

## Cloning and Characterization of cDNAs Encoding Trehalase from Post-Dormant Embryos of the Brine Shrimp, Artemia franciscana

Authors: Tanaka, Shin, Nambu, Fumiko, and Nambu, Ziro

Source: Zoological Science, 16(2): 269-277

Published By: Zoological Society of Japan

URL: https://doi.org/10.2108/zsj.16.269

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at <u>www.bioone.org/terms-of-use</u>.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

### Cloning and Characterization of cDNAs Encoding Trehalase from Post-Dormant Embryos of the Brine Shrimp, *Artemia franciscana*

Shin Tanaka\*, Fumiko Nambu and Ziro Nambu

Biology, Department of Medical Technology, School of Health Sciences, University of Occupational and Environmental Health, Japan, Yahatanishi-ku, Kitakyushu 807-8555, Japan

**ABSTRACT**—To investigate post-dormant regulation of trehalose metabolism in the brine shrimp, we cloned and characterized two trehalase cDNAs from embryos of *Artemia franciscana* using a PCR probe corresponding to a highly conserved region among trehalases. The cDNAs consisted of 2496 and 2485 nucleotides, and had almost the same open reading frame encoding 703 amino acids which showed 46.6–42.6% similarities to trehalases of animals. The calculated molecular mass of the trehalase was 79,995 Da. The deduced sequence had a cleavable signal peptide, a cell adhesion motif, four potential N-glycosylation sites, trehalase signatures and a unique, long carboxyl terminal polypeptide containing a predicted transmembrane region and a potential cAMP-dependent phosphorylation site. Phylogenetic analyses showed a large divergence among trehalases of arthropods. Northern blots revealed the presence of two mRNAs. One of them, a 2.6 kb mRNA, was abundant in the dormant cysts and prenauplii. The other 5.0 kb mRNA was newly synthesized during post-dormant development. Possible mechanisms of trehalase regulation are presented on the basis of the results shown by the Northern blots and developmental changes of trehalase activity.

#### INTRODUCTION

Various organisms accumulate  $\alpha$ ,  $\alpha$ -trehalose, a non-reducing disaccharide consisting of two glucoses, against physiological and environmental changes. The accumulated trehalose has a role as an anti-stress molecule. For instance, in yeast, such anti-stress effects of trehalose against dehydration, heat shock, freezing and ethanol have been reported (reviewed by Nwaka and Holzer, 1998). These anti-stress effects of trehalose are thought to be due to its protective function on the cell membrane or proteins (Crowe *et al.*, 1987, 1988).

Embryos of *Artemia* entering dormancy also accumulate  $\alpha$ ,  $\alpha$ -trehalose. This accumulation is not observed in nondormant embryos (Clegg, 1962, 1965). The amount of trehalose in the encysted dormant gastrula represents 15% of its dry weight (Clegg, 1962, 1965). Under appropriate circumstances, the encysted embryos resume their subsequent development and the trehalose is hydrolyzed by a trehalase (EC 3.2.1.28).

In our previous study (Nambu *et al.*, 1997a), a soluble, alkaline trehalase with a molecular mass of 70 kDa on SDSpolyacrylamide gel electrophoresis was purified from developing embryos of Artemia. The Artemia trehalase showed a significant increase in its activity at the prenauplius stage and the activity decreased after hatching (Nambu et al., 1997a). Hydrolysis of trehalose provides glucose to produce glycogen and glycerol, which are respiratory substrates in the Artemia cysts (Clegg, 1964; Ewing and Clegg, 1969). It is also suggested that accumulated glycerol increases the osmotic pressure in the cysts, resulting in a rupture of the shell followed by the emergence of prenauplii surrounded by hatching membrane (Clegg, 1964). Therefore, the temporal regulation of the trehalase activity is significant in post-dormant development. To investigate the mechanisms of trehalase regulation, we cloned a 228 bp cDNA fragment of the Artemia trehalase using a polymerase chain reaction (PCR) (Nambu et al., 1997b). In the present report, we describe the cloning and characterization of Artemia trehalase cDNAs encoding a full open reading frame (ORF) and show the results of Northern blots and phylogenetic analyses.

#### MATERIALS AND METHODS

#### Animals

Dried cysts of *Artemia franciscana* from the Great Salt Lake in Utah, USA were obtained from Japan Pet Drugs Co. (Tokyo and Los Angeles, CA, USA). The cysts were hydrated and cultured as described by Nambu *et al.* (1997a).

<sup>\*</sup> Corresponding author: Tel. +81-93-603-1611(Ext. 4539); FAX. +81-93-691-7142. E-mail. artemia@med.uoeh-u.ac.jp

#### **RNA** extraction

Total RNA was extracted from 0.5 g of the wet weight of the embryos and nauplii according to the method described by Maniatis *et al.* (1982) with some modifications. Each specimen was homogenized in 2.5 ml of a guanidine thiocyanate solution. The homogenate was centrifuged at 4,000 rpm for 5 min and total RNA was extracted from the supernatant. Poly (A)<sup>+</sup> RNA was purified using Oligotex dT30 <Super> (Roche Japan, Tokyo, Japan).

#### cDNA cloning and nucleotide sequencing

Double stranded cDNA was synthesized from purified poly (A)<sup>+</sup> RNA extracted from 12 hr cultured prenauplii of *A. franciscana* using oligo(dT)<sub>12-18</sub> primer. The double stranded cDNA was ligated with *Eco*RI-*Not*I adaptor and a cDNA library was constructed in  $\lambda$  gt10 vector (Stratagene) using gigapack®III Gold (Stratagene). The 228 bp cDNA fragment of the *Artemia* trehalase was obtained by the degenerated PCR and digoxigenin (DIG)-labelled as described by Nambu *et al.* (1997b). It was used to screen  $2.0 \times 10^5$  clones. Phage plaques were lifted up and immobilized on nylon membranes (Boehringer Mannheim, Mannheim, Germany). Hybridization and detection was performed as described by Engler-Blum *et al.* (1993). Positive clones were isolated and two longer inserts termed *Tre*E2 and *Tre*E3 were fully sequenced using Amersham thermo sequenase fluorescent labelled primer cycle sequence kit (Amersham, Buckinghamshire, UK) and an ABI 373 sequencer (Perkin Elmer, Norwalk, CT, USA).

#### Northern blot analysis

The full length of *Tre*E3 was DIG-labelled according to the random priming method using a DIG DNA Labelling Kit (Boehringer Mannheim). Total RNA was prepared as described in the section of RNA extraction, and 20  $\mu$ g of each total RNA was electrophoresed in a 1.0% agarose gel containing formaldehyde (Sambrook *et al.*, 1989) and blotted on to Hybond N<sup>+</sup> membrane (Amersham). Hybridization was performed at 68°C as described by Engler-Blum *et al.* (1993) followed by washing at 65°C.

#### RESULTS

#### Isolation of trehalase cDNAs from A. franciscana

Screening of the cDNA library derived from the 12 hr cultured embryos of *A. franciscana* with the PCR product resulted in isolation of 12 positive clones. Two longer inserts, *Tre*E2 and *Tre*E3, were used for full-length nucleotide sequencing and further investigation.

#### Comparison of TreE2 with TreE3

*Tre*E2 consisted of 2496 nucleotides while *Tre*E3 contained 2485 nucleotides. These had a common region consisting of 2476 nucleotides (Fig. 1). Identity between the two cDNAs was 99.7%. Seven replacements were found at the nucleotide level (Fig. 1). Both cDNAs contained the same ORF followed by a termination codon and an "AGTAAA" sequence which was similar to the polyadenylation signal, "AATAAA". An oligo (A) sequence was found 14 nucleotides downstream from the "AGTAAA" sequence in *Tre*E2 (Fig. 1, 2). The pro-

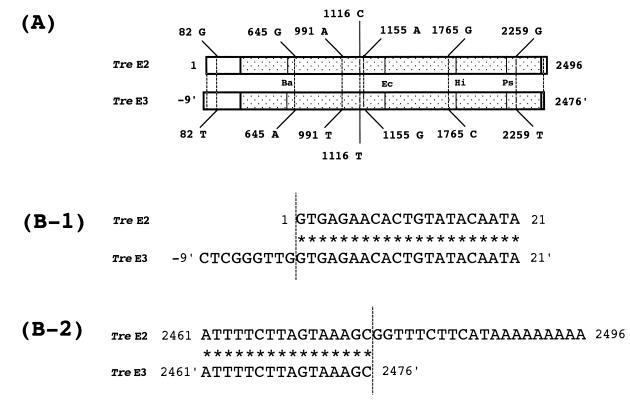


Fig. 1. Profiles of the *Artemia* trehalase cDNAs, *Tre*E2 and *Tre*E3. Nucleotides of *Tre*E2 were numbered from the first nucleotide. Nucleotide numbers of *Tre*E3 were matched to the corresponding nucleotides of *Tre*E2. The sequence identity between the two cDNAs was 99.7%. (A) Varied nucleotides are indicated by their number. Dotted areas indicate ORF regions. Both cDNAs had identical restriction sites. Ba, *Bam*HI; Ec, *Eco*RI; Hi, *Hind*III; Ps, *Pst*I. (B-1) 5'-Termini of both the cDNAs are represented. (B-2) 3'-Termini of *Tre*E3 are aligned. *Tre*E2 had a oligo(A) sequence on its 3'-terminus. Asterisks indicate identical nucleotides between the two cDNAs.

