

# Cloning and sequence analysis of cDNA for a human homolog of eubacterial ATP-dependent Lon proteases

Alexander Yu. Amerik<sup>a,\*</sup>, Galina V. Petukhova<sup>a</sup>, Vitaly G. Grigorenko<sup>a</sup>, Igor P. Lykov<sup>a</sup>, Serge V. Yarovoi<sup>a</sup>, Valery M. Lipkin<sup>a</sup>, Alexander E. Gorbalenya<sup>b</sup>

<sup>a</sup>*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow 117871, Russian Federation*

<sup>b</sup>*Institute of Poliomyelitis and Viral Encephalitis, Russian Academy of Medical Sciences, Moscow Region 142782, Russian Federation*

Received 20 December 1993; revised version received 4 January 1994

## Abstract

Overlapping cDNA clones containing mRNA for a putative Lon protease (LonHS) were isolated from cDNA libraries prepared from human brain poly(A)<sup>+</sup> RNA. The determined nucleotide sequence contains a 2814-bp open reading frame with two potential initiation codons (positions 62–64 and 338–340). The 5′-terminal 337-nucleotide fragment of LonHS mRNA is highly enriched with G and C nucleotides and could direct synthesis of the LonHS N-terminal domain. More likely this region promotes initiation of protein synthesis from the second AUG codon in a cap-independent manner. The amino acid sequence initiated at the second AUG codon includes 845 residues, over 30% of which are identical to those of eubacterial Lon proteases. Residues of the ‘A’ and ‘B’ motifs of NTP-binding pattern and a plausible catalytic serine residue are conserved in LonHS. Northern blot analysis revealed LonHS mRNA in lung, duodenum, liver and heart, but not in thymus cells.

**Key words:** cDNA cloning; Lon protease; ATP-dependent proteolysis; Long 5′-untranslated region; Human mRNA

## 1. Introduction

Ubiquitin, 20S and 26S protease complexes are the best studied ATP-dependent systems which contribute to the control of protein turnover in eukaryotic cells [1]. Besides these, other proteases appear to be involved in ATP-consuming proteolysis. Homologs of the bacterial Clp protease were found in various eukaryotic cells [2,3]. An extensively characterized *E. coli* Lon protease which is involved in the rapid degradation of short-lived regulatory and abnormal proteins [4–7] also appears to have homologs in eukaryotes. ATP-dependent enzymes similar to Lon protease were purified from adrenal cortical [8] and placental [9] mitochondria. In recent studies of human brain mRNAs, a 351 bp fragment was sequenced whose putative amino acid product was similar to a region including the ‘A’ motif [10] of the NTP-binding pattern of the Lon polypeptide [11] (see also below).

\* Corresponding author. Fax: (7) (095) 310 70 07;  
E-mail: amerik@enzyme.iobc.msk.su.

**Abbreviations:** LonBB, *B. brevis* Lon protease; LonEC, *E. coli* Lon protease; LonHS, *H. sapiens* putative Lon protease; LonMX, *M. xanthus* Lon protease; ORF, open reading frame; UTR, untranslated region; NTP, nucleoside triphosphate.

The nucleotide sequence presented here has been deposited in the EMBL Nucleotide Sequence Database under accession numbers X74215 and X76040 (*H. sapiens* mRNA for Lon protease-like protein)

This communication presents unequivocal evidence of existence of eukaryotic Lon protease homolog. Cloning and sequencing of the cDNA for a putative Lon protease from human brain are reported. Tissue specificity of its expression is demonstrated.

## 2. Materials and methods

RNA was purified according to the guanidine thiocyanate method [12]. Poly(A)<sup>+</sup> RNA was isolated and used as a template for cDNA synthesis [13]. Double-stranded cDNA was ligated with *Sma*I-digested plasmid pSP64 and used for the transformation of *E. coli* MH I cells. The library was blotted onto nylon membrane and hybridized with a <sup>32</sup>P-labeled synthetic oligonucleotide probe.

