INFECTION AND IMMUNITY, Feb. 1990, p. 320–331
0019-9567/90/020320-12$02.00/0

Vol. 58, No. 2

Localization of the Cleavage Site Specificity Determinant of *Haemophilus influenzae* Immunoglobulin A1 Protease Genes

FRANK J. GRUNDY,[1]† ANDREW G. PLAUT,[2] AND ANDREW WRIGHT[1]*

Department of Molecular Biology and Microbiology, Tufts University Health Sciences Campus,[1] and Gastroenterology Unit, Department of Medicine, Tufts-New England Medical Center,[2] Boston, Massachusetts 02111

Received 6 June 1989/Accepted 7 November 1989

Immunoglobulin A1 (IgA1) proteases are produced by a number of different species of bacteria which cause infection at human mucosal surfaces. The sole substrate of these proteases is human IgA1. Cleavage is within the hinge region of IgA1, although there is variability in the exact peptide bond within the hinge region that is cut by a particular protease. The cleavage site of the *Haemophilus influenzae* type 1 protease is located four amino acids from the cleavage site of the type 2 enzyme. In this study, the region of the *H. influenzae* IgA1 protease gene (*iga*) that determines the cleavage site specificity was localized through the comparison of the type 1 and type 2 genes and the construction and analysis of type 1-type 2 hybrid genes. The hybrid genes were generated by in vivo and in vitro techniques which facilitated the selection and screening of randomly generated hybrids. The cleavage site determinant was found to be within a 370-base-pair region near the amino-terminal coding region, in one of two large areas of nonhomology between the two types of *H. influenzae iga* genes. DNA sequence analysis of the cleavage site determinant and surrounding regions did not reveal a simple mechanism whereby one enzyme type could be converted to the other type. Comparison of the type 2 gonococcal IgA1 protease gene to the two *Haemophilus* genes revealed a significant amount of homology around the cleavage site determinant, with the two type 2 genes showing greater homology.

Bacterial infection is dependent upon evasion of the host defense mechanisms. In order for a pathogen to become established at mucosal surfaces, it must elude immunoglobulin A (IgA), the primary antibody found at these surfaces (36). *Haemophilus*, *Neisseria*, and *Streptococcus* species, all of which can cause infection at human mucosal surfaces, secrete an enzyme which specifically cleaves human IgA1 (20, 21, 26, 29). This enzyme is likely to be an important factor in pathogenesis, since pathogenic species produce IgA1 protease while nonpathogenic species do not (27). Furthermore, the expected fragments resulting from IgA1 cleavage have been found in the relevant body fluids of infected patients (28).

All of the IgA1 proteases cleave the substrate within the hinge region; however, the exact peptide bond cleaved is strain dependent (20, 25, 26, 29). For instance, certain strains of *Haemophilus influenzae* produce a protease which cleaves at a proline-serine peptide bond (type 1 cleavage specificity), while other strains produce a protease which cleaves four amino acids away at a proline-threonine bond (type 2 specificity). The enzyme type is correlated with the serotype of the *H. influenzae* strain, such that type 1 enzymes are produced by strains of serotypes a, b, d, and f while serotypes c and e strains produce type 2 enzymes (Fig. 1; 16, 25). IgA2 is not cleaved by these proteases (30), as it lacks the portion of the hinge region in which the target site for cleavage is located (10). This may be a factor in the infectivity of IgA1 protease producers at host surfaces having a high ratio of IgA1 relative to IgA2 (19).

The IgA1 proteases are secreted from producing organisms. The isolation and characterization of several IgA1 protease genes from different bacterial sources has led to the identification of a carboxyl-terminal domain critical for the secretion of the *Neisseria gonorrhoeae* (31, 32) and *H. influenzae* (12) IgA1 proteases. The analysis of other functional domains of IgA1 proteases should yield information regarding the mechanism of cleavage of IgA1; this information will aid in the design of inhibitors of protease activity. In this study, the region of the *H. influenzae* IgA1 protease genes (designated *iga*) that determines the exact site of substrate cleavage has been localized through the analysis of type 1-type 2 hybrid *iga* genes.

MATERIALS AND METHODS

Media. L broth (24) was used for culturing *Escherichia coli*. When necessary, kanamycin sulfate or ampicillin (sodium salt, both from Sigma Chemical Co., St. Louis, Mo.) was included at a final concentration of 30 or 100 μg/ml, respectively. *H. influenzae* was routinely grown in brain heart infusion (BHI) medium (BBL Microbiology Systems, Cockeysville, Md.) supplemented with 10 μg/ml of β-NAD and hemin (both from Sigma). Heart infusion (Difco Laboratories, Detroit, Mich.) containing similar supplements was used for *H. influenzae* cultures that would subsequently be made competent in the M-IV medium of Herriott et al. (13). BHI supplemented agar (BBL) was used as a solid growth medium for *H. influenzae* strains. When appropriate, kanamycin (20 μg/ml) was included in the growth medium.

Bacterial and bacteriophage strains. *E. coli* strains used were MM294 (*hrs hrm*$^+$ *thi endA supE*) (2), MC1000 [*leu araD139* Δ(*ara-leu*)7697 Δ*lac*(X74) *galE galK rpsL*] (6), JM103 [Δ(*lac-pro*) *supE thi rpsL sbcB15 endA hspR4*(F′ *traD36 proAB lacI*^{q} ΔM15)] (22), SM525 (MC1000 *lamB560*), and Q1(F$^-$ *thr leu tonA lac supE strA*) (14). *H. influenzae* HF2.8c$^+$(c3) is a type 2 IgA1 protease producer derived from *H. influenzae* Rd (type 1 IgA1 protease producer) via homologous recombination after transformation with DNA from a serotype c, type 2 IgA1 protease-producing *H. influenzae* strain (4, 12). *H. influenzae* Rd (1) and ATCC 9007

* Corresponding author.
† Present address: Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, LA 71130.

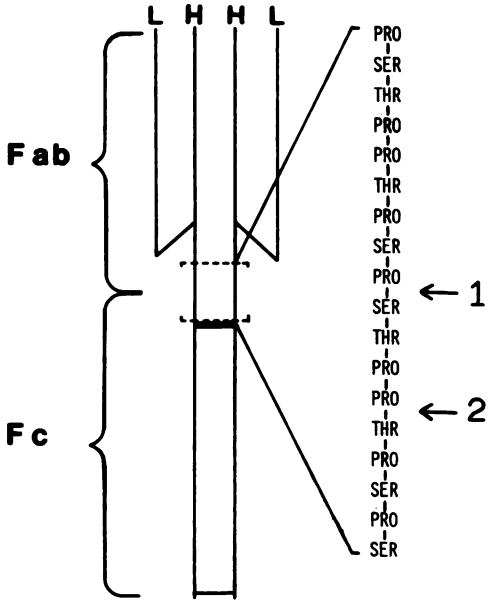

FIG. 1. Cleavage sites of IgA1 proteases. The amino acid sequence of the hinge region of human IgA1 is shown at the right. The location of the hinge region within the heavy chains (H) of IgA1 is bracketed by broken lines. Disulfide bonds between cysteine residues join the heavy chains together and light chains (L) to the heavy chains. Cleavage by IgA1 proteases results in three fragments, two Fab and one Fc. Each Fab, or antigen-binding region, consists of one light chain and the amino-terminal half (Fd) of one heavy chain. The Fc region consists of the carboxyl-terminal ends of two heavy chains. The arrows show the cleavage sites of type 1 and 2 *H. influenzae* IgA1 proteases.

(serotype c, type 2 IgA1 protease) were used in DNA hybridization experiments. Bacteriophage λ*vir* was from the laboratory collection.

In vitro DNA manipulations. All DNA restriction enzymes, synthetic DNA linkers, T4 DNA ligase, and BAL 31 exonuclease were obtained from New England BioLabs, Inc. (Beverly, Mass.). The Klenow fragment of DNA polymerase I was from Dupont, NEN Research Products (Boston, Mass.). Calf intestinal alkaline phosphatase was from Boehringer-Mannheim (Indianapolis, Ind.).

Transformation of *E. coli* strains was carried out by the calcium chloride shock method essentially by the protocol of Cohen et al. (7). The competence method of Herriott et al. (13) was used for transformation of *H. influenzae* strains with modifications as described previously (12).

Plasmids. Plasmid pJB2 is a derivative of plasmid pBR322 containing a 6.9-kilobase (kb) insert of *H. influenzae* DNA at the *Eco*RI site (5). This plasmid contains 3.9 kb of the IgA1 protease gene (*iga*) and encodes a type 1 IgA1 protease that lacks an intact carboxy terminus but is enzymatically active. Plasmids pFG25 and pFG26 are derivatives of pBR322 which encode intact type 2 and type 1 proteases, respectively (12). These plasmids have the *Cla*I-*Nde*I fragment of pBR322, including the ampicillin resistance gene, ligated to *Cla*I-*Nde*I *H. influenzae* DNA fragments. Because of the method of derivation of plasmid pFG25, the type 2 *iga* gene in this plasmid may be a partial hybrid with the type 1 gene (12).

