cDNA cloning of a human 100 kDa de-ubiquitinating enzyme: the 100 kDa human de-ubiquitinase belongs to the ubiquitin C-terminal hydrolase family 2 (UCH2)

Laurent Falquet, Nicole Paquet, Séverine Frutiger, Graham J. Hughes, Khan Hoang-Van, Jean-Claude Jaton*

Department of Medical Biochemistry, Faculty of Medicine, University of Geneva, Medical Center, 1 rue Michel-Servet, CH-1211 Geneva 4, Switzerland

Received 26 September 1995; revised version received 12 October 1995

Abstract The full length cDNA encoding a 100 kDa human de-ubiquitinating enzyme, referred to as de-ubiquitinase was obtained using one clone selected from a randomly sequenced human brain cDNA library and specific primers. The sequence of 18 peptides generated from the de-ubiquitinase isolated from outdated human erythrocytes matched perfectly with the predicted amino acid sequence, which would encode a protein containing 858 amino acids (calculated M_r = 95,743 Da). Homology search disclosed that the protein is a member of a large family of ubiquitin C-terminal hydrolases (UCH2), that was defined on the basis of the presence of two specific patterns, 'the Cys- and His-domains', which are likely to be involved in the de-ubiquitinating activity [7]. An additional conserved region, 'the aspartic acid domain', was also identified, the functional role of which is unknown.

Key words: De-ubiquitinase; Isopeptidase; cDNA

1. Introduction

We have recently reported about a human 100 kDa de-ubiquitinating enzyme which was able to cleave both isopeptidic bond G_{76}-K_{48} in dimeric branched ubiquitin and peptidic bond G_{76}-M_1 in 'head-to-tail' Ub2 and Ub9 polymers [1]. K_m values of the enzyme, which we called 100 kDa de-ubiquitinase, showed however, a 10-fold preference for the cleavage of Ub specific isoamide bonds over 'classical' peptide bonds. On the other hand, the 30 kDa ubiquitin C-terminal hydrolase (a member of UCH family 1) was devoid of such specificity [1], in agreement with earlier reports [2,3]. Because several isopeptidases and ubiquitin C-terminal hydrolases have been described [4,5,6,7], the cDNA cloning of one partially defined 100 kDa de-ubiquitinase offers the opportunity to (i) clarify the structure/function relationship, (ii) possibly unravel a member of new class of ubiquitin specific proteinases, and (iii) to express the protein in amounts sufficient for extensive biochemical characterization and possible biotechnological usage. To this end, partial NH_2-terminal sequence determination of peptides derived from the enzyme isolated from human erythrocytes, was carried out. The data allowed us to select one clone from randomly sequenced cDNA from a human brain library [8]. The full length cDNA clone was obtained by PCR using selected primers and the 5′ anchored PCR amplification technique [9–11].

We report here the predicted amino acid sequence of the 100 kDa human de-ubiquitinase, referred to as UBP1_HUMAN (sp|P45974), which suggests that it is a member of a new, growing family of ubiquitin C-terminal hydrolases, referred to as UCH2 [12].

2. Materials and methods

2.1. Material

Outdated human blood samples from the Central Hematology Laboratory of the Geneva Canton Hospital was the source of the 100 kDa de-ubiquitinating enzyme, which was purified to homogeneity according to Falquet et al. [1].

2.2. Lys-C and trypsin digestions, CNBr cleavage followed by microsequencing of selected peptides from the 100 kDa de-ubiquitinase

Enzymic and chemical cleavages, peptide separation by HPLC or SDS-Tris-tricine gels [13], transfer to PVDF membranes and microsequencing with a ABI 473A sequencer were carried out according to standard methods in our laboratory [14]. Sequences were used to perform a search with the BLAST program [15,16], using the 'tblastn' option which reverse-translate the query peptide and compare the degenerated DNA sequence with all nucleotide databases.

