

cDNA cloning of a new putative ATPase subunit p45 of the human 26S proteasome, a homolog of yeast transcriptional factor Sug1p

Kin-ya Akiyama^a, Kin-ya Yokota^a, Susumu Kagawa^a, Naoki Shimbara^b, George N. DeMartino^c, Clive A. Slaughter^d, Chiseko Noda^e, Keiji Tanaka^{e,*}

^aDepartment of Urology, School of Medicine, Tokushima 770, Japan

^bBiomedical R&D Department, Sumitomo Electric Industries, 1 Taya-cho, Sakae-ku, Yokohama 244, Japan

^cDepartment of Physiology and Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, TX 75235, USA

^dDepartment of Biochemistry and Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, TX 75235, USA

^eInstitute for Enzyme Research, University of Tokushima, Tokushima 770, Japan

Received 2 February 1995; revised version received 10 March 1995

Abstract The nucleotide sequence of a cDNA that encodes a new regulatory subunit, named p45, of the 26S proteasome of human hepatoblastoma HepG2 cells has been determined. The polypeptide predicted from the open reading frame consists of 406 amino acid residues with a calculated molecular weight of 45770 and isoelectric point of 8.35. The sequences of several fragments of bovine p45, determined by protein chemical analyses, spanning 27% of the complete structure, were found to be in excellent accord with those deduced from the human cDNA sequence. Computer analysis showed that p45 belongs to a family of putative ATPases which includes regulatory components of 26S proteasomes. The overall structure of p45 was found to be homologous to that of yeast Sug1p, which has been identified as a transcriptional factor. It is closely similar, but not identical to the sequence reported for Trip1, a functional homolog of Sug1p in human tissues. These results are consistent with the possibility that Sug1-like proteins with distinct sequence function in transcription and protein degradation in human cells. However, the alternative hypothesis, that the same gene locus encodes both p45 and Trip1, cannot be excluded on the basis of such closely similar sequences. In either case, both proteins are likely to function equivalently well in either transcription or protein degradation.

Key words: cDNA cloning; 26S Proteasome; Subunit p45; Putative ATPase family; SUG1

1. Introduction

The 26S proteasome is a eukaryotic ATP-dependent protease, consisting of the 20S proteasome (multifunctional proteinase) and a set of regulatory proteins [1]. It is responsible for selective degradation of abnormal proteins and naturally short-lived proteins related to cell cycle control and metabolic regulation [2]. cDNA cloning of a subunit of human 26S proteasomes named S4 [3] showed that it is a novel member of the putative ATPase family, that includes MSS1 (a suppressor of a temperature sensitive (ts) defect in a kinase that regulates G1 cyclins in *S. cerevisiae*) [4], TBP-1 (a protein that binds HIV TAT in an in vitro assay) [5] and TBP-7 (a protein identified by its close sequence similarity to TBP-1) [6]. In fact, another subunit of

human 26S proteasomes, S7, was found to be identical to MSS1 [7]. In addition, 5 genes belonging to the same family and assuming to encode the 26S proteasome components have been cloned in yeast, including the cell cycle genes MTS2 from *S. pombe* [8] (and its homolog from *S. cerevisiae* YTA5) and CIM5 from *S. cerevisiae* [9], as well as a transcriptional factor SUG1 [10] and YTA1 [11] and YTA2 [11] from *S. cerevisiae*. Computer analyses showed that S4, MSS1, TBP-1 and TBP-7 are homologs of MTS2/YTA5, CIM5, YTA1 and YTA2, respectively. Recently, a human homolog of yeast SUG1 called Trip1 has been cloned, and shown to be functionally equivalent to Sug1p [12]. Antibodies against Sug1p cross-react with a high molecular weight transcriptional mediator complex from *S. cerevisiae*, suggesting that Sug1p is a component of this mediator complex [13]. A distinct complex, called PA700, that activates the latent 20S proteasome has been described as a component of the 26S proteasome [14], and DeMartino et al. [15] have reported that sequences of fragments of one component of PA700, named p45, have high similarity to yeast Sug1p. In the present study, we report cDNA cloning of the human p45 subunit, and show that it is almost identical with the sequence reported for Trip1.