Plasmid pJB2::Tn*5*-209DK was constructed as follows. Plasmid pJB2 was mutagenized with transposon Tn*5*; one isolate, pJB2::Tn*5*-209, was shown by restriction enzyme analysis to have the Tn*5* insertion 2.0 kb downstream from the beginning of the *iga* gene (12; see Fig. 4 for location of Tn*5* insertions). Plasmid pJB2::Tn*5*-209 was digested with the restriction enzymes *Hpa*I, which cleaves within Tn*5*, 185 base pairs (bp) from each end of the element (18), and *Xba*I, which cleaves within *iga* (5), 1.0 kb downstream from the end of the Tn*5*. The staggered ends produced by XbaI were filled in by using the Klenow fragment of DNA polymerase I, and the linear DNA containing deletion(s) was ligated in the presence of synthetic *Xho*I linkers. The DNA was then digested with *Xho*I and ligated to the 2.2-kb *Xho*I internal fragment of Tn*5* containing the kanamycin resistance (Km^r) determinant (18) to form plasmid pJB::Tn*5*-209DK.

Plasmid pJB2::Tn*5*-208DK was derived from plasmid pJB2::Tn*5*-208, which has an insertion of Tn*5* within the *iga* gene, 2.5 kb downstream from the start of the gene. Plasmid pJB2::Tn*5*-208 was cut near the ends of the Tn*5* element with *Hpa*I, digested with nuclease BAL 31, and ligated in the presence of *Xho*I linkers. One of the resulting deletion derivatives lacked the region of *iga* extending from 2.0 to 3.0 kb downstream of the start of the gene. This plasmid was digested with *Xho*I and ligated to the 2.2-kb *Xho*I fragment of Tn*5* containing the Km^r determinant to form plasmid pJB2::Tn*5*-208DK.

Plasmid pFG11, containing the first 1.5 kb of the *iga* gene, was constructed from plasmid pJB2::Tn*5*-203 by subcloning, into pBR322, the *Cla*I-*Sal*I fragment which contains the amino-terminal region of *iga* and a portion of Tn*5* including the Km^r gene. Plasmids pFG16, pFG19, and pFG20 were constructed in a similar manner, but using pJB2::Tn*5*-201, 207, and 211, respectively (see also Grundy et al. [12] for pFG16). Plasmid pFG28, which lacks the first 400 bp of *iga*, was constructed by deleting the *Cla*I-*Xba*I fragment of pFG26 that spans the amino-terminal region of *iga* and replacing it with the *Acc*I(in Tn*5*)-*Xba*I fragment of pJB2::Tn*5*-232 containing the Km^r gene of Tn*5*.

Plasmid pFG36 is a composite plasmid consisting of the amino-terminal region of the type 1 *iga* gene and the carboxyl-terminal region of the type 2 gene separated by a portion of Tn*5* (the Km^r gene) and the *E. coli lamB* gene. This plasmid was constructed in two steps. First, the *Bam*HI (in Tn*5*)-*Cla*I fragment (containing the amino-terminal region of *iga*) from pJB2::Tn*5*-201 was ligated to the large *Bam*HI-*Cla*I fragment of pFG25. The 2.7-kb *Hpa*I-*Hin*dIII fragment of plasmid pCH41 (obtained from C. Hoffman), which contains the *E. coli lamB* gene, was then inserted into the remaining portion of Tn*5*. Plasmid pFG37 is a composite plasmid consisting of the amino-terminal region of the type 2 *iga* gene and the carboxyl-terminal region of the type 1 *iga* gene separated by the *lamB* gene. It was constructed by first linking the *Nde*I-*Bgl*II fragment of pFG25, containing pBR322 sequences and the amino-terminal region of *iga* (type 2), with the *Bam*HI(in Tn*5*)-*Xba*I fragment of pJB2::Tn*5*-225 and the *Xba*I-*Nde*I fragment of pFG26 containing the carboxyl-terminal region of *iga* (type 1). The *lamB* gene from plasmid pCH41 was inserted within the Tn*5* sequences to generate plasmid pFG37.

Plasmid pFG39 has the 370-bp *Bam*HI-*Pvu*II fragment from pFG25, which contains the type 2 specificity region, inserted between the *Bam*HI and *Pvu*II sites of pBR322. Plasmid pFG40 has the *Sal*I (in Tn*5*)-*Xba*I fragment (2.95 kb) from pJB2::Tn*5*-232 inserted between the *Nhe*I and *Sal*I sites of pBR322; this fragment contains 2.7 kb of Tn*5* DNA and 250 bp from the type 1 specificity region (defined below).

Plasmid pFG45 is identical to plasmid pFG25, except for the presence of a *Bcl*I site at a position identical to that of the *Bcl*I site in pFG26; this site was introduced by using oligo-

nucleotide-directed mutagenesis, as described below. Plasmids pFG47 and pFG48 are reciprocal hybrids formed between pFG45 and pFG26 by exchange of *NdeI-BclI* fragments. Plasmids pFG37R5, pFG36R1, and pFG36R2 have hybrid *iga* genes and were constructed as described above.

IgA1 protease assays. For qualitative assays to distinguish protease-negative from protease-positive colonies on agar plates, an overlay assay with ^{125}I-labeled IgA was used (11). For determination of quantity and cleavage site specificity of IgA1 protease, cell-free samples were mixed with an equal volume of 0.05 M Tris hydrochloride (pH 7.7)–0.5% bovine serum albumin containing 0.8 μCi of ^{125}I-labeled IgA; after digestion at 37°C, the products were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis as previously described (25). The IgA1 protease type was determined by comparison of the distance of migration of the Fd fragment of IgA1 to that of bands obtained by using standard protease preparations in control digests.

DNA isolation. Small-scale plasmid isolation from *E. coli* cells was carried out by the method of Ish-Horowicz and Burke (17). Large-scale plasmid isolation was carried out by a modification of the method of Birnboim and Doly (3), followed by ethidium bromide-cesium chloride density gradient separation. Large-scale DNA isolation from *H. influenzae* strains was done as previously described (5).

DNA sequencing. DNA fragments to be sequenced were generated by cleavage at naturally occurring restriction sites in the *iga* gene, at sites within Tn*5* insertions, or at sites generated after ligation of synthetic linkers to BAL 31-treated DNA molecules. These fragments were inserted into the genomes of the M13 phages mp8, -9, -18, or -19 and were sequenced by the dideoxynucleotide chain termination procedure (34), as described by Messing (22). In most cases, the primer used was from Collaborative Research, Framingham, Mass. For sequencing at points of Tn*5* insertions, an oligonucleotide which hybridizes near the ends of Tn*5* (5′-CGTT CAGGACGCTACTT-3′; 35) was used. Sequence information was analyzed by using the Pustell DNA sequence analysis programs (International Biotechnologies, Inc.).

Site-specific mutagenesis. Oligonucleotide-directed mutagenesis was carried out by using the procedure of Inouye and Inouye (15). The mutagenic oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer. Oligonucleotides were as follows: 5′-CCTATAATGATCAAGAA AAATATC-3′ for creation of a *BclI* site in plasmid pFG25 and 5′-CATATAATGATTAGAATAAATATCC-3′ for the simultaneous destruction of the *BclI* site and the generation of a stop codon in plasmid pFG26. The oligonucleotides were hybridized to a DNA molecule containing a single-stranded gap in the region encompassing the target site; the gap was filled in by using the Klenow fragment of DNA polymerase, and the DNA was ligated before introduction into *E. coli* strains by transformation. The nature of the mutations was confirmed by DNA sequencing.

Transfer and hybridization of DNA with type-specific *iga* probes. Chromosomal DNA from *H. influenzae* strains was digested, separated by electrophoresis through agarose gels, and transferred to nitrocellulose filters. The filters were then probed with pFG39 or pFG40 DNA (^{32}P labeled by nick translation as previously described; 35). Hybridizing fragments were visualized by autoradiography.

