*Proc. Natl. Acad. Sci. USA*
Vol. 83, pp. 1106–1110, February 1986
Microbiology

# Genes involved in *Haemophilus influenzae* type b capsule expression are part of an 18-kilobase tandem duplication

**(virulence determinant/high-frequency rearrangement)**

SUSAN K. HOISETH*, E. RICHARD MOXON†, AND RICHARD P. SILVER*

*Division of Bacterial Products, Office of Biologics Research and Review, Center for Drugs and Biologics, Food and Drug Administration, Bethesda, MD 20205; and †Infectious Disease Unit, Department of Paediatrics, John Radcliffe Hospital, Oxford, England, OX3-9DU

*Communicated by Hamilton O. Smith, October 7, 1985*

ABSTRACT **Encapsulated *Haemophilus influenzae* type b produce nonencapsulated variants at high frequency (0.1–0.3%). Cosmid cloning was used to investigate the genetic mechanism responsible for this instability. Analysis of three independently derived cosmid clones showed that the b$^+$ parental strain contains an 18-kilobase tandem duplication of genes involved in type b capsule expression. Loss of one complete copy of the 18-kilobase tandem duplication occurred following transformation of the cosmid clones into Rec$^+$, but not Rec$^-$, *Escherichia coli*, and in *H. influenzae* strains that had spontaneously lost capsule expression. These results suggest that high-frequency loss of type b capsule expression is due to *rec*-dependent recombination between the two copies of the 18-kilobase tandem repeat. This is further supported by our finding that introduction of the *H. influenzae rec-1* mutation stabilized type b capsule expression.**

The *Haemophilus influenzae* type b capsule is considered an important virulence determinant (1–4). Synthesis of the type b capsule, however, is genetically unstable, and we have shown previously that loss of type b capsule expression occurs at frequencies of 0.1–0.3% (5). Hybridization analysis of these high-frequency variants revealed the loss of a 9-kilobase (kb) *Eco*RI restriction fragment homologous to a cloned DNA fragment containing sequences necessary for type b capsule expression. We also found that all of 33 independently isolated type b strains had multiple *Eco*RI fragments sharing homology with this 9-kb probe (5). This probe came from a recombinant phage λ clone that we have shown contains some, but not all, of the genes necessary for type b capsule expression. To identify additional sequences involved in capsule expression, and to attempt to identify the genetic mechanism responsible for the instability of this region, we constructed a cosmid library of *H. influenzae* type b chromosomal DNA. Our results indicate that the b$^+$ parental strain contains an 18-kb tandem duplication of at least some of the genes involved in type b capsule expression. Our data from both the cosmid clones in *Escherichia coli* and hybridization analysis of *H. influenzae* chromosomal DNA suggest that high-frequency loss of type b capsule expression is due to a *rec*-dependent recombinational event between the two copies of the 18-kb tandem repeat.

## METHODS

**Bacterial Strains.** The source of DNA for the *H. influenzae* cosmid library was the prototypic type b strain Eagan (6). The Rec$^-$ *E. coli* strain HB101 and the Rec$^+$ *E. coli* strain C600 are described elsewhere (7). A *rec-1* streptomycin-resistant (Sm$^R$) *H. influenzae* strain (8) was kindly provided by J. Setlow.

**Cosmid Cloning.** High molecular weight *H. influenzae* chromosomal DNA was isolated essentially as described by Smith (9), but the EDTA concentration of the resuspension buffer was reduced to 5 mM to prevent premature lysis. This procedure employs gentle phenol extraction, and there is no ethanol precipitation before the *Sau*3A cleavage reaction. Residual phenol was removed by dialysis, and the DNA was partially cleaved with *Sau*3A (Boehringer Mannheim). The DNA was phenol-extracted, butanol-extracted, and ethanol-precipitated prior to size fractionation in a 5–20% NaCl gradient (SW41 rotor, 35,000 rpm, 4.5 hr, 4°C). Salt was removed by dialysis, and fractions containing DNA in the 30- to 50-kb size range were identified by electrophoresis in a 0.3% agarose gel. DNA from such fractions was mixed with *Bam*HI-cleaved pHC79 (10) vector DNA at an insert/vector molar ratio of approximately 2:1. The DNAs were ligated using T4 DNA ligase (Bethesda Research Laboratories) and the ligated DNA was packaged using λ DNA packaging extracts (also from Bethesda Research Laboratories). Recombinant clones were isolated on LB plates (7) containing ampicillin at 50 μg/ml.

