

# Expression and purification of a truncated recombinant streptococcal Protein G

Christopher R. GOWARD,\* Jonathan P. MURPHY, Tony ATKINSON and David A. BARSTOW

Division of Biotechnology, PHLS Centre for Applied Microbiology and Research, Porton Down, Salisbury SP4 0JG, Wilts., U.K.

---

The gene for Protein G from *Streptococcus* strain G148 was cloned and expressed in *Escherichia coli*. The regions on the gene corresponding to the albumin-binding domains and the Fab-binding region were then deleted by site-directed mutagenesis. The translation of regions corresponding to the cell-wall- and membrane-binding domains was prevented by introduction of stop codons upstream of these domains. This recombinant DNA sequence codes for a protein (G') that contains repetitive regions and that binds only the Fc portion of IgG, analogously to Protein A. Translation of the sequence produces a protein with an  $M_r$  of about 20000. The nucleotide sequence differs from those published previously [Guss, Eliasson, Olsson, Uhlén, Frej, Jornvall, Flock & Lindberg (1986) *EMBO J.* 5, 1567–1575; Olsson, Eliasson, Guss, Nilsson, Hellman, Lindberg & Uhlén (1987) *Eur. J. Biochem.* 168, 319–324]. The protein can be substantially purified on a large scale by chromatography on IgG–Sephacrose 4B. Homogeneous Protein G' can be prepared by anion-exchange f.p.l.c. on Mono Q HR. This Protein G' has a pI of 4.19 and SDS/PAGE gives an apparent anomalous  $M_r$  of 35000.

---

## INTRODUCTION

Proteins which bind to the constant (Fc) region of IgG are located on the surface of a variety of staphylococci and streptococci (Langone, 1982). Protein A from *Staphylococcus aureus* is the best known of these Fc receptor molecules, and its ability to bind IgG has been exploited in many immunochemical methods (Goding, 1978; Langone, 1982). Protein G is a bacterial cell-surface-associated protein of group C and G streptococci and binds the IgG of different subclasses of most mammalian species (Åkerström *et al.*, 1985; Guss *et al.*, 1986; Reis *et al.*, 1986).

In contrast with Protein A from *Staph. aureus*, Protein G from *Streptococcus* strain G148 binds to all four subclasses of human IgG; Protein A does not bind to the IgG<sub>3</sub> subclass (Guss *et al.*, 1986). However, Protein A has a higher overall affinity for human polyclonal IgG than does Protein G (Eliasson *et al.*, 1989), and the two proteins have complementary binding patterns (Guss *et al.*, 1986; Eliasson *et al.*, 1988). Protein G has been shown to interact with the Fab regions of IgG, but with a 10-fold lower affinity than determined for the Fc region (Björck & Kronvall, 1984) and has been reported to bind F(ab')<sub>2</sub> fragments of IgG (Erntell *et al.*, 1988). There are independent and separate binding regions for Fab and Fc fragments of IgG on the Protein G molecule, and the elongated structure of the molecule may permit simultaneous binding of both Fab and Fc (Erntell *et al.*, 1988).

The complete nucleotide sequence of the Protein G structural gene from *Streptococcus* G148 cloned by Olsson *et al.* (1987) indicates an  $M_r$  of 63294 and a pre-protein of 593 amino acids (including the N-terminal signal peptide). Protein and gene sequence analysis show that the Protein G gene has similar features to those of Protein A (Uhlén *et al.*, 1984). Both proteins consist of repetitively arranged domains (Sjödahl, 1977; Fahnestock *et al.*, 1986; Guss *et al.*, 1986). Whereas Protein G from the streptococcal strains GX7809 and G148 consist of two and three IgG-binding domains respectively (Fahnestock *et al.*, 1986; Olsson *et al.*, 1987), Protein A from a range of different *Staph. aureus* strains has five IgG-binding domains (Uhlén *et al.*, 1984; Guss *et al.*, 1985). Protein G (from strain G148) also binds

albumin at several sites which are structurally separate from the IgG-binding sites (Åkerström *et al.*, 1987; Björck *et al.*, 1987; Nygren *et al.*, 1988; Sjöbring *et al.*, 1988). Sjöbring *et al.* (1989) found Protein G from 31 strains of human group C and G streptococci had both IgG- and albumin-binding ability, which suggests both binding regions are essential to the mode of action of these bacteria. Sjöbring *et al.* (1988) isolated several IgG- and albumin-binding proteins, including a Protein G fragment of  $M_r$  14000 which binds only albumin. A Protein G fragment of  $M_r$  7500 binds IgG (Guss *et al.*, 1986), but the exact residues involved in binding IgG have not yet been identified. Despite the similarities of function between Protein A and Protein G, with the exception of a short sequence at the C-terminus of the proteins probably associated with membrane anchorage, the genes and amino acid sequences show no homology.

