INFECTION AND IMMUNITY, Feb. 1995, p. 696–699 Vol. 63, No. 2
0019-9567/95/$04.00+0

The Gene Encoding Protein D (*hpd*) Is Highly Conserved among *Haemophilus influenzae* Type b and Nontypeable Strains

XIN-MING SONG, ARNE FORSGREN, AND HÅKAN JANSON*

Department of Medical Microbiology, Lund University, University Hospital MAS, S-214 01 Malmö, Sweden

Received 11 August 1994/Returned for modification 17 October 1994/Accepted 22 November 1994

The molecular conservation of a surface-exposed lipoprotein, protein D, of *Haemophilus influenzae* was studied by cloning and sequencing of the gene encoding protein D from three encapsulated type b strains and three nontypeable strains of *H. influenzae*. These nucleotide sequences were analyzed with previously reported sequences from one type b strain and one nontypeable strain. The nucleotide sequences and the deduced amino acid sequences for protein D were highly conserved. The deduced amino acid sequence (364 amino acids) of protein D from six strains differed only in two amino acids near the C-terminal end. The remaining two strains, one type b and one nontypeable, differed from the consensus sequence in 7 amino acids each. Protein D is 64 and 36% identical and 77 and 56% similar to the glycerophosphodiester phosphodiesterases (GlpQ) of *Escherichia coli* and *Bacillus subtilis*.

Haemophilus influenzae is an important pathogen among infants and children. The serotype b strains of *H. influenzae* (Hib) are a major cause of meningitis and other invasive infections, including epiglottitis, bacteremia, and pneumonia. Nontypeable strains of *H. influenzae* (NTHi) mainly cause noninvasive mucosal infections such as otitis media and sinusitis (19). These strains are also an important cause of respiratory tract infection in children and adults (3, 19).

Effective vaccines against Hib strains have been used widely (1), but they do not protect children against infections caused by NTHi strains. For developing a vaccine that protects against Hib and NTHi infections, several surface-exposed *H. influenzae* proteins such as pili and outer membrane proteins have been studied intensely (4, 5, 7, 8, 12, 15, 16, 18). One of the possible candidates of a vaccinogen is protein D (22), a 42-kDa surface-exposed lipoprotein (9) with glycerophosphodiester phosphodiesterase (GlpQ) activity (17). In direct binding assays with human myeloma immunoglobulin D 4490, protein D was found on the surface of all 127 *H. influenzae* strains tested (2). By using three different monoclonal antibodies directed against protein D, it was shown that the immunoglobulin D-binding molecule is antigenically conserved (2). Restriction fragment length polymorphism analysis of the *hpd* region of a large number of *H. influenzae* strains and comparison of DNA sequences of one Hib and one NTHi strain indicated that only a limited variation existed within the genes encoding protein D (*hpd*) (11). The present study was performed to obtain more information concerning the molecular conservation within the *hpd* region among Hib and NTHi strains. We determined the DNA sequences of the protein D structural gene from three Hib strains and three NTHi strains. The nucleotide sequences were compared with the previously reported *hpd* sequences (10, 11). A comparison of the deduced amino acid sequences from *hpd* of *H. influenzae*, *Escherichia coli glpQ* (23), and *Bacillus subtilis glpQ* (21) was also made.

Bacterial strains. A total of six *H. influenzae* strains were used for *hpd* sequence determination in this study. The strains represent a variety of biotypes and were collected from different geographic origins (Table 1). The bacteria were grown as described previously (10).

Molecular cloning and DNA sequencing of the protein D gene. Chromosomal DNA was prepared from *H. influenzae* strains by a modification of the method described previously (11). On the basis of the restriction fragment length polymorphism patterns (11), the following DNA fragments containing *hpd* from *H. influenzae* strains were isolated by extraction with Geneclean (Bio 101, La Jolla, Calif.) after *Pst*I restriction endonuclease digestion and agarose electrophoresis: 1.9-kbp fragments from strains NCTC 8468, 3640, and 6-7626; a 2.1-kbp fragment from strain 3639; and 3.3-kbp fragments from strains Eagan and HK695. The DNA fragments were ligated to dephosphorylated *Pst*I-digested pUC18 (25). The ligation mixtures were used to transform competent *E. coli* JM83 cells. Enzymes were obtained from Boehringer Mannheim GmbH (Mannheim, Germany) and Appligene (Illkirch, France) and were used as described in the manufacturers' instructions or in standard methods. Recombinant colonies were screened in a colony immunoassay using the monoclonal antibody 19B4 directed against protein D as described previously (10). Plasmid DNA was prepared by using a Whizard DNA purification system (Promega, Madison, Wis.).

