Gene for an Immunoglobulin-Binding Protein from a Group G Streptococcus

STEPHEN R. FAHNESTOCK,* PATRICK ALEXANDER, JAMES NAGLE, AND DAVID FILPULA

Genex Corporation, Gaithersburg, Maryland 20877

Received 7 April 1986/Accepted 4 June 1986

The gene (spg) for an immunoglobulin G (IgG)-binding protein from a *Streptococcus* clinical isolate of Lancefield group G was cloned and expressed in *Escherichia coli*. The complete nucleotide sequence of the gene and 5'-flanking sequences was determined. The DNA sequence includes an open reading frame which encodes a hypothetical protein of 448 amino acid residues (MW = 47,595). The 5' end of this open reading frame encodes a sequence resembling a typical secretion signal sequence, and the remainder of the encoded protein has features reminiscent of staphylococcal protein A and of streptococcal M6 protein, including repeated sequences and a similar C-terminal structure. Aside from this C-terminal structure, the encoded protein has little direct amino acid sequence homology to either protein A or M6 protein. In *E. coli*, the cloned gene directs the synthesis of a protein which binds to immunoglobulins, including rabbit immunoglobulin, goat IgG, and human IgG3. Its binding properties are similar to those of the protein G described by Björk and Kronvall (L. Björk and G. Kronvall, J. Immunol. 133:969-974, 1984), a type III Fc receptor from a group G streptococcus.

Bacteria of several gram-positive species produce proteins which bind to the constant Fc portion of immunoglobulins (17). The best known of these Fc receptors is protein A of *Staphylococcus aureus*. Protein A has been widely used in laboratory immunochemical procedures which exploit its ability to bind to a variety of immunoglobulin G (IgG) antibodies independently of antigen association. The gene for protein A (spg) has been cloned (10, 18), and its entire DNA sequence has been determined (32).

Fc receptors with broader specificity than protein A are produced by *Streptococcus* species, especially those of Lancefield groups C and G (24, 25, 27). The protein produced by group G *Streptococcus* spp., known as protein G (6), has been shown to bind to all four classes of human IgG, including IgG3, to which protein A does not bind, and to bind more strongly than protein A to several animal IgG classes and mouse monoclonal antibodies (16). The properties suggest that protein G and protein A differ relatively subtly in their mode of IgG binding. A comparison of the structures of the two proteins is therefore of interest in relation to the source of binding specificity and to the evolutionary relationship between them. In addition, it is likely that protein G would be superior to protein A for many immunochemical applications if a more convenient source of the protein was available.

Here we report the cloning and expression in *Escherichia coli* of the gene for a protein with the properties of protein G, derived from a clinical *Streptococcus* isolate of group G. The DNA sequence of the cloned gene predicts a protein product with structural features reminiscent of protein A and of streptococcal M protein (15), but with limited direct amino acid sequence homology to either protein.

MATERIALS AND METHODS

**Bacterial strains and media.** *E. coli* SK2267 (F− gal thi T1' hisD4 recA endA sbcB15) was obtained from M. Nomura. *E. coli* GX3320 (containing spa on plasmid pGX2907) (12), *E. coli* GX3327 (containing spa on plasmid pGX2912) (11), and plasmid pGX1066 (28) has been described previously. *S. aureus* Cowan I (ATCC 12598) was obtained from the American Type Culture Collection, Rockville, Md. Streptococcal strains GX7805, GX7809, and GX7817 were clinical isolates obtained from local hospitals. All were beta-hemolytic and were determined to be of Lancefield group G by means of a commercial typing kit (BBL Microbiology Systems, Cockeysville, Md., or Pharmacia Diagnostics, Piscataway, N.J.). In addition, all produced a positive response in the colony immunopassay described below, and all agglutinated sheep erythrocytes coated with human IgG or with human IgG3.

*E. coli* was grown in L broth containing (per liter): 10 g of tryptone, 5 g of yeast extract, and 10 g of NaCl. L agar contained in addition 1.5 g of agar per liter. Both were supplemented, when so indicated, with ampicillin at 100 µg/ml. Streptococci were grown in Todd-Hewitt broth (Difco Laboratories, Detroit, Mich.) and on Trypticase soy agar containing 5% sheep blood (BBL Microbiology Systems).