Heterogeneous Protein G has previously been prepared by chromatography on IgG–Sephacrose (Guss *et al.*, 1986; Eliasson *et al.*, 1988), by chromatography on DEAE-Sephadex, IgG–Sephacrose and then Sephadex G-100 (Åkerström & Björck, 1986), and by chromatography on DEAE-cellulose, Sephadex G-100 and finally IgG–Sephacrose (Björck & Kronvall, 1984). Falkenberg *et al.* (1988) prepared Protein G by h.p.l.a.(affinity)c., whereas Björck *et al.* (1987) isolated three major IgG-binding proteins from the *Escherichia coli*-cloned *Streptococcus* G148 Protein G gene by chromatography on IgG–Sephacrose and then Sephadex G-200. The major protein of  $M_r$  65000 was separated and described as 96% homogeneous.

The aim of the present study was to clone the Protein G gene in order to be able to express it in a non-pathogenic host without having to use proteolytic enzymes for release of Protein G from the streptococcal cell wall, which appears to result in degradation products (Goward & Barstow, 1989). The map of various regions on the gene coding for different functions (Guss *et al.*, 1986; Åkerström *et al.*, 1987; Sjöbring *et al.*, 1988) was used to design a Protein G' molecule that would bind only the Fc portion of IgG (and neither the Fab region nor albumin) and which would have the cell-wall-spanning and membrane-anchoring regions removed to diminish both non-specific binding of IgG and potential problems with expression (Shuttleworth *et al.*, 1987).

---

Abbreviations used: PBS, phosphate-buffered saline (composition and pH given in the text); HSA, human serum albumin; SPG, *Streptococcus* G148 Protein G.

\* To whom correspondence and reprint requests should be addressed.

We further describe a simple method to prepare homogeneous Protein G' free from fragments of the major protein molecule.

## EXPERIMENTAL

### Materials

Radiochemicals were from Amersham International. X-Omat S X-ray film was from Kodak. Deoxy- and dideoxy-nucleotides, DNA ligase, restriction endonucleases and other DNA-modifying enzymes were from Boehringer. Agarose, acrylamide, bisacrylamide and phenol were from Bethesda Research Laboratories. Chromatography media were from Pharmacia-LKB (Uppsala, Sweden). Immunoglobulins were from Sigma or ICN Biomedicals Ltd., High Wycombe, Bucks., U.K. All other reagents were from Sigma or BDH. Nitrocellulose was purchased from Anderman and Co., Kingston-upon-Thames, Surrey, U.K.

### Media and culture conditions

*E. coli* was cultured in 2xYT broth [2% (w/v) tryptone/1% (w/v) yeast extract/1% (w/v) NaCl] at 37 °C. Media were solidified with 2% (w/v) Bacto-agar (Difco). HT-agar for M13 overlays contained 1% (w/v) tryptone, 0.8% (w/v) NaCl and 0.8% (w/v) Bacto-agar (Difco). Ampicillin (25–50 µg/ml) was used where necessary for the selection and growth of transformants. Functional β-galactosidase was detected by addition of 5-bromo-4-chloroindolyl β-D-galactoside to a final concentration of 600 µg/ml and, where necessary, isopropyl β-D-thiogalactopyranoside to a final concentration of 200 µg/ml.

*E. coli* containing the recombinant Protein G' gene were grown in a 400-litre fermentation on a medium containing yeast extract, Casamino acids and glycerol plus trace elements. The pH was maintained at 7.0 with H<sub>3</sub>PO<sub>4</sub> or NaOH and the temperature at 37 °C. After 8 h of growth the temperature in the vessel was rapidly reduced to below 10 °C and the bacteria were harvested by centrifugation in a Westfalia KA25 centrifuge at a flow rate of 250 litres/h. The cell paste was quick-frozen and stored at –20 °C.

### Isolation of DNA

*Streptococcus* G148 genomic DNA was isolated essentially as described by Guss *et al.* (1986). Plasmids were purified from *E. coli* by Brij lysis (Clewell & Helinski, 1969) and CsCl/ethidium bromide density-gradient centrifugation (Radloff *et al.*, 1967). A rapid small-scale plasmid-isolation technique (Holmes & Quigley, 1981) was used for screening procedures.

### Genetic manipulation procedures

DNA-modifying enzymes were used in the buffers and under the conditions recommended by the supplier (Boehringer). Transformation of *E. coli* was essentially as described by Cohen *et al.* (1972). Electrophoresis of DNA fragments was performed on vertical 1% (w/v)-agarose slab gels in Tris/acetate buffer (40 mM-Tris/20 mM-sodium acetate/2 mM-EDTA, adjusted to pH 7.9 with acetic acid). DNA fragment sizes were estimated by comparison with fragments of λ-phage DNA digested with the restriction endonuclease *Hind*III. DNA fragments were purified by electroelution essentially as described by McDonnell *et al.* (1977). Southern transfers and hybridization conditions were performed by previously described procedures (Southern, 1975). Site-directed mutagenesis was performed by the methods of Carter *et al.* (1985).

### Nucleotide sequencing

Nucleotide sequences were determined by the chain-termination procedure of Sanger *et al.* (1980) on M13 templates generated by the sonication procedure of Deininger (1983). The oligo-

nucleotide sequencing primers were synthesized by using an Applied Biosystems 380B DNA synthesizer. All sequences of the coding and non-coding strands were confirmed with adequate overlap between contiguous sequences. Sequence data were assembled into contiguous sequence by using the computer programs of DNASTAR (Madison, WI, U.S.A.).