2. Materials and methods

Subunit p45 was isolated from the purified PA700 complex of bovine erythrocytes which is an activator complex of the latent 20S proteasome and regulator components of 26S proteasomes [14]. The fragments of p45 were obtained by digestion with trypsin and sequenced by automated Edman degradation as reported previously [15]. A cDNA library of human hepatoblastoma cells HepG2 was constructed in a ZAPII phage expression vector (Stratagene). For isolation of cDNA for p45, about 5×10^5 plaques were screened by hybridization with a cDNA fragment that had been synthesized by the polymerase chain reaction (PCR; for details, see text) and labeled with [α -³²P]dCTP as a probe. Plaque hybridization was carried out and pBluescript plasmid was excised and directly sequenced by a double-strand strategy using a A.L.F. automatic DNA sequencer (Pharmacia LKB Biotechnology Inc.).

3. Results and discussion

3.1. Isolation of a cDNA clone encoding p45

Previously, we reported the separation of multiple components of the 26S proteasomal activator PA700 from bovine red blood cells by reverse-phase high-performance liquid chromatography and the primary structure of p45 internal region has been determined by automated Edman degradation [15]. For

*Corresponding author. Fax: (81) (886) 33-4223.

The nucleotide sequence data reported in this paper will appear in the GSDB, DDBJ, EMBL and NCBI Nucleotide Sequence Databases with the following accession number: D44467.

use as PCR primers, we selected parts of the sequences of two of the proteolytic fragments with ⁶¹EELQLLEQEG⁷⁰ and ²⁴³PSIIFMDEID²⁵², respectively (see Fig. 1). The following oligonucleotides corresponding to the protein sequences were synthesized:

Forward primer: 5'-GAGGAG(T/C)TTCAG(T/C)TT(T/C)T-(T/G)CAGGAGCAGGG-3'

Reverse primer: 5'-TCGATTTCGTCCATGAAGATGATT-(C/G)(A/T)TGG-3'

Using these primers, a fragment of approximately 600 bp was synthesized by PCR against first strand DNA complementary to mRNA from bovine liver as a template.

To isolate cDNA encoding p45, we screened a cDNA library with the ZAPII vector using poly(A)⁺ RNAs extracted from human hepatoblastoma HepG2 cells by hybridization with a cDNA fragment for bovine p45 with approximately 600 bp synthesized by PCR as a probe. We then screened about 5×10⁵ plaques of the human hepatoblastoma HepG2 cells cDNA library with the cDNA fragment. The 4 cDNA clones that gave a strongly positive signal with the probe were isolated from the library by plaque hybridization techniques. The clone carrying the largest cDNA insert of about 1.4 kb length was subjected to cDNA sequencing.

3.2. Primary structure of p45

The nucleotide sequence of the p45 cDNA clone and the primary structure of the p45 protein deduced from the cDNA sequence are shown in Fig. 1. The sequence of 1287 nucleotides included the entire coding region and 5'- and 3'-non-coding regions. The 3'-non-coding region consisted of 58 nucleotides. A putative polyadenylation signal (AATAAA), which is common to eukaryotic mRNAs, was located 18 nucleotides upstream from the poly(A) addition site. We concluded that ATG, located at nucleotides 1–3, is the initiation codon, because it is surrounded by a sequence that is similar to the consensus sequence for translation initiation [16], and because it is the largest open reading frame. Subunit p45 corresponds to a protein of 406 amino acids with a calculated molecular weight of 45770. The amino acid sequence shown in Fig. 1 was confirmed to be that of p45 of bovine 26S proteasomal PA700 by showing that the partial sequence of 7 fragments determined chemically (Fig. 1, continuous lines) was excellently in accord with that deduced from the nucleotide sequence of the cDNA. Only one difference was observed, at amino acid position 61. A further substitution associated with the earlier identification of peptide 139 as Glu [15] was found to represent an error in protein sequencing. The isoelectric point (pI) of p45 was calculated to be 8.35 by the method of Skoog and Wichman [17].

3.3. Similarity of the p45 subunit to yeast Sug1p and its possible functions

In a previous study partial amino acid sequencing of bovine p45 by Edman degradation showed a remarkable similarity of this proteasomal component to the sequence of Sug1p, the product of the yeast *SUG1* gene [15]. *SUG1* was originally identified in a genetic search for transcription factors that interact with the GAL4 regulatory protein of yeast [10], and was identified through a mutation in *Sug1* that suppresses defects in the C-terminal activation domain of the *gal4* protein. A protein or proteins cross-reacting with antibodies against

Sug1p have recently been shown to be present in a high molecular weight complex from yeast that mediates the transcriptional response to the activator proteins GAL4-VP16 and GAL4 in a reconstituted transcription system [13]. The complex consists of some 20 polypeptides that are associated with the core 12-subunit RNA polymerase II [13]. This observation suggests a physical interaction between Sug1p and the transcriptional mediator complex. Fig. 2 shows an alignment between the complete sequence of human p45 reported here and Sug1p sequence. The two proteins display an overall level of identity of 73.6%, confirming that Sug1p and p45 are close homologs.

