JOURNAL OF BACTERIOLOGY, June 1995, p. 3045–3051 Vol. 177, No. 11
0021-9193/95/$04.00+0

ComEA, a *Bacillus subtilis* Integral Membrane Protein Required for Genetic Transformation, Is Needed for Both DNA Binding and Transport

G. S. INAMINE AND D. DUBNAU*

Public Health Research Institute, New York, New York 10016

Received 11 January 1995/Accepted 27 March 1995

The competence-related phenotypes of mutations in each of the four open reading frames associated with the *comE* locus of *Bacillus subtilis* are described. *comEA* and *comEC* are required for transformability, whereas the products of *comEB* and of the overlapping *comER*, which is transcribed in the reverse direction, are dispensable. Loss of the *comEA* product decreases the binding of DNA to the competent cell surface and the internalization of DNA, in addition to exhibiting a profound effect on transformability. The *comEC* product is required for internalization but is dispensable for DNA binding. ComEA is shown to be an integral membrane protein, as predicted from hydropathy analysis, with its C-terminal domain outside the cytoplasmic membrane. This C-terminal domain possesses a sequence with similarity to those of several proteins known to be involved in nucleic acid transactions including UvrC and a human protein that binds to the replication origin of the Epstein-Barr virus.

Competent cells of *Bacillus subtilis* can bind, process, and internalize exogenous DNA via a process requiring several known proteins (for a review, see reference 6). Physical contact is first made with cell surface receptors of unknown identity, resulting in irreversible binding, which is followed by processing of the DNA by double-stranded cleavage and presentation to a transporter apparatus. Following internalization of single-stranded DNA, integration into the recipient chromosome occurs by homologous recombination. Our present understanding of how the competence apparatus functions relies largely on kinetic studies of the binding and uptake of radiolabeled DNA. A more detailed understanding of the specific role that each protein plays in the overall process in *B. subtilis* is the topic of this and other recent investigations (1, 4, 12, 20–22, 26).

comE is unique among four known competence transcription units, in that open reading frames (ORFs) within it have been shown in preliminary experiments to be essential for both the binding and uptake of DNA (11, 12). In these reports we have described the identification, cloning, sequencing, and transcription mapping of the *comE* operon. This locus consists of four ORFs (see Fig. 1). Three are transcribed in the forward direction from a single, major apparent $E\sigma^A$ promoter, with *comEB* and *comEC* additionally transcribed from a minor promoter. The fourth ORF (*comER*) is transcribed in the reverse orientation from that of a promoter that is located near the end of *comEA. comEA* is preceded by a long untranslated leader sequence. Transcription of *comER* overlaps this leader and partially overlaps *comEA*. Insertion of Tn*917-lacZ* in *comEA* or *-C* has revealed that the latter is required only for DNA uptake but that insertion in *comEA* results in a binding defect. Since Tn*917-lacZ* insertions are polar and since transcription of *comEA* and *comER* overlaps, it was unclear whether inactivation of *comEA*, *comEB*, or *comER* was responsible for the observed binding-deficient phenotype of the Tn*917 comEA* insertion.

In this study, we examined the roles of *comEA*, *comEB*, *comER*, and *comEC* following the construction of a set of mutations in the first three of these ORFs. The altered ORFs were individually recombined into the chromosome at the *comE* locus, where the effects on competence expression were evaluated. A single gene product, ComEA, is involved in both the binding and the uptake of DNA, while ComEC is needed only for uptake. The membrane topology of ComEA was also examined.

MATERIALS AND METHODS

Production of antibodies against ComEA. The inferred sequence of ComEA contains a single predicted transmembrane segment near the N terminus. The hydrophilic domain following this hydrophobic segment (residues 33 to 205) was cloned into the expression vector pMAL-c2 between the *Bam*HI and *Hin*dIII sites. Production of the fusion protein was induced in *Escherichia coli* JM109 cells growing exponentially in Luria broth (10 g of tryptone [Difco], 5 g of yeast extract [Difco], 5 g of NaCl, 1,000 ml of water) containing ampicillin (100 μg/ml) by the addition of isopropyl-β-D-thiogalactopyranoside (0.5 mM) and then by growth for 3 h. Cells were harvested from 100 ml of culture, washed in buffer, and disrupted by sonication. The suspension was cleared of unbroken cells and large debris by centrifugation at 10,000 × *g* for 10 min. The cleared supernatant was added to amylose-Sepharose (5 ml) and incubated for 120 min at 4°C. The Sepharose beads were washed twice in a total of 50 ml of buffer and then eluted with 10 mM maltose in buffer. In the description above, buffer refers to 20 mM Tris-Cl (pH 7.4), 200 mM NaCl, 1 mM EDTA, and 10 mM β-mercaptoethanol. The pMAL-ComEA fusion protein was further purified by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis and visualized by incubation of the gel with 0.25 M KCl (10). The protein-containing band was excised and macerated. Antibodies were raised in guinea pigs at CoCalico Biologicals, Inc., Reamstown, Pa., by intramuscular injection of approximately 40 μg of protein emulsified with Freund's adjuvant. Fourteen days after the first injection, the animals were boosted with approximately the same amount of protein without Freund's adjuvant and were then booster injected five additional times at roughly 1-week intervals.

Strains and plasmids. For the construction of strains that contained Campbell-type chromosomal insertions, an integrative vector, pUCCM18, was made by inserting a chloramphenicol resistance cassette prepared by PCR from plasmid pC194 into pUC18 at the unique *Nde*I site. Transcription of the resistance gene is in the direction away from the multiple cloning site of the vector. All of the *B. subtilis* strains were derivatives of strain 168 and are described in Table 1. These strains were all isogenic with BD630 (*hisA1 leu-8 metB5*).

DNA manipulations. Constructions of the in-frame deletions of *comE* were

* Corresponding author. Mailing address: Public Health Research Institute, 455 First Ave., New York, NY 10016. Phone: (212) 578-0842. Fax: (212) 578-0804. Electronic mail address: dubnau@phri.nyu.edu.

