Volume 278, number 2, 217–221 FEBS 09360 January 1991
© 1991 Federation of European Biochemical Societies 0014-5793/91/$3.50
ADONIS 001457939100078T

Cloning and sequencing of the gene encoding the large (α-) subunit of the proteasome from *Thermoplasma acidophilum*

Peter Zwickl[1], Friedrich Lottspeich[1], Burkhardt Dahlmann[2] and Wolfgang Baumeister[1]

[1]*Max-Planck-Institut für Biochemie, D-8033 Martinsried* and [2]*Diabetes-Forschungsinstitut an der Universität Düsseldorf, D-4000 Düsseldorf, Germany*

Received 29 November 1990

The gene encoding the α-subunit of the proteasome from the archaebacterium *Thermoplasma acidophilum* was cloned and sequenced. The gene encodes for a polypeptide with 233 amino acid residues and a calculated molecular weight of 25870. Sequence similarity of the α-subunit with the *Saccharomyces cerevisiae* wild-type suppressor gene *scl1*$^+$ encoded polypeptide, which is probably identical with the subunit YC7 α of the yeast proteasome, lends support to a putative role of proteasomes in the regulation of gene expression. The significant sequence similarity to the various subunits of eukaryotic proteasomes make it likely that proteasomal proteins are encoded by one gene family of ancient origin.

Proteasome, Prosome, Multicatalytic proteinase, Archaebacterium, *Thermoplasma acidophilum*

1. INTRODUCTION

The proteasome is a high molecular mass (approx. 700 kDa) cylinder-shaped particle with multicatalytic proteinase activity (for review, see ref. 1). It is found in all eukaryotes from yeast to man [2]. It is widely accepted now that the proteasome or multicatalytic proteinase [3] is identical to the prosome [4,5], a 19–20 S ribonucleoprotein supposed to be involved in the regulation of gene expression. While there is no evidence for the existence of proteasomes in eubacteria [2,6], we have recently isolated proteasomes from the thermoacidophilic archaebacterium *Thermoplasma acidophilum* [7]. Contrary to the complex subunit composition of eukaryotic proteasomes (8–16 different subunits, all in the molecular mass range between 20 and 35 kDa) proteasomes from *T. acidophilum* comprise only two different subunits. The α-subunit has an apparent M_r of 27 kDa, the smaller β-subunit has an apparent M_r of 25 kDa [7]. Nevertheless, proteasome complexes of eukaryotic origin and from *T. acidophilum* are almost identical in size and shape, as shown by electron microscopy in conjunction with digital image processing [7,8]. In this communication we report the cloning and sequencing of the gene encoding the large (α-) subunit of the *T. acidophilum* proteasome. This is an important prerequisite for future more detailed structural studies and can be expected to give some hints as to the physiological role of proteasomes and the tasks performed by their constituent subunits.

Correspondence address W. Baumeister, Max-Planck-Institut für Biochemie, Abteilung Molekulare Strukturbiologie, Am Klopferspitz 18a, D-8033 Martinsried, Germany

Abbreviations. crl, mutation encoding cycloheximide and temperature sensitive lethality, HPLC, high-performance liquid chromatography, *scl1*$^+$, wildtype allele of *scl1; SCL1-1*, mutant allele of *scl1*, dominant suppressor of *crl3* temperature-sensitive lethality, SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, TFA, trifluoroacetic acid, YC, yeast component of the proteasome

2. MATERIALS AND METHODS

Proteasomes from *T. acidophilum* were purified to homogeneity as reported previously [7], 25 μg were separated by tricine-SDS-PAGE [9] and detected by Coomassie blue staining. For cleavage with cyanogen bromide the α-subunit was electroeluted and precipitated with acetone. The redissolved protein was treated with 10% (w/v) cyanogen bromide in 70% (v/v) formic acid at room temperature for 24 h. The resulting peptides were separated by tricine-SDS PAGE [9], electroblotted onto a siliconized glass-fiber sheet (Glassybond, Biometra, Gottingen, Germany), detected by Coomassie blue staining, excised, and subjected to sequence degradation [10]. For cleavage with trypsin, gel-slices containing the α-subunit were cut into small pieces and suspended in 0.1 M ammonium carbonate. The protein was digested with 5 μg trypsin at room temperature for 24 h. Tryptic peptides were extracted from gel pieces, concentrated and separated by HPLC. The column (Hibar LiChrospher-C_{18} column, 125 × 4 mm, Merck, Darmstadt, Germany) was eluted with a linear gradient of 0–70% acetonitrile (0.1% TFA). The effluent was monitored at 206 nm. Fractions containing tryptic peptides were collected, concentrated and subjected to sequence degradation [11].

