This article has been withdrawn by the authors upon request from the Journal. The Journal raised questions regarding Figs. 4 and 5A. In Fig. 4, the labeling of the lanes corresponding to adult small intestine and ovary were swapped. The actin panels were reused from previous publications of the group using these commercial membranes, and incorrectly oriented. The first two lanes of the Coomassie panel in Fig. 5A appeared to be inappropriately manipulated. The original data were located for some, but not all, gels used to prepare this panel. The authors assert that all of the results reported in this article are valid.

© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

Identification and Characterization of Human Archaemetzincin-1 and -2, Two Novel Members of a Family of Metalloproteases Widely Distributed in Archaea*

Received for publication, April 26, 2005, and in revised form, June 13, 2005
Published, JBC Papers in Press, June 22, 2005, DOI 10.1074/jbc.M504533200

Aarceli Díaz-Perales‡, Víctor Quesada‡, Juan R. Peinado‡, Alejandro P. Ugalde‡, Jesús Álvarez§, María F. Suárez‡, F. Xavier Gomis-Rüth‡, and Carlos López-Otín‡

From the Departamentos de ‡Bioquímica y Biología Molecular y §Biology Celular, Facultad de Medicina, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain and the ‡Institut de Biologia Molecular de Barcelona, Consejo Superior de Investigaciones Científicas, 08034 Barcelona, Spain

Systematic analysis of degradomes, the complete protease repertoires of organisms, has demonstrated the large and growing complexity of proteolytic systems operating in all cells and tissues. We report here the identification of two new human metalloproteases that have been called archaemetzincin-1 (AMZ1) and archaemetzincin-2 (AMZ2) to emphasize their close relationship to putative archaemetzincin-1 (AMZ1) and archaemetzincin-2 (AMZ2) in all cells and tissues. We report here the identification of and growing complexity of proteolytic systems operating.

Proteases mediate many key physiological processes (1). These enzymes play essential roles in a variety of events that determine cell life and death in all living organisms. Thus, proteases participate in the control of cell cycle progression, tissue morphogenesis and remodeling, cell proliferation and migration, ovulation and fertilization, angiogenesis, host defense, hemostasis, apoptosis, and autophagy (2–9). Because of these crucial roles, strict regulatory mechanisms are necessary to avoid aberrant proteolytic systems.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/Eru Data Bank with accession number(s) AJ635357, AJ879912, AJ879913, AJ879914, and AJ879915.

The abbreviations used are: AMZ, archaemetzincin; AMC, 7-amino-4-methylcoumarin; contig, group of overlapping clones; GST, glutathione S-transferase; Mca, (7-methoxycoumarin-4-yl)-acetic acid.
Ametzincin-2 (AMZ2), which are closely related to proteins whose sequence has been predicted by bioinformatic analysis of archaeal genomes. We perform a detailed phylogenetic analysis of these enzymes to clarify the origin and complex evolutionary history of this new family of metalloproteases. Finally, we examine the tissue distribution of AMZ1 and AMZ2 in human tissues and analyze their enzymatic properties.

**EXPERIMENTAL PROCEDURES**

**Materials—**Restriction endonucleases and other reagents used for molecular cloning were from Roche Diagnostics. Double-stranded DNA probes were radiolabeled with [(α-32P)]dCTP (3000 Ci/mmole) from Amersham Biosciences, using a commercial random primer kit purchased from the same company. Human cDNA libraries and Northern blots containing polyadenylated RNAs from different tissues were from Clontech. Fluorogenic substrates and biologically active peptides (neuromagin, angiotensin II, and angiotensin III) were purchased from Bachem, and protease inhibitors and AMC were from Sigma. Albumin, fibrillar collagens, gelatin, plasminogen, and aprotinin were also from Sigma. Antibodies against GST were developed in our laboratory as described previously (22).