**Transformation of *E. coli*.** *E. coli* were made competent and transformed essentially as described previously (9).

**Preparation of streptococcal DNA.** For preparation of DNA, streptococcal strains were grown in 250 ml of Todd-Hewitt broth containing 20 mM DL-threonine (22). After 4 h of cultivation, glycine was added to a concentration of 5% (wt/vol). The cells were harvested at A₆₀₀ = 0.5 to 1.0, after 1 h in the presence of glycine. Harvested cells were washed with phosphate-buffered saline (PBS; per liter, 1.1 g of Na₂HPO₄, 0.27 g of NaH₂PO₄, 8.4 g of NaCl), frozen in liquid N₂, and stored at −70°C. After thawing, the cells were washed with and then resuspended in 10 ml of S7 medium (34) containing 0.5 M sucrose. Mutanolysin (0.1 mg/ml; Sigma Chemical Co., St. Louis, Mo.) was added, and the suspension was incubated at 37°C for 45 min. The resulting protoplasts were recovered by centrifugation and then lysed osmotically by suspension in 100 mM EDTA (pH 8.0)-150 mM NaCl containing 0.5 mg of proteinase K per ml. After 55 min of incubation at 37°C, 2 mM phenylmethylsulfonyl
fluoride was added to inactivate proteinase K, and the mixture was incubated at 70°C for 15 min. The lysate was extracted three times with chloroform-isooamyl alcohol (24:1 vol/vol). DNA was precipitated from the aqueous phase with an equal volume of isopropanol, recovered by spooling, washed with 70% ethanol, and dried in vacuo. DNA was finally dissolved in 0.5 ml of 0.01 M Tris hydrochloride (pH 7.8)-0.05 M NaCl-1 mM EDTA.

**Gene cloning.** Strepococcal DNA (25 μl) was subjected to partial digestion with endonuclease MboI (2 units, 13 min). The partially digested DNA was fractionated by electrophoresis on a 0.8% agarose gel, and fractions with mobility corresponding to lengths between 4 and 9 kilobase pairs (kb) were excised. DNA was eluted from the gel (4) and ligated to the vector plasmid pGX1066, which had been linearized with BamHI and treated with calf intestinal alkaline phosphatase. The ligation mixture was used to transform Ca2+-shocked E. coli SK2267. Transformants were selected for ampicillin resistance on immunoassay plates.

**Colonies immunoassay.** L agar containing ampicillin (100 μg/ml) was overlaid with sheets of nitrocellulose (BA85; Schleicher & Schuell, Inc., Keene, N.H.) and cellulose acetate (OE67; Schleicher & Schuell). The upper, cellulose acetate sheet was inoculated. After colonies had formed, the lower, nitrocellulose sheet was removed, and bound IgG-binding proteins were stained immunochemically by a modification of the procedure of Hawkes et al. (14). The sheet was first blocked with PBS containing 0.5% Tween 20 and then incubated in turn with (i) normal rabbit serum, (ii) horseradish peroxidase-conjugated goat anti-rabbit IgG (IgG fraction; Cappel Laboratories, Cochranville, Pa.), and (iii) 4-chloro-1-naphthol plus H2O2. In some cases the normal rabbit serum step was omitted, and in other cases rabbit anti-goat IgG was used, also omitting the normal rabbit serum step.

**Gel transfer immunochromical analysis.** Cell lysates were prepared from 4-h cultures of E. coli strains by lysozyme treatment (1 mg/ml for 30 min at 37°C) and subjected to electrophoresis (30) on 12% polyacrylamide gels. The sample preparation buffer was modified to contain 4 M urea instead of glycerol. Protein bands were transferred electrochemically (7) to nitrocellulose, where IgG-binding bands were detected immunochemically as described above (colony immunoassay).

**Gel transfer hybridization.** DNA samples were digested with HindIII and subjected to electrophoresis in 1% agarose. DNA was transferred to nitrocellulose (BA85; Schleicher & Schuell) and hybridized to 32P-labeled probe as described by Wahl et al. (36). The final wash was at 60°C in buffer containing 1.5 mM sodium citrate, 15 mM NaCl, and 0.1% sodium dodecyl sulfate (adjusted to pH 6.8 with citric acid). DNA used as probe was a 1.9-kbp HindIII fragment isolated from plasmid pGX4547, purified by gel electrophoresis, and labeled by nick translation with a kit from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.).