**Colony Hybridization.** Recombinant clones were prong-replicated to nitrocellulose filters, and the filters were incubated for 6–18 hr on LB plates containing ampicillin at 50 μg/ml. The colonies were lysed and neutralized according to Hill and Payne (11). The filters were prewashed, prehybridized, and hybridized as described (7).

**DNA Probes.** $^{32}$P-labeled probes were prepared using a nick-translation kit (Bethesda Research Laboratories) and [α-$^{32}$P]dATP (New England Nuclear).

**Plasmid Isolation.** Plasmid DNA was isolated by the rapid alkaline-lysis method (12). Detailed restriction mapping employed DNA that had been further purified by banding in cesium chloride/ethidium bromide gradients.

**Southern Hybridizations.** *H. influenzae* chromosomal DNA or *Eco*RI-cut cosmid DNA was bidirectionally transferred to nitrocellulose, and the filters were hybridized as described (13).

**Antiserum Agar Plates.** LB agar plates containing 2% (vol/vol) antiserum to *H. influenzae* type b were used to screen for the production of capsular polysacchride. The high-titer burro antiserum was obtained from J. B. Robbins (Bethesda, MD). Growth of *H. influenzae* type b or the crossreactive *E. coli* capsule type K100 on such plates produces large precipitin halos surrounding the colonies.

**Construction of a Rec$^-$ Type b Strain.** The characterized *H. influenzae rec-1* mutation (8) was available only in an unencapsulated strain. The *rec-1* mutation was moved into the type b strain, Eagan, by cotransformation with a linked streptomycin-resistance determinant. DNA from the *rec-1* Sm$^R$ strain was used to transform strain Eagan to streptomycin-resistance, and the Sm$^R$ transformants were screened for cotransformation of *rec-1* using UV-sensitivity. Purifica-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

tion of DNA, preparation of competent cells, and the transformation procedure have been described (13).

## RESULTS

**Screening the Cosmid Library and Construction of a Map of the Cap b Region.** Cosmid clones of *Haemophilus* DNA were established in *E. coli* and colonies were screened initially on antiserum agar plates to identify clones synthesizing immunologically reactive type b polysaccharide. None of 2800 clones were positive by this test. Earlier transformation work (14) suggested that a region spanning as much as 50 kb might be required for type b capsule expression. The isolation of the entire sequence in one cosmid clone may thus not be possible. We therefore performed colony hybridization using a fragment from our previously described λ clone as a probe. This clone, Ch.4:48 (13), contains two *Eco*RI fragments of 4.4 and 9 kb and has been shown to contain some, but not all, of the genes necessary for type b capsule expression.

A total of 1700 cosmid clones were tested by colony hybridization using the 4.4-kb probe, and 43 (2.5%) were found to be positive. Plasmid DNA from 13 of these "4.4 kb-positive" clones was digested with *Eco*RI and run on a 0.7% agarose gel. The gel was bidirectionally transferred to nitrocellulose, and one filter was probed with the 9-kb probe, whereas the other filter was probed with the 4.4-kb probe. Previous work showed that the 4.4-kb probe hybridized to a chromosomal *Eco*RI fragment of 4.4 kb, and the 9-kb probe hybridized to three *Eco*RI fragments, of 9, 10.2, and ≈20 kb (5). All 13 cosmid clones contained a 4.4-kb *Eco*RI fragment, homologous to the 4.4-kb probe, and a 9-kb fragment that hybridized to the 9-kb probe. Four of the clones also contained a 10.2-kb *Eco*RI fragment homologous to the 9-kb probe, suggesting that the 9- and 10.2-kb *Eco*RI fragments are closely linked in the chromosome. *Eco*RI fragments that were positive with the 4.4- or 9-kb probe but which were not the predicted 4.4, 9, 10.2, or 20 kb in size were suggestive of ends of the clones in which a portion of the 4.4- or 9-kb homolog was attached to the vector. On ethidium bromide-stained gels, all 13 clones showed *Eco*RI fragments of 2.1, 2.7, 4.4, and 9 kb; this suggested that these were internal fragments. By identifying internal and end fragments, we were able to construct a tentative *Eco*RI map (Fig. 1).

