A physical map of the genome of *Haemophilus influenzae* type b

PETER D. BUTLER* and E. RICHARD MOXON

Molecular Infectious Diseases Group, Department of Paediatrics, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

(Received 2 July 1990; revised 16 August 1990; accepted 24 August 1990)

Contour-clamped homogeneous electric field pulsed-field gel electrophoresis (PFGE) was used in combination with Southern hybridization to construct a genomic restriction map for the pathogen *Haemophilus influenzae* serotype b, strain Eagan. Sites for four restriction endonucleases, *EagI*, *NaeI*, *RsrII* and *SmaI*, were located on the 2100 kbp circular chromosome. Twelve potential virulence loci have been placed on the map together with certain loci essential for growth of the bacteria (e.g. ribosomal RNA operons). PFGE also provided a valuable tool for characterizing ten capsulated, type b isolates (other than Eagan) known to be genetically heterogeneous and two laboratory-derived variants (transformants) derived through complex recombinational events involving random uptake of high-molecular-mass donor genomic DNA.

Introduction

Pulsed-field gel electrophoresis (PFGE) has facilitated the making of physical maps of the chromosomes of bacterial pathogens. Currently physical maps have been published for *Clostridium perfringens* (Canard & Cole, 1989), *Pseudomonas aeruginosa* (Römling et al., 1989) and two species of the genus *Haemophilus*, *H. influenzae* strain Rd (Lee et al., 1989; Kauc et al., 1989) and *H. parainfluenzae* (Kauc & Goodgal, 1989).

In the case of *Haemophilus*, the available physical maps concern relatively non-pathogenic strains which have been sub-cultured over many years. Encapsulated *H. influenzae* type b strains are a major cause of meningitis and other life-threatening infections of childhood. In this paper we report the genomic map of a representative virulent *H. influenzae* b, strain Eagan, which was isolated from the cerebrospinal fluid of a child with meningitis (Anderson et al., 1972) and indicate the positions of several of its virulence loci.

Strain Eagan was chosen because it has been well characterized in the laboratory and has been used extensively in virulence studies (Moxon & Kroll, 1990). However, the population structure of naturally occurring isolates of capsulate *H. influenzae* is characterized by substantial diversity and comprises many genetically distinct families of strains. Each family comprises several closely related clones and, for any particular family, the clones express the identical capsular polysaccharide. Thus strains expressing any one of the six known capsular polysaccharides are represented by one or more distinct clonal clusters (families) unique to that serotype (Musser et al., 1988b). To compare type b clones from genetically distinct clusters, including those belonging to different phylogenetic lineages, we have used PFGE to examine the restriction fragment length polymorphisms of type b strains representative of the different clonal clusters to index similarities or differences in genotype.

Similarly, PFGE proved useful in the further characterization of laboratory-derived type b strains which were constructed by the transformation of a capsular-deficient strain of type d ancestry using donor DNA from strain Eagan.

Methods

Growth of bacteria. Liquid cultures of *H. influenzae* were grown in brain heart infusion (BHI) (Oxoid) broth supplemented with NAD (Boehringer Mannheim) (2 μg ml⁻¹) and haemin (Sigma) (10 μg ml⁻¹) and designated sBHI. Bacteria were also grown on solid media which consisted of BHI agar (BHI plus 1%, w/v, agar) containing 10% (v/v) Levinthal base (Alexander, 1965).

Preparation of DNA. High-molecular-mass DNA was prepared from strain Eagan using phenol as described by Moxon et al. (1984); the procedure was scaled up to work from a 400 ml culture.

To prepare chromosomal DNA in agarose blocks, BHI broth (80 ml)...
was inoculated from a plate and growth followed to OD490 0-4 (approximately 2 x 10^8 colony-forming units ml^-1). Chloramphenicol was added at 20 μg ml^-1 and incubation was continued for 1 h. The culture was chilled on ice for 20 min, then bacteria were spun down at 3000 g at 0 °C in a Beckman JA-20 rotor and resuspended in 20 ml (0-25 vol.) of buffer 21 (Herriott et al., 1970) chilled to 0 °C.