TABLE 1. Strains used in this study

Strain	Genotype	Source or reference
BD630	*hisA1 leu-8 metB5*	
BD204	*hisA1 thyA thyB*	
BD1245	*hisA1 leu-8 metB5 comEA413*	11, 12
BD1246	*hisA1 leu-8 metB5 comEC510*	11, 12
BD1247	*hisA1 leu-8 metB5 comEC518*	11, 12
BD1248	*hisA1 leu-8 metB5 comGA12*	1
BD2327	*hisA1 leu-8 metB5 comEAΔ1*	This report
BD2328	*hisA1 leu-8 metB5 comEAΔ2*	This report
BD2389[a]	*hisA1 leu-8 metB5 comEA413 amyE*::*comEAΔ2*	This report
BD2390[a]	*hisA1 leu-8 metB5 comEA413 amyE*::*comEA*$^+$	This report

[a] Carries the *comEA413* mutation in the *comE* locus and a copy of the indicated *comEA* allele inserted at the *amyE* locus.

carried out by first cloning the region of interest and then by isolation of the fragment, digestion and removal of coding sequence by restriction digestion, and religation and cloning into the vector pUCCM18. In each case, a clone of the undeleted region was retained for testing of its capacity to complement a corresponding competence-deficient deletion strain. For *comEAΔ1*, a PCR product spanning nucleotide residues 928 to 2604 (12) was first cloned, and then a 123-bp fragment was removed by *Hha*I enzyme digestion. *comEAΔ2* was constructed by sequentially digesting a fragment consisting of residues 203 to 2577 with the enzyme *Nsi*I, creating a blunt end by digestion with mung bean nuclease, digesting with the enzyme *Eco*RV, and then ligating, thus removing the coding sequence for 22 aminoacyl residues, adding a single leucine codon, and positioning the last residue just upstream of tandem terminator codons. Removal of the 3′ region of *comEA* also eliminated the putative −35 region of a predicted minor transcription promoter located downstream from *comEA* (12). *comEBΔ3* was made by *Sau*3A digestion, removal of 252 bp from a fragment spanning the region of nucleotide residues 1460 to 3134, and then digestion with *Pst*I prior to cloning into pUCCM18. Construction of *comERΔ4* involved cloning a fragment spanning residues 203 to 1696 and the removal of two restriction fragments, 144 and 120 bp in length, by digestion with *Taq*I and then by ligation.

PhoA fusions were constructed by using a plasmid that encoded PhoA, lacking its 14 amino-terminal aminoacyl residues, that was cloned into the *Pst*I site of vector pUCCM18. Hybrid proteins containing all or part of ComEA fused to the 15th residue of PhoA were constructed by designing PCR primers that placed sections of ComEA in frame with PhoA. The primers contained *Sal*I and *Eco*RI sites which were used for cloning. Constructs consisting of the first 10, 52, or all 205 aminoacyl residues of ComEA fused to the PhoA moiety were made.

Transformation of *B. subtilis*. Competent cells were prepared and transformation was carried out as described previously, using either the one-step (2) or the two-step (8) procedure.

Preparation of radiolabeled DNA and measurement of binding and uptake. *B. subtilis* BD204 (*hisA1 thyA thyB*) was grown in the presence of [*methyl*-^{3}H]thymidine, and its DNA was extracted and purified as described elsewhere (7). Competent cultures were incubated with [^{3}H]DNA, and binding and uptake were measured by liquid scintillation counting of washed samples following incubation with (uptake) or without (binding plus uptake) DNase (50 μg ml^{-1}). Incubation with DNA was carried out for 30 to 45 min at 37°C. DNase treatment was for 5 min at 37°C. These procedures are described elsewhere (21).

Isolation and treatment of membranes. Cells were harvested, washed once, and then resuspended in 10 mM Tris-Cl (pH 8.0)–5 mM EDTA, and then egg white lysozyme was added to 0.5 mg/ml. Crude membranes were prepared by pelleting the resultant lysate. In the case of protoplast preparations, treatment with 0.3 mg of lysozyme per ml was carried out in a solution containing 25% (wt/vol) sucrose, 10 mM Tris-Cl (pH 8.0), and 5 mM $MgCl_2$. Incubation in the presence of lysozyme was at 37°C for 15 to 30 min. Protoplast formation was monitored by microscopic examination. When purified membrane preparations were required, crude pellets or protoplasts initially prepared in sucrose were lysed in 10 mM Tris-Cl (pH 8.0)–5 mM $MgCl_2$. For all membrane preparations, DNase was added to 10 μg/ml, the free Mg^{2+} concentration was adjusted to 5 mM, and the samples were incubated at room temperature for 5 to 10 min. Hypotonically lysed protoplasts were layered onto 42% (wt/vol) sucrose made in 10 mM Tris-Cl (pH 7.5) and centrifuged in a Beckman TLS55 rotor for 60 min. The whitish, turbid band of membrane vesicles was collected, diluted in 10 mM Tris-Cl (pH 8.0)–1 mM EDTA, and pelleted. Membrane preparations were stored at −20°C in 10 mM Tris-Cl (pH 8.0)–1 mM EDTA.

Whole cells or intact protoplasts, prepared as described above, were treated at the indicated concentrations of proteinase K for 30 min at room temperature. Digestions were halted by the addition of phenylmethylsulfonyl fluoride to a final concentration of 4 mM. Protoplast samples were diluted 10- to 50-fold in 10 mM Tris-Cl (pH 8.0)–1 mM EDTA–4 mM phenylmethylsulfonyl fluoride, and membranes were pelleted by centrifugation at 30,000 rpm in a Beckman 50.1 Ti rotor for 30 min. Whole cells were lysed and treated as described above.

Membrane protein was extracted by vortexing in 0.1 N NaOH for 15 s at room temperature, which was followed by centrifugation at 55,000 rpm in a Beckman TLA100.2 rotor for 15 min. Extracted proteins were precipitated from the supernatant by the addition of trichloroacetic acid to 10%. After centrifugation for 15 min, the pellet was washed twice with 100% acetone. The washed precipitate was resuspended in 6 M urea prior to protein estimation and Western blot (immunoblot) analysis.