Sequencing was performed according to the method of Sanger et al. [14]. Sequences were analyzed using MicroGenie software.

For Northern blot analysis poly(A)<sup>+</sup> RNA (5 mg per line) was electrophoresed on 1.2% agarose-formaldehyde gels [13] and transferred to a nylon membrane by vacuum blotting. Blots were hybridized with either LonHS or actin <sup>32</sup>P-labeled cDNA probes at 42°C overnight in the presence of formamide.

Multiple amino acid sequence alignments were generated in a step-wise manner by the OPTAL program [15]. The SwissProt database (release 26) was searched using the BLITZ program, while the EMBL database (release 35) was scanned with the QUICK program.

## 3. Results and discussion

### 3.1. Isolation and sequencing of cDNA encoding putative Lon protease

In order to isolate human cDNA for putative Lon

```

1 GCGCCAGCAGCTGCTACGTGCGACTGTGGGAGCGCGCGGTGCTGGGTGCTGCGCGGCCGATGCTGGCCCGCCCGGGGGCGGGTCCCACATCGACGAGGA
1
104 GCGTGGTGTCCGAGBCCACGGACCTGCGACGCCCTCCTCCTTGGGCACTGTGGGCGGAGCCCGCAATGGGGCCAAATGGCGGGGTTTGGGAA
15 a w l l r g q r t c d a s p p w a l w g r g p a l g g q w r g f w e
206 GCGAGCAGCCCGGGGAGGCGCATTCTCGGGGGGAGGACGCTCCGAGGGCGCGCGGAGGAAGGAGCCGCGGGCGGGGAGCGCGCGCGCGCGCGG
49 a s s r g g g a f s g g e d a s e g g a e e g a g g a g g a g g
308 GAAGCCCGGTATAACGGCGCTCACGCCCATGACGATCCCGATGTTTCCGCACCTGCCGTCATGCCATCACCCGCAACCCGGTGTCCCGCGCTT
83 e g p v i t a l t p m T I P D V F P H L P L I A I T R N P V F P R F
410 ATCAAGATTATCGAGGTAAAAATAAGAAGTTGGTTGAGCTGCTGAGAAGGAAAGTTGCTCTCGCCAGCCCTTATGTCCGGCTCTTTCAAAGAGAGATGAC
117 I K I I E V K N K K L V E L L R R K V R L A Q P Y V G V F L K R D D
512 AGCAATGAGTCGGATGTGGTCGAGAGCCTGGAATCTACCACCGGGACGTTTGGCCAGATCCATGAGATGCAGGACCTTGGGCAAGCTGGCGCAT
151 S N E S D V V E S L D E I Y H T G T F A Q I H E M Q D L K L R M
614 ATGTCATGGGACAGAAAGAGTCCATATCAGCAGACAGCTGGAGGTGGAGCCCGAGGAGCCGAGGGGGAAGAACAGCAAGCCCGCAGGAAGTCAAG
185 I V M G H R R V H I S R Q L E V E P E E P E A E N K H K P R R K S K
716 CGGGCAAGAGGAGGCGGAGGACGCTGAGCGCCAGGACCCGCGGAGCTGGGATGGAGCCACCCCTGAGCTCCGGCTGAGTGTCCGGCTGAGTGTCCGGCT
