Volume 288, number 1,2, 114–118 FEBS 10059 August 1991

ADONIS 001457939100764F

Nucleotide sequence of the *secY* gene from *Lactococcus lactis* and identification of conserved regions by comparison of four SecY proteins

Teija Koivula, Ilkka Palva and Harri Hemilä

Institute of Biotechnology, University of Helsinki, Finland

Received 24 May 1991

SecY is an integral membrane protein which participates in the translocation of proteins through the bacterial cell membrane. We have cloned the *secY* gene of *Lactococcus lactis*, and found its deduced protein sequence, 439 amino acids long, to be similar in length to the previously determined SecY proteins of *Escherichia coli*, *Bacillus subtilis* and *Mycoplasma capricolum*. Comparison of the *L. lactis* SecY to the 3 other SecY proteins revealed 90 conserved amino acid residues (21%). Nearly half of the conserved residues are clustered in 2 of the 10 transmembrane segments, and in 2 of the 6 cytoplasmic regions. Some of the conserved regions are apparently responsible for the interactions of SecY with signal sequences, and the proteins SecE and SecA.

SecY; PrlA; Membrane protein; Protein export; Protein secretion; (*Lactococcus lactis*)

1. INTRODUCTION

SecY protein is an essential component of the protein export machinery in *E. coli*. It was first identified as *prlA* mutations which suppressed defective signal sequences [1], and as temperature sensitive *secY* mutations, which showed a block in protein export at elevated temperature [2]. Accordingly, SecY appears to interact with the signal sequences of secretory proteins, but it has also been shown to interact with at least 2 other components of the protein translocation system: the peripheral membrane protein SecA [3–5], and the integral membrane protein SecE [4–6].

The protein transport system of Gram positive bacteria has been studied less intensively. In *B. subtilis*, a complex of 4 proteins was suggested to directly participate in protein secretion [7], but this complex was proved to be pyruvate dehydrogenase, which undermined its proposed role [8]. Recently, a gene with substantial sequence similarity to the *secY* gene of *E. coli* was cloned from *B. subtilis* [9, 10]. This suggested that other *B. subtilis* secretion components may also be similar to those of *E. coli*, since SecY interacts with SecA and SecE. Subsequently, the *secA* gene of *B. subtilis* has been identified [11], indicating close similarities between the translocation machineries of Gram negative and Gram positive bacteria.

Lactococci are an important group of Gram positive bacteria, which are widely used in the dairy industry, and thus their molecular biology is of importance. Here we report the sequence for the *L. lactis secY* gene and compare its deduced protein sequence to 3 other SecY protein sequences, in order to infer regions of functional significance.

Correspondence address: H. Hemilä, Institute of Biotechnology, Valimotie 7, 00380 Helsinki, Finland. Fax: (358) (0) 4346028.

2. MATERIALS AND METHODS

2.1. *Bacterial strains, plasmids and cultivation conditions*

L. lactis subsp. *lactis* MG1614 [12] was the source for chromosomal DNA. *L. lactis* cells were grown in M17-glucose medium [13]. *E. coli* DH5αF′ [14] was used as a cloning host for pUC19 plasmid constructions. *E. coli* cells were grown in L-broth [15], and transformed by the method of Hanahan [14].

2.2. *DNA methods and sequencing*

Plasmids were isolated from *E. coli* by the alkaline method [15]. The chromosomal DNA of *L. lactis* was isolated according to Marmur [16]. Restriction enzyme digestions, ligations, gel electrophoresis and Southern hybridization were performed by standard methods [15,17]. [α-^{32}P]dCTP radioisotope, for nick translation labeling, was obtained from Amersham International. Polymerase chain reactions (PCR) were done as suggested by the supplier of the AmpliTaq enzyme (Perkin Elmer Cetus). DNA was sequenced directly from alkaline denatured plasmids by the dideoxynucleotide chain termination method as previously described [17,18]. Protein sequences were aligned with the GAP program [19].