Assay of alkaline phosphatase. For analysis of fusions expressed in *E. coli*, strains were streaked on Luria-Bertani agar plates containing 5-bromo-4-chloro-3-indolylphosphate (BCIP) at 40 μg/ml, isopropyl-β-D-thiogalactopyranoside (100 μg/ml), and ampicillin (100 μg/ml). The expression of alkaline phosphatase activity was judged qualitatively by the production of blue product. The expression of alkaline phosphatase fusions in *B. subtilis* was assayed in cells grown in liquid culture as described elsewhere (24), except that permeabilization of the cells with detergent and chloroform was omitted. For the calculation of specific activities, the amounts of protein were estimated from Klett readings on culture samples by using an empirically determined calibration curve.

Western blot analysis. Proteins were solubilized in the sample solution described by Laemmli (16) and then heated to 65°C for 30 min. Polypeptides were resolved in slab gels containing 14% polyacrylamide after stacking in 5% polyacrylamide. The gels were equilibrated in transfer buffer (48 mM Tris [pH 9.2], 39 mM glycine, 0.05% SDS, 20% methanol) for 10 to 15 min and then electrophoretically transferred to prewetted nitrocellulose sheets (Schleicher and Schuell; 0.45-μm pore size) for 30 min at 12 V in a semidry transfer apparatus (Bio-Rad). Transferred proteins were specifically visualized by using chemiluminescent substrate applied to peroxidase-decorated target with an ECL kit (Amersham, Inc.).

RESULTS

Introduction of mutations into *comE*. Deletions were made in each intact ORF of interest by restriction digestion and ligation. Strains containing these deletions were constructed by integration of the resultant plasmids into the chromosome of competent *B. subtilis* IS75. Integration occurred by Campbell-type recombination, and recombinant strains were identified by expression of a chloramphenicol resistance marker carried on the plasmid vector, pUCCM18, which cannot replicate in *B. subtilis*. Spontaneous loss of vector sequence and the encoded drug resistance gene from the chromosome occurred via homologous recombination between the duplicated sequences during growth in the absence of antibiotic selection. Such events result in either replacement of the wild-type gene by the in-frame deletion construct or restoration of the wild-type sequence. Chloramphenicol-sensitive isolates were identified by replica plating and then assayed for truncated versions of the targeted *comE* gene by PCR analysis using chromosomal DNA and the appropriate primers. In addition to these mutations, we have also studied the phenotypes associated with Tn*917-lacZ* insertions in *comEA* (*comEA413*) and in *comEC* (*comEC510* and *-518*) (11, 12). The locations of all of these mutations and the structure of the *comE* locus are summarized in Fig. 1.

Transformation phenotypes of *comE* mutations. A deletion of 88 contiguous aminoacyl residues from ComER did not significantly affect the transformation frequency of the mutant strain (Table 2). Similarly, when an in-frame deletion of 84 residues was made in ComEB, no change in transformability was observed (Table 2). It appears that ComER and ComEB, which resemble pyrroline-5′-carboxylate reductase and dCMP deaminase, respectively (12), are probably not required for the development of competence or for the binding and transport of transforming DNA. We cannot exclude the possibility that the deletions in *comER* and in *comEB*, which remove 28 and 44% of each protein, respectively, leave intact a domain which is sufficient for transformation.

In order to examine the role of ComEA, two deletion derivatives were tested. Δ*1* removes the coding sequence for 41 of 205 amino acids (residues 102 to 140) from the middle of the protein. Δ*2* removes 22 residues (183 to 205) from the extreme

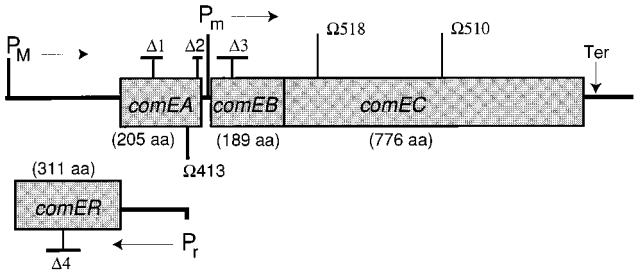

FIG. 1. The *comE* locus. The four ORFs are indicated, together with the approximate locations of the deletion (Δ) and transposon insertion (Ω) mutations referred to in the text. The horizontal bars associated with each deletion indicate the relative sizes of the deletions, and the vertical bars represent their locations. The major (P_M), minor (P_m), and reverse (P_r) promoters are also indicated. The numbers in parentheses refer to the numbers of amino acid (aa) residues predicted for each ORF.

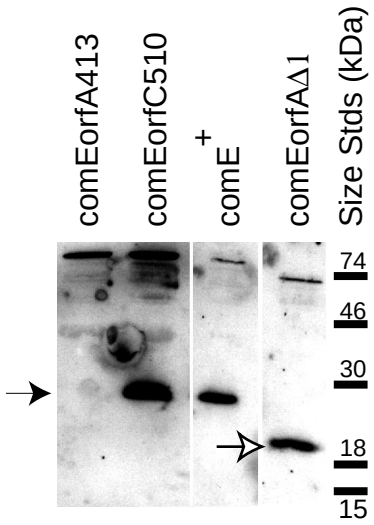

FIG. 2. Western blot analysis of ComEA mutant proteins in membrane preparations. The positions of ComEA and of ComEAΔ1 are indicated by solid and empty arrowheads, respectively. The predicted molecular masses of ComEΔ1 and of ComEA are 16,949 and 21,756 Da, respectively. The predicted size of the truncated ComEA413 protein is approximately 16,345 Da. ComEAΔ1 runs more slowly than predicted from its sequence. The three panels are taken from the same gel and are separated to indicate that irrelevant lanes were deleted. Stds, standards.

