

Human Polyserase-2, a Novel Enzyme with Three Tandem Serine Protease Domains in a Single Polypeptide Chain*

Received for publication, August 10, 2004, and in revised form, October 29, 2004
Published, JBC Papers in Press, November 9, 2004, DOI 10.1074/jbc.M409139200

Santiago Cal‡, Víctor Quesada, María Llamazares, Araceli Díaz-Perales, Cecilia Garabaya,
and Carlos López-Otín

From the Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología,
Universidad de Oviedo, 33006 Oviedo, Spain

We have cloned a human cDNA encoding a new serine protease that has been called polyserase-2 (polyserine protease-2) because it is the second identified human enzyme with several tandem serine protease domains in its amino acid sequence. The first serine protease domain contains all characteristic features of these enzymes, whereas the second and third domains lack one residue of the catalytic triad of serine proteases and are predicted to be catalytically inactive. This complex domain organization is also present in the sequences of mouse and rat polyserase-1, and resembles that of polyserase-1, which also contains three serine protease domains in its amino acid sequence. However, polyserase-2 lacks additional domains present in polyserase-1, including a type II transmembrane motif and a low-density lipoprotein receptor A module. Enzymatic analysis demonstrated that both full-length polyserase-2 and its first serine protease domain hydrolyzed synthetic peptides used for assaying serine proteases. Nevertheless, the activity of the isolated domain was greater than that of the entire protein, suggesting that the two catalytically inactive serine protease domains of polyserase-2 may modulate the activity of the first domain. Northern blot analysis showed that polyserase-2 is expressed in fetal kidney; adult skeletal muscle, liver, placenta, prostate, and heart; and tumor cell lines derived from lung and colon adenocarcinomas. Finally, analysis of post-translational processing mechanisms of polyserase-2 revealed that, contrary to those affecting to the membrane-bound polyserase-1, this novel polyprotein is a secreted enzyme whose three protease domains remain as an integral part of a single polypeptide chain.

Proteolytic enzymes or proteases are a diverse group of proteins that perform a common biochemical reaction, the hydrolysis of peptide bonds. Proteases were first described as gastric juice enzymes involved in the nonspecific degradation of dietary proteins. However, multiple studies have demonstrated that they

may also act as highly specific processing enzymes and perform a selective and limited cleavage of specific substrates (1). These proteolytic processing events are essential for the regulation of multiple events such as cell cycle progression, tissue morphogenesis and remodeling, cell proliferation and migration, ovulation, angiogenesis, hemostasis, apoptosis, and autophagy (1–6). Consistent with these diverse and essential roles of proteases in living organisms, structural changes in these enzymes or alterations in their expression patterns underlie many pathological conditions such as metabolic diseases, neurodegenerative disorders, cardiovascular alterations, arthritis, and cancer (7). This large and growing functional complexity of proteases in both normal and pathological conditions results from the presence in all organisms of multiple enzymes with the ability to catalyze proteolytic processing reactions. To date, more than 550 proteases and protease homologs have been annotated in the human, mouse, and rat genomes (7, 8) (www.merops.ac.uk; web.uniovi.es/degradome). The complexity of protease systems is also apparent in other model organisms such as *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana*, whose genomes encode a similar number of proteases as vertebrate genomes (9) (www.merops.ac.uk).

Recently, and as part of our studies aimed at characterizing mammalian degradomes (the complete set of proteases present in these organisms), we have identified and cloned a human liver cDNA encoding an unusual mosaic protease called polyserase-1 (polyserine protease-1) (10). This protein exhibits a complex modular architecture composed of a type II transmembrane motif, a low-density lipoprotein receptor A module, and three tandem serine protease domains. Interestingly, analysis of post-translational processing mechanisms of polyserase-1 revealed that it is synthesized as a membrane-bound protein that undergoes a series of proteolytic processing events to generate three independent serine protease domains. The complex mosaic architecture of polyserase-1 is conserved in the predicted sequence for its mouse and rat orthologs, although the putative functional advantages derived from this multidomain architecture are still unclear. Nevertheless, similar mosaic structures can also be found in serine proteases from other organisms. This is the case for ovochymase from *Xenopus laevis*, which also possesses three serine protease modules that are post-translationally released from a single polyprotein product (11). Likewise, oviductin from *X. laevis* or *Bufo japonicus* contains two serine protease domains that also undergo proteolytic processing events to generate independent units (12, 13). Furthermore, analysis of the genome sequence of *D. melanogaster* and *C. elegans* has also revealed the occurrence of some predicted genes with several serine protease modules, although it is not known whether they are an integral part of a single polyprotein or are post-translationally cleaved

* This work was supported by grants from the Comisión Interministerial de Ciencia y Tecnología-Spain, Gobierno del Principado de Asturias, Fundación la Caixa, and European Union (FP5, Cancer Degradome-FP6). The Instituto Universitario de Oncología is supported by Obra Social Cajastur-Asturias and Red de Centros de Cancer-Instituto Carlos III, Spain. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AJ627034, AJ784938, and AJ784939.

‡ To whom correspondence should be addressed. Tel.: 34-985-106282; Fax: 34-985-103564; E-mail: santical@uniovi.es.

(14, 15). Taken together, these findings suggest that the evolutionary design of polyproteases may represent an additional strategy for increasing the complexity of the degradome from different species (1). In this work, we have evaluated the possibility that the human genome could encode additional mosaic proteases similar to polyserase-1 and found a gene coding for a protein with three tandem serine protease domains that has been called polyserase-2. We report the cloning of a full-length cDNA for this protein as well as an enzymatic analysis of the recombinant protein produced in eukaryotic cells. We also analyze the tissue distribution of polyserase-2 and demonstrate that, contrary to the membrane-bound polyserase-1, this novel polyprotein is a secreted enzyme whose serine protease domains remain as an integral part of the initial protein product.

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases and other reagents used for molecular cloning were from Roche Applied Science. Double-stranded DNA probes were radiolabeled with [α - 32 P]dCTP (3000 Ci/mmol) from Amersham Biosciences, using a commercial random-priming kit purchased from the same company. Nylon filters containing polyadenylated RNAs from human tissues were from Clontech. PCR amplifications were performed with the ExpandTM High Fidelity PCR system (Roche Applied Science). Reverse transcription-PCRs were carried out using the Thermo-script reverse transcription-PCR system from Invitrogen.

Bioinformatic Screening of the Human Genome and cDNA Cloning—Human polyserase-1 cDNA sequence and the BLAST program were used to look for regions of the human genome that could encode new polyserine proteases. These bioinformatic searches led us to the identification of a region in chromosome 16p11.2 that showed a structural organization similar to that of the gene encoding polyserase-1 (10). Then, a PCR-based strategy was used to clone the full-length cDNA for this putative novel polyprotease. First, specific oligonucleotides derived from the identified genomic sequences were used to screen a panel of human cDNA libraries. The sequences of the designed primers were as follows: poly2-f1 (forward), 5'-TCTCCAGATATTGCCAGGGAT-3'; and poly2-r1 (reverse), 5'-TGTGTCTCAGTCTCCTCCCA-3'. All PCRs were performed in a GeneAmp 2400 PCR system from PerkinElmer Life Sciences for 40 cycles of denaturation (94 °C, 20 s), annealing (63 °C, 20 s), and extension (68 °C, 60 s). After cloning of the PCR-amplified products in pBlueScriptII (Invitrogen), the identities of the products were confirmed by nucleotide sequencing using the DR terminator TaqFS kit and the ABI-PRISM 310 automatic DNA sequencer (PerkinElmer Life Sciences). The 5'- and 3'-ends of the cloned cDNA were extended by successive rounds of rapid amplification of cDNA ends. Finally, the full-length cDNA was obtained by PCR amplification using primers ATGpoly2 (forward; 5'-ATGGCCCGGACCTGTCTCTC-3') and ENDPoly2 (reverse; 5'-TTAGTCTGGATCAGGAGAT-3'). The PCR conditions for this amplification were as described above, but with 280 s of extension. The sequences of the mouse and rat ortholog cDNAs were deduced using the human sequence as a query to search the available mouse and rat genome sequences (16, 17). Then, the deduced sequences were confirmed by reverse transcription-PCR amplifications using RNA extracted from mouse and rat liver, respectively.

Chromosomal Mapping and Gene and Protein Structure Analysis—Identified genomic sequences were mapped to a specific human chromosomal region through the Ensembl genome browser at www.ensembl.org. Gene analysis was carried out using different programs available at www.hgmp.mrc.ac.uk. Protein structural analysis was performed using different programs available at www.expasy.org. All the remaining bioinformatic analysis were carried out using the tools available at www.hgmp.mrc.ac.uk.