RESULTS

Activity encoded by hybrid genes generated in vivo. Strains of *H. influenzae* produce an enzyme, IgA1 protease, encoded by the *iga* gene, that specifically cleaves human IgA1; different strains of *H. influenzae* produce variants of this protease that differ in the exact peptide bond cleavage site. To localize the region of the *H. influenzae* IgA1 protease that is responsible for substrate cleavage site specificity, the enzyme type encoded by type 1-type 2 hybrid genes was correlated with the region of *iga* contributed by each parental gene. Hybrid genes were generated in vivo in *H. influenzae* by the procedure shown in Fig. 2 (12). In this method, a pBR322-derived plasmid containing a type 1 *iga* gene with deletions in either the amino-terminal or carboxyl-terminal region is integrated into the chromosome of a *H. influenzae* type 2 IgA1 protease-producing strain via homologous recombination after transformation of competent cells. Transformants with integrated plasmids are selected for expression of the Km^r determinant of Tn*5*, which is included in

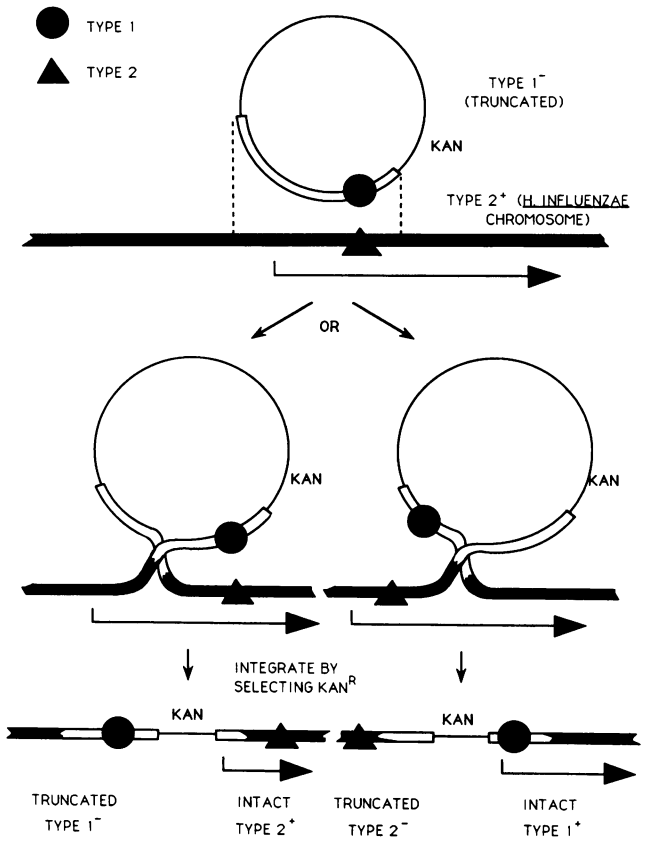

FIG. 2. Plasmid integration events which allow localization of IgA1 protease cleavage determinants. In this example, the incoming plasmid contains the portion of *iga* (▭) which encodes the amino terminus of the type 1 protease, including the region which determines type 1 specificity (●). The recipient *H. influenzae* strain encodes a type 2 protease; the location and direction of transcription of this intact *iga* gene is indicated by the horizontal arrow. Selection for plasmid-encoded kanamycin resistance results in integration of the plasmid into the *H. influenzae* chromosome via *iga* homology, since the plasmid is unable to replicate autonomously in *H. influenzae*. Integration will occur either upstream (left side of diagram) or downstream (right side) of the chromosomal type 2 specificity region (▲). The former case results in no change of the type of protease produced (i.e., type 2), while the latter case results in a switch to type 1 production. The presence of type 1 transformants indicates that the incoming plasmid contains the specificity region. Similar analyses would apply for carboxyl-terminal clones, except that after plasmid integration, the intact *iga* gene would be upstream of the truncated form of the gene.

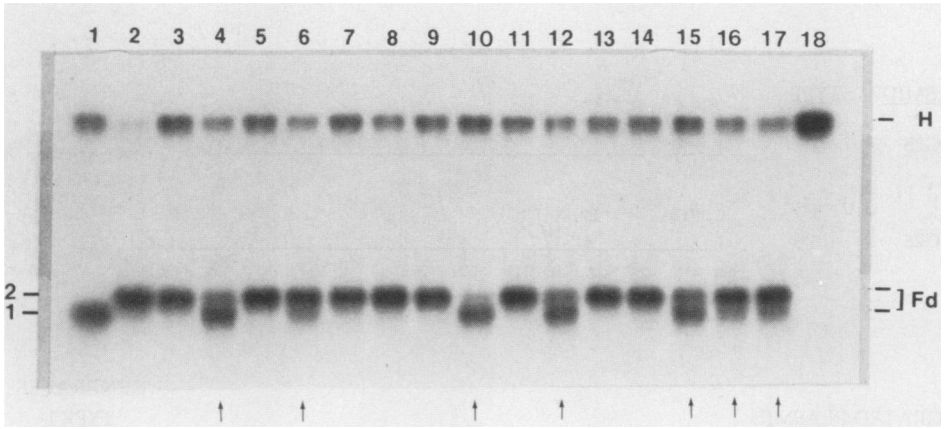

FIG. 3. IgA1 protease assay of transformants resulting from plasmid integration. Plasmid pFG11, containing the portion of the type 1 gene shown in Fig. 4B, was integrated into the chromosome of the type 2 IgA1 protease-producing *H. influenzae* HF2.8c$^+$(c3). Integrants were pooled into groups of three and inoculated into liquid growth media. After overnight growth, ^{125}I-labeled IgA1 was incubated with culture supernatants. The digestion products were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (lanes 3 through 17). The positions of the intact IgA1 heavy chain (H), and the amino terminus of the heavy chain (Fd) are shown. The positions of the Fd fragments which resulted from digestion with type 1 or type 2 enzymes are shown by numbers at the left. IgA light chains are visible only after longer exposure due to preferential ^{125}I labeling of heavy chains. Control digestions are indicated by the following lanes: 1, native *H. influenzae* type 1 protease; 2, native *H. influenzae* type 2 protease; 18, growth medium. Vertical arrows mark lanes which contain both sizes of Fd fragments, indicating that the *iga* DNA in plasmid pFG11 includes the type 1 specificity region.

each plasmid construct; plasmid integration is ensured since derivatives of plasmid pBR322 cannot replicate autonomously in *H. influenzae* (9). If the test plasmid contains the type 1 cleavage site determinant (designated CSD) and integration occurs in the region between the CSD and the terminus provided to the hybrid gene by the chromosomal copy of *iga*, then the resulting transformant will be a type 1 IgA1 protease producer. The frequency of such recombination events depends upon the relative distances from the CSD to the end of the *H. influenzae* sequences contained in the plasmid and the degree of homology between the type 1 and type 2 sequences adjacent to the CSD. Failure to detect type 1 transformants would indicate either that the plasmid does not contain the CSD or that integration is favored to occur on the side of the CSD that does not lead to a switch in specificity type.

An example of the analysis of an experiment in which type 1 transformants were found is shown in Fig. 3. In this experiment, transformants were pooled in groups of three before assay for protease type, to increase the efficiency of the screening procedure. Since several pools were found which contained both type 1 and type 2 proteases, it was concluded that the test plasmid contained the type 1 CSD. The results obtained with a number of plasmid constructs is tabulated in Fig. 4. The frequency of type 1 transformants among all Km^r transformants is shown for each plasmid. Experiments with plasmids containing the amino-terminal 1.1 kb of type 1 *iga* (pFG11, pFG19, pFG20, and pJB2::Tn*5*-208DK) resulted in the generation of type 1 transformants. Type 1 transformants resulting from integration of these plasmids would be predicted to have the amino-terminal region of *iga* contributed by the type 1 gene and the carboxyl-terminal region contributed by the type 2 gene. The negative result for pFG28 could be due to the extensive area downstream available for alternative integration events or to nonhomology between the type 1 and type 2 genes in the region between the amino-terminal end and the adjacent 1.0 kb of *iga*. An 800-bp nonhomologous region starting 200 bp into *iga* was in fact detected by heteroduplex analysis of the type 1 and type 2 genes (12). For these reasons, the results shown in Fig. 4 indicate that the carboxyl-terminal end of the CSD is within the first 1.1 kb of *iga* but do not localize the amino-terminal end of this determinant.

The above results indicate that the frequency of type 1 transformants was too low to be detected when the border of the cloned fragment approached the CSD and the nearby region of nonhomology. An alternative method, outlined in Fig. 5, was therefore developed in an attempt to select for recombination events occurring upstream of the CSD. This procedure used plasmid pFG37, which contains the first 0.7 kb of type 2 *iga* and a copy of the type 1 gene lacking only the first 63 bp of the *iga* coding sequence. *E. coli* strains containing pFG37 produce no active protease. The type 1 and type 2 sequences in pFG37 are in the same relative orientation, have 0.6 kb of *iga* in common, and are separated by the insertion of the *E. coli lamB* gene. Recombination within the 0.6-kb common region would result in the generation of an intact hybrid *iga* gene, with its amino-terminal region contributed by the type 2 gene and its carboxyl-terminal region contributed by the type 1 gene. This recombination event would result in concurrent loss of the intervening DNA, including the *lamB* gene, which encodes the receptor for bacteriophage λ (33). *E. coli* SM525, in which plasmid pFG37 was maintained, contains a mutation in the chromosomal copy of *lamB* and is therefore resistant to infection by bacteriophage λ; introduction of plasmid pFG37 results in sensitivity to λ infection. Cells containing spontaneously arising derivatives of plasmid pFG37 with deletions of *lamB* were selected by challenging cultures with λ*vir*, which will lyse LamB$^+$ λ-sensitive cells. Iga$^+$ isolates resistant to phage λ were found at a frequency of 10^{-5}, and all 12 of the strains tested proved to produce type 1 enzyme. Restriction enzyme mapping of the plasmids contained in these 12 Iga$^+$ isolates showed that 11 of 12 of the plasmids had been generated by recombination upstream of the *Eco*RI site located 200 bp into the type 2 *iga* gene. The remaining recombination event had taken place at a site in the type 2 sequence between the *Eco*RI site and a *Bam*HI site located 58 bp downstream from the *Eco*RI site (see Fig. 4c, pFG37R5, for structure). This result indicated that the CSD was located downstream of the first 200 bp of *iga*.

