JOURNAL OF BACTERIOLOGY, June 1995, p. 3235–3240 Vol. 177, No. 11
0021-9193/95/$04.00+0

DNA Sequence and Characterization of *Haemophilus influenzae* *dprA*$^+$, a Gene Required for Chromosomal but Not Plasmid DNA Transformation

SUREKHA KARUDAPURAM, XUMEI ZHAO,† AND GERARD J. BARCAK*

Department of Biological Chemistry, University of Maryland School of Medicine, Baltimore, Maryland 21201

Received 17 November 1994/Accepted 29 March 1995

Natural genetic transformation in *Haemophilus influenzae* involves DNA binding, uptake, translocation, and recombination. In this study, we cloned and sequenced a 3.8-kbp *H. influenzae* DNA segment capable of complementing in *trans* the transformation defect of an *H. influenzae* strain carrying the *tfo-37* mutation. We used subcloning, deletion analysis, and in vivo protein labeling experiments to more precisely define the gene required for efficient DNA transformation on the cloned DNA. A novel gene, which we called *dprA*$^+$, was shown to encode a 41.6-kDa polypeptide that was required for efficient chromosomal but not plasmid DNA transformation. Analysis of the deduced amino acid sequence of DprA suggested that it may be an inner membrane protein, which is consistent with its apparent role in DNA processing during transformation. Four other open reading frames (ORFs) on the cloned DNA segment were identified. Two ORFs were homologous to the phosphofructokinase A (*pfkA*) and alpha-isopropyl malate synthase (*leuA*) genes of *Escherichia coli* and *Salmonella typhimurium*, respectively. Homologs for the two other ORFs could not be identified.

Natural genetic transformation in bacteria is a complex genetically programmed process involving DNA binding, uptake, translocation, and recombination. *Haemophilus influenzae* is a gram-negative bacterium that in growing cultures can be induced to become competent for transformation by a temporary shift to anaerobic conditions, by physiological change occurring during the late-log phase growth, or by a transfer of cells to a nutrient-poor chemically defined medium such as MIV (10).

Only recently has the molecular cloning of transformation genes allowed identification of several structural and regulatory components of the transformation apparatus (4–6, 16, 17, 26, 39). To identify the genes involved in transformation, Tomb and coworkers (35) performed minitransposon mutagenesis using *H. influenzae* chromosomal DNA and isolated 24 mutant strains that were defective in transformation. These mutant strains were characterized for DNA binding and uptake and for their transformability with *H. influenzae* chromosomal DNA. Among the 24 strains analyzed, only 2 bound and took up radiolabeled DNA in a manner similar to that of the wild type but transformed at frequencies less than 0.1% of that of the wild type. This suggested that the two strains were defective in events occurring subsequent to DNA uptake, i.e., DNA processing. This processing may involve DNA translocation and/or recombination.

In this study, we examine the nature of the defect in one of the two DNA uptake-proficient mutant strains, JG37. The mutated locus derived from strain JG37 was cloned and used to isolate the wild-type locus. The DNA sequence of a 3.8-kbp portion of the locus was determined and shown to encode several polypeptides. At least one of these, DprA, is essential for efficient chromosomal DNA transformation but is not required for plasmid transformation.

* Corresponding author. Mailing address: Department of Biological Chemistry, University of Maryland School of Medicine, 108 North Greene St., Baltimore, MD 21201. Phone: (410) 706-6304. Fax: (410) 706-8297. Electronic mail address: gbarcak@umabnet.ab.umd.edu.

† Present address: 10 Heritage Dr., Salem, MA 01970.

MATERIALS AND METHODS

Bacterial strains, plasmids, and culture conditions. Table 1 describes the bacterial strains and plasmids used in this study. *H. influenzae* strains were grown in brain heart infusion broth supplemented (sBHI) with hemin and NAD (Sigma Chemical Co., St. Louis, Mo.) as described previously (2). Bacto Agar (Difco Laboratories, Detroit, Mich.) was added to sBHI broth medium at a final concentration of 1.0 or 1.5% (wt/vol) for pour plates. *H. influenzae* cells were made competent for DNA transformation by the MIV procedure (13) and were transformed with chromosomal or plasmid DNAs as described elsewhere (2). *Escherichia coli* cells were made competent for plasmid transformation by the calcium chloride method (18). Antibiotics (Sigma) were used at the following final concentrations (in micrograms per milliliter) for selection in *H. influenzae*: kanamycin, 7 to 20; novobiocin, 2.5; chloramphenicol, 2; spectinomycin, 15; and nalidixic acid, 3. All bacterial strains were grown at 37°C.