carboxyl terminus of the predicted ComEA protein and introduces a single leucine immediately before the stop codon. The *comEΔ1* mutant exhibited a transformation frequency 50-fold below that of the parental strain (Table 2). When the *Δ2* mutation was present in the *comE* locus, transformation was reduced by at least 10^7-fold (Table 2). This is identical to the phenotype observed with a Tn*917-lacZ* insertion (the *comEA413* mutation) in the C-terminal region of *comEA* (11, 12). It should be noted that Tn*917-lacZ* insertions in *comEC* also decrease transformability by at least 10^7-fold. However, the effect of the *comEA* Tn*917-lacZ* insertion cannot be explained entirely by polarity on *comEC*, because the minor promoter drives sufficient *comEC* transcription to support about 25 to 50% of the wild-type level of transformability. This follows from the observation that expression of an intact copy of *comEA* in *trans*, under the control of the *comE* promoter, restored the competence of the *comEA413* mutant to about 40% of the wild-type level (Table 2), which is consistent with the presence of this minor promoter. On the other hand, when the carboxyl-terminal deletion mutation (*Δ2*) was placed in *trans*, no complementation was observed. This argument, as well as the profound effect of the nonpolar *Δ2* mutation, shows that ComEA is essential for transformation and that the removal of its C-terminal segment is particularly deleterious.

The stabilities of the various mutant ComEA gene products were examined by Western blot analysis (Fig. 2). The centrally located in-frame deletion (*Δ1*) yielded a signal that was as intense as that of the full-length (wild-type) product, but with the expected faster electrophoretic migration. As expected, the ComEA signal in a *comEC510* mutant was unaffected. Since the location of the Tn*917-lacZ* insertion in *comEA413* (12) and the sequence of Tn*917* (33) are known, the predicted molecular mass of the truncated ComEA413 protein is estimated as 16,345 Da. However, extracts of BD1245, a *comEA413* mutant, exhibited no detectable Western signal with anti-ComEA serum, either on the portion of the gel shown in Fig. 2 or in the lower-molecular-weight range (not shown). Western blot analysis of sucrose gradient-purified membrane fractions from competent cultures carrying the *Δ2* mutation in the carboxyl-terminal region of ComEA also did not exhibit any detectable signal (not shown). It appears that the relatively mild loss of competence associated with *Δ1* was due to the in-frame deletion, whereas the more severe phenotype associated with *Δ2* could be explained by a dramatic loss of ComEA stability. We also infer that deletion of the extreme C terminus of the ComEA protein probably resulted in a loss of protein stability.

Binding of [^{3}H]DNA. Radiolabeled DNA was used to determine the relative extents of DNA binding and uptake by wild-type and mutant cultures grown through the competence regimen (Table 3). Total [^{3}H]DNA associated with cells after washing was used as a measure of binding, and the amount of radioactivity that persisted after DNase treatment was used to determine uptake. The *comGA12* strain, which carries a polar

TABLE 2. Transformation phenotypes of *comE* mutants

Strain	Mutation	ORF affected	No. of residues missing (total no. of wild-type residues)[a]	Transformation frequency[a]
BD630	None	None	None	1.0
BD2327	*Δ1*	*comEA*	41 (205)	0.02
BD2328	*Δ2*	*comEA*	22 (205)	$<5 \times 10^{-8}$
BD2326	*Δ4*	*comER*	88 (311)	0.8
BD2329	*Δ3*	*comEB*	84 (189)	1.0
BD2389[b]	*comEA413 amyE::comEAΔ2*	*comEA*		6.9×10^{-8}
BD2390[b]	*comEA413 amyE::comEA*	*comEA*		0.40

[a] The transformation frequencies are normalized to the wild-type values.
[b] Carries the *comEA413* transposon insertion in *orfA* and a copy of the *comEAΔ2* mutation (BD2389) or of the intact *comEA* gene (BD2390) in the *amyE* locus.

TABLE 3. Binding and uptake of DNA

Strain	Mutation	Total DNA bound[a]	Fraction of DNase-resistant DNA[b]
BD630	None	1.0	0.52
BD2327	*Δ1*	0.54	0.01
BD2328	*Δ2*	0.05	0
BD1245	*comEA413*	0.24	0.07
BD1247	*comEC518*	1.43	0.07
BD1248	*comGA12*	0.13	0

[a] The values are normalized to that of the wild-type strain (BD630) and represent the total [^{3}H]DNA that is irreversibly associated with the cells. Each value in this column as well as the one representing the fraction of DNase-resistant DNA is the average of at least three independent determinations, except for the values for *comEAΔ2*, which are the averages of two experiments.

[b] Values are the fraction of the total cell-associated [^{3}H]DNA which is resistant to the action of exogenous DNase.

insertion of Tn*917-lacZ* in *comEA* of *comG* and is deficient in the binding of DNA (3, 11), was included as a negative control.

A strain carrying the *Δ1* mutation in *comEA* was able to bind DNA about half as well as the wild type but was associated with much less DNase-resistant radioactivity relative to the total bound DNA. The low DNase-resistant radioactivity is consistent with the 50-fold reduction in transformability noted for the *Δ1* strain (Table 2). *comEA413* is a Tn*917-lacZ* insertion in *comEA*. This mutation resulted in both a 4-fold-reduced level of binding and a marked decrease in the fraction of bound DNA rendered DNase resistant. Although the insertion of Tn*917-lacZ* is polar, a minor promoter is present upstream from *comEB* (12). As noted above, we have found that this minor promoter can allow enough ComEC synthesis to support at least 25% of the wild-type level of transformability. Thus, the reduced proportion of DNase-resistant donor DNA associated with *comEA413* cells cannot be entirely explained by this partial polar effect on *comEC*. The latter conclusion is reinforced by the results shown for *Δ2*, which are similar to those obtained with the *comEA413* mutation; binding is reduced and internalization is nearly eliminated (Table 3). We do not know whether the apparently greater reduction of total binding in the *Δ2* mutant compared with the strain carrying *comEA413* is significant, since the former strain has been tested only twice in this assay. We conclude that the loss of 41 aminoacyl residues from the middle of ComEA decreases binding moderately but exhibits a strong effect on the conversion of donor DNA to DNase resistance. The total loss of ComEA protein in the *Δ2* and *comEA413* mutants decreases DNA binding more severely than *Δ1*, and, like the latter, has a drastic effect on the conversion of transforming DNA to DNase resistance.

comEC510 and *comEC518* are Tn*917-lacZ* insertions in *comEC*. Strains carrying *comEC518* exhibited a greater-than-normal level of total DNA binding and a sharply reduced amount of DNase-resistant donor DNA. An increase in total cell-associated DNA is not unexpected in a mutant that is specifically impaired in uptake, since the internalization of DNA is accompanied by the degradation of one strand equivalent. We have also observed increased binding and reduced conversion to DNase resistance in a *comEC510* strain (not shown). These two strains are essentially untransformable (11) (Table 2). We conclude from these studies that ComEC is required for DNA uptake but not for binding.