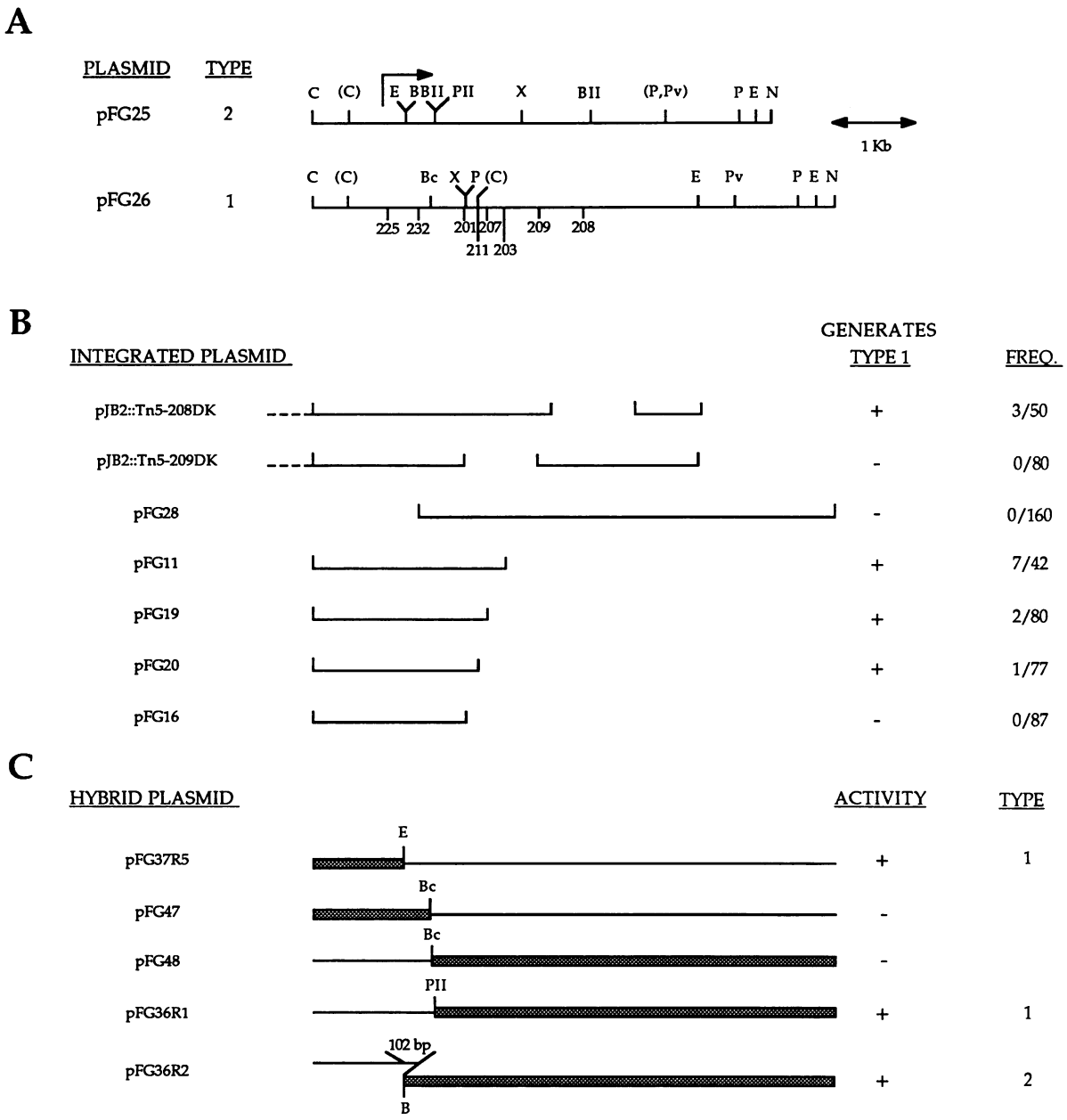

FIG. 4. Activity of type 1-type 2 hybrid *iga* genes. (A) Restriction maps for the plasmids which encode type 1 (pFG26) and type 2 (pFG25) IgA1 proteases. The initiation site and direction of transcription of *iga* is indicated by the bent arrow. Sites for DNA restriction enzymes are indicated above each restriction map with the following abbreviations: C, *Cla*I; (C), *Cla*I which is affected by Dam methylation; E, *Eco*RI; B, *Bam*HI; BII, *Bgl*II; PII, *Pvu*II; X, *Xba*I; P, *Pst*I; Pv, *Pvu*I; N, *Nde*I; Bc, *Bcl*I; (P, Pv), *Pst*I and *Pvu*I sites, the order of which has not been determined. Positions of Tn*5* insertions within pFG26 which were used for various constructions are indicated by numbered bars beneath the pFG26 map. (B) Structures of plasmids used in *H. influenzae* integration experiments to create type 1-type 2 *iga* hybrids. Lines indicate the amount of type 1 DNA which is present for each plasmid. The extent of each plasmid corresponds to the maps shown in panel A. The broken lines for pJB2::Tn*5*-208DK and pJB2::Tn*5*-209DK indicate the presence of additional sequences not shown in panel A. The columns at the right summarize whether or not (+ or −, respectively) the integration of a plasmid into *H. influenzae* HF2.8c$^+$(c3) can result in a switch of the type of protease produced from type 2 to type 1 and the frequency among transformants with which this switching occurs. (C) Structures of type 1-type 2 hybrid *iga* genes. Type 1 (▬) and type 2 (▩) sequences are shown. Restriction enzyme sites which define the junctions are as described above for panel A. The method of plasmid construction is described in detail in the text. The type 1 and type 2 sequences of pFG36R2 overlap by 102 bp. The capability of these plasmids to encode active IgA1 proteases (+ or −) and the specificity of these enzymes (1 or 2) is indicated in the columns at the right.

DNA sequence of the region spanning the CSD. The results from in vivo-generated hybrid *iga* genes showed that the CSD was located within a 0.9-kb region near the amino-terminal coding region of the 5.6-kb *iga* gene. The DNA sequence for this region of the type 1 and type 2 genes was determined and is presented in Fig. 6, together with the predicted amino acid sequences for the two enzymes. The sequence is shown starting 170 bp upstream of an ATG

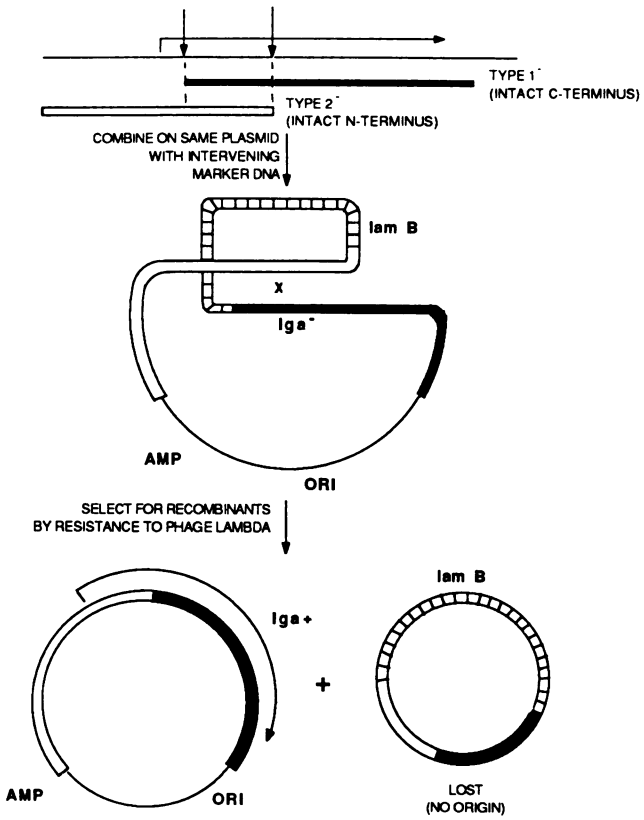

FIG. 5. *E. coli* in vivo recombination to generate hybrid genes. The horizontal arrow at the top indicates the size and direction of transcription of the *iga* gene. The thick bar and the open bar represent type 1 and type 2 sequences, respectively, which contain opposite ends of the *iga* gene. The region between the dashed lines is represented in each of the two sequences. Plasmid pFG37 contains these sequences in direct orientation such that recombination within the area delimited by the dashed lines results in loss of the *lamB* gene (▥) and subsequent resistance to bacteriophage λ infection.

which is followed by a continuous open reading frame to the point of insertion of Tn*5*-211, which defines the carboxyl-terminal limit of the region which encodes the CSD. The reading frame of the type 1 gene was verified by construction (by oligonucleotide-directed mutagenesis) of a variant with a TAG stop codon near the *Bcl*I site. This variant does not encode an active protease unless it is maintained in a suppressor-plus background (*E. coli* Q1). The first 25 amino acids of both proteins make up a region resembling a secretion signal sequence (23). The subsequent sequences show regions of homology separated by regions of no apparent homology in sequence or length. A summary of the overall comparison of the DNA and amino acid sequence between the *Eco*RI site of the type 2 gene, which defines the amino-terminal limit of the CSD, and the carboxyl-terminal limit is shown in Fig. 7. The regions of homology range from 64 to 82% identity at the DNA level and 50 to 80% identity at the protein level. The regions of nonhomology range from an insertion or deletion of 3 bp to a substitution of 45 bp for 9 bp. There does not appear to be a region in either gene that can be easily manipulated (e.g., by inversion across inverted repeats or deletion through direct repeats) to yield the sequence of the other gene. It is important to note that the *H. influenzae* type 2 gene was isolated through a procedure involving integration and excision of a plasmid containing the amino-terminal region of the type 1 gene (12). Therefore, the long region of identical sequence found at the amino terminus of the *H. influenzae* type 1 and type 2 genes may be the result of the potential hybrid nature of the cloned type 2 gene.