The close relationship between p45 and Sug1p raises the question of whether eukaryotes have two closely related genes encoding proteins with different sequence, one of which becomes incorporated into the 26S proteasome and the other into the transcription mediator complex, or whether there is only one gene encoding a protein that functions in both complexes. This question is of considerable importance because of its impact upon our understanding of the mode of action of Sug1p in transcription. Under the two-gene model, Sug1p presumably exercises its transcriptional effects via its role in the mediator complex. Under the one-gene model, some or all of these effects may be caused by the effect of p45 on the degradation of a transcription factor(s) catalyzed by the 26S protease as suggested by Dubiel et al. [3]. Consistent with the one-gene model, Ghislain et al. [9] have isolated in yeast a mutant, *cim3-1*, that is allelic to *SUG1*, in which cell division is arrested at the G2/metaphase transition and the degradation of a ubiquitin-protein conjugate normally degraded rapidly by the 26S protease is retarded, indicating that Sug1p has a role in protein degradation. The cell cycle block may be due to a deficiency in the degradation of cyclins CLB2 and CLB3 which occurs in this mutant. Nevertheless, Swaffield et al. [18], using a yeast strain in which Sug1p molecules are tagged with the bacteriophage T7-S10 epitope, failed to detect the epitope-tagged Sug1p in the 26S protease by Western blotting following gel filtration chromatography of cell extracts or by immunoprecipitation of the 26S proteasome, but were able to identify it associated with a high molecular weight species, presumably a transcriptional mediator complex [18]. Although this result is consistent with the two-gene model, it is possible that the product of a single gene is located both in the 26S protease and in the transcriptional mediator complex, but a much larger amount is associated with the latter.

Recently, a human homolog of *SUG1* has been identified by Moore and colleagues [12]. It was isolated in a two-hybrid selection in yeast using a human thyroid hormone receptor as the target protein, and has been designated Tripl (thyroid hormone receptor interacting protein). Tripl has been shown to be functionally equivalent to the *SUG1* protein in its transcriptional effects in yeast and its ability to bind to the activation domains of the GAL4 protein and VP16 [12]. As shown in Fig. 2, the translated sequences of human p45 and human Tripl are very closely related, and differ at only four positions; 61, 266, 272, and 300. At two of these four, the results of Edman degradation are also available for bovine p45. The cDNA assignment of Ile-300 in p45 is confirmed by the sequence analysis of the protein, thus supporting the existence of a difference between p45 and the reported sequence of Tripl which encodes Met at this position. However, at position 61, where the p45