Northern Blot Analysis—Nylon filters containing poly(A)⁺ RNA of diverse human tissues were prehybridized at 42 °C for 3 h in 50% formamide, 5× saline/sodium phosphate/EDTA, 10× Denhardt's solution, 2% SDS, and 100 µg/ml denatured herring sperm DNA. Hybridization was performed with a radiolabeled 656-base pair BamHI-NheI fragment of polyserase-2 cDNA. After hybridization for 20 h under the same conditions used for prehybridization, filters were washed with 0.1× SSC, 0.1% SDS for 2 h at 50 °C and autoradiographed.

Construction of Expression Vectors and Western Blot Analysis—Full-length cDNA encoding polyserase-2 was first cloned between HindIII and XhoI of a modified pCEP4 expression vector (Invitrogen). To do this, the full-length cDNA encoding polyserase-2 was PCR-amplified with oligonucleotides 5'-TTAAAGCTTCCATGGCCCGGACCTGCTC-

C-3' and 5'-TCACTCGAGCTAGCTCTGGATCAGGAGATC-3', where the HindIII and XhoI sites are underlined. Additionally, two oligonucleotides (5'-CTAGCAGACTACAAGGACGACGATGACAAG-3' and 5'-CTAGCTTGTATCATCGTCTGCTCTGTAGTCTG-3') were used to introduce a FLAG epitope at the NheI site at position 1684, located between the second and third serine protease domains, in a position where the putative activation sites were not affected. The resulting vector, pCEP-pol2, was used to transfect HeLa, COS-7, and 293-EBNA cells using Lipofectamine reagent (Invitrogen). Two additional vectors lacking the activation domain present in the first module of polyserase-2 were prepared in the same modified vector indicated above. One of these constructs contained the first serine protease domain of the polyserase-2 (pCEP-pol2spd1) with a FLAG epitope at the C terminus, whereas the second construct contained the entire protein (pCEP-pol2AD). These new constructs were used for transfection experiments as described above. When indicated, tunicamycin at a final concentration of 1 µg/ml was added to the wells. Conditioned medium obtained from the different cell lines was dialyzed against phosphate-buffered saline buffer and concentrated about 10-fold using Microcon Filter Devices (Amicon). Alternatively, conditioned medium was precipitated with an equal volume of cold acetone. Then, the expression of the recombinant proteins was analyzed by Western blot using the anti-FLAG M2 antibody (Sigma). Blots were visualized with an enhanced chemiluminescence kit according to the manufacturer's instructions (Amersham Biosciences).

Immunocytochemical Analysis—After transfection, COS-7 or HeLa cells were fixed with 4% paraformaldehyde in phosphate-buffered saline. Then, the cells were permeabilized for 5 min with 0.2% Triton X-100 in phosphate-buffered saline. Blocking was performed with 1.5% fetal bovine serum in phosphate-buffered saline. Blocked slices were incubated for 2 h with several dilutions of the primary anti-FLAG antibody, followed by a 2-h incubation with a secondary fluorescein-conjugated sheep anti-mouse antibody (Amersham Biosciences). To visualize DNA in the cell nucleus, 4',6'-diamidino-2-phenylindole hydrochloride was added to the samples at 100 ng/ml. Then, the slides were coverslipped in the presence of Vectashield medium (Vector Laboratories) and imaged by fluorescence microscopy.

Enzymatic Analysis—Different batches of COS-7 and 293-EBNA cells were transfected with pCEP-pol2, pCEP-pol2spd1, pCEP-pol2AD, and empty pCEP vectors. After 48 h, the supernatants of these cultures were harvested, and the presence of similar levels of the recombinant proteins pol2, pol2spd1, and pol2AD was assessed by Western blot and by Coomassie Blue staining of the recombinant deglycosylated proteins, followed by densitometry. The putative enzymatic activity of these proteins was analyzed by using the synthetic fluorescent substrates *N*-t-Boc-Gln-Ala-Arg-AMC,¹ *N*-t-Boc-Gln-Gly-Arg-AMC, *N*-t-Boc-Ala-Phe-Lys-AMC, and *N*-t-Boc-Val-Leu-Lys-AMC. Routine assays were performed at 37 °C at a substrate concentration of 50 µM in an assay buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 1% Me₂SO. The fluorometric measurements were made in an LS55 PerkinElmer Life Sciences spectrofluorometer (λ_{ex} = 360 nm, λ_{em} = 460 nm). The fluorescence units were standardized with AMC (Sigma). The enzymatic activity of each sample was detected as the increase in fluorescence at 37 °C for different times. Conditioned medium from cells transfected with the empty vector was assayed as a control. These control proteolytic activities were subtracted from those of pol2-, pol2spd1-, and pol2AD-containing supernatants. For inhibition assays, the conditioned media were preincubated for 1 h at 37 °C with 2 mM synthetic proteinase inhibitors (Calbiochem), and the hydrolyzing activity of the resulting samples was determined by fluorometric measurements as described above.

RESULTS

Identification and Characterization of a New Human Polyserine Protease—We used the recently described sequence of polyserase-1 and the BLAST algorithm to look for regions in the human genome that could encode serine proteases with several protease domains embedded within the same polypeptide chain. This bioinformatic search allowed us to identify a region in chromosome 16p11.2 that contained coding information for three serine protease domains located very close to each other. After PCR experiments with RNA from human liver, we confirmed that these three protease domains were encoded by

¹ The abbreviation used is: AMC, 7-amino-4-methylcoumarin.