The hydropathic profiles for the two proteins are very similar, although the two profiles are slightly offset from one another in the center of the sequence. The most notable difference between the two profiles occurs in a region near the center of the sequences where the type 2 sequence is hydrophilic while the type 1 sequence is neutral (see boxed area in Fig. 7). This area coincides with a stretch of 51 bp of the type 2 gene and 63 bp of the type 1 gene that show no apparent homology. The possibility that cleavage site specificity could result from the difference in size in this region of nonhomology was tested by insertion of a 12-bp synthetic DNA linker (*Hin*dIII linker, New England BioLabs) at the *Pvu*II site of the type 2 gene. This site is immediately adjacent to the carboxyl-terminal end of the nonhomologous region. While insertion of a single linker resulted in total loss of protease activity, insertion of several tandem copies of the linker sequence resulted only in a decrease in the level of activity with no change in cleavage site specificity (structures verified by DNA sequence analysis; data not shown). These results indicate that an alteration in the length of this region of nonhomology is not sufficient to change cleavage site specificity and that this region of the protein is sensitive to alterations in sequence.

Activity encoded by in vitro-generated hybrid genes. DNA sequence data for the CSD regions of the two *iga* genes made it possible to derive, by site-specific mutagenesis, variants of the genes which could be used for the in vitro construction of hybrid genes. Oligonucleotide-directed mutagenesis was used to construct a variant of the type 2 gene (pFG45) that had a *Bcl*I site at the same position as the *Bcl*I site in the type 1 gene. The nucleotide alterations would change the amino acid sequence from Lys-Asp-Lys to Asn-Asp-Gln, making this variant identical to the type 1 protein in this region (see Fig. 6 for sequence). This variant still produced an active IgA1 protease of type 2 specificity. The *Bcl*I site is in the region of homology that separates the two largest regions of nonhomology between the two genes. Reciprocal hybrids (pFG47 and pFG48) were constructed to contain the amino-terminal region of one type and the carboxyl-terminal region of the other type, with the junction at the *Bcl*I site (see Fig. 4c for structure). Each hybrid therefore retained one of the large regions of nonhomology from each parent gene. Neither hybrid gene resulted in the production of active enzyme (data not shown).

Since the above results indicated that active hybrid genes could not necessarily be produced by fusion at any chosen point, a generalized technique was developed for production of iga^+ hybrid genes in vitro. This approach utilized plasmid pFG36, which is similar to plasmid pFG37 used in the bacteriophage λ resistance selection procedure (Fig. 5). This plasmid contains the amino terminus and the first 1.0 kb of the type 1 gene and all but the first 270 bp of the type 2 gene, separated by spacer DNA consisting of the Km^r determinant of Tn*5* and the *lamB* gene. The rationale was to linearize the plasmid DNA within the spacer region but close to the type 1 amino terminal segment and to digest with exonuclease BAL 31, generating a pool of molecules with deletions. The BAL 31 treatment was followed by digestion with a restriction enzyme at a unique site in the type 2 gene in the vicinity of the analogous region in the type 1 segment in which BAL 31 digestion was expected to have stopped. Intramolecular

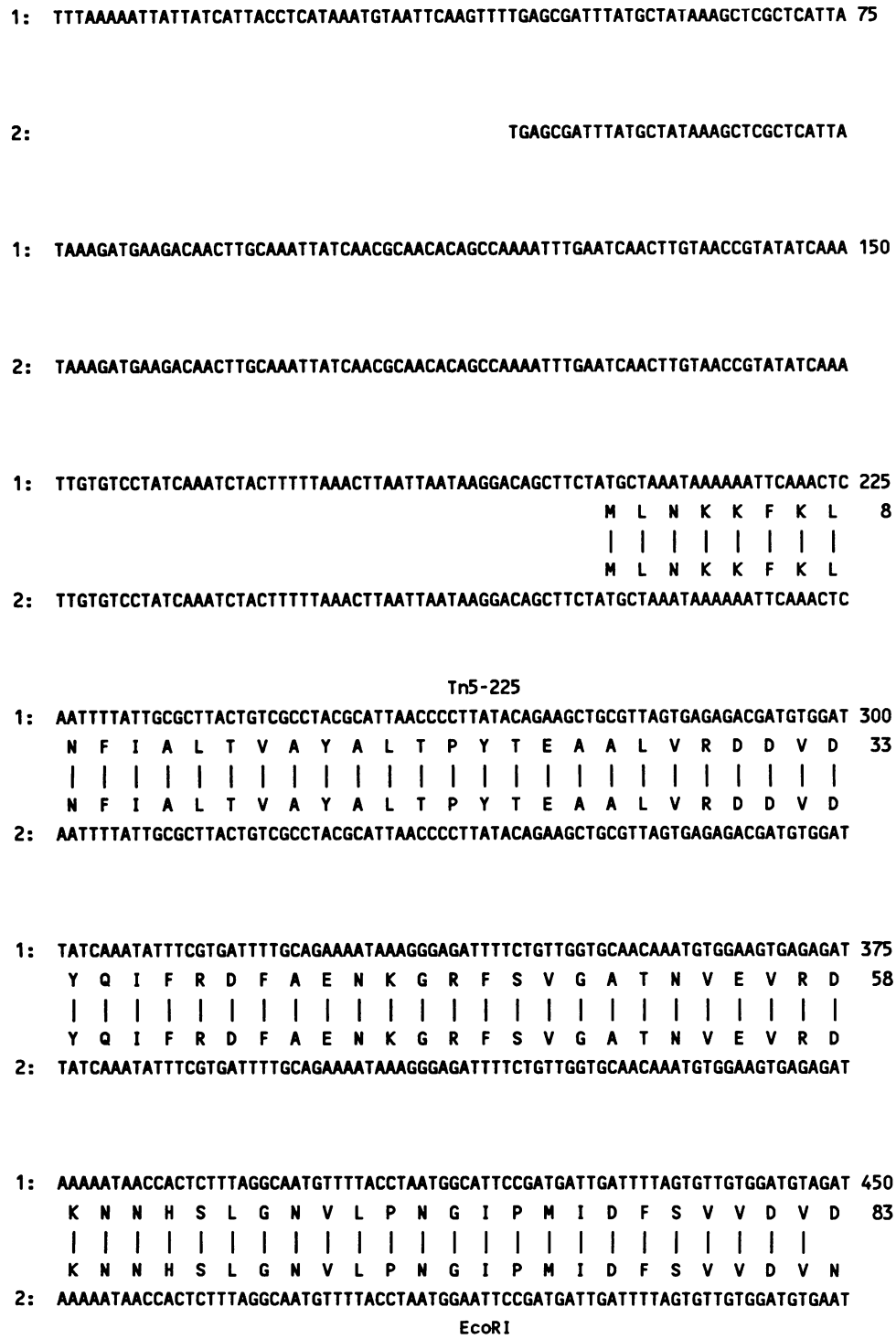

FIG. 6. DNA sequence of *H. influenzae iga* genes. The DNA sequence for the type 1 *iga* gene is shown above the sequence for the type 2 gene. The deduced amino acid sequences are shown between the DNA sequences in single-letter code. Matches in amino acid sequence are shown as vertical bars. Dashes within sequences represent padding characters introduced for optimal alignment. Numbering for amino acid positions starts from the proposed translation initiation codon for the type 1 gene, while numbering for DNA sequence starts with the first base sequenced. Key restriction enzyme sites and Tn*5* insertions are shown above (type 1) and below (type 2) the DNA sequence. The dashed line delimited by arrow heads represents the 102-bp region which was duplicated in pFG36R2 (see Fig. 4c for structure). Sequence data for the type 1 gene is shown to the position of Tn*55*-211 which was found to be beyond the cleavage site determinant region (see pFG20, Fig. 4b).