ComEA is an integral membrane protein with its C terminus outside. ComEA is predicted to contain a single transmembrane segment near the N terminus (residues 10 to 30) (12). This segment is both preceded and followed by several basic residues, although the excess of positive charge preceding the predicted transmembrane segment would seem to favor an arrangement with the C terminus outside (37, 38). We have addressed the cellular location and the topology of this protein more directly using a polyclonal antibody raised against a fusion of ComEA to the maltose binding protein. Figure 3A shows that the bulk of ComEA was located in the cytoplasmic membrane, as predicted, and that it was not significantly solubilized by incubation with 0.1 N NaOH. The latter fact suggests that the protein is organized as an integral membrane protein (30). Treatment of protoplasts with proteinase K revealed that ComEA was accessible from the outside (Fig. 3B), again in accordance with the prediction from sequence analysis. Additionally, the Western blot signal was diminished but not eliminated by the incubation of intact cells with proteinase K, suggesting that either peptidoglycan or other proteins partially blocked access to the protease. Since Coomassie blue staining of the proteinase K-treated and untreated samples showed no gross changes (Fig. 3C), we conclude that the cytoplasmic membrane had remained intact during the incubation with protease. These data, together with the hydropathy analysis, support a model with ComEA arranged as an integral membrane protein with a moiety exposed outside the membrane. Since, as shown below, the major epitope on ComEA recognized by our antiserum is near the C terminus, and since only 10 aminoacyl residues precede the predicted transmembrane segment, the absence of a truncated product in Western blots following proteinase K treatment argues in favor of an arrangement with the C terminus located outside the membrane, rather than the reverse.

The topology of ComEA was further examined by the construction of fusions to the *E. coli* PhoA protein (25, 34). Fusion proteins were prepared to the 10th (f10), 52nd (f52), and 205th (f205) residues of ComEA. The first fusion point was located just before the predicted transmembrane segment, and the second and third fusion points were located downstream. The third fusion attached PhoA to the ComEA C-terminal aminoacyl residue. When it was present on a plasmid in *E. coli* and streaked on an indicator plate for alkaline phosphatase, the strain carrying f10 yielded a negative response and the second two exhibited positive responses (blue colonies). These results were therefore consistent with the topological model described above.

The use of PhoA fusions to determine the topology of membrane proteins in *B. subtilis* has not been reported, although Payne and Jackson (27) showed that a fusion of the *E. coli* PhoA protein to a secreted protein in *B. subtilis* exhibited alkaline phosphatase activity. The ComEA-PhoA fusion plasmids, which were incapable of replication in *B. subtilis*, were used to transform BD630 to chloramphenicol resistance. This resulted in integration at the *comE* locus by Campbell-type recombination and placed the fusion constructs under control of the *comE* promoter. Western blots were carried out using membrane preparations from these *B. subtilis* strains with a monoclonal antibody raised against PhoA. A strong positive Western signal was detected with the f52 strain, a much weaker one was detected with f205 (Fig. 4A), and no signal was observed with the N-terminal fusion. When the anti-ComEA polyclonal antiserum was used, a strong signal was detected with f205 and a very weak signal was detected with f52 (Fig. 4B). This strongly suggests that the major epitope recognized by this antiserum was removed in f52 and was therefore located in the C-terminal domain. This is understandable, since the fusion protein used as antigen contains the C-terminal moiety of ComEA and overlaps f52 by only 20 aminoacyl residues.