A

1 ATGCCCGGCACCTGCTCTCCCTTGTGATGCTTGTTCATCAGTCCCATCCCAGGAGCC
M A R H L L L L P L V M L V I S P I P G A

61 TTCCAGGACTCAGTCTCAGTCTACCCAGGAAGAECTGAAGATCTGGACTCGGGCCG
F Q D S A L S P T Q E E P E D L D C G C R

121 CCTGAGCCCTCGGCCGCTCGTGGGGGCTCAAACGCGCAGCGGGGCACTGGCCTTGG
P E P S A R I V G G S N A Q P G T W P W

181 CAAGTGAAGCTGCACCATGGAGTGGCCACATCTCGGGGGCTCCCTCATCGCCCTCC
Q V S L H H G G G H I C G G S L I A P S

241 TGGTCTCTCCGCTGCTACTGTTTCATGACGAATGGGACGCTGGAGCCCGGGCCGAG
W V L S A A C F M T N G T L E P A A E

301 TGGTCGTACTGTGGCGTGCATCCAGGACGGCCCTGGAGCGCGCACACCAGC
W S V L L G V H S Q D G P L D G A H T R

361 GCAGTGGCCCATCGTGGTGGCGGCACTACAGCAAGTGGAGCTGGGCGCGACCTG
A V A A I V V P A N Y S Q V E L G A L

421 GCCCTGCTGCGCTGGCCTCACCCGAGCTGGGCCCCCGGTGTGGCTGTCTGCTG
A L L R L L A S P A S L G P A V W P V C L

481 CCCCGCCTCACACCCTCGTGCACGGCACCCGCTGTGGCCACCGCTGGGAGAC
P R A S H R F V H G T A C W A T G W G D

541 GTCCAGGAGCAGATCCTGCTCCTCCCTGGTGTACAGGAAGTGGAGTAAAGGCTG
V Q E A D P L P L P W V L Q E V E L R L

601 CTGGGCGAGGCCACTGTCAATGTCTACAGCCAGCCGGTCCCTCAACCTCACTCTC
L G E A T C Q C L Y S Q P G P F N L T L

661 CAGATATTGCCAGGATGCTGTGTGCTGCTACCCAGAGGGCCGAGGACACTGCCAG
Q I L P G M L C A G Y P E G R R D T C Q

721 GGCGACTGTGGGGGCCCTGGTGTGTGAGGAAGCGGGCCCTGGTTCAGGCGAAGATC
G D G G P L V C E E G G R W F Q A G I

781 ACCAGCTTTGGCTTTGGCTGTGGACGGAGAAACCGCCCTGGAGTTTTCACTGTGGCT
T S F G F G C G R R N R P G V F T A V A

841 ACCTATGAGGCATGGATACGGGAGCAGGTGATGGTTTCAGAGCCTGGGCTGCCTTCCC
T Y E A W I R E Q V M G S E P G P A F P

901 ACCAGCCCGAGAAGACCAGTCAAGTCCCGAGGAGCCAGGAGGAGAACTGCACCAT
T Q P Q K T Q S D P Q E P R E N C T I

961 GCCCTGCCTGAGTGGGAGGCCCCCGGCGGAGGGGCTGGCCTGGGAGGCCAGGTG
A L P E C G K A P R P G A W P W E A Q V

1021 ATGGTGCAGGATCCAGACCTGCCATGGGGCGCTGGTGTCTGAAAGCTGGGTCTTGCA
M V P G S R P C H G A L V S E S W V L A

1081 CCTGCCAGCTGCTTTCTGGACCCGAACAGCTCCGACAGCCACCCCGGCACTCGACGCC
P A C F L D P N S S D S P P R D L D A

1141 TGGCGGTGCTGCTGCCCTCGCGCCCGCGGAGCGGTGGCGCGCTGGTGCAGCAC
W R V L L P S R P R A E R V A R L V Q H

1201 GAGAACGCTTCGTGGGACAACGCTCGGACCTGGCGCTGCTGCAGTGGCGACGCGCGTG
E N A S W D N A S L A L L Q L R T P V

1261 AACCTGAGTGGCGCTTCGCGGCCGCTGTGCCTACCCACCCGGAACACTACTTCCTGCC
N L S A A S R P V C L P H P E H Y F L P

1321 GGGAGCGCTGCCGCTGGCCGCTGGGGCCGCGGGGAACCCGCTGGCCAGGCGCG
G S R C R L A R W G R G E P A L G P G A

1381 CTGCTGGAGCGGAGCTGTTAGGCGGCTGGTGGCCACTGCCTGTACGGCCCGCAGGGG
L L E A E L L L G G W W C H C L Y G R Q G

1441 GCGGCACTACCGTGGCCGAGACCCCGGCGCTGCTGCCCTGCCTACCGAGAAAG
A A V P L P G D P P H A L C P A Y Q E K

1501 GAGGAGTGGGCGCTGCTGGAATGACTCGGTTGGAGCCTTTTGTGCCAGGAGGAGGGG
E E V G S C W N D S R W S L L C Q E E G

1561 ACCTGGTTTGGCTGGAAATCAGAGACTTCCCAGTGGCTGTCTACGTCCCGAGCCTTC
T W F L A G I R D F P S G C L R P R A F

1621 TTCCTCTGCAGACTCATGGCCATGGATCAGCCATGTACTCGGGGAGCCTACCTGGAG
F P L Q T H G P W I S H V T R G A Y L E

1681 GACCAGTAGCTGGGACTGGGCGCTGATGGGAGGAGACTGAGACACAGACTTGTCCC
D Q L A W D W G P D G E T E T Q T C L

1741 CCACACAGAGCATGGTGCCTGTGGCTGCGGCTGGAGGCTGCTCCAGTGGGGTCTG
P H T E H G A C G L R L E A A P V G V L

1801 TGGCCCTGGCTGGCAGAGTGCATGTGGCTGGTGTGATCGAGTCTGCATGGATCCTCTG
W P W L A E V H V A G D R V C T G I L L

1861 GCCCAGGCTGGTCTGGCAGCCACTCACTGTCTCAGCCAGGCTTACAACAGT
A P G W V L A A T S C L R S L P I C L

1921 CCTTACATTGAAGTGTATCTGGGCGGGCAGCCCTCCCTCCACAGGGCCACCAG
P Y I E V Y L G R A G A S S L P Q G H Q

1981 GTATCCCGCTGGTGCATCAGCATCCGGCTGCCCGAGCCTGGGACTCAGGCCCCCTG
V S R L V I S I R L P Q H L G L R P L

2041 GCCTCTGGAGTGAAGTCCCGGGTGGAGCCCTCCCATCAGCCCTGCCATCTGTCTC
A L L E L S S R V E P S P A L P I C L

2101 CACCCGGGGTATCCCCCGGGGCGAGCTGTGGTGTGGCTGGAAGAACCAG
H P A G I P P G A S C W V L G W K E P Q

2161 GACCCAGTCCCTGTGGCTGCTGTCTCCTCATTTGACACAACGAATCTGTACTGCCTC
D R V P V A A A V S I L T Q R I C D C L

2221 TATCAGGCATCCTGCCCGTGGAAACCTCTGTCTGTATGACAGAGGGGAGGAGAAC
Y Q G I L P P G T L C V L Y A E G S T E N

2281 AGGTGTGAGATGACTCAGCACCAGCCCTCTGTGCCAGATGACGGAAGGCTCTGGATC
R C E M T A P P L L C Q M T E G S W I

2341 CTCGTGGGATGGCTGTTCAAGGAGCGGGAGCTGTTGCTGCCATTGGTCTGGAAGAG
L V G M A V Q G S R E L F A A I G P E E

2401 GCCTGGATCCTCCAGCAGTGGGAGGCCAATCTCTGCCCGCCAGTGGCTCCACAC
A W I S Q T V G E A N F L P P S G S P H

2461 TGGCCACTGGAGGAGCAATCTGCCCCCAAGAATGGCCAAGGCTCGGGATCCCCG
W P T G G S N L C P P E L A K A S G S P

2521 CATGCAGTCTACTTCTGCTCTGCTGACTCTCTGATCCAGAGCTAA
H A V Y F L L L L T L L I Q S *

B

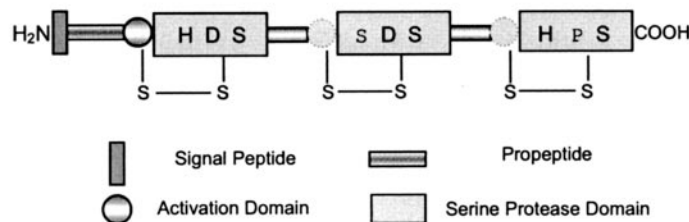


FIG. 1. Deduced amino acid sequence and domain organization of human polymerase-2 cDNA. A, the deduced amino acid sequence is shown in single-code letter. The three serine protease domains are underlined. The putative *N*-glycosylation sites are indicated with a black dot. The residues (His, Asp, and Ser) corresponding to the catalytic triad of the first serine protease domain are shaded in black. Equivalent residues of the second and third serine protease domains are shaded in gray, indicating the Ser (second serine protease domain) and Pro (third serine protease domain) residues that differ from those expected for an active serine protease. B, schematic representation of the domain structure of polymerase-2.

a single gene. Computer analysis of the identified cDNA indicated that it codes for a protein of 855 amino acids, with a predicted molecular mass of 93 kDa, containing nine potential *N*-glycosylation sites (GenBankTM accession number AJ627034) (Fig. 1A).

Further analysis of the cloned cDNA confirmed that the

identified protein possesses structural hallmarks characteristic of serine proteases (14, 18), although it also exhibits some particular features (Fig. 1, A and B). Thus, the sequence contains a recognizable signal peptide (positions 1–22), which predicts that this protein is directed to the secretory pathway. Following this hydrophobic sequence, there is a propeptide

