Eur. J. Biochem. *193*, 325–330 (1990)

Export and purification of a cytoplasmic dimeric protein by fusion to the maltose-binding protein of *Escherichia coli*

Arnaud BLONDEL and Hugues BEDOUELLE

Unité de Biochimie Cellulaire (Unité de Recherche associée D1129 du Centre National de la Recherche Scientifique), Institut Pasteur, Paris, France

(Received April 6/June 28, 1990) – EJB 90 0384

A hybrid between the maltose-binding protein (MalE) of *Escherichia coli* and the gene 5 protein (G5P) of phage M13 was constructed at the genetic level. MalE is a monomeric and periplasmic protein while G5P is dimeric and cytoplasmic. The hybrid (MalE-G5P) was synthesized in large amounts from a multicopy plasmid and efficiently exported into the periplasmic space of *E. coli*. The export was dependent on the integrity of the signal peptide. MalE-G5P was purified from a periplasmic extract by affinity chromatography on cross-linked amylose, with a yield larger than 50000 molecules/*E. coli* cell. The hybrid specifically bound denatured but not double-stranded DNA cellulose, as native G5P. Sedimentation velocity and gel-filtration experiments showed that MalE-G5P exists as a dimer. Thus, it was possible to efficiently translocate through the membrane a normally cytoplasmic and dimeric protein, by fusion to MalE. Moreover, the passenger protein kept its activity, specificity and quaternary structure in the purified hybrid. MalE-G5P will enable the study of mutant G5P that no longer binds single-stranded DNA and therefore cannot be purified by DNA-cellulose chromatography.

Site-directed mutagenesis is a powerful approach to the study of the relationship between structure and function in proteins. However, the range of mutations that can be made in a protein is praetically limited to those that still allow its purification and study. This constraint strongly diminishes the power of the method when the purification of the protein rests on its activity. Such is the case for the gene 5 protein (G5P) of phages fd and M13, which specifically binds single-stranded DNA [1, 2] and is usually purified by affinity chromatography on denatured DNA cellulose [2]. G5P is a cytoplasmic and dimeric protein [3, 4]. Its two monomers are identical and each one includes 87 amino acid residues [5]; they dissociate with a K_d of 1.25 μM [6]. Its crystallographic structure has been determined at high resolution [7]. Its role consists of switching the replication cycle of the phage from the production of double-stranded DNA to that of single-stranded DNA, which is subsequently packaged into virion particles.

A remedy to the above problem lies in using one of the protein vectors that have recently been developed (for a review, see [8]). These vectors are based on the construction of hybrids, at the genetic level, between the target protein and a vector protein which can be easily purified, generally by affinity chromatography. In this way, the desired protein can be produced and purified independently of its properties. Export of the hybrid protein outside of the cytoplasm is generally advantageous because it simplifies the purification.

We have developed a protein vector which is based on protein MalE of *Escherichia coli* [8]. MalE is an abundant, medium-sized (370 residues) protein which is encoded by the *malE* gene [9]. Its expression is controlled by *malEp*, a strong promoter which is activated by protein MalT in the presence of maltotriose and repressed by glucose [10–12]. It is exported into the periplasmic space by means of an N-terminal signal peptide which is cleaved during the export process [13]. MalE binds maltose and maltodextrins and can be purified by affinity chromatography on cross-linked amylose. The elution is performed by competition with free maltose [14]. MalE is normally a monomeric protein. However, when produced in a mutant strain which constitutively expresses the maltose regulon in the absence of inducer, it forms dimers. This dimeric form of MalE dissociates into monomers upon addition of maltose [15].

In this paper, we report the construction and characterization of a hybrid between MalE and G5P. We show that this hybrid protein, MalE-G5P, is efficiently exported into the periplasmic space of *E. coli* and that its G5P portion retains its activity, specificity and quaternary structure. We indicate potential uses of the hybrid in the study of G5P.

MATERIALS AND METHODS

Media, vectors and strains

Plasmid pPD1 carries the wild-type *malE* gene under the control of its own promoter, *malEp* [9]. pPD127 and pPD140 carry mutant alleles of *malE* in which a *Bam*HI restriction site has been inserted into the signal sequence (*malE127*) and the 3′-terminal codon (*malE140*) [16]; phages M13mp19am4 [17],

Correspondence to H. Bedouelle, Unité de Biochimie Cellulaire, Institut Pasteur, 28, rue du Docteur Roux, F-75724 Paris cedex 15, France.

Abbreviations. MalE, maltose-binding protein, product of the *malE* gene of *Escherichia coli*; G5P, gene 5 protein of bacteriophage M13; $s_{20,w}$, sedimentation coefficient at 20°C in water; R_S, Stokes radius.

Enzymes. Bovine erythrocyte carbonic anhydrase (EC 4.2.1.1); bovine liver catalase (EC 1.11.1.6); horseradish peroxidase (EC 1.11.1.7); rabbit muscle phosphorylase *b* (EC 2.7.1.37); restriction endonucleases (EC 3.1.21.4); β_2 subunit of *E. coli* tryptophan synthase (EC 4.2.1.20).