Cloning of DNA-fragments into pUC18-vectors, construction of small gene libraries and colony screening were done as described previously [12]. Inserts of recombinant double-stranded pUC18-plasmids were sequenced by the dideoxy chain-termination method using a Sequenase Sequencing Kit (United States Biochemical, Cleveland, OH).

Amino acid sequence alignments of various subunits of eukaryotic proteasomes and the *T. acidophilum* α-subunit were performed with 'CLUSTAL' [13], a software package for multiple sequence align-

ment. Search for protein motifs in the amino acid sequence was performed with 'PROSITE' [14], a computer-edited dictionary of protein sites and patterns.

3. RESULTS AND DISCUSSION

3.1. *Protein sequencing of the α-subunit*

As the N-terminus of the α-subunit was found to be blocked, the purified protein was cleaved with cyanogen bromide and trypsin, respectively. The isolated peptides were subjected to partial amino acid sequence analysis by automated Edman degradation (underlined sequences in Fig. 1).

3.2. *Cloning and sequencing of the α-subunit-gene*

Southern blots of genomic restriction fragments of *T. acidophilum* DNA were probed with two oligonucleotide mixtures, the first (5'-TTYCARGTNGARTAYGC-3') derived from a tryptic and the second (5'-ATGAARTTYGCNAAYGG-3') derived from a cyanogen bromide peptide; (N = AGCT, R = AC, Y = TC). Fragments in the range from 4 to 4.5 kb of *Hin*dIII-digested DNA, which gave strong hybridization signals with both probes, were ligated in pUC18 and *E. coli* DH5α was transformed with the recombinant plasmids. One positive clone, harbouring a 4.3 kb *Hin*dIII fragment, was identified by colony hybridization of a small gene library of about 900 clones. The nucleotide sequence of the α-subunit gene and its 5'- and 3'-flanking regions was determined (see Fig. 1). There were 3 possible initiation-codons for the open reading frame encoding the α-subunit, but only one was preceded by the putative Shine-Dalgarno sequence GGAGGA, which is complementary to the 3'-end of the *T. acidophilum* 16 S rRNA [15]. Starting with this initiation codon, the α-subunit gene codes for a polypeptide of 233 amino acid residues (Fig. 1), giving rise to a molecular weight of 25870. This is in good agreement with the apparent molecular mass of 27 kDa obtained by SDS-PAGE [7].

3.3. *Analysis of the derived protein sequence of the α-subunit gene*

Scrutinizing the derived amino acid sequence of the *T. acidophilum* α-subunit for known primary structure motifs with 'PROSITE' [14], a computer-edited dictionary of protein sites and patterns, we found a sequence with homology to cAMP/cGMP-dependent phosphorylation sites between amino acid residues 31 and 37 [16,17] (see Table I). As discussed also for some eukaryotic proteasomal proteins [22,24], we identified a potential tyrosine autophosphorylation site at amino acid residue 123 [18], a consensus nuclear location signal (NLS) between amino acid residues 49 and 56 [19] and a region complementary to the NLS-consensus sequence between amino acid residues 201 and 212 [20]