A

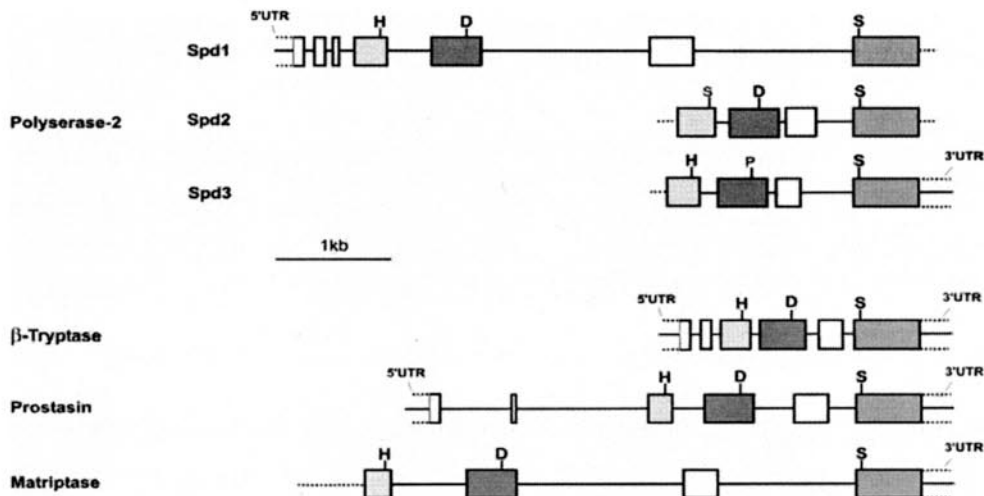
```

Pol2spd1 : IVGGSNRQPTWPNQVSLHHGG--HICGSLIAPSIVLSAHCPTMNTGL--EPAARVSVLLGVHSQDGLD-GAHTRAVAAIVVPANYSQV-ELGADLALLLSPASLGFVAVWVCLPRAS : 118
Pol2spd2 : ---KAPRPQAWFWEAQMVMVPG---SRPCHGALVSESHVLAAPACF-----LDPNSSDSPPRDLDAWRVLLPSRPAERVARLV--QHBNASWDNASDLALLQLRTFVNLSAASRFVCLRHPE : 109
Pol2spd3 : LRLEAAPVGLWFWLAEVHVAG---DRVCTGILLAPGMVLAATHCV-----LRPGSTTVFYIEVYLGRAGASSLPQHQVSRVLSIRLPQHLGLRFPALLESSRVEPSPSALPCLHPAG : 115
Prostasin : ITGSSAVAGQWPNQVSLITTEGV--HVCGSLVSEQWVLSAAHCPPSEHKK--EA---YEVKLGAKHLDYSYS-DAKVSTLKDIIIPHSYLQE-GSQGDIALLCCLSRPITFRYIRIICLPAAN : 115
Matriptase : VVGGTDADBESEWFWQVSEHALS-QCHIIGASLISPNLVSAAHCYIDDRGFRYSDPQTWTFALGLHDQSRSAPOVQERRLRIISHEFFNDF-TFYDIALLELEKFAEYSSMVRICLFDAS : 122
β-Tryptase : IVGGQEAFRSKFWQVSLRVHSPYWMHFCGGSIIHPQWVLAHCVGPDVKD--LAA---LEVQLREQHLIYQ-D-QLLPVSRIVHPQFYTA-GIGADIALLELEEEVNVSHVHTVTLFPAS : 116
Pol1spd1 : IVGGMEASPFSEFPWQASLRE-N--KEHFCGAAIINARMLVSAHCPE-----FQDPTKVAIVGATYLSGSEASTVRAQ-VVQVVKHLYNAD-TADFDVAVLELTSPLFFRHIOVCLPAAT : 115
Pol1spd2 : VVGGFGAASSEVFWQVSLKE-θ--SRHFCGATVVDGRMLLSAAHCPEH-----TKVQVQR-AHLTASLLGLGGSVVKIG-LRVLHLYNPG-TLDFLAVLELASELAFNKYIQVCLLAI : 114
Pol1spd3 : IVGSAAGRSEWFWQVSEMLR--REHRCGAVVEAERMLLSAAHCPEV-----YGDPKWAAFLG--TFPLSGAGQLERVARIYKHFFNYLY-TLDYDVALLELAGVRRSLVRLPCLPEPA : 114

Pol2spd1 : HRFVHGATCAWATGNGDQADPLPLPWLQGEVELRLLEGEATCCCLYSQPG-PFNLTQLPLPGLMAGYPERRRTCCQDSGGPLVC-EEGGRWFQAGITSFGGCGRRNRPGVFTAVATYEAWIR : 241
Pol2spd2 : HYFLPGSRCLRANRGERPALGPGAL----LEAELGGWCHCLYGRQGAAVPLPGDPPHALCPAYQEKVEVGSQWNSRMSLLCO-EGGTWFLAGIRDFPSGCLRPR--AFF-PLQTHGPWIS : 225
Pol2spd3 : --IFPGASCVLGNKPEQDRVPA-----AAVSLIQRICDCLY-OGIL-----PPGTLCLVYAEQENRCMETSAPFLLCQMTBGSMLVGM----AVQGRS--ELFAAIGPEEAWIS : 215
Prostasin : ASFFNGLHCTVTVGNGHVAFSVLLTPKFLQLEVLISRETCLNLYIDA-KPEEPHFVQDMVVCAGYVEGKDACCGDSGGPLSC-PVEGLWYLTGIVSWGDAQCARNRPGVYTLASSYASWIQ : 238
Matriptase : HVFFAGKAIWVTGNGHTQYG--GTGALIQQKGEIRVINQTTCCENLLPQQ-----ITFRMCMVGFSLGGVDSGGPLSSVEADGRIFQAGVSVWGDGCAQRNPKGVYTRLPFRDNIK : 236
β-Tryptase : ETFFPGMPCWVTGNGDNDNERLPPFPFLKQVVPIMENHICDAKYHLGAYTGDDVIRVRDMLCAG--NTRDSCCGDSGGPLVC-KVNGTWLQAGVSVWGGCAQPNRPGIYTRVYTYLDWIH : 238
Pol1spd1 : HIFFPKCKLISGNGYKEDF-LVKPEVLQKATVELLDQALCASLYGHS-----LTDRMVAGYLDGKVDSCGGSGGPLVCEEPSGRFFLAGIVSWGIGCAEARRPGVYARVTRLRDWIL : 230
Pol1spd2 : QKFFVGRKCMISGNGNTQEGN-ATKPELLQKASVGIIDQKTCVSLYNFS-----LTDRMICAGFLEKVDSCGGSGGPLACEEAPGVYLAGIVSWGIGCAQVKKPGVYTRITRLKGWIL : 229
Pol1spd3 : PRPFDTRCVITGNGSVREG--GSMARQLKAAVRLLEQTCTRRFYVQ-----ISSRMLCAGFPQGGVDSGSDAGGPELACREPSGRVVLTVTSWYGGCRPHFPVGVYTRVAAVRCWIG : 228

Pol2spd1 : EQVMGSEPGPAFF----- : 254
Pol2spd2 : HVTGAYLEDQLA----- : 238
Pol2spd3 : QTVGEANFLPFGSPHWPFGGSLNCPPELAKASGSPHAVYFLLLLTLIIQS----- : 266
Prostasin : SKVTELQPRVFPQTESQFDSNLCGSHLAFSSAPAQGLLRPILFLPLGLALGLLSPWLSH : 299
Matriptase : ENTGV----- : 241
β-Tryptase : HYVPKKP----- : 245
Pol1spd1 : E----- : 231
Pol1spd2 : E----- : 230
Pol1spd3 : QHIQE----- : 233
    
```

B



C

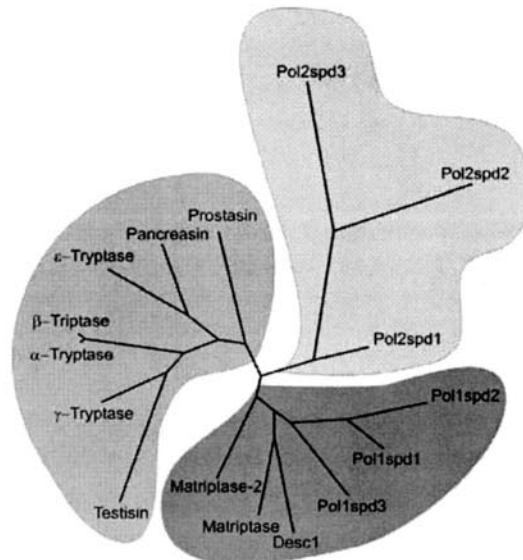


FIG. 2. Relationships between polymerase-2 and other serine proteases. A, comparison of the primary structure of the three serine protease domains of polymerase-2 with other related proteins. Active site amino acids are marked with a black dot. Pol2spd and pol1spd indicate the different serine protease domains of polymerase-2 and polymerase-1, respectively. B, organization of the human polymerase-2 gene and

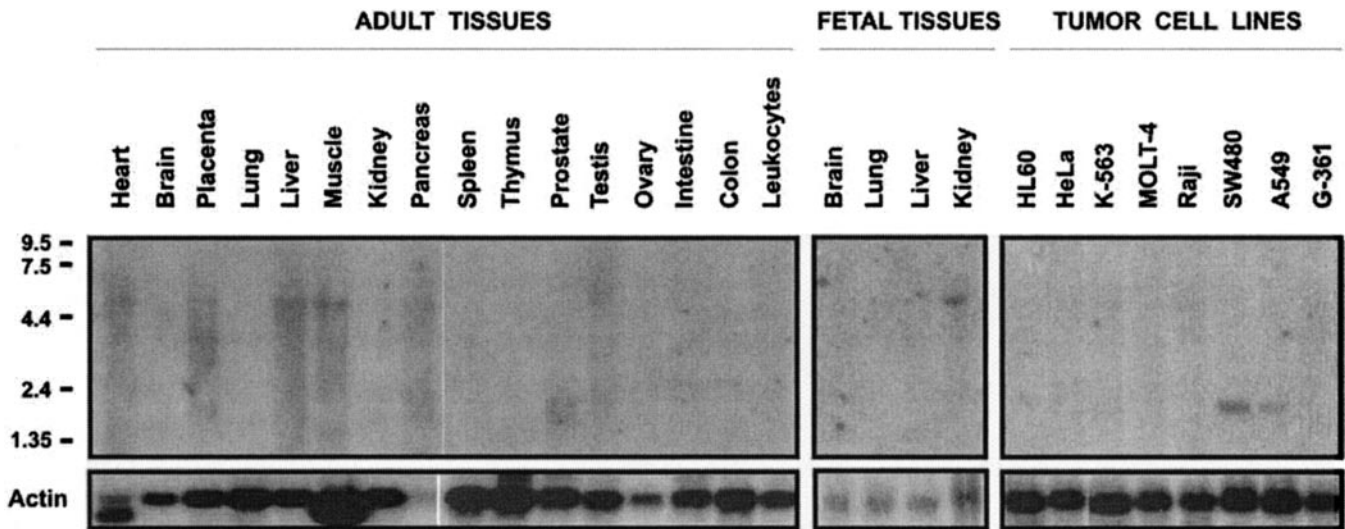


FIG. 3. Analysis of polyserase-2 expression in human tissues and tumor cell lines. Approximately 2 μ g of polyadenylated RNA from the indicated tissues or tumor cell lines was hybridized with a probe specific for polyserase-2. The positions of the RNA markers are shown. The filters were subsequently hybridized with an actin probe to ascertain the differences in RNA loading among the different tissues.