The nucleotide sequences of the *hpd* genes were determined by the dideoxy chain termination method using the Sequenase kit as described by the manufacturer (U.S. Biochemicals, Cleveland, Ohio). For sequencing both strands of the *hpd* gene, 10 oligonucleotide primers (Table 2) based on the previously reported *hpd* sequences (10, 11) were synthesized (Scandinavian Gene Synthesis AB, Köping, Sweden). In addition, a 506-bp *Sau*3A fragment from the plasmid carrying a DNA insert from NTHi strain 3639 and a 432-bp *Eco*RI-*Taq*I fragment from the plasmid carrying a Hib strain NCTC 8468 insert were each ligated to pUC18. The nucleotide sequences of these two fragments were determined for filling up the sequence gaps in the protein D genes from strains 3639 and NCTC 8468 because of failure to obtain sequence data by use of primers *hpd*-7 and *hpd*-1, respectively. The nucleotide sequences and the deduced amino acid sequences were analyzed with the Genetics Computer Group software package from the University of Wisconsin (6).

Conservation of protein D. The nucleotide sequences of the genes encoding protein D from the six *H. influenzae* strains in

* Corresponding author. Mailing address: Department of Medical Microbiology, Lund University, University Hospital MAS, S-214 01 Malmö, Sweden. Phone: +46/40-331321. Fax: +46/40-336234. Electronic mail address: mblhjn@maja.mas.malmo.se.

TABLE 1. Sources of *H. influenzae* strains

Strain	Serotype	Biotype	Clinical source	Geographic source
6-7626	Nontypeable	III	Nasopharyngeal swab	Sweden
3639	Nontypeable	I	Middle ear fluid	United States
3640	Nontypeable	V	Middle ear fluid	United States
Eagan	Type b	I	Cerebrospinal fluid	United States
HK695	Type b	I		Denmark
NCTC 8468	Type b	IV		United Kingdom

this study were compared with the two previously published sequences (10, 11). The eight strains reveal an open reading frame of 1,092 bp starting with a ATG codon and ending with a TAA stop codon (Fig. 1).

In a comparison of the three Hib strains, i.e., Eagan, HK695, and MinnA, only two nucleotides differed within *hpd*. These differences are silent substitutions in the deduced amino acid sequences which, subsequently, were 100% identical among the three strains. The nucleotide sequences of *hpd* and the deduced amino acid sequences of protein D from NTHi strains 772, 3639, and 3640 were also highly homologous, with 99% identity. The nucleotide sequences of *hpd* from Hib strain NCTC 8468 and NTHi strain 6-7626 were a little more divergent in comparison with the sequences of the other strains, with 26 and 22 nucleotide substitutions, respectively (Fig. 1). In the derived amino acid sequence of protein D from these two strains, seven substitutions were found in each strain. These amino acid differences were clustered near the N-terminal end of protein D in NCTC 8468, whereas they were relatively randomly distributed in strain 6-7626 (Fig. 1).

Protein D has been shown to be a lipoprotein preceded by an 18-amino-acid signal peptide at the N terminus which ends with the consensus sequence for lipoproteins, Leu-Ala-Gly-Cys (9, 24). The signal sequences were identical in all strains except NCTC 8468, which has two amino acid substitutions at positions 13 and 16 of the unprocessed translation product (Fig. 1).

An interesting difference was found when comparing the nucleotide sequences in the 5′-flanking regions of the protein D-encoding genes. The nucleotide sequences immediately upstream of the *hpd* start codon of Hib strains Eagan, HK695, and MinnA diverge from the remaining strains (Fig. 1). This result confirms previous results obtained by restriction fragment length polymorphism and nucleotide sequence analyses of the *hpd* region, which indicated that the majority of Hib strains differed from NTHi strains upstream of *hpd* (11). Hib strain NCTC 8468 has been shown previously to differ from the majority of Hib strains in a multilocus electrotyping assay (13, 20). We found that the nucleotide sequence upstream of *hpd* in Hib strain NCTC 8468 is homologous to that in the four NTHi strains of our study. This observation also correlates well with the previous restriction fragment length polymorphism results of the *hpd* region, where NCTC 8468 was shown to carry the protein D gene on a 1.9-kbp *Pst*I fragment as did the majority of NTHi strains. The other three Hib strains compared in this report carry the *hpd* gene on a 3.3-kbp *Pst*I fragment (11). The nucleotide sequences downstream of *hpd* were identical in all strains in this study, with the exception of a few nucleotide substitutions (Fig. 1). This finding further supports the speculation that the majority of Hib strains might contain a 1.4-kbp fragment upstream of *hpd* that is lacking in strains that carry the gene encoding protein D on a 1.9- to 2.1-kbp *Pst*I fragment. This difference is currently being analyzed in our laboratory.