219 R G K K E A E D E L S A R H P A E L A M E P T P E L P A E V L M V E
818 GTAGAGAAGCTTGTCCACGAGGACTCCAGGTCAAGGAGGAGTGAAGCCCTGACTGCAGAGATCGTGAAGACCATCCGGGACATCATTCCTTGAACCT
253 V E N V H E D F Q V T E E V K A L T A E I V K T I R D I I A L N P
920 CTCTACAGGGAGTCAGTGTGATGATGAGCTGGCCAGCGGGTGGTGGACAACCCATCTACCTGAGCGACATGGGCGCGGCTCACCGGGCGGAG
287 L Y R E S V L Q M M Q A G Q R V V D N P I Y L S D M G A A L T G A E
1022 TCCCATGAGTGCAGGACGCTCTGGAAGAGCAATATTCCTAAGCGCTGTACAAGCCCTCCTCCTGCTGAAGAAGGAATTTGAACGAGCAAGCTGCAG
321 S H E L Q D V L E E T N I P K R L Y K A L S L L K K E F E L S K L Q
1124 CAGCGCTGGGGCGGAGGTGGAGGAGAAGATCAAGCAGACCCACCGTAAGTACTGCTGCAGGAGCAGCTAAGATCATCAAGAAGGAGCTGGGCTGGAG
355 Q R L G R E V E E K I K Q T H R K Y L L Q E Q L K I I K K E L G L E
1226 AAGGACCAAGGATGCCATCGAGGAAGTTCGGGAGCGCTGAAGGAGCTCGTGGTCCCAAGCAGCTATGGATGTTGGACGAGCTGAGCAG
389 K D D K D A I E E K F R E R L K E L V V P K H V M D V D E E L S K
1328 CTGGGCTGCTGGACAACCACTCTCGGAGTTCAATGTCAACCGCACTACCTAGACTGGCTCAGCTCCATCCCTTGGGGCAAGTACAGCAACGAGAACTG
423 L G L L D N H S S E F N V T R N Y L D W L T S I P W G K Y S N E N L
1430 GACTGGCCGGGCACAGGCACTGCTGGAGGAAGCACCACAGGCACTGGAGGACGTCAGAAACGCATCCTGGAGTTCATTCGGTATGCCAGCTCCCGGG
457 D L A R A Q A V L E E D H Y G M E D V K K R I L E F I A V S Q
1532 TCCACCAGGGCAAGATCCTCTGCTTCTATGGCCCCCTGGCTGGTAAAGACCAGCATTGTCTGCTCCATCGCCCGGCCCTGAAGCAGAGTACTTCGGC
491 S T Q G K I L C F Y G P P G V G K T S I A R S I A R A L N R E Y F R
1634 TCCAGCTCGGGGACGATGACTGACGTGGCTGAGATCAAGGGCCACAGGCGGACCTADGTGGGCGCCATCGCCGGGAAGATCATCCAGTGTTTGAAGAAGACC
525 F S V G G M T D V A E I K G H R R T Y V G A M P G K I I Q C T L K K T
1736 AAGCAGGAGAACCCCTGATCCTCATCGACGAGGTGACAAGATCGGCCGAGGCTACCAGGGGACCCCTGCTCGGCACTGCTGGAGCTGCTGGACGAG
559 K T E W P L I L I D E V D K I G R G Y Q G D P S S A L L E L L D P E
1838 CAGAATGCCAATCCTGACCACTACTGGACGTGCCGTGGACTGTCCAAGTGTCTTATCGACGGCCAACTCACGGACACCATCCCGAGCG
593 Q N A N F L D H Y L D V P V D L S K V L F I C T A N V T D I P E P
1940 CTGGAGACCGTATGGAGATGATCAACGTGTGAGGCTACTGCGCCAGGAGAAGCTGGCCATGGCGAGCGCTACTGTTGCCCGAGGCTCGCGCCCTGTGT
