Mol Gen Genet (1989) 217:347–354

MGG
© Springer-Verlag 1989

Analysis and DNA sequence of the osmoregulated *treA* gene encoding the periplasmic trehalase of *Escherichia coli* K12

Claude Gutierrez[1], **Maryvonne Ardourel**[1], **Erhard Bremer**[2], **Anke Middendorf**[2], **Winfried Boos**[2], **and Ulrike Ehmann**[2]

[1] Centre de Recherche de Biochimie et de Genetique Cellulaire du C.N.R.S., 118 route de Narbonne, F-31062 Toulouse Cedex, France
[2] Department of Biology, University of Konstanz, P.O. Box 5560, D-7750 Konstanz, Federal Republic of Germany

Summary. The *treA* gene of *Escherichia coli* K12 codes for a periplasmic trehalase that is induced by growth at high osmolarity. The position of *treA* within a cloned chromosomal DNA fragment was identified by subcloning of restriction fragments and analysis of the gene product in minicells. The nucleotide sequence of the *treA* coding region as well as its upstream control region was determined. The *treA* gene consists of 1695 bp encoding 565 amino acids. The amino-terminus of the mature trehalase was found to begin with the amino acid Glu at position 31 of the open reading frame. The first 30 amino acids resemble a typical signal sequence, consistent with trehalase being a secreted periplasmic enzyme. Two previously isolated *phoA* fusions to the *osmA* gene were transferred by homologous recombination on to a *treA*-containing plasmid and found to be within *treA*. Analysis of the hybrid genes and their gene products aided the localization of *treA* and the determination of its direction of transcription within the cloned chromosomal segment. The *treA-phoA* fusions encoded hybrid proteins which could be found in the periplasm. We found that at high osmolarity the normal pathway for the uptake and utilization of trehalose is blocked. Therefore, the function of the periplasmic trehalase is to provide the cell with the ability to utilize trehalose at high osmolarity by splitting it into glucose molecules that can subsequently be taken up by the phosphotransferase-mediated uptake system.

Key words: Osmoregulation – Periplasmic protein – Signal sequence – *phoA* fusions

Introduction

Trehalose is a nonreducing disaccharide composed of two glucose moieties that are linked alpha-glucosidically. This sugar plays an important role as a reserve carbohydrate in yeast and other organisms (Panek 1969). Trehalase (EC 3.2.1.28), the enzyme that liberates glucose from trehalose during the onset of vegetative growth, has been studied intensively and is known to be regulated by cAMP (Thevelein 1984). The gram-negative bacterium *Escherichia coli* K12 can use trehalose as its sole source of carbon and energy. Its specific utilization has been described as phosphotransferase (PTS)-mediated uptake presumably as trehalose-6-phosphate (Maréchal 1984), a conclusion that has been challenged by Postma et al. (1986). Further metabolism of internal trehalose-6-phosphate has been claimed by Maréchal (1984) to occur by a trehalose-6-phosphate hydrolase releasing glucose and glucose-6-phosphate. Our unpublished results do not support this conclusion. They rather indicate hydrolysis in the cytoplasm of trehalose-6-phosphate to trehalose and subsequently to glucose.

E. coli can hydrolyse trehalose by a periplasmic trehalase. This enzyme is induced under conditions of high osmolarity in the growth medium but not by trehalose (Boos et al. 1987). Very likely, this enzyme provides the cell with an alternate pathway for trehalose utilization at high osmolarity, when it splits trehalose in the periplasm into glucose molecules which can then enter the cell as glucose-6-phosphate via a PTS-mediated uptake system for glucose. In addition, *E. coli* also contains an effective system for the internal synthesis of trehalose under conditions of high osmolarity in the growth medium (Giaever et al. 1988); this synthesis represents one of several cellular osmoprotective reactions (Strom et al. 1986). The biochemical basis for the internal synthesis of trehalose is known: trehalose synthetase, an enzyme that is induced under growth conditions of high osmolarity and that is activated in the presence of high concentrations of potassium ions, transfers glucose from UDP-glucose to glucose-6-phosphate, yielding trehalose-6-phosphate (Giaever et al. 1988). The latter is further hydrolysed to trehalose, presumably by a specific phosphatase.

The isolation of gene fusions to *phoA*, the structural gene for the periplasmic alkaline phosphatase, led to the identification of several genes (*osmA* to *osmK*) that code for periplasmic or membrane-bound proteins that are induced under growth conditions of high osmolarity (Gutierrez et al. 1987). Among these, *osmA* was found to lack the periplasmic trehalase (Boos et al. 1987), suggesting that *treA* and *osmA* are the same gene. To learn more about the molecular basis of the complex regulatory phenomena that are connected with the cellular response to osmotic stress and to overproduce trehalase for biochemical studies, we have recently cloned *treA* (Boos et al. 1987). In the present paper we report an analysis of this gene.

Materials and methods

Bacterial strains and plasmids. Bacterial strains and plasmids used in this study are listed in Table 1.

Offprint requests to: W. Boos

Table 1. Bacteria and plasmids

Strain	Description	Origin/Reference
Escherichia coli K12		
MC4100	F^- *araD139 Δ(argF−lac)U169 rpsL150 relA1 deoC1 ptsF25 rbsR flbB5301*	Casadaban (1976)
UE5	Hfr KL16 *thi Δ(ptsHI-crr) galR treA*::Tn*10*	Boos et al. (1987)
MPh2	MC4100 *Δ(brnQ-phoA-proC)*	Gutierrez et al. (1987)
BRE2410[a]	MPh2 *treA*::Tn*10*	This work
CLG1[b]	MPh2 *treA-134*::Tn*phoA*	Gutierrez et al. (1987)
CLG11[c]	MPh2 *treA-259*::Tn*phoA*	C. Gutierrez
CLG190	*araD139 Δ(ara-leu)7697 Δ(lac)X74 galU galK rpsL Δ(malF)3 phoR Δ(phoA)Pvu*II *pcnB zad*::Tn*10 recA1*/ F′*lacI*^{q} *lacZ*$^+$ *Y*$^+$	Laboratory collection
CLG310[d]	CLG1/F′*lacI*^{q} *lacZ*$^+$ *Y*$^+$	This work
KRIM4	MC4100 *treA*::Tn*10 trt*	W. Boos
UE15	MC4100 *treA trt*$^+$	W. Boos
UE41	MC4100 *trt treA*$^+$ *zcg-757*::Tn*10*	W. Boos
SH120	MC4100 *glpR Δ(glpT−glpA) 593 gyrA phoST*	Schweizer and Boos (1983)
DS410T	*minB ara lacY malA mtl xyl rpsL thi fhuA azi Δ(glpT−glpA)593*	Larson et al. (1982)
Plasmids		
pBR322	*bla*$^+$ *tet*$^+$	Bolivar et al. (1977)
pTZ19R	*bla*$^+$ *lacPO*	Mead et al. (1986)
pTRE5	pBR322 *bla*$^+$ *tet*$^+$ *treA*$^+$; Fig. 1A	Boos et al. (1987)
pTRE7	pBR322 *tet*$^+$ *treA*$^+$; Fig. 1A	This work
pTRE8	pBR322 *bla*$^+$ *treA*$^-$; Fig. 1A	This work
pTRE9	pBR322 *tet*$^+$ *treA*$^-$; Fig. 1A	This work
pTRE10	pBR322 *bla*$^+$ *treA*$^+$; Fig. 1A	This work
pTRE11	pBR322 *bla*$^+$ *treA*$^+$; Fig. 1A	This work
pTRE12	pTRE5 *treA-134*::Tn*phoA*; Fig. 1B	This work
pTRE13	pTRE5 *treA-259*::Tn*phoA*; Fig. 1B	This work
pTRE14	pTZ19R *bla*$^+$ *treA*$^+$	This work

