and Characterization of cDNAs from *Bam*HI-H Gene Family RNAs Associated with the Tumorigenicity of Marek's Disease Virus

FANGYU PENG,¹ GUY BRADLEY,²† AKIKO TANAKA,² GERALD LANCZ,¹ AND MEIHAN NONOYAMA²*

Laboratory of Virology, Tampa Bay Research Institute, 10900 Roosevelt Boulevard, St. Petersburg, Florida 33716,² and Department of Medical Microbiology and Immunology, University of South Florida School of Medicine Tampa, Florida 33612¹

Received 5 June 1992/Accepted 14 September 1992

It has been reported that loss of the tumorigenic potential of attenuated Marek's disease virus (MDV) is strongly associated with amplification of the 132-bp repeat sequences found within the *Bam*HI-D and *Bam*HI-H fragments contained within the long terminal repeat and the long internal repeat, respectively. The expansion of this region results in loss of transcripts that are 3.8, 3.0, and 1.8 kbp long that are produced by tumorigenic strains of MDV. This evidence suggests that production of one or more of these three RNAs is strongly associated with the tumorigenic potential of the virus. In this study, we have cloned and sequenced 1.69-, 1.5-, 1.9-, and 2.2-kbp cDNAs from the *Bam*HI-H gene family RNAs associated with tumorigenicity. The 1.69- and 2.2-kbp cDNAs are derived from nonspliced transcripts, whereas the 1.5- and 1.9-kbp cDNAs are from single spliced mRNAs spanning the *Bam*HI-H and *Bam*HI-I2 fragments of MDV DNA. Sequence analysis has shown two potential open reading frames in each of the cDNAs. The putative 63-amino-acid protein encoded by the first open reading frame in the 1.69-kbp cDNA and a putative 75-amino-acid protein encoded by the first open reading frame in the 1.5-kbp cDNA showed limited homology with the mouse T-cell lymphoma oncogene and the *fes/fps* family of kinase-related transforming proteins.

Marek's disease virus (MDV) is an oncogenic avian herpesvirus which induces lymphoma in its natural host, the chicken. It has been demonstrated that serial passage of virulent MDV in primary chicken embryo fibroblasts (CEF) in vitro results in viral attenuation and loss of tumorigenicity (6, 19). The loss of viral tumorigenicity strongly correlates with amplification of 132-bp direct repeat sequences located within the homologous *Bam*HI-D and -H fragments contained within the long terminal repeat and long internal repeat, respectively (6, 14, 19). The ability of the virus to induce tumors in infected chickens is severely compromised if its DNA has undergone amplification of the 132-bp repeats.

Analyses of viral transcription in lymphoma tissue obtained from infected chickens or from MDV-induced lymphoma cell lines have revealed transcriptional activity which is limited to approximately 20% of the viral genome (15, 18, 20, 21). Transcriptional activity in the transformed lymphoid cells is detected within the long and short terminal repeats and the long and short internal repeats, while little activity is detected within either the long unique region or the short unique region (21). Recent studies have demonstrated the presence of a BamHI-H gene family producing 5' coterminal transcripts of 1.8, 3.0, and 3.8 kb from the BamHI-D and -H regions that strongly correlated with the tumorigenicity of the virus (1, 2). These RNAs could be detected in primary CEF infected with oncogenic strains of MDV, tissue from MDV-induced lymphoma, and lymphoma-derived cell lines (2, 21). It has been reported that the BamHI-H 1.8-kbp gene family is produced only by oncogenic MDV, and its transcription map was proposed according to Northern (RNA)

blot and S1 mapping analyses (1). However, it will be essential to isolate and characterize the cDNAs derived from these MDV transcripts to determine their precise locations in the MDV genome. Availability of these cDNA clones can be used to identify the biological roles of these oncogenic virus-specific RNAs in tumor induction by MDV. We report here the isolation and characterization of cDNAs derived from the *Bam*HI-H gene family RNAs associated with tumorigenicity of MDV.

MATERIALS AND METHODS

Cells and viruses. The oncogenic strain of MDV, RBIB, was provided by L. N. Schierman (University of Georgia, Athens). The attenuated strain of MDV is a high-passage number clone of RBIB developed by serial passage in CEF to passage 64. The preparation, propagation, and infection of primary CEF with oncogenic and attenuated MDV were performed as previously described (1).

Isolation of RNA and construction of cDNA libraries. Total cellular RNA was isolated from RBIB-infected CEF according to the method described by Chomczynski and Sacchi (4). Poly(A)⁺ RNA was isolated from total cellular RNA by use of the streptavidin-biotin-magnetic-bead technique provided in the Promega PolyAtract kit (Madison, Wis.) according to the manufacturer's instruction. Construction of cDNA libraries from poly(A)⁺ RNA utilized the cDNA synthesis kits obtained from Pharmacia (Piscataway, N.J.) and Promega. The resultant cDNAs were ligated into *Eco*RI-digested, dephosphorylated lambda gt10 vector DNA (Stratagene, La Jolla, Calif.) or *Eco*RI-digested lambda GEM4 vector DNA (Promega).

Isolation of oncogenic virus-specific cDNAs and cDNA sequencing. Virus-specific cDNAs from cDNA libraries were screened by the standard plaque assay (7). Briefly, each

^{*} Corresponding author.

[†] Present address: All Children's Hospital, Laboratory of Immunosuppressive Retrovirology, St. Petersburg, FL 33701.

library was plated at 10,000 PFU per plate with Escherichia coli C600 (hfl mutant) (Pharmacia) or LE393 (Promega). After being transferred to filters, the plaques were hybridized with appropriate ³²P-labeled DNA probes synthesized by the method of Feinberg and Vogelstein (5). Each positive plaque isolated was subject to rescreening at 1,000, 100, and fewer than 10 PFU per plate. Repeated isolation of single plaques giving positive signals yielded pure virus-specific cDNA clones. The cloned cDNA was subcloned into the pBluescript KS⁺ plasmid vector (Stratagene), which was used to transform E. coli DH5 α (F'). cDNAs cloned into the lambda GEM4 vector were rescued by cutting the phage DNA with SpeI and adding T4 DNA ligase to assemble the pGEM4-cDNA plasmid construct and then used to transform E. coli DH5 α (F'). Double-stranded DNA sequencing of the cloned cDNA was carried out by the Sanger dideoxychain termination method (17) using the Sequenase version 2.0 kit (United States Biochemical Corp., Cleveland, Ohio). Both strands of the cDNA were sequenced.

PCR. Oligonucleotide primers were synthesized with the Applied Biosystems 380B DNA synthesizer (Foster City, Calif.). The polymerase chain reaction (PCR) was performed with a GeneAmp Kit (Perkin Elmer Cetus, Norwalk, Conn.) and a Perkin Elmer Cetus DNA Thermal Cycler (Cetus Corp., Emeryville, Calif.).

Northern blot hybridization. Electrophoresis of $poly(A)^+$ RNA after denaturation with 1 M glyoxal in aqueous 50% dimethyl sulfoxide was performed as described by McMaster and Carmichael (16). An RNA ladder from 0.24 to 9.5 kb (GIBCO-Life Technologies, Inc., Grand Island, N.Y.) was used as the size marker. Hybridization was performed as previously described (1).

Computer analysis of cDNA sequence. Analyses of the cDNA and amino acid sequences of the polypeptides encoded by the potential open reading frames (ORF) contained in the cDNAs were performed by using the IntelliGenetics Suite release 5.4 and PC/Gene Software release 6.01 (Intelligenetics, Inc., Mountain View, Calif.).

Nucleotide sequence accession number. The sequences of the 1.69-, 1.5-, 1.9-, and 2.2-kbp cDNAs were submitted to GenBank and have been assigned accession numbers M62573, M77342, M77343, and L01618, respectively.

RESULTS

Isolation of cDNAs derived from oncogenic virus-specific BamHI-H gene family transcripts. Several cDNAs ranging in size from 1.5 to 2.2 kbp were isolated when the cDNA library made from CEF infected with oncogenic MDV RBIB was screened with the 2.9-kbp EcoRI-BamHI subfragment of the BamHI-H DNA fragment and the 1.0-kbp SstII-XbaI subfragment of BamHI-I2 DNA fragment. On the basis of the sequences of the 5' and 3'-end sequences of each clone and their sizes and locations on the BamHI-H and BamHI-I2 fragments of the MDV genomic DNA, three classes of cDNA were identified (Fig. 1). For complete sequencing, cDNAs 1 (class I, 1.69 kbp), 4 (class I, 1.5 kbp), 3 (class II, 1.9 kbp), and 6 (class III, 2.2 kbp) were chosen as representatives of the classes of cDNAs (Fig. 2). The sequences of these representative cDNAs showed that the cDNAs were derived from rightward-transcribed mRNAs of the BamHI-H gene family (Fig. 1 and 2).

The 1.69-kbp cDNA was derived from a nonspliced mRNA containing two copies of the 132-bp repeat sequence. The initiation site of the 1.69-kbp cDNA was at nucleotide 741 of the previously reported *Bam*HI-H sequence (1), a position 87 nucleotides downstream from the TATAAA

sequence. The 3' end of the 1.69-kbp cDNA was located at nucleotide 2437 of the *Bam*HI-H sequence (1), a position 2 nucleotides downstream from a potential polyadenylation signal sequence (AATAAA) (1). The 1.5-kbp cDNA was derived from a spliced transcript sharing the same initiation site with the transcript represented by the 1.69-kbp cDNA but terminating at nucleotide 1449 of the *Bam*HI-I2 fragment (unpublished *Bam*HI-I2 fragment sequence). This spliced transcript utilized a splice donor, GTATGG, beginning at nucleotide 988 of the previously reported *Bam*HI-H fragment (GenBank accession number M26392) and a splice acceptor (ATCGTTGCAG) located 226 bp into the *Bam*HI-*Sst*II subfragment of the *Bam*HI-I2 fragment (see Fig. 4, unpublished I2 sequence).

The 1.9-kbp cDNA of class II was derived from another spliced transcript. The 5' end of this cDNA was located 113 bp downstream from the 5' end of the first 132-bp repeat. Analysis of the cDNA sequence showed that this spliced transcript utilized a splice donor, GTATGC, beginning at position 2088 of the BamHI-H fragment and the same splice acceptor site utilized by the spliced RNA from which the class I 1.5-kbp cDNA was derived (see Fig. 4). This 1.9-kbp cDNA also had the same termination site as the 1.5-kbp cDNA. The 1.9-kbp cDNA is thought to be a partial cDNA clone of the 3.0-kbp transcript for the following reasons. (i) The 3.0-kbp transcript could be detected by hybridization with the PCR-generated DNA probe representing the 5' end of the BamHI-H gene family (Fig. 3, lane 1) and the SstII-XbaI DNA subfragment (Fig. 3, lane 3) of the BamHI-I2 fragment but not with the PCR-generated probe which detected only the 3.8-kbp transcript (Fig. 3, lane 3). (ii) If 1.9-kbp cDNA is a full-length cDNA, the SstII-XbaI DNA subfragment of the BamHI-I2 fragment used as probe should detect the 2.0- to 2.1-kbp transcripts upon Northern blot hybridization of $poly(A)^+$ RNA of CEF infected with oncogenic MDV. However, no 2.0- to 2.1-kbp transcripts were detected in Northern blot hybridization of $poly(A)^{-1}$ RNA isolated from oncogenic-MDV-infected CEF with the SstII-XbaI DNA subfragment of the BamHI-I2 fragment as probe (Fig. 3, lane 5). (iii) Extension of the 1.9-kbp cDNA linearly from the 5' end of this cDNA to the transcription start site of the BamHI-H gene family makes a 2.72-kbp cDNA, which is close to the expected size of the full-length cDNA derived from the 3.0-kbp transcript.

Class III cDNAs are represented by the 2.2-kbp cDNA which was derived from a nonspliced mRNA joining the BamHI-H and BamHI-I2 fragments of MDV DNA. The 5' end of this cDNA is located at nucleotide 2161 of the previously reported BamHI-H sequence (1), and it shares its 3' termination site with the class I 1.5-kbp cDNA and the class II 1.9-kbp cDNA. The 2.2-kbp cDNA (class III, cDNA 6) is likely a partial cDNA clone of the 3.8-kbp transcripts because (i) when aligned with the MDV genomic sequence, no promoter-like sequence within the appropriate context was found upstream of the 5' end of this cDNA; and (ii) a PCR-generated probe prepared from nucleotides 736 to 957 of this 2.2-kbp cDNA sequence, which represented the first 221 bp of the BamHI-I2 fragment (unpublished BamHI-I2 sequence), detected only a 3.8-kbp band upon Northern blot hybridization of poly(A)⁺ RNA isolated from CEF infected with oncogenic MDV (Fig. 3, lane 3).

Northern blot analysis of oncogenic virus-specific transcripts. To confirm the origins of the cDNA clones, we used PCR-generated DNA fragments prepared from these cDNAs and cloned subfragments of MDV genomic DNA as probes to hybridize with $poly(A)^+$ RNA isolated from CEF infected with oncogenic RBIB or attenuated RBIB, respectively. A



FIG. 1. Genomic structure of MDV and locations of the three classes of oncogenic virus-specific cDNAs from the *Bam*HI-H gene family. (A) Genomic structure of MDV with restriction map of the *Bam*HI-H and *Bam*HI-I2 fragments and the region where the oncogenic virus-specific cDNAs are located. Two vertical bars indicate the 132-bp direct repeats. (B) Three classes of oncogenic virus-specific cDNAs are shown, with curves indicating introns and dashed lines representating unfinished sequences. TRL, long terminal repeat; UL, long unique region; IRL, long internal repeat; IRS, short internal repeat; US, short unique region; TRS, short terminal repeat. The restriction sites shown are *Bam*HI (B), *ClaI* (C), *EcoRI* (E), *SalI* (Sa), *SstII* (Ss), and *XbaI* (X).