The cells were pelleted again as described above and, finally, resuspended in 4 ml (0-05 vol.) of cold buffer 21. The cell suspension was warmed by standing for 5 min at room temperature. After a further 5 min at 37 °C the cell suspension was mixed thoroughly with 4 ml (equal volume) of 1-5% (w/v) low-melting-temperature (LMT) agarose (Bethesda Research Laboratories) in buffer 21 at 37 °C.

The bacteria/agarose mixture was maintained at 37 °C whilst volumes of 215 μl were micro-pipetted into a perspex plug mould. Each space for a plug was 15 x 5 x 3 mm. Plugs were allowed to solidify in moulds at room temperature for about 10 min then hardened at 4 °C for 15 min before being removed. Batches of plugs were washed in 20 ml volumes of NDS buffer (0-5 M-EDTA, 10 mm-Tris pH 9-5, 1%, w/v, sodium laurylsarcosine) (Schwartz & Cantor, 1984) containing Proteinase K (Sigma) at 1 mg ml^-1. After 1 h at room temperature, three washes of approximately 2 h, 17 h and 24 h were performed at 50 °C. Plugs were rinsed with NDS followed by five 2 h washes at 50 °C to remove Proteinase K. This procedure solubilizes everything but the DNA; the latter remains trapped in the agarose. Plugs containing DNA were then stored in NDS at 4 °C for up to 18 months. About 30 plugs were obtained from each 80 ml culture, each plug containing about 9 μg DNA.

This method was also scaled down to process 5 ml cultures, each of which gave two plugs.

**Digestion of DNA in agarose blocks.** Usually digests were carried out on the DNA contained in one-third of a complete plug. Restriction enzyme buffers were diffused into the agarose blocks as outlined below. Plugs or portions of plugs were washed in Eppendorf tubes with 500 μl volumes of buffer (unless stated otherwise). Fresh buffer was used for each wash. Two 30 min washes in TC buffer (10 mm-Tris/HCl, 1 mm-cyclohexanediaminetetra-acetic acid pH 7-5) containing 2 mm-PMSE (phenylmethylsulphonyl fluoride) at 0 °C were followed by a wash in TC at 0 °C for 30 min. Plugs were then transferred to restriction endonuclease buffers (made according to the enzyme manufacturers’ formulations) and incubated for 1 h at 0 °C and then for a further 1 h at room temperature with fresh buffer. Plugs were transferred to new Eppendorf tubes holding 75 μl restriction buffer containing 2-mercaptoethanol (10 mm) and stood at room temperature for 20 min then incubated at 65 °C for 15 min to melt the agarose. Plugs were cooled to 37 °C for 5 min before the addition of the restriction enzyme and incubated at a temperature suitable for optimum enzyme activity. The agarose remained molten at 37 °C but not at room temperature (about 20°C). More enzyme was added over the course of about 30 h so that at least 20 U of enzyme would have been added to each reaction. For Smal, as digests are incubated at 20 °C, it was necessary to melt 65 °C then cool them to 37 °C when more enzyme was added to allow good mixing of the enzyme into the digest (Smal was usually used at about 50 U. in total, per digest because of loss of any remaining activity on repeated melting). Digests were heated at 65 °C before storage at 4 °C or immediately prior to loading on a gel.

**Contour-clamped homogeneous electric field (CHEF) electrophoresis.** Samples for electrophoresis (digests) were melted at 65 °C for 10 min then micro-pipetted into the wells of the gel. λ oligomers were made by incubating λ DNA (New England Biolabs), 2 mg ml^-1 in 2 x SSC, 3%, w/v, Ficol, 0.6 mg Orange G ml^-1 for 30 min at 37 °C then standing for 6-16 h at room temperature (SSC is 0.15 M-NaCl, 0.015 M-trisodium citrate, pH 7.0). It is essential that the λ DNA has never been frozen to achieve good oligomerization. After loading, λ oligomers were overlaid with about 20 μl 1% (w/v) LMT agarose in 0-5 x TBE (TBE is 0-89 m-Tris, 0-89 m-boric acid, 0-02 m-EDTA). Gels (15 x 15 cm) were made from 1% or 1-5% (w/v) high-melting-temperature (HMT) agarose (Sigma) or for preparative purposes 1-5% (w/v) LMT agarose (Bethesda Research Laboratories) in 0-5 x TBE. Half-concentration TBE was also used as the electrode buffer. Typically gels were run in a CHEF electrophoresis apparatus (Chu et al., 1986) with sets of electrodes separated by 27 cm with a potential difference of 225 V for 40-45 h at a temperature of 6 °C. Pulse times of 20 or 30 s proved to be optimal for the size-range of fragments being separated.