[a] This strain was constructed by growing a P1vir lysate on the *treA*::Tn*10* strain UE5 and subsequently transducing the *treA*::Tn*10* insertion into strain MPh2

[b] The Tn*phoA* insertion present in this strain has previously been described as *osmA-134*::Tn*phoA* (Gutierrez et al. 1987)

[c] This PhoA$^+$ Tn*phoA* insertion in *treA* was isolated and characterized as previously described for *osmA-134*::Tn*phoA* (Gutierrez et al. 1987)

[d] This strain is a Lac$^+$ exconjugant from a cross between strain CLG1 and strain CLG190

Media and chemicals. Trehalose MacConkey, MMA and LB media were prepared as described by Miller (1972). Kanamycin (Kan) and ampicillin (Amp) were added to media at 40 μg/ml and 100 μg/ml, respectively. The chromogenic substrate of alkaline phosphatase, 5-bromo-4-chloro-3-indolyl-phosphate (XP), was used at a concentration of 40 μg/ml. PhoA$^+$ colonies developed a blue colour when grown on solid media at pH 7.0 in the presence of this compound.

Genetic procedures. Standard procedures were used for growth of bacteria and bacteriophages, and generalized transduction with phage P1vir was as described by Miller (1972) and Silhavy et al. (1984). The selection for recombination of chromosomal *treA*::Tn*phoA* insertions on to the *treA*$^+$ plasmid pTRE5 was based on the increase in resistance to Kan conferred by the amplification of the Tn*phoA* sequence on a multicopy plasmid (Berg et al. 1983). The *treA*::Tn*phoA* fusion strains CLG1 and CLG11 (Table 1) were transformed with pTRE5 and streaked on LB agar supplemented with XP and containing 300 μg/ml Kan. Large, dark blue colonies which appeared after 24 h of incubation at 37° C were restreaked on the same medium. Plasmids prepared from these clones were used to transform strain CLG190, selecting for Kan^r transformants. The restriction patterns of plasmids prepared from these transformants were determined to verify that these plasmids carried a Tn*phoA* insertion. Strain CLG190 carries a *recA1* and a *pcnB* mutation (Table 1); the latter mutation reduces the copy number of the ColE1-derived plasmids (Lopilato et al. 1986). CLG190 was used as a recipient in the last transformation step in order to avoid a possible deleterious effect of high-level expression of an amplified *treA–phoA* gene fusion.

Since Tn*phoA* is a transposon (Manoil and Beckwith 1985), it was necessary to verify that its insertion into plasmid pTRE5 had resulted from homologous recombination between DNA sequences present both in the *E. coli* chromosome and pTRE5, and not by transposition of Tn*phoA* into this plasmid. Therefore, several independent plasmid candidates were analyzed for the recombination of each *treA*::Tn*phoA* fusion into pTRE5. All plasmids obtained with a particular fusion showed the same restriction pattern, although those obtained from the two different *treA*::Tn*phoA* fusion strains, CLG1 and CLG11, were readily distinguished by restriction analysis. Thus the insertions were assumed to have resulted from homologous recombination.

Methods used with nucleic acids. Isolation of plasmid DNA, restriction digestion of DNA, DNA ligation with T4 ligase, and DNA transformation were carried out as described (Maniatis et al. 1982; Silhavy et al. 1984). The 2.55 kb *Eco*RI-*Bam*HI fragment from pTRE5 that carries *treA* (Fig. 1) was cloned into the multifunctional vector pTZ19R (Mead et al. 1986), resulting in plasmid pTRE14 (Table 1). A series of deletions extending from the *Bam*HI site toward the *Eco*RI site were constructed using exonuclease III/S1 nuclease (Promega). Various DNA fragments were also subcloned into pTZ18R and pTZ19R (Mead et al. 1986). These plasmids were transformed into strain CLG190, and single-stranded DNA was obtained by infecting the resulting strains with the M13KO7 helper phage (Mead et al. 1986). DNA sequences were determined using the enzyme Sequenase (USB Corp.) and ^{35}S-dATP for the labelling, according to the manufacturer's protocol. Double-stranded DNA of the plasmids pTRE12 and pTRE13 (Table 1) was denatured, hybridized with a 13-mer primer specific for the 5′ end of the ′*phoA* sequence carried by Tn*phoA*, and sequenced with Sequenase to determine the *treA*::Tn*phoA* fusion joints. These fusion joints were deduced according