PCR-generated probe prepared from the 1.69-kbp cDNA and located at the 5' end of the *Bam*HI-H gene family (nucleotides 611 to 858) (2) hybridized with the 3.8-, 3.0-, and 1.8-kbp transcripts produced by oncogenic RBIB in CEF (Fig. 3, lane 1). On the other hand, this probe did not detect those 3.8-, 3.0-, and 1.8-kbp transcripts, which could be detected in poly(A)⁺ RNAs isolated from CEF infected with oncogenic MDV, in Northern blot hybridization of poly(A)⁺ RNAs isolated from CEF infected with attenuated MDV.

A PCR-generated probe derived from the 2.2-kbp cDNA and representing the first 220 bp of the *Bam*HI-I2 fragment was found to hybridize specifically to the 3.8-kbp transcripts but not to 3.0- and 1.8-kbp transcripts (Fig. 3, lane 3). This indicated that this DNA probe was specific for the 3.8-kbp RNA and could be used to screen for cDNAs representative of the 3.8-kbp RNA. No 3.8-kbp RNA was detected by this probe when it was hybridized to RNAs from CEF infected with attenuated RBIB (Fig. 3, lane 4).

To confirm the locations of the 3' termination sites of the *Bam*HI-H gene family transcripts spanning the *Bam*HI-H and *Bam*HI-I2 region, subfragments of *Bam*HI-I2 were used as probes in Northern blot analyses of $poly(A)^+$ RNA

isolated from CEF infected by oncogenic or attenuated MDV. The SstII-XbaI subfragments detected the 3.8- and 3.0-kbp transcripts from RNAs isolated from CEF infected by oncogenic RBIB (Fig. 3, lane 5), but not from RNAs isolated from CEF infected with attenuated RBIB (Fig. 3, lane 6). The 1.7-kbp transcripts were detected in Northern blots containing poly(A)⁺ RNAs isolated from CEF infected with oncogenic or attenuated MDV. Since probe I failed to detect the 1.7-kbp transcript from $poly(A)^+$ RNAs isolated from CEF infected with attenuated MDV (Fig. 3, lane 2), the 1.7-kbp transcripts detected in CEF infected with attenuated MDV is different from the oncogenic virus-specific 1.7-kbp transcript of the BamHI-H gene family represented by class I 1.5-kbp cDNA (Fig. 1, cDNA 4). In contrast, the XbaI-ClaI subfragment of BamHI-I2 failed to detect these transcripts (Fig. 3, lane 7 and 8). Taken together, the results from cDNA sequencing and Northern blot hybridization suggest that the SstII-XbaI subfragment contains the 3' end of those rightward transcripts of the BamHI-H gene family spanning the BamHI-H and BamHI-I2 regions.

Computer analysis of putative polypeptides encoded by ORFs within the oncogenic virus-specific cDNAs. Computer-

(A)	1.69	kbp	cDNA 10		20	30	40	50	60	70	80	90	100
	1 101	GGAA AATT	 CGCTAC AATAGA	GCTAGGC ACGGCGA	 GACGAACGA TTTTTTATT	 .GCTGAATTT TACGGCGAT	 CTCCCTTCAT ATTTG ATG AT Met.Me	 CAAATAAGTA GAAGCGGTTC tLysArgPbe	 ATAAAATTAC GTCGGTCAGT ValGlvGlnS	 TAGCATTCGA CCCCGACGAT SerArgArgS	 ATAAGCAAATA CCGGGACCGCA erGlyThrAla	 AAAAAAAAGG GACCACCGTC AsphisArgG	i ATCTC AACAT Inhis
										 (P)			
	201	CGAC ArgP	CCGGGA roGlyM	TGAGCCT etSerLe	GCTGGGCAG uLeuGlyAr	ACCCATGGG gPro <u>MetGl</u>	TGTCGTTTCT yValValSer	GTATGGTCTC ValTrpSerH	CATTTGCTCAI HisleuleuMe	CTTATGGTC: tLeuTrpSe:	TTTTATCGGAG rPhelleGlyG	GAATGAGCGA lyMetSerAs	CGCGG pAlaA
		CTGG laGl	GTCGTG ySerCy	CAATGTA sAsnVal	GAAAGGATG GluArgMet	GGC TGA TAT Gly***	AATCCGATAT	TTTTTTTTT	STTGGGGGGGAC	CAGTTCCACA	ATATAGGACCG	CGATCATTAA	AATGG
	401 501	GAAG ATAG	GTTTAT ATTGAG	TCTGCGA ACGTTAA	ATGTTGATI TATTTTAAA	ACATGGAAA AAATGCGAI	AATATGTATG GAAAGTGCTA	TGTGGGAGAA TGGAGGAATI	AGTATGTCGA1 AAGGGGGGTGG1 > DR	TTTTAAATGT IGCTCGGCGA	AGTTGGAGTTC GCATGTTCTGT	GGTATTACTT AACTGCCCTT	CCTAT GCTAG
	601	GGTT	CTTCAC	ACGAGCC	TCGCCTTAI	TAAATGTGA	AGTTCGGCA AT Me	CTTACGATO	CTGCCGAAAC CysArgAsnl	AGTTTTTAT LysPheLeuC	GTCTACTTCCA ysLeuLeuPrc DF	.CAAGGGGGGTG GlnGlyGlyG < > > DR	GTGCT SlyAla
	701	CGGC ArgA	GAGCAT rgAlaC	GTTCTGI ysSerVa	AACTGCCC1	TTGCTAGGGI auAlaArgVa	TTCTTCACACG alleuHisThr	AGCCTCGCC SerLeuAlai	TTATTAAATGI LeuLeuAsnVa	IGAGTTCGGC. alSerSerAl	AATGCTTACGA aMetLeuThrI	.TCTGCCGAAA leCysArgAs	ACAAGT snLysP
	801	TTTT heLe	ATGTCT uCysLe	ACTTCCA uLeuPro DR	CATTCAAGO HisSerSei >	CGCTTTCTCA AlaPheSer	A TAG GGATTCI c***	CATTAAAGA	CAAGTTCGGT	AACGCTTTCG.	ATTAGAAACTG	TTTTAGGTGC	CCTTA
	901 1001 1101 1201	TATT AAAC TGAT CTTT	ACACCT TGTCTT GCAATA TGCTGC	TACGAGG ATAAGGA AATACAC GTAGCTO	ATCTCCTT GCATGGCT GACATGAA TTTTGTGAC	AAATTCGAG1 AAATGCGAC1 ACTTTCATT1 CGTCAGGTT1	FCCAGTAATTC FTCGGTAGTGC FATTGACGAAT FGCAACTCCTC	CTTTGAGTG TTTCGACCT TGCACCATT CCATCAACGT	ACGGGAAGCC AGCAATCTCT TAAACTATTT CCCCGCGCGCTT CACATTTCAT	ITACGAGTTG ITTTTTTTTT GGCCGTTGTT CCGTCTTTGT ITACTATCCA	CCCCCCCCAP TTGATTTAATC GCACGATTATC ACATGATAAAI ATGACCCAGTC	AAAAAAAAAAAA TCCCTTGTTA GTTATACTTT ACTACGGCAG	AAAAAA AAGATA FACCTG GAAAGA
	1301 1401 1501 1601	AACC ATAG GATI ATGI	CTTCTC TTTTGA GAATGI ACGACG	GACTGCA GACTGCA GTACTAA GTAACAAG	GATTGACAG AGGTGTAGGI ATGAGGTGGG CGATAATAT	AGGCTCTGGC AGGCCTGTAT GATTAATTTC IGCTTCAAAC	ITGGATATGT GAGATGTATCA GGATTTAGCT	CGTACAAGAG ATCAGGCTAT GAGGGGGGCAA	CTGTTGACACO AGATATGTAA AGTCTGAAAA	CTCTGATATA CATGTTGATT TGTTGGCACA	ATCAAGTTATA GAATGTGTACI ATGAACCTTAA	AGATATGTAAC AATGAGGTGA ACAACTGA	CATGTT AGATTA
(B)	1.5	kbp	cDNA										
		-	10)	20	30	40	50	60	70	80	90	100
	1 101	GGCC GATA	GATCCC	CGATATA TAAAAA	 ATTTCCTGT/ AAAAGGATC'	 AAGGAACGC TCAATTAAT	 TACGCTAGGC(AGAACGGCGA!	 GACGAACGAG ITTTTTATTT	 CTGAATTTCT ACGGCGATAT	I CCCTTCATCA TTG ATG ATGA MetMetI	 AATAAGTAAT AGCGGTTCGT .ysArgPheVa	 AAAATTACTAG CGGTCAGTCCG lGlyGlnSeri 	GCATTC CGACGA ArgArg
	201	TCCC	GGACCO	CAGACCA AlaAspH:	ACCGTCAAC isArgGlnH	ATCGACCCG isArgProG	GGATGAGCCT lyMetSerLe	GCTGGGCAGA uLeuGlyArg	CCCATGGGTG ProMetGlyV	TCGTTTCTAC alValSerTh	CCCCGGATGCA	(P) ITGACTGCGGA LeuThrAlaA:	ACCCAC spProL
									(M)	(C)			
	301	euSe	CATCTO	CGAAACG erLysArd	GATACTGCG. gIleLeuAr	gGlnLeuGl	yThrGluHis	GlnValLeuV	allleAlaSe	rGluValPhe	ValArgGlyL	euAlaPro**	*
	401 501 601	AAC	() TTTGTA AGCTAG ICTGTA	?) ACATGGC GCGGGAT. FATACGC.	GTAGTAGAA AATGAGTCG ACGAACATA	ACGAGAGCT CTGTTTGCA TAAGTCTGT	GTGAGGTTCT CATTATCAAA AAGAATGTAA TCCAGGCTTT	GGCAGAGATI GCTACGCATI TGCTTCGTAC	CCACAAGAGA AGATAACTGC AGATCACTGT ACATTGACAT	AAGAAGTGGA AGAAAGACGC TTATTGAAGI	ATCGACCCTC TGCGTATAGT TCAACGGTATA	IGAATCCAGT TATGTATTCT GAAATTTGAG CGATTCGCTT	ATAAAT TAGAAT TATACC ACCCTT
	801 901	GGG	CAACTT: TTCGAG	ICTCTGT AGGGGTG	CGGTCGTGG AGACCTAAA	CATGCAGTC	GCCGGTGTAT GCATGC ATG T MetT	GTAGAGAGTC GGAACACGAI rpAsnThrI]	TACGATCTTC TGGCCGTTGT eGlyArgCys	GATCTCCTTC AGCATACAAC SerlleGln/	CGGATCACATG SCAGTACACAT AlaValHisMe	GAGCGGAGAT GGCGAAAGTT tAlaLysVal	GTTGTA TGCCGT CysArg
	1001	CCG Prol	CCTGTT(ProVal)	CGGTGTG ArgCysA	ACGTCATGT spValMetP	TTAGGTTTG heArgPheG	AGCATGTAAG luHisValAr	AAAAATGGAA gLysMetGlu	CTGTTAACTC ILeuLeuThrL (D)	TAAAAAGAAG LeuLysArgSe	GTATCTCGCCC erIleSerPro	CATTTGTATC: HisLeuTyrH	ATTCGG isSerA
	1101	L CGG laV	TGGGAA alGlyA	ATATAGG snIleGl	TAATAGGAA yAsnArgLy	AAATCATAC sAsnHisTh	CTACGTCAGG hrTyrValArg	CTTCTTCGTC LeuLeuArgI	(P) TGCGATCGCA LeuArgSerGl 	GAAGTGTCTC nLysCysLet	GAGACGCGCA GluThrArgL	AAGAAGGTCT ysGluGlyLe	GGCGCA WAlaHi
	1201	l TTC sSe	CGATAT. rAspIl	AGTTTGC eValCys	AGCCAATGC SerGlnCys	TTGGTCCGC LeuValArg	GGAATCGGGA GlyIleGlyI	TCGGAGCCG leGlyAlaAs	(P) ATTATCGA TAA spTyrArg***	TACGGAAGC	(P,C) AAAGGGGATAA	(M) CTTCCTTGTT.	ТАСАТА
	1301 1401	GAA	IGTATG GGTCGC	ACCGGAC. CGAGACG	ACGACAACA AACGCGGAC	CGGAAATCC	TGTCTCAAAG GGACGTTCAT	TCTTTGTTTG TGTCTTTGTI	ССССАААТАСА ТТТТСТААТТА	ACGGATATGA TTTTGAATGI	ACGCTTCTGTA PATGTATATTT	AACGATTGCG TTCAGCCTCA	GAAGTA TCCTGT

FIG. 2. Sequences of oncogenic virus-specific 1.69-kbp cDNA (A), 1.5-kbp cDNA (B), 1.9-kbp cDNA (C), and 2.2-kbp cDNA (D). Underlining below nucleotide sequences marks the first 6 nucleotides of the second exon of the 1.5- and 1.9-kbp cDNAs. Boldface letters indicate start and stop codons of the ORFs. Amino acid sequences of these ORFs are indicated with the three-letter code. Underlining of the ORF-A amino acid sequence marks the transmembrane domain predicted by computer analysis. (P), potential protein kinase C phosphorylation site; (C), potential casein kinase II phosphorylation site; (M), myristylation site; DR, direct repeat.