AGGGA & & AGGMGGA & A	ATGAACAT															
AGGCAAACGTGCAAG	TAAAAAG	TGAACT	TTTGTT.	AGTGA	CTCTT	GAAAAG	AATAA	AAGAI	CATTT?	FATCA	CTAA					
	•			* *									C		ľ	!
AGTTTTTTTCGTCGT							•									
		F A 1				G D						~				3
IGTGCTAGCGAAAT1	TACTGCC	ATGGAA	CATTCC	IGCAT	ACAAT	ACAGAT	GGCAG	GGCTO	TTTC	GAGAC	TCGA	AACC	TTT	GTAG	ATAAA	450
CASEI	Y C	нсл	C F I	Н	ΤI	QM	AG	3 L	FR	D	s K	т	F	VI	К	6
AAGCTAAAGATAAA	CCAGAAG	AAGTTT	TAGCTA	GCTTT	GAAGT	TTTGAT	GAACT	CAACO	GATC	AAAA	CCCAG	CCGG	GAC	CAAI	TGGCI	540
K L K I N	ΡE	EVI	L A S	5 F	EV	LM	N S	5 Т	DQ	N	P S	R	D	QI	A	9
CTTTTATAAATCT	ACACTTCG	AGCCAG	AAGGGT	CAGAA	<b>FTTGA</b>	AGAATG	GGATC	CAATA	GATT	GCAI	TCTA	CCCI	AGT	TTTT	TGGAT	630
A F I N L	H F	ЕРЕ	GG	5 E	FE	ΕW	DF	? I	D W	н	s N	Р	s	FI	D	12
GTATTAGAGACACO	GAACTTGA	AAATAT	GGGGAA	ATTCT	TCCA	CGAAGC	CTGGA	CATGO	TTAG	GTCG <b>A</b>	AAAA	recgi	GAC	GACG	TGAGA	A 720
G I R D T	N L	K I V	V G 1	N S	L H	ΕA	r w	r w	L G	R	K I	R	D	DV	7 R	15
ATAAATCCCCAATTA	ATATTCGA	TGATGT	ATCTCC	CCAACO	CTTT	CATCAT	TCCTG	GTGGI	CGTT	rccgi	GAAAO	TTAC	TAC	IGGG	ATTCO	810
INPQL	Y S	мму	( L I	PN	P F	ΙI	[k] [	5/ K/	k/x	71	1/1	747	4/	w / 1	{/\$	18
TATTGGATTATCAA	AGGGCTTC	TTATCT	CGGGAA	TGCAT	GAGAC	GGTGAA	GGGTA	TGTTO	TTGA	ACTTI	TTGC	GATO	GTG	GATA	CCATI	900
<b>y</b> w i i k	G L	LIS	5 G 1	4 H	Е Т	V K	G M	I L	LN	F	L L	м	v	D	r I	21
 GCCTTGTACCCAA	IGGTGGAA	GGATAT	ATTATG.	AGAAA	CGATC	TCAGCC	TCCAC	TTCT	TACTCO	CAATO	GTGG	ACT	TAC	GTTA	ATGCI	r 990
GLVPN	GG	RIY	<i>.</i> Y 1	с к	R S	QР	ΡI	гг	ТР	м	V E	L	Y	vi	A	24
ACGGGAGACATTGA	TTTTTAA	AGCAAA	ACATTC	ACCTG	TGGA	AAAAGA	GATGG	ATTTI	TGGC	TCAG	GAGCG	AACG	GTC	AACG	TTGAT	1080
T G D I E	FL	кон	N I H	чL	LE	ΚE	ME	) F	W L	Q	ER	т	v	N V	7 D	27
GGCACCGGCTTAT	CGTTACG	ATGTTA	AAGTTG	CGGT	CCCG	ACCTGA	ATCTT	ACAAG	GAAG	TATA	GAAGO	AATT	CATO	CATA	ACCCT	1170
GHRLI	RУ	DVB	< v (	3 G	PR	ΡE	s y	Y K	EC	) I	ЕА	I	н	ны	I P	30
ATTTAGAAGCACA	ATTGGATT	TCTATA	TGAACA	TAGCAG	CTGG	TGCTGA	AACAG	GATGG	GATT	стса	AGCCG	ATGG	TAT	IGGA	ACGGG	1260
DLEAQ	LD	FYN	4 N 3	E A	A G	A E	то	w e	DF	s	S R	W	Y	w	I G	33
GATATACAAACTAA	TTGTCAC	ATGTTC	GAACCC	GTGAC!	ATACT	ACCTGT	AGACC	TGAAT	TCTT	TATT	GCTTG	GGAC	TTT	JATA	TTATG	1350
DIOTN																
υτΥΥΝ	LS	нин	R T B	R D	ΙL	РV	DI	LN	SF	' I	A W	D	F	DJ	с м	36
ICCCGCTTTGAAAA	CAGTTAG		ACAACG	CATCAC	TTGT	CTATTC		TATAC		ATGG	AAAAC	GTCG		AATG	CCATT	36 1440 39
ICCCGCTTTGAAAA	CAGTTAG Q L	GTCGAG G R I	ACAACG	CATCAG A S	TTGT	CTATTC Y S	TGATT D I	TATAC L Y	AGTGA S E	AATGG W	AAAAC K T	GTCG S	ATTZ I	AATG	CCATT	1440 39
ICCCGCTTTGAAAAG S R F E K ITGTGGGATGATGAA	CAGTTAG Q L AGCTGGTA	GTCGAG G R I	ACAACG N 2 TTGATT	CATCAG A S ACGATI	STTGT VV VCTGC	CTATTC Y S	TGATT: D I GAGAT(	TATAC L Y GGAAC	AGTGI S E ACGAI	ATGG W ACTTI	AAAAC K T	GTCG S TTCT	ATTZ I	AATG N A	CCATT A I CTCCC	1440 39
ICCCGCTTTGAAAAC S R F E K ITGTGGGATGATGAI L W D D E	SCAGTTAG Q L AGCTGGTA A G	GTCGAGA G R I AGCTGGT S W I	ACAACG N TTGATT	CATCAG A S ACGATI Y D	TTGT VV V CTGC SA	CTATTC Y S GCATCG H R	TGATT D I GAGATO R V	TATAC L Y GGAAC N N	CAGTGA S E CACGAA T N	AATGG W ACTTI	K T TATGI Y V	GTCG S TTTCT S	ATTA I AATT N	AATG N A FTGA L 1	CCATT A I CTCCC F P	1440 39 1530 42
ICCCGCTTTGAAAAC S R F E K ITGTGGGATGATGAT L W D D E ITATTCGTAGGATG	SCAGTTAG Q L AGCTGGTA A G	GTCGAGA G R I GCTGGT S W I CGAAAAQ	ACAACG DN TTGATT FD CAGTTC	CATCAG A S ACGAT Y D ATCATO	STTGT VVV FCTGC SA SAAGA	CTATTC Y S GCATCG H R CGTTGC	TGATT D I GAGATO R V CACCAO	TATAC L Y GGAAC N N GAGTA	CAGTGA S E CACGAA T N ACTCGA	ATGG W ACTTI F ATTAC	K T TATGI Y V	GTCG S TTCT S AAAAA	I I N N AGT/	AATG N A FTGA L 1 AACG	CCATT A I CTCCC F P	1440 39 1530 42
ICCCGCTTTGAAAA S R F E K ITGTGGGATGATGAJ L W D D E ITATTCGTAGGATG L F V G C	CAGTTAG Q L AGCTGGTA A G ITACGATC Y D	GTCGAG G R I GCTGGT S W I CGAAAA P K 1	ACAACG N TTGATT F D CAGTTC F V I	CATCAG A S ACGATT Y D ATCATC H H	STTGT V V ICTGC S A SAAGA E D	CTATTC Y S GCATCG H R CGTTGC V A	TGATT D I GAGATO R V CACCAO T F	TATAC L Y GGAAC W N GAGTA R V	CAGTGA S E CACGAA T N ACTCGA L E	AATGG W ACTTI F ATTAC Y	K T K T TATGI Y V CTTGA L E	GTCG S CTTCT S AAAAA K	I N AGT/ S	AATG N A FTGA L 1 AACG N A	CCATT A I CTCCC F P CCTTG A L	1440 39 1530 42 1620 45
CCCCCTTTGAAAA S R F E K TTGTGGGATGATGAA L W D D E TTATTCGTAGGATG L F V G C AAGTTTCCTGGTGGT	CAGTTAG Q L AGCTGGTA A G ITACGATC Y D IGTACCAA	GTCGAGA G R I GCTGGT S W I CGAAAA P K 1 CATCATT	ACAACG DN2 TTGATT. 7 D ? CAGTTC. F V I IGATGC	CATCAC A S ACGATT Y D ATCATC H H AGACAP	STTGT V V CTGC S A SAAGA E D AGCCAA	CTATTC Y S GCATCG H R CGTTGC V A ACAATG	TGATT D I GAGATO R W CACCAO T F GGACTI	TATAC GGAAC V N GAGTA R V FTCCT	CAGTGA S E CACGAA T N ACTCGA L E CAATGG	AATGG W ACTTI F ATTAC Y KCTGG	AAAAAC K T TATGI Y V CTTGA L E CCTCC	CGTCG S TTTCT S AAAAA K TCTCC	I I N AGTZ S CAGC	AATG N A FTGA L 1 AACG N A	CCATT A I CTCCC F P CCTTG A L IGCTG	1440 39 1530 42 1620 45 1710
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATG L F V G C AAGTTTCCTGGTGGT K F P G G	Q L Q L AGCTGGTA A G TTACGATC Y D TGTACCAA V P	GTCGAGA G R I GCTGGT S W I CGAAAAA P K 1 CATCAT1 T S I	ACAACGO DN2 TTGATT FD2 CAGTTC FV1 TGATGC2 LMQ	CATCAC A S ACGAT Y D ATCATC H H AGACAR Q T	STTGT V V S A SAAGA E D AGCCAA S Q	CTATTC Y S GCATCG H R CGTTGC V A ACAATGO	TGATT D I GAGAT R V CACCAC T F GGACTI	TATAC L Y GGAAC W N GAGTA R V FTCCT	CAGTGA S E CACGAA T N ACTCGA L E CAATGG	AATGG W ACTTI F ATTAC Y CTGG	AAAAAC K T TATGI Y V CTTGA L E CCTCC	GTCG S TTTCT S AAAAA K TCTCC	I N AGTZ S CAGC Q	AATG N A ITGA L I AACG N A EACAI H N	CCATT A I CTCCC F P CCTTG A L IGCTG A L	1440 39 1530 42 1620 45 1710 48
ICCCGCTTTGAAAAC S R F E K ITGTGGGATGATGAT L W D D E ITATTCGTAGGATGAT L F V G C AAGTTTCCTGGTGGT K F P G G STTATGGGACTTGA	CAGTTAG Q L AGCTGGTA A G ITACGATO Y D IGTACCAA V P CAAGACCG	GTCGAGA G R I GCTGGT S W I CGAAAAA P K T CATCATT T S I GTGATCC	ACAACG DN 2 TTGATT. 7 D 7 CAGTTC. 1 V 1 IGATGC2 2 M 6 CAAGGG6	CATCAC A S ACGAT Y D ATCATC H H AGACAF Q T CCAAAG	V V CCTGCC S A SAAGA E D AGCCAA S Q SAACTO	CTATTC Y S GCATCG H R CGTTGC V A ACAATG Q:W GCTTT	TGATT D I GAGAT R V CACCA T F GGACT I I GACG	TATAC L Y GGAAC V N GAGTA R V FTCCT F. P. ITGCG	CAGAG	AATGG W ACTTI F ATTAC Y CTGG	AAAAAC K T TATGI Y V CTTGA L E CCTCC P:P GTTTT	GTCG S TTCT S AAAAA K TCTCC ] L CAAT	I N AGTZ S CAGC Q AATT	AATG N 2 FTGA L 7 AACG N 2 AACG N 2 AACG N 2 AACA7 H N 2 ATG2	CCATT A I CTCCC F P CCTTG A L IGCTG A L IGCTG A L	1440 39 1530 42 1620 45 1710 48 1800
TCCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATG L F V G C AAGTTTCCTGGTGGT K F P G G GTTATGGGACTTGAC V M G L D	CAGTTAG Q L AGCTGGTA A G TTACGATC Y D TGTACCAA V P CAAGACCG K T	GTCGAGA G R I GCTGGT S W I CCGAAAA P K 7 CATCATT T S I GTGATCC G D I	ACAACG DN 2 TTGATT 7 D 2 CAGTTC. F V I IGATGC2 CAAGGG 2 R 2	CATCAC A S ACGATT Y D ATCATC H H AGACAF Q T CCAAAG A K	YTTGT VV SA SAAGA ED SCCAA SQ SAACT CEL	CTATTC Y S GCATCG H R CGTTGC V A ACAATG Q:W 3GCTTT A F	TGATT D I GAGATO R V CACCAO T F GGACTI C C CACCAI T F GGACGI D (	TATAC L Y GGAAC V N GAGTA R V FTCCT FTCCG TTGCG V A	CAGTGA S E CACGAA T N ACTCGA L E CAATGG CAGAG Q F	AATGG W ACTTI F ATTAC Y CTGG CTGG ATGG	AAAAAC K T TATGI Y V CTTGA L E CCTCC P:P GTTTT V F	GTCG S TTCT S AAAAA K TCTCC ] L CAAT N	I N AGTZ S CAGO Q AATT N	AATG N A TTGA L T AACG N A CACAT H N CATGA Y I	CCATT CTCCC P CCTTG L IGCTG L IGCTG L AGCCG S A	1440 39 1530 42 1620 45 1710 48 1800 51