M13K11RX [18] and the media [19] have been described. The complete genotypes and origins of strains PD28 (*malT^{c}1*, *ΔmalE444*, *recA*) which constitutively expresses the maltose regulon [16], HB2200 (*malT*, *hsdR*, *hsdM$^+$*) [8], BMH71-18 (*hsd$^+$*) and TG1 (*hsdR*, *hsdM*) [17] can be found in the references. TG2 is a *recA* derivative of TG1 and a gift of T. Gibson.

Mutagenesis and recombinant DNA

We introduced a *Bam*HI site upstream of the M13 gene 5 by oligonucleotide site-directed mutagenesis, using the technique of double priming with amber selection [17] and modifications already described [20]. The sequence of the mutagenic oligonucleotide was 5′-CTTTAATCATTGGGATCCAC-CTTATGCGA-3′.

Preparation of cellular extracts and purification on cross-linked amylose

The PD28 derivatives, harboring the recombinant plasmids, were grown in the presence of ampicillin, at 100 μg/ml. They were kept in conditions where promoter *malEp* is repressed, i.e. on plates of ML agar supplemented with 2% glucose [8]. For the expression of protein MalE and its derivatives, the strains were grown at 30 °C in ML broth without added sugar, and harvested at $A_{600} = 1.0\,cm^{-1}$. The preparation of periplasmic and shocked-cell extracts and the purification of proteins on a column of cross-linked amylose were performed as described in [8] except that we included 1 mM phenylmethylsulfonyl fluoride in the buffers, we adjusted the periplasmic extract to 50 mM Tris/HCl, pH 7.5, 2.5 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (buffer A) immediately after the osmotic shock and we ran the column with this buffer. In addition, the column was washed with several volumes of 0.5 M NaCl in buffer A to elute any bound DNA.

DNA-cellulose chromatography

Chromatography was carried out with buffer B (50 mM Tris/HCl, pH 7.5, 2.5 mM 2-mercaptoethanol, 1 mM EDTA). The columns were loaded with protein preparations that either came directly from the cross-linked amylose column or had been dialysed against buffer B. The bound proteins were eluted with a gradient of 0 – 2 M NaCl. The native-DNA – cellulose and denatured-DNA – cellulose (Pharmacia) had similar contents of DNA (1.9 mg/ml and 1.2 mg/ml cellulose). For analytical chromatography, we used 1-ml columns, loaded at a flow rate of 2 ml/h and rinsed with 15 – 20 vol. buffer B. Quantities of pure MalE-G5P were used that did not saturate the denatured-DNA column.

From the extinction coefficients of MalE-G5P (see below) and DNA, and from the absorbances of the pure MalE-G5P fractions, at 260 nm and 280 nm, we evaluated the content of these fractions in DNA [21]. We found that this content was within the standard error of the method and equivalent, at most, to a few percent of the MalE – G5P-binding sites for DNA, assuming 4 consecutive nucleotides/site on each monomer [22].

Protein concentrations

The MalE-G5P fractions were concentrated by dialysis against a saturated solution of ammonium sulfate in 50 mM Tris/HCl, pH 7.5, collection of the precipitate by centrifugation and resuspension in buffer C (50 mM Tris/HCl, pH 7.5, 50 mM NaCl, 1 mM phenylmethylsulfonyl fluoride, 2.5 mM 2-mercaptoethanol, 1 mM maltose). We checked that MalE-G5P retained the capacity to bind denatured-DNA – cellulose after ammonium sulfate precipitation.

We used the following absorption coefficients at 260 nm and 280 nm, respectively ($cm^{-1} \cdot g^{-1} \cdot l^{-1}$): for MalE, 0.71 and 1.47 [14, 23]; for G5P, 0.41 and 0.73 [1]; for MalE-G5P, 0.65 and 1.32. The absorption coefficients for MalE-G5P were calculated as the weighted averages of those for MalE (40.7 kDa) [9] and G5P (9.7 kDa) [5]. The protein concentrations were also determined by the Bio-Rad protein-assay kit, using bovine serum albumin as standard. The values obtained by the spectroscopic and colorimetric methods were very close.

Centrifugations

For zonal centrifugations, we used preformed linear gradients (11.5 ml) of 5 – 20% sucrose in buffer C, run in a Beckman SW41 rotor at 4 °C. Each of the protein samples (0.2 ml) contained catalase as a marker; they were dialysed against buffer C, layered on top of the gradients and sedimented at 40000 rpm for 18 h. After centrifugation, the gradients were pumped from the bottom of the tubes, using a capillary, through a ultraviolet detector. The positions of the protein peaks were determined by their absorbance at 280 nm. The peaks were collected manually and their content identified by polyacrylamide gel electrophoresis and reaction with hydrogen peroxide. The sedimentation coefficient ($s_{20,w}$) of MalE-G5P was deduced from the values for catalase (11.10 S), bovine serum albumin (4.30 S), hen egg white ovalbumin (3.37 S) and MalE (3.30 S), using a linear relationship between $s_{20,w}$ and the distance from the protein peaks to the meniscus of the gradients on the absorbance profiles.