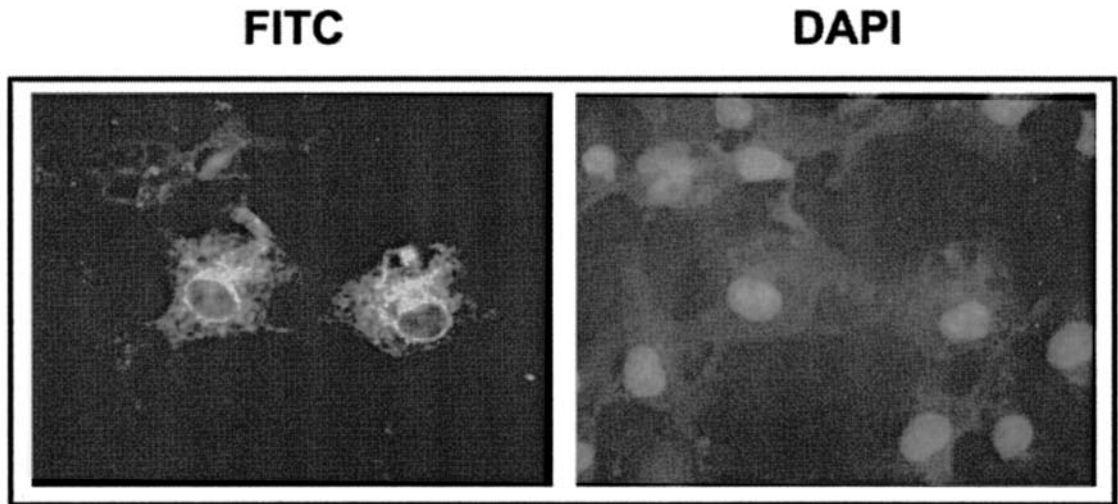
domain that ends in a conserved Arg residue (position 46), which corresponds to the activation site found in most members of this family of proteolytic enzymes (19). This site is followed by the first serine protease domain (positions 47–289) that includes the catalytic triad required for the activity of the serine proteases. This triad comprises the residues His⁸⁷, Asp¹³⁹, and Ser²⁴³. This last residue is included within the conserved motif Gly-Asp-Ser-Gly-Gly (positions 241–245) characteristic of serine proteases. Following the first serine protease domain, two additional serine protease modules can be recognized. However, these additional domains are preceded by sequences unrelated to the conserved activation motif present in the first domain of the identified polyprotease as well as in most serine proteases. In addition, the second and third serine protease domains contain some remarkable changes in residues that would conform their catalytic triad. Thus, the second protease domain of the polyprotease (positions 327–552) displays the characteristic Asp and Ser residues of the triad at positions 410 and 510 but lacks the His residue, which is substituted by a Ser residue at position 363. Likewise, the third serine protease domain contains conserved His and Ser residues (at positions 630 and 766, respectively) but lacks a conserved Asp that is replaced by a Pro residue (position 679). It is also noteworthy that the putative catalytic Ser residues present in the second and third domains of the polyprotease are found within regions (Asn-Asp-Ser-Arg-Trp and Met-Thr-Ser-Ala-Pro), which do not resemble the consensus sequence Gly-Asp-Ser-Gly-Gly characteristic of serine proteases. Based on these structural characteristics, we propose that the second and third serine protease domains of the polyprotease are catalytically inactive. It is also remarkable that the first catalytic domain exhibits some additional features of serine proteases (Figs. 1 and 2). Thus, the six Cys residues involved in the formation of three disulfide bonds in the catalytic region (Cys⁷²-Cys⁸⁸, Cys²⁰⁶-Cys²²⁸, and Cys²³⁹-Cys²⁶⁷) are conserved in the first protease domain of the polyprotein. The second domain also contains six conserved Cys residues (Cys³⁴⁸-Cys³⁶⁴, Cys⁴⁷²-Cys⁴⁹⁴, and Cys⁵⁰⁶-Cys⁵³⁴), whereas the third

domain maintains five of these residues (Cys⁶¹⁵-Cys⁶³¹, Cys⁷³⁷-Cys⁷⁵¹, and Cys⁷⁶²) but lacks the last Cys potentially involved in the formation of the third disulfide bond. As in other serine proteases, a fourth disulfide bond is predicted to be formed between Cys³⁸ located at the propeptide of the first protease domain and Cys¹⁷³ of the catalytic region of this domain; equivalent disulfide bonds should be formed between Cys³²⁵ and Cys⁴⁴⁴ of domain 2 and Cys⁵⁸⁸ and Cys⁷¹¹ of domain 3. The formation of the first of these predicted disulfide bonds would imply that the catalytic domains of the polyprotease should still remain linked to the main polypeptide chain even after cleavage at the activation site. All these structural features are also conserved in the sequence of the putative mouse and rat orthologs, whose sequence was deduced by using the identified human sequence as query and then amplified by reverse transcription-PCR using RNA from mouse and rat liver (GenBankTM accession numbers AJ784939 and AJ784938). The percentage of identity between the human polyprotein and its predicted mouse and rat orthologs is 80.2% and 79.5%, respectively. The genes encoding the mouse and rat polyproteins are located in chromosomes 7F4 and 1p36, within regions syntenic to chromosome 16p11.2, where the human gene is located. In summary, and according to this structural analysis, we can conclude that the cloned human, mouse, and rat cDNAs encode a novel polyprotein with three serine protease domains, which we have tentatively called polyserase-2.

Relationship of Polyserase-2 to Other Serine Proteases—The predicted amino acid sequences of each of the three serine protease domains of polyserase-2 exhibit a significant degree of identity with different enzymes belonging to this catalytic class. Thus, the first domain shows the highest percentage of identity with γ 1-tryptase (42%) (20), pancreasins (42%) (21), different members of the type II transmembrane serine proteinase family such as matriptase (40%) and matriptase-2 (40%) (22–24), and the three serine protease domains of polyserase-1 (41%) (10). The second domain is most closely related to matriptase-2 (33% identities), prostaticin (30%) (25), ϵ -trypsinases (29%) (26), and the three serine protease domains of

comparison with other serine protease genes. The relative positions of the exons are indicated by boxes. In the case of matriptase, only the exon/intron organization of its serine protease domain is represented. *H*, *D*, and *S* refer to positions of the codons that encode the catalytic triad amino acids in each protease gene. *C*, phylogenetic tree of different human serine proteases related to polyserase-2. The analysis was performed using the serine protease domain of the different proteins and the Phylogeny Interface Environment program supplied by the Human Genome Mapping Project.