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAJ L W D D E TTATTCGTAGGATGT L F V G C AAGTTTCCTGGTGGT K F P G G STTATGGGACTTGAC V M G L D TTCACGCAAAGCTTJ	Q L Q L AGCTGGTA A G TTACGATO Y D NGTACCAA V P CAAGACCO K T ACCTAATO	GTCGAGA G R I GCTGGT S W I CCGAAAAA P K 1 CATCAT T S I GTGATCC G D I CCATGT	ACAACGO N 2 TTGATT. 7 D 2 CAGTTC. F V 1 TGATGC2 CAAGGGC P R 2 TTGAAAA	CATCAC A S ACGAT Y D ATCATC H H AGACAP Q T CCAAAC A K AGTATC	STTGT V V CTGCC S A SAAGA E D AGCCAA S Q SAACT E L SACGTZ	CTATTC Y S GCATCG H R CGTTGC V A ACAATGC <u>Q:W</u> SGCTTT A F AACGGT	TGATT D I GAGATC R W CACCAC T F GGACTI IGACGI D ( CGTTGC	TATAC GGAAC W N GAGTA R V FTCCT F::P:: TTGCG V A SCCTC	AGTGA S E ACGAA T N ACTCGA L E AATGG CAGAG Q F CCCGG	AATGG W ACTTT F ATTAC Y CTGG ATGG W	AAAAAC K T TATGI Y V CCTTGA L E CCTCC CTCC CTCC CTCC CTCC CTCC CTCC C	GTCG S TTCT S AAAAA K TCTCC ] L CAAT N AGAG	I NAGTZ S CAGC Q AATT N TACG	AATG N 2 TTGA L 1 AACG N 2 AACG H 1 AACG2 Y 1 AACG2	CCATT I CTCCC P CCTTG L IGCTG I L IGCTG I L IGCTG A L IGCTG A L IGCTG A L IGCTG A L I I I I I I I I I I I I I	1440 39 1530 42 1620 45 1710 48 1800 51 1890
CCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAJ L W D D E TTATTCGTAGGATGA L F V G C AAGTTTCCTGGTGGT K F P G G <u>STTATGGGACTTGAC</u> V M G L D TTCACGCAAAGCTT7 F T Q S L	Q L Q L AGCTGGTA A G TTACGATO Y D NGTACCAA V P CAAGACCO K T ACCTAATO P N	GTCGAGA G R I GCTGGT S W I CCGAAAA P K 2 CATCAT T S I GTGATCC G D I CCCATGT A M I	ACAACGO N 2 TTGATT. T D 2 CAGTTC. F V I TGATGC2 C M ( CAAGGGC P R 2 TTGAAAA F E I	CATCAC A S ACGAT Y D ATCATC H H AGACAF Q T CCAAAC A K AGTATC K Y	YTTGT VV FCTGC SA SAAGA ED AGCCAA SQ SAACTC EL SACGTA DV	CTATTC Y S GCATCG H R CGTTGC V A ACAATG Q::W SGCTTT A F AACGGT T V	TGATT D I GAGATC R W CACCAC T F GGACTI CACCTI CONTROL V C	TATAC L Y GGAAC W N GAGTA R V FTCCT FTCCT FTCCG V A SCCTC 3 L	AGTGA S E ACGAA T N ACTCGA L E PAATGG CAGAG Q F CCCGG P G	ATGG W ACTTI F ATTAC Y CTGG ATGG W TGGT	AAAAAC K T TATGI Y V CTTGA L E CCTCC <u>P:P</u> R GTTTT V F GGGGGG G G	GTCG S TTCT S AAAAA K TCTCC ] L CAAT N AGAG E	I I N AGT/ S CAGC Q AATTI N TACCG Y	AATG N 2 TTGA L 7 AACG N 2 CACA Y 1 FACG Y 1 CACG D 7	CCATT I CTCCC P CCTTG L CCTTG L CCTTG L CCTTG A L CCTTG A L CCTCCC A L CCTCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCC A L CTCCCCC A L CTCCCCC A L CTCCCCC A L CTCCCCC A L CTCCCCC A L CTCCCCC A L CTCCCCC A L CTCCCCCC A L CTCCCCC A L C CTCCCCC A L C CTCCCCC A L C CTCCCCCCCCC A L C CTCCCCC A L C CTCCCCC A L C CTCCCCC A L C C C C C C C C C C C C C	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54
CCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAJ L W D D E TTATTCGTAGGATGA L F V G C AAGTTTCCTGGTGGT K F P G G GTTATGGGACTTGAC V M G L D TTCACGCAAAGCTT7 F T Q S L CTAGGTTTTGGCTGC	Q L Q L AGCTGGTA A G TTACGATC Y D IGTACCAA V P CAAGACCG K T ACCTAATG P N SACAAATG	GTCGAGA G R I GCTGGT S W I CCGAAAA P K 7 CATCATT T S I GTGATCC G D I CCCATGT7 A M I GTGTAG	ACAACGO N 2 TTGATT. T D 1 CAGTTC. T V 1 TGATGC2 CAAGGGC P R 2 TTGAAAA TTGAAAA F E 1 TAATGG.	CATCAC A S ACGAT Y D ATCATC H H AGACAP Q T CCAAAG A K AGTATC K Y ACTTT	YTTGT VVV PCTGC SA SAAGA ED SAACA SQ SAACT EL SACGT DV TTAGT	CTATTC Y S GCATCG H R CGTTGC V A ACAATG Q::W GGCTTT A F AACGGT T V TAAATA	TGATT D I GAGATC R W CACCAC T F GGACTI <u>CGACGI</u> D ( CGTTGC V C CGGTG	TATAC L Y GGAAC N N GAGTA R V TTCCT F:P: TTGCG V A GCCTC G L AGAGA	AGTGA S E CACGAA T N ACTCGA L E CAATGG CAGAG Q F CCCGG P G	ATGG W ACTTI F ATTAC Y CTGG ATGG W TGGT G CTTCA	AAAAAC K T TATGI Y V CTTGA L E CCTCC CTCC CTCC CTCC CTCC GTTTT V F GGGGGG G G G G GTCCC	CGTCG S TTTCT S AAAAA K TCTCC CAAT N AGAGG E TTGAC	I I N AGT/ S CAGC/ Q AATT N TACG Y	AATG N 2 TTGA L 1 AACG N 2 AACG Y 1 BACG Y 1 D V AAACC	CCATT I CTCCC P CCTTG L I CCTTG L I CCTTG L I CCTTG L I CCTTG L I CCTTG L I CCTTG L I CCTCCC P CCTTG L I CCTCCC I P CCTTG L I CCCTG I L I CCCTG I L I CCCTG I L I CCCTG I L I CCCTG I L I CCCTG I L I CCCTG I L I CCCTG I L I CCCTG I L I CCCG I L I CCCG I L I CCC I L I CCC I L I CCC I L I A A CCG Z A L Z A A CCG Z A A C C A A C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C C A A C A A A C A A A A A A A A A A A A A	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 .1980
TCCCGCTTTGAAAAG   S R F E K   TTGTGGGATGATGATGATGATGATGATGATGGGATGATGA	Q L AGCTGGTA A G TTACGATC Y D NGTACCAA V P CAAGACCG K T ACCTAATG P N GACAAATG T N	GTCGAGJ G R I GCTGGT S W I CCGAAAA P K T CATCATT T S I GTGAACC G D I CCATGTT A M I GTGTAAC G V V	ACAACGA TTGATT. T D CAGTTC. T V II IGATGCI C M (C CAAGGGC P R I TTGAAAA F E II TAATGG. V M I	CATCAC A S ACGAT Y D ATCATC H H AGACAA Q T CCAAAG A K AGTATC K Y ACTTT D F	TTGT V V CTGCC S A SAAGA E D GCCA S Q SAACT E L SACGT D V TTAGT L V	CTATTCC Y S GCATCG H R CGTTGC V A ACAATGC Q::W SGCTTTT A F AACGGT T V TAAATA K Y	TGATT D I GAGAT R V CACCAO T F GGACT <u>D ( CGTTGG</u> V C CGGTG G F	TATAC L Y GGAAC V N GAGTA GAGTA R V FTCCT FTCCT FTCCG V A GCCTC G L AGAGA E R	AGTGA S E CACGAA T N ACTCGA L C CAGAG Q F CCCGG P G ATTGAC L 1	AATGG WACTTI F ATTAC O Y CTGG AATGG AATGG W TGGT G CTTCA S	AAAAAC K T TATGI Y V CTTGA L E CCTCC <u>P:P</u> <u>P</u> <u>P</u> <u>P</u> <u>P</u> <u>P</u> <u>P</u> <u>P</u> <u>P</u> <u>P</u> <u></u>	CGTCG S TTTCT S AAAAA K TCTCC I L CAAT N AGAG E TGAC	I N AGT/ S CAGC/ Q AATT N TACG Y SGCA/ A	AATG N / ITTGA L T AACG N / AACG Y I BACG D V AAAAC K (	CCATT I CTCCC P CCTTG L CCTTG L CCTTG L CCTTG A L CCTTG A L CCTTG A L CCTTG A L CCTTG A L CCCTTG A L CCCTCC C C C C C C C C C C C C	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57
TCCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAJ L W D D E TTATTCGTAGGATGA L F V G C AAGTTTCCTGGTGGT K F P G G STTATGGGACTTGAC V M G L D TTCACGCAAAGCTTT F T Q S L CTAGGTTTTGGCTGC L G F G W ATTTGGAGTCCATC	CAGTTAG Q L AGCTGGTA A G TTACGATC Y D NGTACCAA V P CAAGACCG K T ACCTAATG P N SACAAATG T N ATCAGTTT	GTCGAGJ G R I GCTGGT S W I CGGAAAAA P K 7 CATCATT T S I GTGATCC G D I CCATGTT A M I GTGTAG G V V CCAGTTG	ACAACGA TTGATT. T D T CAGTTC. T V II GATGCI C M ( CAAGGGC P R I TTGAAAA F E II TAATGG. V M II GAGTTG	CATCAC A S ACGAT Y D ATCATC H H AGACAR Q T CCAAAC A K AGTATC K Y ACTTT O F GTGCT	TTGT V V V CCTGCC S A SAAGAA E D GGCCAI S Q GAACTI E L V CAACGTI D V TTAGT L V GGTGC	CTATTC Y S GCATCG H R CGTTGC V A ACAATG Q::W SGCTTT A F AACGGT T V TAAATA K Y AACGTT	TGATT D I GAGATY R V CACCAG T F GGACTI D ( CGTTGC CGTTGC CGGTGC G F CCTCGG	TATAC L Y GGGAAC V N GAGTA R V TTC <u>CT</u> TTC <u>CT</u> TTC <u>CT</u> C C C C C C C C C C C C C	AGTGA SE ACGAA T N ACTCGA L C AATGG CAGAG Q F CCCGG P G ATTGAA L T TTGGAA	AATGG WACTTI F ATTAC Y CTGG AATGG AATGG G CTTCA S FTGGT	AAAAAC K T TATGI Y V CTTGA L E CCTCC TC TC TC CTCC GGGGG G G GGGGGG G G GGGCCC V R CTTGC	GTCG S TTTCT S AAAAA K TCTCC CAAT N AGAG G STGAC D CAAGT	I I AATT N AAGT S CAGC Q AATT N TACG Y SGCAA A XACTO	AATG N 2 TTGA L 1 AACG N 2 CACA1 H N AATG2 L AATG2 D V I D V AAAC K ( CGAT	A I CTCCC P P CCCTTG A L ICCTG A L ICCTG A L ICCTG A AACTA 2 L TCTAI	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 .1980 57 2070
TCCCCGCTTTGAAAAC   S R F E K   TTGTGGGATGATGATGATGATGATGATGAGATGGAGTTCCTGGTGGGGGGGG	CAGTTAG Q L AGCTGGTA A G TTACGATC Y D NGTACCAA V P CAAGACCG K T ACCTAATG P N SACAAATG T N ATCAGTTT S V	GTCGAGJ G R I GCTGGT S W I CGAAAAA P K 7 CATCATT T S I GTGATCC G D I CCATGTT A M I GTGTAG G V V CAGTTG S V C	ACAACGA TTGATT. T D T CAGTTC. T V II IGATGCI C M ( CAAGGGC P R I ITGAAAA F E II IAATGG. V M II GAGTTG GAGTTG G V (	CATCAC A S ACGAT Y D ATCATC H H AGACAR Q T CCAAAG A K AGTATC K Y ACTTT O F GTGCTC 3 A	V V VCTGC S A GAAGA E D GGCCA S Q GAACTA E L GAACTA D V TTAGT L V GGTGC G A	CTATTC Y S GCATCG H R CGTTGC V A ACAATG GCTTT A F AACGGT T V TAAATA K Y AACGTT T F	TGATT D I GAGAT R V CACCAO T F GGACTJ D C CGTGO V C CGTGO G F CCTCGO L C	TATAC L Y GGAAC W N GAGTA R V TTCCT F::P:: F::P:: AGCCTC GCCTC GCCTC GCCTC GCCTC GCCTC GCCTC GCCTC GCAC	AGTGA S E ACGAJ T N L C CACGAJ L C AATGGG CAGAG Q F C CCCGG Q F C CCCGG Q F C CCCGG L T TTGGA W I	AATGG WACTTI F ATTAC Y Y CTGG W TGGT G CTTCA S TTGGI S G	K T TATGI Y V CTTGA L E CCCCC C C C C C C C C C C C C C C C C	CGTCG S TTTCT S AAAAA K TCTCC I CAAT N AGAG E TTGAC D CAAGI	I ZAATT N AGT7 S CAGC Q AATT N TACG Y S GCAA A CACTC T	AATG N 2 TTGA L 1 AACG N 2 AAACG H 1 AACG T L AAACG CGAT R 1	CCCATT L CTCCCC P CCCTTG L CCCTGG L CCCTGG L CCCTGG A L CCCCTGG A L CCCCTGG A L CCCCCG P CCCTGG A L CCCCCC P CCCCCC C P CCCCCC C P CCCCCC C P CCCCCC C P CCCCCG A L CCCCCC C P CCCCTGG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L CCCCGG A L C CCCCGG A L C CCCCGG A L C CCCCGG A L C CCCCGG A L C CCCCGG A C C C C C C C C C C C C C	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 .1980 57 2070 60
