Bacillus subtilis sucrose-specific enzyme II of the phosphotransferase system: Expression in Escherichia coli and homology to enzymes II from enteric bacteria

AGNES FOUET, MARYVONNE ARNAUD, ANDRE KLIER, AND GEORGES RAPPORT

ABSTRACT Sucrose is transported into Bacillus subtilis cells by way of a phosphotransferase system, which consists of a specific enzyme II, a nonspecific enzyme I, and a histidine-containing phosphocarrier protein. Mutations in the sacP locus abolish the specific transport of sucrose. The B. subtilis sacP gene was cloned and expressed in Escherichia coli, and transformed cells could transport and phosphorylate sucrose. This indicates that the sacP gene product is enzyme II of the sucrose phosphotransferase system of B. subtilis. The nucleotide sequence of the sacP gene was determined and was found to overlap with the sacA gene at the tetranucleotide ATGA, which may allow a translational coupling between sacP and sacA. The two genes are therefore probably organized in an operon structure with the promoter located 5' to sacP gene. The deduced amino acid sequence gave a Mr of 48,945 for the sacP enzyme II polypeptide. The amino acid sequence was compared to that of three other known enteric bacterial enzymes II (β-glucoside-specific enzyme II, mannitol-specific enzyme II, and glucose-specific enzyme II). Homology was found with β-glucoside enzyme II, and well conserved regions were identified through the comparison of the proteins.

The sucrose metabolic system has been proposed as a model for studying the regulation of gene expression in Bacillus subtilis, and biochemical and genetic studies have established that at least nine different loci are involved (1, 2). Three of these loci, sacT, sacP, and sacA, are clustered; sucrose induces the expression of sacP and sacA. sacC codes for sucrose, an endocellular sucrose-6-phosphate hydrolase. It has been cloned, sequenced, and expressed in Escherichia coli, which contains no endogenous sucrase (3, 4).

In B. subtilis, sucrose is transported exclusively by way of a phosphoenolpyruvate-dependent phosphotransferase system (PTS) (5), which involves two general proteins and one specific protein. Enzyme I of the PTS (enzyme I) and histidine-containing phosphocarrier protein, which are cytoplasmic proteins, are required as general phosphocarrier proteins for the transport and phosphorylation of all PTS sugars. Sugar specificity is determined by enzyme II, which is a specific integral membrane protein of the PTS. Cells harboring mutations in the sacP locus are unable to transport sucrose but are able to transport other PTS sugars, indicating that mutations in sacP act directly or indirectly on sucrose-specific enzyme II (II*). The gene for enzyme I has been mapped on the chromosome, and it is not linked to the sacT sacP sacA cluster. The sacT locus has been defined by a single mutation giving rise to the constitutive high level synthesis of both the PTS and sucrase.

In this paper, we describe a plasmid carrying the sacP gene and the complementation of a sacP mutation in a Rec− strain of B. subtilis. Expression of sacP in E. coli was also obtained, demonstrating that sacP corresponds to the structural gene for II*. The first nucleotide sequence, to our knowledge, of the gene for an enzyme II from a Gram-positive bacteria is reported, and the deduced amino acid sequence is presented. Sequence comparisons were made with enzyme II sequences from Gram-negative bacteria, and homology with E. coli β-glucoside-specific enzyme II (II*) is shown.

MATERIALS AND METHODS

Bacterial Strains, Phages, and Plasmids. B. subtilis QB976 (trpC2, sacP1, recE4) was constructed in this laboratory. E. coli strain TG1 (6) was used as a host for pUC9 derivatives (7), for pMK4 derivatives (8), and for the sequencing vector M13mp19 (9) and its derivative phages. E. coli strain HB101 (10) was used for subcloning experiments. E. coli strain Ts19 (Psl*) (11) was used for assay of sucrose uptake.

DNA Preparation and Plasmid Constructions. Plasmids were isolated from E. coli as described by Birnboim and Doly (12). Single-stranded DNA was purified as described by Messing (7). Vectors were linearized and treated with calf intestine phosphatase, and fragment ends were made blunt by using the Klenow fragment of E. coli polymerase I.

Transformation and Selection of Recombinants. Transformation and selection of E. coli and B. subtilis were performed as described (13, 14). MacConkey plates, on which sugar fermentation gives rise to red colonies, were used.

Assay for Sucrose Uptake. The sucrose uptake was measured as described by Delobbe et al. (15).