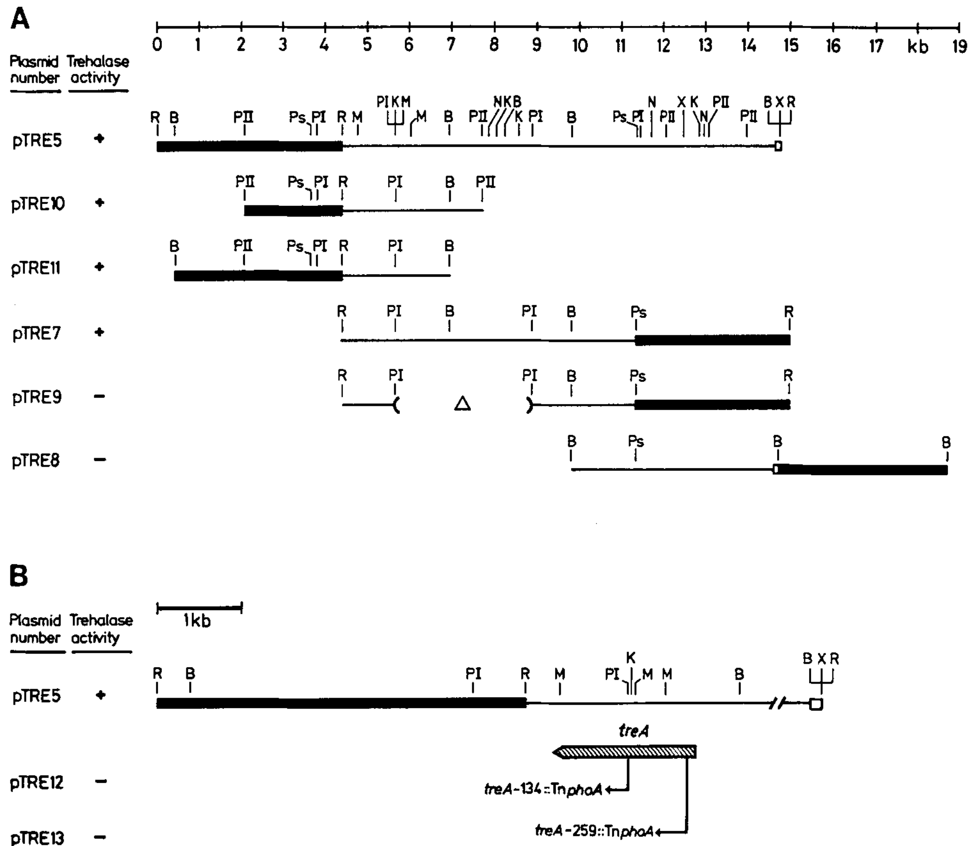

Fig. 1. Structure of the *treA*$^{+}$ plasmid pTRE5 and its derivatives and location of the *treA*::Tn*phoA* insertions. **A** The restriction maps of pTRE5 and its derivatives with their TreA phenotypes are shown. The *heavy lines* represent pBR322 vector DNA and the *thin lines* represent chromosomal DNA. The *open boxes* correspond to a segment of the mini-Mu element pEG5005 (Groisman and Casadaban 1986) which was originally used to clone *treA* (Boos et al. 1987). The *triangle* indicates an in vitro constructed deletion. **B** The direction of transcription of the *treA* gene is indicated by the *hatched arrow* and the positions of the Tn*phoA* insertions are marked by *small arrows*. The other symbols are as in **A**: R, *Eco*RI; B, *Bam*HI; PII, *Pvu*II; Ps, *Pst*I; PI, *Pvu*I; M, *Mlu*I; K, *Kpn*I; N, *Nsi*I; X, *Xma*I

to the Tn*phoA* left-end sequence (Manoil and Beckwith 1985).

Preparation of periplasmic proteins and minicells, gel electrophoresis, and immunological detection of TreA-PhoA hybrid proteins. Periplasmic proteins were prepared as described by Neu and Heppel (1965). The particular strains were grown overnight in 100 ml of LB medium. The shock procedure was done with cells resuspended in 20 ml of the conditioning medium. The final shock fluid was lyophilysed and resuspended in 1 ml of 50 mM TRIS-HCl, pH 7.2, dialysed against the same buffer, and clarified by centrifugation. It routinely contained 2–3 mg of protein/ml. The purification of trehalase and the determination of trehalase activity in crude shock fluids were as described previously (Boos et al. 1987). Polyacrylamide gel electrophoresis with gels containing 12% acrylamide was done according to Laemmli (1970). Immunoblotting was done according to Towbin et al. (1979) using antibodies raised against alkaline phosphatase or trehalase. Minicell preparation from strain DS410T (Table 1) and radiolabelling of newly synthesized proteins with ^{35}S-methionine (1000 mCi/mmol; from Amersham Corp.) were done as described by Boos et al. (1987).

Results and discussion

Location of treA *in pTRE5*

We have previously described a *treA*$^{+}$ plasmid, pTRE5, that carries a 10.3 kb chromosomal *Eco*RI restriction fragment inserted into the unique *Eco*RI site of plasmid pBR322. One end of this fragment originates from an *Eco*RI site present in the chromosomal DNA, and the other end corresponds to an *Eco*RI site in the DNA of the mini-Mu element pEG5005 (Groisman and Casadaban 1986), which was used in the cloning of *treA* (Boos et al. 1987). We first established a restriction map of pTRE5 (Fig. 1A). The *treA* gene was mapped at 26 min on the *E. coli* chromosome, which corresponds to a position of 1260 kb on the *E. coli* chromosomal restriction map (Kohara et al. 1987). The restriction pattern of the chromosomal segment present in pTRE5 is in fairly good agreement with the pattern reported by Kohara et al. (1987) for a DNA fragment extending from positions 1258 kb to 1268 kb on this map.

To locate *treA* within the cloned fragment, we deleted various segments of pTRE5 and also subcloned defined restriction fragments into plasmid pBR322. To detect trehalase activity, we transformed these new plasmids into strain CLG1 carrying the *treA-134*::Tn*phoA* fusion (Gutierrez

et al. 1987) and streaked the transformants on to trehalose MacConkey indicator plates. Strain CLG1 forms pink colonies on these plates; however, when it is transformed with a *treA*$^+$ plasmid like pTRE5, it forms dark red colonies. This phenotype results from the hydrolysis of trehalose by the periplasmic trehalase and the further transport and metabolism of the resulting glucose molecules (Boos et al. 1987). Two plasmids, pTRE10 and pTRE11, were constructed by cleaving pTRE5 with *Pvu*II and *Bam*HI, respectively, and religating. Both plasmids conferred a TreA$^+$ phenotype in strain CLG1 (Fig. 1A). Additional plasmids (pTRE7 to pTRE9; Fig. 1A) were constructed by subcloning segments of pTRE5 into plasmid pBR322 and transformed into strain CLG1. These results narrowed the position of *treA* to the 2.55 kb *Eco*RI-*Bam*HI fragment (position 4.36 kb to 6.91 kb of pTRE5; Fig. 1A), and showed that a segment of *treA* is present between the *Pvu*I and *Bam*HI sites (positions 5.56 kb and 6.91 kb, respectively) of pTRE5 (Fig. 1A). Assays of trehalase activity in crude shock fluids prepared from each of the plasmid-containing strains were used to confirm the TreA phenotypes determined on the trehalose MacConkey indicator plates (data not shown).

Analysis of plasmid-encoded trehalase in minicells

We also expressed the *treA*$^+$ plasmids in minicells and analysed the synthesized proteins by SDS-polyacrylamide gel electrophoresis (PAGE). Purified trehalase is known to have a molecular weight of 58000 on such gels, and pTRE5 has been shown to direct the synthesis of trehalase, as well as that of several other proteins (Boos et al. 1987). A 58 kDa protein that comigrated with purified trehalase was detected in the protein profiles of plasmid-containing strains (Fig. 2B, lane 3). These results show that the various plasmids when present in strain CLG1 not only lead to a TreA$^+$ phenotype but also encode an intact trehalase. It should be noted that trehalase was the only non-vector-encoded protein detected in minicells carrying pTRE11 (Fig. 2). This plasmid, which was used in subsequent DNA sequencing experiments, carries a 2.55 kb chromosomal segment whose coding capacity is only slightly larger than that required for a 58 kDa protein.

