INFECTION AND IMMUNITY, Oct. 1985, p. 279–283
0019-9567/85/1000279-05$02.00/0

Vol. 50, No. 1

Nucleotide Sequences of Four Variants of the K88 Gene of Porcine Enterotoxigenic *Escherichia coli*

COLIN W. DYKES,* ISHBEL J. HALLIDAY, MELANIE J. READ, ADRIAN N. HOBDEN, AND STEPHEN HARFORD

Department of Biotechnology, Glaxo Group Research, Greenford, Middlesex UB6 0HE, United Kingdom

Received 19 February 1985/Accepted 12 June 1985

The nucleotide sequences of four variants of the *Escherichia coli* K88 antigen gene, K88ab1, K88ab2, K88ac, and K88ad, have been determined. The K88ab2 and K88ac sequences have not been reported previously. The K88ab1 sequence is very similar to that determined by other workers, but the K88ad sequence differs considerably from that described in a previous report. Comparison of the amino acid sequences inferred from the gene sequences revealed certain clusters of amino acid substitutions which have been correlated with areas of potential antigenicity in the mature proteins.

The pathogenicity of certain enterotoxigenic *Escherichia coli* strains is due to their ability to bind specifically to host intestinal epithelial cells via filamentous structures known as fimbriae. Gut colonization is of great importance in diarrheal diseases, and there is considerable interest in using vaccines based on these attachment factors (1, 14, 16, 17). The fimbriae of porcine enterotoxigenic *E. coli* have been shown to consist of multimers of the K88 protein, held together by noncovalent interactions thought to be mainly hydrophobic in nature (11). These fimbriae have been classified into three major subtypes, K88ab, K88ac, and K88ad, based on immunological cross-reactivity (8). This representation indicates that each subtype possesses a common "a" factor plus a variable "b," "c," or "d" factor, each of which may consist of one or more individual antigenic determinants.

The amino acid sequence of the K88ab protein (11) and the nucleotide sequences of the K88ab and K88ad genes (6, 7) have been determined. The N- and C-terminal regions of the K88 variants studied are identical, and it has been suggested that these are involved in stabilizing the polymeric structure of the fimbriae (11). Klemm and Mikkelsen (12) have used the hydrophilicity analysis method of Hopp and Woods (9) to identify potential antigenic determinants in K88ab. This method has also been used by Gaastra et al. (6) to compare the amino acid sequences of K88ab and K88ad since any sequence differences which occur within predicted antigenic determinants are likely to contribute to the variable b and d factors. Similarly, it is likely that strongly antigenic regions common to both proteins contribute to the a factor. We have extended this approach by determining the sequence of the third major variant, K88ac, two different K88ab variants, and a K88ad variant which differs considerably from that sequenced previously.

MATERIALS AND METHODS

Restriction endonucleases. Restriction endonucleases were obtained from Boehringer Corp. London.

Bacterial strains and plasmids. *E. coli* strain 2134E was obtained from the Glaxo Group Research Culture Collection and was the source of a K88ab gene, designated K88ab1, carried on plasmid p520. *E. coli* strains E68, Abbotstown A1, and 56/190 were generously provided by I. Orskov, International *Escherichia* and *Klebsiella* Centre, Copenhagen, Denmark. The latter two strains were the source of K88ac and K88ad genes carried on plasmids pA1 and p56/190, respectively. Strain E68 was the source of a different K88ab gene subtype, designated K88ab2, carried on plasmid pE68. *E. coli* HB101 (3) and plasmid pAT153 (19) were used as host and vector, respectively, for cloning.

Plasmid preparation. Plasmid was prepared by the method of Clewell and Helinski (4), using chloramphenicol amplification, and was purified by centrifugation in CsCl-ethidium bromide gradients.

Cloning of K88 genes. Plasmids pD520, pA1, p56/190, and pE68 were each digested to completion with *Hin*dIII. The resulting fragments were inserted into the *Hin*dIII site of pAT153 and transformed into HB101 by the method of Mandel and Higa (13). The resulting clones were analyzed for K88 production, using the hemagglutination assay (10). Plasmids carrying the K88ab1, K88ac, and K88ad genes, on an 11.4-kilobase fragment in each case, were reduced in size by deletion of a nonessential 5.3-kilobase fragment. The K88ab2 gene was carried on a 10.2-kilobase *Hin*dIII fragment, and after insertion into pAT153, the resulting plasmid was reduced by deletion of a 4.1-kilobase *Eco*RI fragment.

