

Complete amino acid sequence of the large subunit of the low-Ca²⁺-requiring form of human Ca²⁺-activated neutral protease (μ CANP) deduced from its cDNA sequence

Kazumasa Aoki⁺, Shinobu Imajoh, Shigeo Ohno, Yasufumi Emori, Morio Koike*, Goro Kosaki* and Koichi Suzuki[†]

*Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science and *Tokyo Metropolitan Komagome Hospital, Bunkyo-ku, Tokyo 113, Japan*

Received 19 June 1986

The complete amino acid sequence of the large subunit (catalytic subunit) of human low-Ca²⁺ requiring-calcium-activated neutral protease (μ CANP) was deduced from its cDNA base sequence. It is composed of 714 amino acid residues and its sequence is highly homologous to the chicken CANP sequence determined previously. Human μ CANP, like chicken CANP, has a clear 4-domain structure, and their fundamental structures are essentially the same, although their Ca²⁺ sensitivities are significantly different. The role of each domain in the Ca²⁺ sensitivity and protease activity of CANP is discussed on the basis of sequence comparison.

Ca²⁺-activated neutral protease Calpain Thiol protease cDNA cloning Sequence homology Ca²⁺

1. INTRODUCTION

Ca²⁺-activated neutral protease (CANP, calpain, EC 3.4.22.17) is an intracellular thiol protease that is ubiquitously found in tissues of higher animals [1–3]. Two types of CANP exist in mammals, μ CANP and mCANP, which respectively require micromolar and millimolar Ca²⁺ for their activity [1–3]. Both CANPs are composed of two subunits, a large (catalytic) (M_r 80000) one and a small (regulatory) (M_r 28000) one [1–3]. As the small subunit is common to both CANPs [4,5], the Ca²⁺ requirement of CANP is determined apparently by the large subunit [6]. We have already shown that the large subunit of chicken CANP

consists of four domains, two of which were identified as a thiol protease domain and a calmodulin-like Ca²⁺-binding domain [7]. Comparison of the structures of μ and mCANPs is important for clarifying the molecular basis of their different Ca²⁺ sensitivities. However, since only one molecular species of CANP, with an intermediate Ca²⁺ sensitivity, has been found in chicken [7–9], we started to analyze the structures of μ and mCANPs from mammalian sources. Recently, cDNA clones corresponding to the C-terminal Ca²⁺-binding domains of rabbit μ and mCANPs were isolated [10]. Here we report the complete amino acid sequence of the large subunit of human μ CANP, which was deduced from the nucleotide sequence of its cDNA clone isolated using rabbit μ CANP cDNA as a probe.

2. MATERIALS AND METHODS

Total RNA was prepared from human skeletal

[†] To whom correspondence should be addressed

⁺ Present address: Department of Biochemistry, Faculty of Pharmaceutical Sciences, Science University of Tokyo, 12 Ichigaya-funagawara-machi, Shinjuku-ku, Tokyo 162, Japan

muscle and spleen [11], and then poly(A)⁺ RNA enrichment was performed by oligo(dT)-cellulose chromatography [12]. Two cDNA libraries were constructed. One was prepared from human skeletal muscle poly(A)⁺ RNA by the conventional method, as in [7]. The other library was constructed from human spleen poly(A)⁺ RNA. Double-stranded cDNA was synthesized as in [13], and fractionated on low-gelling-temperature agarose gel after *Eco*RI linker ligation. cDNA longer than 2.3 kb was collected and ligated to the vector, λ gt10 [14]. Screening of the library was performed as in [12]. DNA was sequenced according to [15].

