

cDNA cloning of rat proteasome subunit RC10-II, assumed to be responsible for trypsin-like catalytic activity**

Chihiro Nishimura^a, Tomohiro Tamura^b, Hiroshi Akioka^b, Fuminori Tokunaga^c, Keiji Tanaka^{b,*}, Akira Ichihara^b

^aLaboratory of Applied Toxicology, Nippon Kayaku Co. Ltd., 239 Iwahana-machi, Takasaki-city, Gunma, 370-12, Japan

^bInstitute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan

^cHimeji Institute of Technology, Faculty of Science, Harima Science Park City, Hyogo 678-12, Japan

Received 12 November 1993

The nucleotide sequence of a cDNA that encodes a new subunit, named RC10-II, of the 20S proteasome of rat embryonic brain has been determined. The polypeptide predicted from the open reading frame consists of 205 amino acid residues with a calculated molecular weight of 22,965 and isoelectric point of 6.15. Computer analysis showed that RC10-II belongs to the β -type subgroup of proteasomes, differing clearly from α -type subunits of the proteasome gene family. The primary structure of RC10-II was found to contain the partial amino acid sequences of several fragments of subunit 13, which has a cysteinyl residue critical for the trypsin-like catalytic activity, as reported by L.R. Dick et al. [Biochemistry 31, 7347–7355, 1992], suggesting that RC10-II is a proteasomal subunit necessary for the expression of tryptic activity.

cDNA cloning; Multicatalytic proteinase; Proteasome; Subunit RC10-II; Trypsin-like activity

1. INTRODUCTION

The proteasome is a 20S multifunctional proteinase complex, consisting of 14–16 non-identical components of 21–32 kDa [1]. The 20S proteasome is associated with a set of regulatory proteins including multiple ATPases to form the 26S proteasome complex that is a eukaryotic ATP-dependent protease [2–4]. The 26S proteasome catalyzes selective breakdown of naturally occurring short-lived proteins related with cell cycle progression such as Mos [5] and ornithine decarboxylase [6]. Moreover, it is a processing enzyme involved in a pathway of MHC class I-restricted antigen presentation [7]. For determining the functions of this proteasomal multi-subunit complex, we are attempting to clarify the entire structure of the rat 20S proteasome by recombinant DNA techniques, and so far we have isolated and sequenced cDNAs for 11 subunits [8–15]. Structural information on proteasomes from various eukaryotes determined by us and others showed that proteasomes form a multi-gene family with a common evolutionary origin [1]. The proteasomal subunits, which have considerably high inter-subunit homology, has been proposed to be classified into two subgroups, α and β , judging from their high similarities to the α - and β -subunit, respectively, of the archaebacterial proteasome

[16]. During structural analyses of the rat proteasome, we recently isolated a cDNA encoding a new subunit named RC10-II which belongs to be the β -type subgroup of the proteasome gene family. The primary structure of the RC10-II was shown to be consistent with the partial amino acid sequences of several fragments of subunit 13, which has a cysteinyl residue critical for the trypsin-like catalytic activity, as reported by L.R. Dick et al. [17]. Thus RC10-II seems to be responsible for trypsin-like activity of the 20S proteasome.

2. MATERIALS AND METHODS

Subunit RC10-II was isolated from purified rat liver proteasomes, and its fragments were obtained by digestion with lysyl-endopeptidase as reported previously [8]. The amino acid sequences of the fragments were determined with a gas-phase sequencer (Applied Biosystems, model 477A), and 120A phenylthio-hydantoin analyzer on-line system [8]. A cDNA library of rat embryonic brain on day 18 of gestation (a gift from Dr. A. Kakizuka, Kyoto University) was constructed in a λ ZAPII phage expression vector (Stratagene). For isolation of cDNA for RC10-II, 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. Preparation of a probe for screening by PCR reaction

Previously, we reported the separation of multiple

*Corresponding author.