Tomb and coworkers (35) reported the isolation and characterization of the *H. influenzae* transformation-deficient (Tfo$^-$) mutant strain JG37, which was produced by mini-Tn*10kan* insertional mutagenesis. We prepared strain GBH37F as an isogenic derivative of this strain by transforming strain KW20 to Kan^r with a limiting amount of chromosomal DNA from strain JG37. GBH37F was found to be phenotypically identical to strain JG37 for DNA uptake and lack of transformability (see below). To clone the transposon-tagged mutation from strain GBH37F, chromosomal DNA from the strain was digested with *Pst*I and ligated to similarly cleaved cloning vector pGJB103 (35), under standard reaction conditions. The ligation mixture was used to transform strain DH5α to Tet^r and Kan^r. Plasmid DNAs from transformant colonies were then screened by restriction enzyme digestion for the presence of a DNA insert containing the 2-kbp mini-Tn *10kan* element. The mutated locus was cloned as an 8-kbp DNA segment. The corresponding wild-type locus (called *dpr*$^+$ for DNA processing) was cloned by recombinational exchange with the strain KW20 chromosome as described elsewhere (35). Plasmid pHX37-1A was isolated as one such plasmid and contained the wild-type version of the *dpr*$^+$ locus as a 6-kbp *Pst*I fragment. Plasmid pXZ184 was constructed by subcloning the 3.8-kbp *Pst*I-*Xho*I fragment from pHX37-1A (Fig. 1) into the cloning vector pGB18. Plasmid pGB18 is a derivative of pSU2718 (19) that lacks the *Eco*RI site in the *cat* gene on account of site-directed mutagenesis (1). Plasmid pXZ184N was constructed by inserting an 8-bp *Not*I linker (GCGGCCGC) into the blunt-end unique *Spe*I site of *dprA*$^+$ (see below) in pXZ184 to produce an in-frame 12-bp insertion. The insertion was confirmed by DNA sequencing and places two arginine and two proline residues at amino acid positions 143 to 146 in DprA (Fig. 1B). It was thought that this mutation would perturb DprA function without affecting transcription of the gene. Plasmid pGJB110 is a spontaneously occurring deletion derivative of pGJB103 that is 1.5 kbp smaller but retains all of the parental phenotypic properties (1).

TABLE 1. Bacterial strains and plasmids used in this study

Bacterial strains	Relevant genotype or phenotype[a]	Source (reference)
H. influenzae		
KW20	Wild-type	H. O. Smith (37)
JG37	KW20::mini-Tn*10kan* Kan^r and Tfo$^-$	This laboratory (35)
GBH37F	KW20 transformed with JG37 chromosomal DNA for Kan^r and Tfo$^-$	This study
MAP7	Nov^r Nal^r	J. K. Setlow (35)
E. coli		
BL21DE3	*hsdS*	F. W. Studier (31)
DH5α	*recA1*	GIBCO-BRL
Plasmids		
pGJB103	*bla*$^+$ *tet*$^+$	This laboratory (35)
pGJB110	*bla*$^+$ *tet*$^+$	This laboratory (1)
pGB18	*cat*$^+$	This laboratory (1)
pHX37-1A	pGJB103 (*Pst*I::*Hin* 6-kbp *dpr*$^+$)	This study
pHXB2	pHX37-1A deleted for 2-kbp *Bgl*II fragment	This study
pXZ184	pGB18 (*Xho*I-*Pst*I::3.8-kbp *dpr*$^+$)	This study
pXZ184N	pXZ184 (*Spe*I::*Not*I linker)	This study
pT7-5	*bla*$^+$	J. B. Kaper (34)
pT373.1	pT7-5 (*Sma*I::*Hin* 1.3-kbp *dprA*$^+$)	This study
pT373.6	pT7-5 (*Sma*I::*Hin* 1.3-kbp *dprA*$^+$)	This study
pTS3.1	pT7-5 (*Sma*I::*Hin* 3.1-kbp *dpr*$^+$)	This study

[a] *Hin*, *H. influenzae*.

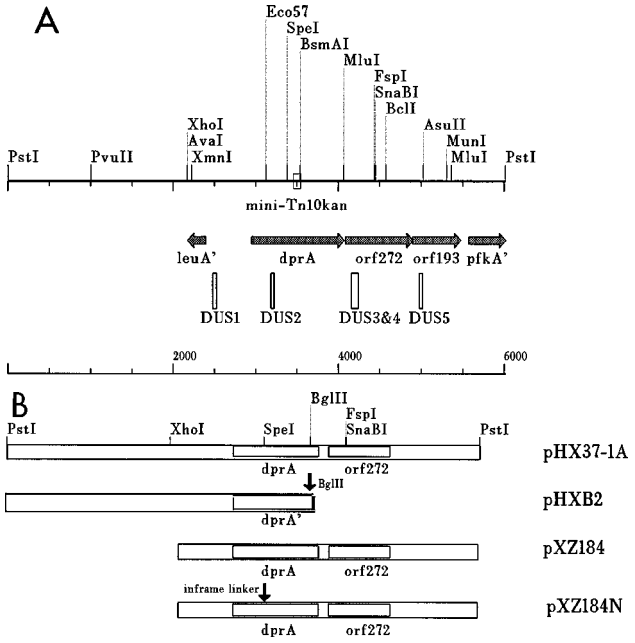

FIG. 1. (A) Structural organization of the cloned *H. influenzae dpr*$^+$ locus. The restriction endonuclease cleavage map is presented, and only selected restriction sites are shown. The arrows indicate the positions of the genes and the directions of transcription. The superscript prime in *pfkA′* and *leuA′* denotes that only the 5′ end of the gene is present. The boxed segment in the coding region of the *dprA*$^+$ gene indicates the approximate position of the mini-Tn*10kan* insertion. The boxes labeled DUS represent the positions of the five DNA uptake sequences. (B) Structural organization of the cloned DNA segments used in the plasmid complementation assay.

Recombinant DNA techniques. *H. influenzae* and *E. coli* chromosomal and plasmid DNAs were prepared as described by Barcak et al. (2) and Silhavy et al. (29), respectively. Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, and shrimp alkaline phosphatase were used according to the recommendations of the respective manufacturers. DNA fragments were isolated and purified from agarose gels and subcloned into plasmids by standard methods (18).

DNA sequence determination and data analysis. The DNA sequence of the 3.8-kbp *dpr*$^+$ locus on both DNA strands was determined by the method described by Sanger et al. (28) by using either a single- or a double-stranded template and primer walking. The reaction products were analyzed on a 7% polyacrylamide denaturing gel. Computer analysis of the DNA and protein sequences was performed with the Genepro software package (Riverside Scientific Enterprises, Bainbridge Island, Wash.).