3. RESULTS AND DISCUSSION

3.1. Identification and DNA sequence of the cDNA clone for the large subunit of human μ CANP

The cDNA library from human skeletal muscle was screened with a *Pst*I cDNA fragment for the large subunit of rabbit μ CANP [10]. One positive clone (pHM42) was isolated from about 45000 transformants and sequenced. Further screening of about 4×10^5 plaques from the human spleen cDNA λ gt10 library with the pHM42 insert as a hybridization probe yielded 18 positive clones. The clone (λ 31) containing the longest insert was subjected to nucleotide sequence analysis after subcloning into the pUC8 vector (resulting subclone; p31). The restriction map for the cDNA inserts in clones pHM42 and p31 is shown in fig.1,

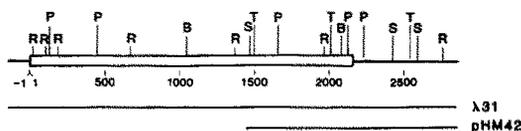


Fig.1. Restriction map of the inserts in cDNA clones p31 and pHM42 for the human μ CANP large subunit. Horizontal lines denote regions covered by the inserts of p31 and pHM42. Nucleotide residues are numbered in the 5' to 3' direction, beginning with the first residue of the ATG triplet that encodes the initiation methionine. An open box indicates the protein-coding region and horizontal lines on both sides of the box indicate the 5'- and 3'-untranslated regions. Vertical lines and letters represent restriction sites. R, *Rsa*I; P, *Pvu*II; B, *Bgl*II; S, *Sma*I; T, *Taq*I.

the total nucleotide sequence of cDNA (3011 bp) being shown in fig.2.

The translational initiation site was assigned to ATG at position 1, because a termination codon, TGA, exists at -57 in phase and the second ATG triplet appears a long way downstream (at position 778). This open reading frame is terminated by TGA at position 2143 and encodes 714 amino acid residues. The calculated M_r (81889) and the amino acid composition agree well with the reported values [16].

The C-terminal amino acid sequence of the large subunit of rabbit μ CANP [10] is almost identical (97% homology) to the deduced sequence for the corresponding position. Furthermore, we have already obtained another cDNA clone that encodes the large subunit of human mCANP (unpublished). Thus, it is concluded that the present cDNA clones encode the large subunit of human μ CANP. Northern blot hybridization of human spleen RNA with the cDNA insert of p31 as a probe gave a single band at a position corresponding to approx. 3.5 kb (fig.3). This indicates that the cloned cDNA, spanning a total of 3011 bp, covers nearly the full length of the μ CANP mRNA.

3.2. Comparison of the amino acid sequences

This is the first complete sequence determination of the large subunit of μ CANP. As for the structure of the large subunit of CANP, the complete sequence of chicken CANP [7] and the C-terminal sequences of rabbit m and μ CANPs [10] have been elucidated.

Comparison of the deduced amino acid sequence of human μ CANP with that of chicken CANP indicated that they are highly homologous, irrespective of the clear difference in their Ca²⁺ sensitivities, and human μ CANP, like chicken CANP, has a clear 4-domain structure (I-IV from the N-terminus). In domain II, the protease domain (amino acid residues 88-327), two active-site amino acid residues, Cys-115 and His-272, are found, and the sequences around these residues are highly conserved among various cysteine proteinases [7,17,18]. In domain IV, the Ca²⁺-binding domain (residues 570-714), 4 consecutive EF-hand sequences are found. The scores for the test sequences [19] are above 11 for all the 4 EF-hand sequences, indicating that these 4 sites can be regarded as potential Ca²⁺-binding sites. The total