**Hemagglutination.** Sheep erythrocytes (Cappel Laboratories) were coated with immunoglobulin essentially as described by Adler and Adler (2). Tannic acid-treated erythrocytes were incubated in PBS containing (i) total human IgG (Sigma Chemical Co.), (ii) a human IgG3(A) myeloma protein (a generous gift from D. Segal, National Cancer Institute), or (iii) no protein and then washed with PBS. For the agglutination assay, 50 μl of a 1% suspension of coated erythrocytes was mixed with 50 μl of test sample, diluted serially with PBS, in a conical well of a multiwell dish. Unagglutinated erythrocytes sedimented to the bottom of the well and formed a small pellet, while agglutinated cells formed a more diffuse precipitate on the walls of the well. For streptococcal strains, samples were heat-killed (80°C, 5 min) intact bacterial cells. For E. coli strains, cell lysates were obtained by lysozyme treatment as described above (gel transfer immunochemical analysis).

**Bacteriophage M13 cloning.** Bacteriophage M13mp8 and M13mp9 derivatives were constructed by methods described by Messing (20).

**DNA sequence analysis.** Restriction fragments of the purified 1.9-kbp HindIII segment were cloned into bacteriophage M13mp8 or M13mp9 (21). The procedure of Dale et al. (8) was used to shorten progressively some of the larger inserts. The phage clones were sequenced by the dideoxy method (5). Sequence was determined independently on both strands throughout.

**RESULTS**

**Cloning of the gene.** DNA isolated from a group G streptococcal clinical isolate (GX7809) was partially digested with MboI and enriched for fragments in the size range of 4 to 9 kb by gel electrophoretic fractionation. These fragments were ligated to the BamHI-cut and phosphatase-treated vector pGX1066 (Fig. 1). E. coli SK2267 cells transformed with the ligation mixture were plated on immunoassay plates, which permit the screening of individual colonies for the production of IgG-binding protein.

This procedure, developed by James Anderson (unpublished data), is similar in principle to the procedure described by Shiroza et al. (29). Cells are plated on a cellulose acetate filter, which is placed atop a nitrocellulose filter on a selective agar plate. Proteins released either by secretion or by spontaneous cell lysis from colonies which form on the cellulose acetate filter pass through that filter and are adsorbed to the nitrocellulose below, while whole cells are retained by the cellulose acetate filter. After colonies have formed, the nitrocellulose filter is removed, and IgG-binding proteins adsorbed to it are detected immunochemically as described in Materials and Methods.

In the screening procedure, several positive spots were identified among approximately 10,000 transformants, but only one of these spots could be correlated with a visible colony. Attempts to restreak from the area of several tiny, intensely positive spots produced no positive progeny. One positive spot could be correlated with a visible colony, but the positive phenotype was extremely unstable. However, repeated propagation on ampicillin plates produced a stable, positive, ampicillin-resistant clone, designated GX7820.

Surprisingly, plasmids isolated from GX7820 consisted predominantly of a species of the same size and restriction properties as the vector pGX1066, with no inserted DNA. In addition, however, there was a minor band of lower mobility, which proved to be pGX1066 plus an insert of 10 kbp of streptococcal DNA. Retransformation of E. coli with an unpurified plasmid preparation from GX7820 produced the pattern shown in Fig. 2. On the immunoassay filter were many small, intensely positive spots, the majority of which could be correlated with no visible colony, or in a few cases with tiny colonies. A smaller number of positive, but less intense spots did correspond to normal colonies and resembled GX7820 in their immunoassay intensity.

The explanation of this phenomenon was clarified by the following experiment. Plasmids isolated from GX7820 were fractionated by agarose electrophoresis, and the two bands,
FIG. 1. Cloning and subcloning of the streptococcal spg gene. Broad segments represent sequences derived from streptococcal DNA. The open segment in pGX4530 represents the location of sequences deleted in pGX4533.

the smaller, vector-sized band designated pGX1066X and the larger band designated pGX4530, were eluted from the gel separately. *E. coli* cells were transformed with each of the bands separately and with a mixture of the two. By itself, pGX1066X produced no ampicillin-resistant transformants. pGX4530 produced only a few tiny colonies, but many intensely positive spots on the immunoassay filter, again mainly corresponding to no visible colonies. However, a mixture of the two bands produced a pattern indistinguishable from that shown in Fig. 2, including positive colonies resembling GX7820.