A



B

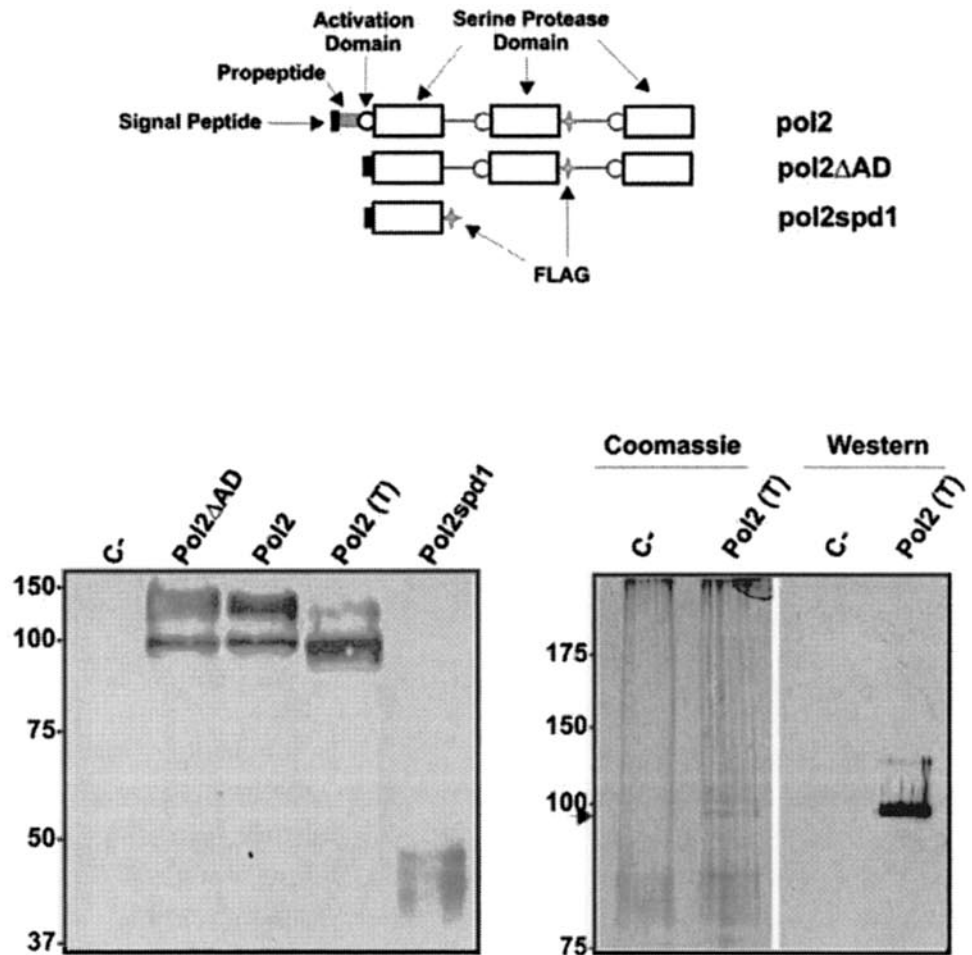


FIG. 4. Immunocytochemical detection and analysis of recombinant polyserase-2 expression in human cell lines. A, image captured by fluorescence microscopy of COS-7 cells transfected with pCEP-pol2 and incubated with an anti-FLAG antibody followed by incubation with a secondary fluorescein-conjugated antibody (*FITC*). This signal was not detected in cells transfected with an empty vector. The observed cytoplasmic fluorescence indicates that polyserase-2 is not a membrane-anchored protease. DNA in the cell nucleus was visualized with 4',6'-diamidino-2-

polyserase-1 (29%) (10). Finally, the third domain also shows a similar percentage of identity with these same proteins (30% with prostatic, 29% with matriptase-2, 29% with ϵ -trypsin, 28% with γ 1-trypsin, and 27% with the serine protease domains of polyserase-1). Alignment of the three protease domains of polyserase-2 with some representative members of these serine proteases confirmed the common features shared by these enzymes, and the sequence conservation around the amino acids forming the catalytic triad of the first protease domain was particularly remarkable (Fig. 2A).

Analysis of the exon-intron organization of the catalytic region of the first serine protease domain showed that it is similar to that found in the equivalent region of matriptase and matriptase-2 genes, as well as to that corresponding to the two active domains of the polyserase-1 gene (Fig. 2B; data not shown). By contrast, the genomic region that comprises the initial exons encoding the signal sequence and the propeptide of polyserase-2 is similar in size to that found in the α/β -trypsin genes, with the exception that three coding exons are found in the case of polyserase-2, but only two exons are present at the equivalent region of α/β -trypsin genes (Fig. 2B). The γ -trypsin or pancreatin genes also contain two coding exons, but they are embedded within much longer genomic sequences. However, exon-intron organization of the second and third domains of polyserase-2 resembles that found in the γ -trypsin and pancreatin genes, as well as that found in the third protease domain of polyserase-1 (Fig. 2B; data not shown). The phylogenetic tree analysis shown in Fig. 2C confirmed that the three serine protease domains of polyserase-2 are closely related to each other and that they are only distantly related to members of both the type II transmembrane serine proteinase and trypsin/pancreatin subfamilies of serine proteinases.

Tissue Distribution of Polyserase-2—Northern blots containing poly(A)⁺ RNAs from a variety of fetal and adult human tissues were used to determine the pattern of polyserase-2 expression (Fig. 3). Using a specific probe for polyserase-2, a band of \approx 5 kb was detected in fetal kidney as well as in adult skeletal muscle, liver, placenta, and heart. An additional polyserase-2 transcript of \approx 2.2 kb and of lower intensity could also be detected in placenta and prostate. This transcript was also observed in the human tumor cell lines SW480 and A549, which were derived from colon and lung adenocarcinomas, respectively. The major band could correspond to a transcript for the full-length polyserase-2 cDNA, whereas the minor transcript could derive from an alternative splicing event and result in a protein lacking one or two of the serine protease domains. Consistent with this, the conceptual translation of the human expressed sequence tag BM768456 indicates the existence of an alternative splicing in exon 13 of the gene. The corresponding full-length cDNA would lack the region comprised between guanine 1981 and guanine 2167 (Fig. 1). The consequence of this splicing event is a frameshift that causes the introduction of a premature stop codon at the beginning of the third domain, resulting in a protein with only two complete serine protease domains. Bioinformatic analysis using different programs available at the NIX tool at www.hgmp.mrc.ac.uk/allowed us to predict a polyserase-2 transcript of \approx 5.5 kb. Nevertheless, several putative polyadenylation sites are predicted in the polyserase-2 gene, suggesting that different mRNA transcripts can be generated for this gene.

Production of Recombinant Polyserase-2 in Transfected Human Cells—COS-7, 293-EBNA, and HeLa cells were transfected with the plasmid pCEP-pol2, and an anti-FLAG antibody was used to immunolocalize the polyserase-2 recombinant protein. As can be seen in Fig. 4A, the fluorescent signal detected in COS-7 cells is predominantly cytoplasmic, and there is no staining at the cell surface. Similar results were obtained in 293-EBNA and HeLa transfected cells (data not shown). In all cases, some cells exhibit an eccentric perinuclear immunofluorescence, which is indicative of the presence of these proteins in the endoplasmic reticulum or in the Golgi apparatus. It is remarkable that the fluorescent signal can be detected only if the polyserase-2 transfected cells have been previously permeabilized but is absent if samples were not treated with Triton X-100 (data not shown). This situation clearly differs from that observed in polyserase-1, in which the corresponding fluorescent signal can be detected without permeabilization of the cells, as a consequence of its membrane localization (10). According to this finding, we conclude that polyserase-2 is not a membrane-anchored protein and that it is likely secreted to the extracellular medium.

To further evaluate this possibility, conditioned medium of cells transfected with pCEP-pol2 and pCEP-pol2 Δ AD vectors was analyzed by Western blot using anti-FLAG monoclonal antibodies. As shown in Fig. 4B, two immunoreactive bands of \approx 100 and 130 kDa were detected in cells transfected with the polyserase-2 constructs. These bands were absent in cells transfected with the empty vector. To determine whether the presence of both polyserase-2 immunoreactive bands could be due to differences in the degree of *N*-glycosylation, transfected cells were treated with tunicamycin and then analyzed by Western blot as described above. As can be seen in the *left panel* of Fig. 4B, a single major band of \approx 100 kDa was detected for polyserase-2 in the presence of this inhibitor of *N*-glycosylation. SDS-PAGE analysis of this conditioned medium from transfected cells treated with tunicamycin also showed the presence of a 100-kDa band that was absent in the medium of cells transfected with the empty vector (Fig. 4B, *right panel*). These results strongly suggest that this enzyme is partially glycosylated and secreted outside the cell, remaining soluble in the extracellular medium. The expression of the first serine protease domain was also analyzed by transfection with the pCEP-pol2spd1 vector (Fig. 4B). Western blot analysis using an anti-FLAG monoclonal antibody revealed the presence of a wide and diffuse immunoreactive band slightly higher than 37 kDa, the expected size for this recombinant truncated protein plus the epitope. This finding again suggests that this first serine protease domain is glycosylated, as can be inferred from the number of potential *N*-glycosylation sites predicted in this region of the polyserase-2 amino acid sequence.