Three clones with different endpoints were chosen for further analysis. The size of these three clones, based on the sum of the *Eco*RI fragments (≈39 kb, ≈39 kb, and ≈37 kb for pSKH3, pSKH4, and pSKH5, respectively; see Figs. 1 and 2), did not equal the sum obtained for the *Bam*HI fragments (≈48 kb, ≈45 kb, and ≈46 kb, respectively; Fig. 2*D*) and suggested that some of the *Eco*RI fragments might be present in more than one copy. That the 4.4-kb *Eco*RI fragment was present in two copies was suggested by pSKH4; this clone had a 4.4-kb *Eco*RI fragment positive with the 4.4-kb probe (Fig. 2*C*, lane 2) but also had a 1.3-kb fragment positive with the 4.4-kb probe, suggesting that this fragment contains part of a second copy, attached to the 0.37-kb *Eco*RI vector end.

Digestion of pSKH3 and pSKH5 with *Bam*HI gave two large fragments, while pSKH4 gave three (Fig. 2*D*). These results indicate that there are two *Bam*HI sites in pSKH3 and pSKH5 and three sites in pSKH4. *Bam*HI/*Eco*RI double digests of pSKH3 and pSKH5, however, showed that only the 2.1-kb *Eco*RI fragment was cut by *Bam*HI (data not shown), which suggests that the 2.1-kb *Eco*RI is also present in two copies. When *Bam*HI-cut cosmid clones were tested with the 4.4-kb probe, two large fragments were found to hybridize with the probe (Fig. 2*E*). As *Bam*HI does not cut within the 4.4-kb fragment itself, these results add further support to the existence of two copies of the 4.4-kb fragment.

In all three clones, the spacing between the *Bam*HI sites was 18 kb (Fig. 2*D*). This requires that the 2.1-kb *Eco*RI fragments be 18 kb apart and further supports the map order shown in Fig. 1. More detailed restriction mapping (Fig. 3) confirmed the *Eco*RI map and showed that the Cap b region contains an 18-kb tandem duplication.

**Transformation of Cosmid Clones into Rec$^+$ *E. coli*.** Tandem duplications are known to be unstable in a recombination-