**Bioinformatic Analysis and cDNA Cloning—**The BLAST program was used to search public (www.ncbi.nlm.nih.gov) and private (www.celera.com) human genome databases, searching for regions with sequence similarity to prokaryotic metalloproteinase sequences (21). We found two partial sequences located in the human chromosomes 7p22.3 and 17q24.2 exhibiting similarity to putative metalloproteinase sequences identified during the course of large scale genome-sequencing projects involving Archaea (23–28). After the identification of these human sequences, we designed specific oligonucleotides to PCR amplify the cDNAs for these metalloproteases using a human brain cDNA library as a template. All PCR amplifications were performed by the GeneAmp 2400 PCR system from PerkinElmer Life Sciences. PCR cloning the PCR products in pBluescript, their identity by nucleotide sequencing.

**Nucleotide Sequence Analysis—**Cloned cDNAs were sequenced (Applied Biosystems). Computer sequences were aligned with the University of Wisconsin Genetics Computer Group.

**Phylogenetic Analysis—**Sequences from archaeal, bacterial, and eukaryotic origin were aligned using ClustalW (www Genetics.Wisc.Edu/Clustalw). A phylogenetic tree based on this alignment was calculated using the analysis package FL WinLab 2.01 (PerkinElmer Life Sciences).

**Northern Blot Analysis—**Nylon membranes containing 2 µg of poly(A+) RNA from diverse human tissues were prehybridized at 42 °C for 3 h in 50% formamide, 5× SSPE (1× SSPE is 150 mM NaCl, 10 mM NaH2PO4, 1× EDTA, pH 7.4) 10× Denhardt’s solution, 2% SDS, and 100 µg/ml denatured herring sperm DNA. Membranes were then hybridized with specific radiolabeled probes containing nucleotides from probe hybridization. Finally, blots were washed once with 2× SSC, 0.5% SDS for 1 hour and three times in 0.1× SSC and 0.1% SDS for 30 min at 50 °C and exposed to autoradiography. RNA integrity and loading was assessed by hybridization with an actin probe.

**Production and Purification of Recombinant Proteins—**cDNAs for the predicted catalytic domains of AMZ1 (positions 1–320) and AMZ2 (positions 1–300) were obtained by PCR amplification using specific oligonucleotide pairs containing defined restriction sites. The AMZ1 catalytic domain oligonucleotides were 5′-GGGGATCCCATGCTGCAG-3′ and 5′-GGGATCCCATGCTGCAG-3′, and the AMZ2 catalytic domain oligonucleotides were 5′-GGGGATCCCATGCTGCAG-3′ and 5′-GGGGATCCCATGCTGCAG-3′ (where the restriction sites are underlined). PCR amplifications were performed with 30 cycles of denaturation (95 °C for 30 s), annealing (60 °C for 30 s), and extension (68 °C for 1 min) using the ExpandTM long template, high fidelity PCR system. PCR products were then digested with the corresponding restriction enzymes and cloned into the appropriate sites of the pGEX-5x-2 expression vector (Amersham Biosciences). The resulting constructs were transformed into BL21 (DE3)-pLysE competent Escherichia coli cells, and expression was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside (final concentration 1 mM), followed by 3 h of incubation at 28 °C. The cells were then harvested by centrifugation, washed with phosphate-buffered saline, and lysed by incubation in phosphate-buffered saline with 100 µg/ml lysozyme, 0.5% SDS for 30 min and three times in 0.1% Triton X-100 overnight at 4 °C. The recombinant catalytic domain proteins contained in the corresponding supernatants were purified by affinity chromatography using a glutathione-Sepharose column. The identity of the recombinant proteins was verified by Western blot and trypsin digestion followed by mass spectrometry analysis.