DNA sequence analysis. The sequencing strategy used is outlined in Fig. 1. The N-terminal portions of each of the genes were sequenced by the chemical cleavage technique of Maxam and Gilbert (15). Restriction fragments were labeled at either the *Eco*RI or the *Bst*EII end with T4 kinase and [γ-^{32}P]ATP before being subjected to chemical cleavage as described before (15). The internal *Hin*dII site in K88ad was labeled with terminal transferase and [α-^{32}P]-cordycepin-5'-triphosphate (18). The remainder of each gene was sequenced by the dideoxy method, using [α-^{35}S]dATP as described by Biggin et al. (2). *Eco*RI-*Hin*dII fragments were inserted into M13mp10 and M13mp11 and sequence information was obtained from both ends (Fig. 1), using the M13 15-base primer (Bethesda Research Laboratories, Cambridge, U.K.). The 265-base pair *Eco*RI fragment from K88ad was inserted into M13mp8 in both orientations and sequenced from both ends. The M13mp10 clones carrying the N-terminal portions of each gene were sequenced by using an internal primer, 5'-GTGGATATCAAGGGGT-3', prepared by the Glaxo oligonucleotide synthesis group. This sequence was selected on the basis of the K88ab sequence of

* Corresponding author.

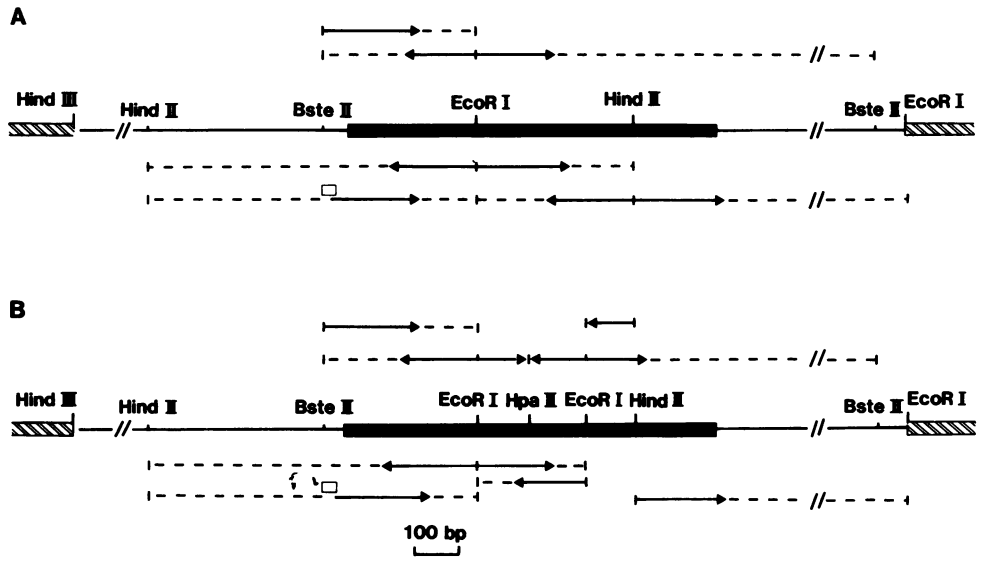

FIG. 1. Sequencing strategy for K88ab1, K88ab2, and K88ac genes (A) and the K88ad gene (B). Solid arrows indicate the direction and extent of sequencing along a given fragment. Broken arrows represent the remaining unsequenced portion of each fragment. Maxam-Gilbert (upper arrows) and dideoxy (lower arrows) methods were used. Hatched areas represent pAT153 DNA. The position of the specific internal primer used is indicated by the open box. bp, Base pairs.

Gaastra et al. (7) such that the primer would hybridize to a region just upstream of the initiator ATG.

Prediction of antigenic determinants. Hydrophilicity profiles for each of the K88 variants were derived by the method of Hopp and Woods (9). Hydrophilicity values were assigned to each amino acid residue, and these values were then averaged, six at a time starting at the amino-terminal residue, along the sequence. Hydrophilic regions are indicated by peaks on a plot of the averaged values against sequence position. Since these regions are likely to be on the surface of the protein, they may be regarded as potential antigenic determinants.