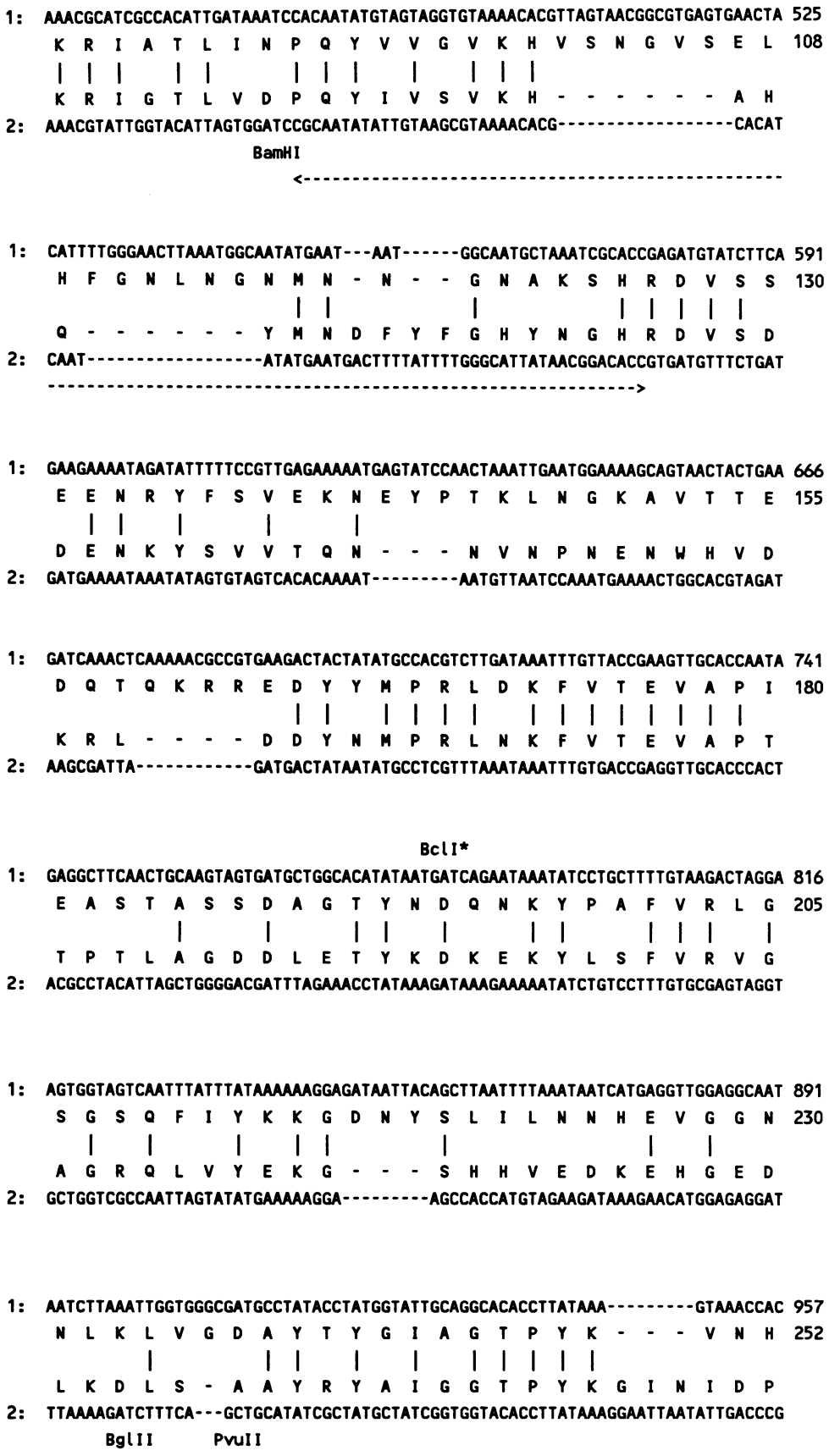

FIG. 6—*Continued*

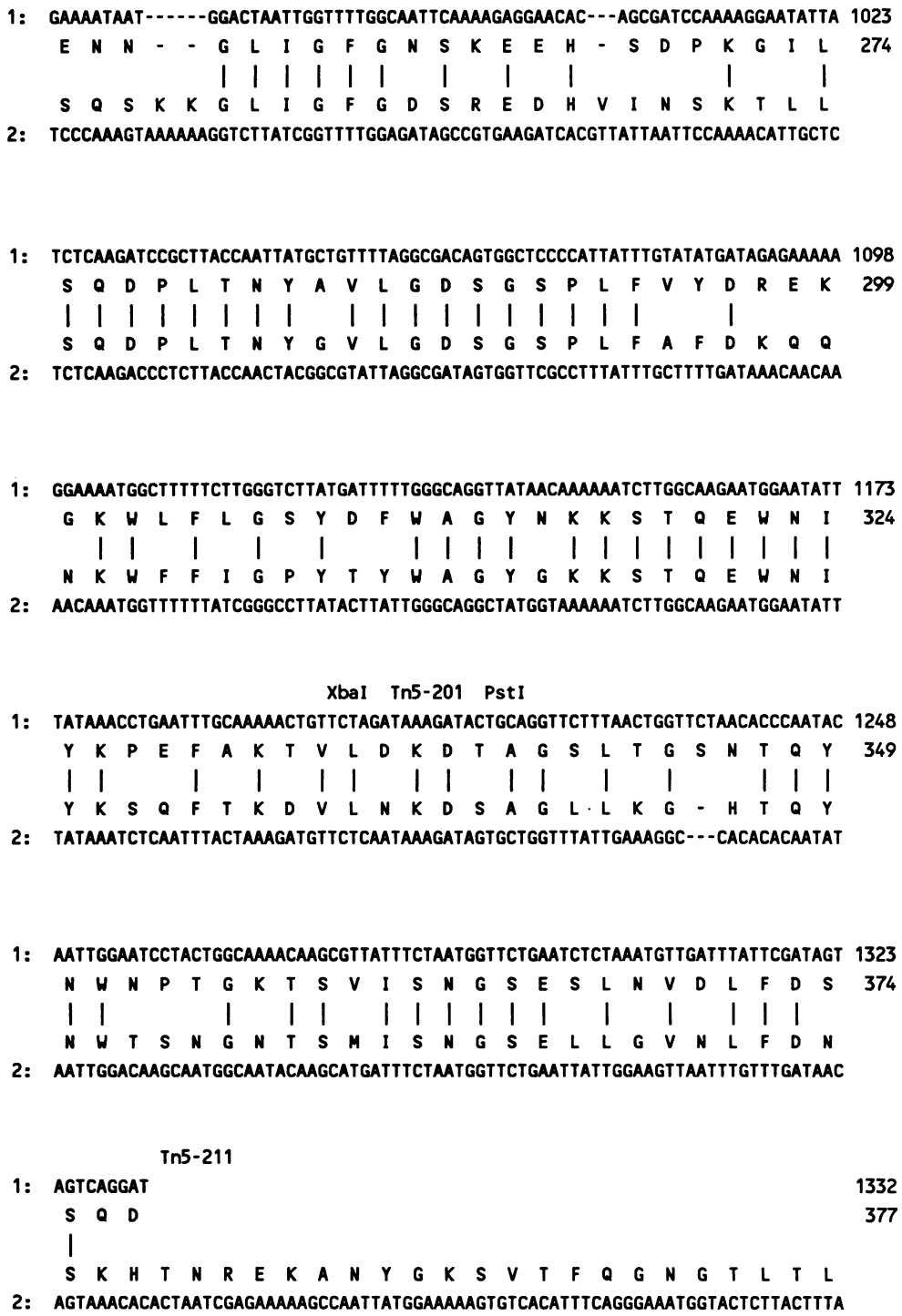

FIG. 6—*Continued*

ligation of these plasmids was predicted to generate some molecules in which the point of ligation was in an equivalent position in each *iga* segment, i.e., the spacer region would be deleted and an intact hybrid *iga* gene would remain. Transformants obtained with this pool of plasmid molecules would then be screened for IgA1 protease activity. The unique restriction site chosen for cleavage of the type 2 segment was the *Pvu*II site adjacent to the large region of nonhomology in the *iga* genes (Fig. 7). The corresponding site in the type 1 segment required that BAL 31 digestion proceed 450 bp from the restriction site in the spacer region chosen for linearization of the plasmid (the *Hpa*I site of Tn*5*). Exonuclease digestion rates were monitored by agarose gel electrophoresis of the DNA, and a pool of plasmids with the appropriate deletion length was generated. After ligation and introduction into *E. coli* MC1000, IgA^+ transformants were found with a frequency of 1%, all with type 1 specificity. DNA sequencing of one isolate verified that the construct had the