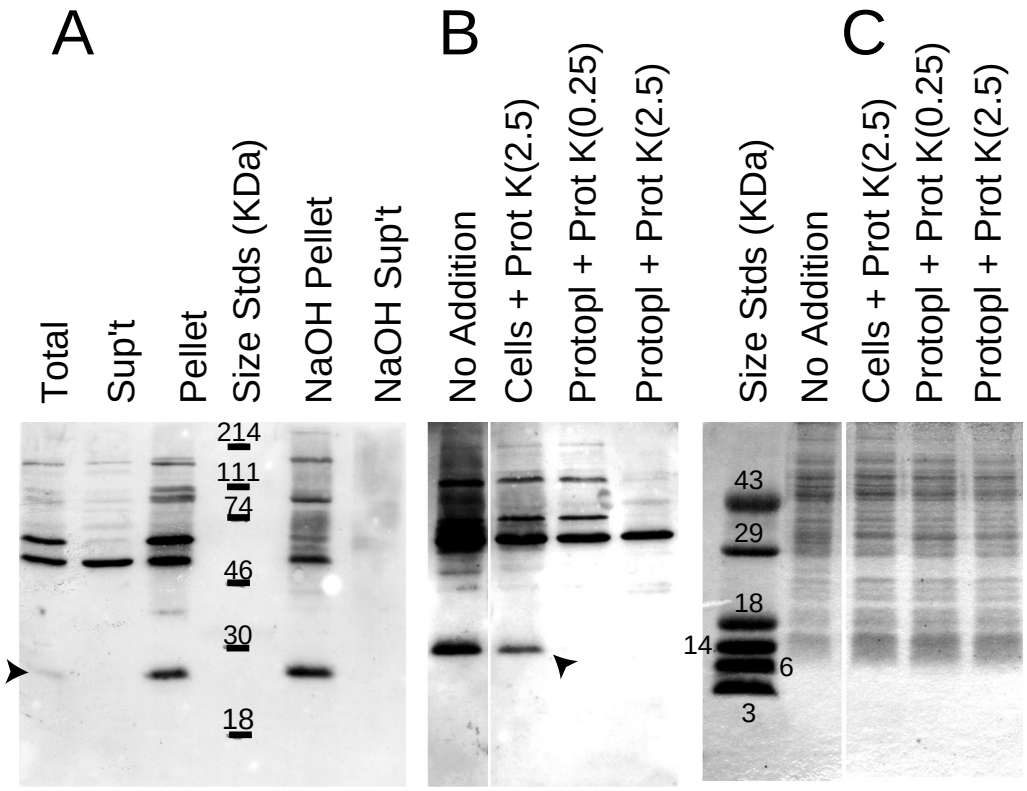

FIG. 3. Localization of ComEA. (A) Western blot analysis of ComEA localization with a total cell extract, the supernatant (supt) and pellet from a centrifuged lysate, and the pellet and supernatant from a membrane preparation extracted with 0.1 N NaOH; (B) Western blot analysis of membrane preparations from untreated cells, from whole cells treated with 2.5 mg of proteinase K (Prot K) per ml, and from protoplasts (protopl) treated with 0.25 mg and 2.5 mg of proteinase K per ml (the arrowheads in panels A and B indicate the positions of the ComEA signals); (C) Coomassie blue-stained polyacrylamide gel with the same samples run in panel B. Within all panels, identical amounts of protein were loaded in each lane. The pellet fraction in panel A is therefore enriched for the ComEA signal compared with the total extract. The predicted M_w of ComEA is 21,756. The separations within panels A and B indicate that irrelevant lanes were deleted. Stds, standards.

The alkaline phosphatase-specific activities of the three fusion strains as well as of the IS75 parent that carried no *E. coli phoA* gene were also measured. For these determinations, the cultures were grown in a medium containing 0.125 mM P_i, which represses endogenous alkaline phosphatase. The medium also contained 0.5% glucose, and the samples were taken 2 h after the end of the exponential growth phase, in order to repress expression of the sporulation-associated alkaline phosphatases yet to permit the expression of competence-related genes. The specific activities were normalized to that of the highest expressing fusion (f52), after the endogenous activity, which was 32% of the lowest fusion activity, was subtracted. The normalized activities were as follows: f52, 1.0; f205, 0.33; and f10, 0.18. The low activity of the f10 fusion construct was consistent with its weak Western signal with the anti-PhoA antiserum, suggesting that this construct turned over rapidly or

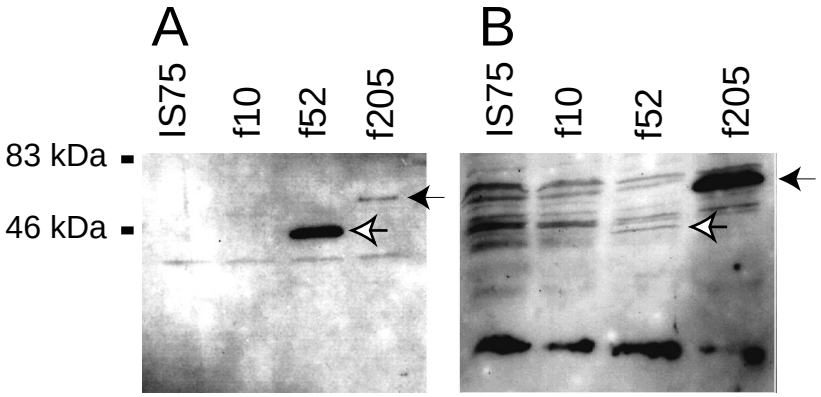

FIG. 4. Western blot analysis of PhoA-ComEA fusions. (A) Results with a monoclonal antibody against PhoA. (B) Results with a polyclonal antibody against ComEA. The solid arrows indicate the signals from the fusions to residue 205. The open arrows represent fusions to residue 52. The lanes contain membrane preparations from the wild-type strain (IS75) and from the strains with fusions to residues 10 (f10), 52 (f52), and 205 (f205). The predicted M_ws of f52 and f205 are 54,305 and 69,418, respectively.

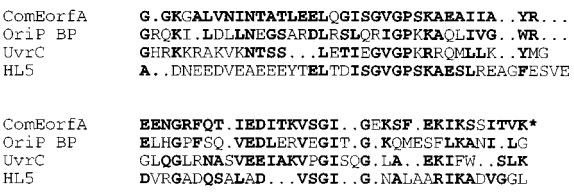

FIG. 5. Similarities of ComEA to other proteins. The sequences of UvrC (31), HL5 (32), and the OriP BP (39) are compared with that of ComEA. Boldface type indicates identity or conservative replacement of amino acid residues compared with the ComEA sequence. The following groupings were used to define conservative replacements: E and D; K, R, and H; M, L, I, and V; W, Y, and F; S, T, A, and G; N and Q; and C. The asterisk indicates the C terminus.

was expressed at a low rate. In a study of an unrelated *B. subtilis* polytopic membrane protein, we have observed that more promoter-distal PhoA fusions are underrepresented in Western blots (28). This may be a general characteristic of PhoA fusions in the *B. subtilis* host. The results obtained with these fusions in *E. coli* and the positive result obtained with f52 in *B. subtilis* support the more direct determination of ComEA topology presented above, and we conclude that the latter is a type II bitopic transmembrane protein (29), which is arranged with its C terminus outside the cytoplasmic membrane.

Similarity of ComEA to several DNA-binding proteins. A segment consisting of about 68 residues near the C terminus of ComEA shows similarity to a number of proteins that are known to interact with nucleic acids (Fig. 5). These include a human protein that binds to the origin of the Epstein-Barr virus (39), UvrC from *E. coli* (31), and a 50S ribosomal protein (HL5) from *Haloarcula marismortui* (32). Strong similarity (not shown) is also present to a predicted protein from *Dichelobacter nodosus*, whose function is unknown (17).