The sedimentation-equilibrium experiments were performed in an MSE 215 rotor, at 12000 rpm, 12 °C in buffer C. The volume of the protein samples, in the cells, was 400 μl. After equilibrium had been reached, the absorbance profiles in the cells were recorded. The samples were subsequently centrifuged at 48000 rpm to obtain the baseline.

The boundary-sedimentation velocity experiment was carried out in similar conditions, except that the speed was 55000 rpm and the absorbance in the cell was scanned every 10 min.

Gel-filtration chromatography

The Stokes radius (R_S) of MalE-G5P was determined by gel filtration. A Superose 12 HR 10/30 column was connected to a Pharmacia FPLC system and run at 1 ml/min at room temperature in buffer C. Sample (0.1 ml) was loaded via a loop and valve assembly and the effluent continuously monitored at 280 nm. Fractions were collected manually. We used a linear relationship between the R_S and the elution volumes of the proteins. This relationship was calibrated with horseradish peroxidase (3.02 nm), hen egg white ovalbumin (3.05 nm), horse haemoglobin (3.11 nm), tryptophan synthase β_2 subunit (4.0 nm), phosphorylase *b* (4.8 nm) and catalase (5.22 nm).

Molecular mass calculations

The molecular masses (m) were deduced from the equation $m = 6\pi\eta R_S N_A s_{20,w}/1\text{-}\varrho\bar{v}$, derived from Svedberg's and Stokes' laws, where $\bar{v}$ is the specific volume of the molecule, N_A,

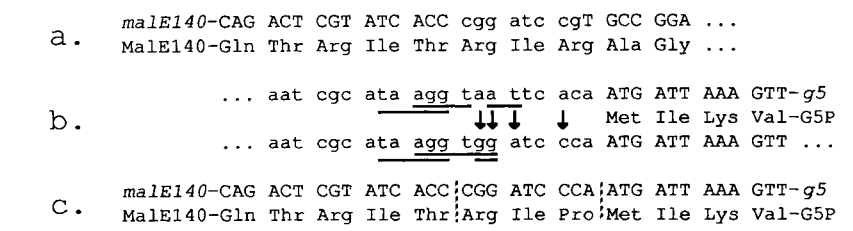

Fig. 1. *Design of the fusion between* malE140 *and gene 5.* (a) Upper line: sequence of the 3′-terminal end of the mutant gene, *malE140*. The *Bam*HI linker is represented in lower case letters. The insertion occurs in the last codon of the wild-type gene, *malE*, and introduces a frameshift [16]. Lower line: sequence of the encoded polypeptide. (b) Upper line: sequence of the 5′-terminus of gene 5 [5]. The coding sequence is represented in upper case letters. Sequences complementary to the 3′-terminus of the 16S ribosomal RNA are underlined. Middle line: sequence of the N-terminus of G5P. Lower line: sequence of the mutant gene 5. The mutations, indicated by arrows, introduce a *Bam*HI site while keeping a sequence highly complementary to the 16S ribosomal RNA. (c) Upper line: sequence of the join between *malE140* and gene 5 after recombination at the *Bam*HI sites. Lower line: deduced amino acid sequence. The linking sequences are bracketed

Avogadro's number, $\varrho = 1$ g/ml and $\eta = 0.001$ Pa · s, the density and viscosity of water at 20 °C [24]. From the amino acid sequence of MalE [9] and sedimentation equilibrium data [23], we calculated $\bar{v} = 0.764$ for MalE and took the same value for MalE-G5P.

Polyacrylamide gel electrophoresis

The gels were run as described [8]. When necessary, the protein samples were concentrated by precipitation with 5% trichloroacetic acid before electrophoresis.

RESULTS

Construction of a hybrid between malE *and gene 5*

To keep the activities of both MalE and G5P in a hybrid protein, we fused complete versions of genes *malE* and 5. For this, we used a mutant gene, *malE140*, which carries a *Bam*HI site inserted into the last codon of *malE* [16], and introduced a *Bam*HI site immediately upstream of gene 5 (Fig. 1). This new *Bam*HI site is located within the translation-initiation site of gene 5 and improves the complementary nature of the sequences at this site and at the 3′ OH-terminus of the 16S ribosomal RNA. It was introduced by oligonucleotide site-directed mutagenesis into an M13 phage derivative which carries only one copy of gene 5, as the wild-type M13. The mutant phage, M13-AB1, was viable but gave clearer plaques than its parent. Thus, the introduction of the *Bam*HI site did not prevent a functional expression of gene 5.