After electrophoresis, gels were stained with ethidium bromide, photographed, and the DNA transferred to a nylon filter (Hybond N, Amersham) essentially as described in the Amersham ‘Membrane Transfer and Detection Methods’ booklet. Ultra-violet irradiation was used to attach DNA to filters.

**Preparation of radioactive probes and hybridization.** Nick-translation was the preferred method when purified plasmid DNA was available (Rigby et al., 1977). Plasmid DNA (500 ng to 1 μg) was used with an Amersham kit with [α-32P]dCTP in quarter-scale reactions (25 μl total volume).

High-molecular-mass DNA from strain Eagan (3 μg) was also used in this type of reaction to make probes to detect all bands on filters with Eagan DNA on them.

Gel-purified fragments were also a source of DNA for nick-translation. Large fragments, < 30 kbp, were phenol-extracted from ethidium-bromide-stained bands cut from LMT agarose CHEF gels. Agarose fragments containing DNA were melted in 200 μl TC buffer at 65 °C for 10 min, then extracted twice with phenol saturated with TC (250 μl). After two ethanol precipitations the DNA was sufficiently pure for use in a nick-translation reaction.

Small DNA fragments, < 20 kbp, from conventional agarose gels, were removed from pieces of HMT agarose using an IBI (International Biotechnologies Inc.) electrophoretion apparatus as described by the manufacturer. A random-primed DNA labelling kit (United States Biochemical Corporation) was then used with [α-32P]dCTP to synthesize radioactive DNA (Feinberg & Vogelstein, 1983, 1984).

Prehybridization and hybridization were carried out for 4 h at 37 °C and 20 h at 42 °C, respectively, as described by Scott et al. (1979). Filters were then washed for three 20 min periods in 2 × SSC, 0.1% SDS at room temperature followed by four 30 min washes in 0 × SSC, 0.1% SDS at 37 °C and four 5 min washes in 1-0 × SSC at room temperature. After autoradiography, filters were stripped of radioactivity for re-use as described in the Amersham booklet, except that an extra wash in 0-4 M-NaOH was included.

**Results and Discussion**

**Selection of restriction enzymes**

H. influenzae DNA has a base composition of 37 mol% G + C (Roy & Smith, 1973); and therefore restriction enzymes with recognition sequences containing only C and G cut the DNA less frequently than enzymes whose sites contain all four bases. It is therefore possible to calculate the theoretical frequency of endonuclease sites of any particular nucleotide sequence (Upholt, 1977). Accordingly enzymes recognizing six-base-pair sequences comprising only G–C base pairs were used to digest H. influenzae DNA isolated in agarose blocks. Four such enzymes were selected which gave a manageable number of fragments (less than 25) whose size was
larger than 20 kbp. Initially small fragments, less than 20 kbp in size, were not considered. Smal (5'CCC'GGG3'), EagI (C'GGCCG) and NaeI (GCC'GGC) gave reproducible results, each generating 15–23 large fragments. ApaI (GGGCC'C) gave inconsistent results and was not used. Two other enzymes with recognition sequences comprising only G-C base pairs are SacII (CCGC'GG) and NarI (GG'CGCC); these were not useful for our mapping project. SacII generated too many fragments, more than 40, and NarI apparently did not cleave H. influenzae DNA. This was demonstrated using double digests with either Smal or RsrII; the restriction patterns obtained in the presence of NarI were identical to those obtained when Smal or RsrII were used alone.

