Human ClpP protease: cDNA sequence, tissue-specific expression and chromosomal assignment of the gene

Peter Bross[a,*], Brage S. Andresen[a], Inga Knudsen[a], Torben A. Kruse[b], Niels Gregersen[a]

[a]*Center for Medical Molecular Biology, Aarhus University Hospital and Faculty of Health Sciences, Skejby Sygehus, Brendstrupgårdsvej, 8200 Aarhus N, Denmark*

[b]*Institute for Human Genetics, Aarhus University, 8000 Aarhus C, Denmark*

Received 31 October 1995

Abstract **We identified three overlapping human expressed sequence tags with significant homology to the *E. coli* ClpP amino sequence by screening the EMBL nucleotide database. With this sequence information we applied 5′ and 3′-rapid amplification of cDNA ends (RACE) to amplify and sequence human clpP cDNA in two overlapping fragments. The open reading frame encodes a 277 amino acid long precursor polypeptide. Two ClpP specific motifs surrounding the active site residues are present and extensive homology to ClpP's from other organisms was observed. Northern blotting showed high relative expression levels of clpP mRNA in skeletal muscle, intermediate levels in heart, liver and pancreas, and low levels in brain, placenta, lung and kidney. By analysis of human/rodent cell hybrids the human clpP gene was assigned to chromosome 19.**

Key words: ATP-dependent protease; Protein degradation; ClpP; cDNA sequence; Tissue-specific expression, *Homo sapiens*

1. Introduction

ATP-dependent proteases have first been identified in *E. coli*. Two such proteases, Lon (also called La) and ClpAP (also referred to as Ti), have been characterized (for recent reviews see [1,2]. Both are involved in ATP-dependent degradation of abnormal (i.e. incorrectly folded or unfolded) proteins. Homologs of the Lon protease have been identified in other bacteria and mitochondria [3,4]. ClpAP protease consists of a regulatory unit (ClpA) with chaperone characteristics which contains the ATPase domain [5] and a proteolytic subunit (ClpP) [6]. The ClpA subunit belongs to a family of proteins with at least three homologs in *E. coli* [7] and homologs in eucaryots (Hsp78, Hsp104 [8,9]. In *E. coli*, ClpP has been found complexed with either ClpA or the homologous ClpX [7]. It appears that, dependent on which of the regulatory subunits ClpP is complexed with, different substrates are processed proteolytically. Recently, it has been shown that clpA can complement the chaperone function of DnaK [10]. This suggests that the function of the ATP-ase subunit in the ClpAP complex may be to unfold the protein and present it to the protease subunit. The ClpP subunit is organized in a tetradecamer consisting of two heptameric rings [11]. ClpA or ClpX are present as dimers which assemble into a hexamer upon substrate binding followed by complex formation with the ClpP double heptamer.

Sequences for bacterial and chloroplast clpP genes have been reported. Immunological experiments indicated the presence of ClpP homologs in mammalian cells [12]. We are interested in inherited defects in enzymes involved in mitochondrial β-oxidation of fatty acids. Many of the mutations in medium-chain acyl-CoA dehydrogenase (MCAD) deficiency cause impaired folding of the proteins [13]. Impaired folding results in premature degradation of the mutant MCAD polypeptides and we have shown that the availability of hsp60/hsp10 chaperones can modulate the level of MCAD enzyme present [13,14]. Recently it has been suggested that an interplay exists between the hsp60/hsp10 system and ClpP protease [15]. This prompted us to determine the cDNA sequence of human ClpP in order to be able to investigate its possible contribution of proteins carrying folding mutations.