3. RESULTS AND DISCUSSION

3.1. *Cloning of the L. lactis secY*

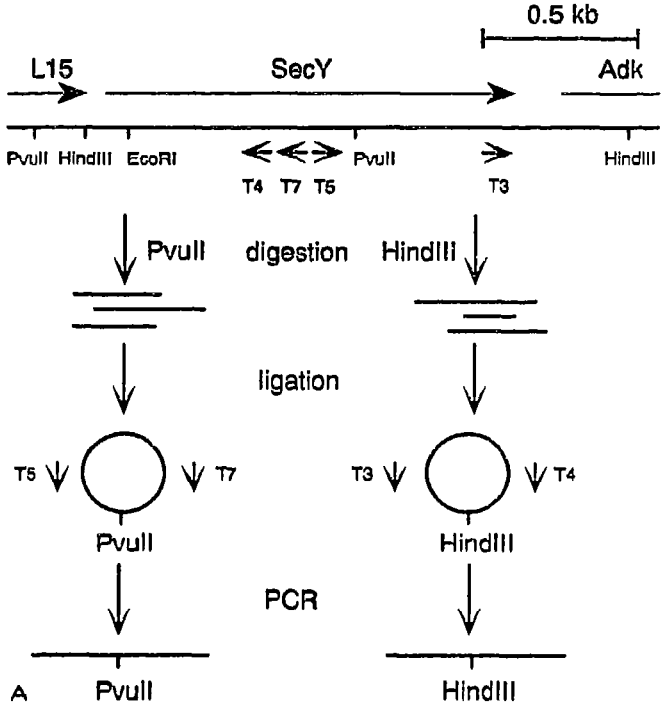
The 3′ end of the *L. lactis secY* gene was identified in a fragment of DNA containing the genes for adenylate kinase (*adk*) and initiation factor 1 of ribosomal protein operons *spc* and *α* (Koivula and Hemilä, submitted). The *secY* gene was located upstream of the *adk* gene, as it is also found in *B. subtilis* [9].

The 5′ end of the *L. lactis secY* gene was cloned by PCR. Southern analysis of *L. lactis* chromosomal

 Published by Elsevier Science Publishers B.V.

DNA, using as a probe a fragment containing the 3′ end of *secY* and the 5′ end of *adk*, revealed *Hin*dIII and *Pvu*II fragments of 1.7 and 1.1 kb, respectively (data not shown). These enzymes appeared to be suitable candidates for inverted PCR. The chromosomal DNA of *L. lactis* was digested with *Hin*dIII, and ligated at low concentration to yield circular molecules, which were then used as templates in PCR (Fig. 1). The PCR fragments were cloned into the *Sma*I site of the pUC19 vector and sequenced. To make the sequencing of the complementary strand easier, the PCR fragment was digested with *Hin*dIII, and the fragment containing the 3′ end of *secY* was isolated using agarose gel electrophoresis, and cloned into *Hin*dIII-*Sma*I digested pUC19. Three independent *Hin*dIII clones were sequenced to reveal possible errors produced by the AmpliTaq enzyme during the PCR reaction. A similar cloning strategy was used for *Pvu*II digested DNA (Fig. 1).

The sequenced region contained an open reading frame of 1317 bp starting at the nucleotide 138 (Fig. 2). A ribosomal binding site is located upstream of the putative initiation codon of *secY*. In *B. subtilis* a gene



T3: 5′-GTAAATATGCAGGATTTA-3′
T4: 5′-AGGAATTTCAGATGGTCTT-3′
T5: 5′-CCAACAAGTTCGTACTTACC-3′
T7: 5′-AATAATGACAATTGCTGAC-3′
B

Fig. 1. (A) Cloning of *L. lactis secY* by inverted PCR. ADK indicates the gene for adenylate kinase and L15 the gene for ribosomal protein L15. (B) Oligonucleotides used for the inverted PCR.

for ribosomal protein L15 is located immediately upstream of the *secY* gene [9,10], and upstream of the *L. lactis secY* gene we found an open reading frame coding for a L15 homologue (Fig. 2). No obvious promoter was found immediately upstream of the *secY* gene, which apparently is transcribed from the promoter of the *spc* operon.