**Trypsin Digestion—**Gel bands were manually excised and placed into 0.5-ml tubes. Then, gel pieces were washed three times with 180 µl of 25 mM ammonium bicarbonate/acetonitrile (70:30) (v/v), dried at 90 °C for 15 min, and incubated with 12 µg/ml trypsin (Promega) in 25 mM ammonium bicarbonate at 60 °C for 1 h. Likewise, soluble proteins were incubated with trypsin (12 µg/ml) in 25 mM ammonium bicarbonate for 1 h at 60 °C. The resulting peptide mixture peptides were placed into ice for 2 min, and 2 µl of 10% trifluoroacetic acid and 1 µl of water were added to each sample. The samples were then desalted by C18 reverse phase chromatography (ZipTip; Millipore). Peptides were eluted with 0.2% trifluoroacetic acid in acetonitrile and 0.1% trifluoroacetic acid (v/v). In a typical experiment, 1 µl of this solution was used for mass spectrometry.

**Mass Spectrometry—**A mechanical pulsed laser desorption ionization mass spectrometer equipped with a linear ion trap (VG, Applied Biosystems) was used to produce a mass spectrum of the purified recombinant proteins using AMC-coupled amino acids (Bachem) for MALDI-MS experiments. Highly charged peptides were analyzed with a MALDI-TOF/TOF mass spectrometer equipped with a linear ion trap (VG, Applied Biosystems).

**RESULTS**

**Cloning and Characterization of Two Human cDNAs Encoding Novel Metalloproteinases Similar to Archaeal Metazincins—**A bioinformatic search of the human genome to look for sequences similar to those of archaeal or bacterial metallopro-
teases led us to identify two DNA contigs located in chromosomes 7p22.3 and 17q24.2 and encoding two uncharacterized proteins with sequence similarity to putative archaeal metalloprotease sequences (21, 23–26). The full-length cDNAs for both human enzymes were PCR-amplified using specific oligonucleotides and a brain cDNA library. These experiments led us to the amplification of 1.5- and 1-kb cDNAs, both containing an in-frame initiator and stop codons. After cloning and sequencing of the PCR-amplified products, we confirmed by conceptual translation that the generated sequences encoded two novel proteins of 498 and 360 amino acids, respectively (Fig. 1A, and GenBank™ accession numbers AJ635537 and AJ635538). Domain analysis with the InterPro (www.ebi.ac.uk/interpro) and SMART (smart.embl-heidelberg.de) programs confirmed the presence in both human protein sequences of a catalytic domain related to neutral zinc metalloproteases. A search for orthologous sequences using the TBLASTN algorithm showed that both human sequences are closely related to members of a family of predicted metalloproteases originally identified during the analysis of archaeal genomes and tentatively called archaemetzincins (21). Accordingly, we propose to call the newly identified proteins human archaemetzincin-1 and -2. The maximum percentages of similarities between the catalytic domains of human and archaeal enzymes were 27% between human AMZ1 and the predicted archaemetzincin from the genome sequence of Thermococcus kodakaraensis and 39% between human AMZ2 and the corresponding enzyme from Pyrococcus abyssi (Fig. 1B). Likewise, the percentage of identities between the catalytic domains of human AMZ1 and AMZ2 is ~40%. Further bioinformatic analysis of available sequences revealed that archaemetzincins are widespread in Archaea as well as in vertebrates including birds, amphibians, and fish (Fig. 1B). However, archaemetzincins were also absent from plants and from bacterial species such as A. aeolicus, and D. melanogaster (28, 29). Finally, data from this analysis were fitted to a taxonomic tree to construct a model that could explain the evolution of AMZ genes (Fig. 3). According to this model, the primordial AMZ arose in a common ancestor of Archaea and Eukaryota. Some bacterial species acquired this gene through lateral gene transfer from archaeal organisms. On the other hand, two duplication events would explain the presence of AMZ2 in Thermococcalesaeae and AMZ3 in Archaea. The lack of AMZ genes in several eukaryotic organisms would be likely explained by multiple independent different times (Fig. 3).