Enzymatic Activity of the Recombinant Polyserase-2—To investigate the enzymatic activity of human polyserase-2, we used about 4 ng of the recombinant proteins detected in the conditioned medium of the different transfections (Fig. 5B, *inset*; data not shown). Only the two constructs lacking the prodomain, pol2spd1 and pol2 Δ AD, showed measurable activities against a panel of synthetic fluorescent peptides commonly used for assaying serine proteinases. Thus, pol2spd1 hydrolyzed the peptides *N*-*t*-Boc-Gln-Ala-Arg-AMC and *N*-*t*-

phenylindole hydrochloride (DAPI). *B*, expression of the different recombinant proteins employed in this work. *Left panel*, Western blot of dialyzed and concentrated medium conditioned by COS-7 cells transfected with a plasmid encoding FLAG-tagged polyserase-2 (pol2), polyserase-2 lacking the activation domain (pol2 Δ AD), and the first serine protease domain lacking the activation domain (pol2spd1). Pol2(T) indicates treatment with tunicamycin, and *C*- indicates conditioned medium from cells transfected with the empty vector. The concentration of the SDS-PAGE gel was 12%, and the molecular mass markers (in kDa) are indicated on the *left*. *Right panel*, 8% Coomassie Blue-stained SDS-PAGE gel of 600 μ l of conditioned medium precipitated with acetone and showing a band corresponding to deglycosylated polyserase-2, indicated by an *arrowhead*. Polyserase-2 identity was confirmed by Western blot analysis with an anti-FLAG M2 antibody.

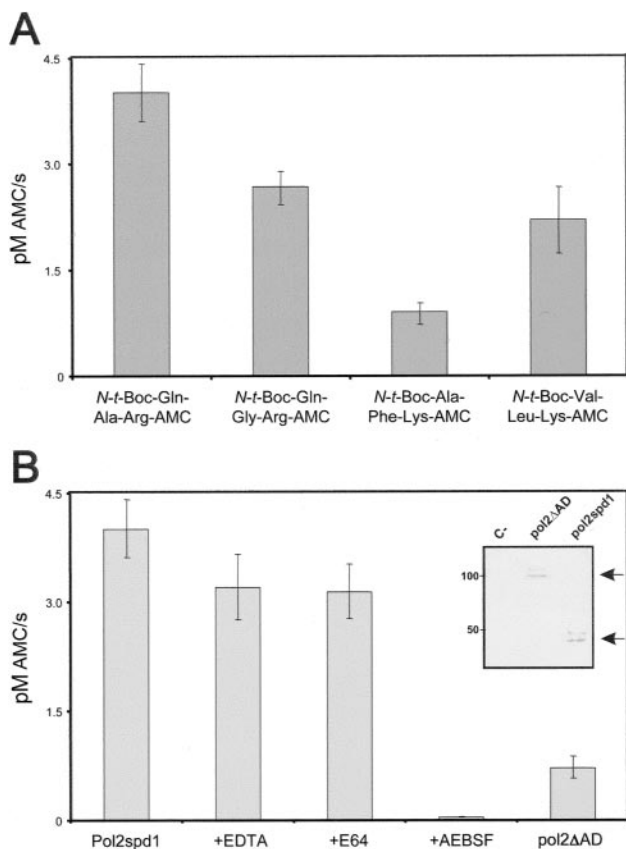


FIG. 5. Analysis of the enzymatic activity of polyserase-2 on synthetic peptides. *A*, synthetic fluorescent peptides *N-t*-Boc-Gln-Ala-Arg-AMC, *N-t*-Boc-Gln-Gly-Arg-AMC, *N-t*-Boc-Ala-Phe-Lys-AMC, and *N-t*-Boc-Val-Leu-Lys-AMC (50 μ M) were incubated with 10 μ l of the first catalytic module of polyserase-2 (*pol2spd1*) in 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, at 37 $^{\circ}$ C, in a final volume of 100 μ l. The fluorometric measurements were made at $\lambda_{\text{ex}} = 360$ nm and $\lambda_{\text{em}} = 460$ nm. *B*, for inhibition assays, synthetic peptide *N-t*-Boc-Gln-Ala-Arg-AMC was incubated with polyserase-2 in the presence or absence of 2.5 mM EDTA, 10 μ M E-64, or 2 mM 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF), and fluorescence was measured as described above. The activity of the full-length polyserase-2 lacking the activation domain (*pol2ΔAD*) against *N-t*-Boc-Gln-Ala-Arg-AMC was also measured. *Inset*, Western blot showing equivalent amounts of Pol2ΔAD and Pol2spd1 proteins (indicated by arrows). *C*-, conditioned medium from cells transfected with the empty vector. The obtained data were derived from four different transfections carried out in COS-7 cells. Similar results were obtained when 293-EBNA cells were used.

Boc-Gln-Gly-Arg-AMC and, to a lesser extent, *N-t*-Boc-Ala-Phe-Lys-AMC and *N-t*-Boc-Val-Leu-Lys-AMC (Fig. 5A). Pol2ΔAD showed a similar pattern of activity, although the activity of this protein was less than that observed for the first module of polyserase-2 (Fig. 5B). These data also indicate a preference of polyserase-2 for substrates with an Arg instead of Lys in position P1 (Fig. 5A), as described previously for other related proteins such as pancreatin (21) or matriptase-2 (23, 24). The activity of both recombinant proteins was substantially abolished with the serine proteinase inhibitor 4-(2-aminoethyl)-benzenesulfonyl fluoride, but not with EDTA or E-64, thereby providing additional evidence for their classification as serine proteases.

DISCUSSION

The detailed exploration of the human genome offers the possibility to identify all members of large and complex protein families, such as proteolytic enzymes that represent about 2% of currently annotated genes in mammalian genomes (7). In this work, we describe the finding of a new human protease tentatively called polyserase-2 to emphasize its structural re-

lationship with polyserase-1, a complex mosaic polyprotein also containing three serine protease domains within a single polypeptide chain. The approach followed to identify polyserase-2 was first based on a genomic search for regions encoding serine protease domains that were very close to each other. After identification of candidate regions and a series of PCR and rapid amplification of cDNA end experiments using liver cDNA as template, a full-length cDNA for human polyserase-2 was finally cloned and characterized.

The structural analysis of the sequence predicted for polyserase-2 revealed some striking similarities compared with that of polyserase-1 but also revealed clear differences in their modular architecture. Thus, both proteases contain three tandem serine protease domains in their respective amino acid sequences. However, the polyserase-2 architecture is less complex because of the lack of additional domains such as a type II transmembrane sequence and a low-density lipoprotein receptor A module present in polyserase-1 (10). Consistent with these structural differences, experimental analysis demonstrated that polyserase-2 is detected as a soluble protease, whereas polyserase-1 is associated with the plasma membrane through its type II transmembrane sequence. Further comparative analysis of the three protease domains of both polyproteins revealed some additional differences between them. Thus, we have reported previously that the first two protease domains of polyserase-1 are proteolytically active, whereas the third module is inactive because of the presence of an Ala residue instead of the catalytic Ser residue in the active site. However, analysis of the polyserase-2 structure indicated that its first protease domain was the only one containing all characteristic features of serine proteases, whereas the second and third domains lacked one characteristic residue of the catalytic triad of these enzymes and were predicted to be catalytically inactive. In agreement with these structural findings, enzymatic analysis demonstrated that both full-length polyserase-2 and its first serine protease domain have the ability to hydrolyze fluorogenic peptides used for assaying serine proteases. In addition, these proteolytic activities were extensively blocked by inhibitors of serine proteases but not by inhibitors of other classes of enzymes, thereby reinforcing the classification of polyserase-2 as a serine protease. Interestingly, kinetic analysis demonstrated that the activity of the first protease domain of polyserase-2 produced as an independent unit was greater than that of the entire protein, which includes the two additional protease domains predicted to be catalytically inactive. According to these results, it is tempting to speculate that these modules located at the C-terminal region of polyserase-2 could act as dominant negative binding proteins with ability to regulate the activity of the first serine protease domain of the polyprotein.

Analysis of putative post-translational processing mechanisms occurring in human polyserase-2 also revealed interesting differences compared with those affecting human polyserase-1 and other polyserine proteases identified in amphibians. Thus, the three protease domains present in polyserase-2 are not cleaved from the original translation product and remain as an integral part of a single polypeptide chain. By contrast, polyserase-1 undergoes a complex series of proteolytic events that lead to the generation of three independent serine protease units (10). Likewise, amphibian oviductin and ovochymase also release their respective protease modules after translation of a single polyprotein product (11–13). Structural analysis of the sequences preceding the catalytic domains of these modules may provide an explanation for the observed inability of polyserase-2 to generate independent protease units. Thus, although the first catalytic domain of this polyprotein is preceded

by a propeptide that ends in a conserved Arg residue present in the activation site of serine proteases, this domain should still remain anchored to the main chain through a conserved disulfide bond. However, the second and third domains are preceded by sequences completely unrelated to these activation motifs, thereby making unlikely that they can be separated from the original translation product.