DISCUSSION

The four ORFs in the *comE* locus were analyzed for their roles in DNA binding and uptake by competent cells. An in-frame deletion of ComEA, which removed aminoacyl residues 102 to 140 (Δ1), lowered the transformation frequency by approximately 50-fold (Table 2). Despite this internal shortening of the protein sequence, its function with regard to DNA binding to the cell was decreased only by a factor of 2 (Table 3). A residual level of binding is also evident in the *comEA413* mutant, which is associated with the apparently complete absence of ComEA protein. A role for ComEA in the conversion to DNase resistance separate from its involvement in binding is implied by the data in Table 3. Taken together, these results suggest that ComEA is absolutely required for uptake but not for binding. Its role in binding may be indirect; perhaps it is needed to stabilize the binding apparatus or to consolidate binding which is directly mediated by another protein. The centrally located 41 aminoacyl residues removed by *Δ1* are not absolutely required for binding or for internalization, as revealed by the residual transformability of the *Δ1* mutant (Table 2). The C-terminal domain of ComEA, on the other hand, is probably needed at least for protein stability.

Our data strongly support a model of ComEA as a transmembrane protein with its C-terminal moiety exposed at the outer surface of the membrane. This was demonstrated by the protease accessibility of the protein in protoplasts and intact cells (Fig. 3) and by the alkaline phosphatase activity of two PhoA fusions. This topology is consistent with a role for ComEA in contacting DNA, since binding occurs on the cell surface, prior to uptake. The similarity of a restricted segment of ComEA to UvrC and to a protein that binds to the Epstein-Barr virus replication origin is also consistent with a role for this protein in contacting DNA, and the region of similarity is located near the C terminus of the protein, outside the cytoplasmic membrane. It is interesting to note that UvrC is a DNA repair excinuclease that cleaves DNA 5′ to the site of DNA damage (18). The region of similarity to ComEA is immediately adjacent to several residues that have been implicated in UvrC nuclease action.

The ComEC product is predicted to be a polytopic integral membrane protein (12) and is required for DNA uptake but not for binding to the cell surface. It is interesting that the loss of ComEC results in an increase in DNA binding (Table 3). This has been interpreted to mean that in this mutant, the normal degradation of one strand equivalent of DNA does not occur. ComEC is similar to DNA transport proteins recently identified in *Neisseria gonorrhoeae* (9) and in *Haemophilus influenzae* (5), suggesting conservation of the uptake machinery in these three organisms. The predicted polytopic membrane organization of ComEC, its role in DNA transport, and its toxicity when expressed in *E. coli* are all consistent with a proposed role for this protein as constituting part of an aqueous channel for DNA internalization. If this is the case, the apparent failure of ComEC mutants to degrade one strand equivalent of donor DNA may result from a requirement that the incoming DNA enter the channel to gain access to an appropriate nuclease, or ComEC protein may itself possess nuclease activity. ComFA is also a membrane-bound competence protein that is required for DNA uptake but not for binding (20–22). ComFA may provide the driving force for transport of DNA through an aqueous channel, since it resembles the so-called DEAD family of ATP-dependent DNA helicases, and its consensus ATP-binding site is required for DNA uptake in vivo (22). Therefore, it may interact with ComEC during uptake.

In contrast to ComEA and ComEC, the gene products ComEB and ComER appear to be inessential for competence, and their inactivation was not associated with any obvious growth phenotype. Why are their coding sequences present in the *comE* locus? The predicted sequences of these proteins exhibit strong similarities to dCMP deaminases and pyrroline-5′-carboxylate reductases, respectively (12), although we do not know if they possess these activities. There are several known cases of gene products that are synthesized as part of the competence regulon, in response to accumulation of the transcription factor ComK, several of which appear to play no direct role in transformation. These include the SOS response genes (14, 23), a nuclease (35), an ATP-dependent nuclease (13), and, in some genetic backgrounds, levansucrase (15). It appears that the *comEB* and *comER* products are additional members of this family. However, *comEB* may play an indirect role in competence. The start codon of *comEC* and the last codon of *comEB* are separated by only a tandem pair of UGA codons, and the *comEC* start is not preceded by an obvious ribosomal-binding site. It is possible that translation of *comEB* is required for the delivery of ribosomes to the translation start codon of *comEC*. Transcription of *comER* decreases as that of *comE* in the forward direction turns on (12). The overlapping arrangement of these transcripts may provide a regulatory mechanism, perhaps based on the positive superhelical turns expected to accumulate between converging RNA polymerase molecules (19). For instance, transcription in the reverse direction may be impeded by the activation of *comE* forward transcription in response to the accumulation of the competence transcription factor ComK (36). In addition, the transcription of *comER* early in growth may serve to inhibit the

basal forward transcription of *comE*, which might otherwise occur in the absence of ComK. This may be important, since even a low level of ComEC expression in *E. coli* is toxic (not shown) and the cell may need to prevent the inappropriate synthesis of this protein. Therefore, we suggest that the presence of both *comEB* and *comER* may serve regulatory functions for competence, although their gene products appear to be dispensable.

ACKNOWLEDGMENTS

We acknowledge useful discussions with all the members of our laboratory, and especially with Y. S. Chung.

This work was supported by NIH grant GM 43756.