(C) 1.9 kbp cDNA

	10	20	30	40	50	60	70	80	90	100
1	GTTTTTATGTCTACT					 3000000000000000000000000000000000000	 666770770		 	
101	AGTTCGGCAATGCTT	ACGATCTGCC	GAAACAAGTT	TTTATGTCTA	CTTCCACATT	CAAGCGCTTT	CTCATAGGG	ATTCTCATTA	AGACAAGTTO	CGGTAA
201	CGCTTTCGATTAGAA	ACTGTTTTAG	GTGCCCTTAT	ATTACACCTT	ACGAGGATCT	CCTTAAATTC	GAGTCCAGT	AATTCCTTTG	AGTGACGGGAA	AGCCTT
401	TTTTTTTTTTGATTTA	ATCTCCCTTG	TTAAGATATG	ATGCAATAAA	TACACGACAT	GAAACTTTCA	TTTATTGAC	GAATTGCACC	ACCTAGCAATO	TTTGGG
501	CGTTGTTGCACGATT	ATGGTTATAC	TTTACCTGCT	TTTGCTGCGT	ACGTGTTTTG	IGACGTCAGG	TTTGCAACT	CCTCCATCAAC	GTCCCCGCGC	CTTCCG
601	TCTTTGTACATGATA	AATACTACGG AspThrThrA	CAGAAAGAAA	CCCTTCTCAG	TTGTCGATTG	ACACGGCTCT	GGGTGGGAA	CGACCCCGGAT	GCATTGACTO	GCGGAC
	Meerre.		Indiangas	in rosergin	Deusertien	spinikiane	UGIYGIYAS	IIASpriogryc	SystieAspc	YSGIYP
		(C)								
701	CCACTTTCCATCTCG.	AAACGGATAC	TGCGACAACT	AGGAACGGAA ArgaspClyT	CATCAAGTTT	IGGTGATCGC	ATCAGAAGT	TTTTGTGAGAG	GGCTCGCTC	CTTAAT
	I of mill menilobedo	runninspin		ArgAshoryr		SOLANDALA	IIERIGSEI	e necysorum	Ighianigsei	Leuli
			(P)		(C)		(C)			
801	eSerAsnPheValTh	rTrpArgSer	AGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	GCTGTGAGGT erCvsGluVa	TCTGGCAGAGA lLeuAlaGlu	ATTCCACAAG IleProGlnG	AGAAAGAAG lulvsGluV	IGGAATCGACG	CCTC TGA ATCO	CAGTAT
		<u>-</u>	···· j·····				142300141		Deu	
0.01) (C)	CTCCCTCTTT							
1001	GAATACGTCTGTATA	TACGCACGAA	CATATAAGTC	TGTAAGAATG	TAATGCTTCG	TACAGATAAC	TGCAGAAAG	ACGUTGUGTA	TAGITATGIA TATGAAATTI	ITCTTA IGAGTA
1101	TACCTAGAATTAGCA	GGAATAATTG	GCATCTCAAT	TTCTCGAGGC	TTTTTTTTTT	IGCACATTGA	CATCTACCG	GAAAAGTGTG	TATGCGATTCO	GCTTAC
1201	CCTTCCCCAACTTTC	TCTGTCGGTC	GTGGTATATT	GAGGCCGGTG	TATGTAGAGAGA	GTCTACGATC	TTCGATCTC TCTACCATA	CTTCGGATCAC	CATGGAGCGG	AGATGT
1301	IGIAGGUICGAGAG	GGGIGAGACC	IMMENIGEN	M	etTrpAsnTh	rIleGlyArg	CysSerIle	GlnAlaValH	isMetAlaLy:	sValCy
									-	-
1401	SArgProProValAr	GTGTGACGTC gCvsAsnVal	ATGTTTAGGT Met Phe ArgPl	TTGAGCATGT heGluHisVa	AAGAAAAATG 1 ArgivsMet(GAACTGTTAA	CTCTAAAAA hrleulysa	GAAGTATCTCO	SCCCCATTTG	CATCAT
		J-1 p			929 0	orazouzour		- goot i i coci	i i oni sheu	lyrnrs
1501					ACCOMMOMMO) 	P)			
1501	SerAlaValGlyAsn	IleGlyAsnA	rgLysAsnHi	sThrTyrVal	ArgLeuLeuA	raLeuAraSe	rGlnLvsCv	sLeuGluThrA	ArgLvsGluG	ivLeuA
	•	-	- ,- ,	· · · · · · · · · · · · · · · · · · ·	···· J _····		11			
1601	CCCATTCCCATATAC	₩₩₩₽ ₽₽₽₽₽	ATCOTTCOTC	CCCCCANTCC	CONTROCACCI	P) בכמששמשככמ) TRA TACCCA 	(P,C)		M) Dommma
1001	laHisSerAspIleV	alCysSerGl	nCysLeuVall	ArgGlyIleG	lyIleGlyAla	aAspTyrArg	***	HOCHAHOGOG	AIAACIICCII	IGIIIA
1701	CATAGAATGTATGAC	CGGACACGAC	AACACGGAAA	TCCTGTCTCA	AAGTCTTTGT	TTGCGGAAAT	AGAACGGAT	ATGACGCTTC	GTAAACGAT	TGCGGA
1801	AGTACGCGGTCGCCG.	AGACGAACGC	GGACGTGTAG	CGGGGACGTT	CATTGTCTTT	GTTTTTCTAA	TTATTTTGA	ATGTATGTAT	ATTTTTCAGCO	CTCATC
1901	CTGTAAATCGGTCGA	GCATTAAAAG	i							
(D) 2.2	kbp cDNA	20	30	40	50	60	70	80	90	100
(D) 2.2	kbp cDNA 10	20 	30 I	40 1	50 I	60 I	70 1	80 . I	90 I	100
(D) 2.2	kbp cDNA 10 CAGGTGTAGGAGGCC	20 TGTATTGGAT	30 I ATGTCTGTAC	40 I AAGAGCTGTT	50 I GACACCTCTG	60 ATATAATCAA	70 I GTTATAGAT	80 ATGTAACATG	90 ITGATTGAAT(100 STGTAC
(D) 2.2 1 101 201	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGAGTT AACGATAATATTGCT	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT	30 I ATGTCTGTAC. GTATCATCAG TAGCTGAGGG	40 I AAGAGCTGTT GCTATAGATA GGCAAAGTCT	50 GACACCTCTG TGTAACATGT GAAAATGTTG	60 ATATAATCAA TGATTGAATG GCACAATGAA	70 İ GTTATAGAT TGTACTAAT CCTTAACAA	80 ATGTAACATG GAGGTGAGAT CTGAGAATAA	90 ITGATTGAAT(IAATGTACGA(ATCGTATCGT(100 STGTAC CGTAAC SATCAT
(D) 2.2 1 101 201 301	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGGGTT AACGATAATATTGCT ACTCAGGTGTATATG	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA	30 ATGTCTGTAC. GTATCATCAG TAGCTGAGGG TGAGGTAATG	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG	50 GACACCTCTG; TGTAACATGT GAAAATGTTG ACTCAGAAGA	60 ATATAATCAA TGATTGAATG GCACAATGAA GGGGATTATC	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT	80 ATGTAACATG GAGGTGAGAT CTGAGAATAAA TATACTACGG	90 Ittgattgaat(IAAtgtacga Atcgtatcgt(IAgt atg gaa(100 STGTAC CGTAAC SATCAT GGCTAT
(D) 2.2 1 101 201 301 401	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGGGGTT AACGATAATATTGCT ACTCAGGTGTATATGCT	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA	30 TATGTCTGTAC. GTATCATCAG TAGCTGAGGG TGAGGTAATG	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG	50 GACACCTCTG; TGTAACATGT GAAAATGTTG ACTCAGAAGAC	60 ATATAATCAA TGATTGAATG GCACAATGAA GGGGATTATC	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCA	80 ATGTAACATG GAGGTGAGAT CTGAGAATAA TATACTACGG	90 ITGATTGAATC IAATGTACGA(ATCGTATCGTA IAGT ATG GAA(MetGlu	100 GTGTAC CGTAAC GATCAT GGCTAT GlyTyr FCCCCC
(D) 2.2 1 101 201 301 401	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGGGGTT AACGATATATGCT ACTCAGGTGTATATGC ATGAGTTTTTCTGCT MetSerPheSerAla	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GluileGlyS	30 TATGTCTGTAC. GTATCATCAGG TAGCTGAGGGG TGAGGTAATG GCGGGGGGAAG erGlyGlyAr	40 AAGAGCTGTT GGCAAAGTCT TCACTGTTAG GGGTATGTTT gGlyMetPhe	50 GACACCTCTG; TGTAACATGT GAAAATGTTG ACTCAGAAGACAG CATGAGACAG HisgluThrV.	60 ATATAATCAA IGATTGAATG GCACAATGAA GGGGATTATC TAGATTTGTC alAspLeuSe	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHishisAr	80 ATGTAACATG GAGGTGAGAT CTGAGAATAAA TATACTACGG ACAACTCCTCC qGlnLeuLeul	90 ITGATTGAATC TAATGTACGAC ATCGTATCGTA TAGT ATG GAAC MetGlu CCTCGTGGCT? ProArqG1yP]	100 GTGTAC CGTAAC GATCAT GGCTAT GlyTyr ICGCCC heAlaH
(D) 2.2 1 101 201 301 401	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGGGTT AACGATAATATTGCT ACTCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla	20 TGTATTGGAT AATTTGAGATT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GluileGlyS	30 TATGTCTGTAC. GTATCATCAGG TAGCTGAGGGG TGAGGTAATG GCGGGGGGAAG GCGGGGGGAAG GCGJyGlyAr	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT gGlyMetPhe	50 GACACCTCTG; TGTAACATGT GAAAATGTTG ACTCAGAAGACAG CATGAGACAG HisGluThrV	60 I ATATAATCAA GGATGAATG GGCACAATCAA GGGGATTATC TAGATTTGTC alAspLeuse	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHishisAr	80 ATGTAACATG GAGGTGAGAT CTGAGAATAAA TATACTACGG ACAACTCCTCC gGlnLeuLeul	90 TTGATTGAATC TAATGTACGAC ATCGTATCGTT TAGT ATG GAAC MetGlu CCTCGTGGCT ProArgGlyPl	100 GTGTAC CGTAAC GATCAT GGCTAT GlyTyr ICGCCC heAlaH
(D) 2.2 1 101 201 301 401 501	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGATT AACGATAATATTGCT ACCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT ieValCueArcAgaSCT	20 I TGTATTGGAT AATTTGAGATT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT	30 TATGTCTGTAC. GTATCATCAGG TAGCTGAGGGAAG GCGGGGGGAAG GCGGGGGGAAG GCGGGGGGAAG GCGGGGGGAAG GCGGGGGGAAG GCGGGGGGAAG TATATTCCC.	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTTA gGlyMetPhe AAGTTCATTCC	50 GACACCTCTG; TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGG HisGluThrV. TGGGACTTTCC SUCLW SUSS	60 I ATATAATCAA GGATGAATG GGCACAATCAA GGGGATTATC TAGATTTGTC allspleuse cctaccaccccc cctaccaccccc	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCGTT HishisAr ATTTTGGTT	80 ATGTAACATG GAGGTGAGAT CTGAGAATAAA TATACTACGG GGALCCCCCC gGlnLeuLeul GGGGCTCTCTAA GGGGCTCTCTAA	90 TTGATTGAATC TAATGTACGAA ATCGTATCGTA TAGT ATG GAA(MetGlu CCTCGTGGCT ProArgGlyPl CC TGA CTTAG	100 STGTAC CGTAAC SATCAT GGCTAT GlyTyr TCGCCC heAlaH
(D) 2.2 1 101 201 301 401 501	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGAGGCC AACCATAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT isValCysArgAspS	20 TGTATTGGAT AATTTGAGATT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe	30 ATGTCTGTACA GGTATCATCAGG TAGCTGAGGGAAGG GGCGGGGGGAAG GGCGGGGGGAAG GGCGGGGGGAAG GGCGGGGGGAAG GGCGGGGGGAAG GGTGIJGIJA CATATATTCCC CATATATTCCC	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGlyMetPhe AAGTTCATTC LysPhelleL	50 GACACCTCTG; TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGG HisGluThrV. TGGGACTTTC: euGlyLeuSe	60 I ATATAATCAA GGATGAATG GGCACAATCAA GGGGATTATC TAGATTTGTC alAspleuSe cctacAccccg rleuHisPro	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCGTT ATTTTGGTT IleLeuVal	80 ATGTAACATG GAGGTGAGAT CTGAGAATAAA TATACTACGG GCAACTCCTCC GGInLeuLeul GGGGTCTCTAA GlyValSerT)	90 I TTGATTGAATC TAATGTACGAC ATCGTATCGTA TAGT ATG GAAC MetGlu CCTCGTGGCT ProArgGlyPl CC TGA CTTAG hr***	100 STGTAC CGTAAC SATCAT GGCTAT GlyTyr TCGCCC heAlaH TAAGTG
(D) 2.2 1 101 201 301 401 501 601	kbp cDNA 10 CAGGTGTAGGAGGCCT TAATGAGGTGGGATT AACGATAATATTGCT ACTCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT isValCysArgAspS TGTCCGGGAAGCTTGA	20 TGTATTGGAT AATTTGAGATT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT	30 ATGTCTGTAC. GGTATCATCAGG TAGCTGAGGGAAG GCGGGGGGAAG GCGGGGGGAAG GCGGGGGGAAG GCGGGGGGAAG CTGAGGAAG CTGAGAAG CCCCTCAAAA	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGlyMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA	50 GACACCTCTG; TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGG HisGluThrV. TGGGACTTTC: euGlyLeuSe GAGGTTTGGT	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe cctACACCCG rleuHisPro TACCCAGGTA	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHiSAr ATTTTGGTT <u>lleLeuVal</u> GTGCCCCCC	80 ATGTAACATG GAGGTGAGAT TATACTACGG GGINLEULEUI GGGGTCTCTAG GlyValSerTI CAGAAAGAAT	90 TTGATTGAATC TAATGTACGAA ATCGTATCGTA TAGT ATG GAA(MetGlu CCTCGTGGCT ProArgGlyP CC TGA CTTAG hr***	100 J STGTAC CGTAAC SATCAT GLYTYr TCGCCC heAlaH TAAGTG TCGAAC
(D) 2.2 1 101 201 301 401 501 601 701 801	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGATT AACGATAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT isValCysArgAspS TGTCGGGAAGCTTGA GGAGCTCGATCATC	20 TGTATTGGAT AATTTGAGAT TTTATTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT ATCCGGGATCT	30 ATGTCTGTACA GGTATCATCAGG TAGCTGAGGTAATG GCCGCGCGCGAAG HETGIYGIYAT CATATATTCCC CCCCCTCAAAA ATGCTTTGGG	40 I AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAT GGATCCCAAG	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGG HisGluThrV TGGGACTTTC euGlyLeuSe GAGGTTTGGT TTTCGCAGAGTA	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG CLEUHISPTC TACCCAGGTA GTAAATATCC	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHiSAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCCCGGAT	80 ATGTAACATG GAGGTGAGAT TATACTACGG GGINLEULEUI GGGGTCTCTAG GlyValSerTI CAGAAAGAAT CTATTTTGTGG	90 TTGATTGAATC TAATGTACGAC ATCGTATCGTA TAGT ATG GAAC MetGlu CCTCGTGGCT ProArgGlyP CC TGA CTTAG hr*** AATCACTGCT TGGCAATGTGC	100 J STGTAC CGTAAC SATCAT GLYTYr TCGCCC heAlaH TAAGTG TCGAAC GACTTT