3.2. *Amino acid sequence of the* L. lactis *SecY and comparison to other SecY proteins*

The *secY* gene of *L. lactis* encodes a protein of 439 amino acids, which shows a close similarity to the SecY sequences from *E. coli* [20], *B. subtilis* [9,10] and *M. capricolum* [21] (Table I). We have compared the 4 SecY sequences in order to identify the conserved residues (Fig. 3). Although several mutants of *E. coli secY* are available [22–26], the gene is essential for growth and therefore identification of residues essential for function may be an unattainable goal using a traditional mutation analysis. Accordingly, the identification of conserved residues may yield better indications of the essential regions and residues than random mutagenesis of the *secY* gene in vivo. In the 4 SecY proteins 90 residues (21%) are conserved.

The mutants that have been obtained of the *E. coli secY* gene are either temperature sensitive (Ts) [22,24], cold sensitive (Cs) [25], or they have altered interactions with the signal sequences (SS) [23,26]. Only some of the *E. coli* mutations correspond to conserved residues: one Cs-mutation (*secY40*), one Ts-mutation (*secA100*) and three SS-mutations (*prlA4-2*, *secY121*, *secY161*) have affected residues conserved in the 4 sequences (Fig. 3). This is consistent with the notion that changes in totally conserved residues have a high probability of making the protein nonfunctional. Some of the mutations have affected a residue that is neutral in all 4 proteins and changed it into a charged one (*secY24*, *secY39*, *prlA401*).

Topology analysis of the *E. coli* SecY has indicated 10 transmembrane segments [27]. The alignment of the four SecY sequences shows that the same transmembrane regions can be identified in all of the 4 proteins (Fig. 3). Two of the transmembrane segments (TM2 and TM7) contain a high number of conserved residues, while 3 (TM1, TM5, and TM10) contain a medium number of conserved residues. Conservation of a residue in transmembrane regions may indicate participation in protein-protein interactions [28]. Accordingly, some of the 5 regions may interact with the signal sequence and/or SecE, but apparently some of them are essential for the structure of SecY. Four of the SS-mutations are located in TM7, and this segment has been proposed to play a crucial role in signal sequence recognition [26]. However, the signal sequence appears to first interact with SecE and only thereafter with SecY [6]. Thus, mutations affecting the SecY-SecE interaction may indirectly affect the signal sequence recogni-