RESULTS

Nucleotide sequence determination. The nucleotide sequences of the K88ab1, K88ac, and K88ad genes and the primary structures of the corresponding proteins are given in Fig. 2. Numbering refers to the K88ab amino acid sequence. The K88ab1 sequence presented here is very similar to that previously reported by Gaastra et al. for K88ab (7). There were three changes in the coding region, none of which caused a change in the amino acid sequence. There were also two differences in the 5′ untranslated region such that the specific primer used for dideoxy sequencing of the amino-terminal regions of the four K88 genes contained a mismatch relative to the template strand. This was detected only when these sequences were redetermined by the Maxam-Gilbert technique and did not significantly affect the priming reaction. The sequence of the K88ab2 gene was similar to that of K88ab1, there being only three base changes at the positions indicated by asterisks in Fig. 2. These differences, GTG to GCG (*1), ACT to AAT (*2), and GCT to TCT (*3), resulted in valine to alanine, threonine to asparagine, and alanine to serine changes, respectively, in the K88ab2 protein. There were 19 base changes between K88ab1 and K88ac, resulting in 17 amino acid changes. There were also three additional bases in K88ac relative to K88ab1, adding an extra lysine residue at position 105, and a region of nine bases deleted, indicated by dashed line in Fig. 2, resulting in the removal of amino acid residues 165 to 167 (threonine, valine, and serine in K88ab1). There were 37 base changes between K88ab1 and K88ad, resulting in 32 alterations in the K88ad amino acid sequence but no insertions or deletions relative to K88ab1.

Nine of the 17 amino acid changes in K88ac (relative to K88ab1) are also found in K88ad. However, there are 24 other amino acid differences between the ac and ad variants in addition to the insertion and deletions previously mentioned. There are 21 nucleotide sequence differences between the K88ad sequence reported here and that of Gaastra et al. (6), resulting in 11 amino acid changes. The previous authors do not name the source of their K88ad gene but it is likely that the differences observed are due to the use of different K88ad+ *E. coli* strains.

Prediction of antigenic determinants. We have applied the hydrophilicity analysis method of Hopp and Woods (9) to the sequences presented above. It is apparent that all four profiles are very similar (Fig. 3) and closely resemble the previously published results for K88ab and K88ad. Klemm and Mikkelsen (12) previously identified the regions 19 to 24, 63 to 68, 87 to 92, 96 to 104, 114 to 120, 140 to 145, 151 to 156, 186 to 191, and 213 to 219 as potential antigenic determinants in K88ab, using a combination of hydrophilicity analysis and secondary structure prediction. Gaastra et al. (6) found that the hydrophilicity profiles of K88ab and K88ad were almost identical despite the 34 amino acid sequence differences between the two proteins, the only major difference being that in K88ad the region bounded by residues 213 to 219 was less hydrophilic than the corresponding region of K88ab. Although our results are very similar, we find that K88ac and K88ad both show diminished hydrophilicity in the region corresponding to 151 to 156 in K88ab as well as in the 213 to 219 region. Furthermore, we observe a new peak in the 165 to 174 region of K88ad and a slightly smaller peak in the corresponding region of K88ac. On the basis of hydrophilicity analysis alone, we would also include the 177 to 182 region as a potential antigenic determinant in all four proteins.