### Large-scale purification

**Disruption of cells.** A 400-litre fermentation yielded 18 kg of cell paste. A 2 kg portion of cell paste was thawed at 4 °C in 4 litres of 50 mM-Hepes/NaOH buffer, pH 8.0, containing 250 mM-NaCl and 0.5 mg of DNAase/litre. The thawed suspension was disrupted with a 15M-8BA Manton-Gaulin homogenizer at 550 kg/cm<sup>2</sup>. The homogenate was centrifuged for 45 min at 13000 g at 4 °C and passed through a 0.45 µm-pore-size filter to remove fine particulate matter.

**Affinity chromatography on IgG-Sepharose 4B.** IgG-Sepharose 4B was prepared by coupling porcine IgG to CNBr-activated Sepharose 4B (5 mg of IgG/ml of matrix). The cell-extract supernatant was applied to a 1-litre IgG-Sepharose 4B column [17.5 cm long × 9 cm internal diameter (i.d.)] operated at a linear flow rate of 30 cm/h and equilibrated with 50 mM-Hepes/NaOH buffer, pH 8.0, containing 250 mM-NaCl. Unbound protein was removed with equilibration buffer and Protein G' was eluted in a single peak with 100 mM-glycine/HCl, pH 2.0. The eluate was immediately made 20 mM with respect to Tris and the pH was adjusted to 7.5 with NaOH. Protein G' was concentrated with an Amicon CH2A ultrafiltration unit fitted with an H10P10 hollow-fibre cartridge.

**Anion-exchange chromatography on Q-Sepharose FF.** The concentrated IgG-Sepharose 4B eluate was further purified in portions on a 32 ml column (16 cm × 1.6 cm i.d.) of Q-Sepharose FF operated at a linear flow rate of 60 cm/h and equilibrated with 20 mM-Tris/HCl, pH 7.5. The column was washed with equilibration buffer and the protein was eluted with a linear gradient of 0–500 mM-NaCl in 20 mM-Tris/HCl, pH 7.5.

**Anion-exchange chromatography on Mono Q.** The major fractions purified as above were separated by f.p.l.c. (Pharmacia LKB, Sweden) on an 8 ml Mono Q HR 10/10 column. The column was operated at a linear flow rate of 230 cm/h and equilibrated with 20 mM-Tris/HCl, pH 7.5. A 500 µl portion of concentrated IgG-Sepharose 4B eluate was applied. The column was washed with equilibration buffer and protein was eluted with a linear gradient of 0–250 mM-NaCl in 20 mM-Tris/HCl, pH 7.5.

### Protein assay

Protein concentrations were determined by the Folin method of Lowry *et al.* (1951), with bovine serum albumin as the standard. The protein content of column eluates was also monitored by absorbance at 280 nm.

### Protein G assay

Protein G concentration was determined functionally by an e.l.i.s.a. procedure. It was assumed that uni-, bi- or poly-valent Protein G fragments are detected, since a fragment containing one IgG-binding domain can be separated by chromatography on IgG-Sepharose (Guss *et al.*, 1986). The method was adapted from that described for Protein A (Warnes *et al.*, 1986). Microtitre plates were coated with capture antibody (human polyclonal IgG) at 2 µg/ml in 15 mM-Na<sub>2</sub>CO<sub>3</sub>/35 mM-NaHCO<sub>3</sub>, pH 9.6 (100 µl/well) at room temperature overnight. The plates were washed six times after each stage of the assay with phosphate-buffered saline (PBS): 8 mM-Na<sub>2</sub>HPO<sub>4</sub>/1.5 mM-KH<sub>2</sub>PO<sub>4</sub>/137 mM-

NaCl/2.7 mM-KCl, pH 7.4, containing 0.1% (v/v) Tween 20 ('polyoxyethylenesorbitan monolaurate'). This solution was also used as a diluent in all stages of the procedure. Protein G standards, prepared by chromatography on IgG-Sepharose 4B followed by anion-exchange f.p.l.c. on Mono Q HR, and other samples, were diluted as appropriate. Samples (100  $\mu$ l) of 2-fold dilution series of each sample were transferred to the microtitre plate and incubated for 90 min at room temperature. Bound Protein G was detected by incubation with goat anti-(rabbit IgG) IgG-horseradish peroxidase conjugate for 90 min at 20 °C. The bound conjugate was incubated for 30 min at room temperature with 0.1% (w/v) 5-aminosalicylic acid/0.006% (v/v) H<sub>2</sub>O<sub>2</sub> in 50 mM-Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 6.0, and the absorbance of the coloured product was determined at 450 nm by using a Titertek Multiscan MCC automatic plate reader. The Protein G concentration was determined by comparison of the samples with Protein G standard. The human serum albumin (HSA) binding capacity of the Protein G was determined by substitution of HSA for the capture antibody.