Our results indicate that protein D is not only antigenically and functionally conserved (2) but also that the nucleotide and deduced amino acid sequences of the *hpd* gene are highly homologous among all *H. influenzae* isolates tested, regardless of their serotypes. Therefore, it is possible that the highly conserved and surface-exposed protein D might be a good candidate for inclusion in a future vaccine against all types of *H. influenzae* infections.

Similarity between protein D and GlpQ. A comparison of protein D from *H. influenzae* Eagan and GlpQ of *E. coli* (23) showed that the deduced proteins were 77% similar and 64% identical. A similar comparison was made with the recently reported GlpQ sequence of *B. subtilis* (21), which was 56% similar and 36% identical to the sequence of protein D. The homologous regions were found mainly in the N-terminal parts of the processed forms of *B. subtilis* GlpQ and protein D, supporting the speculation that the N-terminal region may contain the catalytic domains (23). A problem in using protein D as a component in a vaccine might be cross-reactivity to GlpQ of *E. coli* and other bacteria. However, the risk must be considered relatively small since antibodies directed against protein D would probably not cross-react with GlpQ of viable *E. coli* because the protein is found in the periplasm of *E. coli* (14).

Nucleotide sequence accession numbers. The nucleotide sequences of the *hpd* genes from strains 3639, 3640, 6-7626, Eagan, HK695, and NCTC 8468 have been assigned EMBL accession numbers Z35656, Z35657, Z35658, Z35659, Z35660, and Z35661, respectively.

TABLE 2. Synthesized oligonucleotide primers

Name	Position[a]	Nucleotide sequence
hpd-1	1231–1215	5′ GGT TGT GAG GGA TAT TA 3′
hpd-2	966–950	5′ CAA CAC CAT CGG CAT AT 3′
hpd-3	702–686	5′ GCG TTT CAG CAG CAA TA 3′
hpd-4	457–438	5′ GTA AAG TCG ATG ACA TAG TA 3′
hpd-5	192–176	5′ ATG AAT GGC TGC TAC AA 3′
hpd-6B[b]	58–74	5′ TCA CGG AAT AAT TAA TT 3′
hpd-6N[c]	52–68	5′ GTC GCC TTT TTT GTA AC 3′
hpd-7	322–340	5′ AGC AAG ATT TAG CAA TGA C
hpd-8	600–616	5′ CAA GGC TTA GAA AAA TC 3′
hpd-9	933–949	5′ ATG GCA GAA GTG GTT AA 3′

[a] Range in 5′ to 3′ orientation.
[b] Used for strains Eagan and HK695.
[c] Used for strains 3639, 3640, 6-7626, and NCTC 8468.

We thank Gertrud Hansson for skillful technical assistance.

This work was supported by grants from Swedish Medical Research Council (projects 03163 and 05196), the Medical Faculty of Lund, the Alfred Österlund Foundation, the Crafoordska Foundation, the G. and J. Kock Foundation, and Malmö Medical Service Foundation.