<p>-143 AAGGAGAGAGGGAGGGCCGAGGG</p> <p>-120 CGGAGGGGCGCGGGAGGGCCGGGAGGAGCGCTTCCTGGTGGGCCCTGCCCTGA</p> <p>- 60 GCTGCCACCGGGAAGCCAGCCCTCAGGGACTGCAGCGACCCCAACCCCTCCCCGAGG</p> <p>1 ATGTGGGAGGAGATCATCAGCCGGTACTGCACTGGGGTGCAGCCCAAGTGCAGAAAG MetSerGluGluIleIleThrProValTyrCysThrGlyValSerAlaGlnValGlnLys 1</p> <p>61 CAGCGGGCCAGGGAGCTGGCCGGCCGATGAGAATGCCATCAAGTACCTGGCCAG GlnArgAlaArgGluSerGlyLeuGlyArgHisGluAsnAlaIleLysTyrLeuGlyGln</p> <p>121 GATTATGAGCAGCTGGGGTGCAGTGCCTGCAAGTGGACCCCTCTTCGCTGATGAGGCC AspTyrGluGlnLeuArgValArgCysLeuGlnHisArgValThrLeuPheArgAspGluAla 30</p> <p>181 TTCGCCCGGTACCCAGAGCTGGGTTACAAGGACCTGGGTCCCAATCTCCCAAGACC PheProProValProGlnSerLeuGlyTyrLysAspLeuGlyProAsnSerSerLysThr</p> <p>241 TATGGCATCAAGTGAAGCTCCCAAGCACTGCTGCAACCCCAAGTTCATTGTGGAT TyrGlyIleLysTrpLysArgProThrGluLeuLeuSerAsnProGlnPheIleValAsp 100</p> <p>301 GGAGCTACCCGACAGACATCTGCCAGGGAGCCTGGGGGAGCTGGCTCTGGCGGCC GlyAlaThrArgThrAspIleCysGlnGlyAlaLeuGlyAspLysTrpLeuLeuAlaAla</p> <p>361 ATTGCCTCCCTCACTCAACGACCCCTCCTGCACCGAGTGGTCCCGACGGCCAGAGC IleAlaSerLeuThrLeuAsnAspThrLeuLeuHisArgValValProHisGlyLysSer</p> <p>421 TTCAGAATGGTATGCCGGCATCTCCATTTCCAGCTGTGGCAATTTGGGAGTGGGTG PneGlnAsnGlyTyrAlaGlyIlePheHisLysPheGlnLeuTrpGlnPheGlyGluTrpVal 150</p> <p>481 GAGGTGCTGGTGGTACCTGCTGCCATCAAGACGGGAAGCTAGTGTTCGTCACTCT AspValValValAspAspLeuLeuProIleLysAspGlyLysLeuValPheValHisSer</p> <p>541 GCGGAAGGCAAGGAGTTCTGGAGCGCCCTGCTTGAAGGCCATGCAAGGTAATAGGC AlaGluGlyAsnGluPheTrpSerAlaLeuLeuGluLysAlaTyrAlaLysValAsnGly 200</p> <p>601 AGCTACGAGGCCCTGTGAGGGGACAGCCCTCAGAGGGCTTTGAGGACTTCACAGCGGG SerTyrGluAlaLeuSerGlyCysSerTyrGlyPheCysLeuPheAspPheThrGlyGly</p> <p>661 GTTACCAGTGGTACGAGTTCGCGAAGCTCCCACTGACCTCACAGATCCTCAAG ValThrGluTrpTyrGluLeuArgLysAlaProSerAspLeuTyrGlnIleIleLeuLys</p> <p>721 GCGCTGGAGCGGGCTCCCTGCTGGCTGCTCCATAGACATCTCCAGCGTCTAGACATG AlaLeuGluArgGlySerLeuLeuLysIleAspIleSerSerValLeuAspMet 250</p> <p>781 GAGGCCATCACTTTCAAGAAGTGGTGAAGGGCTATTCCTACTCTGTGACCGGGCCCAAG GluAlaIleThrPheLysLysLeuValLysGlyHisAlaTyrSerValThrGlyAlaLys</p> <p>841 CAGGTGAATACCGAGCCAGGGTGGTGGCTGATCCGGATGCGGAACCCCTGGGGGAG GlnValAsnTyrArgGlyGlnValValSerLeuIleArgMetArgAsnProTrpGlyGlu 300</p> <p>901 GTGGAGTGGACGGGAGCTGGAGCGACAGCTCCTCAGAGTGGAAACAACGTGGACCCATAT ValGluTrpThrGlyAlaTrpSerAspSerSerSerGluTrpAsnAsnValAspProTyr</p> <p>961 GAACGGACCAAGCTCCGGCTCAAGTGGAGGACGGGAGTTCGGATCTCATTCCGAGAC GluArgAspGlnLeuArgValLysMetGluAspGlyCyluPheTrpMetSerPheArgAsp 350</p> <p>1021 TTCATCGGGAGITCACCGGCTGGAGATCTGCAACCTCACACCCGACCGCCCTCAAGACC PheMetArgGluPheThrArgLeuGluIleCysAsnLeuThrProAspAlaLeuLysSer 350</p> <p>1081 CGGACCATCCGAAATGGAACACCACTCTACGAAGGACCTGGCGGGGGAGCACC ArgThrIleArgLysTrpAsnThrThrLeuTyrGluGlyThrTrpArgArgGlySerThr</p> <p>1141 GCGGGGGCTGCCAACTACCCAGCCACTTCTGGGTGAACCTCAGTTCAGATCCGG AlaGlyGlyCysArgAsnTyrProAlaThrPheTrpValAsnProGlnPheLysIleArg 400</p>	<p>1201 CTGGATGAGACGGATGACCCGGACCTACCGGGACCCGACTCAGGCTGCAGCTTCGGG