Identification of plasmid-encoded polypeptides. The bacteriophage T7 RNA polymerase-directed expression system described by Studier and Moffatt (31) was used with *E. coli* BL21DE3 and [^{35}S]methionine as a tracer. After induction of T7 RNA polymerase with isopropyl-β-D-thiogalactopyranoside (IPTG) and radiolabeling, cells were lysed in loading buffer and about 2.5 μg of the whole-cell protein was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis with ^{14}C-labeled protein molecular mass standards. The gel was fixed, stained in 1 M sodium salicylate, and dried under vacuum, and radioactive polypeptides were visualized by autoradiography.

Nucleotide sequence accession number. The GenBank accession number for the DNA sequence reported in Fig. 2 is U18657.

RESULTS

Preparation and phenotype of the transformation mutant, strain GBH37F. Tomb and coworkers found that the *H. influenzae* strain JG37 bound and took up DNA at wild-type levels but transformed at $<10^{-4}$ of the wild-type frequency (35). To ensure that the transformation-deficient phenotype of JG37 was a direct result of the mini-Tn*10kan* insertion and not an unlinked mutation, the backcross strain GBH37F was made by transforming strain KW20 with strain JG37 chromosomal DNA. The phenotype of strain GBH37F was assayed by transformation of MIV-competent cultures with ^{32}P-labeled MAP7 chromosomal DNA. As may be seen in Table 2, strain GBH37F bound and took up approximately the same amount of DNA as strain KW20. However, GBH37F transformed at a frequency that was 10^4-fold lower than the wild-type frequency, as shown in Table 3. These results suggested that GBH37F was defective in the DNA-processing steps of transformation which involve DNA translocation and/or recombination.

The ability of strain GBH37F to be transformed by plasmid DNA was also measured. Plasmid pGJB110 was isolated from strain KW20 and purified by ultracentrifugation through CsCl. Approximately 1 μg of this DNA was added to 1 ml of MIV-competent cells with selection for tetracycline resistance. The results from two independent experiments, each of which was performed in duplicate, showed that both strain KW20 and strain GBH37F gave rise to tetracycline-resistant transformants at nearly identical frequencies (1.1×10^{-5} versus 1.0×10^{-5}, respectively). The data suggested that the defect in strain GBH37F was specific for linear chromosomal DNA. To determine whether an alternate method of plasmid transformation might reveal a different result, the above experiment was repeated by the glycerol stimulation method described by Stuy and Walter (33) with essentially the same result (data not shown).

Plasmid complementation of the GBH37F transformation defect. The *dpr*$^+$ locus was cloned as a 6-kbp DNA fragment as described in Materials and Methods. In order to determine the location of the *dpr*$^+$-complementing activity on this DNA segment, a series of plasmids was used in a complementation assay. Figure 1A shows the restriction map of the *dpr*$^+$ locus and the position of the mini-Tn*10kan* insertion as determined by restriction mapping. Plasmids pHX37-1A and pXZ184 (Fig. 1B) restored the transformation frequency of strain GBH37F to wild-type levels (Table 3). Plasmid pHXB2, which contained the 4-kbp leftward end of the locus (as depicted in Fig. 1B), did