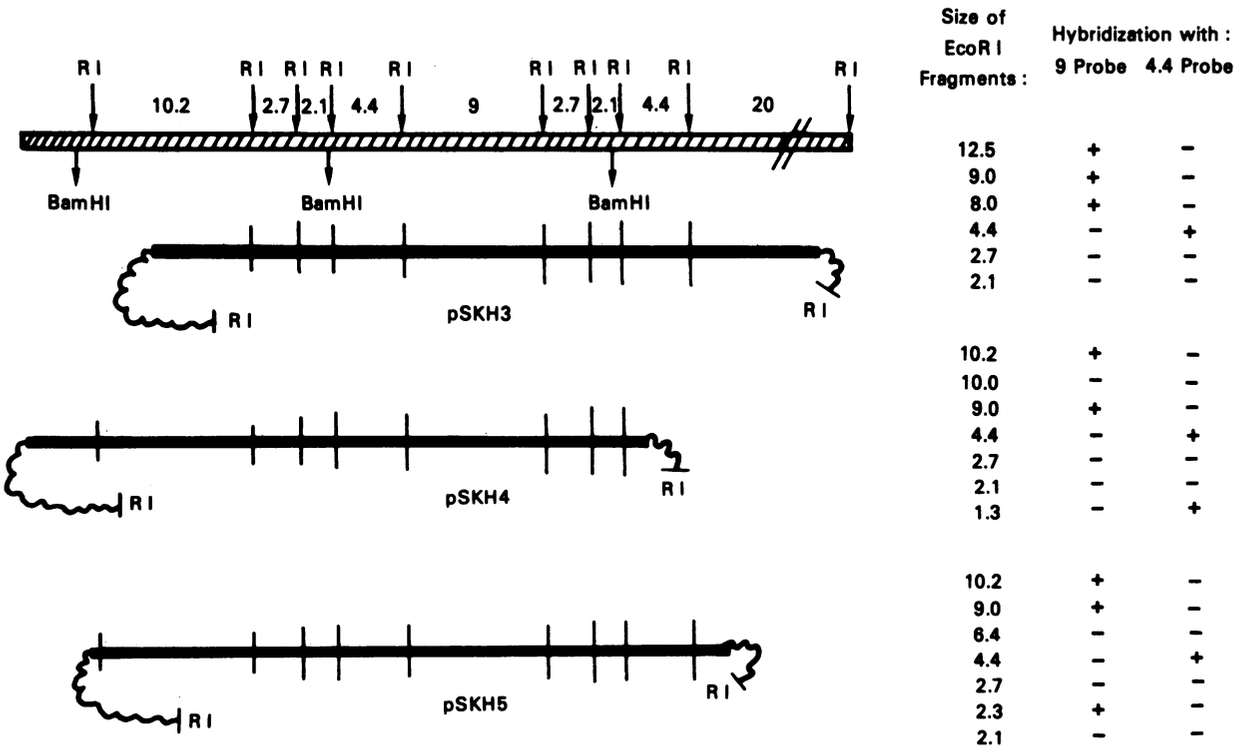


FIG. 1. *Eco*RI restriction map of the Cap b region. The top line represents a composite map, and three independent clones are shown below. The solid lines represent *H. influenzae* insert DNA, and the wavy lines represent vector DNA. Sizes are given in kb.

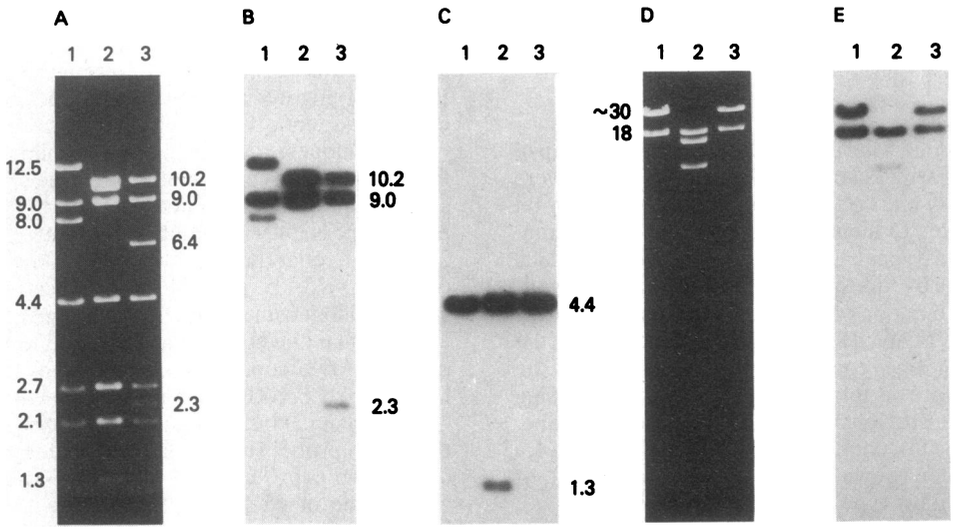


FIG. 2. Hybridization analysis of cosmid clones. (*A*) Ethidium bromide-stained gel of *Eco*RI-cut cosmid clones. (*B*) *Eco*RI-cut cosmid clones probed with the 9-kb probe. (*C*) *Eco*RI-cut cosmid clones probed with the 4.4-kb probe. (*D*) Ethidium bromide-stained gel of *Bam*HI-cut cosmid clones. (*E*) *Bam*HI-cut cosmid clones probed with the 4.4-kb probe. Lanes: 1, pSKH3; 2, pSKH4; 3, pSKH5. Fragment sizes are given in kb.