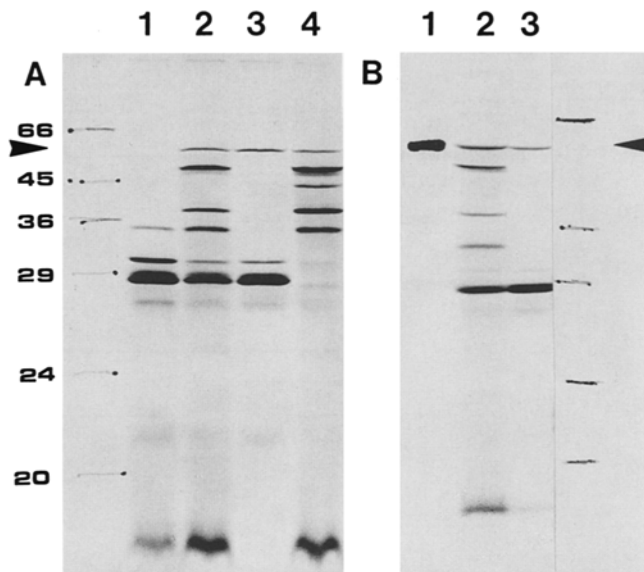

Fig. 2. Plasmid-directed protein synthesis in minicells. Minicells transformed with pBR322 (A, lane 1), pTRE5 (A, lane 2), pTRE10 (A, lane 3), pTRE7 (A, lane 4), pTRE5 (B, lane 2) and pTRE11 (B, lane 3) were labelled with ^{35}S-methionine, dissolved in SDS, and analysed by SDS-PAGE. Molecular weight standards are shown in the left lane of (A) and the right lane of (B); purified trehalase was electrophoresed in lane 1 of (B). The gel was stained with Coomassie blue and dried. The molecular weight standards and purified trehalase were marked with radioactive ink and the gel was autoradiographed. The *arrowheads* indicate the position of the plasmid-encoded trehalase. The *numbers* at the left indicate the molecular weight of the marker proteins in kilodaltons

Recombination of treA::*Tn*phoA *gene fusions on to pTRE5*

Strains CLG1 and CLG11 (Table 1) carry PhoA$^+$ insertions of the transposon Tn*phoA* (Manoil and Beckwith 1985) in a nonessential gene initially named *osmA*, whose expression is stimulated by an increase in the osmotic pressure of the growth medium (Gutierrez et al. 1987). Like *treA*, *osmA* was mapped at 26 min on the *E. coli* chromosome. Furthermore, strains CLG1 and CLG11 lacked the periplasmic trehalase activity (Boos et al. 1987 and our unpublished data). In order to clarify the relationship between *treA* and *osmA*, we transformed strains CLG1 and CLG11 with pTRE5 and transferred the Tn*phoA* insertions by recombination on to this plasmid. The resulting plasmids, pTRE12 and pTRE13, allowed us to map the two Tn*phoA* insertions at different positions within the 2.55 kb chromosomal fragment that carries *treA* (Fig. 1B). One can therefore conclude that *osmA* is identical to *treA*; this was confirmed by subsequent DNA sequence analysis (see below).

We propose to conserve the name *treA* for this gene. Restriction analysis of the *treA*::Tn*phoA* fusions present on plasmids pTRE12 and pTRE13 also allowed us to deduce the direction of transcription of *treA* within the cloned chromosomal fragment (Fig. 1B). According to the position of the *treA*-encoding DNA segment on the *E. coli* chromosomal restriction map (Kohara et al. 1987), this would correspond to a counterclockwise transcription of *treA* on the bacterial genetic map (Bachmann 1987).

Biochemical analysis of the plasmid-encoded TreA-PhoA hybrid proteins

To determine the size of the two plasmid-encoded TreA-PhoA fusion proteins derived from strain CLG1 (carrying *treA-134*::Tn*phoA*) and CLG11 (carrying *treA-259*::Tn*phoA*), we isolated the periplasmic proteins of strains harbouring these two *treA-phoA* fusion plasmids and analysed them by SDS-PAGE (Fig. 3A). The proteins were then transferred to nitrocellulose sheets and probed with antibodies against alkaline phosphatase (Fig. 3B). As can be seen in the figure, the TreA-PhoA 259 fusion protein was slightly larger than mature alkaline phosphatase, whereas the TreA-PhoA 134 fusion protein was larger by about 25 kDa. The latter fusion protein showed extensive proteolytic breakdown. Thus, TreA-PhoA 259 has its fusion joint very early in trehalase, while the TreA-PhoA 134 fusion protein contains about one-third of the amino-terminal portion of trehalase. Since the fusion joint could accurately be determined, the position of the 5′ end of the *treA* coding region on the plasmid DNA became clear (Fig. 1B).