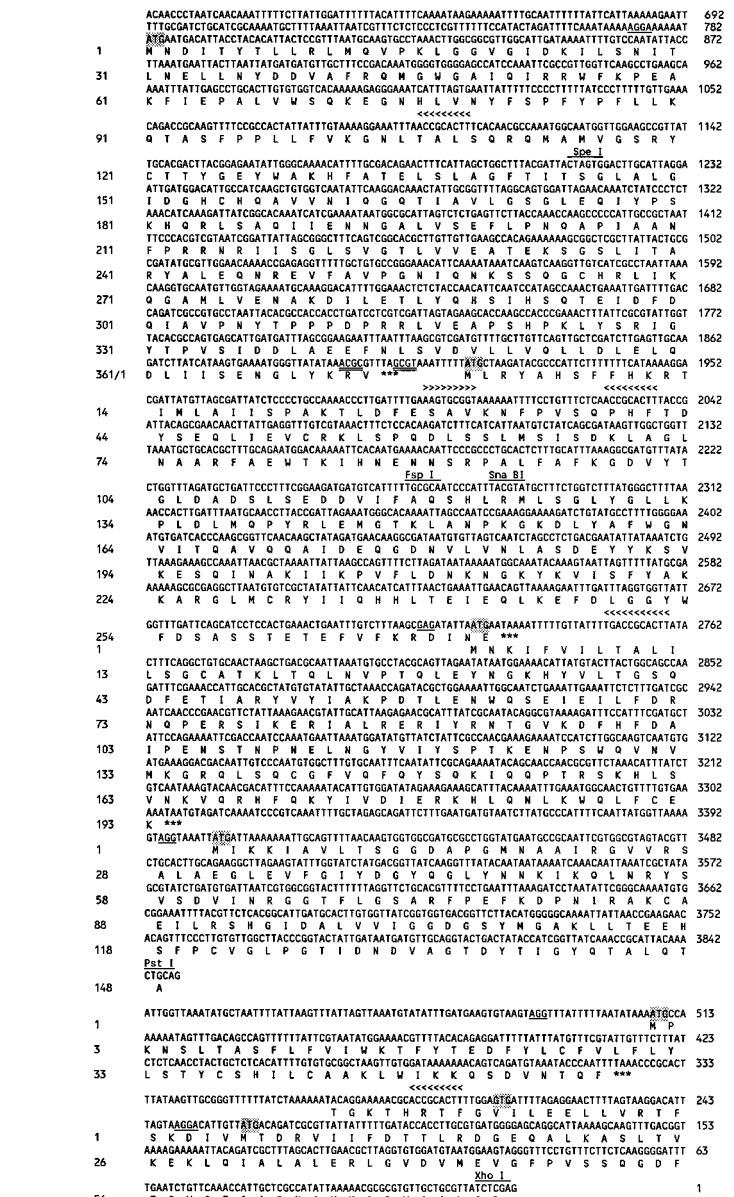

FIG. 2. DNA and deduced amino acid sequences of the *H. influenzae dpr*$^+$ locus. For clarity, the amino acid sequences encoded by the *dprA*$^+$-containing DNA strand are shown first. Numbers at the left refer to amino acid positions. Numbers at the right correspond to the DNA sequence as deposited in GenBank. The start codon of each of the ORFs is depicted by a shaded ATG. The positions of landmark restriction endonuclease cleavage sites are indicated. The five DNA uptake sequences (AAGTGCGGTCA) are indicated by the chevrons above each sequence. The polarity of a DNA uptake sequence is given by the direction of the chevrons. The position of the putative ribosome-binding site is underlined for each ORF as indicated in the text. The inverted repeat sequence that may function as a rho-independent transcription terminator at the end of *dprA*$^+$ is doubly underscored. ***, a translational stop codon.

not complement the mutation. The results from these experiments suggested that the transposon-induced mutation was complemented by an activity present on the rightward 3.8-kbp *Xho*I-*Pst*I segment (Fig. 1B). The DNA sequence of this segment was determined, and the map position of the Tn*10kan* element was found to interrupt a potential open reading frame (ORF) of 373 amino acids, which we named *dprA*$^+$. Furthermore, analysis of the DNA segment in the noncomplementing

TABLE 2. DNA binding and internalization by *H. influenzae* cells after incubation in MIV medium

Strain	Cell-associated radioactive DNA (cpm)[a]	
	Without DNase	With DNase
KW20	16,190	15,464
KW20 (zero time)		394
GBH37F	16,169	16,816
GBH37F (zero time)		158

[a] To measure DNA binding (total counts associated with cells) and DNA uptake (DNase I-resistant counts), a ^{32}P-labeled DNA sample was prepared by filling in the ends of *Bam*HI-digested MAP7 chromosomal DNA with the Klenow fragment of DNA polymerase I. The resulting DNA was predominantly 20 to 40 kbp in size as estimated by agarose gel electrophoresis and had a specific activity of approximately 9×10^6 cpm/μg. About 0.5 μg of this DNA was added to cells passaged through MIV competence-inducing medium. After 30 min at 37°C, a 0.2-ml cell sample was placed into each of two tubes at 0°C, one of which contained 15 μg (300 Kunitz units) of DNase I (Grade II; Boehringer Mannheim). DNase I was added prior to the DNA in the zero time samples. After 5 min, the samples were brought to 0.5 M NaCl and both samples were pelleted and washed once with MIV-NaCl medium. The resuspended pellets were applied to glass fiber filters and dried, and the radioactivity was determined. The values are the averages of duplicate samples. A total of 27 cpm background was subtracted from each sample.

plasmid pHXB2 revealed that the *Bgl*II junction in the plasmid, which was formed by deletion of locus sequences, actually removed DNA sequences encoding the predicted C-terminal 13 amino acids of DprA.

To ensure that the GBH37F mutant phenotype was not due to a polar effect of the mini-Tn*10kan* insertion on ORF272 and perhaps other downstream genes, variants of strain GBH37F bearing plasmids carrying only *dprA*$^+$ or *dprA*$^+$ and a disrupted gene for ORF272 were prepared. However, these strains were found to have aberrantly long doubling times (resulting from a loss of cell viability [data not presented]), and thus these strains could not be studied further.

As an alternate strategy, strain GBH37F/pXZ184N, which contained a plasmid-encoded in-frame 4-amino-acid insertion mutation of *dprA* and otherwise wild-type downstream genes, was assayed for transformability and was found to transform at mutant frequencies (Table 3). These results indicated that *dprA*$^+$ was essential to restore transformability to strain GBH37F.

To characterize the role of ORF272 in transformation, we tried to construct a strain carrying an ORF272 gene disruption. This was attempted by transforming strain KW20 with a DNA fragment that contained a spectinomycin resistance (Sp^r) cassette insertion in the ORF272 gene at its unique *Sna*BI site.

TABLE 3. Plasmid complementation of the transformation defect in *H. influenzae* GBH37F

Strain/plasmid	Transformation frequency[a]
KW20	1.5×10^{-3}
GBH37F	1.5×10^{-7}
GBH37F/pHX37-1A	5.0×10^{-3}
GBH37F/pHXB2	4.5×10^{-6}
GBH37F/pXZ184	1.5×10^{-3}
GBH37F/pXZ184N	2.0×10^{-7}

[a] Strains were transformed by the MIV method by using strain MAP7 chromosomal DNA with selection for novobiocin or naladixic acid resistance. To obtain the transformation frequency, the number of drug-resistant CFU was divided by the total number of viable CFU. The values are the averages of two experiments performed in duplicate. Transformation frequencies of the different strains varied by less than 10-fold between experiments.

Unfortunately, even upon repeated attempts, no Sp^r colonies were obtained. Because the Sp^r cassette has been inserted successfully into other *Haemophilus* genes (e.g., *crp*, and *tfoX* [39]), a plausible explanation for our results is that the gene for ORF272 is essential for cell viability. Thus, we still lack evidence that the potential genes downstream of $dprA^+$ participate in the transformation process.