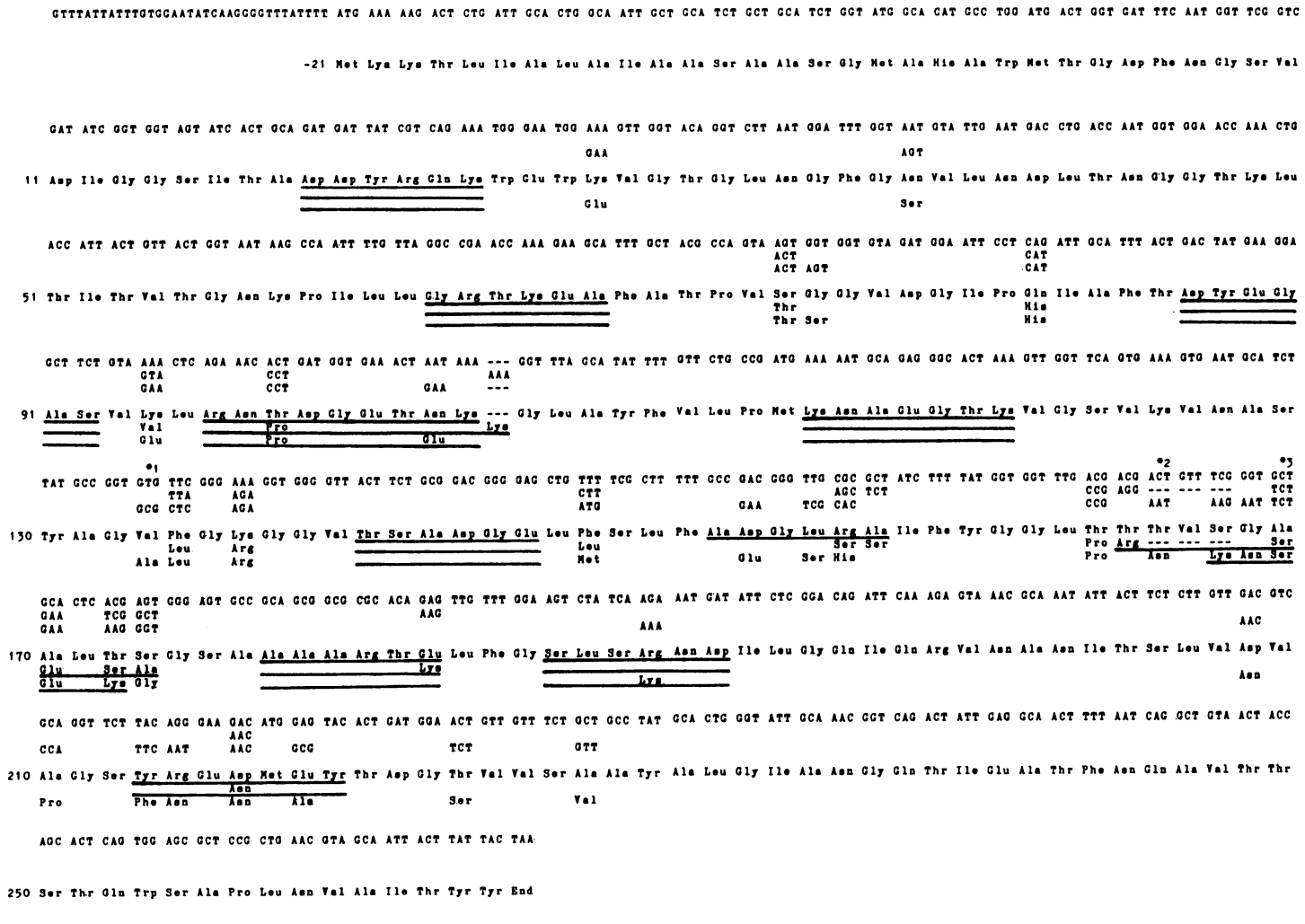

FIG. 2. Nucleotide sequences and inferred amino acid sequences of the K88ab1, K88ac, and K88ad genes. The K88ac and K88ad sequences were the same as that of K88ab except where indicated. Numbering refers to the K88ab1 protein sequence. Asterisks indicate the positions of the three differences between the K88ab1 and K88ab2 genes. In these positions the K88ab2 sequence is the same as that of K88ad. Potential antigenic determinants are underlined.