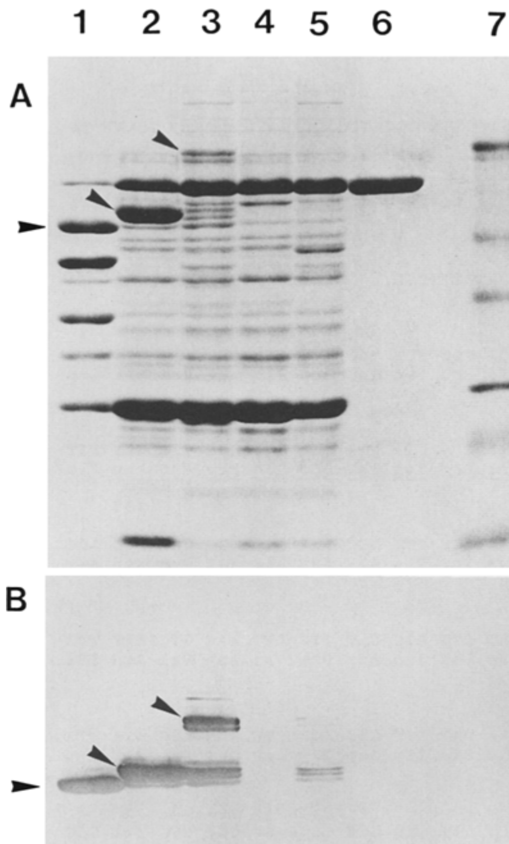

Fig. 3. SDS-PAGE of periplasmic proteins and immunological detection of TreA-PhoA hybrid proteins. **A** The periplasmic proteins of the following strains were applied to a 12% SDS-polyacrylamide gel which was stained with Coomassie blue: SH120, a strain synthesizing alkaline phosphatase constitutively (lane 1); BRE2410 (pTRE13; *treA-259*::Tn*phoA*) (lane 2); BRE2410 (pTRE12; *treA-134*::Tn*phoA*) (lane 3); MPh2, a strain lacking alkaline phosphatase (lane 4); CLG1 (*treA-134*::Tn*phoA*, pTRE5; *treA*$^{+}$) (lane 5); purified trehalase (lane 6). Only the periplasmic protein preparations in lanes 1, 4 and 5 contained trehalase; the protein bands at the position of trehalase in lanes 2 and 3 represent an unrelated polypeptide (Boos et al. 1987). In lane 7 a molecular weight standard is shown. **B** The proteins shown on the gel in A were transferred electrophoretically on to a sheet of nitrocellulose and probed with a rabbit antiserum directed against alkaline phosphatase. The *arrowheads* to the left of **A** the gel and **B** the nitrocellulose sheet mark the position of mature alkaline phosphatase; the other *arrowheads* indicate the position of the TreA-PhoA hybrid proteins

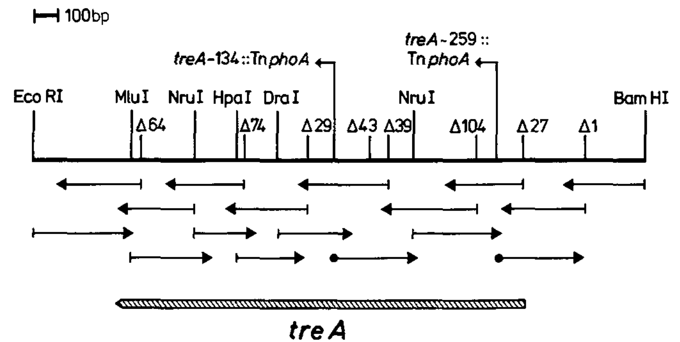

Fig. 4. Strategy for sequencing *treA* and the *treA*::Tn*phoA* fusion joints. Deletions *Δ1* to *Δ64* represent the endpoints of deletions extending from the *Bam*HI site towards the *Eco*RI site. The restriction sites used for the subcloning of DNA fragments are shown and the fusion sites for the two *treA*::Tn*phoA* insertions are indicated. *Thin arrows* show the regions sequenced in each sequencing reaction. *Arrows* originating at a closed circle indicate sequencing reactions performed with denatured double-stranded DNA. The other sequencing reactions were done with single-stranded DNA. The *hatched arrow* indicates the extent of the *treA* coding region

DNA sequence of treA *and the amino-terminus of the mature trehalase*

From plasmid pTRE11, we have sequenced the entire 2.55 kb *Bam*HI-*Eco*RI chromosomal DNA segment that carries *treA* (Fig. 1). The sequencing strategy used is summarized in Fig. 4, and the determined DNA sequence is shown in Fig. 5. Only one open reading frame (ORF) that is long enough to encode a 58 kDa protein can be found in this sequence. This open reading frame is oriented from the unique *Bam*HI site present in pTRE11 towards the *Eco*RI site (Fig. 4), in agreement with the direction of transcription of *treA* as determined from the analysis of the *treA*::Tn*phoA* insertions present in pTRE12 and pTRE13 (Fig. 1B). The full-length ORF starts with an ATG codon (at position 496; Fig. 5) and ends with a TAA stop codon (at position 2191; Fig. 5). The *treA* ORF was also confirmed by sequencing the fusion joints of the two *treA*::Tn*phoA* insertions. They occurred after nucleotides 610 and 1288 for *treA-259*::Tn*phoA* and *treA-134*::Tn*phoA*, respectively (Fig. 5). According to the DNA sequence of the Tn*phoA* left end (Manoil and Beckwith 1985), the created *phoA* fusions are in frame with the proposed *treA* ORF. With an amino-terminus that is identical to that of wild-type trehalase, these fusions would encode hybrid proteins that are larger than alkaline phosphatase by 8 and 234 amino acids. This would increase the size of alkaline phosphatase by 900 and 26839 Da, respectively, in agreement with the sizes of the hybrid proteins determined from their electrophoretic mobility.