**Enzymatic Properties of Human AMZ1 and AMZ2 Produced in E. coli**—To analyze the enzymatic properties of both human AMZs, we produced in E. coli two fusion proteins containing the putative catalytic domains of these enzymes linked to GST at their N termini. The catalytic domains were defined based on the alignments of human AMZs with the related archaeal proteins, which showed the maximum degree of conservation in the N-terminal region of these proteins. Then, these constructs (encoding amino acids 1–320 of AMZ1 and 1–300 of AMZ2) were transformed in E. coli BL21, and, after isopropyl-1-thio-
β-D-galactopyranoside induction, bands of the expected size (55 kDa) were detected by SDS-PAGE and Western blot analysis of protein extracts using antibodies against GST (Fig. 5A). These recombinant GST-proteases were then purified by glutathione-Sepharose chromatography. To assess the identity of the proteases present in these bands, they were digested with trypsin and analyzed by mass spectrometry. The obtained spectra confirmed that the 55-kDa bands corresponded to GST-AMZ1 and GST-AMZ2 fusion proteins.

The recombinant human AMZ1 and AMZ2 proteins were then used in enzymatic assays with the fluorescent substrates commonly employed for assaying other proteases. These assays showed that recombinant AMZ1 exhibits a significant hydrolytic activity against Ala-AMC, whereas recombinant AMZ2 preferentially cleaves Arg-AMC (Fig. 5B). By contrast, we did
FIG. 1. Sequences of human AMZs and comparison to proteases from other organisms. A, sequences of human AMZ1 and AMZ2. The nucleotide and amino acid sequences of AMZ1 and AMZ2 are shown. The characteristic core catalytic motif of archaemetzincins, including the zinc-binding site, is in black. B, amino acid sequence alignment around the zinc-binding site of the catalytic domains of AMZ enzymes. Residues common to all of them are shaded. The alignment was performed using ClustalX (version 1.81). Hsa, Homo sapiens; Mmu, Mus musculus; Rno, Rattus norvegicus; Gga, Gallus gallus; Xtr, Xenopus tropicalis; Tni, Tetraodon nigroviridis; Pfu, Pyrococcus furiosus; Mja, Methanococcus jannaschii; Neq, Nanoarchaeum equitans; Sso, Sulfolobus solfataricus; Aae, A. aeolicus.
The fitting of the resulting data to the Michaelis-Menten equation yielded $k_{cat}/K_m$ values of 46 m$^{-1}$ s$^{-1}$ and 22 m$^{-1}$ s$^{-1}$ for catalytic domain proteins of AMZ1 and AMZ2, respectively, which are similar to the value reported for recombinant aminopeptidase O produced in the same expression system (30). Wenext performed a kinetic analysis of the proteolytic reaction catalyzed by the catalytic domains of AMZ1 and AMZ2 with their preferred substrates ( Ala-AMC and Arg-AMC, respectively).

Figure 6. Enzymatic activity of AMZ1 and AMZ2 against QF35 or QF41, two peptides derived from fibrillar collagens, gelatin, plasminogen, and aprotinin (data not shown). A, B, and C are the same experimental conditions (data not shown). Similarly, neither AMZ1 nor AMZ2 hydrolyzed angiotensin II (data not shown).

To further characterize the enzymatic activity of the identified AMZ metalloproteases, several commercially available bioactive peptides were incubated in the presence of purified catalytic domains of AMZ1 or AMZ2, and the resulting samples were analyzed by mass spectrometry. As shown in Figure 6, these experiments demonstrated that human AMZ1 exhibited aminopeptidase activity against neurogranin, whereas human AMZ2 was active against angiotensin III. Thus, as can be seen in Figure 6A, neurogranin is detected as a 1800.1-Da peak corresponding to the processed peptide KIQASFRGHMARKK, whereas incubation of neurogranin with AMZ1 produced a single additional peak with a mass of 1657.9 Da, corresponding to the processed peptide KIQASFRGHMARKK (Figure 6B). Notably, AMZ2 did not process neurogranin under the same experimental conditions (data not shown). Similarly, AMZ2 was active against angiotensin III (RVYIHPF), albeit with low efficiency, to produce angiotensin IV (VYIHPF) (Figure 6, C and D).