627 L R D R M E M I N V S G Y V A Q E K L A I A E R Y L V P Q A R A L G
2042 GGGTGTGATGAGAGCAAGCGCAAGCTGCTATCGGACGTGCTGACGCTCATCAAGCAGTACTGCCGGAGAGCGGTGTCCGCAACCTCGCAGAGCAAGTG
661 G L D E S K A K L S S D V L T L L I K Q Y C R E S G V R N L Q K Q V
2144 GAGAAGTGTACGGAATCGGCCTACAAGATTGTACGCGCGGAGGCGGAGTCCGTGGAGGTGACGCGGAGAACCTGCAGGACTCTGTGGGAGGCCGTG
695 E K V L R K S A Y K I V S G E A E S V E V T P E N L Q D F V G K P V
2246 TTCACCGTGGAGCGCATGTATGACGTGACACCGCCCGCGGTGGTTCATGGGGTGGCCGGCAATGGGAGGCTCCACGCTGTTTGTGAGACATCCCTG
729 F T V E R M Y D V T P P G V V M G L A W T A M G G S T L F A C T S L
2348 AGACGGCCACAGGACAAGGATGCCAAGGTTGACAAGGATGGCAGCCTGGAGGTGACAGGCCAGCTGGGGAGGTGATGAAGGAGAGCCCGCATAGCCTAC
763 R R P Q D K D A K G D K D G S L E V T G Q L L G E V M K E S A R I A Y
2450 ACCTTCGCCAGAGCCTCCTCATGACGACGCGCCCGCAATGACTACCTGGTACCTCACACATCCACCTGCATGTGCCGAGGGCCACCCCAAGGAC
797 T F A R A F L M Q H A P A N D Y L V T S H I H L H V P E G A T P K D
2552 GGCCCAAGCGCAGGCTGACCATCGTCAAGGCTGCTGCTCCTGGCCATGGGACGCGCTGTCCGGCAGAATCGCCATGATGGCAAGTCTCCCTCACG
831 G P S A G C T I V T A L L S L A M G R P V R Q N L A M T G E V S L T
2654 GGCAAGCTCGCTGTGGTGGCATCAAGGAGAAGCATTGCGGCCAAGCGCGAGGGGTGACGTGCATCATCTGCCAGCCGAGAACAAGAAAGCACTTC
865 G K I L P V G G I K E K T I A A K R A G V T C I I L P A E N K D F
2756 TACGACCTGGCAGCCTTTCATCACCGAGGGCTGGAGGTGCACTTCGTGGAACACTACCGGAGATCTTCGACATCGCCTTCCGGACGAGCAGGCAAGGCG
899 Y D L A A F I T E G L E V H F V E H Y R E I F D I A F P D E Q A E A
2858 CTGGCGTGGAAAGGTGACGGCCACCCGGGACTGCAGGCGGGATGTCAGGCGCTGTGGCCGAGAAGTGGCGCTGTGGGAGCGCGCCGAGCTGG
933 L A V E R
2960 CAGTGGAGCCACGAGCGAGCAGCTCGGTCCAGTGACCCAGATCCAGGGACCTCAGTCGGCTTAATCAGAGTGGCATAGAAGCTATTTAATGATTAAG
3062 TCATTGCGAGTGGAGTATGATCATCACTAAGCTGACAGTGTGTCAGGAATTTGCTTTGTTACTGCTAGTATATTAGAATCCTGATCTCAGATCCAGAA
3164 TAGTAATAAACACAGGGGTCATTTTCTCACTACTCTGTGTGTCAGGTGTGAAATTTCTGTCTCCCAAGGAAATGTGACTTCACCTTTGGTCCCAATG
3266 GACGAAAATTTCTACTGTCTACATAGGAGAAGTTTGAATGCACCTAATAGCTGGTTTTACACCTTGATTTGAGGTGGAAAGAAATGTATCATGAATC
3368 TCAATAAAATTTAAATCTCTAAACCAAAAAA