2. Materials and methods

2.1. Sequencing of human clpP cDNA

Sequence information from three overlapping human expressed sequence tags (accession: T34278, T04899, and T31415) was used to locate a clpP-specific sense (CGGGGCCAAGCCACAGACAT) and an antisense (TGGACCAGAACCTTGTCTAAGATGC) primers. Adaptor ligated cDNAs from human placenta mRNA and total RNA from cultured human fibroblast cells were produced using the Marathon cDNA kit (Clonetech, Palo Alto). Subsequently, rapid amplification of cDNA ends (RACE) was performed using the clpP specific primers and the adaptor primer included in the kit. PCR was performed with cDNA as template in 100 μl with 15 pmol of each primer in 1 × PCR buffer (0.067 M Tris, 0.166 M ammonium sulfate, 0.1 M β-mercaptoethanol and 8% DMSO). An automated cykler (Perkin Elmer Cetus, Nowark) was used with the following program: 1 cylce of 95°C for 5 min and 68°C for 5 min, followed by 35 cycles of 95°C for 1 min and 66°C for 5 min. In the last cycle, polymerization was for 10 min at 74°C. PCR was started after the first denaturation step by addition of 5 units of recombinant Taq polymerase (Perkin Elmer Cetus, Nowark) and 0.5 units of Pfu polymerase (Stratagene, La Jolla). A band of approximately 800 basepairs was obtained in the 5′-RACE reactions and one of approximately 540 basepairs in the 3′-RACE reactions. Both fragments were reamplified using biotinylated adaptor primer and the s-1 or as-1 primer respectively. The PCR products were captured with streptavidin-coated magnetic beads (Dynal, Skøjen) and direct sequencing of the PCR products was performed using the Prism Sequenase Terminator single-stranded DNA sequencing kit (Applied Biosystems). The products from the sequencing reactions were analysed on an Applied Biosystems 373A automated sequencer. In parallel, 5′-and 3′-RACE products from the first PCR reactions were reamplified with non-biotinylated adaptor primer and the respective sense or antisense primer. The reamplified products were gel-purified and

*Corresponding author. Fax: (45) (89) 496 018.
E-mail: bross@biobase.dk

Sequence data from this article have been supplied to the EMBL/GenBank Data Libraries under accession no. Z50853.

Abbreviations: EST, expressed sequence tag; Hsp, heat-shock protein; RACE, rapid amplification of cDNA ends.

SSDI 0014-5793(95)01353-9

cloned into the TA-vector pCRII (InVitrogen, San Diego) as recommended by the supplier. Double-stranded plasmid DNA from four 5'- and four 3'-clones was isolated and sequenced using the Prism Sequenase Terminator double-stranded DNA sequencing kit (Applied Biosystems) and the automated sequencer.

2.2. Tissue-specific expression

Northern blotting was performed on a multiple tissue Northern blot filter (human MTN blot; Clonetech) as recommended by the supplier. The probe used was a 832 basepair *Eco*RI fragment excised from one of the 5'-RACE clones and spanning nucleotides −27 to +740 of the human clpP cDNA labelled by random hexamer labelling [16] using [^{32}P]dCTP (Amersham).

2.3. Chromosome localization

Cell hybrid DNA from the NIGMS Human/Rodent Somatic Cell Hybrid Mapping Panel #1 were obtained from the NIGMS Human Genetic Mutant Cell Repository (Camden, NJ). Restriction enzyme digestion of hybrid cell DNA, Southern transfer, and filter hybridization were carried out as described [17]. The clpP probe used was a restriction fragment comprising nucleotides 553 to the poly-A tail of the cDNA sequence.

3. Results and discussion

3.1. cDNA sequence

Using the TFastA program of the GCG sequence analysis package (University of Wisconsin Genetics Computer Group) three overlapping human expressed sequence tags (EST's) which displayed considerable homology to the *E. coli* ClpP amino acid sequence were identified in the EMBL/GenBank databases. With primers situated in the region covered by the EST's, the RACE technique was used to amplify two overlapping fragments covering the whole cDNA of human clpP from placenta and fibroblast cDNA synthesized from mRNA or total RNA respectively. The PCR fragments were subjected to direct sequencing. In parallel, both fragments from amplified placenta cDNA were cloned into pCRII and eight selected clones were sequenced (four covering the 5'-end and four covering the 3'-end).

The sequencing data are summarized in Fig. 1. The sequence is 1044 basepairs long. It comprises a 19 basepair 5'-untranslated region, an open reading frame starting with an ATG start codon of 831 basepairs, and a 195 basepair long 3' untranslated region. The last two bases preceding the poly-A tail could not be determined unambiguously as the four 3'-clones contained variable sequences at these positions. This is probably due to the fact that the lock docking primer used for first strand cDNA synthesis carries degenerate nucleotides in the two last bases at the 3'-end.