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGAT L F V G C AAGTTTCCTGGTGGT K F P G G STTATGGGACTTGAG V M G L D TTCACGCAAAGCTTT F T Q S L CTAGGTTTTGGCTGG L G F G W ATTTGGAGTCCATCA I W S P S CGCCAACTTTTCAAG	AGCAGTTAG Q L AGCTGGTA A G TTACGATC Y D GTACCAAT V P CAAGACCGA K T ACCTAATG P N SACAAATG T N ATCAGTTT S V STCCGACC	GTCGAGA G R I GCTGGT S W I CCGADAAA P K 7 CATCAT T S I GTGATCC G D I CCATGT A M I GTGTAG G V V CAGTTG S V C CACAAAA	ACAACGA TTGATT TGATT TGATGA CAGTTC TGATGCI CAGTTC CAGTCA TGAAGGGC R TGAAGGGC R TGAAGGCA TGGGAG TGGGAG	CATCAC A S ACGAT: ( D ATCAT( H H AGACAP 2 T CCAAAC A K AGTATC ( Y ACTTT: F GTGCTC 3 A GCATCC	TTGT V V CTGCC S A BAAGA E D GCCA S Q BACTA E L BACGTI D V TTAGT L V GGTGC. G A CATGG	CTATTCC Y S GCATCG H R CGTTGC V A ACAATGG Q::M SGCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGGC	TGATT: D I GAGATY R V CACCAA T F SGACTT CACCAA T F GACGT CACCAA CCCTTGG CCCTCGG CCCTCGG L C CCTTGAGG	TATACC L Y GGAAC N N GAGTA R V TTCCT F::P: F::P: A GCCTC J L AGAGA E R GTGCT G A CAATT	AGTG/ S E ACGAJ T N L C CAGAJ CAGAG Q F CCCGG P G CCCGG P G UTTGAC L T TTGAC W I UCAAG	AATGG WACTTI F ATTAC Y CTGG ATTGG G CTTCA S TTGGI G GACTI	AAAAAC   K T   TTATGT   Y V   V CTTGA   L E   CCTCC E   P::P P   GGTTTT V   V F   GGGGGG G   G G   GGTCCC V   V R   CCTCC L   A ACTTGC   L A   XATTGC A	CGTCG S TTTCT S AAAAA K TCTCC ] L CAAT N AGAG CTGAC CAAGI S SCCTGI	I VAATT N AGTT S CAGC Q AATT N TACG Y CGCAJ A CACTY T CGTTY	AATG N 2 TTGA L 7 AACG N 2 CACA7 V 1 BACG7 D 7 AAAC K ( CGAT R 1 3GTT	CCCATT A I CTCCCC P P CCCTGG A L CCCTGG A L CCCTGG A L CCCTGG A L CCCTGG A L CCCCCC P P CCCTGG A L CCCCCCC P P CCCCCCC P P CCCCCCC P P CCCCCCC P P CCCTGG A L CCCCCCC P P CCCTGG A L CCCCCCC P P CCCTGG A L CCCCCCC P P CCCTGG A L CCCCCCC A L CCCCCCCCCCCCCC P P CCCTGG A L CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 22160
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGAT L F V G C AAGTTTCCTGGTGGT K F P G G STTATGGGACTTGAG V M G L D TTCACGCAAAGCTTT F T Q S L CTAGGTTTTGGCTGC L G F G W ATTTGGAGTCCATCA I W S P S CGCCAACTTTTCAAAG R Q L F K	AGCTAGTTAG Q L AGCTGGTA A G TTACGATC Y D NGTACCAA V P CAAGACCG K T ACCTAATG P N ACCAATG T N ATCAGTTT S V STCCGACC S D	GTCGAGA G R I GCTGGT S W I CCGADAAA P K 7 CATCAT T S I GTGATCC G D I CCATGT A M I GTGTAG G V V CAGTTG S V C CACAAAA H K I	ACAACGA TTGATT TGATT TGATGA CAGTTC TGATGCA CAAGGGC R A CAAGGGC R A TGAAGGC R A TGAAGGC A TGGGAG A G G CAGTCG CAGTCG CAAGGCA CAA	CATCAC A S ACGAT: Y D ATCAT( I H AGACAP Q T CCAAAC A K AGTATC K Y ACTTT: D F GTGCT( 3 A GCATC( 3 I	TTGTU V V V CTGCCC S A SAAGAA S Q SAACTO E L SACGTI D V TTAGT L V GGTGC. G A CATGG H G	CTATTCC Y S GCATCG H R CGTTGC V A ACAATGG Q::M SGCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGGC V A	TGATT: D I GAGATY R V GACCAI T F GACCTI D ( CCCTCGI CCCTCGI CCCTCGI L C TGAGGG E Z	TATACC GGAAC N N GAGTA R V TTCCT TTGCG V A GCCTC G L AGAGA A GTGCT G A CAATT A I	AGTGA SEACGAN T N ACTCGA L C AATGG CAGAG Q F CCCCGG P C CCCGG P C CCCGG P C CCCGG P C CCGGAG V TTGGA V I VCAAGG	AATGG W ACTTI F ATTAC Y CTGG CTGG CTGG G CTTCA S CTGGT S CTTCA S CTTCA S CTGGT C S CTTCA S	AAAAAC   K T   TTATGT   Y V   V CTTGA   L E   CCTGA E   CCTCC E   CCTCC E   CTTGA F   QCTTGA GGTTTT   V F   GGGGGG G   G G   QCTTGC L   L A   XATTGC I   I A	CATCG S TTTCT S AAAAA K TCTCC ] L CAAT N AGAG C TCAAC D CAAGT S CCTGT	I VAATTI N AGTI S CAGCO Q AATTI N TACG Y CACTO T CACTO T	AATG N 2 ITGA L 7 AACG N 2 AACG N 2 AACG Y I I AATG2 Y I I AATG2 X Y I I AACG X CGAT R 1 3GTT G 8	CCATT CCCCAT CTCCCC P CCTTG A L CCCTG A L CCCTG A AAGCG A A CCCTG A A CCCTG A L CCCCCC P CCTTG A L CCCCCC P CCTTG A L CCCCCC P CCTTG A L CCCCCC A L CCCCCC A L CCCCCC A L CCCCCC A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L C CCCCG A C C C C C C C C C C C C C	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 2160 63
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGAT L F V G C AAGTTTCCTGGTGGT K F P G G STTATGGGACTTGAG V M G L D TTCACGCAAAGCTTA F T Q S L CTAGGTTTTGGCTGC L G F G W ATTTGGAGTCCATCA I W S P S CGCCAACTTTTCAAG R Q L F K CTCGTGGCTACTGG	AGCAGTTAG Q L AGCTGGTA A G TTACGATC Y D GTACGAATC V P CAAGACCG K T ACCTAATG P N SACCAATG S V STCCGACC S D AGCACTT	GTCGAGA G R I GCTGGT S W I CGAAAAA P K 7 CATCAT T S I GTGATCC G D I CCATGT A M I GTGTAG G V V CAGTTG S V C CAGTTG S V C CAGTAAA S V C CACAAAA H K I	ACAACGA TTGATT TGATGAT CAGTTC. TGATGCI CAGTTC. TGAAGGGC R I TGAAGGGC R I TGAAGGGC R I TGAAGG TGGAAAI G C AAAGGGC M G C AAAGGGC	CATCAC A S ACGATT Y D ATCATC I H AGACAP Q T CCAAAC A K AGTATC CCAAAC A K AGTATC CCAAAC A K AGTATC CCAAAC A S AGCATC 3 I ATACGI	TTGT V V CTGCC S A SAAGA E D GCCAI S Q SAACTC E L SACGTI D V TTAGT L V GGTGC. G A CATGG H G ACCAA	CTATTCC Y S GCATCG H R CGTTGC V A ACAATG Q::W BGCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGC V A ACGACC	IGATT: D I GAGATY R V CACCAA T F SGACTT D CACCAA CACCAA CACCAA CACCAA CACCAA CACAAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAA CACAAA CACAAA CACAA CACAAA CACAAA CACAAA CACAAA	TATACC L Y GGAAC N N GAGTA R V TTCCT TTGCG V A GCCTC G L A GCCTC G C A C A C A C A C A C C C C C C C C C C C C C	AGTGA SEACGAN TN ACTCGA L C AATGGA CCAGAG Q F CCCCGG P C CCCGGA L T TGGA W I I CCAGAG Q C	AATGG WACTTI F ATTAC Y CTGG ATGG ATGG T CTTCA S TTGGI CTTCA S CTTCA S CTTCA S CTTCA S CTTCA S CTTCA C CCCCI	AAAAAC   K T   TTATGT   Y V   V CTTGA   L E   CCTGA E   CCTCC E   QCTTGA E   QCTTGA E   QCTTGA E   QCTTGA E   QCTTGC E   V F   GGGGGG G   G G   QCTTGC L   L L   VATTGC L   XATTGC L   XATGGC L	GTCG S STTCT S AAAAA K TCTCC ] L CAAT N AGAG E STGAC C CCTGT C CCAGC	I AATTI N AAGTI S CAGC Q AATTI N TACCG Y CACTY T V CCCAI	AATG N 2 TTGA L 7 AAACG N 2 AAACG N 2 AAACG V 1 BACG V 1 BACG V 1 BACG V 1 BACG V 1 BAAAC CGAT R 1 BGGT G 2 AAAA	CCATT CCCCAT CTCCCC P CCTTG L CCTTG L CCTTG A L CCTTG A AAGCG A CTCAA A CTCAA C CTCAA C C C C C C C C C C C C C	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 2160 63 2250
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGT L F V G C AAGTTTCCTGGTGGT K F P G G <u>STTATGGGACTTGAG</u> V M G L D TTCACGCAAAGCTTT F T Q S L CTAGGTTTTGGCTGC L G F G W ATTTGGAGTCCATC I W S P S CGCCAACTTTTCAAA R Q L F K CTCGTGGCTACTGC L V A T A	AGCTAGTTAG Q L AGCTAGTA A G TTACGATC Y D GTACCAAT V P CAAGACCCA K T ACCTAATG P N SACCAATTT S V STCCGACC S D AGCACTTT A L	GTCGAGA G R I GCTGGT S W I CGAAAAA P K 7 CATCAT T S I GTGATCC G D I CCATGT A M I GTGTAG G V V CCAGTTG S V C CAGTTG S V C C CAGTTG S V C C CAGTTG S V C C C C C C C C C C C C C C C C C C C	ACAACGA TTGATT TGATT TGATGA TGATGA TGAAGGG R AAAGGGA TGGGAG M G C AAAAGG K I K I K	CATCAC A S ACGATT ( D ATCAT( I H AGACAP Q T CCAAAC A K AGTATC ( Y A AGTATC C A K AGTATC C A K AGTATC C A K AGTATC C A S C A S C C C A A S C A S C C C A A S C C C A A S C C C C A A S C C C C A S C C C C A S C C C C C A S C C C C C C C C C C C C C C C C C C C	TTGT V V V CTGCC S A SAAGA E D GCCAI S Q AACTA E L V CACGTI C V SGTGC. G A CATGG H G ACGAA T K	CTATTCC Y S GCATCG H R CGTTGC V A ACAATG Q::W BGCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGC V A ACGACCC R P	TGATT: D I GAGATY R V CACCAA T F SGACTT D CACCAA D CCTTGACGT C CCTCGG C CCCCGGGG C CCCCCGGGG C CCCCCGGGG C CCCCCGGGG C CCCCCGGGG C CCCCCGGGG C CCCCCC	TATACC L Y GGAAC N N GAGTA R V TTCCT TTGCG C C GGACA C A GGCTC C C C C C C C C C C C C C	AGTGA S E ACGAN T N ACTCGA L C AATGGC CAGAG Q F CCCCGG P C CCCGGG P C CCCGGG P C CCCGGG NTTGGA U TTGGA C C CAGAG Q C	AATGG W ACTTI F ATTAC O Y CTGG CTGG AATGG W TGGT CTGGT G G CTTCA G G CTTCA G CTTCA G CTTCA C S CTCGGT C CCCCI C CCCCI	AAAAAC   K T   TTATGT   Y V   V CTTGA   L E   CCTCCC E   Q STTTT   V F   GGTTTT V   V F   GGCGGG G   G G   G G   G G   G G   G G   Q S	GTCG S S TTCT S AAAAA K TCTCC I C AAAAT N AGAG E C CAAGT S CCAGC S	I I VAATT N AGTA S CAGCO Q AATT N TACCG Y CGCAA A CACTY V V V CCCAA P	AATGAAATGAAATGAAATGAAATGAAATGAAATGAAACGAAAAACGAATGAAAACCGATRIJ	CCATT CCCCAT CTCCCC P CCTTG L CCTTG L CCTTG A L CCTTG A L CCTTG A L CCTCAA 7 Q AACTA 2 L CCTCAT C C C C C C C C C C C C C	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 2160 63 2250 66