proficient host (15), and we therefore tested the stability of our cosmid clones in a Rec$^+$ *E. coli* strain. Purified cosmid DNA (isolated originally from the Rec$^-$ strain HB101) was used to transform the Rec$^+$ strain C600. Ampicillin-resistant transformants were purified and plasmid DNA was isolated after overnight incubation in LB/ampicillin broth. By cutting the plasmid DNA with *Bam*HI (which cuts only once in each copy of the tandem duplication; Fig. 3), we found that three of three transformants using pSKH3 DNA and three of three using pSKH5 DNA were missing an 18-kb *Bam*HI fragment (Fig. 4). That this was not simply an artifact of the transformation was shown by transformation of the original cosmid DNA into a new Rec$^-$ host (HB101). Three of three ampicillin-resistant pSKH3 transformants and three of three pSKH5 transformants were unaltered (Fig. 4). These results thus suggest that loss of the 18-kb *Bam*HI fragment is dependent on a functional recombination system. The results with pSKH4 were not as clear-cut. Although two of four C600 transformants were missing the predicted 18-kb *Bam*HI fragment, the other two have apparently undergone some other type of rearrangement (data not shown).

**Analysis of Chromosomal DNA from Cap$^-$ Mutants and Effect of Rec$^-$ Mutation on Capsular Stability.** Our previous work showed that high-frequency loss of b capsule expression was associated with the loss of a 9-kb *Eco*RI restriction fragment from the *H. influenzae* chromosome (5). Based on the results from the cosmid clones, we felt that the mutants were probably missing one complete copy of the 18-kb tandem duplication and that only the 9-kb *Eco*RI fragment appeared to be missing because the other *Eco*RI fragments are present in two exact copies (e.g., the 2.1-, 2.7-, and 4.4-kb fragments). The 9-kb-homologous sequences (the 9-, 10.2-, and 20-kb *Eco*RI fragments) appear as *Eco*RI fragments of different sizes because they contain the endpoints of the duplication and are asymmetric about the join between the two copies. Since *Bam*HI cuts only once in each copy, one would predict that loss of one complete copy would result in loss of the 18-kb *Bam*HI fragment that spans the join between the two copies. Comparison of *Bam*HI-cut chromosomal DNA from the b$^+$ strain Eagan and b$^-$ derivatives of it, probed with either the 4.4- or 9-kb probes, showed loss of the predicted 18-kb *Bam*HI fragment (Fig. 5). These results suggest that loss of type b capsule expression is due to loss of one copy of the 18-kb tandem repeat.

To test the idea that a *rec*-dependent recombinational event is responsible for this loss, we introduced the *Haemophilus rec-1* mutation into the type b strain Eagan. The *rec-1* mutation was transformed into strain Eagan using a nearby