We constructed the hybrid gene, *malE140-g5*, by recombination between the *Bam*HI sites of plasmid pAB1, which carries *malE140* under control of promoter *malEp*, and M13AB1 (Fig. 2). We named pAB2, the recombinant plasmid that codes for the desired hybrid protein, MalE-G5P. We also constructed *malE127/140-g5*, a derivative of *malE140-g5* which carries mutation *malE127* in its signal sequence. *malE127* totally blocks export of the wild-type protein MalE [8]. We named pAB3, the recombinant plasmid that codes for the mutant hybrid protein, MalE127-G5P.

Expression and cellular location of MalE-G5P

To avoid toxicity problems during the constructions, all the recombinant plasmids were propagated in HB2200, a strain in which promoter *malEp* is inactive [8]. For expression studies, the plasmids were introduced into PD28, a strain which constitutively expresses the maltose regulon but cannot ferment maltose because it harbours deletion *ΔmalE444* [9]. Plasmids pAB2 and pAB3 were not toxic to PD28; neither at 30 °C nor 42 °C, contrary to other constructs with *malE* [8]. Strain PD28(pAB2) could ferment maltose on McConkey indicator plates, whereas PD28(pAB3) could not. These genetic complementation experiments, performed with appropriate controls [8], were a first indication that MalE-G5P was exported into the periplasm and that MalE127-G5P remained inside the cytoplasm.

To determine the cellular locations of the hybrid proteins, we submitted cultures of the PD28 derivatives to the cold osmotic shock procedure [25] and analysed the cell extracts by polyacrylamide gel electrophoresis (Fig. 3). The periplasmic extract of strain PD28(pAB2) contained a protein with an apparent molecular mass corresponding to the theoretical value for MalE-G5P, 50.6 kDa, calculated from its amino acid sequence. This protein was the major species in the extract. It was absent from the periplasms of strains overproducing the wild-type MalE or MalE127-G5P (Fig. 3, lanes 4, 5 and 15). These results strongly suggested that the 50.6-kDa protein was MalE-G5P.

Purification of the hybrids

Proteins MalE, MalE127 and MalE140 can be purified by affinity chromatography on cross-linked amylose [8, 16]. We applied this technique to the purification of MalE-G5P and MalE127-G5P.

Regarding MalE-G5P, we purified 3.5 mg protein from a periplasmic extract of PD28(pAB2)/1 l culture at $A_{600} = 1.0\ cm^{-1}$. Less than 10% of the MalE-G5P chains flowed through the amylose column during its loading and, therefore, most (> 90%) had a MalE portion which bound amylose (Fig. 3, lanes 5–7). We did not obtain additional amounts of MalE-G5P by a high-salt wash of the shocked cells, but could purify 3.2 mg protein from a soluble extract of the shocked cells (Fig. 3, lanes 10–12). Thus, MalE-G5P was only partially exported into the periplasm, although in large quantities. The purification yields showed that the periplasmic space of each cell contained more than 50000 molecules of MalE-G5P (once the contribution of the degradation products had been removed; see below).

The periplasmic extract of PD28(pAB3) did not contain a protein with an apparent molecular mass close to that of MalE127-G5P (Fig. 3, lane 15). Accordingly, we purified only trace amounts of protein from this extract, which could be due to cell lysis. In contrast, we purified 0.8 mg protein/l culture, from a shocked-cell extract (Fig. 3, lane 18). The