To explain these results, we propose that pGX1066X is a derivative of pGX1066 which has suffered a mutation in its *amp* gene and can no longer confer ampicillin resistance. It
functions in GX7820 as a cryptic helper plasmid, which has
the effect of depressing the copy number of pGX4530, with
which it is coresident. pGX4530 apparently can confer
ampicillin resistance, but in the absence of the cryptic helper
pGX1066X, it assumes a higher copy number and becomes
lethal to the cell. GX7820 is stable under ampicillin selection
because loss of either plasmid is lethal—pGX4530 because it
alone carries an active amp gene, and pGX1066X because
without it pGX4530 increases in copy number to unacceptable
levels. The original transformant had probably acquired
both pGX1066 and pGX4530, and was unstable because
pGX4530 was readily thrown off owing to incompatibility
with pGX1066. Eventually, however, a mutation occurred in
pGX1066, producing pGX1066X, and this created the stable
situation described.

Among the products of retransformation shown in Fig. 2
were a few intensely positive spots which seemed to corre-
spond to tiny but visible colonies. Two of these positive
transformants could be propagated by restreaking on
ampicillin plates and retained the strongly positive pheno-
type. One, designated GX7822, was found to contain a single
plasmid species (pGX4533) which proved to be a derivative
of pGX4530 which had suffered a deletion of approximately
2 kbp of the streptococcal DNA insert. The other, design-
nated GX7822, also contained a single plasmid (pGX4532), a
derivative of pGX4530 which had acquired an additional
approximately 2 kbp of DNA. Based on restriction analysis
(data not shown), the point of insertion in pGX4532 was very
near one end of the deletion in pGX4533, approximately 2.5
to 3.5 kbp from one end of the streptococcal insert. Plasmid
pGX4533 was purified by retransforming E. coli SK2267,
producing strain GX7825.

Location of the gene. Immunoglobulin-binding proteins
produced by the original transformant GX7820 and by its
derivatives GX7822 and GX7823 were analyzed by Western
blot electrophoresis (Fig. 3; data not shown for GX7822). All	hree strains produced identical patterns of multiple IgG-
binding bands. Therefore the deletion in pGX4533 and the
insertion in pGX4532 did not interrupt the gene. To deter-
mine on which side of the deletion the gene is located in
pGX4533, we subjected the plasmid to restriction fragment
deletion analysis. Deletion of EcoRI or HpaI fragments
located counterclockwise (Fig. 1) from the point of sponta-
neous deletion in pGX4533 produced only transformants
which were negative on immunoassay, while deletion of
NruI and EcoRV fragments clockwise from the point of
spontaneous deletion produced transformants which were
positive on immunoassay (data not shown). Finally, a
HindIII fragment of 1.9 kbp isolated from pGX4533 was
recloned in pGX1066 (Fig. 1) and found to produce trans-
formants which were positive on immunoassay. One such
transformant, designated GX7841(pGX4547), was found by
Western blot analysis to produce a pattern of immunoglob-
ulin-binding protein bands identical to that of GX7820 (Fig.
3).

DNA hybridization (Southern blot) analysis indicated that
the 1.9-kbp HindIII fragment cloned in pGX4547 hybridized
to a fragment of the same size in the chromosomal DNA of
the streptococcal strain from which it was derived, in
plasmid DNA from the original clone pGX4530, and in the
spontaneously deleted plasmid pGX4533 (Fig. 4). This indi-
cates that the structure of this fragment was not detectably
altered either in the initial cloning or in generating the
deleted plasmid pGX4533. Included in the experiment shown
in Fig. 4 are chromosomal DNA samples from two additional
independent clinical isolates of group G streptococci and
from S. aureus Cowan I. Both streptococcal DNA samples
contain sequences which hybridize to the 1.9-kbp HindIII

FIG. 2. Plate immunoassay of E. coli tetrasformed with plasmid
isolated from GX7820. E. coli SK2267 was retransformed with a
crude plasmid preparation isolated from GX7820 by the method of
Holmes and Quigley (16). Transformants were selected for ampicil-
in resistance on immunoassay plates. The underlying nitrocellulose
filter was developed as described in Materials and Methods. Visible
colonies on the cellulose acetate filter corresponded only to the
larger, less intense spots.

FIG. 3. Immunoblot analysis of IgG-binding proteins produced by
E. coli strains carrying the cloned spg gene. Cell extracts were
prepared and analyzed as described in Materials and Methods. Lanes:
1, prestained molecular weight standards (Bethesda Re-
search Laboratories); 2, GX7837 carrying the vector plasmid
pGX1066; 3, GX7820(pGX4530), the initial positive spg clone; 4,
GX7825(pGX4533) carrying the spontaneously deleted plasmid; 5,
GX7841(pGX4547) carrying the 1.9-kbp HindIII fragment; 6,
GX7852(pGX4558) carrying the truncated spg gene with duplicated
210-bp PstI fragment; 7, GX3320(pGX2907) carrying the spa gene
encoding protein A; 8, protein A (Pharmacia), 40 ng. Lanes 2 and 4
to 6 contained the equivalent of 0.12 ml of culture, lane 3 contained
the equivalent of 0.25 ml, and lane 7 contained the equivalent of
0.025 ml. Stds, Standards (molecular weight x 10^6).