Structural organization of the dpr^+ locus. DNA sequencing and computer-aided analysis of the 3.8-kbp dpr^+ locus revealed the genetic organization shown in Fig. 1A. DNA sequencing confirmed restriction sites that previously had been mapped to the region. The locus encoded five AUG-initiated ORFs. On one DNA strand, ORF373 ($dprA^+$), ORF272, ORF193, and ORF148 (*pfkA′* [see below]) were located at nucleotide (nt) positions 783 to 1901, 1916 to 2731, 2727 to 3305, and 3404 to 3847, respectively (Fig. 2). On the other strand, ORF74 (*leuA′* [see below]) was located from nt 1 to 226. Repetition of the search for GUG-initiated ORFs resulted in 4-, 2-, and 1-amino-acid extensions, respectively, to ORF373, ORF272, and ORF148. ORF193 was extended to 214 amino acids, while ORF74 was extended to 119 amino acids. A potential ribosome-binding site that was identical to the *E. coli* sequence (AGGA) was located at positions 773 to 776 for $dprA^+$ and positions 234 to 237 for ORF74. Only trinucleotide matches to potential ribosome-binding sites could be found for two of the other ORFs. AGG was found at positions 3395 to 3397 for ORF148, and GAG was found at positions 2718 to 2720 for ORF193; no such match could be found for ORF272 (Fig. 2). Our inability to find a consensus ribosome-binding site upstream of ORF272 may indicate translational coupling between this gene and $dprA^+$. The DNA sequence also revealed five DNA uptake sites (AAGTGCGGTCA [30]) at positions 280 to 290, 1096 to 1106, 1999 to 2009, 2037 to 2047, and 2749 to 2759 in the 3.8-kbp fragment (Fig. 2).

We previously showed that the *H. influenzae* $rec\text{-}1^+$ (38) and $tfoX^+$ (39) genes contained σ^{70}-like promoter sequences upstream of their respective transcription start sites. Thus, DNA sequences upstream of the coding regions of the various ORFs of the dpr^+ locus were analyzed for homology to the bacterial σ^{70} consensus promoter by using the O'Neill program (22). Two potential promoters for $dprA^+$ were identified. First, a 4-of-6-base match to the *E. coli* −35 consensus sequence (TTGACA) was found from bp 497 to 502 (CTGTCA), which was followed by an 18-base spacing and a 5-of-6-base match to the −10 consensus sequence (TATAAT) from bp 521 to 526 (TATATT) (Fig. 2). The start of the mRNA could be 8 bases downstream at the purine at bp 534, resulting in a leader sequence of 248 nt; second, a 4-of-6-base match to the −35 consensus sequence from bp 563 to 568 (TTAACT) which was followed by a 17-bp spacing and a 4-of-6-bp match to the −10 consensus sequence from bp 586 to 591 (TAGCAT) were also found. The start of the mRNA could be 5 bases downstream at the purine at position 596, which would give a leader sequence of 186 nt. Analysis of the upstream regions of the other ORFs did not reveal any 4-of-6-bp matches to the *E. coli* consensus sequence for the −35 region and −10 region that were separated by a 17-base spacing.

Recent findings (5) have suggested that the Rec-2 protein might be involved in transformation as part of a membrane-bound pore complex at the base of the transformasome, participating in DNA translocation by interacting directly with the transforming DNA. Since a mutation in $dprA^+$ caused GBH37F to be defective in DNA processing, a possible role of DprA could be as a membrane-associated (or periplasmic) protein that interacts with Rec-2 and is part of the DNA translocation apparatus. To provide support for this hypothesis, the deduced amino acid sequence of $dprA^+$ was analyzed by the method described by Kyte and Doolittle (15) to obtain a hydropathy profile. As may be seen in Fig. 3, six regions that are 10 or more amino acids long have a hydrophobic character. Hydrophobic amino acids 3 to 29 may function as a signal sequence to allow DprA transport past the cytoplasmic membrane. The region with the greatest average hydrophobicity spans amino acids 341 to 352. We speculate that if DprA is a membrane-associated protein, this segment could anchor DprA to the membrane. Also striking are the stretches of 19 and 17 hydrophobic amino acids in the center of the protein from residues 130 to 148 and 154 to 170, respectively, (Fig. 3), which could constitute a transmembrane domain, although the average hydrophobicity index of the two regions is lower than those found for other transmembrane domains (15). Other hydrophobic segments of DprA are found at residues 91 to 101 and 215 to 226. From our analysis, we conclude that at least two of the hydrophobic segments that we have identified (regions 3 to 29 and 341 to 352) are consistent with DprA functioning as a membrane or periplasmic protein in a complex of proteins that constitute the DNA translocation apparatus of the competent cell.

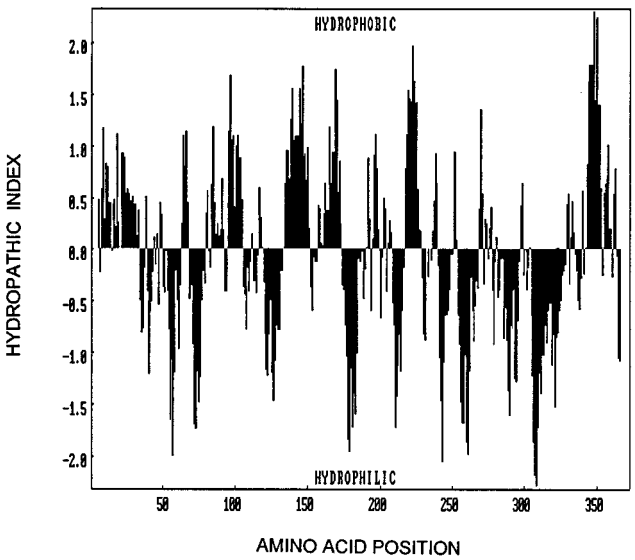

FIG. 3. Hydropathy analysis of *H. influenzae* DprA. Hydropathic indices were calculated by the method described by Kyte and Doolittle (15) with Genepro software. A 9-amino-acid window was used in the analysis.