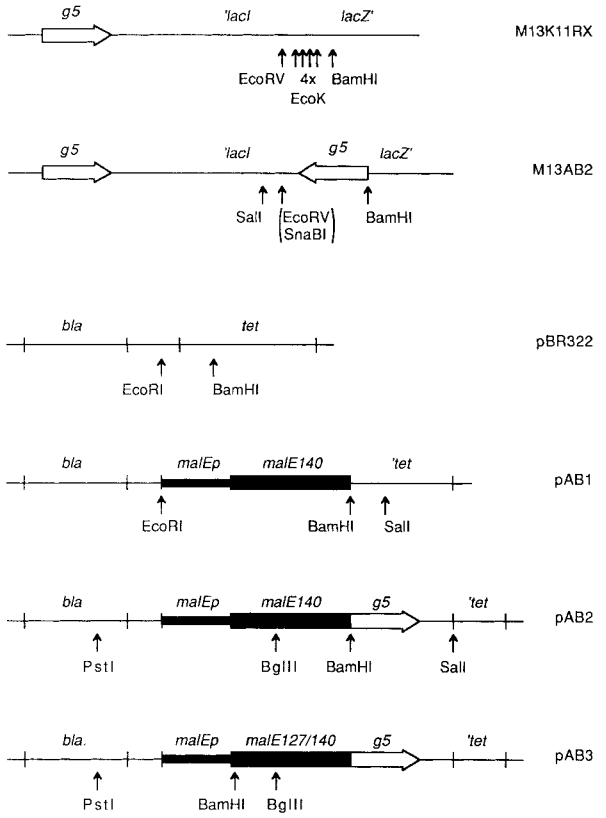

Fig. 2. *Construction of hybrid gene* malE140-g5. The four *Eco*K restriction sites which are present between the cloning sites of phage M13K11RX, provide a positive selection for the insertion of foreign DNA fragments [18]. The *Bam*HI – *Sna*BI restriction fragment of phage M13AB1 that contains gene 5 (*g5*, see Fig. 1), was inserted between the *Bam*HI and *Eco*RV sites of M13K11RX to give the recombinant phage, M13AB2. The *Eco*RI – *Bam*HI fragment of plasmid pPD140 that contains the mutant gene, *malE140*, under control of promoter *malEp* [16], was inserted between the *Eco*RI and *Bam*HI sites of plasmid pBR322 to give the recombinant plasmid, pAB1. The *Bam*HI – *Sal*I fragment of M13AB2 that contains *g5*, was inserted between the *Bam*HI and *Sal*I sites of pAB1 to give pAB2, carrying the hybrid gene, *malE140-g5*, under control of *malEp*. Finally, mutation *malE127* was introduced into *malE140-g5* by replacing the *Pst*I – *Bgl*II fragment of pAB2 that contains the 5′ end of *malE*, by the corresponding fragment of plasmid pPD127. *malE127* creates a *Bam*HI restriction site within the signal sequence of *malE* [16]

results showed that MalE127-G5P remained in the cytoplasm, as MalE127, and that the export of MalE-G5P was dependent on the integrity of its signal peptide.

The eluates from the amylose column contained two main protein species: the major species had an apparent molecular mass of around 50 kDa, corresponding to those of the hybrids; the minor species had an apparent molecular mass of 42 kDa, intermediate between those for the wild-type MalE and the hybrids, and was probably a degradation product (Fig. 3, lanes 7, 12 and 18). Since the 42-kDa species still bound amylose and because mutations in the N-terminal part of MalE result in the loss of its maltose-binding activity [16], the degradation must have occurred within the G5P portion of the hybrids. It was important to stop the cultures before saturation to minimize the degradation. The presence of a protease inhibitor (1 mM phenylmethylsulfonyl fluoride) during the cultures had no effect. The respective amounts of MalE-G5P and of its degradation product did not change in the periplasmic extract after 4 h of incubation at 37 °C. This indicated that MalE-G5P was not degraded during the following steps of purification.

The MalE-G5P chains had the same size when purified from the periplasmic or the shocked-cell extract of PD28(pAB2) (Fig. 3, lanes 7 and 12). Thus, the signal peptide seems to have been proteolysed in both cases. The hybrid protein that fractionated with the shocked cells might represent molecules that had been partially translocated across the bacterial inner membrane but not released into the periplasm. This type of behaviour has already been observed [26]. Both MalE127-G5P and its main degradation product were copurified with slightly heavier polypeptides (Fig. 3, lane 18), probably corresponding to intermediates in the degradation of the defective signal peptide. Such intermediates have already been observed for MalE127 [8].

Binding of hybrids to DNA

To test if the G5P portion of the hybrids could bind DNA, we loaded the material that had been partially purified with the amylose column, onto a column of denatured-DNA – cellulose. The contaminant proteins flowed through the column during its washing whereas most of MalE-G5P or MalE127-G5P (70 – 80%) remained bound to the column and could be eluted with a salt gradient, around 0.2 – 0.3 M NaCl, in pure form (Fig. 3, lanes 7 – 9, 12 – 14 and 18 – 20). This result demonstrated that MalE-G5P and MalE127-G5P could bind denatured DNA, whatever the extract from which they were purified, and that the binding was specific to the full-length hybrids. It provided a convenient means of purifying the hybrids to homogeneity.

In the remainder of the work, the experiments were carried out with hybrid protein purified from the periplasmic extract of PD28(pAB2). To test if the binding of MalE-G5P to DNA was specific to single-stranded DNA, we dialysed a pure fraction from the denatured-DNA column against a low-salt buffer and loaded it on to denatured-DNA and native-DNA columns, in parallel. The pure MalE-G5P flowed through the native-DNA column and was again retained by the denatured-DNA column. As a control, we showed that pure Klenow polymerase was retained by the native DNA-column. These experiments demonstrated that our pure MalE-G5P was functional for DNA binding and had the same specificity as native G5P for single-stranded DNA.

Oligomerization of periplasmic MalE-G5P

We performed sedimentation-equilibrium experiments with MalE and MalE-G5P. At equilibrium, the plot of ln[MalE] as a function of r^2, where r is the distance to the rotation axis of the centrifuge, gave a straight line. When equilibrium was reached from an initial protein concentration of 6.5 μM in the cell, the same plot for MalE-G5P had an upward curvature, indicating the presence of associating monomers (not shown).

The sedimentation coefficient of MalE was determined by a boundary-sedimentation-velocity experiment. We found $s_{20,w}$(MalE) = 3.30 ± 0.05 S. The sedimentation coefficient of MalE-G5P was determined by zonal centrifugation through a sucrose gradient. Pure MalE-G5P was layered on top of the gradients at a concentration of 19 μM. At the end of the run, the maximal local concentration of MalE-G5P, in the gradients, was 3 μM, i.e. a concentration at which native G5P would be 65% dimeric, assuming K_d = 1.25 μM [6]. In these conditions, we found $s_{20,w}$(MalE-G5P) = 5.22 ± 0.05 S.