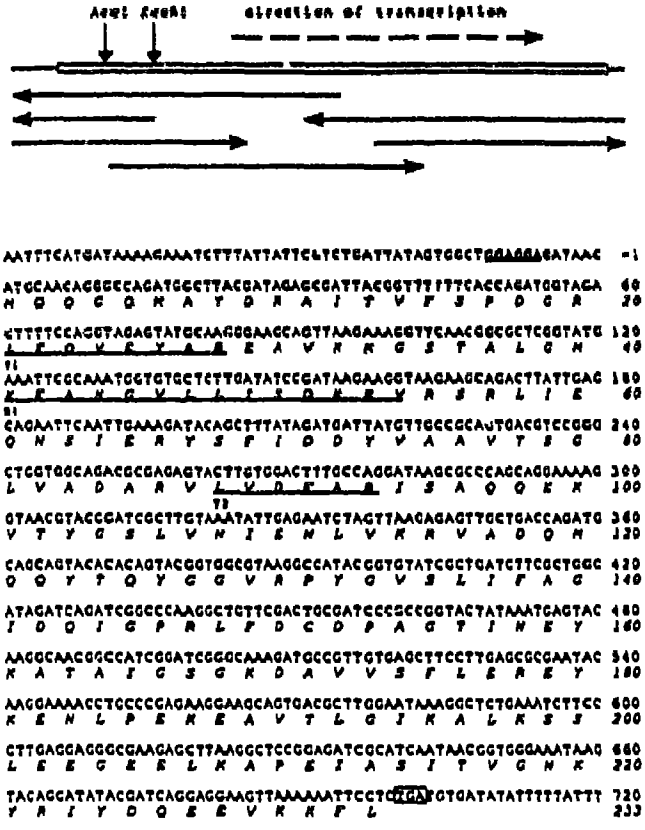

Fig. 1. Restriction map, sequencing strategy, nucleotide sequence and derived amino acid sequence of the DNA-fragment containing the α-subunit-gene. The upper panel shows the restriction endonuclease map of the cloned DNA-fragment containing the α-subunit gene. The open box indicates the coding region of the gene. Continuous lines indicate 5' and 3'-non-coding regions. Vertical arrows indicate restriction endonuclease cleavage sites. Horizontal arrows indicate the individual sequence runs. The dotted arrow shows the direction of transcription of the α-subunit-gene. The lower panel shows the nucleotide sequence and the derived amino acid sequence of the DNA-fragment containing the α-subunit gene. Nucleotides are numbered beginning with the first residue of the putative initiation codon ATG. The nucleotides upstream from the initiation codon are indicated by negative numbers. The underlined nucleic acids represent the putative Shine-Dalgarno sequence. The stop codon is boxed. The predicted amino acid sequence is shown below the nucleotide sequence in the one letter code. Amino acid residues are numbered beginning with the putative N-terminal methionine. Underlined residues were determined by protein sequencing. T1 and T2 mark the sequences of two tryptic peptides. B1 marks the sequence of a cyanogen bromide peptide.

(see Table I). The identification of a consensus NLS-sequence (X-K-K(R)-X-K(R); X standing for any amino acid [19]) in a protein from an archaebacterium, i.e. an organism without a distinct nucleus, is of course puzzling. It is unlikely that the sequence similarity at this particular sequence position is entirely coincidental. It is more reasonable to assume, that either the target of the NLS-sequence exists also in *T. acidophilum*, or that this

TABLE I

Sequence regions of proteasomal proteins, which show homology to consensus sequences of phosphorylation sites or nuclear location signals (NLS) or NLS complementary sequences

Proteasomal protein	Putative tyrosine autophosphorylation sites	Putative cAMP/cGMP-dependent phosphorylation sites	Putative NLS sequence	Amino acid sequence complementary to NLS sequence
Tac	112-LVKRVADQMQQY TQ-YGG--VRPY-132	32-KKGST-36	50-SDKKVR-55	202-EEGEELKAPE-211
RnC2	nf	nf	nf	218-DLEFTIYDDDD-228
RnC3	110-LVQRVASVMQE Y TQ-SGG-VRPF-130	nf	48-TEKKQK-53	nf
RnC8	nf	nf	nf	246-EEDESDDD-253
RnC9	110-LVTALCDIKQA Y TQ-FGG- -KRPF-130	nf	249-REKKEK-254	241-EEEEAKAERE-250
Dm28.1	nf	29-RKGST-33	243-EKKKQK 248	nf
Dm29	110-LVSHLCDIKQA Y TQ-YGG-KRPF-130	nf	252-AAKKEK-257	nf
ScC1	nf	nf	nf	250-DDDEDEDDSD-259
ScC7α	117-LAKRMANLSQI YTQR- -AYMRPL-137	nf	185-HFKKSK-190	nf

Sequence designations are as given in Fig. 2, nf stands for not found. Numbers indicate position of the first or last residue of the peptide in the respective sequence. Underlined amino acid residues mark potential phosphorylation sites. Bold type residues show consensus of the NLS sequence or the cluster of acidic amino acid residues complementary to the NLS sequence.

particular sequence motif existed before the corresponding receptor had evolved. However, we can also not completely rule out the possibility that *T. acidophilum* has acquired the α-subunit-gene by horizontal gene transfer from an eukaryote. As with the eukaryotic proteasomal proteins investigated so far, we have been unable to identify any known proteinase motif in the sequence of the α-subunit of *T. acidophilum*.