In Vitro Transcription/Translation of Plasmid-Encoded Proteins. Covalently closed circular plasmids were used as templates for a prokaryotic coupled transcription/translation system as recommended by the manufacturer (Amersham). Products were separated by NaDodSO4/PAGE (10% acrylamide/0.27% bisacrylamide), and [3S]methionine-labeled proteins were visualized by autoradiography.

DNA Sequencing. Nucleotide sequencing by the dideoxy chain termination method (16) was carried out using phage M13mp19 (9) and the synthetic 17-mer primer, dATP[α-35S] (22,200 GBq/mmol) was supplied by Amersham. Overlapping deletions were obtained using the technique described by Dale et al. (17). The experiment was carried out as directed by the manufacturer (International Biotechnologies, New Haven, CT).

Abbreviations: PTS, phosphotransferase system; enzyme I, enzyme I of the PTS; enzyme II, specific integral membrane protein of the PTS; II*, mannitol-specific enzyme II; II*, β-glucoside-specific enzyme II; II*, sucrose-specific enzyme II.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03006).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
RESULTS AND DISCUSSION

Isolation and Characterization of a DNA Fragment Harboring the sacP Gene. The 2.1-kilobase (kb) EcoRI DNA fragment harboring the sacA gene and part of the sacP gene (4) was used as a probe to screen the library constructed in E. coli using the shuttle vector pMK4 (18). One of the clones giving a positive signal was isolated, and the plasmid contained a 3.9-kb B. subtilis DNA fragment and was designated pBSG8-10 (Fig. 1).

The B. subtilis strain QB976 (sacPl, recE4) was transformed by pBSG8-10. All the chloramphenicol-resistant transformants tested grew when the sole carbon source was sucrose (0.1%). It has been shown that the 2.1-kb EcoRI fragment harbors the entire sacA gene and the 3′ end of the sacP gene (4). It can therefore be concluded that pBSG8-10 codes for a functional sacP gene product and for sucrase.

Characterization of the sacP Gene Product in E. coli. Genetic and biochemical studies have shown that sacP mutants do not take up sucrose (1), but no direct evidence has been reported that sacP codes for \textsuperscript{II}S\textsuperscript{Sac}. In order to test this hypothesis, the sacP gene was expressed in E. coli, which is devoid of sucrose transport and saccharolytic activities, but contains enzyme I and histidine-containing phosphocarrier protein. The sacP gene was isolated on recombinant plasmid pBSG8-12 by subcloning the 1.7-kb \textit{Sma} I–\textit{ Ava} I DNA fragment in pUC9 (Fig. 1). An \textit{E. coli} Rec\textsuperscript{−} strain (HB101) was transformed by pBSG8-10 and by pBSG8-12, and the capacity of the transformants to metabolize sucrose was tested. Cells transformed by the vector alone or by the recombinant plasmids containing either the sucrose gene or the sacP gene (pBSG8-12) remained white on MacConkey plates containing 2% sucrose, whereas those transformed by pBSG8-10, which carries both the sacA and sacP genes, grew as red colonies on MacConkey plates due to the hydrolysis of sucrose. These cells were also replica plated onto M9 agar plates containing 0.1% sucrose and ampicillin. The pBSG8-10 transformants grew well, thus demonstrating that they could use sucrose as the sole carbon source. Moreover, the growth curves obtained when strain TG1 (pBSG8-10) was grown on M63 liquid medium supplemented with ampicillin (25 \mu g/ml), chloramphenicol (1.5 \mu g/ml), and either 0.1% glucose or 0.2% sucrose were identical. The generation time was 90 min at 37°C in both cases. The sacP gene product is therefore sufficient to allow the transport of sucrose into the \textit{E. coli} cytoplasm.