Other enzymes are commercially available which cleave recognition sites comprising more than 6 base pairs rich in G and C. RsrII (CG'GCGGCCG) gave only four fragments; SfiI (GGCCN₂NGGCC) did not cleave the DNA; NotI (GC'GGCCGC) made a single cut detected by using double/sequential digests with Smal or RsrII. This single NotI site was of value in construction of the genomic map because the recognition sequence of NotI contains an EagI recognition site (C'GGCCG) and this could be located on each of the maps constructed with the different endonucleases. Thus, it was possible to align the different endonuclease maps around the NotI site. In contrast, double digests involving EagI, NaeI or Smal gave too many fragments to be of use in lining up the maps relative to one another.

Analysis of digests

Fractionation of digests was achieved using CHEF gel electrophoresis (Chu et al., 1986). Running conditions and loads were such that fragments larger than about 20 kbp could be detected with optimal resolution. Fig. 1 (a, b) shows typical EagI, NaeI, Smal and RsrII digests,
and the fragments produced with each of the four enzymes are shown in Table 1. Of these enzymes, SmaI proved to be most reproducible. Two of the bands in the SmaI digest lane (Fig. 1a, lane C), at about 100 kbp and 120 kbp, stained quite bright with ethidium bromide and suggested the presence of more than one band. In these examples, and in other cases, information gained from radioactive probing experiments was used to assess the number of DNA species in a band.

A recurring problem was the incomplete digestion of DNA. (For example, several faint bands due to this can be seen in the upper part of the NaeI track in Fig. 1a, lane B.) Such bands would be subject to variation between gels and they were attributed to partial digestion. However, a large EagI fragment at about 325 kbp (the uppermost band in Fig. 1a, lane A) was consistently observed. When DNA from this band was used to generate a radioactive probe, it hybridized preferentially to several smaller bands but only to a small extent to the original (325 kbp) band. Two bands, also in the EagI lane, were often visible at about 35 and 40 kbp. They were not always present and were never detected in any radioactive probing experiment.

Routinely all filters were hybridized with radioactive DNA made from strain Eagan by nick-translation. The pattern obtained was in all cases the same as that obtained with ethidium bromide staining of the original gel.

High-molecular-mass DNA, prepared by phenol extraction, was digested with EagI or NaeI to generate small fragments (<20 kbp). These fragments were fractionated by electrophoresis through 1% HMT agarose gels run using normal (non-CHEF) conditions. After sizing these fragments, the DNA was extracted and used to prepare radioactive probes which were hybridized to CHEF gel filters containing endonuclease-digested genomic DNA from strain Eagan. This gave the sizes of these fragments and the number of types of DNA species in each band could be determined. The smallest fragment seen in SmaI digests was 36 kbp.

A digest obtained with RsrII is shown in Fig. 1(b). Although the size of the largest RsrII fragment was not measured directly from a CHEF gel, the size of the chromosome determined by summing the fragment sizes (Table 1) obtained with each of the other three enzymes was in very close agreement. Double digests of strain Eagan DNA made using RsrII and EagI or SmaI showed that there were four RsrII sites in the chromosome. When DNA from the largest RsrII fragment was used to prepare a radioactive probe the sum of the sizes of the SmaI fragments to which it hybridized was about 1400 kbp. This figure probably overestimates the size of the large RsrII fragment since the two SmaI fragments which contain RsrII sites also contain DNA from the

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Fragment</th>
<th>Size (kbp)</th>
<th>Multiplicity</th>
<th>Total size (kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EagI</td>
<td>A</td>
<td>305</td>
<td>2</td>
<td>305</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>215</td>
<td>1</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>158</td>
<td>2</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>120</td>
<td>2</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>105</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>96</td>
<td>2</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>90</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>80</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>70</td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>62</td>
<td>2</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>54</td>
<td>2</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>41-5</td>
<td>1</td>
<td>41-5</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>8-5</td>
<td>2</td>
<td>8-5</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>7-6</td>
<td>1</td>
<td>7-6</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>3-6</td>
<td>4</td>
<td>3-6</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>3-3</td>
<td>4</td>
<td>3-3</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>5-7</td>
<td>1</td>
<td>5-7</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>4-2</td>
<td>2</td>
<td>4-2</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>4-0</td>
<td>5</td>
<td>4-0</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>3-3</td>
<td>2</td>
<td>3-3</td>
</tr>
</tbody>
</table>