**DISCUSSION**

In this work we describe two new human proteases that have been tentatively called archaemetzincin-1 and -2. According to a series of structural and enzymatic features, these proteins belong to a new family of metalloproteases characterized by a conserved motif (HEXXHXXXG, CXX,CXX,CXXX) that contains an archetypal zinc-binding site and four Cys residues that contribute to defining the specific signature of this novel metalloprotease family. Furthermore, enzymatic assays performed with human recombinant AMZs have provided the first evidence that these proteins are catalytically active metalloproteases that exhibit substrate specificity and sensitivity to inhibitors to block the enzymatic activity of both human AMZs.
hibitors, which appears to indicate that both proteases may act predominantly as aminopeptidases.

An additional distinctive feature of this family of metalloproteases is the complex series of evolutionary events that have contributed to its creation and diversification in different organisms. In fact, our bioinformatic analysis revealed that these enzymes are widely distributed in vertebrate and archaeal organisms but are absent in the genomes of a number of model organisms such as *E. coli*, *S. cerevisiae*, *A. thaliana*, *D. melanogaster*, and *C. elegans*. The occurrence of genes shared by prokaryotes and vertebrates but absent in other eukaryotes has been widely considered as an indication of lateral gene transfer events from prokaryotes to vertebrates (31). Accordingly, AMZs could represent novel and interesting examples of these rare evolutionary events. However, the recent accumulation of data questioning many cases of lateral gene transfer to the vertebrate lineage (32–34) prompted us to perform an exhaustive bioinformatic search for AMZ genes in all available genome sequences. This analysis led us to identify additional AMZ-related sequences in other non-vertebrate eukaryotes and in two bacterial species, as well as to uncover a series of complex evolutionary events underlying the formation of this metalloprotease family (Fig. 3). According to this phylogenetic analysis, the evolutionary history of AMZs is best described by...
a scenario in which the primeval AMZs and eukaryotic AMZs share a common ancestor, which is consistent with the proposed origin of AMZs after the appearance of the primordial bacterial organism. The presence of an AMZ in a hyperthermophilic bacteria that occupies an ecological niche dominated by Archaea, should be explained by lateral gene transfer from some of these archaeal organisms. Furthermore, the clear phylogenetic relationship between Archaea and eukaryotic organisms has also involved a series of diverse mutations, which would explain its location as an out group in the phylogenetic tree. The evolutionary history of AMZs in eukaryotic organisms has also involved a series of diverse events since their separation from their common ancestor with Archaea. First, the absence of AMZ genes in plants, nematodes, or insects is remarkable, suggesting the occurrence of multiple gene loss events in these organisms. Consistent with this proposal, codon usage or nucleotide composition analysis of AMZ genes failed to provide any evidence of lateral transmission from Archaea to vertebrates. Finally, our phylogenetic analysis also revealed that eukaryotic AMZ1 diverged from AMZ2 recently, probably by gene duplication, again illustrating the genomic plasticity of this family of metalloproteases.

To further explore the functional relevance of AMZ1 and AMZ2, we performed an enzymatic analysis of both recombinant enzymes produced in E. coli. This analysis revealed that the recombinant proteins are catalytically active and that their activities seem to correspond to those of aminopeptidases.
FIG. 6. Mass spectrometry analysis of the biological peptide proteolysis catalyzed by AMZs. Human neurogranin or angiotensin III were incubated alone (A and C, respectively) or in the presence of recombinant GST-AMZ1 (B) or GST-AMZ2 (D) for 2 h. The resulting peptide mixture was analyzed by mass spectrometry. The peaks corresponding to neurogranin (1800.1 Da), processed neurogranin (1657.9 Da), angiotensin III (931.6 Da), and angiotensin IV (775.4 Da) are shown.
tissues. Additional clues about the physiological and pathological roles of these enzymes may be derived from their chromosome locations at 7p22 and 17q24, respectively. Alterations in these regions have been frequently associated with cancer and other diseases such as hypertension or multiple sclerosis (38–41). Further studies will be required to ascertain whether AMZs could be a direct target of any of these genetic abnormalities resulting in cancer or other pathological conditions. Likewise, further experimental work, including the three-dimensional structural analysis of these enzymes and the generation of mutant organisms deficient in these proteases, will be necessary to clarify their functional roles and to define their precise relevance in the context of the growing complexity of proteolytic systems operating in all living organisms.