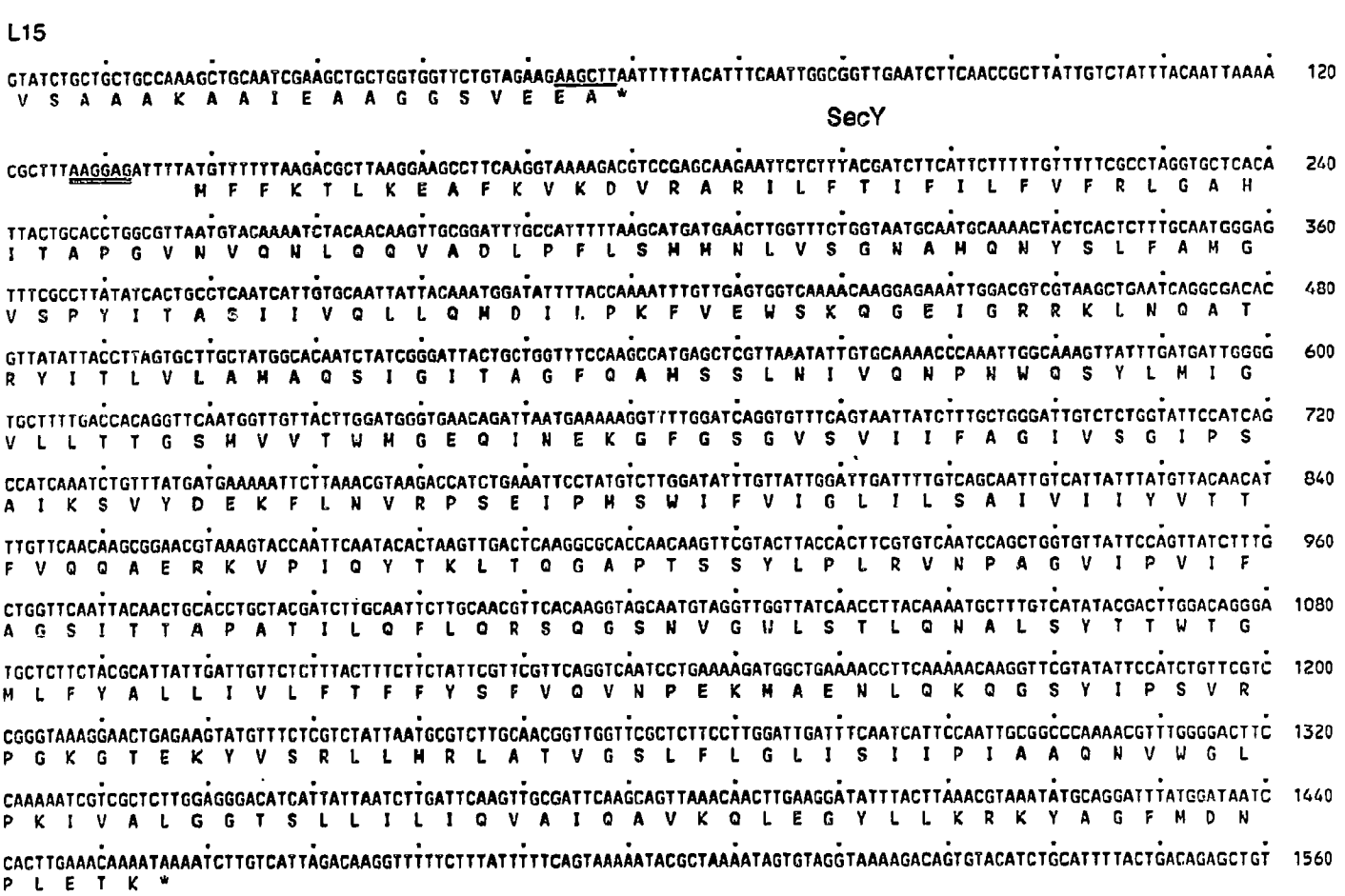


Fig. 2. Nucleotide sequence and the deduced amino acid sequence of the *L. lactis secY* gene. The ribosomal binding site is double underlined. The upstream *Hin*dIII site is underlined. The putative fragment of L15 (19 aa) has 11 identical amino acids with the C-terminus of *B. subtilis* L15 [9,10]. The DNA sequence has been submitted to the EMBL nucleic acid database (X59250). Both strands of the DNA were sequenced, except for the region of 1–53 bp.

tion. In fact, it has been found that the *prlA4* mutation affects the interaction between SecY and SecE [6]. Unfortunately, *prlA4* has 2 mutations, one in TM7 (*prlA4-1*) and another in TM10 (*prlA4-2*) and thus either of these regions, or both, may interact with SecE. It is notable that none of the mutations have affected TM2, the most conserved of the ten transmembrane regions (Fig. 3). Whether TM2 plays a structural role in SecY or has catalytic activities in translocation cannot be concluded.

Table 1

Percentage of identical amino acids upon optimal alignment of SecY proteins

	B. subtilis	*M. capricolum*	*E. coli*
L. lactis	48	39	39
B. subtilis	-	39	41
M. capricolum		-	33

Two regions of SecY which are directed to the cytoplasm, IN2 and IN5, contain a large number of conserved residues, and they may correspond to the SecA binding site of SecY. Antiserum against the N-terminus of *E. coli* SecY has been shown to cause a partial inhibition of the SecA binding, suggesting that the SecA binding site is located close to, but not exactly at, the N-terminus [5]. Region IN2 is most probably located close to the N-terminus, but there are no data to suggest the relative location of IN5 with respect to the N-terminus. Interestingly, two Cs-mutations (*secY39* and *secY40*), which cause elevated *secA* expression, are located in the IN5 region. No specific interactions (i.e. evolutionary constraints) are known for the regions of the protein that point to the outside of the membrane (OUT-regions). For the most part the known insertions/deletions of SecY are located in the regions facing the outside, and few of the conserved residues are located in the OUT regions. Although the conserved residues give indications for the essential regions of the SecY protein, there are not yet sufficient data to con-