Identification of polypeptides encoded by the dpr^+ locus. We wanted to confirm our observations concerning the size and number of polypeptides potentially encoded by the dpr^+ locus. We previously showed that the size of the *H. influenzae* $rec\text{-}1^+$ protein synthesized in *E. coli* under the regulation of a bacteriophage T7 promoter was the same as that of the native protein in *H. influenzae* (3, 38). Accordingly, polypeptides predicted by the dpr^+ locus were assayed in *E. coli* by using a T7 RNA polymerase-directed system (31). Plasmid pT373.6 carrying only $dprA^+$ produced a 39.5-kDa polypeptide that was IPTG inducible (Fig. 4A, lanes 6 and 7). Plasmid pT373.1, encoding the $dprA^+$ gene in an opposite orientation with respect to the T7 promoter, did not produce significant levels of polypeptide upon induction with IPTG (Fig. 4A, lanes 2 and 3). Plasmid pTS3.1, which contained a 3.1-kbp fragment from nt 716 to 3848 (Fig. 2) that encoded DprA and all of the downstream genes, produced four IPTG-inducible polypeptides with gel-estimated sizes of 40, 29, 23, and 16 kDa (Fig. 4B,

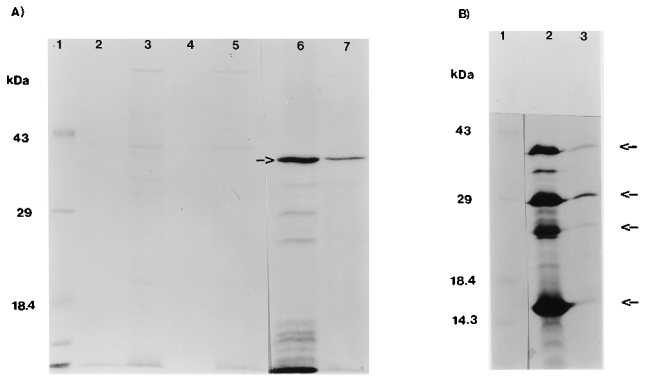

FIG. 4. Polypeptides encoded by *E. coli* BL21DE3 containing various expression plasmids. Odd-numbered lanes contain polypeptides produced without IPTG. Even-numbered lanes contain polypeptides produced with the addition of IPTG. (A) Lane: 1, M_rs (in thousands), as determined by reference to ^{14}C-labeled protein standards (given at the far left); 2 and 3, pT373.1; 4 and 5, pT7-5; 6 and 7, pT373.6. The position of DprA is indicated. (B) Lanes: 1, same as lane 1 for panel A; 2 and 3, pTS3.1. The positions of DprA, ORF272, ORF193, and PfkA′ are indicated from top to bottom by arrows at the right.

lanes 2 and 3). These sizes are consistent with the predicted sizes of the polypeptides encoded by ORF373 (41.6 kDa), ORF272 (30 kDa), ORF193 (22.7 kDa), and an ORF148 protein that contained at its C terminus 5 additional amino acids encoded by vector sequences (16.2 kDa), respectively. Vector pT7-5 did not produce any significant level of polypeptide (Fig. 4A, lanes 4 and 5). We conclude from our experiments that the sizes and numbers of polypeptides predicted from DNA sequence analysis of the dpr^+ locus are consistent with the in vivo results from *E. coli* and that these results likely extend to *H. influenzae*.

Sequence homologies of the polypeptides of the dpr^+ locus. Each of the potential ORFs was analyzed for homology to any known genes or proteins in the GenBank (Release 79) and PIR (Release 38) databases. DprA was 43% identical (52% similar, allowing for conservative amino acid substitutions) to the *smf* ORF of *E. coli* (20). This ORF codes for a 374-amino-acid hypothetical protein and is located upstream of the *fmt* gene at 72.4 min on the *E. coli* chromosome. No function has yet been identified for *smf*. An alignment of the two deduced protein sequences indicated that the central region is most highly conserved between the two proteins and is perhaps most important for the function of the proteins.

Analysis of ORF272 revealed a 61% identity to a 258-amino-acid hypothetical protein encoded downstream of the *thrC* gene of the *E. coli* threonine operon (23). Again, no function has been assigned to this hypothetical *E. coli* protein.

ORF148 is 76% identical to the first 148 amino acids of the 340-amino-acid PfkA protein of *E. coli* (12). PfkA is the main phosphofructokinase in *E. coli*.

The 74-amino-acid ORF was found to be 84% identical to the amino terminus of the enzyme alpha-isopropylmalate synthase, which is encoded by $leuA^+$ of *Salmonella typhimurium* (27). Examination of the DNA sequence upstream of the *H. influenzae leuA* gene revealed a potential leucine-containing 58-residue leader peptide and potential attenuation signals, as have been found in other amino acid biosynthetic operons (9).

We are tentatively designating ORF148 as the *H. influenzae* $pfkA^+$ gene and ORF74 as the *H. influenzae* $leuA^+$ gene.

