

# Cloning and sequencing of the gene encoding the large ( $\alpha$ -) subunit of the proteasome from *Thermoplasma acidophilum*

Peter Zwickl<sup>1</sup>, Friedrich Lottspeich<sup>1</sup>, Burkhardt Dahlmann<sup>2</sup> and Wolfgang Baumeister<sup>1</sup>

<sup>1</sup>Max-Planck-Institut für Biochemie, D-8033 Martinsried and <sup>2</sup>Diabetes-Forschungsinstitut an der Universität Düsseldorf, D-4000 Düsseldorf, Germany

Received 29 November 1990

The gene encoding the  $\alpha$ -subunit of the proteasome from the archaeobacterium *Thermoplasma acidophilum* was cloned and sequenced. The gene encodes for a polypeptide with 233 amino acid residues and a calculated molecular weight of 25.870. Sequence similarity of the  $\alpha$ -subunit with the *Saccharomyces cerevisiae* wild-type suppressor gene *sc11*<sup>+</sup> encoded polypeptide, which is probably identical with the subunit YC7 a 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, Archaeobacterium, *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 archaeobacterium *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  $\alpha$ -subunit has an apparent  $M_r$  of 27 kDa, the smaller  $\beta$ -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 ( $\alpha$ -) 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.

## 2. MATERIALS AND METHODS

Proteasomes from *T. acidophilum* were purified to homogeneity as reported previously [7]. 25  $\mu$ g were separated by tricine-SDS-PAGE [9] and detected by Coomassie blue staining. For cleavage with cyanogen bromide the  $\alpha$ -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, Göttingen, Germany), detected by Coomassie blue staining, excised, and subjected to sequence degradation [10]. For cleavage with trypsin, gel-slices containing the  $\alpha$ -subunit were cut into small pieces and suspended in 0.1 M ammonium carbonate. The protein was digested with 5  $\mu$ g trypsin at room temperature for 24 h. Tryptic peptides were extracted from gel pieces, concentrated and separated by HPLC. The column (Hibar LiChrospher-C<sub>18</sub> column, 125  $\times$  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*  $\alpha$ -subunit were performed with 'CLUSTAL' [13], a software package for multiple sequence align-

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

Abbreviations: *cr1*, mutation encoding cycloheximide and temperature sensitive lethality; HPLC, high-performance liquid chromatography; *sc11*<sup>+</sup>, wildtype allele of *sc11*; *SCL1*-1, mutant allele of *sc11*, dominant suppressor of *cr13* temperature-sensitive lethality; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid; YC, yeast component of the proteasome.



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
Tat	112-LVKRVADQMQQY	TQ-YGG-VRPY-132	32-KKGGT-36	50-SDKKVR-55
RnC2	nf	nf	nf	202-EKQNELKAPD-211
RnC3	110-LVQRVASVMQE	TQ-SGG-VRPY-130	48-TEKKQK-53	218-DLEFTIYDDDD-228
RnC8	nf	nf	nf	246-EKDESDDD-253
RnC9	110-LVTALCDIKQA	TQ-FGG-KRPF-130	249-REKKK-254	241-EKKAKAERE-250
Dm28.1	nf	29-RKGGT-33	243-EKKKQK-248	nf
Dm29	110-LVSHLCDIKQA	TQ-YGG-KRPF-130	252-AAKKK-257	250-DDDEDEDSD-259
ScC1	nf	nf	nf	nf
ScC7a	117-LAKRMANLSQI	YTQR-AAYMRPL-137	185-HPKKSK-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  $\alpha$ -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  $\alpha$ -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 *sc11<sup>+</sup>* 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 *sc11<sup>+</sup>* gene is almost identical with the yeast proteasomal subunit YC7- $\alpha$  [26]. Insertion of one thymidine and one adenosine at position 129 and change of adenosine-228 to guanosine in the nucleotide sequence of the *sc11<sup>+</sup>* gene leads to absolute identity of the nucleotide sequences of the *sc11<sup>+</sup>* and the YC7- $\alpha$  genes, in the coding as well as in the 5'- and 3'-flanking regions. Independently it was shown by gene disruption experiments that the *sc11<sup>+</sup>* gene as well as the YC7- $\alpha$  gene are essential for growth [25,26]. The *SCL1-1* mutation of the *sc11<sup>+</sup>* gene is a dominant suppressor of the cycloheximide-resistant, temperature-sensitive lethal mutation, *cr13* [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*  $\alpha$ -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 archaeobacterial  $\alpha$ -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*  $\alpha$ -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.