**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: D21800.

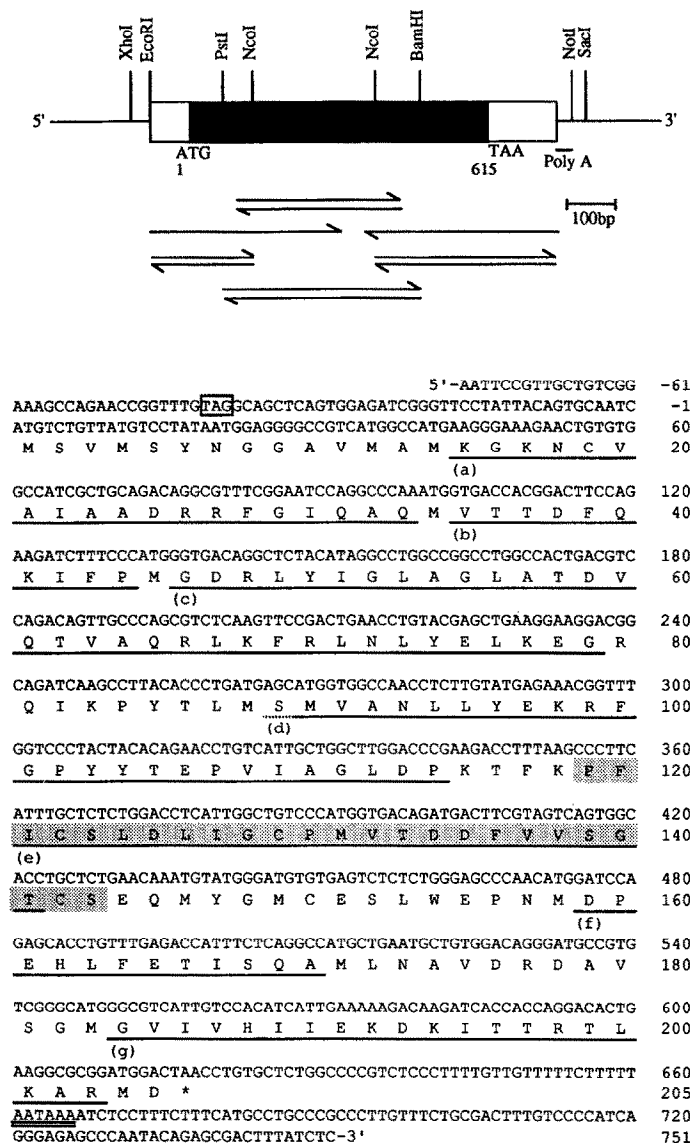


Fig. 1. Structure of cDNA for subunit RC10-II of rat proteasomes. Upper panel: Restriction endonuclease map of cloned cDNA for RC10-II and sequencing strategy. The solid and open boxes show the coding region and 5'- and 3'-noncoding regions, respectively. Continuous lines indicate the sequence of the 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, TAA. Sequenced regions are shown by horizontal arrows. The bar represents 100 bp. Lower panel: Nucleotide sequence of the cDNA encoding component RC10-II 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 ATG³. The nucleotides on the 5' side are indicated by negative numbers. Stop codon -43TAG-40 in the 5'-noncoding region is boxed. The predicted amino acid sequence of RC10-II is shown below the nucleotide sequence. Amino acid residues are numbered from the N-terminus. The amino acid sequence obtained by Edman degradation of a major fragment cleaved with lysyl-endopeptidase of RC10-II is shaded. Continuous underlines (a-g) show the amino acid sequences corresponding to those obtained by Edman degradation of fragments of bovine subunit 13 cleaved with trypsin, chymotrypsin and cyanogen bromide [17]. Note that the single amino acid shown by a dotted line was not identical with that found by chemical analysis. The termination codon TAA is marked with an asterisk. The possible polyadenylation signal (AATAAA) is doubly underlined.

components of proteasomes from rat liver by reversed-phase high-performance liquid chromatography [18]. First 10 major components were separated on a Cosmosil 5C₄-300 column and named component 1 (C1) to component 10 (C10) in order of their elution. These components from rats were recently renamed RC1 to RC10 to distinguish them from those of other species. RC10 was eluted from the column with 61% acetonitrile.

This RC10 fraction gave bands of two components of different sizes (I and II) on SDS-PAGE and these were separated by rechromatography on a TSK-GEL phenyl-5PW column. For determination of the primary structure of its internal region, samples of RC10-II were reduced, S-pyridylethylated and digested with lysyl-endopeptidase. The resulting fragments were resolved by reversed-phase HPLC on a Chemcosorb 7-ODS-H col-

umn and their partial primary structures were analyzed by automated Edman degradation. Surprisingly, one sequence of PFICSLDLIGCPMVTDDFVVSQTCS is completely in accord with the ICSLDLIGCPMVTDDFVVSQT of a fragment derived from the subunit 13 of bovine proteasomes which was reported by Dick et al. [17]. Thus we concluded that rat RC10-II is a homologue of subunit 13 and used the partial amino acid sequences of the subunit 13 to prepare a probe for screening. For use as PCR primers, we selected parts of the sequences of two of the proteolytic fragments with ³⁴MVTTFDFQKIFP⁴⁴ and ¹²⁸GCPMVTDDFV¹³⁷, respectively (see Fig. 1). The following oligonucleotides corresponding to the protein sequences were synthesized:

Forward primer: 5'-ATGGTTACTACTGATTTTCA-GAAGATTTTCC-3'.