| RsrII  | A        | 1225      | 1            | 1225 (36)       |
|        | B        | 365       | 7            | 365             |
|        | C        | 290       | 6            | 290             |
|        | D        | 230       | 5            | 230             |

| SmaI   | A        | 330       | 3            | 330             |
|        | B        | 310       | 2            | 310             |
|        | C        | 270       | 2            | 270             |
|        | D        | 210       | 4            | 210             |
|        | E        | 132       | 2            | 132             |
|        | F        | 120       | 3            | 120             |
|        | G        | 100       | 3            | 100             |
|        | H        | 80        | 2            | 80              |
|        | I        | 70        | 2            | 70              |
|        | J        | 48        | 3            | 48              |
|        | K        | 36        | 1            | 36              |

* The size of the largest RsrII fragment was determined by subtraction using the average genome size obtained using the other three enzymes (†).
adjacent Rsrl fragments. Therefore, the size of the largest Rsrl fragment determined by subtraction was used. The data in Table 1 were used to generate an idealized picture of a CHEF gel showing all restriction fragments (Fig. 2).

The size of the *H. influenzae* strain Eagan chromosome, obtained with three restriction enzymes, was 2110 ± 15 kbp. This is larger than the *H. influenzae* Rd genome, 1834–1980 kbp (Kauc et al., 1989; Lee & Smith, 1988; Lee et al., 1989). The genomes of strains Eagan and Rd are relatively closely related based on metabolic enzyme typing (Musser et al., 1988a). However, comparison of the restriction fragment length polymorphisms obtained by digesting DNA from Eagan and Rd using Smal were dissimilar (compare Fig. 1a, lane C to Fig. 5, lane C). Because strain Rd has been in laboratory use for many decades and has undergone serial passaging, we investigated the genomes of an infrequently passaged serotype d strain, RM7033, isolated from a patient with pneumonia (Kroll et al., 1989) and a nontypable strain derived from it (RM7051) using Smal to digest the DNA (data not shown). These strains gave similar banding patterns to strain Rd, corresponding to a chromosome size of about 1955 kbp. However, there was a clear difference between the capsulate and non-capsulate strains. A band of 209 kbp in RM7033 decreased to 192 kbp in RM7051, indicating that RM7033 had sustained a deletion of about 17 kbp giving rise to the capsule-deficient strain RM7051.

**Ordering of fragments**

To confirm that the chromosome of *H. influenzae* is circular and to establish a framework on which to build the remainder of the map, the relative positions of the four large Rsrl fragments were determined. Smal fragments were isolated from LMT gels and used to probe filters to which Rsrl-digested DNA had been bound. Four Smal fragments, each of which hybridized to two Rsrl fragments, were found. However, the identification of these linking fragments was not as straightforward as was first hoped since two of the Smal linking fragments were present in Smal bands which contained more than one species (S100 a, b, c and S120a, b) (see Table 1). Most of the information about the order of fragments was gained by the use of radioactive probes made from DNA prepared from LMT agarose. Bands thought to be doublets or triplets were used during the later stages of constructing the map to confirm the putative positions of these fragments. Bands of unknown size were identified by superimposing two autoradiographs, one obtained by probing endonuclease-digested genomic DNA with itself, the other obtained using a specific probe (e.g. pU038 from the type b capsule locus). A calibration curve was constructed to determine the molecular mass of unknown bands.

Radioactive probes (listed in Table 2) of cloned *Haemophilus* DNA provided further information about the order of fragments and their map positions as shown in Fig. 3. The use of these probes served to confirm and expand details of the map and also indicated the relative positions of several virulence loci known or suspected to be important in pathogenicity.