Trehalase is a periplasmic enzyme and therefore is most likely synthesized as a precursor molecule. We thus sequenced the amino-terminus of the purified trehalase. The determined sequence with the exception of Pro-34 matches that of the predicted product of the identified ORF from Glu-31 to Leu-43 (Fig. 5), demonstrating that this proposed ORF does encode trehalase. The mature trehalase would then comprise 535 amino acids and its calculated molecular weight would be 60463 Da, in good agreement with the value of 58000 Da for trehalase estimated from its electrophoretic mobility on SDS-polyacrylamide gels (Boos et al. 1987). Upstream of the GAA codon encoding Glu-32 in the *treA* ORF, four codons are found that can function as translational starts (Stormo 1986), a GUG, a UUG, and two AUGs at nucleotides 577, 550, 529 and 496, respectively (Fig. 5). AUG-496 is the only one of these preceded by a putative ribosome-binding site: 5′-AAGGAG-3′ (marked with asterisks in Fig. 5). This sequence is located 8 bp upstream of the AUG codon, a distance close to the consensus spacer in translational starts (Stormo 1986). The amino-terminus of the putative trehalase precursor exhibits the characteristic features of a bacterial signal sequence (von Heijne 1985). An amino-terminal segment with three positive charges (from Met-1 to Lys-11; Fig. 5) is followed by a hydrophobic core of fifteen amino acids. The sequence Val-Gln-Ala-Glu from amino acids 28 to 31 fits the minimal requirement for a signal peptidase cleavage site with a cut between Ala and Glu (von Heijne 1985). Furthermore, this

```
                                                                                                          100
                                                                                                           .
CGGTTGTAAAAGGTATCTGCCTTAGCGCCGGAAGTCCGGTATCCCACAGCGCCCT AATCGCCCGTGAACTGGGGATTGGCTGGATTTGCCAGCAGGGTGAGAAACTGTAT
                                                                                                200
                                                                                                 .
GCGATACAACCAGAAGAAACGCTAACGCTGGACGTTAAAACGCAACGTTTCAACC GTCAGGGTTAAGCCACTGTTCGGATGGCATCGTTCTGATGTCATCCGGCAATTTA
                                                                                        300
                                                                                         .
CTCAGCACTAATGGTTTTTAACTACCTGTTATATTTTTCACCAATTTATGATGAT TTTTCTTCGCTAATCACGAACCCAACGCCGCTGGCTGGCTAATAAACAGACGTTA
                          -35                     -10               400
                                                                     .
TTTGATCACTCTATTTCGCCGTTATTTATCCCTTTAATTCCCTTTTCTAAAATGC CTGACAGTTCGCAGAATGAGATTTCGATCATGCAGCTAGTGCGATCCTGAACTAA
                                                              500
                                                  SD           .
GGTTTTCTGATACTTGAATACCGTTTTATTCGGTTTCGCCAAAGGAGAATGATTG ATG AAA TCC CCC GCA CCT TCT CGC CCG CAA AAA ATG GCG TTA
                                                  *******  Met Lys Ser Pro Ala Pro Ser Arg Pro Gln Lys Met Ala Leu
                                                                           600
                                                                            .               ▼
ATT CCA GCC TGT ATC TTT TTG TGT TTC GCT GCG CTA TCG GTG CAG GCA GAA GAA ACA CCG GTA ACA CCA CAG CCG CCT GAT ATT
Ile Pro Ala Cys Ile Phe Leu Cys Phe Ala Ala Leu Ser Val Gln Ala Glu Glu Thr Pro Val Thr Pro Gln Pro Pro Asp Ile
                                                               ▲                                           700
                                                                                                            .
TTA TTA GGG CCG CTG TTT AAT GAT GTG CAA AAC GCC AAA CTT TTT CCG GAC CAA AAA ACC TTT GCC GAT GCC GTG CCG AAC AGC
Leu Leu Gly Pro Leu Phe Asn Asp Val Gln Asn Ala Lys Leu Phe Pro Asp Gln Lys Thr Phe Ala Asp Ala Val Pro Asn Ser

GAT CCG CTG ATG ATC CTT GCT GAT TAT CGG ATG CAG CAA AAC CAG AGC GGA TTT GAT CTG CGC CAT TTC GTT AAC GTC AAT TTC
Asp Pro Leu Met Ile Leu Ala Asp Tyr Arg Met Gln Gln Asn Gln Ser Gly Phe Asp Leu Arg His Phe Val Asn Val Asn Phe
        800
         .
ACC CTG CCG AAA GAA GGC GAG AAA TAT GTT CCG CCA GAG GGG CAG TCA CTG CGC GAA CAT ATT GAC GGA CTT TGG CCG GTA TTA
Thr Leu Pro Lys Glu Gly Glu Lys Tyr Val Pro Pro Glu Gly Gln Ser Leu Arg Glu His Ile Asp Gly Leu Trp Pro Val Leu
                            900
                             .
ACG CGT TCT ACC GAA AAC ACC GAA AAA TGG GAT TCT CTG TTA CCG CTG CCG GAA CCT TAT GTC GTG CCG GGC GGA CGC TTT CGC
Thr Arg Ser Thr Glu Asn Thr Glu Lys Trp Asp Ser Leu Leu Pro Leu Pro Glu Pro Tyr Val Val Pro Gly Gly Arg Phe Arg
                                                    1000
                                                      .
GAG GTA TAT TAC TGG GAC AGT TAC TTC ACC ATG TTA GGA CTT GCC GAA AGC GGT CAC TGG GAT AAA GTC GCG GAT ATG GTG GCC
Glu Val Tyr Tyr Trp Asp Ser Tyr Phe Thr Met Leu Gly Leu Ala Glu Ser Gly His Trp Asp Lys Val Ala Asp Met Val Ala
                                                                                1100
                                                                                  .
AAT TTT GCT CAT GAA ATA GAC ACT TAC GGT CAT ATT CCC AAC GGC AAC CGC AGT TAC TAT TTA AGC CGC TCG CAA CCG CCC TTC
Asn Phe Ala His Glu Ile Asp Thr Tyr Gly His Ile Pro Asn Gly Asn Arg Ser Tyr Tyr Leu Ser Arg Ser Gln Pro Pro Phe
                                                                                                    1200
                                                                                                      .
TTT GCC CTG ATG GTA GAG TTA CTG GCG CAG CAT GAA GGC GAT GCC GCG TTG AAG CAA TAC CTG CCG CAA ATG CAA AAA GAA TAT
Phe Ala Leu Met Val Glu Leu Leu Ala Gln His Glu Gly Asp Ala Ala Leu Lys Gln Tyr Leu Pro Gln Met Gln Lys Glu Tyr
                                                                                                   ▼
GCT TAC TGG ATG GAC GGT GTT GAA AAC CTG CAA GCC GGA CAA CAG GAA AAA CGC GTT GTC AAA CTT CAG GAT GGT ACC CTT CTC
Ala Tyr Trp Met Asp Gly Val Glu Asn Leu Gln Ala Gly Gln Gln Glu Lys Arg Val Val Lys Leu Gln Asp Gly Thr Leu Leu
    1300
     .
AAC CGC TAC TGG GAC GAT CGC GAT ACG CCA CGA CCA GAG TCA TGG GTG GAA GAT ATT GCC ACC GCC AAA AGC AAT CCG AAT CGA
Asn Arg Tyr Trp Asp Asp Arg Asp Thr Pro Arg Pro Glu Ser Trp Val Glu Asp Ile Ala Thr Ala Lys Ser Asn Pro Asn Arg
                        1400
                          .
CCT GCC ACT GAA ATT TAC CGC GAC CTG CGC TCT GCC GCT GCG TCT GGC TGG GAT TTC AGC TCG CGC TGG ATG GAC AAC CCG CAG
Pro Ala Thr Glu Ile Tyr Arg Asp Leu Arg Ser Ala Ala Ala Ser Gly Trp Asp Phe Ser Ser Arg Trp Met Asp Asn Pro Gln
                                            1500
                                              .
CAG TTA AAT ACC TTA CGC ACC ACC AGC ATC GTA CCG GTC GAT CTG AAC AGC CTG ATG TTT AAA ATG GAA AAA ATC CTC GCC CGC
Gln Leu Asn Thr Leu Arg Thr Thr Ser Ile Val Pro Val Asp Leu Asn Ser Leu Met Phe Lys Met Glu Lys Ile Leu Ala Arg
                                                                    1600
                                                                      .
GCC AGC AAA GCT GCC GGA GAT AAC GCG ATG GCA AAC CAG TAC GAA ACG CTG GCA AAT GCC CGT CAA AAA GGG ATC GAA AAA TAC
Ala Ser Lys Ala Ala Gly Asp Asn Ala Met Ala Asn Gln Tyr Glu Thr Leu Ala Asn Ala Arg Gln Lys Gly Ile Glu Lys Tyr
                                                                                                1700
                                                                                                  .
CTG TGG AAC GAT CAA CAA GGC TGG TAT GCC GAT TAC GAC CTG AAA AGT CAT AAA GTG CGC AAT CAG TTA ACC GCG GCC GCC CTG
Leu Trp Asn Asp Gln Gln Gly Trp Tyr Ala Asp Tyr Asp Leu Lys Ser His Lys Val Arg Asn Gln Leu Thr Ala Ala Ala Leu

TTC CCG CTG TAC GTC AAT GCG GCA GCG AAA GAT CGC GCC AAC AAA ATG GCG ACG GCG ACG AAA ACA CAT CTG CTG CAA CCC GGC
Phe Pro Leu Tyr Val Asn Ala Ala Ala Lys Asp Arg Ala Asn Lys Met Ala Thr Ala Thr Lys Thr His Leu Leu Gln Pro Gly
```