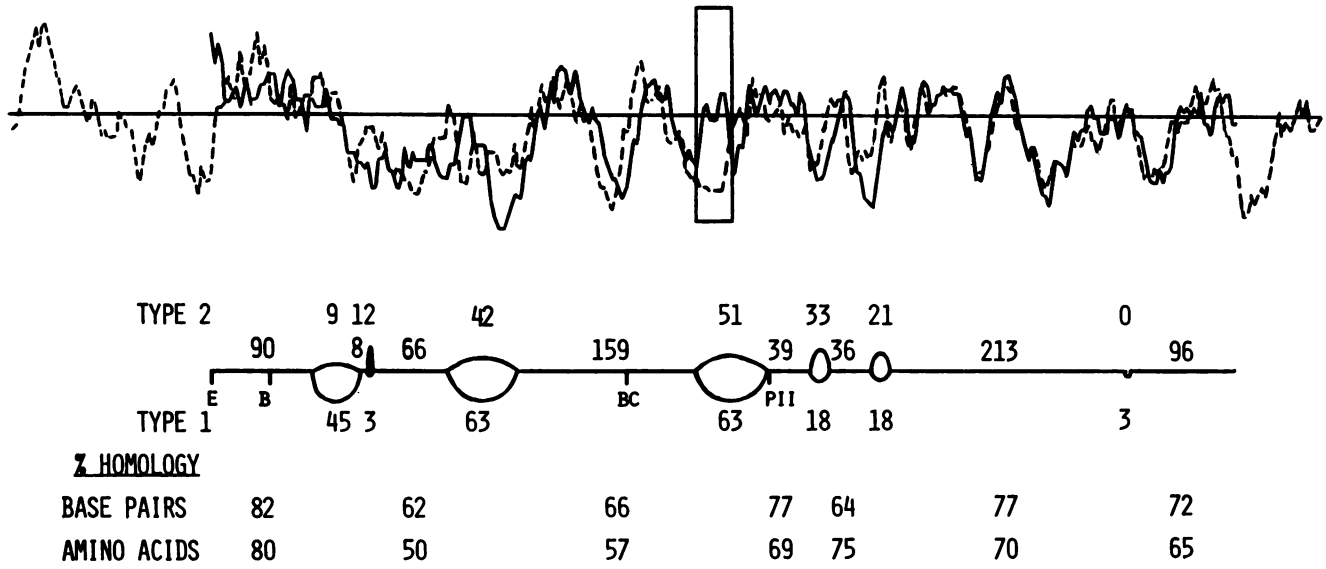

FIG. 7. Hydropathy and DNA sequence comparison of specificity regions. The top portion of the figure shows the hydropathy of the type 2 protein (broken line) from the amino terminus to the end of the sequence shown in Fig. 6. The solid line is the type 1 profile shown for the specificity region only. Points below the horizontal line indicate hydrophilic portions of the protein; points above the line indicate hydrophobic regions. The curves were aligned starting from the carboxyl-terminal end of the specificity region where the two proteins have extensive regions of uninterrupted homology. Therefore, the amino-terminal portion of the two specificity regions are offset from one another due to the differences in sequence length shown in the lower portion of the figure. The boxed area indicates a region in which the two types differ markedly. The lower portion of the figure shows a schematic comparison of the two DNA sequences in the CSD region. Solid lines are regions of the same length in each type which show a high degree of homology. The length of the region, in number of nucleotides, is indicated directly over the line and the degree of homology is tabulated at the bottom. Loops represent areas with no detectable homology. The lengths of nonhomologous loops in numbers of nucleotides are shown above (type 2 gene) and below (type 1 gene) the loops. The locations of several restriction enzyme sites are shown as vertical bars labeled as follows: E, *Eco*RI; B, *Bam*HI; BC, *Bcl*I; PII, *Pvu*II.

expected structure (see Fig. 4c, pFG36R1, for structure). Since this hybrid gene of type 1 specificity contained only the first 700 bp of type 1 DNA, this result narrowed the limits of the CSD to a 430-bp region of the type 1 gene, corresponding to the 370-bp region between the *Bam*HI and *Pvu*II sites of the type 2 gene.

A modification of the exonuclease digestion procedure, in which digestion proceeded from each end of the overlap region, was used in an attempt to generate a different hybrid gene with the two largest regions of nonhomology within the specificity region derived from different parental genes. However, no IgA^+ transformants were obtained in which the resulting gene had the expected structure. The only active hybrid obtained resulted from an *iga* gene with a partial duplication of the specificity region. This *iga* gene, of type 2 specificity, had the entire type 2 CSD region (starting from the *Bam*HI site) at the carboxyl-terminal end of the duplication and the amino-terminal 382 bp of the type 1 gene, resulting in a 102-bp duplication (see Fig. 4c, pFG36R2, for structure and Fig. 6 for exact sites of hybrid formation). This result indicates that although certain variations in the structure of the *iga* gene did not destroy enzyme activity, other variations, such as certain combinations of type 1 and type 2 *iga* sequences, resulted in loss of protease activity.

DNA hybridization with type-specific *iga* probes. Two plasmids, pFG39 and pFG40, were constructed that contained 370- and 250-bp segments of the type 2 and type 1 genes, respectively. These segments contained the CSD of each type and were within the region of nonhomology between the two genes. When these plasmids were used under stringent hybridization conditions as DNA probes of chromosomal DNA from *H. influenzae* strains, each probe hybridized only to chromosomal DNA from the strain of the same IgA1 protease type as the probe; the sizes of the hybridizing bands were consistent with the sizes predicted for each *iga* gene. This indicates that there is not a silent copy of the CSD located in the chromosome at a site different from that of the active CSD, which could potentially be used to switch the IgA1 protease specificity.

Sequence comparison with *N. gonorrhoeae* protease gene. The sequence of the 370-bp region delimited above was compared for the *H. influenzae* type 1 and type 2 genes and the *N. gonorrhoeae* type 2 gene (31). There was a high degree of sequence conservation for all three genes (Fig. 8). However, the two type 2 genes showed greater homology to each other than to the type 1 gene within the conserved-length regions, as well as greater homology in terms of amino acid sequence and segment length within the *H. influenzae* variable-length regions. Despite the high homology between the two type 2 genes, the number of differences is still too great to permit rapid identification of critical specificity-determining sequence elements.

DISCUSSION

The type 1 and type 2 *H. influenzae* IgA1 protease genes are relatively large genes of 5.6 and 4.8 kb, respectively. These genes are largely homologous, on the basis of heteroduplex analysis (12), except for two regions of 0.8 and 0.9 kb. The CSD of the type 1 and type 2 *H. influenzae iga* genes was localized through the analysis of hybrid genes generated in vivo and in vitro. The CSD was found to be within a 370-bp segment near the amino-terminal coding region of the type 2 *iga* gene, within the 0.8-kb region of nonhomology between the two *H. influenzae* genes. The role of the other nonhomologous region is not yet known, although it is near

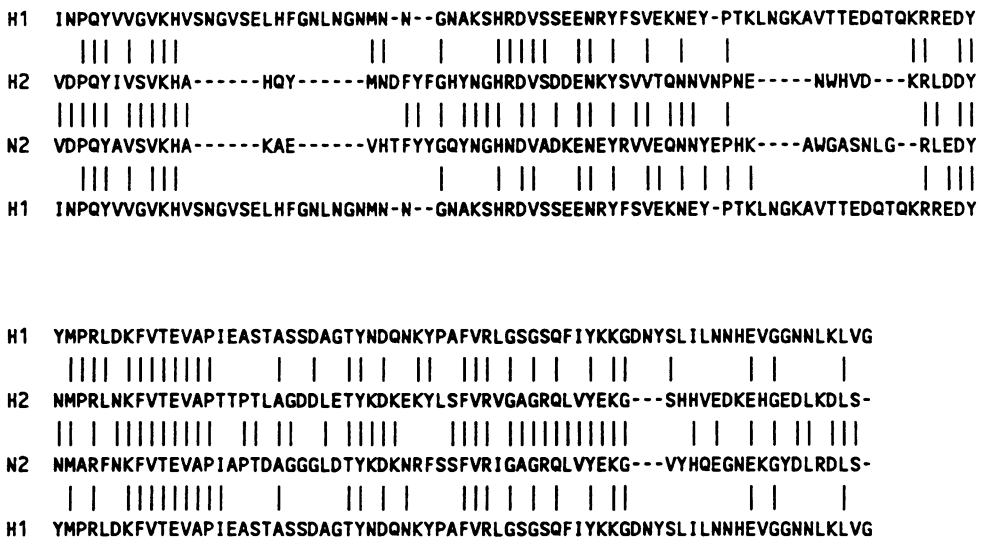

FIG. 8. Comparison of specificity regions of *H. influenzae* enzymes with *N. gonorrhoeae* sequences. The amino acid sequences for the *H. influenzae* type 1 (H1), *H. influenzae* type 2 (H2), and *N. gonorrhoeae* type 2 (N2) IgA1 proteases are shown for the region defined in this work to be the specificity determining region (amino acids 90 through 236 for the type 1 sequence shown in Fig. 6). Sequences are shown in single-letter code with dashes introduced as padding characters for optimal alignment. Vertical bars represent matches in sequence. Vertical bars beneath the *N. gonorrhoeae* sequence represent matches with the *H. influenzae* type 1 sequence. The sequence shown is between the *Bam*HI and *Pvu*II sites of the *H. influenzae* type 2 gene.

or within a region of the protease that may be processed during secretion of the protein.

Since strains of *H. influenzae* that are otherwise very similar produce IgA1 proteases of different specificities, it was of interest to determine whether there was a mechanism whereby protease production could switch from one type to another. Possible switching events could include DNA rearrangements within *iga* or alternate expression of different copies of *iga*. DNA sequence analysis did not reveal a simple mechanism, such as DNA inversion or deletion, for generation of one specificity region from the other. Also, in strains producing one type of protease, there does not appear to be a silent copy of the other specificity region located elsewhere in the chromosome.