(D) 2.2 1 101 201 301 401 501 601 701 801 901	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGGATT AACCATAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT isValCysArgAspS TGTCGGGAAGCTTGA GGAGCTCGATTCATC TCTATGCCTCAGGAA TACTAACAAAGTGTG	20 TGTATTGAGAT AATTTGAGAT TTTATTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT ATCCGGGATT GAACATATCT CCGGTGTACC	30 ATGTCTGTACA GGTATCATCAGG TAGCTGAGGGAAGG GCCGCGCGCGAAG GCCGCGCGCGAAG CTTYT <u>11ePro</u> CCCCCTCAAAA ATGCTTTGGG CACTTGTTGG TACCATCCGT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAAAAAAAT GGATCCCAAT ATTTTGGAG GTTGATTACA	50 GACACCTCTGG GADAATGTTG GADAATGTTG ACTCAGAAGACAG (CATGAGACCAG CATGAGACTTTC :euGlyLeuSe GAGGTTTGGT TTTCGCAGAG TGAGAATTAA GCTGTGATAT	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG CTACACCCG TLeuHisPro TACCCAGGTA GTAAATATCC GAACGTGGTAT CGTTGCAGAC	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCCCGCAT ATGAAGCAG CCCGGAATGC	80 ATGTAACATG GAGGTGAGAT TATACTACGG GGGTCTCTAG GGGGTCTCTAG GlyValSerT CAGAAAGAAT CTATTTGTGG GACTGAAAAA	90 TTGATTGTACGAC ATCGTATCGTA TAGT ATG GAAC MetGlu CCTGGTGGTY ProArgGlyP CC TGA CTTAG hr*** AATCACTGCT TGGCAATGTG AAAAACTATT	100 J STGTAC CGTAAC SATCAT GLYTYr TCGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTC CCATCT
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001	kbp cDNA 10 CAGGTGTAGCAGGCCT TAATGAGGTGGGATT AACCATAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT isValCysArgAspS TGTCGGGAAGCTTGA GGAGCTCGATTCATC TCTATGCCTCAGGAA TACTAACAAAGTGTG CGAAACGGATACTGC	20 I TGTATTGAGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT ATCCGGGATT GAACATATCT CCGGTGTACC GACAACTAGG	30 ATGTCTGTACA GGTATCATCAGG TAGCTGAGGGAAGG AGCGGGGGGAAGG AGCGGGGGGGGAAGA CCCCCTCAAAA ATGCTTTGGG TACCATCCGT AACTTGGGAACAT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAAAAAAT GGATCCCAAT ATTTTGGAG GTTGATTACA CAAGTTTTGG	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGA HisGluThrV. TGGGACTTTCC GAGGTTTGGT TTCGCAGAG GCTGTGATATAA GCTGTGATATA	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG rLeuHisPro TACCCAGGTA GTAAATATCC GAACGTGTAT CGTTGCAGAC AGAAGTTTTT	70 J GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCCGCATGC GTGGCAGAGGG GTGAGAGGGG	80 ATGTAACATG GAGGTGAGAT TATACTACGG GGGTCTCTA GGGGTCTCTA GIyValSerT CAGAAAGAAT CTATTTGTGG GACTGAAAAA ATTGACTGCG CTCGCTCCTT	90 TTGATTGAATC TAATGTATCGTA TAGT ATC GTATCGTA CCTCGTGGCT PrOArgGlyPl CC TGA CTTAG hr*** AATCACTGCT TGGCAATGTG AAAAACTATTT GACCCACTTTT AATTTCCAAC	100 J STGTAC CGTAAC SATCAT GLYTYr TCGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTTC CCATCT TTTGTA
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1101	kbp cDNA 10 CAGGTGTAGCAGGCCT TAATGAGGTGGGATT AACGATAATATTGCT ACTCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAAGCTTGA GGAGCTCGATTCATC TCTATGCCTCAGGAA TACTAACAAAGTGTG CGAAACGGATACTGC ACATGGCGTACAGAC	20 TGTATTGAGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GluileGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT ATCCGGGATT GACAATCTGCG GACAACTAGG AACGAGAGCT	30 CATGTCTGTACA GGTATCATCAGG TAGCTGAGGGAAGG GGCGGGGGAAGG GGCGGGGGAAGG CCCCCTCAAAA CATGCTTTGGG CACTTTGTTGGG TACCATCCGT CACTATGTTCAAA	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAAAAAAAT GGATCCCAAT GGTAGAGATTACA CAAGTTTTGGGG GGCACAGAAT	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGA HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGT TTCGCAGAG GCTGTGATATAA GCTGTGGATAT TGATCGCATC CCACAAGAA	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHisPro TACCCAGGTA GTAAATATCC GAACGTGTAT CGTTGCAGAC AGAAGTGTAA	70 J GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCCCGATGC GTGAGAGGG ATCGAGAGGG ATCGACCT	80 ATGTAACATG? GAGGTGAGATT CTGAGAATAAA TATACTACGG? GCAGAACTCCTCC GGGGTCTCTAC GLyValSerT1 CAGAAAGAAT7 CTATTTGTGG? GACTGAAAAAA ATTGACTGCG CTCGCTCCTT7 CTGAATCCAG	90 TTGATTGAATC TAATGTACGAA ATCGTATCGTA TAGT ATG GAA(MetGlu CCTCGTGGCT' PrOArgGlyP) CC TGA CTTAG' hr*** AATCACTGCT' TGGCAATGTG GACCACTTTT GACCACTTTT AATTTCCAAC' TATAAATAGT'	100 J STGTAC CGTAAC SATCAT GLYTYr TCGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTC CCATCT TTTGTA AGCTAG
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1101 1301	kbp cDNA 10 CAGGTGTAGCAGGCCT TAATGAGGTGGGATT AACGATAATATTGCT ACTCAGGTGTATATG ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATTCATC CCAAACGGATACTGG CGAAACGGATACTGG GCGGGATAATGAGTC TATACGCACGACAACAT	20 TGTATTGAGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GluileGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT ATCCGGGATT CCGGTGTACC GACAACTAGG AACGAGAGCT CTGTTTGCA	30 CATGTCTGTAC. GGTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAGG GCCGCGGGGAAGG CCCCCTCAAA CATGCTTTGGG CACTTTGTTGG TACCATCCGT CACTATCATCGG CACTTACCAACA AAGGAACGTA	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAAAAAAAT GGATCCAAT GGTAGAGAGATT GCTACGCATT GCTACGCAT	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAGA HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGT TTCGCAGAG GCTGTGATATA TGATCGCATC CCACAAGAA AGATAACTGC AGATCACTGT	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHisPro TACCCAGGTA GTAAATATCC GAACGTGTAT CGTTGCAGAC AGAAGTGGA AGAAAGACGA	70 J GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCCGTGAT GTGAGAGGG ATCGACCGT TGCGTATAG TCAACGGTA	80 ATGTAACATG? GAGGTGAGATT CTGAGAATAAA TATACTACGG? ACAACTCCTCC gGlnLeuLeul GGGGTCTCTAG GlyValSerT] CAGAAAGAATT CTATTTGTGG? GACTGAATACAG CTCGCTCCTT CTGAATCCAG? TTATGTATTCAG	90 TTGATTGAATC TAATGTACGAA ATCGTATCGTA CAGT ATG GAAA MetGlu CCTCGTGGCT' PrOArgGlyP) CC TGA CTTAG' hr*** AATCACTGCT' TGGCAATGTG GACCAATTT' GACCACATTT' AATTTCCAAC' TATAAATAGT' TTAGAATACCTAG	100 J STGTAC CGTAAC SATCAT GLYTYr ICGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTTC CCATCT TTTGTA AGCTAG ICTGTA AATTAG
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401	kbp cDNA 10 CAGGTGTAGGAGGCCT TAATGAGGTGGAATATTGCT ACCCAGGTGTATATGCT ACCCAGGTGTATATGCT ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATTCATC CCAGAACGGATACTGCC ACATGGCGTACTGCA ACCGGATAATGGCA	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACCGGATT GACAATATCT GCGGTGTACC GACAACTAGG AACGAGAGC TCTCAATTTC	30 ATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAGG GCCGCGGGGAAGG GCCGCGCGGGAAGG TATATATTCCC. DUTyrIlePro CCCCTCAAAA ATGCTTTGTGG TACCATCGT GTGAGGTTCT GTGAGGTTCT GTGAGGTTCT CATTATCAAAA TCCGAGGCTTT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA GGATCCAAT GGTTCCAAT GGCTACGAGGATT GGCTACGCATT TGCTTCGTAC CTTCTTTTTCG	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG CATGAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGCAATGA GCTGTGATATT TGCACGCATC CCACAAGGA AGATCACTGC AGATCACTGT ACATTGACCTGT	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHi SPro TACCCAGGTAA GGAAGTGTAT CGTTGCAGAC AGAAGAGTGTTTT AAGAAGTGCA TATTGAAGT CTATCGGAAA	70 J GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGCGATGC GTGAGAGGG ATCGACCCT TGCGTATAG TCCACGGTA CCACCTATAG	80 ATGTAACATG? GAGGTGAGAT TATACTACGG? ACAACTCCTCC gGlnLeuLeul GGGGTCTCTAC GlyValSerT] CAGAAAGAAT? CTATTTGTGG? GACTGAATCAG CTGAATCCAG? TGAAATTTGAC CGGATTCGC?	90 TGATTGAATC TAATGTACGAA ATCGTATCGTG CTCGTGGCT ProArgGlyP CC TGA CTTAG hr*** AATCACTGCT GGCAATGGT AAAAACTATT GACCAACTTT AATTTCCAAC TATAAATAGT TAGAATACG GTATACCTAG	100 J STGTAC CGTAAC SATCAT GGLYTYT TCGCCC heAlaH TAAGTG TCGAAC GACTT TTTGTA AAGTTC CCATCT TTTGTA AAGTAG TCTGTA AAGTAG CAACT
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601	kbp cDNA 10 CAGGTGTAGAGAGGCC TAATGAGTGGATATATGCT ACCAAGGTGTATATGCT ACCAGGTGTATATGCT ACCAGGTGTATATGC ATGAGTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTTGA GGAGCTCGATCATC CCAGAACGGATACTGCC ACATGGCGTAACAGTC CCAGAACGGATACTGCC ACATGGCGATACTGCC ACATGGCGATACTGCC ACATGGCGACACAT. CAGGAATAATTGCCA CCGGCTGAGCCTGA	20 TGTATTGGAT AATTTGAGAT TTTTTTTTA GAGATAGGAT GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTGT ATCCGGGATT GCACATATCT GCCGTTTGCA AACGAGAGCT GCTGTTTGCA ATAAGTCTGT TCTCAATTGC GTATATTGAG	30 ATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAG GCCGCGCGAAAG CCCCTCAAAA ATGCTTTGTGG TACCATCCGT TACCATCCGT GTGAGGTTCT CATTATCAAA AAGAATGTAA TCGAGGCTAT GCCAGCGCATGT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA GATCCCAAT GGATCCCAAT GGCAGGAGTTTGGG GGCAGGAGTTTGGG GCTACGCATT TGCTTCGTAC GTAGAGAGTC GGAACACGAT	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAG GCTGTGATTTG TGACAATGACAGA AGATCACTGC AGATCACTGC AGATCACTGC TACGATCTCC	60 I ATATAATCAA IGATTGAATG GGCACAATGAA GGGATTATCC TAGATTTGTCC alAspleuSe CCTACACCCCG TLEUHISPICO TACCCAGGTAA GGAAGGTGTAT CGTTGCAGGAC AGAAGGTGTTTT AAGAAGGCCGTTATTGAAGT CTATCGGAAA GATCTCCTTCC ACCATACAG	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGCGATGC GTGACACGGA ATCGACGGT TCAACGGTA AGTGTGTATA GGATCACAT	80 ATGTAACATG? GAGGTGAGATT CTGAGAATAAA TATACTACGG? QGInLeuLeul GGGGTCTCTA(GIyValSerT) CAGAAAGAAT? CTATTTGTGG? GACTGAATCCGG CTGGATCCGT? TGAAATTTGAC GCGATCGGAAGT? GGCGAAAGT?	90 TGATTGAATC TAATGTACGAA ATCGTATCGTG (AGT ATG GAAT CCTCGTGGCT ProArgGlyP) CC TGA CTTAG' hr*** AATCACTGCTT GGCCATGGT AAAAACTATT GACCCACTTC AATTTCCAAC' TTAGAATAGT TAGAATAGT TAGCCATCCCG GTTGTGTAGGG' TGCCGTCCGG	100 J STGTAC CGTAAC SATCAT GGLYTYT ICGCCC heAlaH TAAGTG TCGAAC GACTT TTTGTA AAGTAC CAACT TTTGTA AAGTAG CAACTT TTCGAG CCACTT
(D) 2.2 1 101 201 301 401 501 601 701 1001 1001 1201 1301 1401 1501 1601	kbp cDNA 10 CAGGTGTAGAGAGGCC TAATGAGGTGGATATATGCT AACGATAATATTGCT ACCCAGGTGTATATGC ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGAGACT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATTCATC CCAAACGGATACTGC ACATGGCGTACAGACT CAAGGGATAATGACCA ACTGCCGATACTGC ACATGCCGATACTGC ACATGCCGATCATGCCA ACGGGATAATGCCA	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTCCGGGATT GACAACTAGGAT CCGGTGTACC GACAACTAGGAAGCT GCTGTTTGCA ATAAGTCTGT TCTCAATTTG GTATATTGAG ACAGCAGTC	30 ATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAG GCCGCGCGAAAG CCCCTCAAAA ATGCTTTGTGG TACCATCCGT TACCATCCGT GTGAGGTTCT CATTATCAAA AAGAATGTAA TCGAGGCTAT GCCGGCGTAT GCCAGCC ATG T	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA GATCCCAAT GGATCCCAAT GGTTGATTACA CAAGTTTGGAG GGCAGGATT GCTACGCATT TGCTTCGTAC GTAGAGAGTC GGAACACGAT rpAsnThrIl	50 GACACCTCTGG GAAAATGTTG GAAAATGTTG ACTCAGAAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAG GCTGTGATTGT TGACAATGAA GCTGTGATATG TGACAACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC	60 I ATATAATCAA IGATTGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCCG TLeuHiSPro TACCCAGGTAA GAACGTGTAT CGTTGCAGGAA AGAAGAGTTTTT AAGAAGGCGTAT CTATCGAAG GATCTCCTTCC SerIleGlnA	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGGATGC GTGACAGGGA ATCGACGTATAG TCAACGGTA AGTGTGTATA GGATCACAT GGATCACAT LAVALHISM	80 ATGTAACATG? GAGGTGAGATT CTGAGAATAAA TATACTACGG? QGInLeuLeul GGGGTCTCTA(GIyValSerT) CAGAAAGAAT? CTATTTGTGG? GACTGAATCCGG? CTGAATCCGG? TGGAATCCGAT GGGATTCGCT? GGACGGAAGA? TGGCGAAAGT? etAlaLysVa?	90 I TGATTGAATC TAATGTACGAA ATCGTATCGTG TAGT ATG GAAT CCTCGTGGCT ProArgGlyPl CC TGA CTTAG' hr*** AATCACTGCTT GACCAATGGT GACAATGTT GACCCACTTC AATAACTAGT TTAGAATAGT TAGAATAGCT TAGCAATGGCT TGCCGTCCGG IGTTGTAGGGC ICysArgProJ	100 J STGTAC CGTAAC SATCAT GLYTYT ICGCCC heAlaH TAAGTG TCGAAC GACTT TTTGTA AAGTAC CAACT TTTGTA AAGTAC CAACT TTTGGAG CCACTT TTCGAG CCACTT TTCGAG