LeuAspGluThrAspAspProAspAspTyrGlyAspArgGluSerGlyCysSerPheVal</p> <p>1261 CTCGCCCTTATGCAGAAGCACCGCTCGCCGAGCGCCGCTTCGGCCGACATGGAGACT LeuAlaLeuMetGlnLysHisArgArgArgGluArgArgPheGlyArgAspMetGluThr</p> <p>1321 ATTGGCTTCGGGCTCAGAGAGCTCCCTCGGAGCTGGTGGCCAGCCGCGCTACACTTG IleGlyPheAlaValTyrGluValProProGluLeuValGlyGlnProAlaValHisLeu 450</p> <p>1381 AAGCGTCACTTCCTCGCCAAATGGCTCTCGGGCGGCTCAGAGCAGTTCATCAACCTG LysArgAspPhePheLeuAlaAsnAlaSerArgAlaArgPheSerGluGlnPheIleAsnLeu</p> <p>1441 CGAGGCTCAGCACCCGCTTCGCCCTGCCACCGGGGATATGTGTGTGTGCCCTCCACC ArgGluValSerThrArgPheArgLeuProProGlyGluTyrValValValProSerThr 500</p> <p>1501 TTCGAGCCCAACAAGGAGGGCGACTTCGTGCTGCTTCTCTCAGAGAAGAGTCTGGG PheGluProAsnLysGluGlyAspPheValLeuArgPhePheSerGluLysSerAlaGly</p> <p>1561 ACTGTGGAGCTGGATGACCAGATCCAGCCCAATCTCCCCGATGAGCAAGTGTCTCAGAA ThrValGluLeuAspAspGlnIleGlnAlaAsnLeuProAspGluGlnValLeuSerGlu</p> <p>1621 GAGGACATTGACGAGAATTCGAAGGCCCTTCACCGAGCTGCCAGGGGACGACATGGAG GluGluIleAspGluAsnPheLysAlaLeuPheArgGlnLeuAlaGlyGluAspArgMet 550</p> <p>1681 ATCAGCGTGAAGGAGTGGCGCAATCTCAATAGGATCATCAGCAACACAAAGACCTG IleSerValLysGluLeuArgThrIleLeuAsnArgIleIleSerLysHisLysAspLeu 600</p> <p>1741 CGGACCAAGGCTTCAGCTAGAGTGGCGCCGAGCATGGTGAACCTCATGGATCGTGAT ArgThrLysGlyPheSerLeuGluSerCysArgSerMetValAsnLeuMetAspArgAsp 600</p> <p>1801 GCGAATGGGAAGTGGGCTGGTGGAGTCAACATCTCTGGAACCGCATCCGGAATAC GlyAsnGlyLysLeuGlyLeuValGluPheAsnIleLeuTrpAsnArgIleArgAsnTyr</p> <p>1861 CTGTCCATCTCCCGAAGTTTACCTGGCAAGTGGCGAGCATGATGGCTACGAGATG LeuSerIlePheArgLysLeuAspLeuAspLysSerGlySerMetSerAlaTyrGluMet</p> <p>1921 CGGATGGCCATTGAGTGGCGAGCTCAAGCTCAACAAGAAGCTGTACGAGCTCATCATC ArgMetAlaIleGluSerAlaGlyPheLysLeuAsnLysLeuTyrGluLeuIleIle 650</p> <p>1981 ACCCGCTACTCGGAGCCGACCTGGCGGCTGACTTTGCAAAATTTCTGCTGCTGGTG ThrArgTyrSerGluProAspLeuAlaValAspPheAspAsnPheValCysCysLeuVal</p> <p>2041 CGGCTAGACACATGTTCCGATTTTCAAACTCTGGACACAGATCTGGATGGAGTTGTG ArgGluLeuThrMetPheArgPhePheLysThrLeuAspThrAspLeuAspGlyValVal 700</p> <p>2101 ACCTTTGACTTGTAAAGTGGTGCAGCTGACCATGTTTGCATGAGCCAGGGACTCGGT ThrPheAspLeuPheLysTrpLeuGlnLeuThrMetPheAla</p> <p>2161 CCCCTTGGCTGCTCCCTCCCTCCCTGCTGCCAAGCTCGCCCTTACCACACCACAC</p> <p>2221 CAGGCCACCCGAGTCAAGTCCCTCTGGAGCAGAGGACGAGCTCGTCTCTCTGTC</p> <p>2281 CCCTCTCTCCAGCCACCATCGTTCAATCTGCTCGGGCAGAAGTGTGGCCCTCGCT</p> <p>2341 GTGCCAGCATGGGCTGGGATGGACTCCCTGGGCCCCACCATGGCCAAAGCAGGAAG</p> <p>2401 CAGCTTTCGCTTGTCTGCTCCGGACACCCCGGGTTTCCCGACATCTGATGTGTCT</p> <p>2461 CCCTCTCCCACTTCAGAGGCCACCCACTCAGCACCCAGCCCTGGCTTGGCTGAGAC</p> <p>2521 TATAAATAAACCCTAGCTCGACACAGCTGCGATCCAGGGCTGTGGAGCCGCTCC</p> <p>2581 GGCTCGGGAGGCCCGGGCTGGGAACCGCTGCGCTTCTCGCCGAAGCCACCGCC</p> <p>2641 CCTCTGCTTCCCTGGCTGCTGGCCGACAGGAGCTGCCAGCTGTGGGCTGGCC</p> <p>2701 CTTCCTCTCTGCTCTTTTTTATATTAGTATTTAAAGGGGACTTTCAGGACTTG</p> <p>2761 TGTACTGGTATGGGGTGGCAGAGGCACTAGGCTGGGGTGGGAGTCCCGTGTCCA</p> <p>2821 TATAGAGGAACCCCAATATAAAGGCCCCACATCTGCTGTGA</p>
---	--