2.3. cDNA cloning

A human brain cDNA clone, HIBAA66, was obtained from the American Type Culture Collection (ATCC) library (see section 3). The clone being incomplete, two peptide sequences, not found in the clone HIBAA66 were used to design degenerated primers (ID1F, ID2F), which were combined with 3 perfectly matching primers based on the cDNA sequence of HIBAA66 (I1R, I2R, I3R) (see Fig. 1). All of these primers were used for RT-PCR on human brain mRNA (Clontech) with Superscript II reverse transcriptase (BRL) and Dynazyme II DNA polymerase (Fynnzyme Oy). The second nested reaction, using primers ID2F and I3R, allowed us to amplify a band of 1300 bp, which was purified, blunted with T4 DNA polymerase (Biofinex) and ligated into *Sma*I site of vector pBluescript II KS. Out of 48 colonies, 7 independent clones were obtained and were labeled RST1, -4, -8, -10, -11, -12, -13. As the primer ID2F did not correspond to the blocked N-terminal region of the enzyme and that some peptide sequences were still not assigned to the determined cDNA sequence, the 5′ anchored-PCR method using the 5′ RACE-PCR kit from BRL was carried out to clone the ultimate 5′ end. The kit was used with 2 perfectly matching primers designed from the RST4 sequence: I6R and I4R (Fig. 1). A band at around 300 bp was obtained and cloned into *Sal*I site of pBluescript II KS, as *Sal*I site was present on both UAP (5′ RACE kit) and I4R primers. One clone only was obtained (BBS4). This clone was sequenced on both strands and was found to encompass a putative AUG initiation codon.

The 3 clones (BBS4-RST4-HIBAA66), were fused in two steps; HIBAA66 and RST4 clones were fused first using the unique *Hin*cII

*Corresponding author. Fax: (41) (22) 702 55 02.

Abbreviations: Ub, ubiquitin; UCH, ubiquitin C-terminal hydrolase; PVDF, polyvinylidene difluoride; sp, Swissprot; gb, Genbank; gp, Genpept; pir, protein identification resource; RACE-PCR, rapid amplification of cDNA ends-PCR.

SSDI 0014-5793(95)01287-7

site (pos 1447), while the unique *Nar*I site (pos 230) was used to add the last BBS4 base pairs. DNA sequencing was performed in pBluescript vector with primers M13R and M13–48 using a protocole based on exonuclease III digestion [17]. The T7-sequencing kit from Pharmacia was used with [α^{35}S]dATP from Amersham.

Primer' sequences used (see Fig. 1).
ID2F, 5'-CA(A/G)TA(C/T)GTIGA(A/G)(A/C)G(A/G/C/T)CA(C/T)TT(C/T)AA-3'
I1R, 5'-TCGTGTGGTCTTGACAGC-3'
I2R, 5'-[AATGGATCC]GTGCTCCAGAAGTCATCG-3'
I3R, 5'-GATGTAGTCAACTCGCTGGG-3'
I4R, 5'-[GGGTCGAC]CTCGCTAAGGTCAAATCCG-3'
I6R, 5'-[GGGATCC]TCCTGCTTGCGGGAGG-3'

2.4. PROSITE consensus patterns

Patterns for Cys- and His-domain sequences of UCH2 family members, are: [LIVMFY]-x(3)-[AGC]-[NA]-x-C-[FY]-[LIVMC]-[NS]-[SC]-x-[LIVM]-Q and Y-x-L-x-[SAG]-[LIVMT]-x(2)-H-x-G-x(4,5)-G-H-Y, respectively [12].