The hypothesis that the sacP gene product is enzyme II was tested by assaying [\textsuperscript{U-\textsuperscript{14}C}]sucrose transport in strains of \textit{E. coli} containing sacP and either a thermosensitive enzyme...
I (Ts19) at 30°C and 42°C (Fig. 2) or a wild-type enzyme I. It can be seen that both the rate of sucrose uptake and the maximal intracellular level of sucrose were temperature dependent in strain Ts19, indicating that the transport of sucrose occurs by way of the temperature-sensitive enzyme I. The thermostability of the cloned sacP gene product was tested by assaying sucrose transport in strain TGI, which contains wild-type enzyme I. The initial rate of sucrose uptake by strain TGI (pBSG8-12) was 6 nmol/min per mg of protein at 30°C and 8 nmol/min per mg of protein at 42°C, values that represent 30% and 40% of the rate of wild-type B. subtilis. The initial rate in strain Ts19 at 30°C is only of 1 nmol/min per mg of protein, showing that, even at permissive temperature, the thermosensitive enzyme I is less active than the wild-type enzyme I. The initial rate values indicate that the complementation is very efficient, which contrasts with results reported earlier for in vitro complementation experiments by Simoni et al. (19). Two reasons may explain this difference. The first one comes from the in vivo situation where intact membrane may stabilize proteins or reactive intermediates. The second one may be that the stoichiometry could be shifted in favor of the number of 1^10^6 molecules since the gene is on a plasmidic plasmid. The rate of sucrose uptake by the control strain Ts19 (pUC9) corresponds to diffusion. If sucrose transport occurs by way of the PTS, the sugar would be expected to be phosphorylated.

**Fig. 3.** Analysis of plasmid-encoded polypeptides. Circular plasmid DNA (1 μg) was used in our assay of plasmid-encoded gene products. Tranlational products were separated by NaDodSO4/PAGE and were detected by autoradiography. (A) Lanes: 1, pUC9; 2, pBSG8-11. (B) Lanes: 1, pUC9; 2, pBSG8-12. The estimated molecular weights are indicated. Arrowheads indicate bands of approximate Mr values of 50,000 and 38,000–40,000.

**Fig. 4.** Nucleotide sequence of a portion of the 2.8-kb Smal I fragment spanning sacP and showing the deduced amino acid sequence of the gene product and the N-terminal part of succrase. The potential ribosome binding site of the sacP gene and the internal EcoRI site are underlined (thick line and thin line, respectively). The stop codon is symbolized by three asterisks. The sacA and the putative sacP initiation codons are indicated by arrows.
when transported into cells containing sacP but lacking sacA (Fig. 2B). Sucrose-6-phosphate was detected, which gives evidence that sucrose had been phosphorylated during transport. The presence of free sucrose was probably due to the presence of a nonspecific phosphatase as reported by Haguenauer and Képès (20).

The results presented above show that E. coli cells containing the sacP gene transport sucrose by way of a phosphotransferase system; the sacP gene product is therefore likely to be enzyme II of the sucrose PTS of B. subtilis. The II\textsuperscript{Sac} is most likely located in the inner membrane in E. coli, and sucrose presumably reaches the periplasm by way of diffusion through the outer membrane porins, as OmpF and OmpC (21).

In order to estimate the Mr, of II\textsuperscript{Sac}, the 2.8-kb Smal fragment encoding sacP and the 454 amino-terminal amino acids of sacA was cloned in pUC9 to give pBSG8-12 (Fig. 1). pBSG8-11 and pBSG8-12 were transcribed and translated in vitro (Fig. 3). Three major additional bands were observed with pBSG8-11 compared to the pattern obtained after transcription and translation of pUC9: a doublet with an approximate Mr of 38,000–40,000 and a weak band of Mr of ≈50,000. The pattern obtained with pBSG8-12 showed the presence of only the additional doublet (Mr of 38,000–40,000). This indicated that the doublet (Mr of 38,000–40,000) is due to the expression of the sacP gene and that the sacP gene product has an apparent Mr of 38,000–40,000. The presence of a doublet rather than a single band could be due to translation reinitiation at two ATG codons.

**Nucleotide Sequence of the sacP Gene.** The complete nucleotide sequence of the sacP gene is shown in Fig. 4. The 1.3-kb Smal–EcoRI fragment was subcloned in both orientations in M13mp19, and 95% of the fragment was sequenced on both strands. A 750-base-pair DNA fragment overlapping the internal EcoRI site was cloned, and deleted phosphates enabled sequencing through the EcoRI site. The sequence data obtained were compiled by computer using the DB system of Staden (22) to give a continuous overlapping sequence. The sequence shown in Fig. 4 starts about 50 nucleotides from the Smal 1 site and harbors the region containing the entire sacP gene and the first 131 nucleotides of the sacA gene. The sequence 3' to the EcoRI site (i.e., the last 270 nucleotides) has already been published (4).

The sequence was then examined for open reading frames in the six possible frames. On one strand, stop codons were found regularly in the three possible frames. In the other orientation, stop codons regularly interrupted two phases, and in the last frame, an open reading frame was found (starting at position 19 with an ATG start codon and ending at position 1398 at a TGA stop codon). In the same frame (at position 118) there is another ATG codon.