#### PAGE

Acrylamide (12.5%, w/v) slab gels were run in an LKB vertical electrophoresis unit (Laemmli, 1970). Proteins were stained with Coomassie Brilliant Blue R-350, and protein bands were scanned with a Chromoscan-3 laser optical densitometer (Joyce-Loebl, Gateshead, Tyne and Wear, U.K.), to estimate the apparent  $M_r$ .

#### Western blotting

Proteins were applied to nitrocellulose membranes by electrophoretic transfer from SDS/polyacrylamide gels as described by Towbin *et al.* (1979). The nitrocellulose membranes were incubated with <sup>125</sup>I-labelled human polyclonal IgG or HSA to detect the Protein G. Autoradiography was performed at -70 °C with Kodak X-Omat S X-ray film.

#### Determination of pI

The pI of Protein G was determined with a Pharmacia PhastGel apparatus and broad-pH-range gels (pH 3.5–9.5), followed by narrow-pH-range gels (pH 4.0–6.5) as described by the manufacturer (Pharmacia-LKB). The appropriate Pharmacia-LKB calibration protein kits were used, and the pI was estimated from densitometer scans of the protein bands.

#### N-Terminal sequencing

N-Terminal sequence analysis was performed to locate the proteins on the gene sequence. Sequences were determined on an Applied Biosystems 470A protein sequencer by automated Edman phenylthiohydantoin degradation (Hunkapiller *et al.*, 1983). Protein G samples were dialysed against 50 mM-NaCl, and about 500 pmol was applied to the gas-phase sequencer. The equipment was operated essentially according to the manufacturer's instructions. Repetitive Edman degradations provided sequential removal of amino acids from the peptide, which were identified by using reversed-phase h.p.l.c. (Hunkapiller & Hood, 1983).

## RESULTS AND DISCUSSION

#### Cloning and characterization of the Protein G gene from *Streptococcus* G148

The strategy to clone the *Streptococcus* G148 Protein G (SPG) gene was to use the previously determined incomplete DNA sequence (Guss *et al.*, 1986) to design specific oligonucleotide probes for detection of the gene. Thus two synthetic oligo-

nucleotides were constructed; SPG1 (sequence 5'-GGTAAAA-CATTGAAAGGCGAA) specific for the C repeat regions (Guss *et al.*, 1986) and SPG2 (sequence 5'-AAATATGGAGT-AAGTGACTAT) specific for the A repeat regions, each being represented three times in the *Streptococcus* G148 sequence (Fig. 1a).

*Streptococcus* GX7805 was shown to have a 2.3 kb *Hind*III fragment containing the SPG gene (Fahnestock *et al.*, 1986; Filpula *et al.*, 1987). As the nucleotide sequence of GX7805 was shown to be identical with that of the SPG gene (Guss *et al.*, 1986; Filpula *et al.*, 1987; Olsson *et al.*, 1987) we attempted to clone the entire SPG gene as a single *Hind*III fragment. *Hind*III-digested fragments of genomic DNA isolated from *Streptococcus* G148 were subjected to Southern-blot analysis (Southern, 1975) and probed independently with <sup>32</sup>P-labelled oligonucleotides SPG1 and SPG2. However, both probes hybridized to a 4.2 kb fragment. *Streptococcus* G148 genomic DNA was digested with *Hind*III, separated by electrophoresis on a 1% (w/v)-agarose gel, and DNA fragments of 4.0–4.4 kb were excised and purified by electroelution. The purified genomic DNA fragments were ligated to *Hind*III dephosphorylated plasmid vector pUC8 and transformed into *E. coli* TG2.

In all 1000 colonies were probed by colony hybridization *in situ* (Grunstein & Hogness, 1975) using the SPG1 and SPG2 oligonucleotides, and nine positives were detected. Plasmid DNA was isolated from four by CsCl density-gradient centrifugation (plasmids pSPG2, 3, 5 and 6) and characterized by restriction-endonuclease analysis. Clones pSPG3, 5 and 6 showed identical restriction-endonuclease cleavage patterns after digestion with *Pst*I, *Eco*RI, *Dra*I and *Hind*III. Single and double restriction enzyme digests were performed on pSPG3 and a 'crude' restriction endonuclease cleavage map was constructed (Fig. 1b), which correlated well with the restriction-endonuclease cleavage maps of previously cloned Protein G genes (Guss *et al.*, 1986; Fahnestock, 1987). The SPG gene in the plasmid construct pSPG3 was found to be in the opposite orientation to the *lac* promoter of the vector pUC8; thus expression of Protein G from this clone was from its own promoter. The clone produced a protein which bound human polyclonal IgG and HSA, as confirmed by Western-blot analysis and e.l.i.s.a.

#### Sequencing of clone pSPG3

The 4.2 kb *Hind*III fragment from pSPG3 (Fig. 1c, i) was purified by gel electroelution; M13 templates isolated from sonicated fragments of this circularized *Hind*III fragment were DNA-sequenced. Three changes in the sequence were found from that previously published (Guss *et al.*, 1986; Olsson *et al.*, 1987), as shown in Fig. 2. All were located in the 5' non-coding region of the gene; two of these were direct changes, whereas the third was an insertion.