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGAT L F V G C AAGTTTCCTGGTGGA K F P G G <u>3TTATGGGACTTGAG</u> V M G L D TTCACGCAAAGCTTA F T Q S L CTAGGTTTTGGCTGC L G F G W ATTTGGAGTCCATCA I W S P S CGCCAACTTTTCAAC R Q L F K CTCGTGGCTACTGCA	AGCAGTTAG Q L AGCTGGTA A G TTACGATC Y D GTACGATC Y D CATACGATC R T ACCTAATG P N ACCTAATG P N ACCTAATG S V STCCGACC S D AGCACTTT A L CGGCTTC7	GTCGAGA G R I GCTGGT S W I CGAAAAA P K 7 CATCAT T S I GTGATCC G D I CCATGT A M I GTGTAG G V V CCAGTTG S V C CAGTTG S V C CAGTTG S V C CAGTTG F S I TTATTGT	ACAACGA TTGATT TGATT CAGTTC. TO CAGTTC. TGATGC2 C ACAAGGG R TGAAGGG C TGGAAAA TGGGAG M G C AAAAGG C K 1 TCCCAAC	CATCAC A S ACGATT Y D ATCATC I H AGACAPA Q T CCAAAC A K AGTATC X Y ACTTTT D F GTGCTC GTGCTC GGCATCC G I ATACGI D T GAGGAGAI	TTGT V V V CCTGCC S A SAAGAA E D GCCAA S Q AAACTA C C CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGA	CTATTCC Y S GCATCG H R CGTTGC V A ACAATG Q::W 3GCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGC V A ACGACCC R P AGGCCA	TGATT: D I GAGATX R V CACCAA T F SCACCAA T F SCACCAA T F SCACCAA CACCAA CCCAA	TATACC GGAAC N N GAGTA R V TTCCT TTGCG GGAGTA GGCTC GGAGTA GGCCTC GGAGTA A GCCTCC GGAGTA A CCAATTI A CCAATTI A CCGGAC CCAATTI A CCGGAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCCCC CCCCC CCCCC CCCCCC CCCCCC CCCCCC	AGTGA SEACGAN T N ACTCGA L C AATGG CAGAGA Q F CCCCGG P C CCCGG Q F CCCGGG VTTGAC V I VCCAGAC Q I VCCCT	AATGG W ACTTI F ATTAC Y CTGG ATGG AATGG W TGGT C CTTCA G G CTTCA G G CTTCA G CTTCA G CTTCA C C CCCCI C P TAAGGC	AAAAAC   K T   TATGI   Y V   V CTTGA   L E   CCTCCC E   QCTTGA E   QTTGA F   GGTTT V   V F   GGCGGG G   G G   G G   G GCGCGG   V R   CTTGC V   V R   CTTGC L   A ATTGC   I A   VATTGC Q   SAACCC SAACCC	GTCG S TTTCT S AAAAA K TCTCC I C CAAT S CCAAT S CCAGC C CCAGC S SCCCA	I I N AGT/ S CAGC Q AATT N TACG Y CGCA/ T CCCA/ P N ATT/	AATG AATG N 2 TTGA L 1 AACG N 2 AAACG V 1 AAACG V 1 AAAAC CGAT CGAT AAAAC CGAT CGAT AATAA I 1 AGTG	CCATT CCCATT CCCCCC P CCTTG L CCTTG L CCTTG A L CCTTG A AAGCG A CCTCTAI TCCAA CCTCTAI TCCAA CCTCTAI CCTCTAI CCCCCCC A L CCCCCCC C P CCTCCCC C P CCTTG A L CCCCCC C P CCTTG A L CCCCCC C P CCTTG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCG A C C CCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C A CCCCG A C A CCCCG A C A A C A A C A C A A C A A A A A A A A A A A A A	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 2160 63 2250 66 22340
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGT L F V G C AAGTTTCCTGGTGGT K F P G G <u>STTATGGGACTTGAG</u> V M G L D TTCACGCAAAGCTTT F T Q S L CTAGGTTTTGGCTGC L G F G W ATTTGGAGTCCATC I W S P S CGCCAACTTTTCAAA R Q L F K CTCGTGGCTACTGC L V A T A	AGCAGTTAG Q L AGCTGGTA A G TTACGATC Y D GTACGATC Y D CATACGATC R T ACCTAATG P N ACCTAATG P N ACCTAATG S V STCCGACC S D AGCACTTT A L CGGCTTC7	GTCGAGA G R I GCTGGT S W I CGAAAAA P K 7 CATCAT T S I GTGATCC G D I CCATGT A M I GTGTAG G V V CCAGTTG S V C CAGTTG S V C CAGTTG S V C CAGTTG F S I TTATTGT	ACAACGA TTGATT TGATT CAGTTC. TO CAGTTC. TGATGC2 C ACAAGGG R TGAAGGG C TGGAAAA TGGGAG M G C AAAAGG C K 1 TCCCAAC	CATCAC A S ACGATT Y D ATCATC I H AGACAPA Q T CCAAAC A K AGTATC X Y ACTTTT D F GTGCTC GTGCTC GGCATCC G I ATACGI D T GAGGAGAI	TTGT V V V CCTGCC S A SAAGAA E D GCCAA S Q AAACTA C C CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGA	CTATTCC Y S GCATCG H R CGTTGC V A ACAATG Q::W 3GCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGC V A ACGACCC R P AGGCCA	TGATT: D I GAGATX R V CACCAA T F SCACCAA T F SCACCAA T F SCACCAA CACCAA CCCAA	TATACC GGAAC N N GAGTA R V TTCCT TTGCG GGAGTA GGCTC GGAGTA GGCCTC GGAGTA A GCCTCC GGAGTA A CCAATTI A CCAATTI A CCGGAC CCAATTI A CCGGAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCGCAC CCCCC CCCCC CCCCC CCCCCC CCCCCC CCCCCC	AGTGA SEACGAN T N ACTCGA L C AATGG CAGAGA Q F CCCCGG P C CCCGG Q F CCCGGG VTTGAC V I VCCAGAC Q I VCCCT	AATGG W ACTTI F ATTAC Y CTGG ATGG AATGG W TGGT C CTTCA G G CTTCA G G CTTCA G CTTCA G CTTCA C C CCCCI C P TAAGGC	AAAAAC   K T   TATGI   Y V   V CTTGA   L E   CCTCCC E   QCTTGA E   QTTGA F   GGTTT V   V F   GGCGGG G   G G   G G   G GCGCGG   V R   CTTGC V   V R   CTTGC L   A ATTGC   I A   VATTGC Q   SAACCC SAACCC	GTCG S TTTCT S AAAAA K TCTCC I C CAAT S CCAAT S CCAGC C CCAGC S SCCCA	I I N AGT/ S CAGC Q AATT N TACG Y CGCA/ T CCCA/ P N ATT/	AATG AATG N 2 TTGA L 1 AACG N 2 AAACG V 1 AAACG V 1 AAAAC CGAT R 1 GGTT G 2 AAAAC CGAT AAAAC CGAT AATAA I 1 AGTG	CCATT CCCATT CCCCCC P CCTTG L CCTTG L CCTTG A L CCTTG A AAGCG A CCTCTAI TCCAA CCTCTAI TCCAA CCTCTAI CCTCTAI CCCCCCC A L CCCCCCC C P CCTCCCC C P CCTTG A L CCCCCC C P CCTTG A L CCCCCC C P CCTTG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCCG A L CCCCG A C C CCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C CCCCG A C C A CCCCG A C A CCCCG A C A A C A A C A C A A C A A A A A A A A A A A A A	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 2160 63 2250 66
ICCCGCTTTGAAAAG S R F E K TTGTGGGATGATGAT L W D D E TTATTCGTAGGATGAT L F V G C AAGTTTCCTGGTGGA K F P G G <u>STTATGGGACTTGAG</u> V M G L D TTCACGCAAAGCTTA F T Q S L CTAGGTTTTGGAGTCCATCA L G F G W ATTTGGAGTCCATCA I W S P S CGCCAACTTTTCAAAG R Q L F K CTCGTGGCTACTGCA	AGCTAGTAG Q L AGCTGGAA A G TTACGAATC Y D CAAGACCGA K T CACCTAATG P N CACCAATG P N CACCAATG S V STCCGACCC S D AGCACTTCA A L CGGCTTCCAG	GTCGAGA G R I GCTGGT S W I CGAAAAA P K T CATCATT T S I GTGATCC G D I CCATGTT A M I GTGTAG G V V CCAGTTGC S V C CCAGTTGC S V C CCAGTTGC F S I TTATTGT L L I	ACAACGA TTGATT. TGATGAT. CAGTTC. TV I IGATGCI CAGTTC. CAGTC. TGAAGGG PR 2 ITGAAGGG PR 2 ITGAAGGG PR 2 ITGAAGGG VM 1 GAGTTGGAG MG 0 AAAAGG K 1 TCCAAC FQ 1	CATCAC A S ACGAT: Y D ATCATX I H AGCAR Q T CCAAAC Y CCAAAC Y A K AGTATC X Y A CTTT: O F GGGCGTC GGCATCC G GGCATCC G GGCATCC Q T GGAGGAI R G	TTGT V V V CCTGCC S A BAAGAA E D AGCCAA S Q AAGCCAA CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGG C A CATGC C	CTATTCC Y S GCATCG H R CGTTGC V A ACAATG Q::W 3GCTTT A F AACGGTT T V TAAATA K Y AACGTT T F TGTGGC V A ACGACCC <u>R P</u> AGGCGA G D	TGATT: D I I GAGATC R V CACCAG T F SGACTT D ( CGTTGC CGTTGC CGTTGC CGGTGG C F CCCCCGG L C CCCCGG L C CCCCGG E 2 TACCTT TACCTT TACCTC E F F	TATACC GGAAC GGAAC N N GGAGTA C C TTCCC TTCCC TTCCC C C C C C C C C C C C C	AGTGA S E ACGAN T N ACTCGA L C AATGG CCAGAG Q F CCCGG Q F CCCGG Q F CCCGG Q F CCCGG Q F CCCGG Q F CCCGG Q F CCCGG Q T CCCAGAG Q T CCCCTT L I	AATGG W ACTTI F ATTAC Y CTGG ATTGG W TGGT G CTTCA S CTTCA S CTTCA S CTTCA S CTTCA C CTTCA C CTCCC C C C C C C C C C	AAAAC K T TTATGT Y V CTTGA L E CCTCC C CTTGA GGTTTT V F GGGGGG G G GGCCC V R CTTGC C C C C C C C C C C C C C C C C C	GTCG S TTTCT S AAAAA K TCTCC ] L CAAAT N AGAG TGAC C CAAGI S CCAGC S SCGCA A	I VATTA I N AGTA S CAGCO Q AATTI N TACCO Y CACTO T CCCAA P AATTA I I	AATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT	CCATT CCCATT CCTCCC P CCTTG CCTTG L CCTCG CA CCTCG CCTCG CCTCTG CCTCTG CCTCTG CCTCTG CCTCTG CCTCTG CCCCCC C A AAAACAA C A AAAAAA C CCCCCCCC	1440 39 1530 42 1620 45 1710 48 1800 51 1890 54 1980 57 2070 60 2160 63 2250 63 2250 66 2340 69