3. Results

Digestions of human erythrocyte de-ubiquitinase with Lys-C enzyme and/or CNBr yielded well resolved peptides by SDS-PAGE and Tris-tricine gels. After transfer to PVDF membranes, bands were sliced out and subjected to automated Edman degradation; 18 partial NH_2-terminal sequences were clearly identified. The BLAST search unraveled the presence of an expressed sequence tag (EST) derived from a human brain cDNA library [8]; sequences from 9 peptides were found to match regions of the translated protein sequence predicted from a 1.7 kb cDNA clone, HIBAA66 (Fig. 1). Degenerated primers, synthesized on the basis of peptide sequences not found in HIBAA66 clone, were used to amplify by PCR a human brain poly A^+ mRNA previously subjected to reverse transcription. Several clones of 1.3 kb were isolated and designated RST. They were all sequenced on both ends (2 × 300 bp) with M 13R and M 13–48 primers and did not exhibit sequence difference, except for the region of the degenerated primer ID2F. Clone RST4 was sequenced on both strands and fused to HIBAA66 clone, thus yielding a 3 kb cDNA clone encoding 803 amino acids. The 5' end of the cDNA was obtained through the 5' anchored PCR technique [8–10]; the full length cDNA sequence comprises 3158 bp and a putative AUG codon, which suggest that the cDNA clone encodes a protein of 858 amino acids (calculated M_r = 95,743 Da) with a pI of 4.90 (Fig. 1). The predicted human de-ubiquitinase sequence exhibits several characteristic structural features: (1) it contains one 'cysteine domain' and one 'histidine domain', which have been postulated as consensus sequences involved in de-ubiquitinating activity (Fig. 2) [7,18]. The Cys-domain is N-terminal from the His-domain. A third, but short region of homology located in between Cys- and His-domains contains a conserved QQDXXE 'Asp-domain' sequence (Fig. 2), the functional role of which is unknown. (2) The overall predicted protein sequence of the 100 kDa de-ubiquitinase is poorly homologous to the yeast DOA4 (sp|P32571) or the human *tre*-2 oncogene products (sp|P35125), the latter resembling the 100 kDa human enzyme in their enzymatic activity and sequence similarity in the Cys- and His-domains only. The closest yeast homologue of the 100 kDa human de-ubiquitinase is UBPC protein (sp|P38237). (3) Homology search in the latest data bank updates disclosed that the human 100 kDa de-ubiquitinase is a member of a new, yet, large family of ubiquitin C-terminal hydrolases, designated UCH2, which was defined by 2 specific patterns, the Cys- and His-domains [7,12,18]. This family contains at least 25 proteins, some of which are still hypothetical because they are derived from genomic sequences without identification of the gene product. Surprisingly, proteins HUM_TRNAGT and RAB_TRNAGT were referred to as tRNA transglycosylases (gp|U30888 and sp|P40826); however, based on their sequences, they definitely belong to UCH2 family. (4) Because the primary structure also suggests consensus sequences for metal binding as drawn in Fig. 1, susceptibility