STREPTOCOCCAL IgG-BINDING PROTEIN G GENE
were analyzed for immunoglobulin-binding protein by hemagglutination of sheep erythrocytes coated with IgG (Fig. 5). Tannic acid-treated erythrocytes were coated with total human IgG, with an IgG3 myeloma protein, or with no protein. Extracts of spg-containing strains, including GX7825 and GX7841 (data not shown), agglutinated erythrocytes coated with either total IgG or IgG3 but not uncoated cells. In contrast, extracts of E. coli GX3320 and GX3327 (11, 12), which produce protein A, agglutinated cells coated with total IgG but not those coated with IgG3. Protein G has been shown to bind to IgG as well as to the other classes of human IgG, while protein A does not bind to IgG3. Therefore, the immunoglobulin-binding protein produced by E. coli GX7825 and GX7841 has the distinguishing property of protein G.

**DNA sequence.** The 1.9-kbp HindIII fragment carrying the putative protein G gene was recloned in bacteriophage M13mp9, and the DNA sequence of the entire fragment was determined on both strands, as described in Materials and Methods.

The DNA sequence, along with its proposed translation, is shown in Fig. 6. The sequence reveals an open reading frame encoding a hypothetical protein of 448 amino acid residues. Preceding the large open reading frame are sequences resembling the consensus sequences for a gram-positive vegetative promoter (23), the −10 sequence, TATAAT, is identical to the consensus sequence, and the −35 sequence, TTGATT, compares with the consensus sequence TTGACA. The spacing between the putative −10 and −35 sequences is 16 bp, compared with the favored value of 17 bp. The sequence preceding the coding region is very AT-rich (78% AT) between positions 120 and 600, compared with 63% for the entire fragment and 59% in the coding region. Upstream from the initiator ATG codon is a 7-base sequence (AAAGGAG) complementary to the 3′ end of 16S rRNA of both Bacillus subtilis and E. coli, satisfying the requirements for a gram-positive ribosome-binding sequence (ΔG = −14 kcal/mol) (19, 23). This putative ribosome-binding sequence is separated from the proposed ATG initiator codon by five A residues. The average spacing measured in this way for 12 gram-positive sequences com-

---

**FIG. 4.** Gel blot hybridization analysis of DNA from E. coli, S. aureus, and Streptococcus strains. Plasmid DNA prepared from E. coli strains (16), S. aureus chromosomal DNA (18), or chromosomal DNA prepared from streptococcal strains was digested with HindIII and analyzed as described in Materials and Methods. The probe was the 1.9-kbp HindIII fragment, gel purified and nick translated. Lanes: 1, S. aureus Cowan I DNA; 2, Streptococcus strain GX7805 DNA; 3, Streptococcus strain GX7817 DNA; 4, Streptococcus strain GX7809 DNA; 5, plasmid DNA from E. coli GX7820; 6, pGX4533; 7, pGX4329 (a negative control pBR322 derivative).

**FIG. 5.** Hemagglutination of IgG-coated erythrocytes with extracts of E. coli strains. Sheep erythrocytes coated with IgG3 (azy) myeloma protein, with total human IgG fraction, or with no protein were incubated with extracts of E. coli strains as described in Materials and Methods. Successive wells down each column contained serial one-third dilutions of the E. coli extract sample indicated at the bottom of the column: A, GX3327 (spa); B, GX7825 (spg); C, GX7835 (spg); D, GX7837 (negative control containing pGX1066); E, GX3320 (spg); F, GX7830 (spg). GX7825 contains the spontaneously deleted spg plasmid pGX4533. GX7830 and GX7835 are subclones derived from pGX4533.
pared by Moran et al. (23) is six residues, generally AT rich in gram-positive sequences.