#### Site-directed mutagenesis of the Protein G gene

Attempts to clone the entire 4.2 kb *Hind*III fragment from pSPG3 directly into M13 repeatedly failed, as did attempts to subclone the large blunt-ended 2.1 kb *Dra*I-*Hind*III fragment directly into M13. However, it proved possible to clone the blunt-ended *Dra*I-*Hind*III fragment into *Sma*I-cleaved pMTL22. The Protein G gene was then excised on a *Bam*HI-*Bgl*II fragment, ligated into *Bam*HI dephosphorylated M13mp8 and transformed into *E. coli* TG2. Templates were made from 12 plaques, and the DNA was sequenced by using oligonucleotides spanning the entire gene as primers. Two templates (9 and 13) were found to contain the Protein G gene, but in different orientations from the *lac* promoter of M13.

Deletion of the cell-wall and membrane-spanning regions from the translated protein was achieved by synthesis of an oligo-

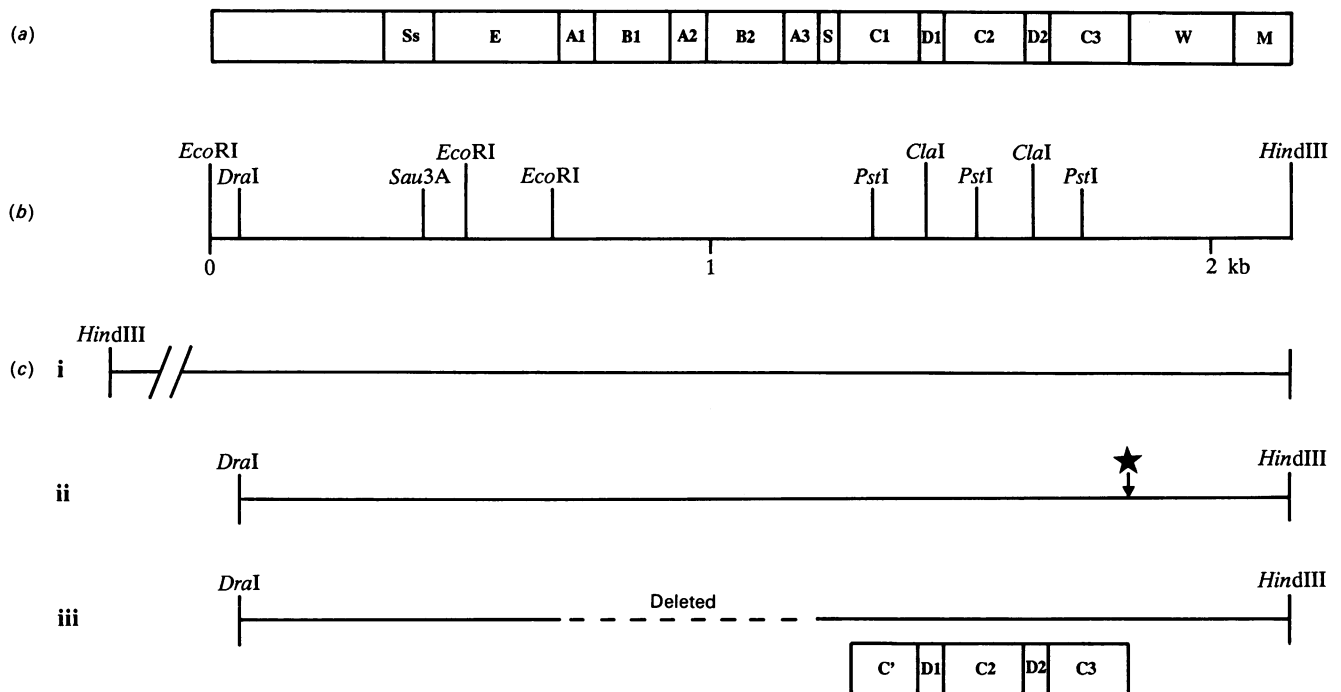


Fig. 1. Schematic representation of Protein G, restriction map and structure of the inserts of the Protein G gene

(a) Schematic representation of the Protein G molecule. Ss is the signal sequence, A and/or B regions are responsible for albumin binding, C regions are responsible for IgG binding, D and S are spacer regions, W is the cell-wall-spanning region and M is the hydrophobic membrane-anchoring region. (b) Restriction map of the 2.16 kb fragment containing the Protein G gene. (c) Structure of the inserts of the Protein G gene: i, full-length 4.2 kb recombinant Protein G, pSPG3 insert; ii, insert (pMSPG594-12) with regions coding for cell-wall spanning and membrane anchoring deleted. A stop codon was introduced at the position marked ★; iii, insert (pMSPG631) used for the production strain with regions coding for albumin binding deleted. A schematic representation of the Protein G' is shown adjacent to the proposed site for initiation and termination of translation.

nucleotide (5'-CCTCTGTAACCTTATTCAGTT) in which, by site-directed mutagenesis, the ATG codon between nucleotide positions 1259 and 1261 was replaced with the stop codon TAA (Fig. 2). To prevent read-through, a guanine residue was inserted at nucleotide position 1262 after the TAA stop codon to alter the reading frame of the gene and introduce further stop codons downstream (Fig. 2). Mutants were confirmed by DNA sequencing. The mutated gene was designated pMSPG594-12 (Fig. 1c, ii). The albumin-binding domains were deleted from pMSPG594-12 by looping out the relevant regions with an oligonucleotide (5'-TAAAATTCATCTATGAAGAATTCT-TTAAG) that hybridized to the sequence 15 bp before the A1 domain and 15 bp after the A3 domain. Removal of the albumin-binding domains was confirmed by DNA sequencing of the final mutant, pMSPG631 (Protein G'), using primers from the mutation (Fig. 1c, iii).