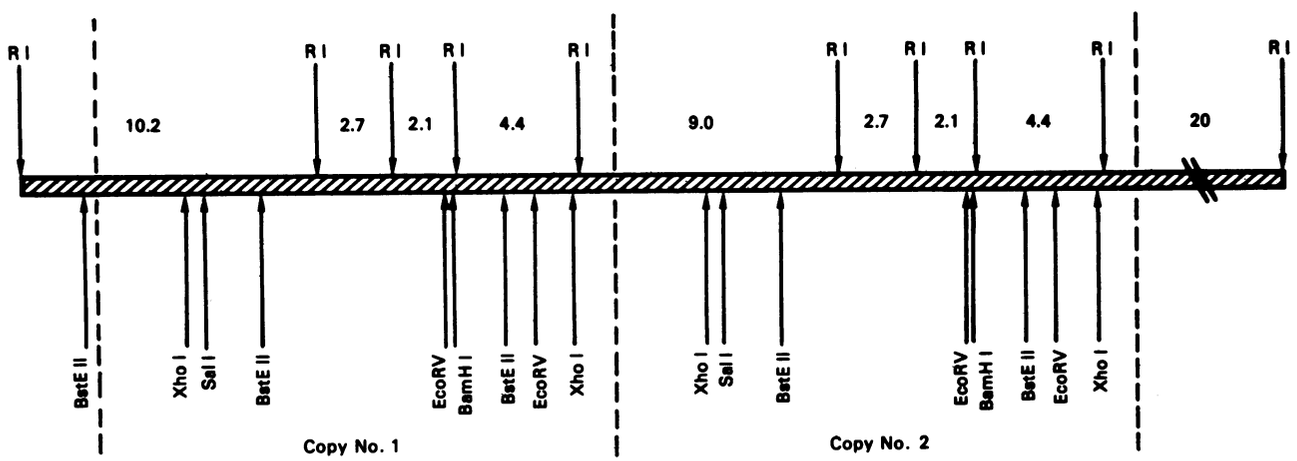


FIG. 3. Composite map of the Cap b region. This map was derived from double and single digests of pSKH3, pSKH4, and pSKH5. It has been independently confirmed by hybridization analysis of *H. influenzae* chromosomal DNA cut with *Bam*HI, *Bst*EII, or *Eco*RV and probed with either the 4.4- or 9-kb probe (data not shown).

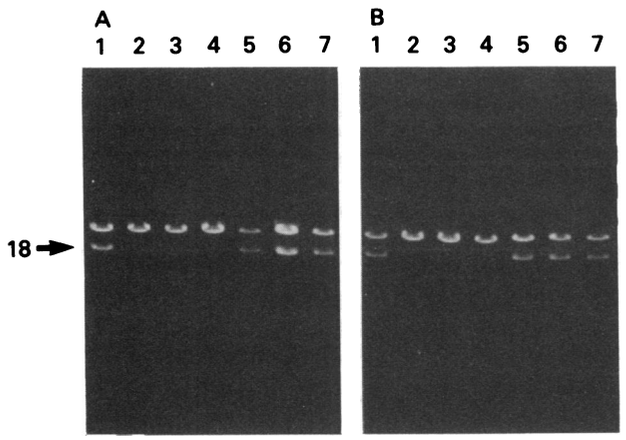


FIG. 4. Ethidium bromide-stained gels, showing loss of the 18-kb *Bam*HI fragment after transformation of the cosmid clones into Rec$^+$ *E. coli*. (*A*) *Bam*HI-cut pSKH3. (*B*) *Bam*HI-cut pSKH5. Lanes 1: the original clone, isolated from HB101. Lanes 2–4: independent C600 (Rec$^+$) transformants. Lanes 5–7: independent transformants from transformation of the original CsCl-banded DNA into a new HB101 (Rec$^-$) host (transformation controls).

streptomycin-resistance determinant as the selectable marker. A streptomycin-resistant, Rec$^+$ transformant was isolated to serve as the Rec$^+$ control. The Sm$^R$ Rec$^+$ transformant and two independent Sm$^R$ Rec$^-$ transformants were grown in parallel and screened for capsule-deficient mutants. The Rec$^-$ strains grew somewhat slower than the Rec$^+$ strain, but by incubation of the Rec$^-$ strains for 26 hr and the Rec$^+$ strain for 18 hr, colonies equal in size and indistinguishable by visual inspection were obtained. The plates were coded and screened "blindly" for capsule-deficient mutants. From six plates containing a total of approximately 10,000 Rec$^+$ colonies, we found 38 capsule-deficient mutants. From 12 plates containing a total of approximately 20,000 Rec$^-$ colonies, we found no capsule-deficient mutants.