3.2. ClpP amino acid sequence

The human ClpP amino acid sequence deduced from the cDNA sequence comprises an 277 amino acid open reading frame starting with a methionine codon (Fig. 1). There is no other methionine codon within the first 87 amino acids suggesting that this is the authentic start codon. The first 59 amino acids preceding proline-60, which constitutes the first amino acid conserved in all known ClpP sequences (see below), have the characteristics of a mitochondrial transit peptide with the recurrent occurrence of arginine residues and the absence of negatively charged amino acids [18]. There are several potential cleavage sites for the mitochondrial processing protease three of which are marked in Fig. 1.

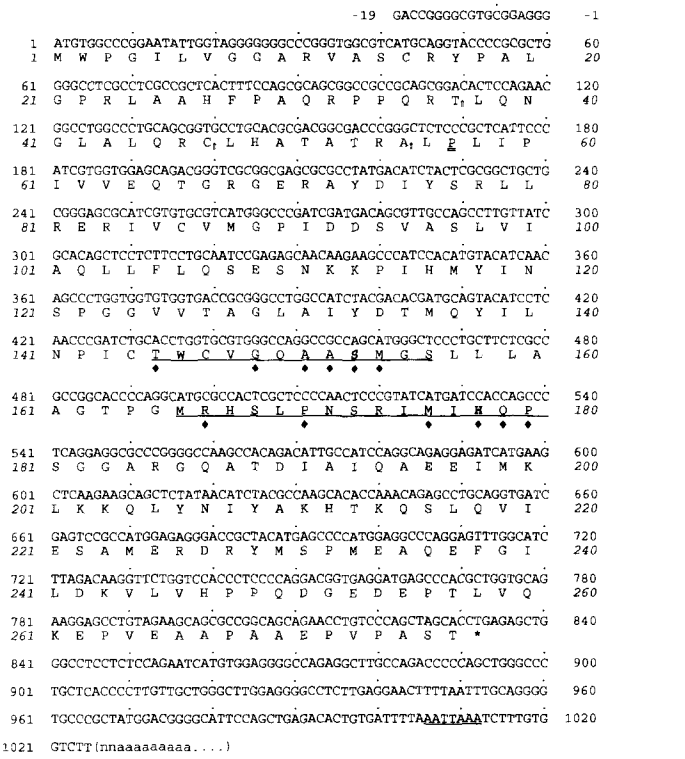

Fig. 1. cDNA sequence and deduced amino acid sequence of human clpP. The numbering of nucleotides and amino acids start at the first residue of the coding sequence. The last two nucleotides preceding the poly-A tail determined by the cDNA synthesis primers and the start of the poly-A tail are shown in brackets (see text). The putative polyadenylation site is underlined. The first amino acid (proline-60) conserved in all known ClpP sequences is denoted by a double underline. Three possible cleavage sites for the mitochondrial processing protease are indicated by arrows. The consensus sequence around the catalytically active serine and histidine residues (in boldface) are underlined and positions conserved in human ClpP are denoted with diamonds.

Database search identified twelve full-length sequences with considerable homolgy to human ClpP: seven chloroplast encoded ClpP's, three eubacterial ones and two others. The sequences align well from proline-60 to aspartic acid 242 in human pre-ClpP. Both these residues are a fully conserved in all sequences. Only few short gaps have to be introduced. Alignment of human ClpP with one representative each of the chloroplast and bacterial sequences and the two other ClpP sequences is shown in Fig. 2. The 22 residues fully conserved in all thirteen ClpP sequences available to us are marked. Human ClpP displays highest homology towards ClpP from the nematode *Cænorhabditis elegans* (C.e.; 58% identity), followed by the bacterial ClpP's (48–56% identity) and lowest toward chloroplast ClpP (32–38% identity). The C.e. ClpP sequence in the database which originates from a genome sequencing project, predicts methionine 16 as the starting amino acid. If this were the case, the sequence would lack a mitochondrial leader peptide. However, there is another possible methionine start codon further upstream which would result in a fifteen amino acids longer amino terminus as shown in Fig. 2.