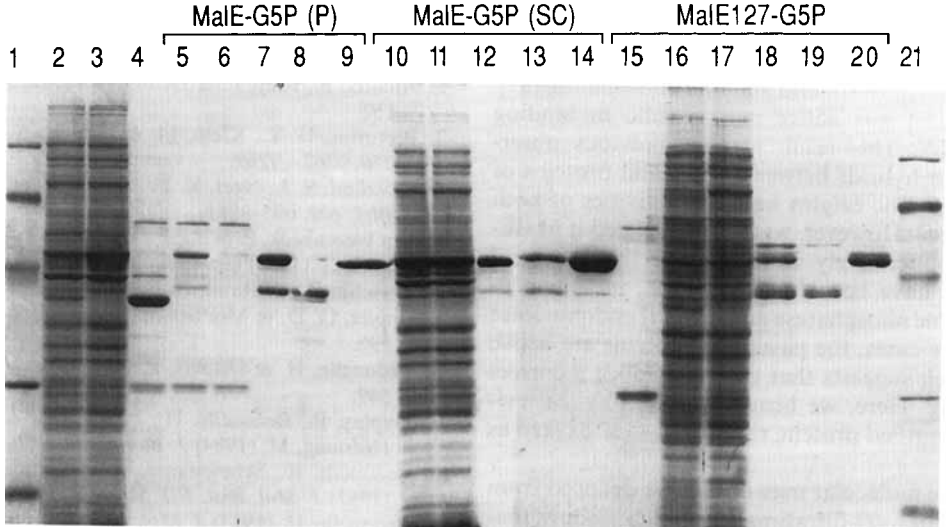

Fig. 3. *Cellular location and purification of* MalE-G5P. Cellular location. The figure compares the shocked-cell and periplasmic extracts of strains overproducing the wild-type MalE (lanes 2 and 4), MalE-G5P (lanes 3 and 5) and MalE127-G5P (lanes 16 and 15). Purification. Lanes 5–9, MalE-G5P from a periplasmic extract; lanes 10–14, MalE-G5P from a shocked-cell extract; lanes 16–20, MalE127-G5P from a shocked-cell extract. For each purification, we analysed the initial material (lanes 5, 10, 16), the fraction flowing through the amylose column (lanes 6, 11, 17), a fraction purified on the amylose column (lanes 7, 12, 18), the fraction flowing through the denatured-DNA column (lanes 8, 13, 19) and a fraction purified on the DNA column (lanes 9, 14, 20). Lanes 1 and 21, molecular mass markers: phosphorylase *b*, 97.4 kDa; bovine serum albumin, 66.2 kDa; hen egg white ovalbumin, 42.7 kDa; carbonic anhydrase, 31.0 kDa; soybean trypsin inhibitor, 21.5 kDa. We loaded amounts of the extracts and flow-through fractions originating from 200 µl of a culture at $A_{600} = 1.0\ cm^{-1}$

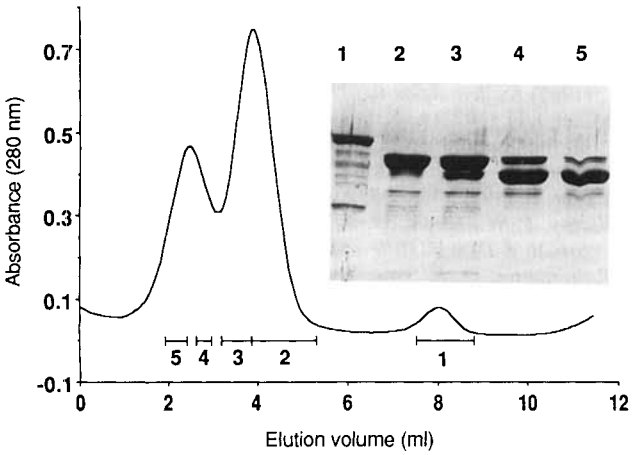

Fig. 4. *Separation of* MalE-G5P *and its main contaminant by zonal centrifugation.* A protein preparation (0.6 mg in 200 µl, containing about 45 µM), partially purified by chromatography on cross-linked amylose, was layered on top of a 5–20% sucrose gradient and centrifuged as described in Materials and Methods. The volumes are plotted from the top of the gradient. The bracketed lines indicate which fractions of the gradient were manually collected. 50 µl fraction 1, containing catalase, and 10 µl fractions 2–5 were analysed by polyacrylamide gel electrophoresis, as shown in the insert

When the same experiment was repeated with an impure preparation of MalE-G5P, corresponding to a fraction that had been eluted from the amylose column and contained both MalE-G5P and its main degradation product (see above), the profile of ultraviolet absorbance showed two peaks. Analysis of the peak contents by polyacrylamide gel electrophoresis revealed that the fast-sedimenting species mainly corresponded to MalE-G5P and the slow one to the degradation product (Fig. 4).