## DISCUSSION

We (5) and others (1, 16) have shown previously that expression of the *H. influenzae* type b capsule is genetically unstable. The results presented here show that at least some of the genes involved in expression of the type b capsule are contained within an 18-kb tandem duplication. The well-documented instability of tandem duplications (15) suggested that high-frequency loss of b capsule expression might be due to *rec*-dependent recombination resulting in loss of one copy. Our data from both the cosmid clones in *E. coli* and from hybridization analysis of *H. influenzae* chromosomal DNA are compatible with the idea that loss of b capsule expression is due to a *rec*-dependent recombinational event between the two copies of the 18-kb tandem repeat. Additional support for this model comes from our finding that introduction of the *H. influenzae rec-1* mutation stabilized capsule expression.

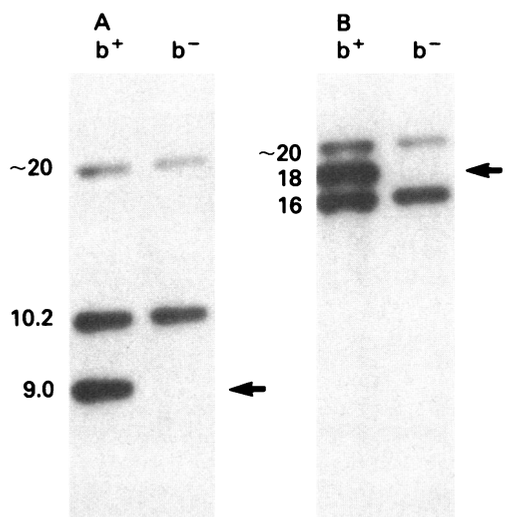


FIG. 5. Southern hybridization of b$^-$ variant showing loss of the 18-kb *Bam*HI fragment from the *H. influenzae* chromosome. (*A*) *Eco*RI-cut chromosomal DNA probed with the 9-kb probe. (*B*) *Bam*HI-cut chromosomal DNA probed with the 9-kb probe.

Our results do not explain why we see loss of b capsule expression when one copy of the repeat is lost. They do suggest, however, that the two copies are not functionally equivalent. Although at the crude level of a restriction map the two copies look similar, there may be some differences. Alternatively, expression may be dependent on the presence of the sequence at the join between the two copies. The join region represents a sequence not present in the single copy, and thus loss of this sequence might account for loss of type b capsule expression.

The generation of tandem duplications has been hypothesized as a means of allowing organisms to adapt to short-term environmental stress without stably altering their genetic constitution (15). Duplications are quite common in *E. coli* and *Salmonella* and have been estimated to occur at frequencies of $10^{-3}$–$10^{-4}$ for almost any region examined (15). They are particularly observed under conditions where it is beneficial to have more than one copy of a gene [e.g., ribosomal RNA genes under conditions of rapid growth (17) and antibiotic-resistance genes (18)]. Another example is cholera toxin. Work done by Mekalanos showed that some El Tor strains have two tandem copies of the gene for cholera toxin (19). When he selected *in vivo* (in rabbits) for hypertoxigenic strains, he found three to five copies of the gene.

In the case of the *H. influenzae* capsule, we see the reverse phenomenon. The type b strains have two copies and tend to lose one copy. Although in most cases the capsule is probably necessary for survival of the organism in the bloodstream (2–4), there may be benefits from a mechanism that allows it to lose its capsule at such a high frequency. The capsule certainly must impose a great biosynthetic drain on the organism. Also, some reports suggest that unencapsulated *Haemophilus* adhere better to epithelial cells (20, 21) and may thereby have an advantage for maintaining colonization of mucosal surfaces.

The question thus arises as to the evolutionary relationship between naturally occurring nontypable and type b strains. We have reported (5) the *in vivo* conversion of a type b strain to the nonencapsulated form in the nose of an infant rat. Hybridization analysis of nontypable disease isolates, however, suggests that only a small percentage represent capsule-deficient mutants derived from type b strains (5). Lack of a common ancestry for type b strains and most nontypable isolates is also suggested by the lack of similarity between outer membrane protein profiles and lipopolysaccharide patterns for type b and nontypable strains (22–24). Thus, although some nontypable strains may have been derived from type b ancestors, the majority probably are not.