**Fig. 2.** Nucleotide and deduced amino acid sequences of *Tre*E2. The nucleotide and amino acid sequences are numbered from the first nucleotide and the first methionine codon. Asterisks indicate stop codons. A region used as a probe in the screening is underlined. A sequence similar to the polyadenylation signal is heavily underlined. Cleavable N-terminal signal peptide is outlined. The cell adhesion motif is boxed. Potential N-glycosylation sites are dotted. The slashed box indicates the "trehalase signature 1" and the dotted box the "trehalase signature 2" (Henrissat and Bairoch, 1993). Both the signatures are submitted on the PROSITE database. The Thr<sup>246</sup> and Val<sup>504</sup> with circles are replaced by Ser<sup>246</sup> and Leu<sup>504</sup> in the polypeptide coded by *Tre*E3. Predicted transmembrane region is printed in reverse. Potential cAMP-dependent phosphorylation site is doubly underlined.

AR HO OR TE BO	MCAFLSFFVVFGFATTVSCRGDLVLSERAKQLPFVCASEIYCHGTFLHTIQMAGLFRDSKTFVDKKLKINPEEVL MPG-RTWELCL-L-LLLG-LGLG-SQEALPFPCESEIYCHGELLNQVQMAKLYQDDKQFVDMPLSIAPEQVL MPG-STWELHL-L-LLLG-LGLG-SEQALPFPCESQIYCHGELLHQVQMARLYPDDKQFVDMPLSTAPDQVL MIPFLLMVAFADTVL-QVSAQSQBSCDSKVYCQCKLLHVVEMSRIFNDSKTFVELKMINDEQTTL MRLFLLVGLT-TVIADDLPPTCIRPVYCNSTLLHYVQMARLYPDSKTFVDFQMRKDENATL	75 67 67 64 61
AR HO OR TE BO	ASEEVLMNSTDQNPSRDQLAAFINLHFEPEGSBFEEMDPIDWHSNESFLDGIRDTNLKIWGNSLHEAMTWLGRKI QTFTELSRDHNHSIPREQLQAFVHEHFQAKGQELQPWTPADWKDSPOFLQKISDAKLRAWAGQLHQLWKKLGKKM QSFAELAATYNNTVPREQLEKFVQEHFQAVGQELBSWTPGDWKESFQFLQKISDPKLRAWAEQLHLLWKKLGKKI ENFDNFLRDTNHKRTRADLMKFVSDNF-KQENBFESWTPTDFTDNFTLLSRIEDKTIRQFAQDLVKIMPTLARKV SAFQEELDRTNHNPTKEDLQEFVVDFF-DETSELEEWKPDDHKENPPFLAKIRDQGFREFAKALNDIMPTLARRV	142
	Trehalase signature 1	
AR HO OR TE BO	RDDVRINDQLYSMMYLPNPFIIDGGRFRETYYWDSYWIIKGLLISCMHETVKGMLLNFLLMVDTICLVPNGCRIY KPEVLSHPERFSLIYSEHPFIVPGGRFVBFYYWDSYWVMEGLLLSEMAETVKGMLONFLDLVKTYCHVPNGCRVY KPEVLSQPERFSLIYSQHPFIVPGGRFVBFYYWDSYWVMEGLLLSEMAETVKGMLONFLDLVTAYCHIPNGCRVY KKEVLDYPEHYSLLPVDNCFIIPGGRFTBFYYWDSYWIVEGLLLSDMHETVRGMLDNFLSIVEKYCFIPNCARVF KPSVLEKPEQSSLVPMTHCFIVPGGRFKEIYYWDAYWIIEGLLITDMTETAKGMIENLIELLYKFCHIPNGSRWY	225 217 217 213 210
AR HO OR TE BO	YEKRSOPPLLTPMVELYVNATGDIEFLKONIHLEKEMDFWLOERTVNVDGHRLIRYDVKVGGPRPESYKE YLORSOPPLLTLMMDCYLTHTNDTAFLOENIETLALELDFWTNNTVSVSLEGKNYLLNRYYVPYGGPRPESYSK YLORSOPPLLTLMMDRYVAHTGDLAFLRENIETLALELDFWAENRTISVSSGGNSHTLNRYHVPYGGPRPESYSK YLNRSOPPLLTLMVSLYVSATNDMEWLAKNIRTIDTELRFWLNNRLVDVVKDGIVYKLAQYNSNSGSPRPESYYE YQERSOPPLLAAMIKLYYEKTKDIEFIRKYISALEKELEYWLDTHLIAFNKDDRVYTLLRYYIPSAGPRPESY VO	292
AR HO OR TE BO	DIEAIHHNPDLEAQLDFYMNIAAGAETGWDFSSRWYWNCDIQTNLSHVRTRDIIEVDLNSFIAWDFDIMSRFE DVELADTLPEGDRE-ALWAELKAGAESGWDFSSRWLIGCPNPNSLSGIRTSKLVPVDLNAFLCQAEELMSNFY DTELAHTLPEGSWE-TLWAELKAGAESGWDFSSRWLVGSPNPDSLGSIRTSKLVPVDLNAFLCQAEELLSGFY DVTTASVFSDERDKAELYMDLKSAAESGWDFSSRWIVDEYCGTRGNLSALHTRRIIEVDLNAFLCQAFQKLSEFY DYELAQKLDKNTDPNDIYADLKSAAESGWDFSTRWFISESCDNSGNLTNLNTKNVIEVDLNAFLCQAFQKLSFY	369 364 364 363 345
AR HO OR TE BO	KQLCRDNASVVYSDLYSEWKTSINAILWDDEACSWFDYDSAHRRWNTNFYVSNLTPLFVGCYDPKTVHHEDVATR SRLCNDSQATKYRILRSQRLAALNTVLWDEQTCAWFDYDLEKKKKNREFYPSNLTPLWAGCF-SDP-GVADK SRLCNESQATKYRNLRAQRIAALTALLWDEDKCAWFDYDLENQKKNHEFYPSNLTPLWAGCF-SDP-AIADK QTLCDYPNATFWSKLVKIWQHSIEMVHYNRDDCIWYDWDNELSQHRRMFFPSNFAPLWSETFDSRNAEILGEM AILKNPRRAAHWGYMAEQWRSSIEQALWDEEDCVWHDYDILNNKPRYFYTSNLAPLWMNAVEKPFLAKHGAR	434 434 436
AR HO OR TE BO	Trehalase signatute 2 VLDYLEKSNALKFPGGVPTSLMQTSQQWDFPNGWPPLQHMLVMGLDKTGDPRAKELAFDVAQRMVFNNYEAFTQS ALKYLEDNRILTYQYGIPTSLQKTGQQWDFPNAWAPLQDLVIRGLAKAPLRRAQEVAFQLAQNWIRTNFDVYSQ- ALQYLQDSQLINHRHGIPTSLQNTGQQWDFPNAWAPLQDLVIRGLAKSPSARTQEVAFQLAQNWIRTNFDVYSQ- AAEYFITQNMMDYHGGIPTSLSHTGEQWDYPNAWPPMQSIIVMGLDKSGSYRAKQLARELARRW-VKANLIGF-R VLEYLHESQALEYPGGVPVSLVNSGEQWDFPNAWPPEVSIVVTAIQNIGSEESSKLAKELAQVW-VRACKSGF-T	508
AR HO OR TE BO	LPNAMFEKYDVTVVCLPGGGGEYDVOLGFGWTNGVVMDFLVKYGERUTSVRDAKQLIWSPSSVSVGVGAGATFLG -KSAMYEKYDVSNGCQPGGGGEYEVOEGFGWDEGVVLMLLDRYGDRUTSGAKLAFLE-PHCLAATLLPSLLLSLL -RSAMYEKYDISN-AQPGGGGEYEVOEGFGWTNGVALMLLDRYGDRLSSGTQLALLE-PHCLAAALLLSFLTR QTGEMFEKYNVEVPCQNGGGGEYVVOSGFGWTNGVVLEFINQFFTT	581 578 555
AR HO OR TE BO	AW <mark>IGLASTRFYRQLFKSDHKMGGIHGVAEAIQGLIACVGSLLVATAALFSEKDTTKRPTSGQTPQSSPIRRTKSL</mark> PW	669 583
AR HO OR TE BO	PGFLLFQRGSKGDEEHLLRNRAISEECSYSSIEE	703