	10	20	30	40	50																	
Taa	QOOQH----A	DRAITVF	QRLP	V	REKV	KRST-A	LGM	KPA	HGV	LV	LS											
RnC2	---FRN----Q	DNDVTTVW	QRLP	V	MEKV	KQSA-T	VGL	KSK	THAV	LV	LA											
RnC3	---AER----Q	SFELTTT	SKLV	Q	LA	AGGAP-S	VGI	KAA	HGV	LV	LA											
RnC5	LSTAAVROPD	RLE--VHG	Q	SAGF	VQHR	FSFY	AFHOOTV	LA	AGE	DF	SV	VA	AS									
RnC8	SSIGT----Q	DLSASTFF	D	RVF	V	MEKV	ENST-A	I	RCM	DGV	V	FO	V									
RnC9	---SR----R	DRTSTTF	E	RLY	V	MEKV	OMACT-C	L	GI	LAM	DGV	LV	LA	A								
Dm28.1	---SS----R	DRAVTTF	D	HLL	V	MEKV	AKGST-A	V	GI	RGA	NCV	LV	LA	A								
Dm29	---AR----R	DRTSTTF	E	RLY	V	MEKV	SHAOT-C	L	GI	LAE	DGV	LV	LA	A								
Dm35	---FRN----Q	DSDVTTVW	QRLP	V	MEKV	KLSTA-T	VGL	KHK	DY	AV	LV	LA	A									
ScC1	TSIGT----G	DLSNSTVF	D	RNF	V	MEKV	ENGT-S	I	GI	KCN	DGV	LV	LA	A								
ScC7a	SGAAASAG	D	RHI	T	IF	E	R	LY	V	FK	A	T	NOTWINS	L	AV	RCK	D	CT	V	V	I	S

	60	70	80	90	100																	
Taa	D K K VRSR	I EQN-SIE	R Y SF	I DOY	V A VTS	L VA	D A	RV	L VDF	A R	IS	A	QQ	E	KVT	Y	GSL	V				
RnC2	L K R AQSE	I AAHQ---K	R T LH	V DNH	I G I	STA	L TA	D A	RL	L CNF	M R	QE	C	LD	S	RFV	F	DRP	L			
RnC3	E K K QKSI	I YDER-SVH	R V EP	I TKH	I G L	VYS	M GP	D Y	RV	L VHR	A R	KL	A	QQ	Y	YLV	Y	QEP	I			
RnC5	D T R LSEQ	F SIHTRDSP	R C YK	L TDK	T V I	GCS	F HG	D C	LT	L TKI	I E	AR	L	KM	Y	KHS	N	NKA	M			
RnC8	E K L VLSK	I YEEG-SHK	R L FN	V DRH	V G M	AVA	L LA	D A	RS	L ADI	A R	EE	A	SN	F	RSH	F	GYN	I			
RnC9	E R R NIHK	I LDEVFSE	R I YK	L NED	M A C	SVA	I TS	D A	MV	L TNE	L R	LI	A	QR	Y	LLQ	Y	QEP	I			
Dm28.1	E K K SVAQ	I QEDR-KVR	K I CM	L DNH	V V M	AFA	L TA	D A	RI	M INR	A Q	VE	C	QS	H	RLN	V	EDP	V			
Dm29	E C R STNK	I LDSAIPSE	K I YR	R NDH	M V C	SVA	I TS	D A	MV	L TSE	L R	LI	A	QR	Y	QFS	Y	GEV	I			
Dm35	L C K PTSE	I SDTQ---R	K I IP	I DDH	L G I	STA	L TA	D A	RV	L SRY	L R	SE	C	LN	Y	KHS	Y	DTT	Y			
ScC1	E K L ITSX	I LVPO-KNV	K I QV	V DRH	I G C	VYS	L IP	D G	RM	L VNR	G R	EE	A	AS	F	KKL	Y	KTP	I			
ScC7a	Q K K VPDK	I LDPT-TVS	Y I FC	I SRT	I G M	VVN	G P	IP	D A	RN	A	ALR	A	K	AE	A	AE	F	RYK	Y	GYD	M