The percentage of identity of human ClpP towards ClpP from tomato (L.e.; 34%) is in the same range as that towards chloroplast ClpP. However, tomato ClpP displays slightly

```
H.s.        1   ........................MWPGILVGGARVASCRYPALGPRLAAHFPAQRPPQRTLQNGLALQRCLHAT
C.e.        1   ..............................................................MLR LvTSS sAS SmS s
L.e.        1   MVTSAIAGTSIVPVSSRHQTSFSSLLSSRSlRKNVVS LRSP SDSSDIGFsSKKLGI LKFNEYESG HTNSSYGV
E.c.        1   ...........................................................................MSYSGErDN
T.ae.       1   ........................................................................................
conserved*  1

H.s.       52   ATRALPLIPIVVEQTGRGERAYDIYSRLLRERIVCVMGPIDDSVASLVIAQLLFLQSESNKKPIHMYINSPGGVVTA
C.e.       20   VQSRvG-  F idnE k   t        d    l T v  Fi  Al        G              S
L.e.       78   IVAKEGAN PIMPAVMTPVG L lStV F N  iFiGQ vnSa  QK  s  vT AtIDENAD Li l C   STYs
E.c.       12   FtPHmA v m i   S    sf       k  viFlT Qve Hm N iv  m   ea NPE D Yl       i
T.ae.       1   MPIgv Kv YRiPRDEEaTW-V l NVmY   TlFlGQE RCEitNHiTgLmvy SI DGISD Flf      WlIs
conserved* 78         P          D       R                    L           N  GG

H.s.      129   GLAIYDTMQYILNPICTWCVGQAASMGSLLLAAGTPGMRHSLPNSRIMIHQPSGG-ARGQATDIAIQAEEIMKLKKQ
C.e.       96         i M SA vS  Vi   s       C  EK   Sa       v      - Q TCs  V R     Tr  r
L.e.      155   V     C SW KPKvG V F A   Q a    g EK   Yam  a       QS -CG HVE vRR VN AVQSrQK
E.c.       87    ms      f KPDvS I m       aF  t  aK K FC    v      L  -YQ      E H R  l v GR
T.ae.      77    m  f    TvTPD Y R l I    a Fi Lg E TK IaF Ha   l   aSaYY aRtPeSLieV  lH vrEM
conserved* 155              T...G.A.S...     G   .R.....M.HQP

H.s.      205   LYNIYAKHTKQSLQVIESAMERDRYMSPMEAQEFGILDKVLVHPPQDGEDEPTLVQKEPVEAAPAAEPVPAST*
C.e.      172    NE  Vh  GM YDE  Ktld   f  AH  LK  lv QiET NGSMPS *.....................
L.e.      231   vDKm  AF G  iem qtYT    f  SA  M   li G  ETEY*.........................
E.c.      163   mNElM L  G   eQ  RDT    fl AP  V y lv Si T RN*.........................
T.ae.     154   iTRv  Vr GKPFW vSED    Vf  AD  KAyElv I GDEMIDKHC TDPvWFP MFKDW*.........
conserved* 232                      RD  S  E      D
```

Fig. 2. Alignment of ClpP sequences (EMBL accession numbers in brackets). The human (H.s.; Z50853) ClpP amino acid sequence deduced from the cDNA sequence was aligned with ClpP from the nematode *Caenorhabditis elegans* (C.e.; Z49073), *Lycopersicon esculentum* (L.e.; L38581), *Escherichia coli* (E.c.; J05534) and *Triticum aestivum* (T.ae.; X54484) using the pileup program (GCG package). Only differences in relation to the human sequence are shown. Gaps are indicated by hyphens. Conservative replacements are indicated by lower case letters. *The conserved residues given in the bottom line are those positions fully conserved in the above four sequences and the following other ClpP sequences available in the database: X04465, Z00044, X15901, L28807, D17510, L07793, U16135 and X86563. The consensus sequences surrounding the catalytic histidine and serine residues are underlined and conserved positions given in boldface. A motif comprising the putative third residue in the catalytic triad (aspartic acid) is underlined.

lower homology towards chloroplast ClpP's (32–34% identity) than towards bacterial ClpP's (36–42% identity). It is therefore probable that the tomato ClpP as the Caenorhabditis elegans and the human ClpP represents a mitochondrial ClpP homolog.