**Fig. 3.** Alignment of the animal trehalases. The amino acid sequences of precursors of animal trehalases were aligned by the Clustal W program (Thompson *et al.*, 1994). Predicted N-terminal cleavage sites are indicated by arrowheads. Identical regions among the trehalases are printed in reverse. Positions of trehalase signatures are indicated by bold lines. The position of N-glycosylation site conserved among arthropod trehalases is indicated by a blank arrowhead. A unique, long C-terminal polypeptide of the *Artemia* trehalase is boxed. AR, *Artemia*; HO, human; OR, rabbit; TE, mealworm beetle; BO, silkworm.

teins deduced from *Tre*E2 and *Tre*E3 comprised 703 amino acids and showed the same molecular mass of 79,995 Da and the same isoelectric point of 5.22. Proteins encoded by these cDNAs were identical except for replacements of two amino acids (Fig. 2, *Tre*E2: Thr<sup>246</sup>  $\rightarrow$  *Tre*E3: Ser<sup>246</sup> and *Tre*E2: Val<sup>504</sup> $\rightarrow$ *Tre*E3: Leu<sup>504</sup>).

#### Characterization of Artemia trehalase cDNA

In a computer-assisted search of databases for SWISS-PROT and PIR, Artemia trehalase showed no significant similarity to other proteins than trehalases. The protein was highly similar to animal trehalases: 46.6% to human (Ishihara et al., 1997), 46.1% to rabbit (Ruf et al., 1990), 43.4% to mealworm beetle (Takiguchi et al., 1992), and 42.6% to silkworm (Su et al., 1993, 1994). The identities of the Artemia trehalase to trehalases of E. coli (Gutierrez et al., 1989; Horlacher et al., 1996) and neutral trehalases of fungi (Kopp et al., 1993, 1994; Feldmann et al., 1994; Wolfe and Lohan, 1994; Nwaka et al., 1995; Amaral et al., 1997; Eck et al., 1997) were relatively low (24.6–32.0%). No significant similarity was observed between the Artemia trehalase and acid trehalases of fungi, such as ATH1 gene product of S. cerevisiae (Destruelle et al., 1995) and treA gene product of Aspergillus nidulans (d'Enfert and Fontaine, 1997).

The weight matrix method described by von Heijne (1986) showed the presence of a cleavable signal sequence consisting of 19 amino acids at the amino(N)-terminal end of the *Artemia* trehalase (Fig. 2, 3). A cell adhesion motif, Arg-Gly-Asp (reviewed by Ruoslahti and Pierschbacher, 1986, and D'Souza *et al.*, 1991) was found next to the N-terminal cleavable site as shown in Fig. 2. The protein had four potential N-glycosylation sites (Fig. 2). Alignment of animal trehalases showed that the location of the third potential N-glycosylation site was conserved among arthropods (Fig. 3). Trehalase signatures submitted to the PROSITE database (Henrissat and

Bairoch 1993) were found on the deduced amino acid sequences (Fig. 2, 3). A unique, long polypeptide was found at the carboxyl terminus (Fig. 2, 3, 4). The PSORT program (Nakai and Kanehisa, 1992) strongly suggested that the carboxyl(C)-terminal polypeptide contained a transmembrane region (Gly<sup>620</sup>-Glu<sup>645</sup>, Fig. 2, 4). The predicted transmembrane region was highly hydrophobic as was the N-terminal signal sequence (Fig. 4). A potential cAMP-dependent phosphorylation site was included in a cytoplasmic region of the C-terminal polypeptide (Fig. 2).

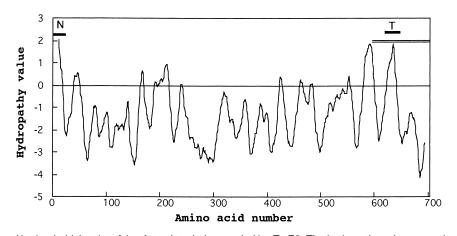
#### Phylogenetic analysis

Phylogenetic trees were drawn by the neighbor-joining method (Saitou and Nei, 1987) (Fig. 5). The results revealed that all the trehalases could be classified into three groups as follows: animal trehalases including the *Artemia* trehalase, trehalases of *E. coli*, and neutral trehalases of fungi (Fig. 5A). Genetic distances between the *Artemia* trehalase and mammalian trehalases were shorter than those between the *Artemia* trehalase and trehalase and trehalases of insects (Fig. 5B).

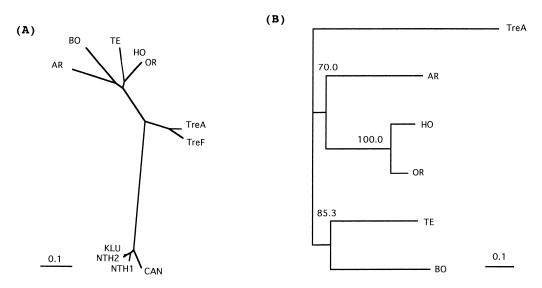
#### Developmental expression of Artemia trehalase

Since the nucleotide sequences of *Tre*E2 and *Tre*E3 were almost identical to each other, the labelled *Tre*E3 was used as a probe to detect both types of mRNA.

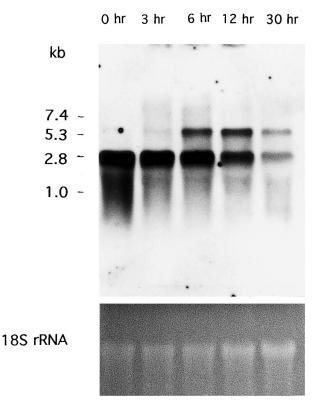
A 2.6-kb mRNA was detected in all the samples of *A. franciscana* (Fig. 6). The 2.6-kb transcript was abundant in the dormant (0 hr) cysts and embryos at 3 hr and 6 hr of the culture (Fig. 6). The transcript then decreased and was weakly detected in the 30 hr sample. A 5.0-kb transcript was observed after 3 hr, which increased by 6-12 hr and decreased thereafter. The amount of the 5.0-kb transcript was lower than that of the 2.6-kb transcript in all the samples.



**Fig. 4.** Hydrophilicity and hydrophobicity plot of the *Artemia* trehalase coded by *Tre*E2. The hydropathy value was calculated using the method of Engelman *et al.*, (1986). Areas showing negative hydropathy values represent hydrophilic regions. Unique C-terminal region is indicated by double lines. Cleavable N-terminal signal sequence (N) and a transmembrane region (T) predicted by PSORT program are indicated by bold lines.



**Fig. 5.** Phylogenetic relationship among trehalases. A neighbor-joining tree of trehalase is represented in (**A**). AR, *Artemia* trehalase; HO, human trehalase; OR, rabbit trehalase; TE, mealworm beetle trehalase; BO, silkworm trehalase; TreA, *Escherichia coli* periplasmic trehalase; TreF, *E. coli* cytoplasmic trehalase; NTH1, *Saccharomyces cerevisiae* neutral trehalase; NTH2, protein product of *S. cerevisiae* NTH2 (*YBR0106*); CAN, *Candida albicans* neutral trehalase; KLU, *Kluyveromyces lactis* neutral trehalase. Detailed neighbor-joining tree among the animal trehalases is represented in (**B**) by employing TreA of *E. coli* as an outgroup. Scale bar indicates 0.1 of branch length. Values on each branching represents percent of bootstrap probability.



# **Fig. 6.** Ontogenic changes of gene expression of *Artemia* trehalase. Total RNA samples were extracted from *A. franciscana* at indicated periods of post-dormant development, and $20 \,\mu g$ of each RNA sample was electrophoresed, transferred to nylon membrane, and hybridized with the DIG-labelled PCR probe. Positions of molecular weight markers are indicated. 18S ribosomal RNA is detected by ethydium bromide staining and shown at the bottom.

#### DISCUSSION

In the present report, we described cloning and characterization of the trehalase cDNAs from crustacean Artemia. Most parts these two cDNAs, termed TreE2 and TreE3, were the same and both contained the same ORF of 2109 nucleotides. Seven replacements were found at the nucleotide level between TreE2 and TreE3, resulting in two replacements on coded 703 amino acids. The rate of change at the nucleotide level between TreE2 and TreE3 was 0.3%, much lower than those among alleles of Na/K-ATPase α1 subunit of A. franciscana reported by García-Sáez et al. (1997). Therefore, TreE2 and TreE3 appear to be derived from point-mutated alleles of one trehalase gene rather than from two different trehalase genes. Sequencing analyses of the PCR products showed that the mRNAs were in the ratio of 1:1 (data not shown). As the dried cysts which we used in the present report were from a wild population, rather than from an established strain, variants of trehalase gene might be observed in individuals or in sub-populations.

The deduced amino acid sequence had two "trehalase signatures" (Fig. 2, 3) and showed high similarities to animal trehalases (Fig. 3), revealing itself as a trehalase.

The presence of potential N-glycosylation sites on the predicted peptide is consistent with our previous conclusion that the *Artemia* trehalase is a glycoprotein (Nambu *et al.*, 1997a).

The most significant feature of the *Artemia* trehalase is the unique, long C-terminal polypeptide. The region consisted of 107 amino acids (Ile<sup>597</sup>-Glu<sup>703</sup>) and had no significant similarity to any other proteins. The predicted transmembrane region (Gly<sup>620</sup>-Glu<sup>645</sup>) was present on the C-terminal polypeptide (Fig. 2, 4) and about 60 amino acids from the C-terminal region were possibly located in the cytoplasm. Membranebound trehalases have been reported in silkworm (Azuma and Yamashita, 1985; Takesue *et al.*, 1989) and mammals (Takesue *et al.*, 1986; Ruf *et al.*, 1990; Sasai-Takedatsu *et al.*, 1996; Oesterreicher *et al.*, 1998). The presence of the cell adhesion motif also suggests the association of *Artemia* tre-halase to the cell membrane.

The predicted molecular mass of the Artemia trehalase without the N-terminal signal sequence was 77,936 Da and substantially larger than that of the purified soluble protein, 70 kDa (Nambu et al., 1997a). The molecular mass of the soluble trehalase decreased to 66 kDa after endoglycosidase H treatment in our previous study (Nambu et al., 1997a). The solubility of the Artemia trehalase is incompatible with the presence of the predicted transmembrane region. As similar sequences to the C-terminal polypeptide of the Artemia trehalase were not found among other trehalases (Fig. 3), it is probable that the C-terminal polypeptide consisting of 107 amino acids is not necessary for trehalase activity and might be removed. The molecular mass of the deduced Artemia trehalase without the N-terminal signal sequence and the C-terminal polypeptide was calculated to be 66,272 Da, showing a good agreement with the molecular mass of the purified Artemia trehalase after the endoglycosidase H treatment.

A possible cAMP-dependent phosphorylation site was found on the predicted C-terminal cytoplasmic region of the *Artemia* trehalase, although we do not have any evidences that the *Artemia* trehalase is regulated by cAMP-dependent phosphorylation. The cAMP-dependent phosphorylation site has not been found in other animal trehalases. In yeasts, cAMP-dependent phosphorylation sites in the N-terminal region and cAMP-dependent regulation of yeast trehalase have been reported (Kopp *et al.*, 1993, 1994; Amaral *et al.*, 1997; Eck *et al.*, 1997). Although the site may be removed by the post-translational processing discussed above, it is possible that the potential cAMP-dependent phosphorylation site plays a role in an immature *Artemia* trehalase.

Phylogenetic trees of trehalase are shown in Fig. 5. On the phylogenetic tree of animal trehalase, the *Artemia* trehalase was closely related to the mammalian trehalases (Fig 5B). The distributions and the physiological meanings of the arthropod trehalases were reported to be different from each other (Azuma *et al.*, 1985; Takesue *et al.*, 1989; Takiguchi *et al.*, 1992; Yaginuma *et al.*, 1996; Nambu *et al.*, 1997a). Adaptation of each trehalases to different roles and physiological environments might result in their variety in amino acid sequences.

The results of Northern blot analysis confirmed the presence of the two transcripts previously described (Fig. 6). The 2.2 kb mRNA of the *Artemia* trehalase found in the previous report (Nambu *et al.*, 1997b) was not detectable when we changed a batch of the cysts.

Three working hypotheses on the two transcripts might be proposed as follows. First, the transcripts of *Artemia* trehalase might originated from alternative splicing of one gene. The second possibility is that the 5.0 kb transcript is a precursor and processed into the 2.6 kb transcript. The third possibility is that the two transcripts are derived from two highly homologous genes.

The abundance of the 2.6 kb transcript in the dormant cysts of *A. franciscana* suggests its importance in early period of the post-dormant development. It may be called as a cryptic RNA. It was observed that 18S rRNA gradually increased during development (Fig. 6). This means that the ratio of rRNA to total RNA is relatively low in the dormant cysts. This observation may be partly due to accumulations of cryptic RNA, such as trehalase mRNA, in the dormant cysts.