Acknowledgments—We thank G. Velasco, J. P. Freije, A. M. Pendás, S. Cal, L. M. Sánchez, and X. S. Puente for helpful comments, and S. Alvarez for excellent technical assistance.

REFERENCES
1. Lopez-Otín, C., and Overall, C. M. (2002) Nat. Rev. Mol. Cell Biol. 3, 509–519
2. Glickman, M. H., and Ciechanover, A. (2002) Physiol. Rev. 82, 373–428
3. Inada, M., Wang, Y., Byrne, M. H., Rahman, M. U., Miyaera, C., Lopez-Otín, C., and Krane, S. M. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 17192–17197
4. Blobel, C. P. (2005) Nat. Rev. Mol. Cell Biol. 6, 32–43
5. Curry, T. E., Jr., and Osteen, K. G. (2003) Endocr. Rev. 24, 32–43
6. Di Cera, E. (2003) J. Biol. Chem. 278, 428–465
7. Overall, C. M., Lopez-Otin, C. (2002) Nat. Rev. Cancer 2, 189–192
8. Puente, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) Genes Dev. 17, 2805–2817
9. Riedl, S. J., and Shi, Y. (2004) Nat. Rev. Mol. Cell Biol. 5, 275–286
10. Rabbitts, T. H. (2005) Nature 432, 544–558
11. Shire, S. (2005) J. Med. Genet. 42, 331–334
12. Smith, P. J., and Kucherlapati, R. (2005) Curr. Opin. Genet. Dev. 15, 14310–14317
13. Smit, A. F., and Hubley, R. K. (2005) Genome Res. 15, 1359–1369
14. Stange-Thomann, N., and Stojanovic, N. A. (2005) J. Clin. Pathol. 58, 577–584
15. Su, Z. O., Kent, S. W., Diepolder, H., et al. (2005) Genome Res. 15, 1483–1492
16. Wang, C. H. (2005) Nat. Rev. Mol. Cell Biol. 6, 10109–10119
17. Williams, A., Gray, A. R., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., et al. (2005) Genomics 1528, 631–638
18. Winter, R., and Disteche, C. M. (2005) Science 308, 53–57
19. Xie, W., et al. (2005) Nature 432, 3671–3678
20. Xing, J., et al. (2005) Nature 432, 509–519
21. Yamada, K. M., and Stiles, C. D. (2005) Genome Res. 15, 428–465
22. Yang, H., et al. (2005) Cell 120, 267–277
23. Zhang, Z., and Ferretti, V. (2005) Science 310, 401–402
24. Zhao, Y., et al. (2005) Hum. Mol. Genet. 14, 2339–2348
25. Zhang, Z., and Ferretti, V. (2005) Science 308, 496–499
26. Zhao, Y., et al. (2005) J. Clin. Pathol. 58, 577–584
27. Northrop, D. B. (1999) Adv. Enzymol. Relat. Areas Mol. Biol. 73, 25–55
28. Forterre, P., and Bouvier De La Tour, C., Philippe, H., and Duguet, M. (2000) Trends Genet. 16, 152–154
29. Nesbo, C. L., and L’Haridon, S., Stetter, K. O., and Doolittle, W. F. (2001) Mol. Biol. Ecol. 18, 362–375
30. Diaz-Perales, A., Sánchez, L. M., and Imanaka, T. (2005) Hum. Archaemetazincins 30375
31. S. Alvarez, for excellent technical assistance.

Acknowledgments—We thank G. Velasco, J. P. Freije, A. M. Pendás, S. Cal, L. M. Sánchez, and X. S. Puente for helpful comments, and S. Alvarez for excellent technical assistance.