Fig.2. Nucleotide sequence of the large subunit of human μ CANP cDNA and the deduced amino acid sequence. Two residues at the active site (Cys-115, His-272) are boxed. The amino acid residues corresponding to the Ca^{2+} -binding loop of the 4 EF-hand structures are underlined. Closed triangles indicate the boundaries of domains. The polyadenylation signal is double-underlined.

sequence homology between chicken CANP and human μ CANP is 70%. Furthermore, the homologies of the two proteins are 54, 78, 72 and 65% for domains I, II, III and IV, respectively.

The homology in domain II is higher than those for the other domains.

A clear structural difference was found between the N-terminal sequences of human μ CANP and

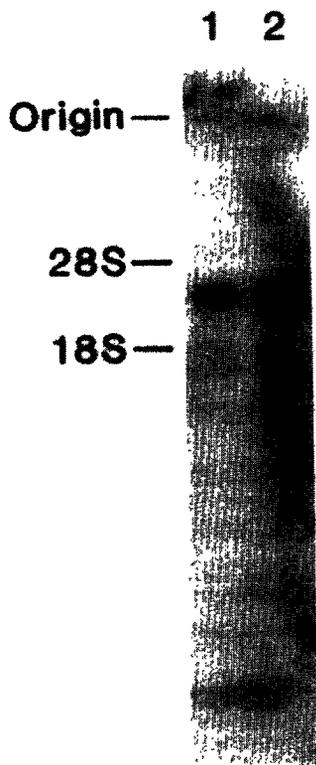


Fig.3. Blot hybridization of human spleen RNA. Poly(A)⁺ RNA (lane 1, ~0.3 µg) and total RNA (lane 2, ~2.0 µg) were analyzed by the procedure described in [12] using nylon membrane. The probe used for hybridization was the *Rsa*I fragment (nos 1967–2763 in fig.2) labelled by nick-translation with [α -³²P]dCTP. The size marker was human rRNA.

chicken CANP. The former is larger than the latter by 7 amino acid residues. The Ca²⁺ sensitivity of chicken CANP increases significantly upon removal of the N-terminal 17-residue peptide by autolysis [7,18]. Thus, the N-terminal region is also important for the Ca²⁺ sensitivity of CANP. In this respect, a clear structural difference in the N-terminal region is noticeable.