**Ribosomal RNA operons and small restriction fragments**

Ribosomal RNA operons are the only repeated DNA sequences in most bacterial chromosomes and can give rise to various kinds of rearrangements (Smith et al., 1987). From their positions the putative location of the origin of replication can be inferred (Brewer, 1988;
Fig. 3. Genomic map of *H. influenzae* serotype b strain Eagan. Fragments of the chromosome produced with the restriction enzymes *EagI*, *NaeI*, *RsrII* and *SmaI* which are listed in Table 1 were arranged to form the genomic map. The circular chromosome (drawn as four restriction maps) has been linearized to aid its presentation. The long vertical lines, extending through all four individual enzyme maps, indicate where the circular molecule has been ‘cut’. Small *EagI* (L–P) and *NaeI* (O–T) fragments have been labelled by the alphabetic nomenclature system used in Table 1. Larger fragments are labelled with the first letter of the enzyme which gave rise to them followed by their size, e.g. S132. Bands which comigrate have been distinguished from one another by suffixing them with a lowercase letter, e.g. S100b. The single *NotI* site is indicated on the *RsrII* map. A dashed vertical line is drawn to indicate its position in the other maps. Above the maps the positions of the markers listed in Table 2 are shown. Ribosomal RNA operons are indicated beneath the *RsrII* map.

Nomura *et al.,* 1984). Filters of strain Eagan DNA, digested with *EagI*, *NaeI*, *RsrII* or *SmaI*, were probed with a radioactive probe made by a random priming reaction using a 3-55 kbp *SacII* fragment of *H. influenzae* Rd DNA (Lee *et al.,* 1989; kindly provided by Rosie Redfield). Only one *EagI* fragment (90 kbp) showed any homology to this probe, yet six were detected in both *NaeI* or *SmaI* digests and the possibility that there was homology of the ribosomal RNA probe to a series of small *EagI* fragments (not visible on the CHEF gel filter) was investigated. When the small *EagI* fragments O and P, 3.6 kbp and 3.3 kbp respectively, were mapped they coincided with regions which contained ribosomal RNA homology. An exception was the S36(K)–E90(Ga) region where there appears to be a ribosomal RNA operon but no small *EagI* fragment; hybridization occurred within a large *EagI* fragment. A very strong signal was obtained when the 36 kbp *SmaI* fragment was probed with the *SacII* ribosomal RNA probe. The 36 kbp *SmaI* fragment was used as a probe so that the overlap of large fragments in this region was determined. *EagI* digests were fractionated on a conventional gel which was then blotted and probed with the 3.55 kbp *SacII* ribosomal RNA probe. Strong homology was found to the 3.6 and 3.3 kbp *EagI* fragments (fragments O and P). Thus the positions of ribosomal RNA operons were narrowed down to the *EagI* O or P fragment (Fig. 3). The only other small fragment to show any homology to the *SacII* probe
was the 4.0 kbp NaeI S fragment. The EagI O fragment (between EagI E80a and EagI E62b) does not coincide with any homology found by the SacI probe. Since the O fragment probe was made from a mixture of 3.6 kbp fragments it is quite possible that one species is not ribosomal DNA.

Efficient replication of E. coli is facilitated by clustering ribosomal RNA operons around the origin of replication so that the direction of transcription follows each replication fork (Brewer, 1988). About half of the ribosomal protein genes are located in this region (Nomura et al., 1984). The gene for streptomycin resistance/sensitivity is usually part of this cluster of ribosomal protein genes, at least in those bacterial genomes which have been examined. In H. influenzae strain Eagan, the clustering of the ribosomal RNA genes, including at least one ribosomal protein involved in streptomycin susceptibility, suggest that the origin of replication may be located between ribosomal RNA operons B and C (Fig. 3). This concept has already been advanced by Lee et al. (1989) in proposing the possible position of the origin of replication of strain Rd; in addition, these authors have determined the direction of transcription of the ribosomal RNA operons.

**Location of the capsule locus in the H. influenzae chromosome**

Strain Eagan has been used extensively for investigating the role of capsule as a virulence factor. The location of the cap b locus in the genomic map was determined using an 18 kbp probe, pUO38 (Ely et al., 1986). This plasmid contains an 18 kbp BamHI fragment from cap b and has been mapped in detail using endonucleases different from those used to construct the physical map of the genome (Hoiseth et al., 1986). In order to orientate it correctly in the genome, pUO38 was digested with NaeI. Two NaeI sites were present in the BamHI fragment (Fig. 4) but further studies, including probing with the cloned 4-4 kbp EcoRI fragment from cap b showed that the 39 kbp NaeI fragment (N39c in Fig. 3) includes the 20 kbp EcoRI fragment at the left-hand end of cap b (Fig. 4).