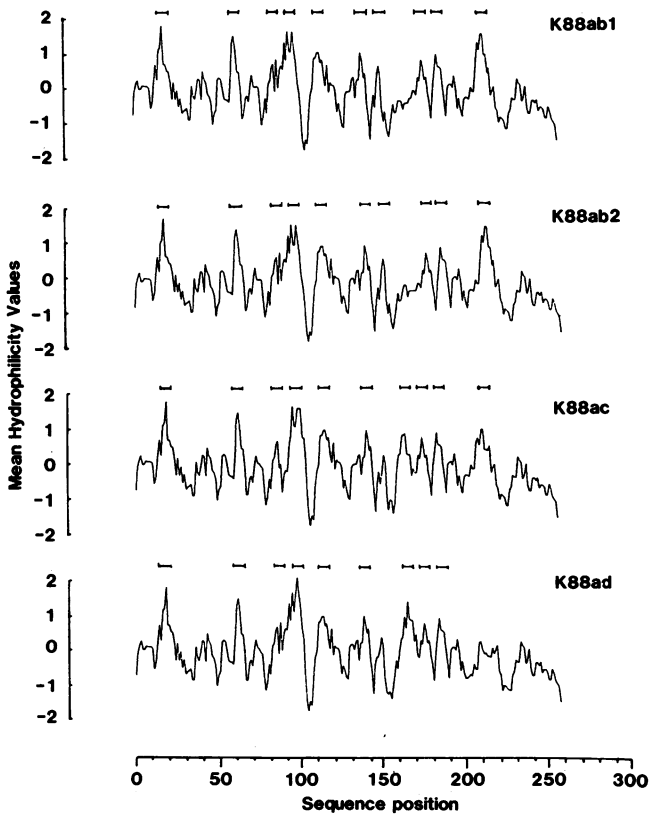

FIG. 3. Hydrophilicity profiles for the four K88 variants. Peaks corresponding to potential antigenic determinants are indicated by bars.

DISCUSSION

All of the sequence differences between the four K88 proteins occur between amino acid residues 28 and 227 in the mature proteins (K88ab numbering). Some clustering of amino acid changes is observed particularly in regions 133 to 136, 152 to 156, 163 to 173, 208 to 218, and, to a lesser extent, 74 to 82 and 94 to 105. These six clusters contain all 3 amino acid sequence differences between the K88ab variants, all but 2 of the differences between K88ab1 and K88ac, and 26 of the 32 differences between K88ab1 and K88ad.

Prediction of an antigenic determinant corresponding to the highest peak in a hydrophilicity profile has a success rate approaching 100% but predictions based on lower peaks is less certain (9). In addition, this method cannot be used to predict discontinuous, or "conformational," antigenic determinants. Nevertheless, the finding that four of the six clusters of amino acid changes coincide with regions which would be expected to be antigenic on the basis of hydrophilicity analysis is a further indication that these regions may in fact contribute significantly to the antigenic properties of the K88 variants.

By comparing the amino acid sequence of K88ab with a partial sequence for K88ad, Klemm and Mikkelsen (12) suggested that the 213 to 219 region might correspond partly or completely to the b factor and that the 19 to 24 and 114 to 120 predicted determinants might contribute to the a factor. The results of Gaastra et al. (6) indicated that regions 87 to 92 and 140 to 145 might also contribute to the a factor and that the difference between the b and d factors was probably due to differences in the 96 to 104 and 151 to 156 regions as well as the 213 to 219 region.

Our results agree with the predictions relating to the a factor since the four variants studied here are also unchanged in the regions corresponding to the 19 to 24, 87 to 92, 114 to 120, and 140 to 145 determinants. Although the predicted 63 to 68 determinant is also the same in the four sequences presented here, the K88ad sequence of Gaastra et al. (6) contains one amino acid change in this region. The region between residues 177 and 182 is the same in the K88ab and K88ad proteins but has a glutamic acid-to-lysine mutation in K88ac. Similarly, the region between residues 186 and 191 is identical in the K88ab and K88ac proteins but contains an arginine-to-lysine mutation in K88ad. These regions may thus contribute to the b, c, and d factors in the respective proteins. However, it may be unwise to attach too much significance to such small changes. More striking changes are evident in the predicted determinants at positions 96 to 104, 151 to 156, 165 to 174, and 213 to 219. Thus, although our results are again consistent with the earlier predictions as to the nature of the b and d determinants, we believe that the 165 to 174 region is also important. These results suggest that the b factor is specified largely by the potential determinant in the 213 to 219 region of K88ab, a region which has much less antigenic potential in K88ac and K88ad, and that the major components of the c and d factors are specified by the new hydrophilic region at 165 to 174 (K88ab1 numbering).