To date, it is difficult to understand the putative functional advantages derived from the complex polyprotein designs occurring in both polyserases as well as in two unrelated human metalloproteases, such as angiotensin-converting enzyme (27–29) and carboxypeptidase D (30), that also contain different protease domains in their amino acid sequences. In this work, and as a very preliminary attempt to elucidate the physiological functions of the newly identified polyserase-2, we have examined the expression pattern of polyserase-2 in human tissues and tumor cell lines. Northern blot analysis of human tissues showed that this polyprotein is predominantly expressed in fetal kidney and in adult skeletal muscle, liver, placenta, prostate, and heart. This expression pattern somewhat resembles that of polyserase-1, which is also produced by all these fetal and adult tissues. Similarly to polyserase-1, polyserase-2 is also detected in several cancer cell lines, opening the possibility that both polyproteins could participate in some of the protease-mediated processes associated with tumor development and progression (31–35). Also in this regard, it is remarkable that the 16p11.2 chromosomal region, where the polyserase-2 gene is located, has been associated with different genetic abnormalities linked to human malignancies such as myxoid liposarcoma (36) and acute myeloid leukemia (37). Unidentified loci for paroxysmal kinesigenic choreoathetosis (38), familial juvenile hyperuricemic nephropathy (39), and autosomal dominant myxomatous mitral valve prolapse (40) have also been linked to this region. The identification of the *in vivo* substrates of polyserase-2 and the elucidation of the functional role of this protease in normal tissues could help to ascertain whether it is a direct target of any of these genetic abnormalities. In this regard, identification of the murine ortholog of human polyserase-2 raises the possibility of generating mice deficient in this gene, which would contribute to clarification of the role of this enzyme in both physiological and pathological processes.

Acknowledgments—We thank Drs. X. S. Puente, A. Fueyo, and G. Velasco for helpful comments and support; M. Fernández, S. Alvarez, and P. Martín Bringas for excellent technical assistance; and Prof. B. A. Connolly (University of Newcastle upon Tyne) for providing oligonucleotides.

REFERENCES

- Lopez-Otin, C., and Overall, C. M. (2002) *Nat. Rev. Mol. Cell. Biol.* **3**, 509–519
- McCawley, L. J., and Matrisian, L. M. (2001) *Curr. Opin. Cell Biol.* **13**, 534–540
- Sternlicht, M. D., and Werb, Z. (2001) *Annu. Rev. Cell Dev. Biol.* **17**, 463–516
- Blasi, F., and Carmeliet, P. (2002) *Nat. Rev. Mol. Cell. Biol.* **3**, 932–943
- Stennicke, H. R., Ryan, C. A., and Salvesen, G. S. (2002) *Trends Biochem. Sci.* **27**, 94–101
- Marino, G., Uria, J. A., Puente, X. S., Quesada, V., Bordallo, J., and Lopez-Otin, C. (2003) *J. Biol. Chem.* **278**, 3671–3678
- Puente, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) *Nat. Rev. Genet.* **4**, 544–558
- Puente, X. S., and Lopez-Otin, C. (2004) *Genome Res.* **14**, 609–622
- Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) *Nucleic Acids Res.* **32**, D160–D164
- Cal, S., Quesada, V., Garabaya, C., and Lopez-Otin, C. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9185–9190
- Lindsay, L. L., Yang, J. C., and Hedrick, J. L. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11253–11258
- Lindsay, L. L., Wieduwilt, M. J., and Hedrick, J. L. (1999) *Biol. Reprod.* **60**, 989–995
- Hiyoshi, M., Takamune, K., Mita, K., Kubo, H., Sugimoto, Y., and Katagiri, C. (2002) *Dev. Biol.* **243**, 176–184
- Krem, M. M., and Di Cera, E. (2001) *EMBO J.* **20**, 3036–3045
- Ross, J., Jiang, H., Kanost, M. R., and Wang, Y. (2003) *Gene (Amst.)* **304**, 117–131
- Mouse Genome Sequencing Consortium. (2002) *Nature* **420**, 520–562
- Rat Genome Sequencing Project Consortium. (2004) *Nature* **428**, 493–521
- Krem, M. M., Rose, T., and Di Cera, E. (2000) *Trends Cardiovasc. Med.* **10**, 171–176
- Oberst, M. D., Williams, C. A., Dickson, R. B., Johnson, M. D., and Lin, C. Y. (2003) *J. Biol. Chem.* **278**, 26773–26779
- Wong, G. W., Yasuda, S., Morokawa, N., Li, L., and Stevens, R. L. (2004) *J. Biol. Chem.* **279**, 2438–2452
- Bhagwandin, V. J., Hau, L. W., Mullen-St Clair, J., Wolters, P. J., and Caughey, G. H. (2003) *J. Biol. Chem.* **278**, 3363–3371
- Lin, C. Y., Anders, J., Johnson, M., Sang, Q. A., and Dickson, R. B. (1999) *J. Biol. Chem.* **274**, 18231–18236
- Velasco, G., Cal, S., Quesada, V., Sanchez, L. M., and Lopez-Otin, C. (2002) *J. Biol. Chem.* **277**, 37637–37646
- Hooper, J. D., Campagnolo, L., Goodarzi, G., Truong, T. N., Stuhlmann, H., and Quigley, J. P. (2003) *Biochem. J.* **373**, 689–702
- Yu, J. X., Chao, L., and Chao, J. (1995) *J. Biol. Chem.* **270**, 13483–13489
- Wong, G. W., Yasuda, S., Madhusudhan, M. S., Li, L., Yang, Y., Krilis, S. A., Sali, A., and Stevens, R. L. (2001) *J. Biol. Chem.* **276**, 49169–49182
- Soubrier, F., Alhenc-Gelas, F., Hubert, C., Allegrini, J., John, M., Tregear, G., and Corvol, P. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9386–9390
- Ehlers, M. R., Fox, E. A., Strydom, D. J., and Riordan, J. F. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7741–7745
- Wei, L., Clauser, E., Alhenc-Gelas, F., and Corvol, P. (1992) *J. Biol. Chem.* **267**, 13398–13405
- Novikova, E. G., Eng, F. J., Yan, L., Qian, Y., and Fricker, L. D. (1999) *J. Biol. Chem.* **274**, 28887–28892
- Egeblad, M., and Werb, Z. (2002) *Nat. Rev. Cancer* **2**, 161–174
- Overall, C. M., and Lopez-Otin, C. (2002) *Nat. Rev. Cancer* **2**, 657–672
- Sidenius, N., and Blasi, F. (2003) *Cancer Metastasis Rev.* **22**, 205–222
- Netzel-Arnett, S., Hooper, J. D., Szabo, R., Madison, E. L., Quigley, J. P., Bugge, T. H., and Antalis, T. M. (2003) *Cancer Metastasis Rev.* **22**, 237–258
- Roshy, S., Sloane, B. F., and Moin, K. (2003) *Cancer Metastasis Rev.* **22**, 271–286
- Rabbits, T. H., Forster, A., Larson, R., and Nathan, P. (1993) *Nat. Genet.* **4**, 175–180
- Ichikawa, H., Shimizu, K., Hayashi, Y., and Ohki, M. (1994) *Cancer Res.* **54**, 2865–2868
- Tomita, H., Nagamitsu, S., Wakui, K., Fukushima, Y., Yamada, K., Sadamatsu, M., Masui, A., Konishi, T., Matsuishi, T., Aihara, M., Shimizu, K., Hashimoto, K., Mineta, M., Matsushima, M., Tsujita, T., Saito, M., Tanaka, H., Tsuji, S., Takagi, T., Nakamura, Y., Nanko, S., Kato, N., Nakane, Y., and Niikawa, N. (1999) *Am. J. Hum. Genet.* **65**, 1688–1697
- Stiburkova, B., Majewski, J., Hodanova, K., Ondrova, L., Jerabkova, M., Zikanova, M., Vylet'ál, P., Sebesta, I., Marinaki, A., Simmonds, A., Matthijs, G., Fryns, J. P., Torres, R., Puig, J. G., Ott, J., and Kmoeh, S. (2003) *Eur. J. Hum. Genet.* **11**, 145–154
- Disse, S., Abergel, E., Berrebi, A., Houot, A. M., Le Heuzey, J. Y., Diebold, B., Guize, L., Carpentier, A., Corvol, P., and Jeunemaitre, X. (1999) *Am. J. Hum. Genet.* **65**, 1242–1251