The protein encoded by the open reading frame includes the following features. (i) A typical secretion signal sequence, including an amino-terminal sequence of net positive charge, an unbroken stretch of apolar amino acid residues, a sequence around Gly-28 which may form a beta turn, and a site at Ala-33 which satisfies the empirical rules noted by Perlman and Halvorson (26) and by von Heijne (35) for a signal peptide cleavage site. This putative signal sequence is longer than is typical for E. coli signal sequences, but is more typical of known gram-positive signal sequences, especially those of streptococcal and staphylococcal proteins (1, 15).

(ii) A region especially rich in alanine residues (25 of 43 residues from positions 69 to 111).

(iii) Two sets of repeated sequences, each represented twice, designated A1, A2 and B1, B2, respectively. Sequences A1 and A2 are 37 amino acid residues long and separated by 38 unique residues; B1 and B2 are 35 residues long and separated by 15 unique residues. This feature is apparent in the dot-matrix self-comparison in Fig. 7a, giving rise to two diagonal arrays off the main diagonal. Amino acid residues which are duplicated in the repeated sequences are

Fig. 6. DNA sequence of the 1.9-kbp HindIII fragment and its proposed translation. \(-10\) and \(-35\), Proposed promoter sequences; \(rbs\), proposed ribosome-binding sequence; *, amino acid residue common to a pair of repeat sequences; +, amino acid residues with conservative replacements in repeat sequence; ------, inverted repeat DNA sequence.

```
AACTTGGCTGGGAAATTTCTTGGGCCCCCAAGGGAATTTTTCGGTGGGGTTTTTACCAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Fig. 6. DNA sequence of the 1.9-kbp HindIII fragment and its proposed translation. \(-10\) and \(-35\), Proposed promoter sequences; \(rbs\), proposed ribosome-binding sequence; *, amino acid residue common to a pair of repeat sequences; +, amino acid residues with conservative replacements in repeat sequence; ------, inverted repeat DNA sequence.

```
FIG. 7. Dot-matrix self-homology plots. The protein (a) or DNA (b) sequence (left to right on abscissa) is compared with itself (top to bottom on ordinate). In panel a, the coordinates of each dot represent the positions of single identical amino acid residues. In panel b, the coordinates of each dot represent the positions of the midpoints of identical trinucleotide sequences. ▲, Possible end of signal sequence. The plots were generated by the Dotmatrix program provided by the Protein Identification Resource, Georgetown University Medical Center.
indicated in Fig. 6 by *, those with conservative replacements are indicated by +. The duplications are also apparent at the DNA level (Fig. 7b). The degree of sequence homology of A1 and A2 is 79% at the DNA level and 73% at the protein level. For B1 and B2, homology is 88% at the DNA level and 89% at the protein level.

(iv) A region rich in relatively evenly spaced proline residues (8 of 25 residues between positions 358 and 382).

(v) A highly charged region consisting of a sequence resembling the pentapeptide Asp-Asp-Ala-Lys-Lys, repeated five times (C1 to C5).

(vi) A C-terminal sequence with strong homology to the so-called membrane anchor, which has been identified in the sequences of staphylococcal protein A (32) and of streptococcal M6 protein (15). *+, Identical residues; +, conservative replacement.

**Preliminary localization of sequence encoding the IgG-binding site.** To identify sequences in the 1.9-kbp HindIII fragment which are essential for the IgG-binding activity of the encoded protein, we constructed deletions by making use of the PstI sites indicated in Fig. 1 and 6. Deleting sequences distal to the PstI site at position 1333 (Fig. 6) should leave a coding sequence which preserves intact the A1, A2 set of repeats, but includes only a portion of B1 and lacks B2 as well as the proline-rich, charged repeat C, and membrane anchor sequences. Such deletions were constructed by digesting double-stranded DNA of an M13mp9 derivative (Fig. 1), in which the 1.9-kbp fragment was inserted at the HindIII site, with endonuclease PstI and then religating. Cells transfected with the intact phage mGX4547 DNA produced a positive immunochemical response when patched on assay plates, while all PstI-deleted phage DNA produced clones which were negative in the plate immunoassay.

One positive clone was identified in this experiment which proved, upon DNA sequencing, to have lost the 419-bp PstI fragment encoding the C-terminus of the putative protein product, but had acquired two tandem copies of the 210-bp PstI fragment. The two PstI sites which define this fragment are located at identical positions in the respective repeated coding sequences B1 and B2 and in the same relationship to the reading frame. Therefore, when this fragment is duplicated, the structure of the encoded protein includes two complete copies of the repeated amino acid sequence B (B1 and a hybrid B2/B1).