Reverse primer: 5'-ACAAAATCATCAGTAACCAT-AGGACAACC-3'.

Using these primers, a fragment of approximately 300 bp was synthesized by PCR against first strand cDNA complementary to mRNA from human hepatoblastoma HepG2 cells as a template.

3.2. Isolation of a cDNA clone encoding RC10-II

To isolate cDNA encoding RC10-II, we screened a cDNA library with the λ ZAPII vector using poly(A)⁺RNAs extracted from rat embryonic brain by hybridization with a cDNA fragment with approximately 300 bp synthesized by PCR as a probe. We then screened about 5×10^5 plaques of the rat embryonic brain cDNA library with the cDNA fragment. The

14 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 0.85 kb length including a poly(A) tail, and was subjected to cDNA sequencing.

3.3. Primary structure of RC10-II

The nucleotide sequence of the RC10-II cDNA clone and the primary structure of the RC10-II protein deduced from the cDNA sequence are shown in Fig. 1. The sequence of 828 nucleotides included the entire coding region and 5'- and 3'-noncoding regions. The 3'-noncoding region consisted of 133 nucleotides, excluding the poly(A) tail. A putative polyadenylation signal (AATAAA), which is common to eukaryotic mRNAs, was located 91 nucleotides upstream from the poly(A) addition site. We concluded that ATG, located at nucleotides 1 to 3, is the initiation codon, because it is surrounded by a sequence that is similar to the consensus sequence for translation initiation [19], and because stop codon TAG was found in the 40 bp upstream from the largest open reading frame. Subunit RC10-II corresponds to a protein of 205 amino acids with a calculated molecular weight of 22,965. The amino acid sequence shown in Fig. 1 was confirmed to be that of RC10-II of proteasomes by showing that the partial sequence of a fragment, PFICSLDLIGCPMVTDDFVVSQTCS, determined chemically (Fig. 1, shaded residues) was completely in accord with that deduced from the nucleotide sequence of the cDNA. The isoelectric point (pI) of RC10-II was calculated to be 6.15 by the method of Skoog and Wichman [20].

Table I
Similarities of primary structures of various rat proteasome subunits

| | Ta- α | Ta- β | RC2 | RC3 | RC8 | RC9 | rIOTA | rZETA | RC1 | rLMP2 | rDELTA | RC5 | RN3 | RC7-I | RC10-II |
|--|--------------|-------------|------------|------------|------------|------------|------------|-------|------------|------------|--------|------|------|-------|---------|
| α-Type Subunits | | | | | | | | | | | | | | | |
| RC2 | <u>384</u> | 163 | 1279 | | | | | | | | | | | | |
| RC3 | <u>419</u> | 150 | <u>336</u> | 1148 | | | | | | | | | | | |
| RC8 | <u>323</u> | 148 | <u>262</u> | <u>299</u> | 1238 | | | | | | | | | | |
| RC9 | <u>382</u> | 162 | <u>263</u> | <u>405</u> | <u>298</u> | 1271 | | | | | | | | | |
| rIOTA | <u>387</u> | 96 | <u>292</u> | <u>355</u> | <u>336</u> | <u>355</u> | 1221 | | | | | | | | |
| rZETA | <u>435</u> | 180 | <u>333</u> | <u>362</u> | <u>280</u> | <u>338</u> | <u>317</u> | 1115 | | | | | | | |
| β-Type Subunits | | | | | | | | | | | | | | | |
| RC1 | 116 | <u>296</u> | 40 | 139 | 38 | 98 | ns | 102 | 1056 | | | | | | |
| rLMP2 | 79 | <u>195</u> | 44 | 36 | 42 | 93 | 60 | 95 | <u>249</u> | 1047 | | | | | |
| rDELTA | 42 | <u>220</u> | ns | ns | 54 | 115 | 50 | 85 | <u>256</u> | <u>659</u> | 977 | | | | |
| RC5 | 36 | <u>191</u> | 56 | 61 | 33 | 90 | 85 | 50 | <u>103</u> | <u>101</u> | 85 | 1169 | | | |
| RN3 | 50 | 63 | 36 | 46 | 30 | 38 | 35 | 40 | 91 | 83 | 59 | 76 | 1150 | | |
| RC7-I | 96 | 180 | 72 | 62 | 100 | 76 | 68 | 64 | 106 | 93 | 102 | 157 | 120 | 1004 | |
| RC10-II | 59 | 193 | 70 | 35 | 39 | 87 | 87 | 36 | 132 | 195 | 147 | 208 | 49 | 121 | 1039 |