The Stokes radii (R_S) were determined by gel filtration. Pure MalE-G5P was injected onto the chromatography column at concentrations between 1.8 µM and 38.5 µM. At elution, the corresponding maximum, local concentrations were 0.3 µM and 6.5 µM. We found R_S(MalE) = 2.6 ± 0.3 nm and R_S(MalE-G5P) = 2.8 ± 0.3 and 3.9 ± 0.3 nm for the lowest and highest concentrations tested. The difference between these values is consistent with the hypothesis of an equilibrium between a monomeric and a dimeric form of MalE-G5P, giving a majority of monomers at low protein concentration and of dimers at high concentration.

From the above values of $s_{20,w}$ and R_S, we calculated values for the molecular mass of MalE as 41 ± 5 kDa and that of MalE-G5P as 98 ± 9 kDa at high protein concentration.

DISCUSSION

In the present work, we showed that hybrid MalE-G5P was synthesized and exported in large quantities in *E. coli* by a signal-peptide-dependent mechanism. The hybrid was purified on cross-linked amylose; it had the same specificity for single-stranded DNA as native G5P and formed dimers.

Our results confirm that normally cytoplasmic proteins can be translocated across the bacterial inner membrane. Previous examples include a fragment of TrpA [27], the Klenow fragment of DNA polymerase I [8], the human superoxide dismutase [28], the *Pst*I endonuclease [29] and the chicken muscle triose-phosphate isomerase [26]. β-Galactosidase can be translocated when synthesized in low amounts [30]. In the case of triose-phosphate isomerase, the exported enzyme is inactive.

Cytoplasmic proteins of widely different lengths have been fused to MalE: 1017 residues for β-galactosidase [31], 605 residues for the Klenow polymerase [8], 318 residues for the *Pst*I endonuclease [29] and 87 residues for G5P (this work). So far, the short polypeptides appear to be exported more efficiently than the long ones. These differences might correspond to the chance occurrence of sequences which inhibit export, clusters of basic residues for example [32]. A similar

conclusion was deduced from the study of tripartite proteins in which fragments of β-galactosidase were inserted between an N-terminal portion of MalE and alkaline phosphatase [33].

Hybrid MalE-G5P was active and specific in binding single-stranded DNA. This result extends previous observations showing that hybrids between MalE and proteins of prokaryotic or eukaryotic origins keep the activities of both partners [8, 29, 34, 35]. However, we have not tested if MalE-G5P binds DNA cooperatively, as native G5P [2, 6]. Several oligomeric proteins have been fused to MalE, including β-galactosidase, alkaline phosphatase and the *Pst*I endonuclease [29, 34]. In the three cases, the passenger proteins are active in the hybrids, which suggests that they can adopt a correct quaternary structure. Here, we brought direct physical evidence, with highly purified protein, that MalE-G5P existed as a dimer.

The values of the molecular masses that we deduced from our sedimentation and gel-filtration experiments show that MalE-G5P dimerized in conditions where MalE was a monomer, i.e. in the presence of maltose, and thus that the dimerization was due to the G5P portion of the hybrid. Note that the ratio of the molecular masses for MalE-G5P and MalE in our experiments, equal to 2.4, is independent of the particular value of the protein-specific volume that we used. We did not determine the dissociation constant for the dimer of MalE-G5P, but our results suggest that it is within the same order of magnitude as that for native G5P.

From the R_S values determined in this work and the theoretical molecular masses, we calculated the R_S/R_0 ratios for MalE and MalE-G5P, respectively 1.11 and 1.25, where R_0 is the radius of the sphere having the same mass, density and thus volume as the protein of interest. The ratio for MalE-G5P is too high to be explained only by solvation and indicates an important assymetry of the molecule [24], consistent with a model in which two large MalE domains are linked by a small G5P domain.

Thus, we have shown that a cytoplasmic and dimeric protein can be exported to the periplasmic space and attain a correct folding, having normal activity, specificity and quaternary structure.

G5P has been extensively studied. The analysis of the relationships between its structure and function has been undertaken by site-directed mutagenesis [36]. However, because its purification scheme includes a chromatographic step on denatured-DNA – cellulose [2], mutant proteins having an impaired DNA-binding activity might be difficult to study. Such mutations could be either in the DNA-binding domain or in the dimerisation domain of the molecule. Hybrid MalE-G5P should be useful in the study of such mutant G5P. We have shown that it is possible to dimerise a protein vector, MalE, by fusion to G5P, a 87 residues long polypeptide. G5P or a portion of it could be used to manipulate the quaternary structure of proteins. Finally, because the crystallographic structures of both MalE and G5P are known [7, 37] and because their stability, folding and, in the case of MalE, export have been extensively studied [36, 38], MalE-G5P could be used to further understand the relationships between folding and export and to compare the properties of a hybrid protein with those of its constituents.

We thank Alain Chaffotte for advice and Michel Goldberg for help in the analytical centrifugation experiments. This work was supported in part by grants from the *Institut Français du Pétrole* and the University Paris 7.