Although no significant homology of ClpP [12] towards classical serine proteases can be observed, the proteolytic activity of the ClpP proteases appears to be accomplished by a charge relay system similar to that known for these enzymes with serine, histidine and aspartic acid forming the active site triad [12]. Consensus sequences surrounding two of the three catalytic residues in ClpP sequences have been identified (Prosite PS00381 and PS00382). As indicated in Fig. 1, these motifs are present in the human ClpP sequence confirming that it represents an authentic ClpP protease. Interestingly, the first residue in the consensus sequence around the catalytically active histidine residue is methionine in the three mitochondrial ClpP's whereas lysine is present at this position in all bacterial and chloroplast sequences. This methionine residue may therefore be characteristic for mitochondrial ClpP sequences. The third residue in the catalytic triad in serine proteases is an aspartic acid. There are three aspartic acid residues fully conserved in all ClpP sequences. Two of these are located in a stretch of conserved residues (residues 226–241 in human ClpP) close to the carboxyl end. This stretch may thus represent the motif containing the third residue of the catalytic triad. Whether this is the case and which aspartic acid residue is involved has to be investigated by site-directed mutagenesis.

From *E. coli* ClpP it is known that it cleaves off the 14 amino terminal residues autocatalytically [19]. The long amino terminal extensions of human and tomato mitochondrial ClpP's preceding the homologous region would allow a two step processing, first by mitochondrial leader peptidase and then autocatalytically. Whether this is the case has to be addressed experimentally.

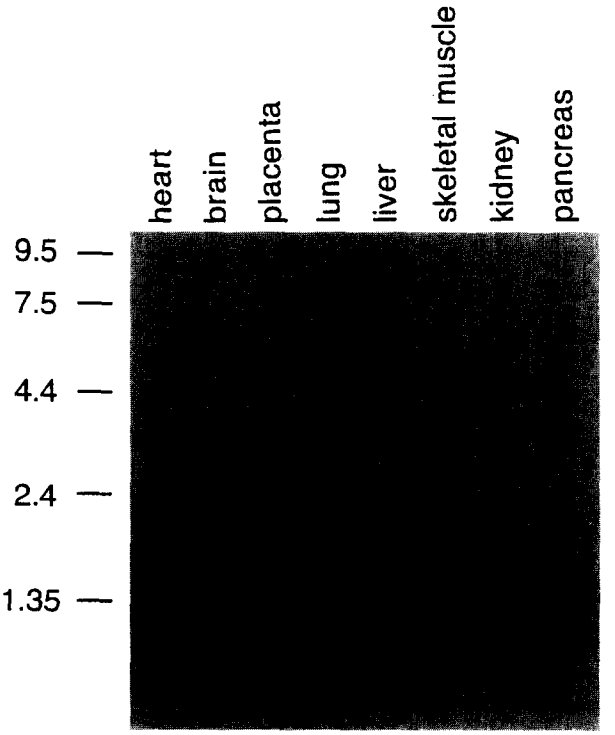

Fig. 3. Expression pattern of clpP in human tissues. A ^{32}P-labelled probe covering nucleotides −27 to 780 of the cDNA sequence was hybridized to a Multiple Tissue Northern blot as described in section 2. The lenght of molecular weight standards in kilobases are given at the left margin.

3.3. Tissue specific expression

A Northern blot filter with lanes containing similar quantities of human mRNA from different tissues was hybridized with a cDNA probe covering nucleotide positions −27 to +740 of the cDNA sequence. As shown in Fig. 3, one specific band with an apparent length of approximately 1.35 kilobases was observed in all tissues analysed. This is in good agreement with the length of the cDNA assuming an approximately 150–200 bases long poly-A tail. No other bands were detected indicating the presence of one single type of spliced transcript. The strongest signal was observed in muscle tissue (skeletal muscle and heart). The order of decreasing intensity was: skeletal muscle > heart > liver and pancreas > brain, placenta, lung and kidney. The expression pattern is similar to that observed by Wang et al. for human Lon protease mRNA with the exception that relatively higher amounts of lon transcript were observed in brain and liver. Interestingly, the pattern for clpP mRNA matches with that for medium-chain and very-long-chain acyl-CoA dehydrogenase (Brage S. Andresen, unpublished), two enzymes involved in the mitochondrial β-oxidation of fatty acids. An involvement of ClpP in the degradation of misfolded intermediates of these proteins will be investigated.