By searching for homologous proteins in the database of the Martinsried Institute for Protein Sequences (MIPS) we found a significant sequence similarity to the wild-type suppressor gene *scl1*$^+$ encoded polypeptide of *S. cerevisiae* [25], as well as to the various proteasomal proteins from eukaryotes deposited in the database. Detailed sequence analysis revealed, that the nucleotide sequence of the *scl1*$^+$ gene is almost identical with the yeast proteasomal subunit YC7-α [26]. Insertion of one thymidine and one adenosine at position 129 and change of adenosine-228 to guanosine in the nucleotide sequence of the *scl1*$^+$ gene leads to absolute identity of the nucleotide sequences of the *scl1*$^+$ and the YC7-α genes, in the coding as well as in the 5'- and 3'-flanking regions. Independently it was shown by gene disruption experiments that the *scl1*$^+$ gene as well as the YC7-α gene are essential for growth [25,26]. The *SCL1*-1 mutation of the *scl1*$^+$ gene is a dominant suppressor of the cycloheximide-resistant, temperature-sensitive lethal mutaton, *crl3* [25]. This mutation-elicited suppressor function of one of the *S. cerevisiae* proteasomal proteins sheds new light on the possible involvement of the particle in the regulation of gene expression [27], repression or modulation of mRNA translation [28,29].

3.4. *Alignment of proteasomal proteins*

The predicted amino acid sequence of the *T. acidophilum* α-subunit shows significant sequence similarity with subunits from all the eukaryotic proteasomes investigated so far (Fig. 2). The similarity is pronounced in the N-terminal and central part of the proteins (up to position 166 in Fig. 2), where in some subunits the potential NLS targeting sequences and the putative phosphorylation sites are located, whereas the C-terminal part appears to be rather variable (Fig. 2). The conserved regions can be expected to be either critically involved in proteasome function, or — in view of the highly conserved structure of the complex in spite of a rather different subunit composition — in providing the regions of contact between the different subunits within the proteasome. Similarity scores provided by pairwise alignment of different proteasomal subunits (Table II), imply that the genes for the archaebacterial α-subunit and the hitherto cloned and sequenced eukaryotic proteasomal proteins belong to one gene family of ancient origin. Rat proteasome subunit C2 [30] and *D. melanogaster* 35 kDa subunit [21] (similarity score 135), rat subunit C8 [31] and yeast subunit YC1 [26] (similarity score 121), as well as rat subunit C9 [32] and *D. melanogaster* 29 kDa subunit [33] (similarity score 178) (see Table II), are likely to be corresponding subunits in the proteasomes of the respective organisms. The rat subunit C5 [34] shows the lowest similarity scores with any of the other sequences; nevertheless it appears to belong to the family of proteasomal proteins, as is shown by the alignment (Fig. 2). Remarkably the rat subunit C3 [23] yields the highest similarity score with the *T. acidophilum* α-subunit (Table II). This might be taken as an indication that several gene duplication events of the C3 encoding gene led to the complex subunit composition of eukaryotic proteasomes. These events can be expected to be consequences of more sophisticated functional requirements and functional fine-tuning enforced during evolution of the eukaryotic cell.