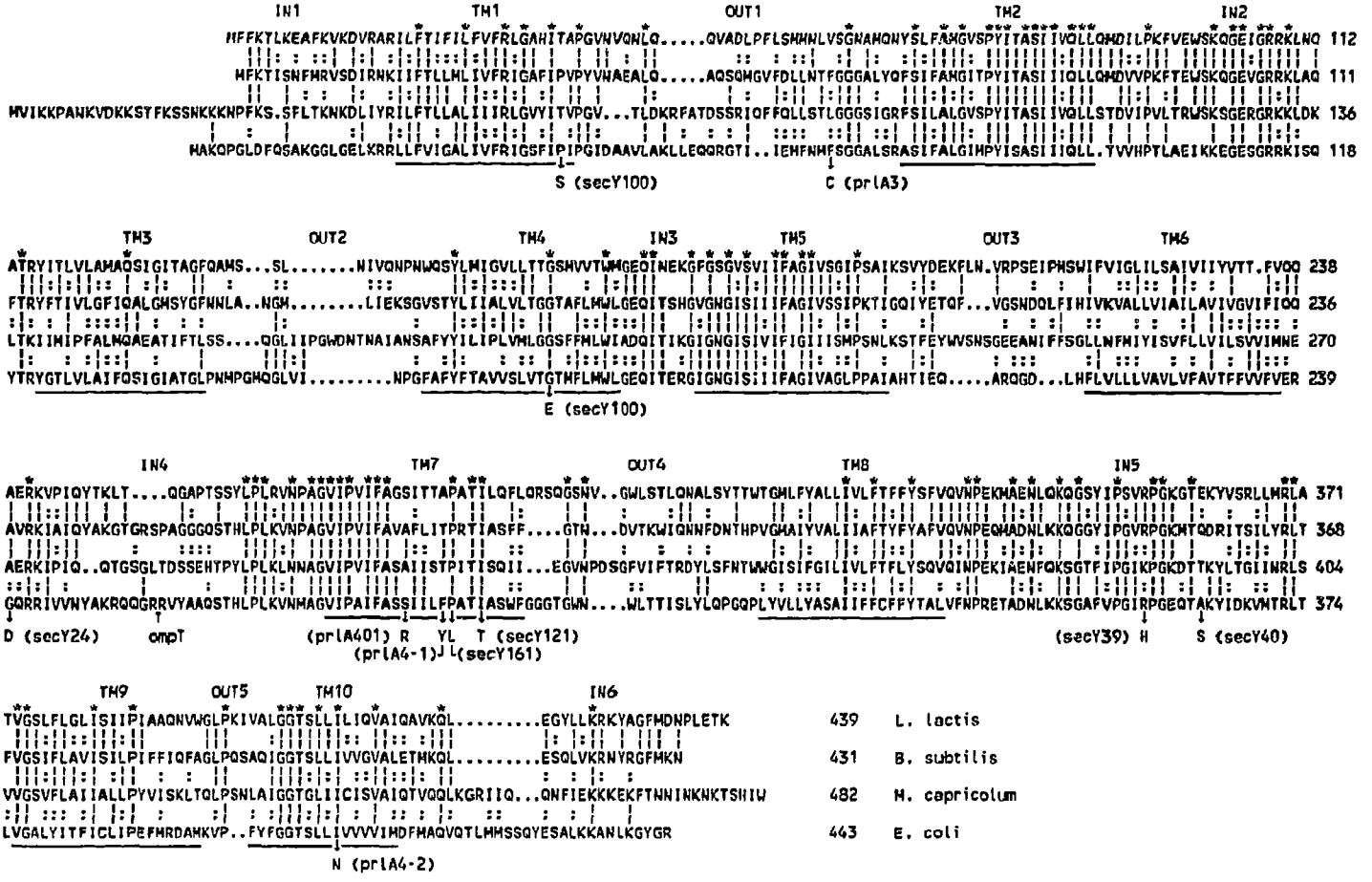


Fig. 3. Comparison of the SecY proteins of *B. subtilis* [9,10], *M. capricolum* [20] and *E. coli* [19]. Identical (!) and closely related amino acids (:), with a comparison value ≥ 0.5 in the GAP mutation data matrix [19], are indicated. The conserved residues are indicated by an asterisk. Underlinings of the *E. coli* SecY indicate the putative α-helical transmembrane regions (TM) [29] for the 4 proteins. The regions between the transmembrane segments pointing to the cytoplasm (IN) or to the outside (OUT) have been derived by topology analysis of the *E. coli* SecY [27]. The mutations in the *E. coli* SecY protein are marked by arrows and the names of the mutations are indicated in parentheses. References for the mutations: Ts, *secY24* [22], *secY100* [24]; Cs, *sec39*, *secY40* [25]; SS, *prlA3*, *prlA401*, *prlA4-1*, *prlA4-2* [23], *secY121* and *secY161* [26]. The ompT cleavage site [4,30] is also indicated.

clude the relative orientations of the ten transmembrane regions, except that the transmembrane regions that have few conserved residues probably face towards the lipids [28].