REFERENCES

1. Oey, J. L. & Knippers, R. (1972) *J. Mol. Biol. 68*, 125–138.
2. Alberts, B., Frey, L. & Delius, H. (1972) *J. Mol. Biol. 68*, 139–152.
3. Pretorius, H. T., Klein, M. & Day, L. A. (1975) *J. Biol. Chem. 250*, 9262–9269.
4. Cavalieri, S. J., Neet, K. E. & Goldthwait, D. A. (1976) *J. Mol. Biol. 102*, 697–711.
5. van Wezenbeek, P. M. G. F., Hulsebos, T. J. M. & Schoenmakers, J. G. G. (1980) *Gene 11*, 129–148.
6. Pörschke, D. & Rauh, H. (1983) *Biochemistry 22*, 4737–4745.
7. Brayer, G. D. & McPherson, A. (1983) *J. Mol. Biol. 169*, 565–596.
8. Bedouelle, H. & Duplay, P. (1988) *Eur. J. Biochem. 171*, 541–549.
9. Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W. & Hofnung, M. (1984) *J. Biol. Chem. 259*, 10606–10613.
10. Bedouelle, H., Schmeissner, U., Hofnung, M. & Rosenberg, M. (1982) *J. Mol. Biol. 161*, 519–531.
11. Bedouelle, H. (1983) *J. Mol. Biol. 170*, 861–882.
12. Raibaud, O., Vidal-Ingigliardi, D. & Richet, E. (1989) *J. Mol. Biol. 205*, 471–485.
13. Bedouelle, H., Bassford, P. J. Jr, Fowler, A. V., Zabin, I., Beckwith, J. & Hofnung, M. (1980) *Nature 285*, 78–81.
14. Kellermann, O. & Ferenci, T. (1982) *Methods Enzymol. 90*, 459–463.
15. Richarme, G. (1982) *Biochem. Biophys. Res. Commun. 105*, 476–481.
16. Duplay, P., Szmelcman, S., Bedouelle, H. & Hofnung, M. (1987) *J. Mol. Biol. 194*, 663–673.
17. Carter, P., Bedouelle, H. & Winter, G. (1985) *Nucleic Acids Res. 13*, 4431–4443.
18. Waye, M. M. Y., Verhoeyen, M. E., Jones, P. T. & Winter, G. (1985) *Nucleic Acids Res. 13*, 8561–8571.
19. Miller, J. H. (1972) *Experiments in molecular genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
20. Labouze, E. & Bedouelle, H. (1989) *J. Mol. Biol. 205*, 729–735.
21. Layne, E. (1957) *Methods Enzymol. 3*, 447–454.
22. Kansy, J. W., Clack, B. A. & Gray, D. M. (1986) *J. Biomol. Struct. & Dyn. 3*, 1079–1110.
23. Kellermann, O. & Szmelcman, S. (1974) *Eur. J. Biochem. 47*, 139–149.
24. Marshall, A. G. (1978) *Biophysical Chemistry*, pp. 181–205, John Wiley & Sons, New York.
25. Neu, H. C. & Heppel, L. A. (1965) *J. Biol. Chem. 240*, 3685–3692.
26. Summers, R. G. & Knowles, J. R. (1989) *J. Biol. Chem. 264*, 20074–20081.
27. Copeland, B. R., Su, T. Z. & Oxender, D. L. (1986) *Microbiology*, pp. 266–269, American Society for Microbiology, Washington DC.
28. Takahara, M., Sagai, H., Inouye, S. & Inouye, M. (1988) *Bio/Technology 6*, 195–198.
29. di Guan, C., Li, P., Riggs, P. D. & Inouye, H. (1988) *Gene 67*, 21–30.
30. Freudl, R., Schwarz, H., Kramps, S., Hindennach, I. & Henning, U. (1988) *J. Biol. Chem. 263*, 17084–17091.
31. Bassford, P. J. Jr, Silhavy, T. J. & Beckwith, J. R. (1979) *J. Bacteriol. 139*, 19–31.
32. Von Heijne, G. (1988) *Biochim. Biophys. Acta 947*, 307–333.
33. Lee, C., Li, P., Inouye, H., Brickman, E. R. & Beckwith, J. (1989) *J. Bacteriol. 171*, 4609–4616.
34. Maina, C. V., Riggs, P. D., Grandea III, A. G., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A. & di Guan, C. (1988) *Gene 74*, 365–373.
35. Szmelcman, S., Clément, J. M., Jehanno, M., Schwartz, O., Montagnier, L. & Hofnung, G. (1990) *J. AIDS*, in the press.
36. Sandberg, W. S. & Terwilliger, T. C. (1989) *Science 245*, 54–57.
37. Quiocho, F. A., Meador, W. E. & Pflugrath, J. W. (1979) *J. Mol. Biol. 133*, 181–184.
38. Park, S., Liu, G., Topping, T. B., Cover, W. H. & Randall, L. L. (1988) *Science 239*, 1033–1035.