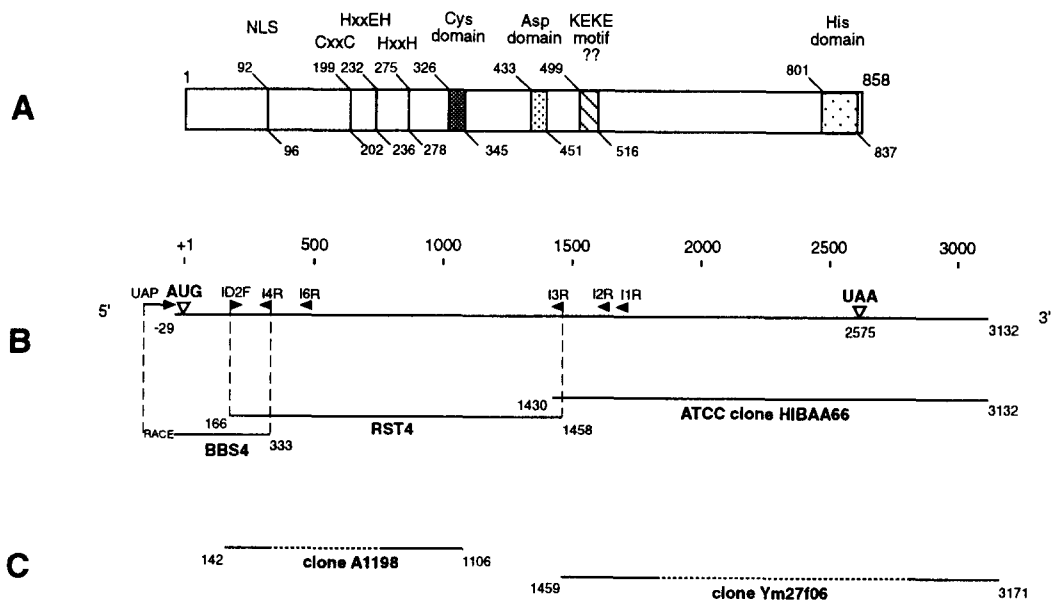

Fig. 1. Full length cDNA sequence encoding the 100 kDa human de-ubiquitinase. (A) Schematic representation of the predicted amino acid sequence with identified homology domains; (B) cloning strategy (see section 2); (C) clones identified from expressed sequence tag (EST) libraries [23], (gb|T11766, gb|T11767, gb|H15561, and gb|H15620); (D) cDNA (EMBL X91349) and predicted amino acid sequences (sp|P45974). Underlined residues in D refer to identified peptide sequences; boxed sequence refers to polyA signal.