On the basis of these observations, we tentatively identified the repeated sequence B1, B2 as responsible for IgG binding. Sequences encoded in the 419-bp PstI fragment, including the proline-rich region, the highly charged repeat structure C, and the membrane anchor, are apparently not required for IgG-binding activity, and sequences encoded upstream from the 210-bp PstI fragment, including the entire repeat structure A1, A2, are not sufficient for IgG-binding activity. However, sequences in this region, which includes the putative sites of initiation of transcription and translation and the putative signal sequence, are required for expression of IgG-binding activity. This is shown by the observation that a deletion in pGX4533 from the vector EcoRI site to the EcoRI site at position 917 of the DNA sequence (Fig. 1 and 6) produces clones with a negative immunoassay phenotype (data not shown).

**DISCUSSION**

We cloned the gene spg from a group of G Streptococcus sp. which encodes an immunoglobulin-binding protein. In E. coli, this gene directs the synthesis of a protein which binds to rabbit immunoglobulin, to goat IgG, and to human IgG3. These are definitive properties of immunoglobulin-binding proteins of type III (25), which includes the protein G described by Bjorck and Kronvall (6), also isolated from a group G Streptococcus sp.

The DNA sequence of spg encodes an acidic protein of Mr 47,595 (43,973 after removal of the putative signal sequence). The protein G of Bjorck and Kronvall (6) is also acidic, but its reported molecular weight of 30,000 was determined for protein released from cells by papain and probably does not apply to the intact protein. If the spg product with signal removed was cleaved by papain between the proposed IgG-binding domain B2 and the proline-rich region which may be the site of cell wall attachment (see below), the product would be a protein of approximately 34,000 daltons. Definitive evidence that the cloned spg gene encodes the protein G of Bjorck and Kronvall must await comparison of the translated spg sequence with the amino acid sequence of the Bjorck and Kronvall protein, which is not yet available. In any case, because the cloned gene and the Bjorck and Kronvall protein were derived from independent clinical isolates of group G Streptococcus sp., strain differences are likely, especially in view of the differences among independent group G streptococcal isolates in the sizes of restriction fragments which cross-hybridize with the cloned spg gene (Fig. 4). Strain differences have been reported among proteins A from different staphylococcal strains (13, 33).

The translated DNA sequence of the cloned gene reveals a putative protein with several features expected of a protein G. The organization of the sequence is strongly reminiscent of protein A. Overall, the protein is relatively hydrophilic and rich in alpha-helix-forming residues, especially Ala.
There is a putative secretion signal sequence, and much of the following sequence is constructed of several repeated domains. Nearer the C terminus is a region rich in evenly spaced proline residues, followed by an N-terminal membrane anchor of protein A (Fig. 8). The proline-rich region and membrane anchor of protein A are believed to function in cell wall attachment (13), although this function has not been demonstrated directly. Similar features in the spg product would suggest that it too is a cell wall protein, as is protein G.

All these features are also shared by the M6 protein of *Streptococcus pyogenes* (15). The C-terminal membrane anchor of M protein is more closely related to that of the spg product than either is to protein A (Fig. 8). However, aside from a more limited homology in the signal sequence region, the membrane anchor sequence represents the only noticeable direct amino acid sequence homology among the spg product, protein A, and M protein. Nevertheless, it is clear from the basic similarity of organization that the three proteins are related members of a family of gram-positive cell wall proteins.

On the basis of this level of similarity to both protein A and M protein and of the observed IgG-binding activity of the gene product produced in *E. coli*, we propose that the cloned spg gene encodes a protein G. We will refer to the product of the cloned spg gene as protein Gc.

It is perhaps surprising that the areas of protein Gc implicated in IgG binding, namely, the B1, B2 repeat, show no noticeable direct amino acid sequence homology to the IgG-binding domains of protein A (32). It may be that the tertiary structure of the two molecules display more subtle homology. Alternatively, it may be that the observed differences in specificity between protein A and protein G reflect different modes of IgG binding. In that case it would seem likely that the proteins arose independently by convergent evolution.