```

Fig. 1. Nucleotide sequence of the LonHS cDNA and deduced amino acid sequence. Numbering of the cDNA starts with the first nucleotide of clone lhs1.6. ATG codons implicated in initiation of translation are double underlined. The possible N-terminal region of the LonHS amino acid sequence starting with the first ATG codon (position 62–64) is depicted in small letters. The polyadenylation signal sequence AATAAA is bold underlined. The regions corresponding to the previously reported 351 and 308 bp pieces of LonHS mRNA [11] are underlined.

protease, (LonHS) the oligonucleotide ATAGAAG-CAGAGGATCTTGCCCTGG was used. This was derived from a previously sequenced 351-bp piece of human mRNA which was claimed to encode a homolog of *E. coli* Lon protease [11]. A random hexanucleotide and oligo(dT) primed cDNA libraries prepared from human brain were hybridized with the oligonucleotide. From about 1,500,000 independent clones analyzed, eight clones, lhs1.6, lhs5, lhs8, lhs20, lhs29, lhs37, lhs51 and lhs64 gave a positive hybridization signal. Sequences of these clones confirmed that they were from

mRNA for a protein similar to *E. coli* Lon protease. Based on these data the LonHS cDNA sequence was reconstructed. It contains 3393 bp and corresponds to apparently the entire mRNA for the human Lon protease homolog (Fig. 1), the length of which determined by Northern blotting is approximately 3500 bases long (Fig. 3). LonHS cDNA contains three potential polyadenylation signals in the 3'-UTR. At least two of them —ATTA (positions 3055–3060, the first signal) and AATAAA (positions 3371–3376, the third signal) seem to be used in vivo. The poly(A) tail was found twelve

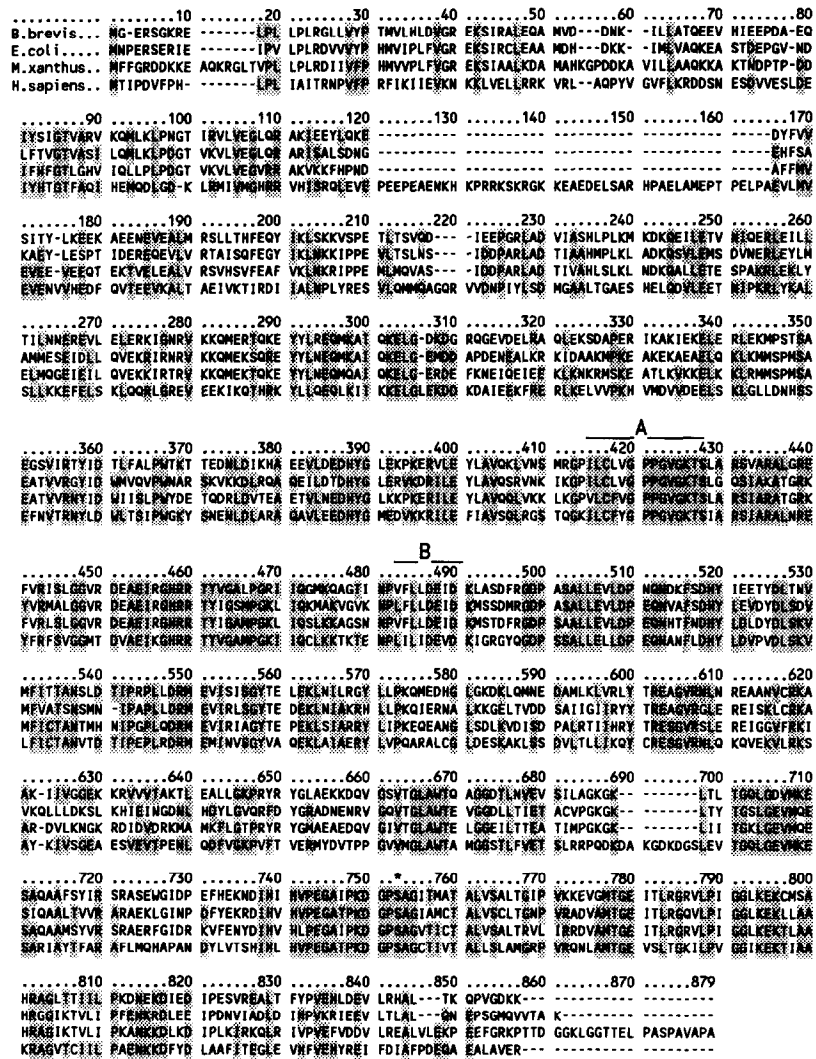


Fig. 2. Multiple amino acid sequence alignment of *E. coli* Lon protease [7] and its homologs from *B. brevis* [19], *M. xanthus* [20] and *H. sapiens* cells. Complete amino acid sequences of bacterial proteins and fragment of LonHS ORF starting with position 93 (see Fig. 1) were aligned. Residues identical in LonHS and in either of the bacterial proteases are highlighted by shading. The plausible catalytic serine residue of the protease domain is denoted by \*. 'A' and 'B' motifs of the NTP-binding pattern are also illustrated. The standard one-letter amino acid code is used.