(D) 2.2 1 101 201 301 401 501 601 701 1001 1001 1201 1301 1401 1501 1601	kbp cDNA 10 I CAGGTGTAGAGAGCCC TAATGAGTGGAGAGCCC AACGATAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATCATCGC ACATGGCGTATCATCC ACATGGCGATACTGCC ACATGGCGATACTGCC ACATGGCGATACTGCCA CCGGTGCGGCCCTCATC	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTCT ATCCGGGATT GACAACTAGGAT CCGGTGTACC GACAACTAGGAAGCT GCTGTTTGCA ATTAGGTTTGCA ACTGCAGTCG TCTCAATTCG GACATGCAGTC	30 ATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAG GCCGGGGGAAG TATATATTCCC. DUTYTILePro CCCCCTCAAAA ATGCTTTGTGG TACCATCCGT TACCATCCGT GTGAGGTTCT GCCGGGGTTCT GCCGGGGTAT GCCAGCC ATG MetT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA CAAGTTTGTGAG GGCAGAGATT GGCACGATT TGCTTCGTAC GTAGAGAGGTC GGAACACGAT rpAsnThrIl AAAAATCCAA	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGT TTTCGCAGAG GCTGTGATTTAA GCTGTGATATT TGGATCGCATC CCACAAGAGA AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC GGCCGTTGT eGlyArgCys	60 I ATATAATCAA IGATTGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHisPro TACCCAGGTAA GAACGTGTAT CGTTGCAGGCA GAACGTGTAT CGTTGCAGAC GAACGTGTAT CTATCGAAAG GATCTCCTTC CTACCGGAAA GATCTCCTTC SerIleGInA TAABAACACC	70 J GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGCGATGC GTGACACGGA ATCGACGGA TGCGTTATAG TCAACGGTA AGTGTCTATAG GACTACACA laValHisM TATCTCCCC	80 ATGTAACATG? GAGGTGAGATT CTGAGAATAAA TATACTACGG? ACAACTCCTCC gGlnLeuLeul GGGGTCTCTA(GlyValSerT) CAGAAAGAATT CTGATTGGC? GCTGAATCCGG? CTGAATCCGG? TGGAGTCGCT? GGAGCGGAAGT? etAlaLysVa? CCATTGCT272	90 TGATTGAATC TAATGTACGAA ATCGTATCGTG TAGTATGGAA ATCGTAGGUY ProArgGlyP CCTGACTTAG ATCACTGCT GGCAATGGT AAAAACTATTI GACCACTGTT GACCACTTC AATTCCACC TTAGAATAGT TAGAATAGT TAGCATCCAG TGCTGTAGGGC ICysArgProl	100 STGTAC CGTAAC SATCAT GGTAT GGTAT GGTAT GGTAT GGACG AAGTAC CAACT TTTGTA AAGTAG CAACTT TTCGAA CCTGTT TTCGAG CCTGTT ProVal TGGGAA
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601 1701	kbp cDNA 10 CAGGTGTAGGAGGCCT AATGAGGTGGAAGGCCT ACTCAGGTGGTATATGCT ACTCAGGTGTATATGCT ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATTCATC CCGAACGGATACTGC ACATGGCGTACTGCA ACTGCGATAATGAGTC TATACGCACGAACAT. CAGGATAATGGCA ACGGGTGAGACCTCATG ArgCysAspValMet	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTGT ATCCGGGTATCT GACATATCT GCGGTGTACC GACAACTAGGAGT CTGTATTGCA ATTAGGTTTGCA ACAAGCAGTC GTATATTGAG ACATGCAGTC TTTAGGTTTG PheArgPheG	30 CATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAGG GCGGGGGGAAG GCCCCCCCAAAA ATGCTTTGTGG TACCATCCGT TACCATCCGT CATTATCAAA AAGAATGTAA CGCAGGCATGT GCCAGGCATGT MetT CAGCATGTAAG. LUHISVALAT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAATCGAAT GGTACGATTTGG GGCAGGATTTGG GGCAGGATTTGG GGCAGGATTTTGG GGCACGAT TPASNThrII AAAAATGGAA GLysMetGlu	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGT TTTCGCAGAG GCTGTGATTAA GCTGTGATAACTGC AGATCACTGT ACATTGACATG AGATCACTGT ACATTGACATG AGATCACTGT ACATTGACATG CGCCGTTGT eGlyArgCys CTGTTAACTCC LeuLeuThrL	60 I ATATAATCAA IGATTGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHiSPro TACCCAGGTA GGAAGTGGTAT CGTTGCAGGC GGAAGTGTTTT AAGAAGTGGAA GAACATGCAAG GATCTCCTTC CTATCGAAAG GATCTCCTTC SerIleGlnA TAAAAAGAAG euLysArgSe	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT lleLeuVal GTGCCCCCC ATGCGCGATGC GTGACACGGA ATGAGCACG TCAACGGTA GGATCACAT GGATCACAT GGATCACAT GGATCACAT IaValHisM TATCTCCGCC rIleSerPr	80 ATGTAACATG: GAGGTGAGAT: CTGAGAATAAA TATACTACGG gGlnLeuLeul GGGGTCTCTA(GLyValSerT) CAGAAAGAAT: CTATTTGTGG GACTGAATCAG CTGAATCCAG CTGAATCCAG CTGAATCCAG CTGAATCCAG CGGAGCGAAAGT etAlaLysVa CCATTTGTATG	90 TGATTGAATC TAATGTACGAA ATCGTATCGTA (AGTATGGAAT ProArgGlyP) CCTGACTTAG' hr*** AATCACTGCT' GGCCAATGGTG AAAAACTATT GACCACTTC' AATTTCCAAC' TTAGAATAGT TAGAATAGT TAGCAATGGC GTTGTAAGGC GTTGTAGGCC ICysArgProJ CATTCGGCGCG' HisSerAlaV	100 STGTAC CGTAAC SATCAT GGLTAT GGLTAT GGLTAT GGLAC GACTTT AAGTAC CAACTT TTTGTA AGTAG CAACTT TTCGAG CCTGTT ProVal TGGGAA alGLYA
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601 1701	kbp cDNA 10 I CAGGTGTAGAGAGCCC TAATGAGGTGGAATATTGCT ACCCAGGTGTATATGC ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAAGCTTGA GGAGCTCGATTCATC CCAAACGGATACTGC ACATGGCGTATACAGCC TATACGCACGATACTGC ACAGGATAATGGCA TCTCTGTCGGCTCGTC ACGGGTGAGACCTCATG ArgCysAspValMet	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTGT ATCCGGGTATCT GACATATCT GCGGTGTACC GACAACTAGGATT CCCGGTGTACC GTATATTGAG ACATGCAGTC TTTAGGTTTG PheArgPheG	30 CATGTCTGTAC. GGTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAGG GCGGGGGGAAG GCCGCGCGAACA TATATATTCCC. DUTYTILEPro CCCCCTCAAAA ATGCTTTGTGG TACCATCCGT TACCATCCGT GTGAGGTTCT GCGAGGCTTT GCCGGGGTAT GCCAGGCATGTAG GCCATGCATGTAAG. iluHisValAr	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAATCGAAT GGTACGATTTGG GGCAGGATTTGG GGCAGGATTTGG GGCAGGATTTGG GGCACGAT TGCTTCGTAC GTAGGAGAGCC GGAACACGAT rpAsnThrIl AAAAATGGAA GLysMetGlu	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG HisGluthrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAG GCTGTGAATTAA GCTGTGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC CCCACAGAGAGA AGATCACTGC AGATCACTGC AGATCACTGC CGCCGTTGT CGCCGTTACTCC LeuLeuThrL	60 I ATATAATCAA IGATTGATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHiSPro TACCCAGGTA GGAAGTGGTAT CGTTGCAGAC GGAAGTGTTTT AAGAAGTGGAA GAACATGCAAG GATCTCCTTC CTATCGAAAG GATCTCCTTC SerIleGlnA TAAAAAGAAG euLysArgSe	70 I GTATAAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGCGATGC GTGACACGGA ATCGACGGAT GCGTGAACGGA AGTGTGTATA GGATCACAT GGATCACAT IaValHisM TATCTCGCC rIleSerPr	80 ATGTAACATG: GAGGTGAGAT: CTGAGAATAAA TATACTACGG? ACAACTCCTCC gGlnLeuLeul GGGGTCTCTA(GlyValSerT) CAGAAAGAAT: CTGATTGGCG GCTGAATCCAG GCGATCGGT TGGAAATTGAC GGAGCGAAAGT: etAlaLySVa: CCATTTGTATC	90 TGATTGAATC TAATGTACGAA ATCGTATCGTG TAGT ATG GAAT CCTCGTGGCT ProArgGlyPl CC TGA CTTAG' hr*** AATCACTGCTT GACAATGGTG AAAAACTATT GACCCACTTC AATTTCCAAC' TTAGAATAGT TAGAATAGT TAGCAATGGC GTTGTAAGGG' GTTGTGAGGG' LCysArgProJ CATTCGGCGGG'	100 STGTAC CGTAAC SATCAT GLYTYr ICGCCC heAlaH TAAGTG TCGAAC GACTT TTTGTA AAGTC CCATCT TTTGTA AAGTAG CAACTT TTCGAG CAACTT TTCGAG CAACTT TTCGAG CAACTT TTCGAG CAACTT TTCGAG CAACTT TTCGAG AAGTAC CAACTT TTCGAG AAGTAC
(D) 2.2 1 101 201 301 401 501 601 701 1001 1001 1201 1301 1401 1501 1601 1701 1801	kbp cDNA 10 I CAGGTGTAGAGAGCCC TAATGAGTGGAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAAGCTTGA GGAGCTCGATTCATC CCAAACGGATACTGC ACATGGCGATACTGC ACATGCCGATACTGC ACAGGATAATGGCA CCGGTGTGACGCTCATG ACGGGTGAGACCTCATG ArgCysAspValMet	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTCCGGATT GACATATCT GCCGGTGTACC GACAACTAGGAT CCGGTGTACC GTATATTGAG ACATGCAGTC TTTAGGTTTG PheArgPheG AAAATCATAC	30 CATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAG GCCGGGGGAAG CATGTTGTGG TACCATCCATCA CATTATCTCCC TACCATCGT CATGATGTAG GCCGGGGTAT GCCGGGGGTAT GCCAGGCATGTAG CATGCTGTAAG CATGTAGAAG	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA GGATCCCAAT GGTTCGATAC GGCAGGATTTGG GGCAGGATTTGG GGCAGGATTTGG GGCACGAT TPASNThrII AAAAATGGAA GLYSMetGlu CTTCTTCGTC	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGT TTTCGCAGAG GCTGTGAATTAA GCTGTGAATTAACTGC AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT CCCACAAGAGA AGATCACTGT AGATCACTGC CCCACAGGACA AGATCACTGC CCCACAGGACA AGATCACTGC AGATCACTGC AGATCACTGC CGCGTTAACTCC LeuLeuThrL (P) TGCGATCGCA	60 I ATATAATCAA IGATTGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHiSPro TACCCAGGTAA GAACGTGTTTT AGAAGGTGTTTT AAGAAGGTGTTTT AAGAAGGCG TTATTGAAGT GATCTCCTTCC SerIleGlnA TAAAAAGAAG euLysArgSe GAAGTGTCTC	70 I GTATAAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGTGAT AGTGCGTGAT GCGTGACACG TCAACGGTA AGTGTGTATA GGATCACAT GGATCACAT IaValHisM TATCTCGCC rIleSerPr GAGACCCCC	80 ATGTAACATG' GAGGTGAGATT CTGAGAATAAA TATACTACGG' QGInLeuLeul GGGGTCTCTA(GIyValSerT) CAGAAAGAACTCCTGC GACTGAATCAG CTGAATCCAG' GGAGTGGATCGCTT GGAGCGAAAGTT GGAGCGAAAGT etAlaLySVa CCATTTGTATC	90 TGATTGAATC TAATGTACGAA ATCGTATCGTG TAGTATGGAA MetGlu CCTCGTGGCT ProArgGlyP CCTGACTTAG ACTCACTGCT CTGCCATGGC ADAACTATTI GACCCACTTC AATACCATGCT TAGAATAGT TAGCAATGGC GTTGTAAGGC GTTGTGAGGC ICysArgProJ CATTCGGCGC HisSerAlaV	100 STGTAC CGTAAC SATCAT GGTAT GGTAT GJYTyr ICGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTAC CACTT TTTGTA AAGTAC CACTT ProVal IGGGAA alGlyA CGATAT
<pre>(D) 2.2 1 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601 1701 1801</pre>	kbp cDNA 10 I CAGGTGTAGAGAGCCC TAATGAGTGGATATATGCT ACCCAGGTGTATATGCT ACCCAGGTGTATATGCT ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGAGCTTGA GGAGCTCGATCCATCATC CCGGATACTGCCGATCATGC GCAAACGGATACTGCC ACATGGCGATACTGCC ACATGGCGATACTGCC ACATGCCGATACTGCC ACAGGATAATGGCA CCGGTGTGACGCTCATG ACGGCTGGACGCTCATG ACGGCTGACACCTATG ACGGCTAATAGGCAATACGCA SnIleGlyAsnArgL	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTCCGGATT GACATATCT GCCGGTGTACC GACAACTAGGAT CCGGTGTACC GTATATTGAG ACATGCAGTC TTTAGGTTTG PheArgPheG AAAATCATAC ysAsnHisTh	30 CATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGG CTGAGGTAATG CATGTTGTGG CATGTTGTGGG CACTTGGTGAGGTTCT CATTATCAAA AAGAATGTAA CAGCATGTATG GCCAGGCATGT MetT CAGCATGTAAG CUHISVALAT CTACGTCAGG	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGCAAAGTCT TCACTGTTAG GGCAGTCATTG GGLYMetPhe AAGTTCATTC LysPheIleL AAAAAAAAAA GGATCCCAAT GGTTCGATTACA GGCAGGAGTTTGGG GGCAGAGATT GGCTACGCATT TGCTTCGTAC GTAGAGAGGCG GGCAACACGAT rpAsnThrIl AAAAATGGAA gLysMetGlu CTTCTTCGTC LeuLeuArgL	50 GACACCTCTGG GAAAATGTTG GAAAATGTTG ACTCAGAAGAC CATGAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAG GCTGTGATTGT TGCACAAGGAG AGATCACTGT ACATTGACATG AGATCACTGT ACATTGACATG GCTGTTAACTGC AGATCACTGT CCCACAGGAGA AGATCACTGT CCCACAGGACAGA AGATCACTGC AGATCACTGC CCCACAGGACAGA AGATCACTGC AGATCACTGC CCCCACAGGACAGA AGATCACTGC CCCCCACAGGACA AGATCACTGC CCCCCACGCAC CCGCTCGCCAC CCGCACCGCA	60 I ATATAATCAA GGGATTATCAA GGGAATTATCA AGAGATTTGTC alAspleuSe CCTACACCCCG TLEUHISPIC CCTACACCCGGTAA GAACGTGAAATATCCC GTAAATATCCC CGTTGCAGACA GAACGTGTTTT AAGAAGGTGTTTT AAGAAGACCGT TTATTGAAGT GATCTCCTTCC SerIleGlnA TAAAAAGAAG SerIleGlnA TAAAAAGAAG SerIleGlnA	70 J GTATAAGAT TGTACTAAT GTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGGATGC GTGACACGGA ACCGGATGC GTGACACGGA ACGATCACAT GGATCACAT GGATCACAT IaValHisM TATCTCGCC rIleSerPr GAGACCGCC GluThrArg	80 ATGTAACATG' GAGGTGAGATT CTGAGAATAAA TATACTACGG' gGlnLeuLeul GGGGTCTCTA(GlyValSerT) CAGAAAGAAGTC CTGATCGGCTCCTGA ATTGACTGCGG CTGGATCCGGT TGAAATTTGAC GCGATCGGAAGTC' etAlaLysVa CCATTTGTATC' TGGCGAAAGTC' etAlaLysVa	90 TGATTGATCGTA TGATATCGTA ATCGTATCGTA ATCGTATCGTA ATCGTATGGAT ProArgGlyP CCTGACTTAG' hr*** AATCACTGCTT GACAATGGTG AAAAACTATTI GACCCACTTCC AATTCCAAC' TTAGAATAGT TTGCGACCACG ICysArgProJ CATTCGGCGG LCysArgProJ CATTCGGCGGG' HisSerAlaVi EGGCGCATTCC euAlaHisSe	100 STGTAC CGTAAC SATCAT GGTAT GJYTyr ICGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTAC CAACTT TTTGTA AAGTAC CAACTT ProVal IGGGAA alGlyA CGATAT rASpl1