**Transformants of H. influenzae Rd**

Transformation of H. influenzae Rd (a capsule-deficient variant derived from a type d strain) with DNA from the type b strain Eagan can produce two distinct categories of transformants (designated 01 and 02) expressing type b capsule (Zwahlen et al., 1986). These different transformants, both of which are highly virulent in experimental infections of rats, elaborate different amounts of capsular polysaccharide, that of 02 being about 50% compared to the 01 transformant. Endonuclease mapping of cap b has shown that these differences correlate with the acquisition by Rd of either two copies (01 transformant) or one copy (02 transformant) of capsule genes which are duplicated in the chromosome of strain Eagan. Therefore PFGE was used to characterize
differences in the genomes of the 01 and 02 transformants (Fig. 5).

DNA from the recipient strain and the two Rd transformants (01 and 02) (Zwahlen et al., 1983, 1985, 1986) was prepared in agarose plugs and digested with SmaI prior to fractionation by CHEF electrophoresis (Fig. 5). The recipient, strain Rd, gave a similar pattern to that observed by Lee and co-workers (Lee & Smith, 1988; Lee et al., 1989). The 01 transformant differed in that band E (sized at 187 kbp in Rd and described by Lee et al., 1989) was found to be about 195 kbp. Both band E (of strain Rd) and the novel, larger fragment in the 01 transformant hybridized to the pU038 probe. The observed increase in the relevant SmaI fragment (8 kbp) is less than predicted since the duplicated genes for type b capsule comprise approximately 35 kbp. The most reasonable explanation for the observed discrepancy is that approximately 27 kbp of DNA found in the Rd genome is effectively deleted in the corresponding SmaI fragment (containing cap b) in Eagan. The recombinational event involved in the transformation would occur by a double cross-over involving regions of homology between donor and recipient so as to introduce the novel cap b sequences (only a small remnant of which is found in Rd) while ‘looping out’ a substantial amount of DNA present in the recipient Rd genome but absent in the donor (Eagan) strand.

Changes accounting for the 02 transformant are more complex. Bands B (254 kbp) and E (187 kbp) are not present. Instead, a novel doublet of 229 kbp which has homology to the pU038 probe (probably in only one of the 229 kbp species) is present. This represents a net increase of approximately 17 kbp in the Rd genome, which is the increase that would be expected for the addition of a single set of type b capsule genes from strain Eagan DNA. Thus, exchange of DNA involving a polymorphic SmaI site would account for the observed findings. In this case, one band increases by 17 kbp (cap b) and a further 25 kbp (due to the SmaI polymorphism) and the contiguous SmaI fragment is decreased in size by 25 kbp; this results in a doublet of SmaI fragments each of 229 kbp.

Analysis of other type b strains

Enzyme electrophoresis has shown that H. influenzae strains Rd and Eagan are relatively closely related (Musser et al., 1985, 1988a). However, there are no obvious similarities in the sets of restriction fragments generated by the endonuclease SmaI. Accordingly the maps obtained with this restriction enzyme ought to be dissimilar and this was found to be the case. SmaI digests of 10 serotype b strains, selected so as to be representative of the several distinct families of clones which have been described for encapsulated H. influenzae b strains (Musser et al., 1988a), were separated by CHEF electrophoresis. Four strains (including strain Eagan)
from the A1 cluster (Fig. 6) gave similar patterns. A further three strains from the A2 cluster gave patterns which differed from those of the A1 cluster, but were similar to one another although not to the same extent as seen in the A1 cluster. Single strains from each of the B1, J1 and J3 clusters gave unique patterns quite distinct from those seen from strains from the A major lineage.

The B family of clones in lineage I is quite diverse in terms of the serotypes represented and includes strains of serotype a, b and d. Although serotype b and d organisms are found within the B1 cluster of this major lineage, no similarities were detected between SmaI digests of the type b strain and those of type d.

We wish to thank Mr Peter Goddard for constructing the CHEF gel apparatus. Thanks are also due to Dr H. O. Smith and members of his group for providing data prior to publication and for several DNA probes used in this work. We also wish to thank Mrs Patricia Hook in helping to prepare the manuscript.

This work was supported by a MRC Programme grant (PG832535).

References


Physical map of H. influenzae genome