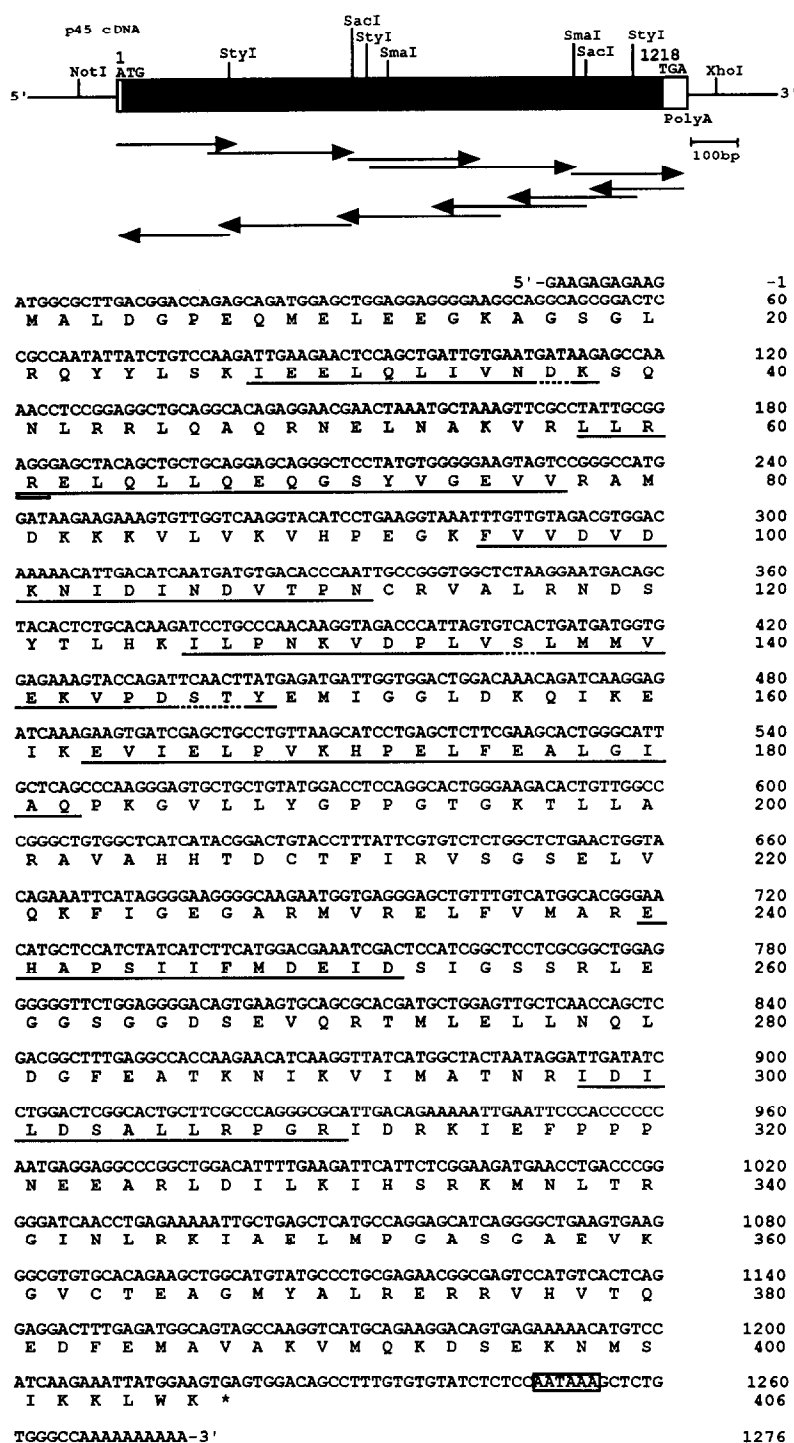


Fig. 1. Structure of cDNA for p45 of human 26S proteasomes. Upper panel: restriction endonuclease map of cloned cDNA for p45 and sequencing strategy. The solid and open boxes show the coding region and 5'- and 3'-non-coding regions, respectively. Continuous lines indicate the sequence of the vector, pBluescript SK⁻. The numbers below boxes indicate the nucleotide numbers of the first nucleotide of the initiation codon, ATG, and the nucleotide before the termination codon, TGA. Sequenced regions are shown by horizontal arrows. The bar represents 100 bp. Lower panel: nucleotide sequence of the cDNA encoding component p45 and the amino acid sequence deduced from its open reading frame. Nucleotides are numbered in the 5'-to-3' direction, beginning with the first residue of the putative initiation methionine codon 'ATG'. The nucleotides on the 5' side are indicated by negative numbers. The predicted amino acid sequence of p45 is shown below the nucleotide sequence. Amino acid residues are numbered from the N-terminus. Continuous underlines show the amino acid sequences corresponding to those obtained by Edman degradation of fragments of bovine p45 cleaved with trypsin [15]. Note that the single amino acid shown by a dotted line was not determined by chemical analysis of bovine p45. The amino acid underlined doubly was not identical with that found by the protein-chemical analysis. The termination codon TGA is marked with an asterisk. The possible polyadenylation signal (AATAAA) is boxed.