Fig. 5

```
1800
GGC CTG AAC ACC ACG TCG GTG AAA AGT GGG CAA CAA TGG GAT GCG CCA AAT GGC TGG GCA CCG TTA CAG TGG GTC GCG ACA GAA
Gly Leu Asn Thr Thr Ser Val Lys Ser Gly Gln Gln Trp Asp Ala Pro Asn Gly Trp Ala Pro Leu Gln Trp Val Ala Thr Glu

                        1900
GGA TTA CAA AAC TAC GGG CAA AAA GAG GTG GCG ATG GAC ATT AGC TGG CAC TTC CTG ACC AAT GTT CAG CAC ACC TAT GAC CGG
Gly Leu Gln Asn Tyr Gly Gln Lys Glu Val Ala Met Asp Ile Ser Trp His Phe Leu Thr Asn Val Gln His Thr Tyr Asp Arg

                                             2000
GAG AAA AAG CTG GTG GAA AAA TAT GAT GTC AGC ACC ACC GGA ACG GGG GGC GGC GGT GGC GAA TAT CCA TTA CAG GAT GGC TTT
Glu Lys Lys Leu Val Glu Lys Tyr Asp Val Ser Thr Thr Gly Thr Gly Gly Gly Gly Gly Glu Tyr Pro Leu Gln Asp Gly Phe

                                                                  2100
GGC TGG ACC AAT GGC GTG ACG CTG AAA ATG CTG GAT TTG ATC TGC CCG AAA GAG CAA CCG TGT GAC AAT GTT CCG GCG ACG CGT
Gly Trp Thr Asn Gly Val Thr Leu Lys Met Leu Asp Leu Ile Cys Pro Lys Glu Gln Pro Cys Asp Asn Val Pro Ala Thr Arg

                                                                                       2200
CCG ACC GTT AAG TCA GCA ACG ACG CAA CCC TCA ACC AAA GAG GCA CAA CCC ACA CCT TAA CCAGCGCTTACTCCGACAACGTCATTCGGCT
Pro Thr Val Lys Ser Ala Thr Thr Gln Pro Ser Thr Lys Glu Ala Gln Pro Thr Pro ---                ------------->

                                                                               2300
TCTCCGGATGGCGTTGTCGGCAGATAACACCATAGCTATGGTTTTTCTACACAG ACGGTTTTTCAAGTCCTGGCGGGAAGTATTGCGCAAATGTTTGCCATACGTGGTT
  <------------

                                                                  2400
GCCGAAGATAGCGTCATTTCTACGGGTAAGTGTTTCACTTTCATCACCCGGCGAT TTACGCCGACGTTGGCTATATAAAGCATCAATTGCTACCTGAATATTGTTACGGT

                                                        2500
GTTCACTCATCAACGTTTTTTTGCTTTCTGGAGTTAAATTAATAATCACATAATC CAGACACTCCAGAGAATTTAGCAACCAAAAAAAGAATTC
```

Fig. 5. Nucleotide sequence of *treA* and its flanking regions. The nucleotide sequence of the *treA* region and the deduced amino acid sequence for the precursor of trehalase are shown. The putative cleavage site for the signal sequence peptidase is indicated by the *arrowhead below* the line. *Underlined* amino acids were determined experimentally. The ribosome-binding site (SD) is marked by *asterisks* below the line, and a possible promoter region (−10 and −35) is *overlined*. The proposed transcription terminator for *treA* is indicated by the two *arrows below* the line. The *treA-259*::Tn*phoA* and the *treA-134*::Tn*phoA* fusions are marked by *solid triangles* after positions 610 bp and 1288 bp, respectively

putative signal sequence-processing site is highly homologous to the processing site for the signal sequence of the outer membrane protein TolC (Hackett and Reeves 1983). This suggests that the trehalase is synthesized as a precursor larger then the mature protein found exported into the periplasmic space.

A search for similarity was done on the mature trehalase by scanning the "Swissprot" protein sequence bank by the method of Lipman and Pearson (1985). No significant similarity was found in this search.

treA *appears to be transcribed as a single cistron*

A computer search using the matrix of Mulligan et al. (1984) identified several putative promoters within the region of 240 to 480 bp. The −35 and −10 hexanucleotides of the strongest predicted promoter with a homology score of 59% are overlined in Fig. 5. We compared the *treA* control region with those of the osmoregulated *ompC* and *ompF* genes (Inokuchi et al. 1982; Mizuno et al. 1983). No apparent similarity was detected, consistent with the observation that the osmotic control of *treA* expression is independent of the products of the *ompB* locus (Gutierrez et al. 1987). Obviously, more work is needed to identify the actual *treA* promoter and the DNA sequences involved in the osmotic regulation of *treA*.

A potential stem-loop structure is found 11 bp downstream of the *treA* stop codon (Fig. 5). This structure is not followed by a stretch of T's typical for consensus-factor-independent transcription terminators (Brendel et al. 1986), but it should be very stable (ΔG of −28.6 kcal) and it is a good candidate for a transcription terminator element. In any case, no reading frame of a significant size and preceded by a potential translation start site can be found either downstream of this stem-loop structure or upstream of the *treA* translational start. Therefore, it appears that *treA* must be transcribed as a single cistron.