DISCUSSION

The use of modern microbial genetic methods for transposon-tagging genes has resulted in a wealth of new information in both gram-positive and -negative transformation systems (reviewed in reference 8 for *Bacillus subtilis*; see also references 14, 32, and 35). *H. influenzae* GBH37F contains a mini-Tn*10kan* insertion that confers a transformation mutant phenotype, although cells are able to bind and take up DNA normally (Table 2). In this study, we cloned the mutant locus from strain GBH37F and the corresponding wild-type cognate from strain KW20. DNA sequencing of a 3.8-kbp segment identified several potential ORFs (Fig. 1). The polypeptides encoded by the ORFs were confirmed by protein expression experiments with *E. coli* (Fig. 4). The original transposon insertion in strain GBH37F was found to be in the largest ORF, *dprA*, encoding a 41.6-kDa protein which was essential for chromosomal transformation based on complementation studies (Table 3).

DprA could be a structural component of the cellular apparatus that participates in DNA translocation or recombination. The hydropathy plot of DprA is consistent with it being a membrane-associated or periplasmic protein (Fig. 3).

Almost a decade ago, Pifer (25) studied the establishment of plasmids in naturally competent *H. influenzae*. On the basis of experiments with topologically different forms of plasmid DNA, she concluded that plasmids can circumvent the normal mode of DNA translocation involving degradation of the DNA, and can arrive as intact double-stranded molecules in the cytoplasm. A similar proposal was made by Stuy and Walter (32). Pifer referred to this mode of transformation as illegitimate transformation. We would predict from her hypothesis that a mutation in a transformation gene whose product is involved in DNA processing after uptake will not affect plasmid transformation frequencies. When Notani and coworkers (21) measured plasmid transformation in the *rec-2* mutant strain (Rec-2 is known to be involved in DNA translocation), they found that the frequency of plasmid transformation did not differ from that of the wild type. When we tested the transformability of strain GBH37F by plasmid DNA, the frequency of plasmid transformation was similar to that of the wild type, indicating that plasmid transformation was unaffected by the mutation in $dprA^+$. Thus, our plasmid transformation data further support our contention that DprA is involved in donor DNA processing during transformation.

While DprA was not homologous to any protein in the various databases whose function is known, it was homologous to the Smf ORF of *E. coli* (20). Since *E. coli* does not transform naturally (11), we wondered what similar process(es) Smf and DprA might perform. It occurred to us that aspects of DNA transformation and conjugation in the two organisms were similar (7). For instance, during transformation, as double-stranded DNA translocates out of the protected compartment (transformasome), the leading 5′→3′ strand is degraded and the leading 3′→5′ single-stranded DNA is partially degraded and translocated until it finds a homologous partner for nucleation and integration. In *E. coli* conjugation, single-stranded DNA is translocated from donor to recipient cell before nucleation and integration. If the biological roles of Smf and DprA are similar, we think it reasonable in future experiments to test for the ability of smf^+ to suppress the transformation defect in a *dprA* strain, even though the two proteins share just 43% identity. Interestingly, the SOS mutagenesis protein of *E. coli*, UmuD, is only 41% identical to a plasmid-encoded analog, MucA (24). However, $mucA^+$ can suppress the nonmutability of *umuD* strains (24, 36).

Although it is unlikely, if DprA does not interact with the DNA as part of the translocation apparatus, our data are also consistent with DprA being a regulatory factor that modulates the expression of a linked or unlinked gene(s) whose product interacts with the incoming DNA.

At 15 bp downstream of *dprA*$^+$, we found another ORF of 272 amino acids that was 61% identical to a hypothetical protein from *E. coli* whose function is unknown. Since we were unable to find a consensus promoter upstream of the ORF272 gene, it is possible that this gene is cotranscribed with *dprA*$^+$. Although *dprA*$^+$ and ORF272 are separated by just 15 bp in *H. influenzae*, their homologs in *E. coli* are 1.5 Mbases apart. We suspect that linkage may be an indication of their functional dependence in *H. influenzae*. Our attempts to make antibiotic insertion mutants of the ORF272 gene have thus far been unsuccessful, which suggests that the gene may be essential.

ACKNOWLEDGMENTS

We thank Hongxia Xu for technical assistance with cloning and complementation analyses and Nicholas Ambulos for assistance with DNA sequencing. We thank Eric Naumann and Susan Heimer for critical reading of the manuscript.

This work was supported by NIH research grant AI30083 to G.J.B.