p45	1:	MALDGP	QMELEEGKAGSGLRQYYLSKIEELQLIVNDKSNLRLRQAQRNELNAKVRLLR		
Trip1	1:				
Sug1p	1:	MTAAVTSSNIVLETHE	IKP-FEQ-Q-TE-KIRS-TE-G-E-A-D-FIK		
p45	61:	RELQLLQE	QGSYVGEVVRAMDKKKVLVKVHPEGKFVVDVDKNIDINDVTPNCRVALRNS		
Trip1	61:	E			
Sug1p	60:	D-R	P- IKIVSD-Q-YI-A-D-NVK-LKASQ-C-S-		
p45	121:	YTLHKILPNKVDPLVSLMMVEKVPDSTYEMIGGLDKQIKEIKEVIELPVKHP	ELFEALGI		
Trip1	121:				
Sug1p	120:	-M-V-E-A	-D-V-T-S-		
p45	181:	AQPKGVLLYGPPGTGKTL	LARAVAHHTDCTFIRVSGSELVQKF	IGEGARMVREL	FVMARE
Trip1	181:				
Sug1p	180:	I	K	A	Y-S
p45	241:	HAPSIIFMDEIDSIGSSRLEGGSGGDSEVQRTMLELLNQLDGFEATKNIKVIMATNRIDI			
Trip1	241:	S-Q		M	
Sug1p	240:	T-V-SG		TS	I-L-
p45	301:	LDSALLRPGRIDRKIEFPPPNEEARLDILKIHSRKMNLTRGINLRKIAELMPGASGA	EVK		
Trip1	301:				
Sug1p	300:	P	SVA-AE-R	V-K-N-C	D-
p45	361:	GVCTEAGMYALRERRVHVTQEDFEMAVAKVMQKDSEKNMSIKKLWK			
Trip1	361:				
Sug1p	360:	I	L-G	N-NQ-TAI-VA	F-

Fig. 2. Comparison of the protein sequences of p45 of human proteasomes, human Trip1p and yeast Sug1p. Continuous lines indicate identical amino acid residues compared with those in the p45. Numbers are residue numbers. Sequence data of human Trip1p and yeast Sug1p are taken from Lee et al. [12] and Swaffield et al. [10], respectively. For details, see text.

cDNA encodes Arg, the amino acid observed by protein sequence analysis was Glu. This is the same as the amino acid encoded in *Trip1*, suggesting a possible cloning error in the p45 cDNA. In the remaining two positions only cDNA sequences are available, but these suggest that Asp-266 and Thr-272 in p45 are replaced by Ser and Gln in Trip1. The nucleotide sequence of the Trip1 cDNA (D. Moore, personal communication), when compared to that of p45 cDNA, also shows a high level of similarity between the two genes. The coding region sequences differ at only nine nucleotide positions, of which seven are involved in coding for the amino acid sequence differences just described, and two are silent substitutions. The 5'-untranslated regions are identical over the 11-residue region of overlap available, and the 3'-untranslated regions differ at only three positions out of 59. Unfortunately, the interpretation of these data is still problematic, because although such differences may be ascribed to the presence of two genes with different functions, one encoding a polypeptide component of the 26S proteasome and the other a component of the transcription mediator complex, some or all of the differences may also represent allelic substitutions at a single gene locus and/or sequencing/cloning errors arising in the analysis of a single gene encoding a protein that forms part of both complexes. Even if two genes are present, the small number of the amino acid differences between the proteins that they encode suggests the possibility that both may be able to function equivalently in either complex. This possibility is amenable to direct test.

3.4. Some properties of the putative proteasomal ATPase family

The protein sequence of p45 has considerable similarity to those of the family of the proteasomal putative ATPases [1]. As shown in Fig. 3, the middle parts of these 5 ATPases, including the sequences conserved as putative ATP binding amino acid

residues (see asterisks) have high similarities, but their NH₂- and COOH-terminal regions differ markedly. The divergence of their terminal regions may have been associated with the acquirement of specific functions of the respective ATPases. Some properties of this ATPase family are listed in Table 1. All of them are conserved structurally and functionally during evolution, because human S4 and MSS1 could complement growth-arrest due to temperature-sensitive mutation of *S. pombe* MTS2 and *S. cerevisiae* CIM5, respectively (see below) [8,9]. These 5 ATPases are similar in their sizes in both human and yeast. It is of particular interest, however, that the pI of p45 as well as its homolog, yeast Sug1p, is considerably basic, differing from the other 4 ATPases which have weakly acidic pIs. Swaffield et al., [18] have noted that the subgroup of ATPases believed to interact with the transcriptional mediator complex, Sug1p and Trip1, as well as TBPb and Let1p, are distinguished from other subgroups of the ATPase family, including those involved in protein degradation, by peculiarly basic pIs. However, the pI of 8.35 calculated here for p45 is close to the mean of the transcriptional ATPases surveyed by Swaffield, indicating that pIs may not provide a clear prediction of function.