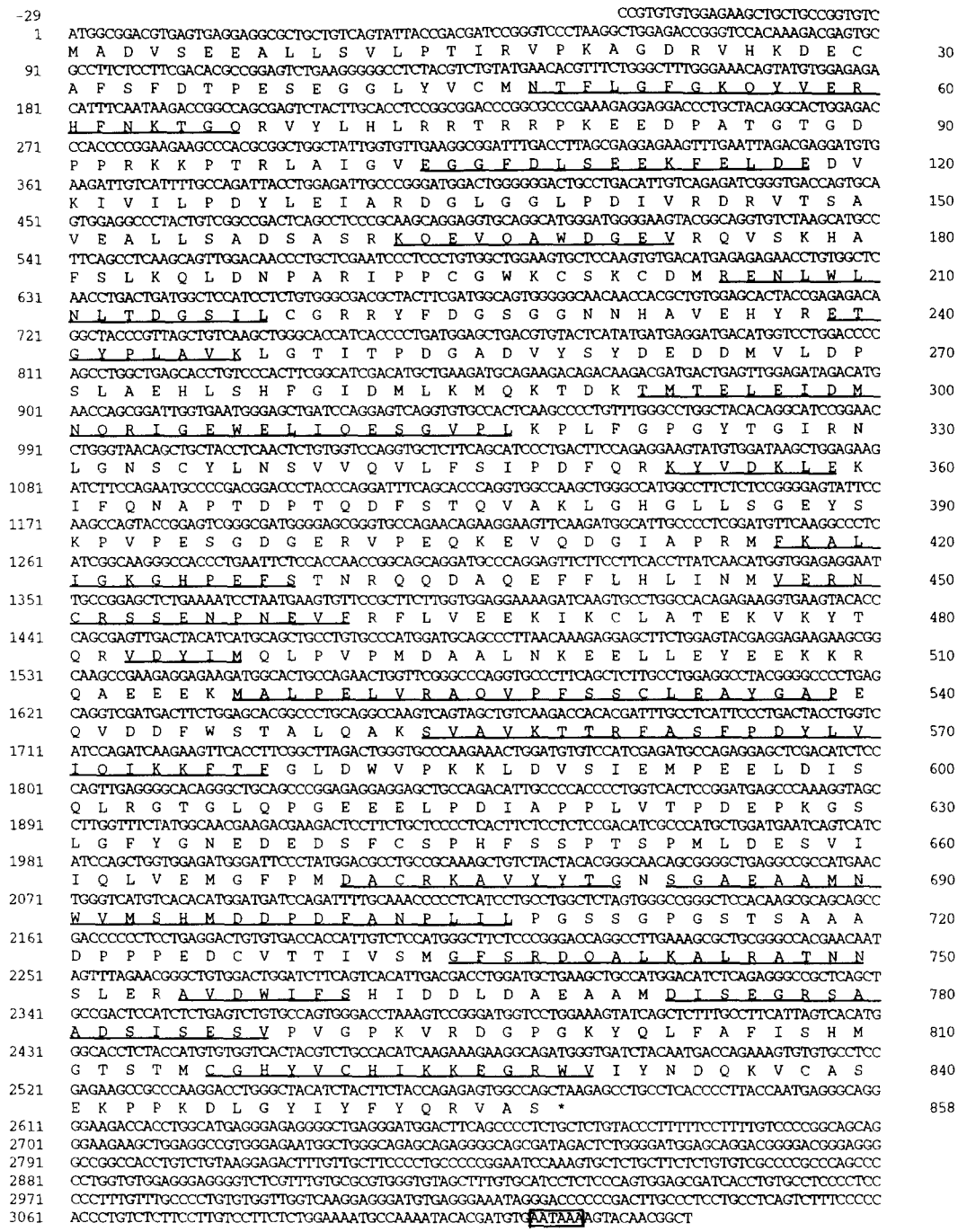

Fig. 1. (*Continued*).

of the 100 kDa de-ubiquitinase towards ion chelators was determined. De-ubiquitinating activity was not affected by EDTA, EGTA or 1,7-phenanthroline, whereas the chelating isomer, 1,10-phenanthroline, markedly inhibited cleavage into monomeric ubiquitin (to be published elsewhere).

4. Discussion

The predicted primary structure of the 100 kDa de-ubiquitinase isolated from human erythrocytes has been determined from cDNA cloning using human brain mRNA and cDNA library. Because the entire cDNA sequence is based on 2 mRNA's each obtained from a different individual, the existence of polymorphism cannot be ruled out. The finding of fit of 18 peptide sequences with the predicted sequence (Fig. 1) supports the notion that the 3 cDNA clones (BBS4, RST, HIBAA66) encode the complete polypeptide chain of the de-ubiquitinase, but does not represent the final evidence.

The flanking nucleotide sequence around the initiation codon, GUCAUGG, exhibits a G at −3 and +4 (Kozak consensus sequence), which are required otherwise AUG is bypassed [19]. Because of the absence of an in-frame stop codon, up-