It has been proposed that the serological variation of the K88 antigen is the result of selective pressures imposed on the bacterium by the widespread use of K88-containing vaccines or possibly by the occurrence of pigs resistant to a particular serotype because of altered receptors on the intestinal epithelial cells (for review, see reference 5). The former theory is consistent with the finding that most K88-positive *E. coli* strains isolated from infected piglets now possess either the K88ac variant or the more recent K88ad type, whereas the "older" K88ab type is being found less frequently (5). That there has been an intense selection of some sort is further indicated by our finding that every nucleotide sequence difference observed here is associated with an amino acid sequence change. There are no silent mutations. Our data indicate that there is not a straightforward progression from ab to ac to ad since, although there are a number of positions where both K88ac and K88ad show the same amino acid change relative to K88ab1, there are also many instances where the K88ab1 and K88ad sequences are identical but different from that of K88ac (e.g., the insertion and deletions). K88ab2 appears to be intermediate to K88ab1 and K88ad since the three amino acid changes found in the ab2 variant relative to ab1 are also found in the ad variant. However, the precise evolutionary relationship between the K88 variants remains to be determined.

ACKNOWLEDGMENTS

We thank D. C. Humber, B. A. Coomber, and C. J. Hazelhurst for providing the K88-specific sequencing primer and J. P. R. Herrmann and M. V. Hayes for helpful discussions and critical reading of the manuscript.

LITERATURE CITED

1. **Acres, S. D., R. E. Isaacson, L. A. Babiuk, and R. A. Kapitany.** 1979. Immunization of calves aginst enterotoxigenic colibacillus by vaccinating dams with purified K99 antigen and whole-cell bacterins. Infect. Immun. **25**:121–126.
2. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and 35S label as an aid to DNA sequence deter-

mination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
3. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
4. **Clewell, D. B., and D. R. Helinski.** 1970. Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. Biochemistry **9**:4428–4440.
5. **Gaastra, W., and F. K. de Graaf.** 1982. Host-specific fimbrial adhesins of noninvasive enterotoxigenic *Escherichia coli* strains. Microbiol. Rev. **46**:129–161.
6. **Gaastra, W., P. Klemm, and F. K. de Graaf.** 1983. The nucleotide sequence of the K88ad protein subunit of porcine enterotoxigenic *Escherichia coli*. FEMS Microbiol. Lett. **18**:177–183.
7. **Gaastra, W., F. R. Mooi, A. R. Stuitje, and F. K. de Graaf.** 1981. The nucleotide sequence of the gene encoding the K88ab protein subunit of porcine enterotoxigenic *Escherichia coli*. FEMS Microbiol. Lett. **12**:41–46.
8. **Guinee, P. A. M., and W. H. Jansen.** 1979. Behavior of *Escherichia coli* K antigens K88ab, K88ac, and K88ad in immunoelectrophoresis, double diffusion, and hemagglutination. Infect. Immun. **23**:700–705.
9. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
10. **Jones, G. W., and J. M. Rutter.** 1974. The association of K88 antigen with hemagglutinating activity in porcine strains of *Escherichia coli*. J. Gen. Microbiol. **84**:135–144.
11. **Klemm, P.** 1981. The complete amino-acid sequence of the K88 antigen, a fimbrial protein from *Escherichia coli*. Eur. J. Biochem. **117**:617–627.
12. **Klemm, P., and L. Mikkelsen.** 1982. Prediction of antigenic determinants and secondary structures of the K88 and CFA1 fimbrial proteins from enterotoxigenic *Escherichia coli*. Infect. Immun. **38**:41–45.
13. **Mandel, M., and A. Higa.** 1970. Calcium dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
14. **Marx, J. L.** 1980. Vaccinating with bacterial pili. Science **209**:1103–1106.
15. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
16. **Nagy, B., H. W. Moon, R. E. Isaacson, C. C. To, and C. C. Brinton.** 1978. Immunization of suckling pigs against enteric enterotoxigenic *Escherichia coli* infection by vaccinating pigs with purified pili. Infect. Immun. **21**:269–274.
17. **Rutter, J. M., and G. W. Jones.** 1973. Protection against enteric disease caused by *Escherichia coli*-a model for vaccination with a virulence determinant. Nature (London) **242**:531–532.
18. **Tu, C.-P. D., and S. N. Cohen.** 1980. 3′-End labelling of DNA with [α-32P]cordycepin-5′-triphosphate. Gene **10**:177–183.
19. **Twigg, A. J., and D. Sherrat.** 1980. Trans-complementable copy-number mutants of ColE1. Nature (London) **283**:216–218.