Seven nucleotides upstream of the first ATG codon (position 19) is a potential ribosome binding site (AAAGGG-GGA) (23). The calculated free energy of interaction of this sequence with the 3' end of B. subtilis 16S RNA is $\Delta G = -20.8$ kcal/mol ($\simeq 87$ kcal/mol) (24). The distance between this putative ribosome binding site and the ATG is 11 nucleotides, which is in good agreement with that usually found in B. subtilis (mean value of 10.6) (25). These results strongly suggest that this ATG is the in vivo initiation codon.

The nucleotide sequences encoding II\textsuperscript{Sac} and sucrase overlap (Fig. 4) at the sequence ATGA, where ATG is the initiation codon of the sucrose gene and TGA is the stop codon of the II\textsuperscript{Sac} open reading frame. This result supports the hypothesis that sacP and sacA are part of an operon in which sacA is distal to sacP (4). Postma and Lenayer (28) pointed out that the structural genes for the enzymes II are not part of a pts operon or regulon but form units with the corresponding metabolic enzymes. The existence of translational coupling in B. subtilis has been demonstrated, as has the optimization of such coupling by the sequence ATGA (26, 27).

**Comparison of II\textsuperscript{Sac} Sequence With That of Other Known Enzymes II.** The open reading frame defined by the nucleo-

---

**Fig. 5.** Comparison of B. subtilis II\textsuperscript{Sac} with E. coli II\textsuperscript{Sac}, S. typhimurium II\textsuperscript{Sac}, and E. coli II\textsuperscript{Sac}. The amino acid sequences of the four polypeptides are given in the one-letter code and have been aligned by introducing gaps (hyphens) to maximize identities. Identical and similar residues are boxed and stippled (accepted conservative replacements are I, L, V, and M; D and E; A and G; R and K; S and T; and F and Y). Numbers at the ends of the lines indicate the number of the nearest residue in the sequence. Numbers above the lines indicate the position number; 1 corresponds to the N-terminal methionine from II\textsuperscript{Sac}.
The amino acid sequence of the II \text{Suc} was compared to that of three other enzymes II—namely, mannitol-specific enzyme II (I\text{II}^{\text{MII}}) from Salmonella typhimurium (29), glucose-specific enzyme II (I\text{II}^{\text{Glc}}) from E. coli (30), and II\text{Bl} from E. coli (31). The alignments are shown in Fig. 5.

I\text{II}^{\text{Suc}} and I\text{II}^{\text{Glc}} are 460 and 477 residues long, respectively, whereas II\text{Bl} and II\text{MII} are 625 and 637 residues long, respectively. Gaps have been introduced to optimize alignment.

When I\text{II}^{\text{Suc}} and I\text{II}^{\text{Glc}} were compared, it appeared that the C-terminal part of I\text{II}^{\text{Suc}} shared similarity with the N-terminal part of I\text{II}^{\text{Glc}}, and a weaker similarity could be displayed encompassing the remaining parts of the sequences.

The greatest similarity is between I\text{II}^{\text{Suc}} and II\text{Bl} (most of the gaps introduced in these two proteins were at the same positions, in order to allow comparison with I\text{II}^{\text{MII}}). These two enzymes II, therefore, share homology, and their genes most probably derive from a common ancestor. The same conclusion may probably be drawn for the genes for the other enzymes II because of the blocks of homology found, although the similarity is not as great.

Indeed, the comparison between the four proteins shows four major blocks of similarity (Fig. 6). One is located in the N-terminal region of two proteins (I\text{II}^{\text{Suc}} and II\text{Bl}) and in the C-terminal part of I\text{II}^{\text{Glc}} at a second is near position 160, a third is near position 270, and the last one is near position 370. The last three regions have been described by Bramley and Kornberg (31). The four proteins share common properties, such as interactions with the cytoplasmic membrane and the phosphorylated histidine-containing phosphocarrier protein as well as transport and phosphorylation of a sugar. This may explain the regions of sequence conservation.

In conclusion, the results presented here demonstrate that sacP codes for B. subtilis I\text{II}^{\text{Suc}}, which shares homology with enzymes II from enteric bacteria. The sacP and sacA genes are probably organized in an operon structure in which the promoter is expected to be located 5' to sacP. The regulatory locus sacR, which is contiguous to sacP, is probably located in the vicinity of the promoter. Cloning the entire operon would provide useful tools to study the regulation of the gene expression in the B. subtilis sucrose regulon.

![Diagram](image-url)