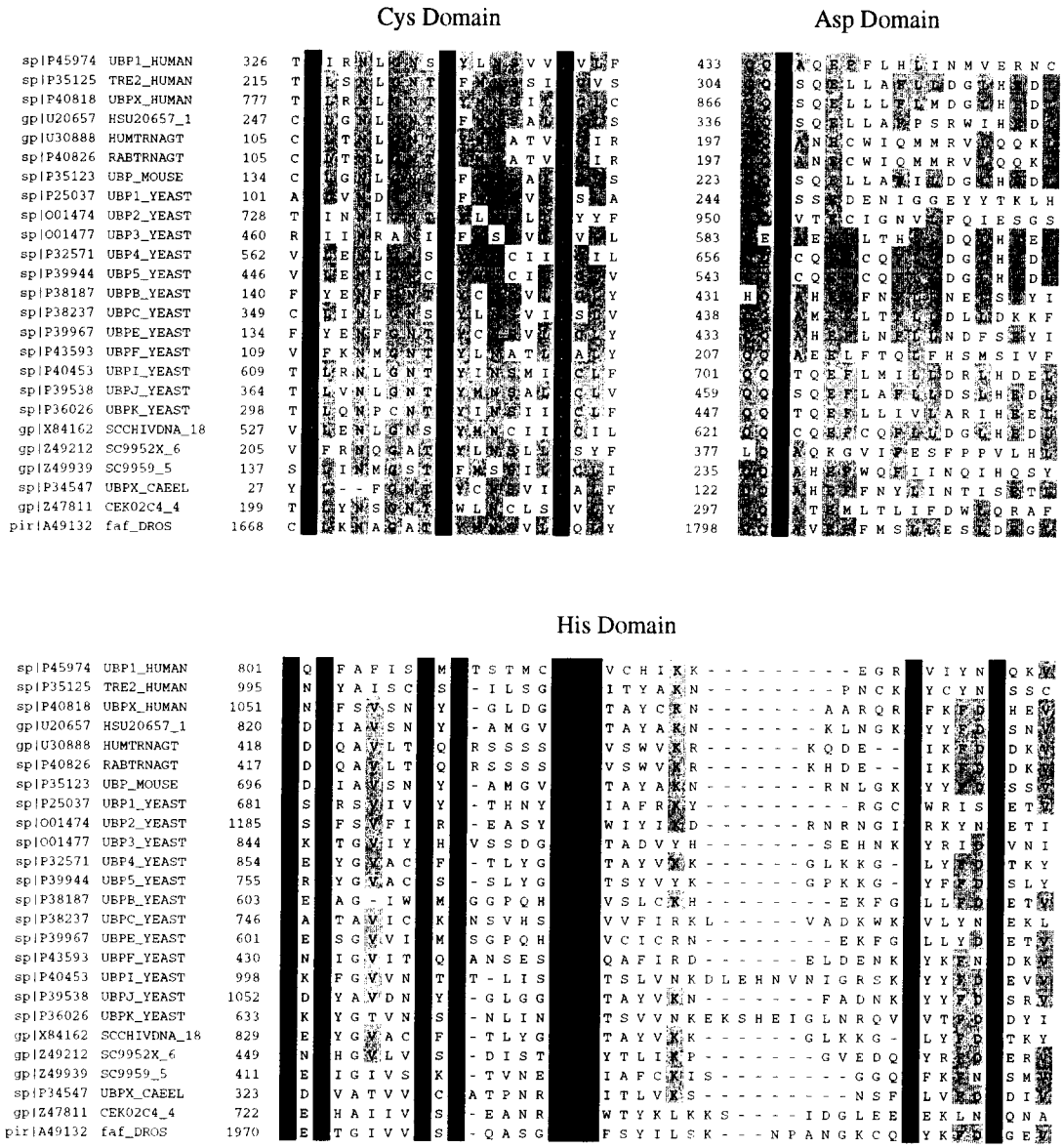

Fig. 2. Alignments of Cys-, Asp- and His-domains within the UCH2 family members. Multiple sequence alignments were carried out using CLUSTAL W program [24]. From left to right, data bank accession number, protein's name, position of the NH_2-terminal residue of the domains; black boxes, 100% identity; gray boxes, >50% identity.

stream of this AUG, the NH_2-terminal Met residue remains putative. Since the most proximal peptide sequence found in the mature protein is located 47 residues C-terminal from the predicted N-terminus of the protein, and because the mature protein exhibited an apparently blocked NH_2-terminal residue, any processing mechanisms which would take place between Met-1 and Asn-48, leading to a new blocked NH_2-terminal residue, cannot be formally excluded. Identification of the NH_2-terminal blocked peptide is underway in the laboratory.

The enzyme displays several distinct consensus patterns which may be the signature of a particular function (Fig. 1). The so-called Cys- and His-domains (Figs. 1 and 2) may be involved in the catalytic activity of the enzyme as mutations of the conserved Cys in the Cys-domain of DOA4 [7] to Ser or Ala resulted in catalytic inactivity; in addition, the presence of both Cys- and His-domains appears a necessary requirement for activity [7]. Treatment of the enzyme with 1,10-phenanthroline almost completely abolished enzymatic activity. Thus, the presence of metal ion(s) can be suspected; in fact, compatible consensus sequences involving Cys and His residues have been pinpointed (see Fig. 1). A nuclear localization signal (NLS) has also been predicted in the NH_2-terminal part of the 100 kDa human de-ubiquitinase (Fig. 1) [20,21].