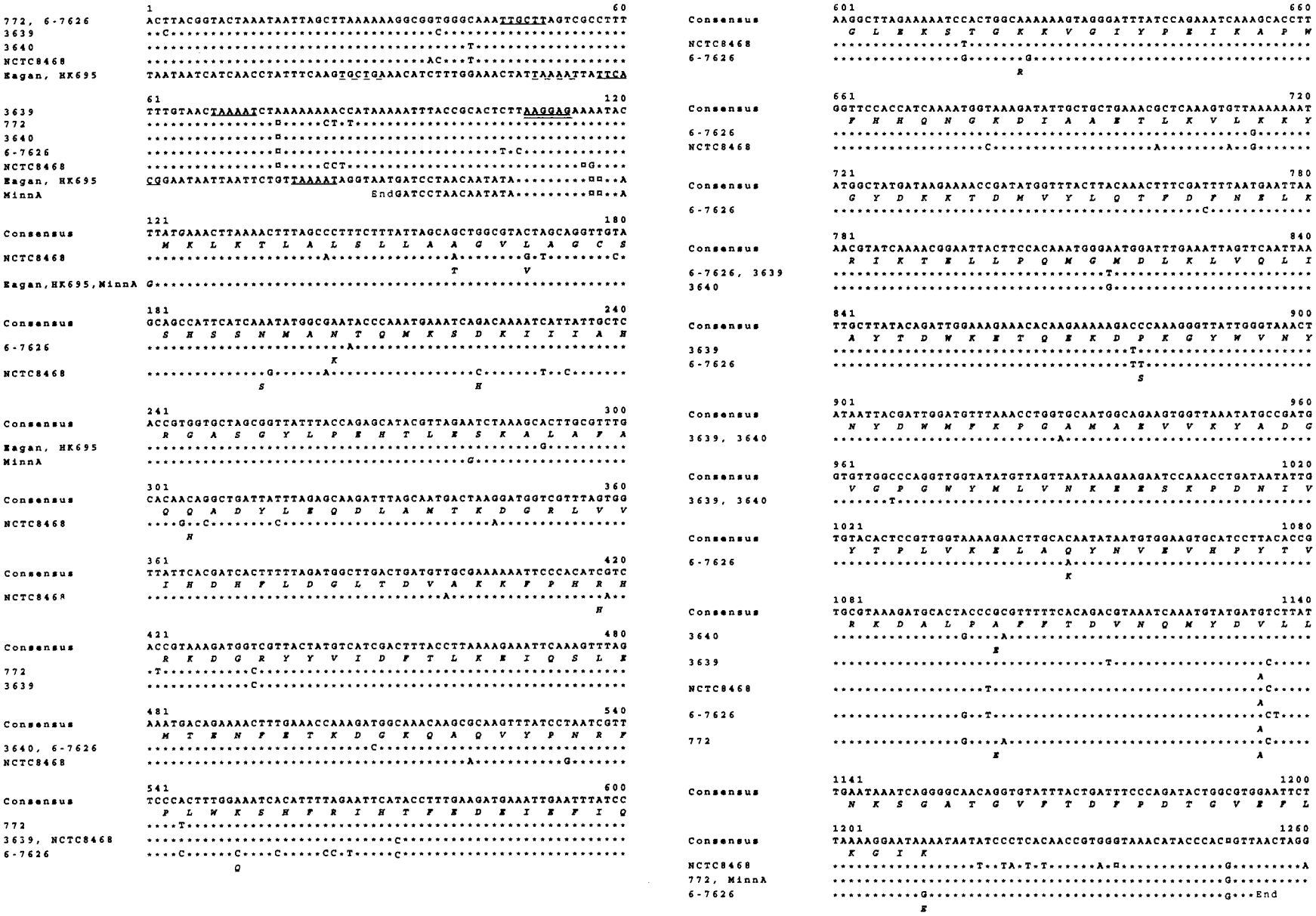

FIG. 1. Comparison of nucleotide (roman) and deduced (italic) amino acid sequences of protein D from eight *H. influenzae* strains. Sequences of strains differing from the consensus sequence in each 60-bp interval are shown below the consensus sequence. Asterisks denote nucleotide sequence identity. Sequence gaps are indicated with the symbol □. Start and stop codons for protein D are indicated in italics. A putative ribosomal binding site is indicated with a double underline. The −10 and −35 sequences of possible promoter regions are underlined. An alternate possible promoter region of strains Eagan and HK695 is indicated with a broken underline. "End" represents the end of *H. influenzae* DNA insert for strain MinnA and the end of available nucleotide sequence of strain 6-7626.