#### Construction of a production strain expressing the recombinant Protein G'

The recombinant Protein G' gene was removed from pMSPG631 on a *KpnI*-*HindIII* fragment and ligated into *KpnI*-*HindIII*-cleaved dephosphorylated pMTL22. After transformation into *E. coli* TG2, 250 recombinant clones were screened for the Protein G' gene by colony hybridization *in situ*, and a positive colony (pPSPG29) was cultured to purify and quantify any translated product.

Cell-free extracts from 100 ml of culture were passed through a 1 ml column (1.6 cm × 0.9 cm i.d.) of IgG-Sepharose 4B equilibrated and washed with 50 mM-Hepes/NaOH, containing 250 mM-NaCl. Protein G' was eluted with 100 mM-glycine/HCl, pH 2.0; it showed no binding to HSA and gave only one

predominant protein, of apparent  $M_r$  35000, which bound to human polyclonal IgG.

#### Nucleotide-sequence analysis of pSPG29

The original aim of this work was to isolate a derivative of the Protein G gene and to express a protein that lacked the portion containing albumin-binding domains A1, B1, A2, B2 and A3 and terminated after the C3 domain (Fig. 1). However, *N*-terminal protein sequence analysis (Fig. 2) showed translation was initiated several residues into one of the C domains. To clarify this anomaly, the nucleotide sequence of clone pSPG29 was determined. The *Bam*HI-*Bgl*III fragment from pSPG29 containing the Protein G' was purified by gel electroelution; M13 templates isolated from sonicated fragments of this circularized fragment were sequenced (Fig. 2). Sequencing confirmed the two mutations introduced by site-directed mutagenesis and the three differences already discussed that existed between our sequence and the sequences previously published (Guss *et al.*, 1986; Olsson *et al.*, 1987). However, a major additional and unsought change was the deletion of a single guanine residue from the original sequence between nucleotide positions 564 and 565 (Fig. 2), which changed the reading frame of the gene. The deletion was found to have probably occurred during construction of pMSPG594-12. As a consequence, the open reading frame of the gene was not preserved, and stop codons were introduced immediately downstream of this deletion (Fig. 2). This accounts for the absence of full-length Protein G. The TTG codon at nucleotide positions 705-707 was found (by *N*-terminal sequence analysis of pure protein) to have acted as an alternative initiation codon within the gene, giving rise to a functional IgG-binding protein (Protein G').

*DraI*  
AAAAAGCTCTGTTTCTTAAAGAAGAAAATAATTGTTGAAAAATTATAGAAAAAT 54

1 2

CATTTTATACTAATGAAATAACATAAGGCTAAATTGCTGAGGTGATGATAGGAGATTTATTTGTAAGGATTCCTTAATTTTATTAATCAACAAAAATTGATAGAAAAA 165

3

TTAATGAAATCCTTGATTAAATTTTATAAGTTGTATAATAAAAAAGTGAATTATTAATCGTAGTTTCAAATTTGTCGGCTTTTATATATGCTGCTGGCATATAAAAT 276

Ss

AAAAAAGGAGAAAAA ATG GAA AAA GAA AAA AAG GTA AAA TAC TTT TTA CGT AAA TCA GCT TTT GGG TTA GCA TCC GTA TCA GCT GCA 363  
Met Glu Lys Glu Lys Lys Val Lys Tyr Phe Leu Arg Lys Ser Ala Phe Gly Leu Ala Ser Val Ser Ala Ala

Sau3A E

TTT TTA GTG GGA TCA ACG GTA TTC GCT GTT GAC TCA CCA ATC GAA GAT ACC CCA ATT ATT CGT AAT GGT GGT GAA TTA ACT AAT 447  
Phe Leu Val Gly Ser Thr Val Phe Ala Val Asp Ser Pro Ile Glu Asp Thr Pro Ile Ile Arg Asn Gly Gly Glu Leu Thr Asn

EcoRI

CTT CTG GGG AAT TCA GAG ACA ACA CTG GCT TTG CGT AAT GAA GAG AGT GCT ACA GCT GAT TTG ACA GCA GCA GCG GTA GCC GAT 531  
Leu Leu Gly Asn Ser Glu Thr Thr Leu Ala Leu Arg Asn Glu Glu Ser Ala Thr Ala Asp Leu Thr Ala Ala Ala Val Ala Asp