The analysis of hybrid genes showed that the CSD was in a region of DNA that contains segments with no apparent homology between the two types, separated by regions of relatively high homology. Active hybrid genes could not be constructed which had junction points between the two largest regions of nonhomology within the CSD. This may reflect differing specific contributions to protein structure by each of these regions. However, there is some flexibility in this region of the enzyme, in that certain insertions and large duplications can be tolerated.

One goal in defining the specificity region of the IgA1 proteases is to localize amino acids which are most important for function. For example, Craik et al. (8) were able to change single amino acids near the catalytic site of trypsin, thereby altering the substrate specificity of the enzyme. Although the construction of active hybrid genes was instrumental in delimiting the CSD of IgA1 protease to a 370-bp region, additional work will be necessary to define the crucial amino acids for enzyme specificity. The inability to form certain specific hybrids, as noted above, will hinder the further use of this approach. Furthermore, difficulties in identifying one or a few critical amino acids within the CSD may be encountered for several reasons. For instance, interacting amino acids that are important for function or specificity may not be adjacent to one another in the primary sequence of the enzyme. Also, although there are a number of major differences between the two specificity regions of the *H. influenzae iga* genes, it may be that the more subtle differences are crucial for determining cleavage site specificity.

It is interesting to note that the CSD of the two IgA1 proteases of the same type but from different species, *H. influenzae* and *N. gonorrhoeae*, are more similar in structure and sequence than are the two of different types from the same species. It has been suggested previously (31) that the active site of the *N. gonorrhoeae* protease is in a region located 1.5 kb downstream from the region of homology to the *H. influenzae iga* CSDs described here. This conclusion was based upon the metal dependence of the gonococcal enzyme and the fact that the only two cysteine residues of the enzyme are encoded within this region. It is not known whether the region delineated in this work overlaps with the active site of the molecule. An alternative possibility is that the CSD is a substrate-binding site, such that the type 1 enzyme binds to a different region of the IgA1 molecule than does the type 2 enzyme, thus altering the site of cleavage by the active site of the enzyme. Another possibility is that site specificity differences are caused by variations in the region separating the binding site from the active site such that a larger spacer region would result in cleavage farther away from the binding site. This ruler mechanism was tested by inserting DNA linkers into the type 2 gene at a position that would result in better alignment with the type 1 gene. Although these insertions did not alter cleavage site specificity, other regions that could have a similar effect have yet to be tested. Further analysis to narrow down the specificity region and substrate- and inhibitor-binding studies, will be needed to determine the nature of the critical differences among proteases of different types and the function of the CSD in enzyme activity.

ACKNOWLEDGMENTS

We thank Joanne Gilbert for assistance with IgA1 protease assays and Tina Henkin for assistance with DNA sequencing.

This work was supported by Public Health Service grant AI20337 from the National Institutes of Health (A.W.) and an American Cancer Society postdoctoral fellowship (F.J.G.).

LITERATURE CITED

1. **Alexander, H. E., and G. Leidy.** 1953. Induction of streptomycin resistance in sensitive *Haemophilus influenzae* by extracts containing desoxyribonucleic acid from resistant *Haemophilus influenzae*. J. Exp. Med. **97**:17–31.
2. **Backman, K., M. Ptashne, and W. Gilbert.** 1976. Construction of plasmids carrying the *c*I gene of bacteriophage λ. Proc. Natl. Acad. Sci. USA **73**:4174–4178.
3. **Birnboim, H., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Bricker, J., M. H. Mulks, E. R. Moxon, A. G. Plaut, and A. Wright.** 1985. Physical and genetic analysis of DNA regions encoding the immunoglobulin A proteases of different specificities produced by *Haemophilus influenzae*. Infect. Immun. **47**:370–374.
5. **Bricker, J., M. H. Mulks, A. G. Plaut, E. R. Moxon, and A. Wright.** 1983. IgA1 proteases of *Haemophilus influenzae*: cloning and characterization in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **80**:2681–2685.
6. **Casadaban, M. J., and S. N. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. **138**:179–207.
7. **Cohen, S. N., A. Chang, and L. Hsu.** 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *E. coli* by R factor DNA. Proc. Natl. Acad. Sci. USA **69**:2110–2114.
8. **Craik, C. S., C. Largman, T. Fletcher, S. Roczniak, P. J. Barr, R. Fletterick, and W. J. Rutter.** 1985. Redesigning trypsin: alteration of substrate specificity. Science **228**:291–297.
9. **Danner, D. B., and M. L. Pifer.** 1982. Plasmid cloning vectors resistant to ampicillin and tetracycline which can replicate in both *E. coli* and *Haemophilus* cells. Gene **18**:101–105.
10. **Frangione, B., and C. Wolfenstein-Todel.** 1972. Partial duplication in the "hinge" region of IgA1 myeloma proteins. Proc. Natl. Acad. Sci. USA **69**:3673–3676.
11. **Gilbert, J. V., and A. G. Plaut.** 1983. Detection of IgA protease activity among multiple bacterial colonies. J. Immunol. Methods **57**:247–251.
12. **Grundy, F. J., A. Plaut, and A. Wright.** 1987. *Haemophilus influenzae* immunoglobulin A1 protease genes: cloning by plasmid integration-excision, comparative analyses, and localization of secretion determinants. J. Bacteriol. **169**:4442–4450.
13. **Herriott, R. M., E. M. Meyer, and M. Vogt.** 1970. Defined nongrowth medium for stage II development of competence in *Haemophilus influenzae*. J. Bacteriol. **101**:517–524.
14. **Howe, M. M.** 1973. Prophage deletion mapping of bacteriophage Mu-1. Virology **54**:93–101.
15. **Inouye, S., and M. Inouye.** 1987. Oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA, p. 181–206. *In* S. Narang (ed.), DNA and RNA synthesis. Academic Press, Inc., New York.
16. **Insel, R. A., P. Z. Allen, and I. D. Berkowitz.** 1982. Types and frequency of *Haemophilus influenzae* IgA1 proteases. Semin. Infect. Dis. **4**:225–231.
17. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9**:2989–2998.
18. **Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. Mol. Gen. Genet. **177**:65–72.
19. **Kett, K., P. Brandtzaeg, J. Radl, and J. J. Haaijman.** 1986. Different subclass distribution of IgA-producing cells in human lymphoid organs and various secretory tissues. J. Immunol. **136**:3631–3635.
20. **Kilian, M., J. Mestecky, and R. E. Schrohenloher.** 1979. Pathogenic species of the genus *Haemophilus* and *Streptococcus pneumonia* produce immunoglobulin A1 protease. Infect. Immun. **26**:143–149.
21. **Male, C. J.** 1979. Immunoglobulin A1 protease production by *Haemophilus influenzae* and *Streptococcus pneumoniae*. Infect. Immun. **26**:254–261.
22. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
23. **Michaelis, S., and J. Beckwith.** 1982. Mechanism of incorporation of cell envelope proteins in *Escherichia coli*. Annu. Rev. Microbiol. **36**:435–465.
24. **Miller, J.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Mulks, M. H., S. J. Kornfeld, B. Frangione, and A. G. Plaut.** 1982. Relationship between the specificity of IgA proteases and serotypes in *Haemophilus influenzae*. J. Infect. Dis. **146**:266–274.
26. **Mulks, M. H., S. J. Kornfeld, and A. G. Plaut.** 1980. Specific proteolysis of human IgA by *Streptococcus pneumoniae* and *Haemophilus influenzae*. J. Infect. Dis. **141**:450–456.
27. **Mulks, M. H., and A. G. Plaut.** 1978. IgA protease production as a characteristic distinguishing pathogenic from harmless *Neisseria*. N. Engl. J. Med. **299**:973–976.
28. **Plaut, A. G.** 1983. The IgA1 proteases of pathogenic bacteria. Annu. Rev. Microbiol. **37**:603–622.
29. **Plaut, A. G., J. V. Gilbert, M. S. Artenstein, and J. D. Capra.** 1975. *Neisseria gonorrhoeae* and *Neisseria meningitidis*: extracellular enzyme cleaves human immunoglobulin A. Science **190**:1103–1105.
30. **Plaut, A. G., R. Wistar, Jr., and J. D. Capra.** 1974. Differential susceptibility of human IgA immunoglobulins to streptococcal IgA protease. J. Clin. Invest. **54**:1295–1300.
31. **Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature (London) **325**:458–462.
32. **Rahr, S., R. Halter, H. Muller, J. Pohlner, and T. F. Meyer.** 1985. Genetic analysis of neisserial immunoglobulin A proteases, p. 157–163. *In* G. K. Schoolnik, G. F. Brooks, S. Falkow, C. E. Frasch, J. S. Knapp, J. A. McCutchan, and S. A. Morse (ed.), The pathogenic *Neisseria*. American Society for Microbiology, Washington, D.C.
33. **Randall-Hazelbauer, L., and M. Schwartz.** 1973. Isolation of the bacteriophage lambda receptor from *Escherichia coli*. J. Bacteriol. **116**:1436–1446.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schofield, P. R., and J. M. Watson.** 1986. DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. Nucleic Acids Res. **14**:2891–2903.
36. **Tomasi, T. B., and H. M. Grey.** 1972. Structure and function of immunoglobulin A. Prog. Allergy **16**:81–213.