REFERENCES

1. **Adams, W. G., K. A. Deaver, S. L. Cochi, B. D. Plikaytis, E. R. Zell, C. V. Broome, and J. D. Wenger.** 1993. Decline of childhood *Haemophilus influenzae* type b (Hib) disease in the Hib vaccine era. JAMA **269:**221–226.
2. **Akkoyunlu, M., M. Ruan, and A. Forsgren.** 1991. Distribution of protein D, an immunoglobulin D-binding protein, in *Haemophilus* strains. Infect. Immun. **59:**1231–1238.
3. **Berk, S. L., S. A. Holtsclaw, S. L. Wiener, and J. K. Smith.** 1982. Nontypeable *Haemophilus influenzae* in the elderly. Arch. Intern. Med. **142:**537–539.
4. **Brinton, C. C., Jr., M. J. Carter, D. B. Derber, S. Kar, J. A. Kramarik, A. C. C. To, S. C. M. To, and S. W. Wood.** 1989. Design and development of pilus vaccines for *Haemophilus influenzae* diseases. Pediatr. Infect. Dis. J. **8**(Suppl. 1)**:**554–561.
5. **Deich, R. A., A. Anilionis, J. Fulginiti, B. J. Metcalf, S. Quataert, T. Quinn-Dey, G. W. Zlotnick, and B. A. Green.** 1990. Antigenic conservation of the 15,000-dalton outer membrane lipoprotein PCP of *Haemophilus influenzae* and biologic activity of anti-PCP antisera. Infect. Immun. **58:**3388–3393.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
7. **Green, B. A., J. E. Farley, T. Quinn-Dey, R. A. Deich, and G. W. Zlotnick.** 1991. The *e* (P4) outer membrane protein of *Haemophilus influenzae*: biologic activity of anti-*e* serum and cloning and sequencing of the structural gene. Infect. Immun. **59:**3191–3198.
8. **Green, B. A., T. Quinn-Dey, and G. W. Zlotnick.** 1987. Biologic activities of antibody to a peptidoglycan-associated lipoprotein of *Haemophilus influenzae* against multiple clinical isolates of *H. influenzae* type b. Infect. Immun. **55:**2878–2883.
9. **Janson, H., L.-O. Hedén, and A. Forsgren.** 1992. Protein D, the immunoglobulin D-binding protein of *Haemophilus influenzae*, is a lipoprotein. Infect. Immun. **60:**1336–1342.
10. **Janson, H., L.-O. Hedén, A. Grubb, M. Ruan, and A. Forsgren.** 1991. Protein D, an immunoglobulin D-binding protein of *Haemophilus influenzae*: cloning, nucleotide sequence, and expression in *Escherichia coli*. Infect. Immun. **59:**119–125.
11. **Janson, H., M. Ruan, and A. Forsgren.** 1993. Limited diversity of the protein D gene (*hpd*) among encapsulated and nonencapsulated *Haemophilus influenzae* strains. Infect. Immun. **61:**4546–4552.
12. **Kimura, A., P. A. Gulig, G. H. McCracken, Jr., T. A. Loftus, and E. J. Hansen.** 1985. A minor high-molecular-weight outer membrane protein of *Haemophilus influenzae* type b is a protective antigen. Infect. Immun. **47:** 253–259.
13. **Kroll, J. S., and E. R. Moxon.** 1990. Capsulation in distantly related strains of *Haemophilus influenzae* type b: genetic drift and gene transfer at the capsulation locus. J. Bacteriol. **172:**1374–1379.
14. **Larson, T. J., M. Ehrmann, and W. Boos.** 1983. Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the *glp* regulon. J. Biol. Chem. **258:**5428–5432.
15. **Loeb, M. R.** 1987. Protection of infant rats from *Haemophilus influenzae* type b infection by antiserum to purified outer membrane protein *a*. Infect. Immun. **55:**2612–2618.
16. **Munson, R. S., Jr., and D. M. Granoff.** 1985. Purification and partial characterization of outer membrane proteins P5 and P6 from *Haemophilus influenzae* type b. Infect. Immun. **49:**544–549.
17. **Munson, R. S., Jr., and K. Sasaki.** 1993. Protein D, a putative immunoglobulin D-binding protein produced by *Haemophilus influenzae*, is glycerophosphodiester phosphodiesterase. J. Bacteriol. **175:**4569–4571.
18. **Munson, R. S., Jr., J. L. Shenep, S. J. Barenkamp, and D. M. Granoff.** 1983. Purification and comparison of outer membrane protein P2 from *Haemophilus influenzae* type b isolates. J. Clin. Invest. **72:**677–684.
19. **Murphy, T. F., and M. A. Apicella.** 1987. Nontypeable *Haemophilus influenzae*: a review of clinical aspects, surface antigens, and the human immune response to infection. Rev. Infect. Dis. **9:**1–15.
20. **Musser, J. M., J. S. Kroll, E. R. Moxon, and R. K. Selander.** 1988. Clonal population structure of encapsulated *Haemophilus influenzae*. Infect. Immun. **56:**1837–1845.
21. **Nilsson, R.-P., L. Beijer, and B. Rutberg.** 1994. The *glpT* and *glpQ* genes of the glycerol regulon in *Bacillus subtilis*. Microbiology **140:**723–730.
22. **Ruan, M., M. Akkoyunlu, A. Grubb, and A. Forsgren.** 1990. Protein D of *Haemophilus influenzae*. A novel bacterial surface protein with affinity for human IgD. J. Immunol. **145:**3379–3384.
23. **Tommassen, J., K. Eiglmeier, S. T. Cole, P. Overduin, T. J. Larson, and W. Boos.** 1991. Characterization of two genes, *glpQ* and *ugpQ*, encoding glycerophosphoryl diester phosphodiesterases of *Escherichia coli*. Mol. Gen. Genet. **226:**321–327.
24. **Wu, H. C., and M. Tokunaga.** 1986. Biogenesis of lipoproteins in bacteria. Curr. Top. Microbiol. Immunol. **125:**127–157.
25. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.