nucleotides downstream from the first signal in clone *lhs5* (not shown) and seventeen nucleotides downstream from the third one in clone *lhs29* (Fig. 1). Noteworthy, a 308-bp fragment of human cDNA derived from the 3'-region of LonHS mRNA was identified in the EMBL database (Fig. 1). Both 351-bp and 308-bp sequences deviate in several positions from the sequence presented in Fig. 1 (not shown).

### 3.2. Sequence analysis of the cDNA of putative Lon protease

Three regions can be defined in the sequence of LonHS cDNA: a 61-bp 5'-terminal fragment, a 2814-bp ORF and a 518-bp 3'-UTR (Fig. 1). Actual translation of the ORF may be initiated from the first (positions 62–64) or second (positions 338–340) AUG codons in the 5'-region of the mRNA. The second AUG codon of the

ORF encodes a Met residue which corresponds to the initiator Met residues of three known eubacterial Lon proteases (Fig. 2). This correspondence suggests that initiation of translation does not occur at the first AUG codon, but rather occurs at the second AUG codon itself. The initiation of translation from this, internal, AUG codon is expected to proceed in a cap-independent manner characteristic of a narrow class of mRNAs [16]. The 5'-terminal 337-nucleotide sequence preceding the second AUG codon contains a high proportion of G and C nucleotides and hence could facilitate a high-ordering structure essential for the initiation of translation from this codon. Despite this, the alternative option, initiation of translation from the first AUG codon of the ORF cannot be ruled out, but in this case the LonHS is to include an N-terminal domain with no counterpart in bacterial Lon proteases (Fig. 1). The first and second

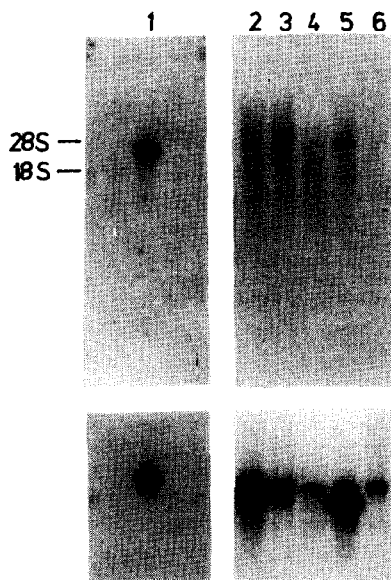


Fig. 3. Northern blot analysis of poly(A)<sup>+</sup> RNA from human brain and various rat organs. The lhc8 cDNA fragment was used as a probe; the same blots were reprobated with the actin cDNA probe (shown below each panel). The samples of poly(A)<sup>+</sup> RNA from the various organs are presented in the following order: lane 1, human brain; lanes 2–6, rat organs, lung, duodenum, liver, heart and thymus. The positions of 28S and 18S rRNAs are shown on the left side.

AUG codons of the ORF as well as a single AUG codon in another reading frame (positions 187–189), which may function as a negative translation control element [16], are in a poor Kozak context predicting ineffective initiation of translation [17].