<pre>(D) 2.2 1 1 101 201 301 401 501 601 701 801 901 1001 1001 1201 1301 1401 1501 1601 1701 1801</pre>	kbp cDNA 10 I CAGGTGTAGAGAGCCC TAATGAGGTGGATATATGCT ACTCAGGTGTATATGCT ACTCAGGTGTATATGCT ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATTCATC CCGAACGGATACTGC ACATGCCGATACTGC ACATGCCGATACTGC ACAGGATAATGGCA CCGGTGTGACGCTCATG ArgCysAspValMet ATATAGGTAATAGGA	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACGGATT GACATATCT GCGGTGTACC GACAACTAGGAGGT CTGTATTGCG AACGAGAGCT TCTCCAATTCG TATAAGTCTGT TCTCAATTGG ACATGCAGTCG TTTAGGTTTGG PheArgPheG AAAATCATAC ysAsnHisTh	30 CATGTCTGTAC. GGTATCATCAGG TAGCTGAGGG TAGCTGAGGG CTGAGGTAATG GCCGCGGGGAAG CATATATTCCC. DUTYTILEPro CCCCTCAAAA ATGCTTTGTGG TACCATCCGT CATGAGGTTCT GGCAGGCATCT GCCAGGCATGTAG GCCAGGCATGT MetT CGCCGGTGATAG CUHISVALAR	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIleL AAAATCGAAT GGTTCGATACA GGCAGAGATTTGG GGCAGAGATTTGG GGCAGAGATTTGG GGCAGAGATTTGG GGCACGAT TPASNThrII AAAAATGGAA gLysMetGlu CTTCTTCGTC LeuLeuArgL	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTTGGT TTTCGCAGAG GCTGTGATTAA GCTGTGATACTGC AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT CCCACAAGAGA AGATCACTGT AGATCACTGC CCCACAGGACAGA AGATCACTGC AGATCACTGC CCGTTAACTCC LeuLeuThrL (P) TGCGATCGCAA	60 I ATATAATCAA GGATGAATG GGCACAATGAA GGGGATTATCC TAGATTTGTC alAspleuSe CCTACACCCG TLeuHisPro TACCCAGGTAA GGAAGGTGTTTT AGAAGGTGTTTT AGAAGAGGC TTATTGAAGT GATCTCCTTCC SerIleGlnA TAAAAAGAAG serIleGlnA TAAAAAGAAG SerIleGlnA	70 GTTATAGAT TGTACTAAT GTACTAAT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGTGAT AGTGCGTGAT GCGTGACACG TCAACGGTA AGTGTGTATA GGATCACAT GGATCACAT IaValHisM TATCTCGCC GluThrArg (DC)	80 ATGTAACATG: GAGGTGAGAT: CTGAGAATAAA TATACTACGG? QGInLeuLeul GGGGTCTCTA(GIyValSerT] CAGAAAGAAT: CTATTTGTGG? GACTGAATCCAG? GCGATCGGTCCTA; GGAGCGAAAGT? etAlaLysVa: CCATTTGTATC? TGGCAAAGT? etAlaLysVa: CCATTTGTATC? HISLEUTYT AAAGAAGAGGTC? LysGluGlyLe 	90 TTGATTGAATC TAATGTACGAA ATCGTATCGTG (TAGTATGGAAT ATCGTAGGUY) CCTGGCTTAG' ProArgGlyP CCTGACTTAG' ATTACCTGCT' GACAATGGT GACAATGTG GACAATGTG GACAATGTG GACAATGTG GACAATGGC GTATACCAGC ICYSArgProJ CATTCGGCGGC HisSerAlaVi EGGCGCATTCC EUALAHISSE:	100 STGTAC CGTAAC SATCAT GLYTYr ICGCCC heAlaH TAAGTG TCGAAC GACTT TTTGTA AAGTTC CCATCT TTTGTA AAGTAG CAACTT TTCGAG CAACTT ProVal IGGGAA alGlYA CGATAT rASpI1
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601 1701 1801 1901	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGAAGGCC AACGATAATATTGCT ACCCAGGTGTATATG ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTT isValCysArgAspS TGTCGGGAGACTTGA GGAGCTCGATTCATC CCAAACGGATACTGC ACATGCCGATACTGC ACATGCCGATACTGC ACAGGATAATGGCA CCGGTGTGACGCCAACAT CGGTGTGACGCTCATG ArgCysAspValMet ATATAGGTAATAGGA	20 TGTATTGGAT ATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTGT ATCCGGGTGTACC GACATATCT GCGGTGTACC GACATATCG ACAAGCAGAGCT GTTTAGGTTTGCA ACATGCAGTC TTTAGGTTTG PheArgPheG AAAATCATAC ysAsnHisTh	30 CATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAG CTGAGGTAATG CATGTTYTILEPro CCCCTCAAAA ATGCTTTGTGG TACCATCCGT CATGACATCGT CATGACATCGTA CAGCATGCATGT GCCGGGGTAT GCCGGGGTAT GCCATGCATGTAAG CUHISVALAT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAGTTCATTC GGATCCCAAT ATTTGTGAG GGCAGGATTTGG GGCAGGATTTGG GGCAGGAGTC GGCACGAT rpAsnThrIl AAAAATGGAA gLysMetGlu CTTCTTCGTC LeuLeuArgL TCGGAGCCGA	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGAC CATGAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAGA GCTGTGATTGT TGACGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGC CCCACAAGAGA AGATCACTGC AGATCACTGT ACATTGACATGC (P) TTATCGA TAA	60 I ATATAATCAA GGGATTATCAA GGGAATTATCA GGGAATTATCC alAspleuSe CCTACACCCCG TLEUHISPIC CCTACACCCGG CTACCAGGTAT CGTGCAGACG GAACGTGTTTT AAGAAGGTGTTTT AAGAAGACCG TTATTGAAGT GATCTCCTTCC SerIleGINA TAAAAAGAAG SerIleGINA TAAAAAGAAG SerIleGINA TAAAAAGAAG	70 GTTATAGAT TGTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGTGAT GTGCCGCG GTGACACGG ATCGACGGT GACACGGTA AGTGTGTATA GACTACACA laValHisM TATCTCGCC GluThrArg (P,C) AAGGGATA	80 ATGTAACATG: GAGGTGAGAT: CTGAGAATAAJ TATACTACGG? QGInLeuLeul GGGGTCTCTA(GIyValSerT) CAGAAAGAAT: CTATTTGTGG? GACTGAATCCAG? GCGATCGGTCCTGT GGAGCGAAAGT? etAlaLysVa: CCATTTGTATC? TGGCGAAAGT? etAlaLysVa: CCATTTGTATC? HISLEUTYT AAAGAAGAGGTC? LysGluGlyLd (M) ACTTCCTTGT	90 TTGATTGAATC TAATGTACGAA ATCGTATCGAC ATCGTATCGAC ATCGTATCGAC CCTGGTGGCT ProArgGlyP CCTGACTTAG' hr*** AATCACTGCTT GACAATGGTG AAAAACTATT GACCCACTTCC AATTTCCAAC TTGCCATCGCCG CATTCGGCGGC HisSerAlaV TGGCGCATTCC euAlaHisSe: TTACATAGAA	100 STGTAC CGTAAC SATCAT GGLYTY TCGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTTC CCATCT TTTGTA AAGTAC CAACTT TTCGAG CCTGTT ProVal TGGGAA alGlyA CGATAT rASpl1 TGTATG
<pre>(D) 2.2 1 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601 1701 1801 1801 1901</pre>	kbp cDNA 10 CAGGTGTAGGAGGCC TAATGAGGTGGAAGGCC AATGAGTTATTGCT ACCCAGGTGTATATGC ATGAGTTTTTCTGCT MetSerPheSerAla ATGTTTGCAGAGACT isValCysArgAspS TGTCGGGAGCTTGA GGAGCTCGATTCATC CCAAACGGATACTGC ACATGGCGTATACAGTC TATACGCACGATACTGC ACATGCGATAATGGCA TCTCTGTCGGTCGTCATG AGGGGTGAGACCTAA CGGTGTGACGCTCATG ArgCysAspValMet ATATAGGTAATAGGA snIleGlyAsnArgL AGTTTGCAGCCAATG eValCysSerGlnCy	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTTCACTGT ATCCGGGATT GACATATCT GCGGTGTACC GACAACTAGGAT CCGGTGTACC GTATATTGAG ACATGCAGTC TTTAGGTTTG PheArgPheG AAAATCATAC ysAsnHisTh CTTGGTCCGC	30 CATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGGAAG CTGAGGTAATG CATGTTYTILEPro CCCCTCAAAA ATGCTTTGTGG TACCATCCGT CATGATGAGGTTCT CATGAGGTTCT GCCAGGCATGTAG GCCAGGCATGT GCCAGGCATGT CGCAGGCATGTAG CUHISVALAr CTACGTCAGG CUHISVALAR	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAATCGAAT GGATCCCAAT ATTTGTGAG GGCAGAGATTTGG GGCAGAGATTTGG GGCAGAGATTTGG GGCACGAT TGCTTCGTAC TTAGAGAGCGA GLYSMetGlu CTTCTTCGTC LeuLeuArgL TCGGAGCCGA	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGAC CATGAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAG AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGT AGATCACTGC AGATCACTGT AGATCACTGC I GD TTTCGCATGATAACTGC (P) TTATCGATAAC	60 I ATATAATCAA GGGATTATCAA GGGAATTATCA GGGAATTATCC alAspleuSe CCTACACCCCG CTLEUHISPIC CCTACACCCG GTAAATATCCC GGAAGGTGTTTT AAGAAGGTGTTTT AAGAAGGTGTTTT AAGAAGAGCG TTATTGAAGT GATCTCCTTCC SerIleGlnA TAAAAAGAAG SerIleGlnA TAAAAAGAAG SerIleGlnA TAAAAAGAAG	70 GTATAAGAT TGTACTAAT GTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGCGATGC GTGACACGG ATCGACGCC TGCGTATAG TCACGGCA IaValHisM TATCTCGCC rIleSerPr GAGACCGCC GluThrArg (P,C) AAGGGGATA	80 ATGTAACATG' GAGGTGAGATT CTGAGAATAAJ TATACTACGG' gGlnLeuLeul GGGGTCTCTA(GlyValSerT) CAGAAAGAATT CTATTTGTGG' GACTGAATCCAG' TTATGTATTC' TGGAATCCAG' TGGCGAAAGT' etAlaLysVa CCATTTGTATC TGGCGAAGGTC' LysGluGlyLe (M) ACTTCCTTGT'	90 TTGATTGAATC TAATGTACGAA ATCGTATCGTG (TAGTATCGAC ATCGTATCGAC (TAGTATGAAT ProArgGlyP) CCTGACTTAG' ProArgGlyP) CCTGACTTAG' AATACACTGCT' GGCAATGGC GTATACCACACTTC CATTCCACCCCCG ICysArgProJ CATTCGGCGGG' HisSerAlaV FIGGCGCATTCC euAlaHisSe: TTACATAGAA'	100 STGTAC CGTAAC SATCAT GLYTYr ICGCCC heAlaH TAAGTG TCGAAC GACTT TTTGTA AAGTTC CCATCT TTTGTA AAGTAG CAACTT TTCGAG CAACTT ProVal IGGGAA alGlYA CGATAT rASpl1 TGTATG
(D) 2.2 1 101 201 301 401 501 601 701 801 901 1001 1201 1301 1401 1501 1601 1701 1801 1901 2001	kbp cDNA 10 CAGGTGTAGGAGGCCT AACGATAATATTGCT AACGATAATATTGCT ACTCAGGTGTATATG ATGAGTTTTTCTGCT Met SerPheSerAla ATGTTTGCAGAGACTTGA GGAGCTCGATCATCA GGAGCTCGATACTAGC GGAACGGATACTGC ACATGGCGTATACTGC ACATGCGTAACAACTC CAGGATAATAGTCA CGGTGTGACGCTCATG AGGGGTGAGACCTAA CGGTGTGACGCTCATG AGGGTGTGACGCTCATG ATATAGGTAATAGGA SnIleGlyAsnArgL AGTTTGCAGCCAATG eValCySSerGlnCy	20 TGTATTGGAT AATTTGAGAT TCAAAGGATT TTTTTTTTTA GAGATAGGAA GlulleGlyS CCTGCAATCT erCysAsnLe ATTCCGGATT GACATATCT GCGGTGTACC GACAACTAGGATT CCGGTGTACC GACAACTAGGATT CCGGTATTGCA ACATGCAGTC TTTAGGTTTGC PheArgPheG AAAATCATAC ysAsnHisTh CTTGGTCCGC SLeuValArg	30 CATGTCTGTAC. GTATCATCAGG TAGCTGAGGG TAGCTGAGGG CTGAGGTAATG CATGTTYTILEPro CCCCTCAAAA ATGCTTTGTGG TACCATCCGT CATGAGGTTTT GTGAGGTTCT GCCGGGGTGAT GCCATGCATGTAG CATGTTACAAAC AAGAATGTAA CAGCATGTAAG CUHISVALAT CTACGTCAGGGA CUUHISVALAT CTACGTCAGGGA CUUHISVALAT	40 AAGAGCTGTT GCTATAGATA GGCAAAGTCT TCACTGTTAG GGGTATGTTT GGIYMetPhe AAGTTCATTC LysPheIlel AAAAAAAAAA CAAGTTTGTGAG GGCAGAGATT GGCAGGAGTT GGCACGAAT GCTACGCAGT TCACTCGTAC TTAGTAGAAGT CTTCTTCGTC LeuLeuArgL TCGGAGCCGA 1eGIYALAA	50 GACACCTCTGG TGTAACATGTT GAAAATGTTG ACTCAGAAGAC CATGAGACAG HisGluThrV. TGGGACTTTCC euGlyLeuSe GAGGTTGGT TTTCGCAGAG GCTGTGATAT TGCACAAGGAC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC AGATCACTGC CCCACAAGGAC AGATCACTGC AGATCACTGC AGATCACTGC GIYArgCys CTGTTAACTCC LeuLeuThrL (P) TGCGATCGCA PTyrArg*** CGGAAATAC	60 I ATATAATCAA IGATTGAATG GGCACAATGAA GGGGATTATC TAGATTTGTC alAspleuSe CCTACACCCG rLeuHisPro TACCCAGGTA GTAAATATCCC GAACGTGTTT AGAAGAGGTGTTTT AAGAAGAGTGTTTT AAGAAGAGGC SerIleGlnA TAAAAAGAAG SerIleGlnA TAAAAAGAAG SerIleGlnA TAAAAAGAAG EuLysArgSe GAAGTGTCTC nLysCysLeu TACGGAAGCA	70 I GTATAAGAT TGTACTAAT GTACTAAT CCTTAACAA TATAGTTTT TCATCATCG rHisHisAr ATTTTGGTT IleLeuVal GTGCCCCCC ATGCGCGATGC GTGACACGG ATGCGCCCC GTGACACGGT AGTGCGTATAG TCACGGATAAC GGATCACAT IaValHisM TATCTCGCC CIUThrArg (P,C) AAGGGGATA	80 ATGTAACATG' GAGGTGAGATAAA TATACTACGG' QGINLeuLeul GGGGTCTCTA(GIyValSerT) CAGAAAGAATT CTATTTGTGG' GACTGAATCCGG' CTGATCGCTCCTTT TGAAATTTGACTGCGA GCGATCGGAAGT' etAlaLysVa CCATTTGTATTC TGGCGAAGGTC' LysGluGlyLa (M) ACTTCCTTGT' AAAGAAGATGC' AAACGATTGC'	90 I TTGATTGAATC TAATGTACGAA ATCGTATCGTG TAGTATCGAC ATCGTATCGAC TAGTATGGAT ProArgGlyPl CCTGACTTAG' ProArgGlyPl CCTGACTTAG' AATACACTGCT' GGCAATGGC GTATACCAGC GTATACCAGC CATTCGGCGGG' HisSerAlaV FGGCGCATTCC EUAlaHisSe: TTACATAGAA' GGAAGTACCGC	100 STGTAC CGTAAC GATCAT GCTAT GLYTYT TCGCCC heAlaH TAAGTG TCGAAC GACTTT AAGTAC CAACTT TTTGTA AAGTAC CCTGTT ProVal TGGGAA alGlYA CGATAT rASpl1 TGTATG GGTCCC