From these results we assume that various CANPs have a common fundamental mechanism for the expression of protease activity and that their Ca²⁺ sensitivities are determined mainly by differences in the structure of the C-terminal Ca²⁺-binding domain (domain IV). The N-terminal region (domain I) of native CANP might interact with domains II and/or IV, and thus

repress the Ca²⁺-dependent proteolytic activity. Autocatalytic cleavage of the N-terminal region affects the interaction between functional domains and changes the Ca²⁺ sensitivity of CANP.

ACKNOWLEDGEMENTS

We are grateful to Drs Y. Sukenaga and Y. Shibasaki for providing the λ gt10 cloning system and for useful discussions. This study was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, and Grants-in-Aid for New Drug Development from the Ministry of Health and Welfare, and the National Center for Nervous, Mental and Muscular Disorders of the Ministry of Health and Welfare of Japan.

REFERENCES

- [1] Imahori, K. (1982) in: Calcium and Cell Function (Cheung, W.Y. ed.) vol.3, pp.473–485, Academic Press, New York.
- [2] Murachi, T. (1983) in: Calcium and Cell Function (Cheung, W.Y. ed.) vol.4, pp.377–410, Academic Press, New York.
- [3] Suzuki, K., Kawashima, S. and Imahori, K. (1984) in: Calcium Regulation in Biological Systems (Ebashi, S. et al. eds) pp.213–226, Academic Press, New York.
- [4] Yumoto, N., Kikuchi, T., Sasaki, T. and Murachi, T. (1984) *J. Biochem.* 96, 1531–1537.
- [5] Kawasaki, H., Imajoh, S., Kawashima, S., Hayashi, H. and Suzuki, K. (1986) *J. Biochem.* 99, 1525–1532.
- [6] Kikuchi, T., Yumoto, N., Sasaki, T. and Murachi, T. (1984) *Arch. Biochem. Biophys.* 243, 639–645.
- [7] Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. and Suzuki, K. (1984) *Nature* 312, 566–570.
- [8] Kawashima, S., Nomoto, M., Hayashi, M., Inomata, M., Nakamura, M. and Imahori, K. (1984) *J. Biochem.* 95, 95–101.
- [9] Suzuki, K., Ohno, S., Imajoh, S., Emori, Y. and Kawasaki, H. (1985) *Biomed. Res.* 6, 323–327.
- [10] Emori, Y., Kawasaki, H., Sugihara, H., Imajoh, S., Kawashima, S. and Suzuki, K. (1986) *J. Biol. Chem.* 261, 9465–9471.
- [11] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 18, 5294–5299.
- [12] Maniatis, T., Fritsch, E.F. and Sambrook, J.

- (1982) in: *Molecular Cloning*, Cold Spring Harbor Laboratory, NY.
- [13] Gubler, U. and Hoffman, B.J. (1983) *Gene* 25, 263–269.
- [14] Huynh, T., Young, R. and Davis, R.W. (1985) in: *DNA Cloning, A Practical Approach* (Glover, D. ed.) vol.1, pp.49–78, IRL Press, London.
- [15] Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499–560.
- [16] Hatanaka, M., Kikuchi, T. and Murachi, T. (1983) *Biomed. Res.* 4, 381–388.
- [17] Suzuki, K., Hayashi, H., Hayashi, T. and Iwai, K. (1983) *FEBS Lett.* 152, 67–70.
- [18] Suzuki, K., Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H. and Kisaragi, M. (1985) in: *The Biological Role of Proteinases and Their Inhibitors in Skin* (Ogawa, H. et al. eds) pp.111–120, University of Tokyo Press, Tokyo.
- [19] Tufty, R.M. and Kretsinger, R.H. (1975) *Science* 187, 167–169.