	110	120	130	140	150	160																
Taa	N I EN	L V KR	V ADQM	Q Q Y T	QYGGV	R P Y	V	SL	I	FA	G	I	D	-QIGPR	L F	DCD	V A	G	T	I	NEY	K
RnC2	P V SR	L V SL	I GSKT	Q I P T	QRYGR	R P Y	V	GL	L	IA	G	Y	D	-DMGPH	V F	QTC	D S	A	N	Y	FDC	R
RnC3	P T AQ	L V QR	V ASVM	Q E Y T	QSGGV	R P Y	V	GL	L	IC	G	W	N	-EGRPY	L F	QSD	D S	G	A	Y	FAW	K
RnC5	T T GA	I A AM	L STIL	- Y S	RRFF	- U Y Y	V	YN	I	IE	G	L	D	EKGKA	V Y	SFD	D V	G	S	Y	QRD	S
RnC8	P L KH	L A DR	V AMYV	H A Y T	LYSAV	R P Y	V	SL	L	YI	G	W	D	KHYGFQ	L Y	QSD	D S	G	N	Y	GGW	K
RnC9	P C EQ	L V TA	L CDIK	Q A Y T	QFGGK	R P Y	V	SL	L	YI	G	W	D	KHYGFQ	L Y	QSD	D S	G	N	Y	GGW	K
Dm28.1	T L EY	I T RF	I AQLK	Q K Y T	QSNGR	R P Y	V	SL	L	IG	G	F	D	ADGSAH	L F	QTE	D S	G	I	F	YFY	R
Dm29	P C EQ	L V SH	L CDIK	Q A Y T	QYGGK	R P Y	V	SL	L	YI	G	W	D	KHYGFQ	L Y	QSD	D S	G	N	Y	GGW	K
Dm35	P V SR	L I TN	L GNKM	Q T T T	QRYDR	R P Y	V	GL	L	VA	G	Y	D	-ERGPY	I Y	QVT	D S	A	T	F	FNC	K
ScC1	P I PA	F A DR	L QQYV	Q A H T	LYNSV	R P Y	V	ST	I	FG	G	V	D	-KNGAH	L Y	MLE	D S	G	S	Y	HGY	K
ScC7a	P C DV	L A KR	M ANLS	Q I Y T	QRAYM	R P Y	V	IL	T	FV	S	V	D	EELGPS	I Y	KTD	D A	G	Y	Y	VGY	K