Acknowledgements: This work was supported by a grant from the Foundation for Biotechnical and Industrial Fermentation Research. The authors are grateful to Dr Lars Paulin for the synthesis of oligonucleotides.

REFERENCES

[1] Emr, S.D., Hanley-Way, S. and Silhavy, T.J. (1981) Cell 23, 79-88.
[2] Ito, K., Wittekind, M., Nomura, M., Shiba, K., Yura, T., Miura, A. and Nashimoto, H. (1983) Cell 32, 789-797.
[3] Fandl, J.P., Cabelli, R., Oliver, D. and Tai, P.C. (1988) Proc. Natl. Acad. Sci. USA 85, 8953-8957.
[4] Brundage, L., Hendrick, J.P., Schiebel, E., Driessen, A.J.M. and Wickner, W. (1990) Cell 62, 649-657.
[5] Hartl, F.-U., Lecker, S., Schiebel, E., Hendrick, J.P. and Wickner, W. (1990) Cell 63, 269-279.
[6] Bieker, K.L. and Silhavy, T.J. (1990) Cell 61, 833-842.
[7] Caulfield, M.P., Horiuchi, S., Tai, P.C. and Davis, B.D. (1984) Proc. Natl. Acad. Sci. USA 81, 7772-7776.
[8] Hemilä, H., Palva, A., Paulin, L., Arvidson, S. and Palva, I. (1990) J. Bacteriol. 172, 5052-5063.
[9] Nakamura, K., Nakamura, A., Takamatsu, H., Yoshikawa, H. and Yamane, K. (1990) J. Biochem. 107, 603-607.
[10] Suh, J.-W., Boylan, S.A., Thomas, S.M., Dolan, K.M., Oliver, D.B. and Price, C.W. (1990) Mol. Microbiol. 4, 305-314.
[11] Sadaie, Y., Takamatsu, H., Nakamura, K. and Yamane, K. (1991) Gene 98, 101-105.
[12] Gasson, M.J. (1983) J. Bacteriol. 154, 1-9.
[13] Terzaghi, P.E. and Sandine, W.E. (1975) Appl. Environ. Microbiol. 29, 807-813.
[14] Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual, Cold Spring Harbor, New York.
[16] Marmur, J. (1961) J. Mol. Biol. 3, 208-218.
[17] Koivula, T., Sibakov, M. and Palva, I. (1991) Appl. Environ. Microbiol. 57, 333-340.
[18] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
[19] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.

[20] Cerretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M. and Nomura, M. (1983) Nucleic Acids Res. 11, 2599-2616.
[21] Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F. and Osawa, S. (1987) Mol. Gen. Genet. 210, 314-322.
[22] Shiba, K., Ito, K., Yura, T. and Cerretti, D.P. (1984) EMBO J. 3, 631-635.
[23] Sako, T. and Iino, T. (1988) J. Bacteriol. 170, 5389-5391.
[24] Ito, K., Hirota, Y. and Akiyama, Y. (1989) J. Bacteriol. 171, 1742-1743
[25] Baba, T., Jacq, A., Brickman, E., Beckwith, J., Taura, T., Ueguchi, C., Akiyama, Y. and Ito, K. (1990) J. Bacteriol. 172, 7005-7010.
[26] Sako, T. (1991) J. Bacteriol. 173, 2289-2296.
[27] Akiyama, Y. and Ito, K. (1987) EMBO J. 6, 3465-3470.
[28] Rees, D.C., Komiya, H., Yeates, T.O., Allen, J.P. and Feher, G. (1989) Annu. Rev. Biochem. 58, 607-633.
[29] Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984) J. Mol. Biol. 179, 125-142.
[30] Akiyama, Y. and Ito, K. (1990) Biochem. Biophys. Res. Commun. 167, 711-715.