The activity of the trehalase was low in the hydrated cysts and began to increase after 6 hr in culture (Nambu *et al.*, 1997a) in spite of the abundant presence of the 2.6 kb transcript in the dormant cysts. The discrepancy between the amount of the mRNA and the trehalase activity suggests a probable presence of post-transcriptional regulation of the enzyme. Similar inconsistency between the amount of mRNA and protein products have been reported in the early period of post-dormant development of *Artemia* (Fisher *et al.*, 1986; Díaz-Guerra *et al.*, 1989).

The Artemia embryos have mechanisms inactivating protein synthesis during their dormancy. The presence of protein synthesis inhibitor (Warner *et al.*, 1977), a lack of template activity of some mRNAs in the dormant cysts by restricted location of mRNAs into mRNP particles (Grosfeld and Littarue, 1975), and low levels of initiation factors (Sierra *et al.*, 1974) have been reported. Moens and Kondo (1976) reported that the specific template activity of the polysomes in the cysts was low and increased linearly up to 13 hr of incubation. The discrepancy between the amount of mRNA and the trehalase activity in the embryos of *A. franciscana* might be explained by such inhibitions.

The appearance of the 5.0 kb transcript in 3 hr embryos of *A. franciscana* is a sign of initiation of post-dormant transcription of the *Artemia* trehalase. If the 5.0 kb transcript is the precursor of the 2.6 kb transcript, the appearance of the 5.0 kb transcript indicates a supplement of the 2.6 kb transcript. The amount of the 5.0 kb transcript increased in accordance with the increase in the trehalase activity (Nambu *et al.*, 1997a). Therefore, it is still possible that transcriptional regulation of the 5.0 kb mRNA directly controls the trehalase activity.

The structural and functional differences of the two transcripts and the mechanisms of the regulation of the trehalase activity are open for future investigation. Further analysis of transcriptional or translational regulations of trehalase would help us to understand the trehalose metabolism in resuming development of *Artemia*.

#### ACKNOWLEDGMENTS

We are deeply indebted to Drs. Shimada, Akasaka and Mitsunaga-Nakatsubo, Hiroshima University, for their helpful discussions and supports. Thanks are also offered to Drs. Satoh and Makabe and their co-workers, Kyoto University, for their advice on phylogenetic analysis of trehalases.

#### REFERENCES

- Amaral FC, Van Dijck P, Nicoli JR, Thevelein JM (1997) Molecular cloning of the neutral trehalase gene from *Kluyveromyces lactis* and the distinction between neutral and acid trehalase. Arch Microbiol 167: 202–208
- Azuma M, Yamashita O (1985) Immunohistochemical and biochemical localization of trehalase in the developing ovaries of the silkworm, *Bombyx mori*. Insect Biochem 15: 589–596
- Clegg JS (1962) Free glycerol in dormant cysts of the brine shrimp *Artemia salina*, and its disappearance during development. Biol Bull Woods Hole 123: 295–301
- Clegg JS (1964) The control of emergence and metabolism by external osmotic pressure and the role of free glycerol in developing cysts of *Artemia salina*. J Exp Biol 41: 879–892
- Clegg JS (1965) The origin of trehalose and its significance during the formation of encysted dormant embryos of *Artemia salina*. Comp Biochem Physiol 14: 135–143
- Crowe JH, Crowe LM, Carpenter JF, Aurell Wistrom C (1987) Stabilization of dry phospholipid bilayers and proteins by sugars. Biochem J 242: 1–10
- Crowe JH, Crowe LM, Carpenter JF, Rudolph AS, Aurell Wistrom C, Spargo BJ, Anchordoguy TJ (1988) Interaction of sugars with membranes. Biochim Biophys Acta 947: 367-384
- d'Enfert C, Fontaine T (1997) Molecular characterization of the *Aspergillus nidulans treA* gene encoding an acid trehalase required for growth on trehalose. Mol Microbiol 24: 203–216
- Destruelle M, Holzer H, Klionsky DJ (1995) Isolation and characterization of a novel yeast gene, *ATH1*, that is required for vacuolar acid trehalase activity. Yeast 11: 1015–1025
- Díaz-Guerra M, Quintanilla M, Palmero I, Sastre L, Renart J (1989) Differential expression of a gene highly homologous to *c-ras* during the development of the brine shrimp *Artemia*. Biochem Biophys Res Comm 162: 802–808
- D'Souza SE, Ginsberg MH, Plow EF (1991) Arginyl-glycyl-aspartic acid (RGD): a cell adhesion motif. TIBS 16: 246–250
- Eck R, Bergmann C, Ziegelbauer K, Schönfeld W, Künkel W (1997) A neutral trehalase gene from *Candida albicans*: molecular cloning, characterization and disruption. Microbiol 143: 3747–3756
- Engelman DM, Steitz TA, Goldman A (1986) Identifying nonpolar transbilayer helices into amino acid sequence of membrane proteins. Annu Rev Biophys Biophys Chem 6: 120–121
- Engler-Blum G, Meier M, Frank J, Müller GA (1993) Reduction of background problems in nonradioactive Northern and Southern blot analyses enables higher sensitivity than <sup>32</sup>P-based hybridizations. Anal Biochem 210: 235–244
- Ewing RD, Clegg JS (1969) Lactate dehydrogenase activity and anaerobic metabolism during embryonic development in *Artemia salina*. Comp Biochem Physiol 31: 297–307
- Feldmann H, Aigle M, Aljinovic G *et al.* (1994) Complete DNA sequence of yeast chromosome II. EMBO J 13: 5795–5809
- Fisher JA, Baxter-Lowe LA, Hokin LE (1986) Regulation of Na, K-ATPase biosynthesis in developing *Artemia salina*. J Biol Chem 261: 515–519
- García-Sáez A, Perona R, Sastre L (1997) Polymorphism and structure of the gene coding for the α1 subunit of the *Artemia franciscana* Na/K-ATPase. Biochem J 321: 509–518
- Grosfeld H, Littauer UZ (1975) Cryptic form of mRNA in dormant Artemia salina cysts. Biochem Biophys Res Comm 67: 176–181
- Gutierrez C, Ardourel M, Bremer E, Middendorf A, Boos W, Ehmann U (1989) Analysis and DNA sequence of the osmoregulated *treA* gene encoding the periplasmic trehalase of *Escherichia coli* K12. Mol Gen Genet 217: 347–354
- Henrissat B, Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities.

Biochem J 293: 781–788

- Horlacher R, Uhland K, Klein W, Ehrmann M, Boos W (1996) Characterization of a cytoplasmic trehalase of *Escherichia coli*. J Bacteriol 178: 6250–6257
- Ishihara R, Taketani S, Sasai-Takedatsu M, Kino M, Tokunaga R, Kobayashi Y (1997) Molecular cloning, sequencing and expression of cDNA encoding human trehalase. Gene 202: 69–74
- Kopp M, Müller H, and Holzer H (1993) Molecular analysis of the neutral trehalase gene from *Saccharomyces cerevisiae*. J Biol Chem 268: 4766–4774
- Kopp M, Nwaka S, Holzer H (1994) Corrected sequence of the yeast neutral trehalase-encoding gene (*NTH1*): biological implications. Gene 150: 403–404
- Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning: A Laboratory Manual (1st ed.). Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York
- Moens L, Kondo M (1976) Polysome-dependent synthesis of embryonic proteins of *Artemia salina* during cell differentiation and analysis of heme-containing protein. Dev Biol 49: 457–469
- Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics14: 897–911
- Nambu Z, Nambu F and Tanaka S (1997a) Purification and characterization of trehalase from *Artemia* embryos and larvae. Zool Sci 14: 419–427
- Nambu Z, Tanaka S, Nambu F (1997b) Gene expression of trehalase during post-dormant development of the brine shrimp, *Artemia*: Comparison of the two species. J Univ Occupational Environmental Health 19: 255–264
- Nwaka S, Kopp M, Holzer H (1995) Expression and function of the trehalase genes *NTH1* and *YBR0106* in *Saccharomyces cerevisiae*. J Biol Chem 270: 10193–10198
- Nwaka S, Holzer H (1998) Molecular biology of trehalose and the trehalases in the yeast *Saccharomyces cerevisiae*. Prog Nuc Acid Res Mol Biol 58: 197–237
- Oesterreicher TJ, Nanthakumar NN, Winston JH, Henning SJ (1998) Rat trehalase: cDNA cloning and mRNA expression in adult rat tissues and during intestinal ontogeny. Am J Physiol 274: R1220– R1227
- Ruf J, Wacker H, James P, Maffia M, Seiler P, Galand G, von Kieckebusch A, Semenza G, Mantei N (1990) Rabbit small intestinal trehalase. Purification, cDNA cloning, expression, and verification of glycosylphosphatidylinositol anchoring. J Biol Chem 265: 15034–15039
- Ruoslahti E, Pierschbacher MD (1986) Arg-Gly-Asp: A versatile cell recognition signal. Cell 44: 517–518
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual (3rd ed.). Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York
- Sasai-Takedatsu M, Taketani S, Nagata N, Furukawa T, Tokunaga R, Kojima T, Kobayashi Y (1996) Human trehalase: Characterization, localization, and its increase in urine by renal proximal tubular damage. Nephron 73: 179–185
- Sierra JM, Meier D, Ochoa S (1974) Effect of development on the translation of messenger RNA in *Artemia salina* embryos. Proc Natl Acad Sci USA 71: 2693–2697
- Su Z-H, Sato Y, Yamashita O (1993) Purification, cDNA cloning and Northern blot analysis of trehalase of pupal midgut of the silkworm, *Bombyx mori*. Biochim Biophys Acta 1173: 217–224.
- Su Z-H, Ikeda M, Sato Y, Saito H, Imai K, Isobe M, Yamashita O (1994) Molecular characterization of ovary trehalase of the silkworm, *Bombyx mori* and its transcriptional activation by diapause hormone. Biochim Biophys Acta 1218: 366–374
- Takesue Y, Yokota K, Nishi Y, Taguchi R, Ikezawa H (1986) Solubilization of trehalase from rabbit renal and intestinal brush-border membranes by a phosphatidylinositol-specific phospholipase C.

FEBS Lett 201: 5-8

- Takesue Y, Yokota K, Miyajima S, Taguchi R, Ikezawa H (1989) Membrane anchors of alkaline phosphatase and trehalase associated with the plasma membrane of larval midgut epithelial cells of the silkworm, *Bombyx mori.* J Biochem (Tokyo) 105: 998– 1001
- Takiguchi M, Niimi T, Su Z-H, Yaginuma T (1992) Trehalase from male accessory gland of an insect, *Tenebrio molitor*. DNA sequencing and developmental profile of the gene expression. Biochem J 288: 19–22
- Thevelein JM (1984) Regulation of trehalose mobilization in fungi. Microbiol Rev 48: 42–59
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc Acids Res 22: 4673–4680

- von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nuc Acids Res 14: 4683–4690
- Warner AH, Shridhar V, Finamore FJ (1977) Isolation and partial characterization of a protein synthesis inhibitor from brine shrimp embryos. Can J Biochem 55: 965–974
- Wolfe KH, Lohan AJE (1994) Sequence around the centromere of *Saccharomyces cerevisiae* chromosome II: Similarity of *CEN2* to *CEN4*. Yeast 10: S41–S46
- Yaginuma T, Mizuno T, Mizuno C, Ikeda M, Wada T, Hattori K, Yamashita O, Happ GM (1996) Trehalase in the spermatophore from the bean-shaped accessory gland of the male mealworm beetle, *Tenebrio molitor*: purification, kinetic properties and localization of the enzyme. J Comp Physiol B 166: 1–10

(Received September 17, 1998 / Accepted December 14, 1998)