3.4. Chromosome localization

DNA samples from 18 different human/rodent cell hybrids were analysed by Southern blotting using a cDNA probe comprising nucleotides 553 to the poly-A tail. A human-specific signal was observed which allowed to assign the human clpP gene to chromosome 19. Strong cross-hybridization signals towards mouse and hamster DNA sequences were observed reflecting the high conservation of clpP sequences between species.

Acknowledgements: This work was supported by grants from the Danish Medical Research Council and the Danish Center for Human Genome Research.

References

[1] S. Gottesman and M.R. Maurizi (1992) Microbiol. Rev. 56, 592–621.
[2] A.L. Goldberg (1992) Eur. J. Biochem. 203, 9–23.
[3] N. Wang, S. Gottesman, M.C. Willingham, M.M. Gottesman and M.R. Maurizi (1993) Proc. Natl. Acad. Sci. USA 90, 11247–11251.
[4] C.K. Suzuki, K. Suda, N. Wang and G. Schatz (1994) Science 264, 273–276.
[5] C. Squires and C.L. Squires (1992) J. Bacteriol. 174, 1081–1085.
[6] M.R. Maurizi, M.W. Thompson, S.K. Singh and S.H. Kim (1994) Methods Enzymol. 244, 314–331.
[7] S. Gottesman, W.P. Clark, V. de Crecy-Lagard and M.R. Maurizi (1993) J. Biol. Chem. 268, 22618–22626.
[8] S.A. Leonhardt, K. Fearson, P.N. Danese and T.L. Mason (1993) Mol. Cell Biol. 13, 6304–6313.
[9] F. Perier, C.M. Radeke, K.F. Raab-Graham and C.A. Vandenberg (1995) Gene 152, 157–163.
[10] S. Wickner, S. Gottesman, D. Skowyra, J. Hoskins, K. Mckenney and M.R. Maurizi (1994) Proc. Natl. Acad. Sci. USA 91, 12218–12222.
[11] M. Kessel, M.R. Maurizi, B. Kim, E. Kocsis, B.L. Trus, S.K. Singh and A.C. Steven (1995) J. Mol. Biol. 250, 587–594.
[12] M.R. Maurizi, W.P. Clark, S.H. Kim and S. Gottesman, (1990) J. Biol. Chem. 265, 12546–12552.
[13] P. Bross, C. Jespersen, T.G. Jensen, B.S. Andresen, M. J. Kristensen, V. Winter, A. Nandy, F. Krautle, S. Ghisla, L. Bolund, J.J. P. Kim and N. Gregersen (1995) J. Biol. Chem. 270, 10284–10290.
[14] P. Bross, B.S. Andresen, V. Winter, F. Krautle, T.G. Jensen, A. Nandy, S. Kølvraa, S. Ghisla, L. Bolund and N. Gregersen (1993) Biochim. Biophys. Acta 1182, 264–274.
[15] O. Kandror, L. Busconi, M. Sherman and A.L. Goldberg, (1994) J. Biol. Chem. 269, 23575–23582.
[16] A.P. Feinberg and B. Vogelstein (1984) Anal. Biochem. 137, 66–267.
[17] T.A. Kruse, L. Bolund, K. Grzeschik, H.H. Ropers, J. Olsen, H. Sjöström and O. Norén (1988) FEBS Lett. 239, 305–308.
[18] Y. Gavel and G. von Heijne (1990) Protein Eng. 4, 33–37.
[19] M.R. Maurizi, W.P. Clark, Y. Katayama, S. Rudikoff, J. Pumphrey, B. Bowers and S. Gottesman (1990) J. Biol. Chem. 265, 12536–12545.