REFERENCES

1. **Barcak, G. J.** 1994. Unpublished results.
2. **Barcak, G. J., M. S. Chandler, R. J. Redfield, and J. F. Tomb.** 1991. Genetic systems in *Haemophilus influenzae*. Methods Enzymol. **204:**321–342.
3. **Barcak, G. J., J. F. Tomb, C. S. Laufer, and H. O. Smith.** 1989. Two *Haemophilus influenzae* Rd genes that complement the *recA*-like mutation *rec-1*. J. Bacteriol. **171:**2451–2457.
4. **Chandler, M. S.** 1992. The gene encoding cyclic AMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. Proc. Natl. Acad. Sci. USA **89:**1626–1630.
5. **Clifton, S. W., D. McCarthy, and B. A. Roe.** 1994. Sequence of the *rec-2* locus of *Haemophilus influenzae*: homologies to *comE*-ORF3 of *Bacillus subtilis* and *msbA* of *Escherichia coli*. Gene **146:**95–100.
6. **Dorocicz, I. R., P. M. Williams, and R. J. Redfield.** 1993. The *Haemophilus influenzae* adenylate cyclase gene: cloning, sequence, and essential role in competence. J. Bacteriol. **175:**7142–7149.
7. **Dreiseikelmann, B.** 1994. Translocation of DNA across bacterial membranes. Microbiol. Rev. **58:**293–316.
8. **Dubnau, D.** 1993. Genetic exchange and homologous recombination, p. 555–584. *In* A. L. Sonenshein, R. Losick, and J. A. Hoch (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
9. **Gemmill, R. M., S. R. Wessle, E. B. Keller, and J. M. Calvo.** 1979. Leu operon of *Salmonella typhimurium* is controlled by an attenuation mechanism. Proc. Natl. Acad. Sci. USA **76:**4941–4945.
10. **Goodgal, S. H., and R. M. Herriott.** 1961. Studies on transformations of *Haemophilus influenzae*. I. Competence. J. Gen. Physiol. **44:**1201–1227.
11. **Hanahan, D.** 1987. Mechanisms of DNA transformation, p. 1177–1183. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
12. **Hellinga, H. W., and P. R. Evans.** 1985. Nucleotide sequence and high-level expression of the major *E. coli* phosphofructokinase. Eur. J. Biochem. **149:** 363–373.
13. **Herriott, R. M., E. M. Meyer, and M. Vogt.** 1970. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. J. Bacteriol. **101:**517–524.
14. **Hui, F. M., L. Zhou, and D. A. Morrison.** 1995. Competence for genetic transformation in *Streptococcus pneumoniae*: organization of a regulatory locus with homology to two lactococcocin A secretion genes. Gene **153:**25–31.
15. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
16. **Larson, T. G., and S. H. Goodgal.** 1991. Sequence and transcriptional regulation of *com*101A, a locus required for genetic transformation in *Haemophilus influenzae*. J. Bacteriol. **173:**4683–4691.
17. **Larson, T. G., and S. H. Goodgal.** 1992. Donor DNA processing is blocked by a mutation in the *com101A* locus of *Haemophilus influenzae*. J. Bacteriol. **174:**3392–3394.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Martinez, E., B. Bartolome, and F. de la Cruz.** 1988. pACYC184-derived cloning vectors containing the multiple cloning site and *lacZ alpha* reporter gene of pUC8/9 and pUC18/19 plasmids. Gene **68:**159–162.
20. **Meinnel, T., J.-M. Guillon, Y. Mechulam, and S. Blanquet.** 1993. The *Escherichia coli fmt* gene, encoding methionyl-tRNA formyltransferase, escapes metabolic control. J. Bacteriol. **175:**993–1000.
21. **Notani, N. K., J. K. Setlow, D. McCarthy, and N.-L. Clayton.** 1981. Transformation of *Haemophilus influenzae* by plasmid RSF0885. J. Bacteriol. **148:** 812–816.
22. **O'Neill, M. C.** 1991. Training back-propagation neural networks to define and detect DNA-binding sites. Nucleic Acids Res. **19:**313–318.
23. **Parsot, R., P. Cossart, I. Saint-Girons, and G. N. Cohen.** 1983. Nucleotide sequence of *thrC* and of the transcription termination region of the threonine operon of *E. coli* K-12. Nucleic Acids Res. **11:**7331–7345.
24. **Perry, K. L., S. J. Elledge, B. Mitchell, L. Marsh, and G. C. Walker.** 1985. The *umuDC* and *mucAB* operons whose products are required for UV and chemical mutagenesis; UmuD, MucA and LexA share homology. Proc. Natl. Acad. Sci. USA **82:**4331–4335.
25. **Pifer, M. L.** 1986. Plasmid establishment in competent *Haemophilus influenzae* occurs by illegitimate transformation. J. Bacteriol. **168:**683–687.
26. **Redfield, R. J.** 1991. *sxy-1*, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. J. Bacteriol. **173:**5612–5618.
27. **Ricca, E., and J. M. Calco.** 1990. The nucleotide sequence of *leuA* from *Salmonella typhimurium*. Nucleic Acids Res. **18:**1290.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
29. **Silhavy, T., M. Berman, and L. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Sisco, K. L., and H. O. Smith.** 1979. Sequence-specific DNA uptake in *Haemophilus* transformation. Proc. Natl. Acad. Sci. USA **76:**972–976.
31. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
32. **Stuy, J. H., and R. B. Walter.** 1986. Homology-facilitated plasmid transfer in *Haemophilus influenzae*. Mol. Gen. Genet. **203:**288–295.
33. **Stuy, J. H., and R. B. Walter.** 1986. Effect of glycerol on plasmid transfer in genetically competent *Haemophilus influenzae*. Mol. Gen. Genet. **203:**296–299.
34. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
35. **Tomb, J. F., G. J. Barcak, M. S. Chandler, R. J. Redfield, and H. O. Smith.** 1989. Transposon mutagenesis, characterization, and cloning of transformation genes of *Haemophilus influenzae*. J. Bacteriol. **171:**3796–3802.
36. **Walker, G. C.** 1987. The SOS response of *Escherichia coli*, p. 1352–1353. *In* F. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. Edwin Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. American Society for Microbiology, Washington, D.C.
37. **Wilcox, K. W., and H. O. Smith.** 1975. Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5′-triphosphate-dependent deoxyribonuclease activity. J. Bacteriol. **122:**443–453.
38. **Zulty, J. J., and G. J. Barcak.** 1993. Structural organization, nucleotide sequence, and regulation of the *Haemophilus influenzae rec-1*$^+$ gene. J. Bacteriol. **175:**7269–7281.
39. **Zulty, J. J., and G. J. Barcak.** 1995. Identification of a DNA transformation gene required for *com101A*$^+$ expression and supertransformer phenotype in *Haemophilus influenzae*. Proc. Natl. Acad. Sci. USA **92:**3616–3620.