The Lon-related amino acid sequence initiated at the second in-frame AUG codon includes 845 residues, more than 30% of which are identical with those of eubacterial Lon proteases (Fig. 2). The N-terminal region is most divergent and contains 6 inserts including the longest, a 45 amino acid span. The middle region of LonHS is pronouncedly conserved and includes counterparts of NTP-binding pattern [10] consisting of amino acid sequences ILCFYGPPGVGKT and LILIDEVD for 'A' and 'B' motifs, respectively. The C-terminal region contains a likely catalytic serine residue of the protease domain [18] located in an extremely conserved amino acid stretch IHLHVPEGATPKDGPASAG (Fig. 2). Taken together, results of phylogenetic analysis show that LonHS retains all features earlier claimed essential for Lon protease function. Thus the primary structure of the human protein shown in Fig. 2 appears to be the structure of authentic Lon protease.

### 3.3. Tissue specificity of LonHS gene expression

The tissue distribution of LonHS mRNA was tested in Northern blot experiments using poly(A)<sup>+</sup> RNA from several rat organs. Poly(A)<sup>+</sup> RNA from human brain served as a positive control. A putative Lon protease

transcript was detected in most organs tested (duodenum, heart, lung and liver), thymus being apparently an exception (Fig. 3). This result suggests a ubiquitous role for the Lon protease homolog in various tissues, which is well compatible with the function of Lon protease in degradation of short-lived proteins in eubacterial cells.

**Acknowledgements:** We thank Dr. Robert Casad for assistance in preparation of the manuscript and Ramune Paulauskaite and Ieva Marcinkiene for technical assistance. We are greatly indebted to Dr. Nikolai S. Bystrov for providing the oligonucleotide, to Dr. Eugene V. Koonin for making available to us sequence information, and to Dr. Vadim I. Agol for helpful discussions. This work was supported by the Human Genome Project of the Russian Academy of Sciences and the Russian Foundation for Support of Fundamental Science.

### References

- [1] Hershko, A. and Ciechanover, A. (1992) *Annu. Rev. Biochem.* 61, 761–807.
- [2] Maurizi, M.R., Clark, W.P., Kim, S.H. and Gottesman, S. (1990) *J. Biol. Chem.* 265, 12546–12552.
- [3] Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Mattick, J.C., Dalrymple, B., Kuramitsu, H., Shiroza, T., Foster, T., Clark, W.P., Ross, B., Squires, C. and Maurizi, M.R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 3513–3517.
- [4] Gottesman, S. and Maurizi, M.R. (1992) *Microbiol. Rev.* 56, 592–621.
- [5] Amerik, A.Yu., Chistyakova, L.G., Ostroumova, N.I., Gurevich, A.I. and Antonov, V.K. (1988) *Bioorganicheskaya Khimiya* 14, 408–411.
- [6] Chin, T.D., Goff, S.A., Webster, T., Smith, T. and Goldberg, A.L. (1988) *J. Biol. Chem.* 263, 11718–11728.
- [7] Amerik, A.Yu., Antonov, V.K., Ostroumova, N.I., Rotanova, T.V. and Chistyakova, L.G. (1990) *Bioorganicheskaya Khimiya* 16, 869–880.
- [8] Watabe, S. and Kimura, T. (1985) *J. Biol. Chem.* 260, 14498–14504.
- [9] Ishii, A., Watabe, S. and Hamada, H. (1992) *Placenta* 13, 343–347.
- [10] Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.* 1, 945–951.
- [11] Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992) *Nature*, 355, 632–634.
- [12] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W. (1979) *Biochemistry* 18, 5294–5299.
- [13] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd edition), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [15] Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1989) *J. Mol. Evol.* 28, 256–268.
- [16] Agol, V.I. (1991) *Adv. Virus Res.* 40, 103–180.
- [17] Kozak, M. (1987) *Nucleic Acids Res.* 15, 8125–8148.
- [18] Amerik, A.Yu., Antonov, V.K., Gorbalenya, A.E., Kotova, S.A., Rotanova, T.V. and Shimbarevich, E.V. (1991) *FEBS Lett.* 287, 211–214.
- [19] Ito, K., Udaka, S. and Yamagata, H. (1992) *J. Bacteriol.* 174, 2281–2287.
- [20] Tojo, N., Inouye, S. and Komano, T. (1993) *J. Bacteriol.* 175, 2271–2277.