FIG. 3. Northern blot hybridization to confirm the origins of cDNAs. Poly(A)⁺ RNA isolated from CEF infected with oncogenic or attenuated MDV was hybridized with the PCR-generated probe representing the 5' end of the 1.69-kbp cDNA (probe I, lanes 1 and 2), a 3.8-kbp RNA-specific PCR-generated probe (probe II, lanes 3 and 4), an *SstII-XbaI* subfragment of the *Bam*HI-I2 fragment DNA probe (probe III, lanes 5 and 6), and a *XbaI-ClaI* subfragment of the *Bam*HI-I2 fragment DNA probe (probe swithin the *Bam*HI-H and -I2 fragments are shown. Sizes of the hybridizing RNAs are given at the left in kilobases. R, poly(A)⁺ RNA from RBIB; A, poly(A)⁺ RNA from attenuated strain of RBIB. The restriction sites shown are *Bam*HI (B), *ClaI* (C), *EcoRI* (E), *SalI* (Sa), *SstII* (Ss), and *XbaI* (X).

assisted translation of the class I 1.69-kb cDNA identified two ORFs, one encoding 63 amino acid residues (Fig. 4, ORF-A) and another encoding 64 amino acid residues (Fig. 4, ORF-C). ORF-A (nucleotides 141 to 329) encodes a putative protein with a molecular size of 6,962 Da. Analysis of the putative protein with the SOAP program as described by Klein et al. (11) suggested that ORF-A encodes an integral membrane protein which contains a transmembrane domain from residue 31 to 47 (MGVVSVWSHLLMLW SFI). A potential protein kinase C phosphorylation site was found at position 9 of ORF-A. Protein sequence homology searches did not reveal highly significant homology with the protein sequences deposited within the Protein Identification Resource (PIR) and Swiss-Prot data bases. However, limited homology of the ORF-A was found with the mouse T-cell lymphoma (TLM) oncogene protein. The N terminus of ORF-A exhibits a 50% homology (17 of 34 residues from residue 8 to residue 41) with the N terminus of the mouse TLM oncogene protein (Fig. 5). Limited homology of ORF-A with domain IX of the family of kinase-related transforming proteins (7) exemplified by the feline c-fes/fps proto-oncogene, feline sarcoma virus v-fes oncogene, and avian sarcoma virus v-fps oncogene was also found (Fig. 5).

ORF-C (nucleotides 644 to 835) encodes a putative protein with a molecular size of 6,891 Da. No significant homology between ORF-C and the protein sequences currently deposited in the PIR and Swiss-Prot data bases was found.

Analysis of the 1.5-kbp cDNA sequence indicated that this spliced mRNA contains an ORF encoding 75 amino acid residues and another ORF encoding 107 amino acid residues (Fig. 4, ORF-B and ORF-F, respectively). ORF-B utilizes the same start codon as ORF-A (contained in the 1.69-kb cDNA), but because of the splicing, ORF-B contains a



FIG. 4. Schematic representation of rightward transcriptional units within the BamHI-H and -I2 regions of MDV shown with pertinent transcription-associated signals and locations of oncogenic virus-specific RNAs. Transcriptional signal sequences are shown within horizontal boxes that connect the sequences to their locations on the genome. The two vertical boxes indicate the positions of the two copies of the 132-bp repeat. Locations of the three classes of rightward transcribed RNAs are indicated. Curves represent intron regions not present in the mature RNAs. Black boxes represent locations of the potential ORFs labeled A through F according to the locations of their start codons on the MDV genome. Box A represents ORF-A, encoding 63 amino acid residues; interrupted box B represents ORF-B, encoding 75 amino acid residues; box C represents ORF-C, encoding 64 amino acid residues; interrupted box D represents ORF-D, encoding 93 amino acid residues; box E represents ORF-E, encoding 66 amino acid residues; and box F represents ORF-F, encoding 107 amino acid residues. The dotted line indicates the predicted cDNA sequence from 3.0- and 3.8-kb RNAs. Several restriction sites are located on the map: BamHI (B), EcoRI (E), SstII (S), and XbaI (X).

different carboxyl terminus. ORF-B encodes a putative protein of 8,220 Da with two potential protein kinase C phosphorylation sites (positions 9 and 48), a myristylation site (position 32), and a casein kinase II phosphorylation site (position 35). ORF-B was predicted not to be a transmembrane protein. Since ORF-B shares some sequences with ORF-A, it also exhibits limited homology with the mouse TLM oncogene (Fig. 5). ORF-F (107 codons) is contained entirely within the second exon of the 1.5-kbp cDNA located in the BamHI-I2 region. ORF-F codes for a putative protein with a molecular size of 12,332 Da which contains three protein kinase C phosphorylation sites (positions 41, 74, and 80), a casein kinase II phosphorylation site (position 80), and a myristylation site (position 84). Limited homology (39%; 13 of 33 amino acid residues) of ORF-F with the feline c-fes/fps proto-oncogene and the feline sarcoma virus v-fes oncogene was found within domain VII of the kinase family of transforming proteins (Fig. 5).

Analysis of the 1.9-kbp cDNA sequence suggested that this class II spliced mRNA contained two ORFs: ORF-D, encoding 93 amino acid residues (Fig. 4), and ORF-F, which was also contained in the 1.5-kbp cDNA. ORF-D spans two exons, while ORF-F is contained entirely within the second exon.