	170	180	190	200	210	220									
Taa	ATAI	G SGKDA	V VSF	L E	REYK-EN---	L PEKEA	V TLG	I K A	L	KSSLEEGEELK--AP	E	IA--S	I	TVGNKY	
RnC2	AMSI	G ARSQS	A RTY	L E	RHMS-EFMQCN	L --DEL	V KHG	L R A	L	RETLPAEQDLT-TK-	-	-N-VS	I	GIVGKD	
RnC3	ATAM	G KNYVN	G KTF	L E	KRYN-ED---	L ELED	I HTA	I L T	L	KESFEGQMTED--NI	E	VG---	I	CNEAGF	
RnC5	FKAG	G SASAM	L QPL	L D	NQVGFKNQNV	E HVFLT	L DRA	M R L	V	KDVFISAAERDVYTG	D	ALRIC	I	VTKEGI	
RnC8	GCAI	G KARQA	A KTE	I E	KL----	QMKE	M TCRDV	V KEV	A K I	I	YIVHDEVKDKA-FEL	E	LSWVG	-	ELTKGR
RnC9	ATCI	G NNSAA	A VSM	L K	QDYK-EGE---	M TLKSA	L AIA	V K V	L	NKTMDSKLSA-EKV	E	IA--T	L	TRENG-	
Dm28.1	ANAT	G RSAKV	V REF	F E	KSYR-EE-EV-	A NEHGH	V KLA	I R A	L	LEVAQSGQ--N--NL	E	VA--I	M	ENGKPL	
Dm29	ATCI	G NNFGA	A ISM	L K	QELA-DKENVK	L TLADA	K DLA	I K V	L	SMTLDTKLTP-EKV	E	MA--T	L	QRVDN-	
Dm35	ANSI	G SRSQS	A RTY	L E	KNLN-KFLDSS	K --DEI	I RHG	I R A	I	LGTLPDEQGK-DAG	Q	YD-IT	V	AIVGKD	
ScC1	GAAT	G KGRQS	A KAE	L E	KLVD-HHPEG-	L SAREA	V KQA	A K I	I	YLAHEDNKEKD-FEL	E	ISWCS	L	SETNGL	
ScC7a	ATAT	G PKQGE	I TTH	L E	NHFK-KSKIDH	I N----	-	-	-	-	-	E	ESW--	-	----

230

Taa	RIYDQEEVK----	KFL	-----
RnC2	LEFTIY-----	DDDDVSFFLDGLEERPQKAQPSQAADPAEKADPEME---	H
RnC3	RLTPTTEVR----	DYLAATA	-----
RnC5	R-----	-----	EETVPLRKD
RnC8	HEIVPKDVREEAEKYAKESLKEEDESDDDNM-----	-----	-----
RnC9	KTVIRVLKQKEVEQLIKKHEEEAKAEREKKEKEQREKD-----	K	-----
Dm28.1	KMLDQDVIT----	DYVKIIEKEKEELEKKKQK-----	K
Dm29	KTVYSVLEKPDVEKLIERYTKVQAEAEAKKEKQAKQPT-----	K	-----
Dm35	QFFTILSNKDSAKHVIAKENDNDTPRNDDDDDRPSFPEEPAAGPRDFEVLVATEQR---	P	-----
ScC1	KHFVKGOLLQEAIDFAQKEINGDDDEDDSDNVMSDDENAPVATNANATTDQEGDIHLE	-----	-----
ScC7a	EKVVEFAITHMIDALGTETFSKNDLEVGVATKOKFFTLSAENIE---	ERLVAIAEQ----	D

TABLE II

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

	Taa	RnC2	RnC3	RnC5	RnC8	RnC9	Dm28.1	Dm29	Dm35	ScC1	ScC7a
Taa											
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		
ScC7a	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 (Martinried) for providing *T. acidophilum* cells and H.-P. Klenk (Martinried) 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., Kloeitzel, 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., Kloeitzel, 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] Schaeffer, H. and von Jagow, G. (198) 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] Barroch, 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. 258, 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., Kito, Y. and Uchida, T. (1988) Science 242, 275-278.
- [21] Haas, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloeitzel, P.-M. (1989) EMBO J. 8, 2373-2379.
- [22] Haas, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloeitzel, 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., Kohler, 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 Kloeitzel, 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*  $\alpha$ -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: Taa, *T. acidophilum*  $\alpha$ -subunit; RnC, *Rattus norvegicus* (rat) proteasome component, Dm, *Drosophila melanogaster* proteasomal protein, ScC, *Saccharomyces cerevisiae* (yeast) proteasome component.