Presumably one role of the ATPase is to supply energy continuously for the selective degradation of ubiquitinated proteins by the active 26S proteasome complex. Although the multiple ATPases in proteasomes of eukaryotes ranging from yeast to mammals have common features, it is unclear why so many ATPases are associated with the 26S proteasome complex. Because a single ATPase subunit may be sufficient to supply energy for proteolysis, these ATPases may have other functions besides supplying energy for proteolysis. Alternatively, each ATPase may be associated with a distinct 26S complex, each with different functional properties. Thus, 26S proteasomes

Fig. 3. Comparison of the protein sequences of p45, TBP-1, TBP-7, S4 and MSS1 of human putative ATPases. Identical amino acid residues are shown by black box. Numbers are residue numbers. Sequence data for TBP-1 [5], TBP-7 [6], S4 [3] and MSS1 [4] are cited. Asterisks indicate putative ATP binding sequences.

The possible functional significance of such multiple 26S proteasomes is unclear. Obviously, the connection between the roles of ATPase family members in proteasome regulation and other cellular processes such as transcription remains to be determined.

Acknowledgements: This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan (to K.T.), the National Institutes of Health (DK 46181 to G.N.D.) and the National Science Foundation (MCB 9219352 to C.A.S.). We thank Stephen A. Johnston and Jonathan C. Swaffield for critical reading of the manuscript, and Carolyn R. Moomaw for amino acid sequencing.

**Saccharomyces cerevisiae*. For details of these ATPases, see text. Note that YTA5 is a homolog of *Saccharomyces pombe* MTS2.

- [1] Rechsteiner, M., Hoffman, L. and Dubiel, W. (1993) *J. Biol. Chem.* 268, 6065–6068.
- [2] Ciechanover, A. (1994) *Cell* 79, 13–21.
- [3] Dubiel, W., Ferrell, K., Pratt, G. and Rechsteiner, M. (1992) *J. Biol. Chem.* 267, 22669–22702.

- [4] Shibuya, H., Irie, K., Ninomiya-Tsuji, J., Goebel, M., Taniguchi, T. and Matsumoto, K. (1992) *Nature* 357, 700–702.
- [5] Nelbock, P., Dillon, P.J., Perkins, A. and Rosen, C.A. (1990) *Science* 248, 1650–1653.
- [6] Ohana, B., Moore, P.A., Ruben, S.M., Southgate, C.D., Green, M.R. and Rosen, C.A. (1993) *Proc. Natl. Acad. Sci. USA* 90, 138–142.
- [7] Dubiel, W., Ferrell, K. and Rechsteiner, M. (1993) *FEBS Lett.* 323, 276–278.
- [8] Gordon, C., McGurk, G., Dillon, P., Rosen, C. and Hastle, N. (1993) *Nature* 366, 355–357.
- [9] Ghisla, M., Udvardy, A. and Mann, C. (1993) *Nature* 366, 358–362.
- [10] Swaffield, J.C., Bromberg, J.F. and Johnston, S.A. (1992) *Nature* 357, 698–700.
- [11] Schnall, R., Mannhaupt, G., Stucka, R., Tauer, R., Ehnle, S., Schwarzlose, C., Vetter, I. and Feldmann, H. (1994) *Yeast* 10, 1141–1155.
- [12] Lee, J.W., Ryan, F., Swaffield, J.C., Johnston, S.A. and Moore, D.D. (1995) *Nature* 374, 91–94.
- [13] Kim, Y.-J., Björklund, S., Li, Y., Sayre, M.H. and Kornberg, R.D. (1994).
- [14] Chu-Ping, M., Vu, J.H., Proske, R.J., Slaughter, C.A. and DeMartino, G.N. (1994) *J. Biol. Chem.* 269, 3539–3547.
- [15] DeMartino, G.N., Moomaw, C.R., Zagniko, O.P., Proske, R.J., Chu-Ping, M., Afendis, S.J., Swaffield, J.C. and Slaughter, C.A. (1994) *J. Biol. Chem.* 269, 20878–20884.
- [16] Kozak, M. (1984) *Nucleic Acids Res.* 12, 857–872.
- [17] Skoog, B. and Wichman, A. (1986) *Trends Anal. Chem.* 5, 82–83.
- [18] Swaffield, J.C., Melcher, K. and Johnston, S.A. (1995) *Nature* 374, 88–91.
- [19] Dawson, S.P., Arnold, J.E., Mayer, N.J., Reynolds, S.E., Billett, M.A., Gordon, C., Colleaux, L., Kloetzel, P.M., Tanaka, K. and Mayer, R.J. (1995) *J. Biol. Chem.* 270, 1850–1858.