Two potential ORFs were found within the 2.2-kbp cDNA: ORF-E, encoding 66 amino acid residues (Fig. 4), and ORF-F, which is present in both the 1.5- and the 1.9-kbp cDNAs. No significant homology between ORF-D or ORF-E

(A)			
8	QSRRSGTADHR-QHRPGMSLLGRPMGVVS-VWSHLL 4	11	ORF-A
11	QSSRGPAAGHRARHTDLLVLLESPAPVLSTVMMCLL 4	16	TLM ONCOGENE
			ODE D
8	QSRRSGTADHR-QHRPGMSLLGRPMGVVST-PDALT	#1	ORF-B
(B)			
`´440	PEALNYGRYSSESDVWSFGILLWEAFSLGAVPYANLSN	477	V-fns
22	DOMONI LODDMCUNICUMENT LMLW CETCOMODAACCON	EO	OPE A
22	PGHSLLGRPHGVVSVWSHLLHLW-SFIGGHSDAAGSCN	50	ORF-A
516	PEALNYGRYSSESDVWSFGILLWETFSLGASPYPNLSN	553	V-fes
	: :: ::: :: : : : :		
727	PEALNYGRYSSESDVWSFGILLWETFSLGASPYPNLTN	764	C-fes/fps
(C)			
160		400	V-for
409	RUBARC-NCLVTERNVERISDF GHSREARDGI	499	v-res
68	RLLRLRSQKCLETRKEGLAHSDIVCSQCLVRGI	100	OKE-F
680	RDLAARNCLVTEKNVLKISDFGMSREEADGI	710	C-fes/fps

FIG. 5. Homology of potential ORFs with the transforming protein. (A) Homology of ORF-A (1.69-kbp cDNA) and ORF-B (1.5-kbp cDNA) with mouse TLM oncogene protein. (B) Homology of ORF-A (1.69-kbp cDNA) with domain IX of the family of kinase-related transforming proteins, feline c-fes/fps proto-oncogene, and avian sarcoma virus v-fes and v-fps oncogenes. (C) Homology of ORF-F (1.5-, 1.9-, and 2.2-kbp cDNAs) with domain VII of the family of kinase-related transforming proteins, feline c-fes/fps oncogene. The amino acid sequence is shown with the single-letter code, with sequence numbers of the residues on each end of the sequences. Double dots indicate identical amino acid homology, and double vertical lines indicate conserved amino acid homology.

and the protein sequences contained in the PIR and Swiss-Prot data bases was found.

DISCUSSION

Previous results have shown that the 3.8-, 3.0-, and 1.8-kbp transcripts of the BamHI-H gene family are expressed by oncogenic MDV. These RNAs are not produced by attenuated MDV that has undergone amplification of the 132-bp repeat sequences located in the BamHI-D and -H regions (1). It has been reported on the basis of S1 mapping that the BamHI-H 1.8-kbp gene family transcripts produced in MSB-1 cells are products of splicing within the BamHI-H region, contain two exons, and exclude the 132-bp repeats (1). The nonspliced transcripts were not included in that report because S1 protection of nonspliced RNA could not be distinguished from S1 protection due to self-annealing of the probe. The existence of nonspliced transcripts within the BamHI-H region remains to be clarified by the cloning of cDNAs derived from BamHI-H gene family transcripts. In this study, we have undertaken the cloning of cDNAs derived from these oncogenic virus-specific transcripts in order to identify their precise transcription patterns and locations on the MDV genome. Also, availability of the cDNAs derived from these oncogenic virus-specific transcripts is useful for studying their biological significance in tumor induction by MDV. We have chosen CEF lytically infected with RBIB to construct the cDNA library for cloning cDNAs representative of the BamHI-H gene family because (i) transcription and the splicing pattern of the BamHI-H gene family in CEF lytically infected with RBIB are similar to those in iododeoxyuridine-treated MSB-1 cells according to S1 mapping analysis (unpublished data), and (ii) an abundance of BamHI-H gene family transcripts was produced in CEF infected with RBIB.

Our data from cDNA analysis indicated that two kinds of transcripts belong to the 1.8-kbp RNAs of the *Bam*HI-H gene family: (i) a nonspliced transcript (1.69-kbp cDNA) within the *Bam*HI-H fragment which contains two copies of

the 132-bp repeats and (ii) a spliced RNA (1.5-kbp cDNA) with exon 1 at the *Bam*HI-H region and exon 2 at the *Bam*HI-I2 fragment. The 1.69- and 1.5-kbp cDNAs share the same initiation site at nucleotide 741 of the *Bam*HI-H sequence, as was previously predicted by S1 mapping analysis (1). The 1.69-kbp cDNA terminated at nucleotide 2437 of the *Bam*HI-H sequence, as was previously predicted by S1 mapping (1), whereas the 1.5-kbp cDNA terminated at nucleotide 1449 of the *Bam*HI-I2 fragment (unpublished I2 sequence) through use of the polyadenylation signal (AT-TAAA) located at nucleotides 1463 to 1468 of the *Bam*HI-I2 fragment.

The 1.69-kbp cDNA derived from nonspliced RNA is identical to those cDNA clones of linear transcripts reported recently by Iwata et al. (9). A comparison of sequence data indicated a difference between amino acid 23 of ORF-C (Cys) and amino acid 23 of ORF-2 reported by Iwata et al. (9). This difference is likely due to the different MDV strains (RBIB versus Md5) used to isolate cDNAs. Cloning of these cDNAs demonstrated the existence of nonspliced transcripts within the BamHI-H 1.8-kbp gene family. So far, cDNAs which are derived from the spliced transcripts and have exon 2 located within the BamHI-H fragment have not been identified. However, S1 mapping analysis using a 5'-endlabeled single-stranded probe (as previously described in reference 1) prepared from 1.69-kbp cDNA to hybridize with poly(A)⁺ RNA isolated from iododeoxyuridine-treated MSB-1- and RBIB-infected CEF did confirm the presence of spliced transcripts of the BamHI-H 1.8-kbp gene family in both type of cells (data not shown). This S1 mapping analysis also excluded the possibility that the probes used in a previous study (1), which were prepared from the BamHI-H fragment of DNA from the GA strain of MDV that had been cloned and maintained for a long time in bacteria, had significant base changes which might give errors in S1 mapping. The cloning and sequencing of cDNAs derived from the spliced mRNAs which contain two exons located within the BamHI-H region are in progress. It should be noted that both spliced mRNAs (predicted by previous S1 mapping) and nonspliced mRNAs transcribed entirely within the BamHI-H region (identified by cDNA cloning) of the BamHI-H 1.8-kbp gene family contain the 63-amino-acid ORF (ORF-A).

Kawamura and colleagues (10) prepared sense and antisense oligonucleotides directed against the first splice donor sequence (nucleotide 1178 of GenBank sequence M26392) of the *Bam*HI-H 1.8-kbp gene family and added the oligonucleotides to MSB-1 cultures. They observed a rapid decrease in cell proliferation only when the antisense oligonucleotide, A1, was used. They also observed that antisense oligonucleotide A1 inhibited transcription of the *Bam*HI-H gene family, suggesting that A1 inhibition may be caused by some other novel mechanisms of antisense RNA.

Chen and Velicer (3) have described the isolation of several cDNAs derived from duck embryo fibroblasts infected with the GA strain of MDV. Their analysis revealed multiple bidirectional initiations and terminations of transcription in the *Bam*HI-D and *Bam*HI-H regions. It is therefore possible that transcripts from the *Bam*HI-H fragment of MDV DNA include mRNAs transcribed in both directions.

Since the 3.0-kbp mRNA could be detected by hybridization with the PCR-generated DNA probe representing the 5' end of the *Bam*HI-H gene family, the initiation site of the 3.0-kbp transcript is possibly identical to those of the transcripts from which class I cDNAs are derived. Precise genomic mapping of the 3.0-kbp transcripts awaits the isolation and characterization of full-length cDNAs. Meanwhile, we could not exclude a possibility that some class II cDNA clones are full-length cDNAs derived from the transcripts initiated at the first unit of the 132-bp repeats, because a TATAA-like sequence (TATTAAA) was found within the proper context of the MDV genomic DNA sequence upstream of the 5' end of this cDNA.

Sequence analysis of the putative proteins translated from ORF-A and ORF-B show a 50 or 44% homology, respectively, with the N terminus of the mouse TLM oncogene protein (12, 13). The TLM oncogene is a stage-specific transforming gene which is frequently altered during mouse T-cell leukemia and lymphoma by a single base mutation or by a more-complex gene rearrangement. MDV could induce T-cell lymphoma within 6 to 8 weeks in chickens. The significance of the limited homology of ORF-A and ORF-B with TLM oncoprotein in tumor induction by MDV is not clear, because no information about the functional domain of the TLM oncoprotein is available. It should be noted that ORF-A is likely a transmembrane protein, whereas ORF-B is predicted to be intracytoplasmic. Thus, the locations of these two proteins in MDV-infected cells would be different. The computer-predicted location and structural characteristics of ORF-A suggest that it might encode a receptor-like protein.

Protein sequence homology searches have shown 48% homology (15 of 31 residues) between ORF-A and domain IX of the *fes/fps* kinase-related transforming protein. However, the significance of this homology is unclear because (i) domain IX is not a catalytic domain of the family of kinase-related transforming proteins, and (ii) the invariable residues <u>SDVMxxGxxxxE</u>, which are characteristic of all kinase family transforming proteins and are present in domain IX, are not conserved within ORF-A. It should be noted that we cannot find 70% homology (23 amino acids) of ORF-A with domain IX of v-*fms*, a kinase-related transforming protein, reported recently by Iwata et al. (9).

ORF-F, in 1.5-, 1.9-, and 2.2-kbp cDNAs, shows 39% homology with domain VII of the *fes/fps* family of kinase-related transforming proteins. Domain VII is a catalytic domain of the kinase-related transforming protein. However, the biological significance of this homology between domain VII of *fes/fps* transforming protein and ORF-F is uncertain because of the absence of homology of ORF-A with the invariable residues DFGxxR of domain VII.

In this report, we have described the isolation and characterization of several cDNA clones from CEF infected with oncogenic MDV RBIB. These cDNA clones are representatives of rightward transcripts of the BamHI-H gene family. A number of ORFs in these cDNA clones have been identified; they may encode polypeptides which have homology with mouse TLM oncogene and the kinase family of transforming proteins. It is not known whether one or more genes of these transcripts, which are specifically expressed by oncogenic RBIB in lytically infected CEF, are expressed in MDVinduced chicken T-cell lymphoma cells, although transcripts of similar sizes have been detected in MDV-induced tumor cells. The isolation and characterization of these oncogenic virus-specific cDNAs will facilitate identification of the biological function of the BamHI-H gene family and aid our investigations into the oncogenic mechanism of MDV.

ACKNOWLEDGMENTS

We thank Patrick Lai for a critical review of the manuscript. This work was supported by grant GAM 8901763 from the U.S. Department of Agriculture and by grants RO1 CA31949 and CA 50523 from the National Institutes of Health.

REFERENCES

- 1. Bradley, G., M. Hayashi, G. Lancz, A. Tanaka, and M. Nonoyama. 1989. Structure of the Marek's disease virus *Bam*HI-H gene family: genes of putative importance for tumor induction. J. Virol. 63:2534-2542.
- Bradley, G., G. Lancz, A. Tanaka, and M. Nonoyama. 1989. Loss of Marek's disease virus tumorigenicity is associated with truncation of RNAs transcribed within *Bam*HI-H. J. Virol. 63:4129-4135.
- 3. Chen, X., and L. F. Velicer. 1991. Multiple bidirectional initiations and terminations of transcription in the Marek's disease virus long repeat regions. J. Virol. 65:2445-2451.
- Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium. Anal. Biochem. 162:156– 159.
- 5. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.
- Fukuchi, K., A. Tanaka, L. W. Schierman, R. L. Witter, and M. Nonoyama. 1985. The structure of Marek's disease virus DNA: presence of unique expansion in nonpathogenic viral DNA. Proc. Natl. Acad. Sci. USA 82:751-754.
- Glover, D. M. 1985. DNA cloning: a practical approach. IRL Press Ltd., London, England.
- 8. Hanks, S. K., A. M. Quinn, and T. Hunter. 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241:42-52.
- Iwata, A., S. Ueda, A. Ishihama, and K. Hirai. 1992. Sequence determination of cDNA clones of transcripts from the tumorassociated region of the Marek's disease virus genome. Virology 187:805-808.
- Kawamura, M., M. Hayashi, T. Furuichi, M. Nonoyama, E. Isogai, and S. Namioka. 1991. The inhibitory effects of oligonucleotides, complementary to Marek's disease virus mRNA transcribed from the BamHI-H region, on the proliferation of transformed lymphoblastoid cells, MDCC-MSB1. J. Gen. Virol. 72:1105-1111.
- 11. Klein, P., M. Kanehisa, and C. DeLisi. 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta 815:468–476.
- Lane, M. A., A. Sainten, K. M. Doherty, and G. M. Cooper. 1984. Isolation and characterization of a stage-specific transforming gene, *Thym-I*, from T-cell lymphomas. Proc. Natl. Acad. Sci. USA 81:2227-2231.
- Lane, M. A., and M. B. Tobin. 1990. Genomic sequence of the mouse oncogene tlm. Nucleic Acids Res. 18:3410.
- Maotani, K., A. Kanamori, K. Ikuta, S. Ueda, S. Kato, and K. Hirai. 1986. Amplification of a tandem direct repeat within inverted repeats of Marek's disease virus DNA during serial in vitro passage. J. Virol. 58:657-659.
- 15. Maray, T., M. Malkinson, and Y. Becker. 1988. RNA transcripts of Marek's disease virus (MDV) serotype-1 in infected and transformed cells. Virus Genes 2:49-68.
- McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA 74:4835–4838.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Schat, K. A., A. Buckmaster, and L. J. N. Ross. 1989. Partial transcription map of Marek's disease herpesvirus in lytically infected cells and lymphoblastoid cell lines. Int. J. Cancer 44:101-109.
- 19. Silva, R. F., and R. L. Witter. 1985. Genomic expansion of Marek's disease virus DNA is associated with serial in vitro passage. J. Virol. 54:690-696.
- Silver, S., A. Tanaka, and M. Nonoyama. 1979. Transcription of the Marek's disease virus genome in a nonproductive chicken lymphoblastoid cell line. Virology 93:127–133.
- Sugaya, K., G. Bradley, M. Nonoyama, and A. Tanaka. 1990. Latent transcripts of Marek's disease virus are clustered in the short and long repeat regions. J. Virol. 64:5773-5782.