Ultimately, a stretch of 18 amino acids containing more than 60% Lys and Glu residues, known as the KEKE motif (see Fig. 1A) was recognized. We designate it as pseudo KEKE motif, however, because it contains an internal Tyr residue not recorded in other KEKE positive proteins [22]. The presence of this motif was postulated to mediate non-covalent association between proteins; if this assumption turned out to be true, the putative, transient, binding of the 100 kDa de-ubiquitinase to the 26S proteasome [25] is of major interest.

Acknowledgements: This work was supported, in part, by Grants 31-33658.92 and 3100-043373.95/1 from the Swiss National Science Foundation.

References

[1] Falquet, L., Paquet, N., Frutiger, S., Hughes, G.J., Hoang-Van, K. and Jaton, J.-C. (1995) FEBS Lett. 359, 73–77.
[2] Pickart, C.M. and Rose, I.A. (1986) J. Biol. Chem. 261, 10210–10217.
[3] Wilkinson, K.D., Deshpande, S. and Larsen, C.N. (1992) Biochem. Soc. Trans. 20, 631–637.
[4] Hadari, T., Warms, J.V.B., Rose, I.A. and Hershko, A. (1992) J. Biol. Chem. 267, 719–727.
[5] Zhang, N., Wilkinson, K.D. and Bownes, M. (1993) Dev. Biol. 157, 214–223.
[6] Eytan, E., Armon, T., Heller, H., Beck, S. and Hershko, A. (1993) J. Biol. Chem. 268, 4668–4674.
[7] Papa, F.R. and Hochstrasser, M. (1993) Nature 366, 313–319.
[8] Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993) Nature Genet. 4, 373–380.
[9] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998–9002.
[10] Ohara, O., Dorit, R.L. and Gilbert, W. (1989) Proc. Natl. Acad. Sci. USA 86, 5673–5677.
[11] Loh, Y., Elliot, J.F., Cwirla, S., Lanier, L.L. and Davies, M.M. (1989) Science 243, 217–220.
[12] Bairoch, A. and Bucher, P. (1994) Nucleic Ac. Res. 22, 3583–3589.
[13] Schägger, H. and von Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[14] Hughes, G.J., Frutiger, S., Paquet, N. and Jaton, J.-C. (1990) Biochem. J. 271, 641–647.
[15] Gish, W. and States, D.J. (1993) Nature Genet. 3, 266–272.
[16] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[17] Henikoff, S. (1984) Gene 28, 351–359.
[18] Baker, R.T., Tobias, J.W. and Varshavsky, A. (1992) J. Biol. Chem. 267, 3364–3375.
[19] Kozak, M. (1991) J. Biol. Chem. 260, 19867–19870.
[20] Dingwall, C. and Laskey, R.A. (1991) TIBS 16, 478–481.
[21] Roberts, B. (1989) Biochim. Biophys. Acta 1008, 263–280.
[22] Realini, C.L., Rogers, S.W. and Rechsteiner, M. (1994) FEBS Lett. 348, 109–113.
[23] Liew, C.C. (1994) Proc. Natl. Acad. Sci. USA 91, 10645–10649.
[24] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Ac. Res. 22, 4673–4680.
[25] Peters, J.-M., Franke, W.W. and Kleinschmidt, J.A. (1994) J. Biol. Chem. 269, 7709–7718.

Note added to the proof

Since this manuscript was submitted, the cDNA sequence of a clone encoding isopeptidase T from human liver cell line HL60 appeared (accession no. gb/U35166). The sequence was 98% identical to ours.