REFERENCES

1. **Albano, M., R. Breitling, and D. Dubnau.** 1989. Nucleotide sequence and genetic organization of the *Bacillus subtilis comG* operon. J. Bacteriol. **171:** 5386–5404.
2. **Albano, M., J. Hahn, and D. Dubnau.** 1987. Expression of competence genes in *Bacillus subtilis*. J. Bacteriol. **169:**3110–3117.
3. **Breitling, R., and D. Dubnau.** 1990. A pilin-like membrane protein is essential for DNA binding by competent *Bacillus subtilis*. J. Bacteriol. **172:**1499–1508.
4. **Chung, Y. S., and D. Dubnau.** 1994. ComC is required for the processing and translocation of ComGC, a pilin-like competence protein of *Bacillus subtilis*. Mol. Microbiol. **15:**543–551.
5. **Clifton, S. W., D. McCarthy, and B. A. Roe.** 1994. Sequence of the *rec-2* locus of *Hemophilus influenzae*: homologies to *comE*-ORF3 of *Bacillus subtilis* and *msbA* of *Escherichia coli*. Gene **146:**95–100.
6. **Dubnau, D.** 1993. Genetic exchange and homologous recombination, p. 555–584. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
7. **Dubnau, D., and C. Cirigliano.** 1972. Fate of transforming deoxyribonucleic acid after uptake by competent *Bacillus subtilis*: size and distribution of the integrated donor segments. J. Bacteriol. **111:**488–494.
8. **Dubnau, D., R. Davidoff-Abelson, and I. Smith.** 1969. Transformation and transduction in *Bacillus subtilis*: evidence for separate modes of recombinant formation. J. Mol. Biol. **45:**155–179.
9. **Facius, D., and T. F. Meyer.** 1993. A novel determinant (*comA*) essential for natural transformation competence in *Neisseria gonorrhoeae*. Mol. Microbiol. **10:**699–712.
10. **Hager, D. A., and R. R. Burgess.** 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA polymerase, and other enzymes. Anal. Biochem. **109:**76–86.
11. **Hahn, J., M. Albano, and D. Dubnau.** 1987. Isolation and characterization of competence mutants in *Bacillus subtilis*. J. Bacteriol. **169:**3104–3109.
12. **Hahn, J., G. Inamine, Y. Kozlov, and D. Dubnau.** 1993. Characterization of *comE*, a late competence operon of *Bacillus subtilis* required for the binding and uptake of transforming DNA. Mol. Microbiol. **10:**99–111.
13. **Haijema, B.-J., L. W. Hamoen, J. Kooistra, G. Venema, and D. van Sinderen.** 1995. Expression of the ATP-dependent deoxyribonuclease of *Bacillus subtilis* is under competence control. Mol. Microbiol. **15:**203–211.
14. **Haijema, B. J., L. W. Hamoen, G. Venema, and D. van Sinderen.** Regulated expression of the *Bacillus subtilis dinR* and *recA* genes during competence. Mol. Microbiol., in press.
15. **Kunst, F., T. Msadek, J. Bignon, and G. Rapoport.** 1994. The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in *Bacillus subtilis*. Res. Microbiol. **145:**393–402.
16. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
17. **La Fontaine, S., and J. Rood (Monash University).** 1995. Personal communication.
18. **Lin, J.-J., and A. Sancar.** 1992. Active site of (A)BC excinuclease. I. Evidence for 5′ incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466 and His538 residues. J. Biol. Chem. **267:**17688–17692.
19. **Liu, L. F., and J. C. Wang.** 1987. Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA **84:**7024–7027.
20. **Londoño-Vallejo, J. A., and D. Dubnau.** 1993. *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. Mol. Microbiol. **9:**119–131.
21. **Londoño-Vallejo, J. A., and D. Dubnau.** 1994. Membrane association and role in DNA uptake of the *Bacillus subtilis* PriA analog ComF1. Mol. Microbiol. **13:**197–205.
22. **Londoño-Vallejo, J. A., and D. Dubnau.** 1994. Mutation of the putative nucleotide binding site of the *Bacillus subtilis* membrane protein ComFA abolishes the uptake of DNA during transformation. J. Bacteriol. **176:**4642–4645.
23. **Love, P. E., M. J. Lyle, and R. E. Yasbin.** 1985. DNA-damage-inducible (*din*) loci are transcriptionally activated in competent *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **82:**6201–6205.
24. **Manoil, C.** 1991. Analysis of membrane protein topology using alkaline phosphatase and β-galactosidase gene fusions. Methods Cell Biol. **34:**61–75.
25. **Manoil, C., and J. Beckwith.** 1985. Tn*phoA*: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA **82:**8129–8133.
26. **Mohan, S., J. Aghion, N. Guillen, and D. Dubnau.** 1989. Molecular cloning and characterization of *comC*, a late competence gene of *Bacillus subtilis*. J. Bacteriol. **171:**6043–6051.
27. **Payne, M. S., and E. N. Jackson.** 1991. Use of alkaline phosphatase fusions to study protein secretion in *Bacillus subtilis*. J. Bacteriol. **173:**2278–2282.
28. **Piazza, F., and D. Dubnau.** Unpublished data.
29. **Pugsley, A. P.** 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. **57:**50–108.
30. **Russel, M., and P. Model.** 1982. Filamentous phage pre-coat is an integral membrane protein: analysis by a new method of membrane preparation. Cell **28:**177–184.
31. **Sancar, G. B., A. Sancar, and W. Dean Rupp.** 1984. Sequences of the *E. coli uvrC* gene and protein. Nucleic Acids Res. **12:**4593–4608.
32. **Scholzen, T., and E. Arndt.** 1991. Organization and nucleotide sequence of ten ribosomal protein genes from the region equivalent to the spectinomycin operon in the archaebacterium *Halobacterium marismortui*. Mol. Gen. Genet. **228:**70–80.
33. **Shaw, J. H., and D. B. Clewell.** 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn*917* in *Streptococcus faecalis*. J. Bacteriol. **164:**782–796.
34. **Traxler, B., D. Boyd, and J. Beckwith.** 1993. The topological analysis of integral cytoplasmic membrane proteins. J. Membr. Biol. **132:**1–11.
35. **van Sinderen, D., R. Kiewiet, and G. Venema.** 1995. Differential expression of two closely related deoxyribonucleases, *nucA* and *nucB* in *Bacillus subtilis*. Mol. Microbiol. **15:**213–223.
36. **van Sinderen, D., A. Luttinger, L. Kong, D. Dubnau, G. Venema, and L. Hamoen.** 1994. *comK* encodes the competence transcription factor (CTF), the key regulatory protein for competence development in *Bacillus subtilis*. Mol. Microbiol. **15:**455–462.
37. **von Heijne, G.** 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J. **5:**3021–3027.
38. **von Heijne, G.** 1989. Control of topology and mode of assembly of a polytopic protein by positively charged residues. Nature (London) **341:**456–458.
39. **Zhang, S., and M. Nonoyama.** 1994. The cellular proteins that bind specifically to the Epstein-Barr virus origin of plasmid DNA replication belong to a gene family. Proc. Natl. Acad. Sci. USA **91:**2843–2847.