The sources of sequence data are as follows: RC1 [14]; RC2 [8]; RC3 [9]; RC5 [10]; RC8 [11]; RC9 [12]; rIOTA, rDELTA, rZETA, rLMP2 [13]; RC7-I [15]; RN3 [22]; Ta- α and Ta- β [16]. The sequences of the subunits were aligned to achieve maximal homology, and scores of the resulting pairs were determined by computer analysis according to Dayhoff et al. [28]. The extent of similarity between pairs of subunits was deduced by comparing the scores of the pairs with those of the two subunits. ns indicates no significant homology (Dayhoff score < 28). High scores (> 220) for similarities of β -type subunits are underlined.

3.4. Identity of the RC10-II with a subunit 13 responsible for the trypsin-like catalytic activity

Computer analysis showed no obvious overall structural similarity of component RC10-II with most previously reported proteins. Interestingly, however, RC10-II is assumed to be a homologue of the subunit 13 from bovine proteasomes, because almost all the sequences of 7 fragments of subunit 13 of bovine proteasomes reported by Dick et al. [17] were in excellent accordance with those deduced from the nucleotide sequences (solid underlines a–g, in Fig. 1), which was observed throughout almost the entire sequence and because 141 amino acid residues of subunit 13 determined chemically were identical to those of RC10-II except one amino acid residue (Fig. 1, the dotted line), which corresponds to 69% of total amino acid residues of RC10-II. Interestingly, subunit 13 was found to have a cysteinyl residue for trypsin-like catalytic activity, because leupeptin, a reversible competitive inhibitor of the 'trypsin-like' activity, selectively inhibited tryptic activity without affecting the chymotrypsin-like and peptidylglutamyl-peptide hydrolyzing (PGPH) activities of the proteasome, and because subunit 13 was the only subunit protected by leupeptin from chemical modification with *N*-ethylmaleimide [17]. Thus an *N*-ethylmaleimide-modified, leupeptin-protected cysteine residue of the subunit 13 contributes to the active site responsible for the proteasome's trypsin-like activity. Judging from the similarity of subunit 13 with RC10-II, the subunit RC10-II may be involved in the expression of the trypsin-like activity in the rat proteasome complex.

3.5. Inter-subunit homology of RC10-II in rat proteasomes

Proteasomal subunits, which have considerably high inter-subunit homology, can be classified into two subfamilies with high similarities to the α - and β -subunit, respectively, of the archaeobacterial proteasome [16]. Table I shows a computer-assisted homology analysis of the rat proteasomal subunits sequenced so far [8–15,21]. Interestingly, members of the α -subunit family with high similarity to the Ta- α subunit of the archaeobacterial proteasome [16] are highly homologous with each other. On the other hand, all other subunits have been classified as the β -subunit family [15,21], but all these subunits have a similar score against the Ta- β subunit of the archaeobacterial proteasome [16] except RN3 which has less similarity. Interestingly in the β -subunit family, RC1 (rLMP7) and rLMP2, which are encoded within the MHC class II gene cluster and assumed to be responsible for processing of endogenous antigens [16], are highly homologous. Moreover, it is of note that the protein sequence of rDELTA has considerably high similarity to that of rLMP2, suggesting their functional relationship. In addition, since the similarity score of the Ta- α and Ta- β subunits is 215, the α -subunit family have been evolutionally conserved, whereas the β -subunit family may have diverged at an earlier stage and conserved during evolution.