4 5

ACT GTG GCA GCA GCG GCA GCT GAA AAT GCT GGG CAG CAG CTT GGG AAG CAG CGG CAG CAG CAG ATG CTC TAG CAAAAGCCAAAGCAG 618  
Thr Val Ala Ala Ala Ala Glu Asn Ala Gly Gln Gln Leu Gly Lys Gln Arg Gln Gln Gln Met Leu End

EcoRI 6 C'

ATGCCCTTAAAGAATCAACATAGATGAAATTTTAACTGCATTACCTAAGACTGACACTTACAAATTAATCCTTAATGGTAAAACA 722  
End End End End TG AAA GGC GAA ACA ACT  
Leu Lys Gly Glu Thr Thr  
Met

PstI

ACT GAA GCT GTT GAT GCT GCT ACT GCA GAA AAA GTC TTC AAA CAA TAC GCT AAC GAC AAC GGT GTT GAC GGT GAA TGG ACT TAC 806  
Thr Glu Ala Val Asp Ala Ala Thr Ala Glu Lys Val Phe Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp Thr Tyr

DI ClaI C2

GAC GAT GCG ACT AAG ACC TTT ACA GTT ACT GAA AAA CCA GAA GTG ATC GAT GCG TCT GAA TTA ACA CCA GCC GTG ACA ACT TAC 890  
Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu Lys Pro Glu Val Ile Asp Ala Ser Glu Leu Thr Pro Ala Val Thr Thr Tyr

PstI

AAA CTT GTT ATT AAT GGT AAA ACA TTG AAA GGC GAA ACA ACT ACT GAA GCT GTT GAT GCT GCT ACT GCA GAA AAA GTC TTC AAA 974  
Lys Leu Val Ile Asn Gly Lys Thr Leu Lys Gly Glu Thr Thr Thr Glu Ala Val Asp Ala Ala Thr Ala Glu Lys Val Phe Lys

D2

CAA TAC GCT AAC GAC AAC GGT GTT GAC GGT GAA TGG ACT TAC GAC GAT GCG ACT AAG ACC TTT ACA GTT ACT GAA AAA CCA GAA 1058  
Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp Thr Tyr Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu Lys Pro Glu

ClaI C3

GTG ATC GAT GCG TCT GAA TTA ACA CCA GCC GTG ACA ACT TAC AAA CTT GTT ATT AAT GGT AAA ACA TTG AAA GGC GAA ACA ACT 1142  
Val Ile Asp Ala Ser Glu Leu Thr Pro Ala Val Thr Thr Tyr Lys Leu Val Ile Asn Gly Lys Thr Leu Lys Gly Glu Thr Thr

PstI

ACT AAA GCA GTA GAC GCA GAA ACT GCA GAA AAA GCC TTC AAA CAA TAC GCT AAC GAC AAC GGT GTT GAT GGT GTT TGG ACT TAT 1226  
Thr Lys Ala Val Asp Ala Glu Thr Ala Glu Lys Ala Phe Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly Val Trp Thr Tyr

8 9 10 11

GAT GAT GCG ACT AAG ACC TTT ACG GTA ACT GAA TAA GGTACAGAGGTTCTCTGGTGATGCACCAACTGAACCAGAAAAACCAGAAGCAAGTATCCCTCT 1325  
Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu End End End End

TGTTCCGTTAACTCCTGCAACTCCAATTGCTAAAGATGACGCTAAGAAAGACGATACTAAGAAAGAAGATGCTAAAAAACAGAAAGCTAAGAAAGACGCTAAGAAAGC 1436

TGAAACTCTTCTACAACCTGTAAGGAAGCAACCCATTCTTACAGCAGCTGCGCTTGCAGTAATGGCTGGTGGGGTGCTTTGGGGTTCGCTTCAAAACGTAAGAAGA 1547

HindIII

CTAATTGTCATTATTTTGGACAAAAAGCT 1576

**Fig. 2. Nucleotide sequence and deduced amino acid sequence of the Protein G' gene**

The sequence was compared with that of Olsson *et al.* (1987). The C' region corresponds to C1, but translation of the protein was initiated part of the way into this domain. The boxed regions show the stop codons. Differences between this sequence and that described by Olsson *et al.* (1987) are: 1, substitution of G for C; 2, insertion of C; 3, substitution of T for A; 4, deletion of G and a corresponding shift in the reading frame; 5, stop codon introduced due to the shift in reading frame; 6, HSA-binding domains deleted; 7, translation initiated again; 8, stop codon introduced; 9, insertion of G to prevent any possible continuation of translation by introduction of putative stop codons, 10 and 11. Differences 1, 2 and 3 are also found in the native Protein G gene. The N-terminal amino acid sequence of the purified protein is indicated by a continuous line.

Guss *et al.* (1986) reported translation of two proteins from an *EcoRI*–*HindIII* fragment, both of which had the same *N*-terminal sequence as the Protein G' we have prepared; translation was thought to be initiated in the system using the TTG codon at nucleotide positions 705–707 and 915–917 (Fig. 2). Guss *et al.* (1986) also postulated a poor ribosomal binding sequence of GGT complementary to the 16S rRNA of prokaryotes upstream from the TTG start codon. However, the –10 and –35 promoter sequences recognized by *E. coli* that Guss *et al.* (1986) suggested, must be ruled out in our gene, since they lie within the A and B repeat regions which have been removed by site-directed mutagenesis.