Another question remaining to be answered is the source of type b organisms in nature, given the high frequency with which *H. influenzae* loses its capsule. We have not detected encapsulated revertants from the high-frequency variants but have used methods capable of detecting a reversion frequency of only about 1 in $10^7$ bacteria (5).

Examination of many independently isolated type b strains suggests that the tandem duplication is extremely conserved. That the tandem duplication results in such genetic instabil-

ity, and yet appears to be so highly conserved, argues that the duplication is biologically important, or it would not have been so precisely maintained throughout evolution. The loss of type b capsule expression via loss of one copy of a tandem duplication appears to be a novel mechanism for loss of a virulence determinant. The role that the seemingly paradoxical loss of a major virulence determinant plays in the overall biology of this microorganism remains to be determined. The finding that the Cap b region contains an 18-kb tandem duplication does, however, offer a plausible explanation for the observed high-frequency loss of type b capsule expression.

We thank Dr. Jane Setlow for providing the *rec-1* $Sm^R$ *H. influenzae* strain, Mrs. Giovanna Connor for manuscript preparation, and Drs. William Habig, Kathryn Stein, and Charles W. Finn, Jr., for helpful discussion and manuscript review.

1. Pittman, M. (1931) *J. Exp. Med.* **53,** 471–492.
2. Moxon, E. R. & Vaughn, K. A. (1981) *J. Infect. Dis.* **143,** 517–524.
3. Sutton, A., Schneerson, R., Kendall-Morris, S. & Robbins, J. B. (1982) *Infect. Immun.* **35,** 95–104.
4. Zwahlen, A., Winkelstein, J. A. & Moxon, E. R. (1983) *J. Infect. Dis.* **148,** 385–394.
5. Hoiseth, S. K., Connelly, C. & Moxon, E. R. (1985) *Infect. Immun.* **49,** 389–395.
6. Anderson, P., Johnston, R. B. & Smith, D. H. (1972) *J. Clin. Invest.* **51,** 31–38.
7. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
8. Setlow, J. K., Boling, M. E., Beattie, K. L. & Kimball, R. F. (1972) *J. Mol. Biol.* **68,** 361–378.
9. Smith, M. G. (1967) *Methods Enzymol.* **12,** 545–550.
10. Hohn, B. & Collins, J. (1980) *Gene* **11,** 291–298.
11. Hill, W. E. & Payne, W. L. (1984) *J. Assoc. Off. Anal. Chem.* **67,** 801–807.
12. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
13. Moxon, E. R., Deich, R. A. & Connelly, C. (1984) *J. Clin. Invest.* **73,** 298–306.
14. Catlin, B. W., Bendler, J. W. & Goodgall, S. H. (1972). *J. Gen. Microbiol.* **70,** 411–422.
15. Anderson, R. P. & Roth, J. R. (1977) *Annu. Rev. Microbiol.* **31,** 473–505.
16. Catlin, B. W. (1970) *Am. J. Dis. Child.* **120,** 203–210.
17. Anderson, P. & Roth, J. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3113–3117.
18. Hashimoto, H. & Rownd, R. H. (1975) *J. Bacteriol.* **123,** 56–68.
19. Mekalanos, J. J. (1983) *Cell* **35,** 253–263.
20. Lampe, R. M., Mason, E. O., Kaplan, S. L., Umstead, C. L., Yow, M. D. & Feigin, R. D. (1982) *Infect. Immun.* **35,** 166–172.
21. Pichichero, M. E. (1984) *J. Med. Microbiol.* **18,** 107–116.
22. Loeb, M. R. & Smith, D. H. (1980) *Infect. Immun.* **30,** 709–717.
23. Barenkamp, S. J., Munson, R. S. & Granoff, D. M. (1982) *Infect. Immun.* **36,** 535–540.
24. Inzana, T. J. (1983) *J. Infect. Dis.* **148,** 492–499.