The α - and β -subgroups probably have distinct functions. As summarized in Table II, most α -type subunits contain a consensus signal sequence for nuclear translocation (NLS) and NLS complementary sequence

Table II
Properties of rat proteasome subunits and their yeast counterparts

| Rat subunits Name | Amino acid residues | M_r | pI | Possible functions or property | Yeast subunits |
|-------------------------------------|------------------------|-------|------|-----------------------------------|-------------------|
| α-Subunits | | | | | |
| RC2 | 263 | 29526 | 6.15 | cNLS motif | |
| RC3 | 234 | 25925 | 7.29 | NLS motif | Y7 |
| RC8 | 255 | 28417 | 5.16 | cNLS motif | PRS1 |
| RC9 | 261 | 29496 | 7.69 | NLS and cNLS motifs | Y13 |
| rIOTA | 246 | 27399 | 6.37 | NLS motif | PRS2 |
| rZETA | 241 | 26391 | 4.65 | | PUP2 |
| β-Subunits | | | | | |
| RC1 (rLMP7) | 208 | 23130 | 8.48 | Antigen processing | PRE2 |
| rLMP2 | 219 | 23324 | 4.70 | Antigen processing | |
| rDELTA | 202 | 21649 | 4.84 | | |
| RC5 | 240 | 26479 | 7.15 | | PRS3 |
| RN3 | 232 | 25746 | 5.94 | PGDH activity | PRE4 |
| RC7-I | 201 | 22912 | 7.25 | Chymotrypsin-like activity | PRE1 |
| RC10-II | 205 | 22965 | 6.15 | Trypsin-like activity | |

PGPH activity, peptidylglutamyl-peptide hydrolyzing activity. Sequence data of rat proteasomes are cited from the references shown in Table I. The sources of sequence data of yeast proteasomes are as follows: PRS1 and PRS2 [29], PRS3 [30]; Y7 and Y13 [26]; PRE1 [23]; PRE2 [27]; PRE4 [24] and PUP2 [31]. Identity of rat subunits with their yeast counterparts was determined by computer analysis. Isoelectric points (pI) were calculated by the method of Skoog and Wichman [20]. The subunits of yeast (*Saccharomyces cerevisiae*) proteasomes indicate strongly conserved residues compared with those of rat proteasome subunits.

(cNLS), which consist of basic and acidic amino acid residues, respectively [1,22]. Thus one role of the α -subunit family may be in control of the intracellular distribution of the proteasome. In contrast, several subunits related with proteolytic functions belong to the β -subunit family, suggesting that most β -type subunits in the proteasome complex have catalytic functions [7]. For examples, RC7-I and RN3 are homologues of yeast PRE1 and PRE4, which are necessary for chymotryptic and PGPH activities, respectively [23,24]. In this paper, we showed that RC10-II may be involved in the tryptic activity. Moreover, RC1 (rLMP7) and rLMP12 are suggested to be involved in antigen processing [16,25]. As shown in Table II, many counterparts of rat proteasome subunits were found in yeast. Interestingly, almost all the yeast proteasomal genes are essential for yeast cell proliferation except Y13 gene [1,26]. However, the RC1 (LMP7) genes of the mouse and human proteasomes are not essential [25], a fact which seems to be related with acquirement of immunity, but gene disruption of its yeast counterpart PRE2 is lethal [27]. Thus, the divergence of β -type subunits including RC10-II may have been associated with the acquirement of specific functions of proteasomes.