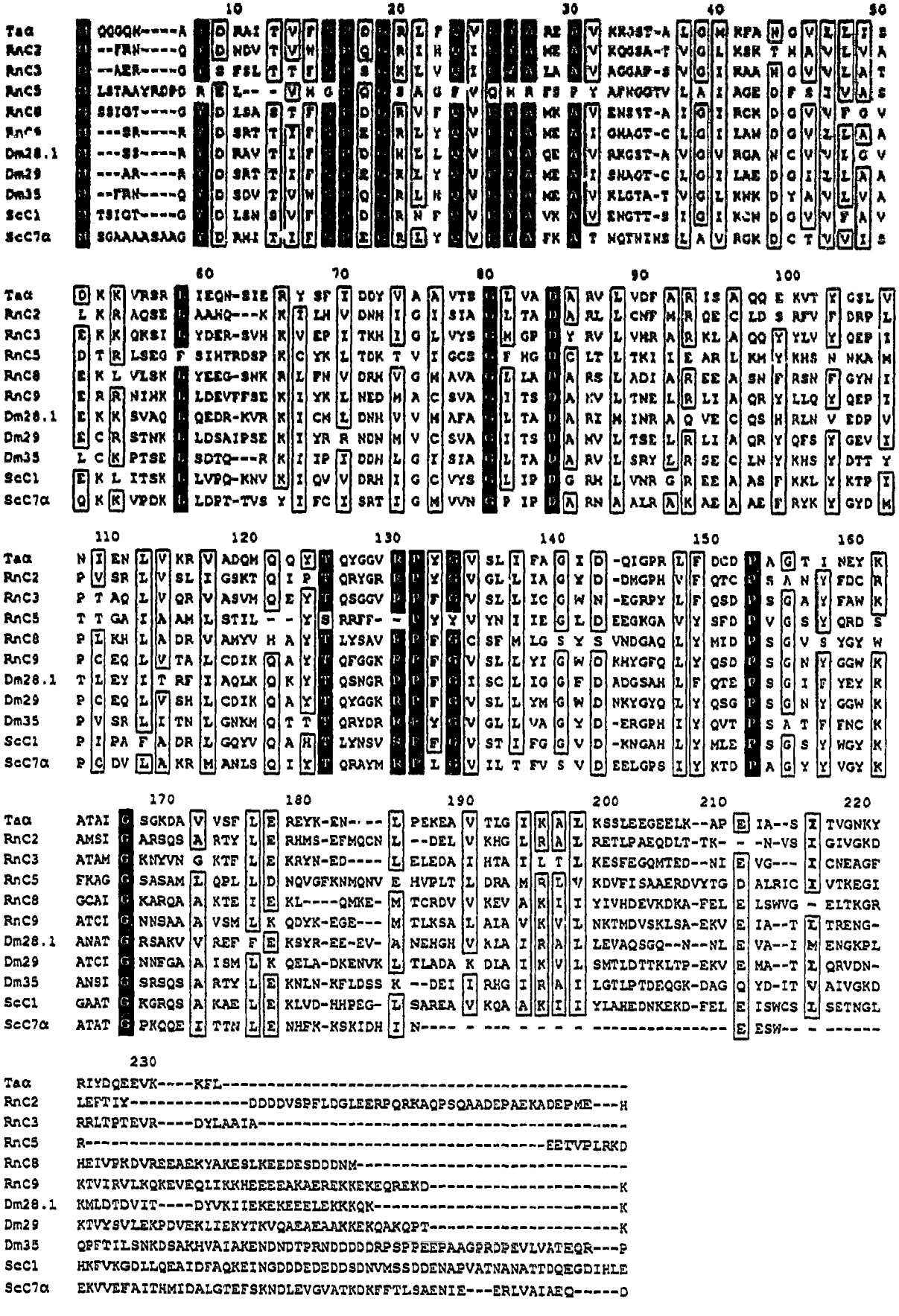

TABLE II

Numbers of identical amino acid residues, elucidated by pairwise comparison of the respective sequences

	Taα	RnC2	RnC3	RnC5	RnC8	RnC9	Dm28.1	Dm29	Dm35	ScC1	ScC7α
Taα											
RnC2	67										
RnC3	87	66									
RnC5	30	30	32								
RnC8	71	54	64	29							
RnC9	72	61	75	41	68						
Dm28.1	77	75	74	32	72	71					
Dm29	67	62	72	40	63	178	65				
Dm35	69	135	66	33	56	70	65	70			
ScC1	73	65	77	28	121	68	69	63	73		
ScC7α	65	53	62	39	68	59	59	58	62	68	

Sequence designations are as given in Fig. 2.

Acknowledgements The authors wish to thank W. Zillig (Martinsried) for providing *T. acidophilum* cells and H.-P. Klenk (Martinsried) for providing genomic DNA from *T. acidophilum*