The periplasmic trehalase is essential for growth on trehalose at high osmolarity

As reported earlier (Boos et al. 1987), a mutation in *treA* did not abolish the ability of *E. coli* to grow on trehalose. In view of the observation that *E. coli* produces from internal metabolites large amounts of trehalose at high osmolarity (Giaever et al. 1988), the presence of a periplasmatic trehalase that is induced under the same conditions appeared illogical. This paradox was solved when we observed that *treA* mutants grew on trehalose only at low osmolarity (Table 2). Apparently, uptake and/or further metabolism of trehalose is turned off in the presence of more than 250 mM NaCl in the growth medium. In addition, a mutant (KRIM4) that was unable to grow on trehalose due to a defect in the trehalose uptake system (*trt; t*rehalose trans*port*) was isolated from strain UE15, a *treA* derivative of strain MC4100 (Boos et al.; unpublished results). Introduction of a wild-type *treA* gene into strain KRIM4 by P1 transduction yielded UE41. This strain grows on trehalose at both low and high osmolarity, although with reduced efficiency in comparison to a $trt^{+}treA^{+}$ strain. The effect of different combinations of *treA* and *trt* on the growth rates with 0.2% trehalose as carbon source is shown in Table 2. From this analysis it is clear that the major function of the periplasmic trehalase is to provide a means of

Table 2. Generation time with trehalose as sole source of carbon in different strains grown at low and high osmolarity

Strain	Relevant genotype[a]	Generation time at	
		low osmolarity[b]	high osmolarity[b]
MC4100	trt^+ $treA^+$	63 min	220 min
UE15	trt^+ *treA*	74 min	–[c]
UE41	*trt* $treA^+$	231 min	234 min
KRIM4	*trt treA*	–[c]	–[c]

[a] The strains are all derivatives of MC4100
[b] The growth medium was MMA, 0.2% trehalose, without or with 250 mM NaCl
[c] No growth of the culture was observed

trehalose utilization at high osmolarity by hydrolysing trehalose to glucose, which is then taken up via a PTS-mediated uptake system.

Acknowledgements. We thank D. Boyd and F. Payre for the gift of bacterial strains and bacteriophages and J. Gebert for kindly providing the antiserum against the PhoA protein. We are grateful to R. Frank (EMBL, Heidelberg) for the sequencing of the aminoterminal end of trehalase. We thank J.M. Louarn for helpful suggestions and J.P. Bouche for help with the computer work and many useful discussions. We appreciate the help of V. Koogle in the editing of the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft through SFB 156 and the Fonds der Chemischen Industrie.

References

Bachmann BJ (1987) Linkage map of *Escherichia coli* K-12, Edn. 7. In: Neidhardt FC (ed), *Escherichia coli* and *Salmonella typhimurium.* American Society for Microbiology, Washington, D.C., pp 807–876

Berg D, Schmandt M, Lowe J (1983) Specificity of transposon Tn*5* insertion. Genetics 105:813–828

Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Heynecker HL, Boyer HW, Crosa JH, Falkow S (1977) Construction and characterization of new cloning vehicles: II. A multipurpose cloning system. Gene 2:95–113

Boos W, Ehmann U, Bremer E, Middendorf A, Postma P (1987) Trehalase of *Escherichia coli.* Mapping and cloning of its structural gene and identification of the enzyme as a periplasmic protein induced under high osmolarity growth conditions. J Biol Chem 262:13212–13218

Brendel V, Hamm G, Trifonov E (1986) Terminators of transcription with RNA polymerase from *Escherichia coli:* what they look like and how to find them. J Biomol Str Dyn 3:705–723

Casadaban MJ (1976) Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J Mol Biol 104:541–555

Giaever HM, Styrvold OB, Kaasen I, Strom AR (1988) Biochemical and genetic characterization of osmoregulatory trehalose synthesis in *Escherichia coli.* J Bacteriol 170:2841–2849

Groisman EA, Casadaban MJ (1986) Mini-Mu bacteriophage with plasmid replicons for in vivo cloning and *lac* gene fusing. J Bacteriol 168:357–364

Gutierrez C, Barondess J, Manoil C, Beckwith J (1987) The use of transposon Tn*phoA* to detect genes for cell envelope proteins subject to a common regulatory stimulus. J Mol Biol 195:289–297

Hackett J, Reeves P (1983) Primary structure of the *tolC* gene that codes for an outer membrane protein of *Escherichia coli* K-12. Nucleic Acids Res 11:6487–6495

Inokuchi K, Mutoh N, Matsuyama S, Mizushima S (1982) Primary structure of the *ompF* gene that codes for a major outer membrane protein of *Escherichia coli* K-12. Nucleic Acids Res 10:6957–6968

Kohara Y, Akiyama K, Isono K (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50:495–508

Laemmli UK (1970) Cleavage of structural proteins during assembly of bacteriophage T4. Nature 227:680–685

Larson TJ, Schumacher G, Boos W (1982) Identification of the *glpT*-encoded *sn*-glycerol-3-phosphate permease of *E. coli,* an oligomeric integral membrane protein. J Bacteriol 152:1008–1021

Lipman D, Pearson W (1985) Rapid and sensitive protein similarity search. Science 227:1435–1441

Lopilato J, Bortner S, Beckwith J (1986) Mutations in a new chromosomal gene of *Escherichia coli, pcnB,* reduce plasmid copy number of pBR322 and its derivatives. Mol Gen Genet 205:285–290

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Manoil C, Beckwith J (1985) Tn*phoA:* a transposon probe for protein export signals. Proc Natl Acad Sci USA 82:8129–8133

Maréchal LR (1984) Transport and metabolism of trehalose in *Escherichia coli* and *Salmonella typhimurium.* Arch Microbiol 137:70–73

Mead D, Szczesna-Skorupa E, Kemper B (1986) Single-stranded DNA "blue" T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Engineering 1:67–74

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mizuno T, Chou MY, Inouye M (1983) A comparative study on the genes for three porins of the *Escherichia coli* outer membrane: DNA sequence of the osmoregulated *ompC* gene. J Biol Chem 258:6932–6940

Mulligan M, Hawley D, Entriken R, McClure W (1984) *Escherichia coli* promoter sequences predicted *in vitro* RNA polymerase selectivity. Nucleic Acids Res 12:789–800

Neu HC, Heppel LA (1965) The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J Biol Chem 240:3685–3692

Panek AD (1969) Adenosine triphosphate inhibition of yeast trehalase. J Bacteriol 3:904–905

Postma P, Keizer HG, Koolwijk P (1986) Transport of trehalose in *Salmonella typhimurium.* J Bacteriol 168:1107–1111

Schweizer H, Boos W (1983) Cloning of the *ugp* region containing the structural genes for the *pho* regulon-dependent *sn*-glycerol-3-phosphate transport system of *Escherichia coli.* Mol Gen Genet 192:177–186

Silhavy TJ, Berman ML, Enquist LW (1984) Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Stormo G (1986) Translation initiation. In: Reznikoff W, Gold L (eds) Maximizing gene expression. Butterworth, Stoneham, Mass, pp 195–224

Strom AR, Falkenberg P, Landfald B (1986) Genetics of osmoregulation in *Escherichia coli:* uptake and biosynthesis of organic osmolytes. FEMS Microbiol Rev 39:79–86

Thevelein JM (1984) Regulation of trehalose mobilization in fungi. Microbiol Rev 48:42–59

Towbin H, Staehlin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA 76:4350–4354

von Heijne G (1985) Signal sequences limits of variation. J Mol Biol 184:99–105

Communicated by J. Lengeler

Received October 31, 1988