Acknowledgements: This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- [1] Tanaka, K., Tamura, T., Yoshimura, T. and Ichihara, A. (1992) *The New Biologist* 4, 173–187.
- [2] Goldberg, A.L. (1992) *Eur. J. Biochem.* 204, 9–23.
- [3] Hershko, A. and Ciechanover, A. (1992) *Annu. Rev. Biochem.* 61, 761–807.
- [4] Rechsteiner, M., Hoffman, L. and Dubiel, W. (1993) *J. Biol. Chem.* 268, 6065–6068.
- [5] Ishida, N., Tanaka, K., Tamura, T., Nishizawa, M., Okazaki, K., Sagata, S. and Ichihara, A. (1993) *FEBS Lett.* 324, 345–348.
- [6] Murakami, Y., Matsufuji, S., Kameji, T., Hayashi, S., Igarashi, T., Tamura, K., Tanaka, K. and Ichihara, A. (1992) *Nature* 360, 597–599.
- [7] Zwickl, P., Grziwa, G., Puhler, G., Dahlmann, B., Lottspeich, F. and Baumeister, W. (1992) *Biochemistry* 31, 964–972.
- [8] Fujiwara, T., Tanaka, K., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A. and Nakanishi, S. (1989) *Biochemistry* 28, 7332–7340.
- [9] Tanaka, K., Fujiwara, T., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A. and Nakanishi, S. (1990) *Biochemistry* 29, 3777–3785.
- [10] Tamura, T., Tanaka, K., Kumatori, A., Yamada, F., Tsurumi, C., Fujiwara, T., Ichihara, A., Tokunaga, F., Aruga, R., and Iwanaga, S. (1990) *FEBS Lett.* 264, 91–94.
- [11] Tanaka, K., Kanayama, H., Tamura, T., Lee, D. H., Kumatori, A., Fujiwara, T., Ichihara, A., Tokunaga, F. and Iwanaga, S. (1990) *Biochem. Biophys. Res. Commun.* 171, 676–683.
- [12] Kumatori, A., Tanaka, K., Tamura, T., Fujiwara, T., Ichihara, A., Tokunaga, F., Onikura, A. and Iwanaga, S. (1990) *FEBS Lett.* 264, 279–282.
- [13] Tamura, T., Shimbara, N., Aki, M., Ishida, N., Bey, F., Scherrer, K., Tanaka, K. and Ichihara, A. (1992) *J. Biochem.* 112, 530–534.
- [14] Aki, M., Tamura, T., Tokunaga, F., Iwanaga, S., Kawamura, Y., Shimbara, N., Kagawa, S., Tanaka, K. and Ichihara, A. (1992) *FEBS Lett.* 301, 65–68.
- [15] Nishimura, C., Tamura, T., Tokunaga, F., Tanaka, K., and Ichihara, A. (1993) *FEBS Lett.* 332, 52–56.
- [16] Goldberg, A. and Rock, K.L. (1992) *Nature* 357, 75–379.
- [17] Dick, L.R., Moomaw, C.R., Premanik, B.C., DeMartino, J. and Slaughter, C.A. (1992) *Biochemistry* 31, 7347–7355.
- [18] Tanaka, K., Yoshimura, T., Ichihara, A., Ikai, A., Nishigai, M., Morimoto, M., Sato, M., Tanaka, N., Katsube, Y., Kameyama, K. and Takagi, T. (1988) *J. Mol. Biol.* 203, 985–996.
- [19] Kozak, M. (1984) *Nucleic Acids Res.* 12, 857–872.
- [20] Skoog, B. and Wichman, A. (1986) *Trends Anal. Chem.* 5, 82–83.
- [21] Thompson, S., Balson, D.F. and Rivett, A.J. (1993) *FEBS Lett.* 322, 135–138.
- [22] Tanaka, K., Yoshimura, T., Tamura, T., Fujiwara, T., Kumatori, A. and Ichihara, A. (1990) *FEBS Lett.* 271, 41–46.
- [23] Heinemeyer, V., Kleinschmidt, J.A., Saidowsky, J., Escher, C. and Wolf, D.H. (1991) *EMBO J.* 10, 555–562.
- [24] Hilt, W., Enenkel, C., Gruhler, A., Singer, T. and Wolf, D.H. (1993) *J. Biol. Chem.* 268, 3479–3486.
- [25] Monaco, J.J. (1992) *Immunol. Today* 13, 173–179.
- [26] Emori, Y., Tsukahara, T., Kawasaki, H., Ishiura, S., Sugita, H. and Suzuki, K. (1991) *Mol. Cell Biol.* 11, 343–353.
- [27] Heinemeyer, W., Gruhler, A., Mohrle, V., Mahe, Y. and Wolf, D.H. (1993) *J. Biol. Chem.* 268, 5115–5120.
- [28] Dayhoff M.O., Schwartz, R.M., Orcutt, B.C. (1978): in *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., Ed.) Vol. 5, Suppl. 3, National Biomedical Research Foundation, Silver Spring, MD.
- [29] Fujiwara, T., Tanaka, K., Orino, E., Yoshimura, T., Kumatori, A., Tamura, T., Chung, C.H., Nakai, T., Yamaguchi, K., Shin, S., Kakizuka, A., Nakanishi, S. and Ichihara, A. (1990) *J. Biol. Chem.* 265, 16604–16613.
- [30] Lee, D.H., Tanaka, K., Tamura, T., Chung, C.H. and Ichihara, A. (1992) *Biochem. Biophys. Res. Commun.* 182, 452–460.
- [31] Georgatsou, E., Georgakopoulos, T. and Thireos, G. (1992) *FEBS Lett.* 299, 39–43.