REFERENCES

[1] Rivett, A.J. (1989) Arch. Biochem. Biophys. 268, 1-8
[2] Arrigo, A.P., Simon, M., Darlix, J.L. and Spahr, P.F. (1987) J. Mol. Evol. 25, 141-150
[3] Dahlmann, B., Kuehn, L., Ishiura, S., Tsukahara, T., Sugita, H., Tanaka, K., Rivett, J., Hough, R.F., Rechsteiner, M., Mykles, D.L., Fagan, J.M., Waxman, L., Ishii, S., Sasaki, M., Kloetzel, P., Harris, H., Ray, K., Behal, F.J., DeMartino, G.N. and McGuire, M.J. (1988) Biochem. J. 255, 750-751
[4] Arrigo, A.P., Tanaka, K., Goldberg, A.L. and Welch, W.J. (1988) Nature 331, 192-194
[5] Falkenburg, P.-E., Haass, C., Kloetzel, P.-M., Niedel, B., Kopp, F., Kuehn, L. and Dahlmann, B. (1988) Nature 331, 190-192
[6] Zwickl, P., Pfeifer, G., Lottspeich, F., Kopp, F., Dahlmann, B. and Baumeister, W. (1990) J. Struct. Biol. 103, in press
[7] Dahlmann, B., Kopp, F., Kuehn, L., Niedel, B., Pfeifer, G., Hegerl, R. and Baumeister, W. (1989) FEBS Lett. 251, 125-131
[8] Baumeister, W., Dahlmann, B., Hegerl, R., Kopp, F., Kuehn, L. and Pfeifer, G. (1988) FEBS Lett. 241, 239-245
[9] Schaegger, H. and von Jagow, G. (1987) Anal. Biochem. 166, 368-379
[10] Eckerskorn, C., Mewes, W., Goretzki, H. and Lottspeich, F. (1988) Eur. J. Biochem. 176, 509-519
[11] Eckerskorn, C. and Lottspeich, F. (1989) Chromatographia 28, 92-94.
[12] Zwickl, P., Fabry, S., Bogedain, C., Haas, A. and Hensel, R. (1990) J. Bacteriol. 172, 4329-4338
[13] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237-244
[14] Bairoch, A. (1990) PROSITE a Dictionary of Protein Sites and Patterns, Fifth Release, University of Geneva
[15] Ree, H.K., Cao, K., Thurlow, D.L. and Zimmermann, R.A. (1989) Can. J. Microbiol. 35, 124-133.
[16] Fremisco, J.R., Glass, D.B. and Krebs, E.G. (1980) J. Biol. Chem. 255, 4240-4245.
[17] Glass, D.B. and Smith, S.B. (1983) J. Biol. Chem., 1497-1503
[18] Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) Science 241, 42-52.
[19] Roberts, B. (1989) Biochim. Biophys. Acta 1008, 263-280.
[20] Yoneda, Y., Imamoto-Sonobe, N., Matsuoka, Y., Iwamoto, R., Kiho, Y. and Uchida, T. (1988) Science 242, 275-278.
[21] Haas, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloetzel, P.M. (1989) EMBO J. 8, 2373-2379
[22] Haas, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloetzel, P.-M. (1990) Gene 90, 235-241
[23] 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.
[24] Tanaka, K., Yoshimura, T., Tamura, T., Fujiwara, T., Kumatori, A. and Ichihara, A. (1990) FEBS Lett. 271, 41-46
[25] Balzi, E., Chen, W., Capieaux, E., McCusker, J.H., Haber, J.E. and Goffeau, A. (1989) Gene 83, 271-279.
[26] 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
[27] Scherrer, K. (1990) Mol. Biol. Rep. 14, 1-9
[28] Schmid, H.P., Akhayat, O., Martins de Sa, C., Puvion, F., Köhler, K. and Scherrer, K. (1984) EMBO J. 3, 29-34
[29] Kuehn, L., Dahlmann, B. and Kopp, F. (1990) FEBS Lett. 261, 274-278
[30] 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
[31] Tanaka, K., Kanayama, H., Tamura, T., Lee, D.H., Kumatori, A., Fujiwara, T., Ichihara, A., Tokunaga, F., Aruga, R. and Iwanaga, S. (1990) Biochem. Biophys. Res. Commun. 171, 676-683.
[32] Kumatori, A., Tanaka, K., Tamura, T., Fujiwara, T., Ichihara, A., Tokunaga, F., Onikura, A. and Iwanaga, S. (1990) FEBS Lett. 264, 279-282
[33] Haas, C., Pesold-Hurt, B. and Kloetzel, P.-M. (1990) Nucleic Acids Res. 18, 4018
[34] 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.
[35] Argos, P. (1987) J. Mol. Biol. 193, 385-396

Fig. 2. Alignment of proteasomal proteins. Amino acid residues are given in the one letter code. Dashes indicate gaps, introduced for better alignment. Numbers above the sequences refer to the position of the respective amino acid residue in the *T. acidophilum* α-subunit sequence. Residues are boxed in black if they are identical in all or in all but one sequence at the respective position. Residues are shaded if they are conserved in at least 9 sequences at the respective position, according to the following grouping of Argos [35]. (PG), (ST), (KR), (EQND), (FWHY) and (AIVLMC). The abbreviations are as follows: Taα, *T. acidophilum* α-subunit; RnC, *Rattus norvegicus* (rat) proteasome component, Dm, *Drosophila melanogaster* proteasomal protein, ScC, *Saccharomyces cerevisiae* (yeast) proteasome component.