In *E. coli*, the cloned spg gene directs the synthesis of multiple IgG-binding species, many of which are probably generated by proteolysis. The most prominent band on an immunoblot runs with the mobility of a protein of molecular weight 57,000. This is significantly larger than the predicted spg product (Mr = 47,595, or 43,973 after signal removal). However, both protein A and M6 protein display anomalously low mobility on sodium dodecyl sulfate gels, and this property has been attributed (30) to the proline-rich region, a feature which is shared by protein Gc. M protein, for example, runs with the mobility of a protein of 57,000 molecular weight, although its molecular weight, deduced from the DNA sequence, is 48,956 (15). It is not clear whether this aberrant migration is due to abnormal sodium dodecyl sulfate binding in the proline-rich region or to posttranslational modification, as suggested by Hollingshead et al. (15). Strain GX7852, which carries a truncated derivative of spg lacking the sequences encoding the proline-rich region and membrane anchor, produces a major IgG-binding band running close to the expected size, which supports the assertion that the anomalous mobility of protein Gc is attributable to these regions. The identity of minor bands of lower mobility, which are present in all positive samples in Fig. 3, including the GX7852 product and protein A, is unclear.

The sequences identified as possible promoter and translation initiation sites are typical of such sequences in gram-positive genes (23). Upstream from the proposed promoter is a region rich in runs of A and T, which is apparent as a dark square in the DNA sequence dot-matrix self-homology plot (Fig. 7b). The proposed promoter sequence (Fig. 6) is close to the consensus sequence for vegetative promoters of *B. subtilis* and for *E. coli* promoters and would therefore be expected to function efficiently in *E. coli* or *B. subtilis*. The upstream AT-rich region includes several inverted repeat sequences (Fig. 6), two of which overlap the putative promoter sequence.

The pattern of codon utilization conforms well to the pattern described by Hollingshead et al. (15) for streptococcal and staphylococcal genes. The frequency of A or T in the third positions of codons in spg is 74%, compared with 72% for the seven streptococcal and staphylococcal genes discussed by Hollingshead et al. (15). Furthermore, codons identified (15) as underutilized in gram-positive genes (CTC, CTG, TCC, TCG, CCC, ACG, GCC, CAG, CGG, AGG, and GGG) are all used relatively infrequently in spg.

The process of obtaining a stable spg clone presented some unusual features. Apparently the originally cloned 10-kbp fragment of streptococcal DNA could not be tolerated in *E. coli* at high copy number. Stable clones were obtained, first with the help of a spontaneously generated cryptic helper plasmid derived from the vector pGX1066, and finally by spontaneous insertion and deletion mutations affecting the original plasmid. Since the insertion and deletion did not affect the structure of the spg gene itself, it is likely that the deleterious effects of pGX4530 at high copy number were due to an adjacent region of the DNA insert and not to spg itself.

Even though the DNA fragment which was sequenced was derived from a deleted plasmid, the deletion did not affect the structural organization of the sequenced DNA. This is supported by two arguments. First, the 1.9-kbp HindIII fragment cloned in pGX4547 is identical in gel mobility to a HindIII fragment present in the original plasmid pGX4530 and to a cross-hybridizing fragment present in the chromosomal DNA of the donor strain. Second, comparison of the translated amino acid sequence with those of streptococcal M protein and protein A at the C terminus (Fig. 8) indicates that the entire coding sequence is probably present. However, it is unclear whether there is a transcription terminator in the sequenced HindIII fragment, which ends 26 bp beyond the end of the coding sequence. There is an imperfect inverted repeat (Fig. 6) which would form an mRNA hairpin (△G = -6 kcal/mol) (31) and which may continue beyond the HindIII site. Whether this is, in fact, a terminator, or whether a terminator is present in the adjacent sequence of pGX533 remains to be determined.

Hybridization analysis (Fig. 4) indicates that the sequenced HindIII fragment exists intact in the chromosomal DNA from which the cloned fragment was obtained. This excludes the possibility that the 1.9-kbp HindIII fragment (which contains three *MboI* sites) was assembled from noncontiguous *MboI* fragments before cloning. Two independent group G streptococcal isolates, both of which produce an IgG-binding protein, contain DNA sequences homologous to the cloned HindIII fragment, but these sequences are found in restriction fragments of different sizes. This suggests that the spg genes of all three isolates are homologous but that subtle strain differences are to be expected.

**ACKNOWLEDGMENTS**

We thank Marshetta Lucas (Washington Adventist Hospital), Sylvia Brake and Amelia Maters (The Johns Hopkins Hospital), and their colleagues for providing streptococcal clinical isolates, D. Segal (National Cancer Institute) for a sample of IgG3, Doug Sims...
and Bob Ladner for computer assistance, including a program for depicting plasmids, and Mark Guyer, Charlie Saunders, Vasantha Nagarajan, Tom Poulos, and Jeff McGuire for discussions and suggestions.

LITERATURE CITED