### Purification

The protein was heterogeneous after affinity chromatography, possibly due to proteolytic 'nicking' of the molecule. Anion-exchange chromatography on Q-Sepharose FF removed some of the smaller contaminating Protein G' fragments, but anion-exchange chromatography on Mono Q HR (Fig. 3) was used to prepare Protein G' with an overall recovery of 45%, which gave a single protein band after SDS/PAGE and isoelectric focusing (Figs. 4 and 5). Results in Table 1 show Protein G' was expressed at 0.4% total soluble cell protein; and in different cultures results have ranged between 0.3 and 0.6%.

### SDS/PAGE

SDS/PAGE showed the IgG–Sepharose 4B eluate contained a heterogeneous mixture of protein (Fig. 4), but Western-blot analysis with <sup>125</sup>I-human polyclonal IgG showed the minor bands to bind IgG and therefore they may be post-translational products. The apparent *M<sub>r</sub>* of 35000 is in conflict with the predicted *M<sub>r</sub>* of about 20000. Other authors observed that recombinant IgG-binding proteins had lower mobilities on SDS/PAGE than the full amino acid sequence would predict (Guss *et al.*, 1986; Nygren *et al.*, 1988). This anomalous behaviour may be attributable to an excessively elongated structure of the molecule in SDS, low SDS binding to the protein or a C-terminal post-translational modification (limited proteolytic cleavage) of the protein removing disproportionate capacity for SDS binding.

### Specificity of binding

The Protein G' was shown to bind Fc fragments, but not Fab fragments, of human polyclonal IgG by e.l.i.s.a. using Fc or Fab as the capture molecule. The autoradiograph of a Western blot probed with <sup>125</sup>I-HSA showed no evidence of HSA binding; this

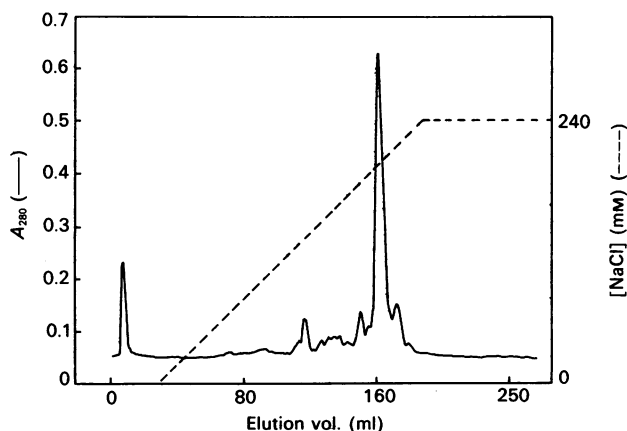


Fig. 3. Chromatography on Mono Q HR

The major peak was collected as a single fraction and was retained on the basis of electrophoretic homogeneity.

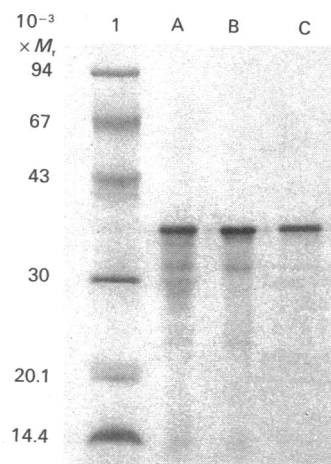


Fig. 4. SDS/PAGE

The IgG–Sepharose 4B (A), Q-Sepharose FF (B) and Mono Q HR (C) eluates were electrophoresed on a 12.5% (w/v) acrylamide gel in the presence of SDS with the following standards (1): phosphorylase *b* (*M<sub>r</sub>*, 94000), bovine serum albumin (*M<sub>r</sub>*, 67000), ovalbumin (*M<sub>r</sub>*, 43000), carbonic anhydrase (*M<sub>r</sub>*, 30000), soybean trypsin inhibitor (*M<sub>r</sub>*, 20100) and  $\alpha$ -lactalbumin (*M<sub>r</sub>*, 14400).

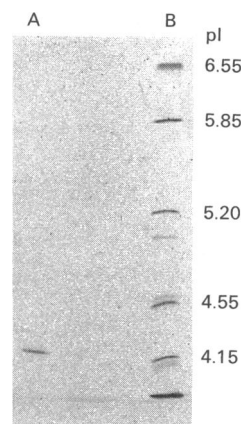


Fig. 5. Isoelectric focusing

Protein G' (A) was focused on a Phastgel IEF (pH 4.0–6.5) with standards (B): glucose oxidase (pI 4.15), soybean trypsin inhibitor (pI 4.55),  $\beta$ -lactoglobulin A (pI 5.20), bovine carbonic anhydrase B (pI 5.85), and human carbonic anhydrase B (pI 6.55).

was further confirmed by e.l.i.s.a. using HSA as the capture molecule when again no HSA binding was observed.

### pI

The pI of Protein G' was shown to be 4.19 by using narrow-range (pH 4.0–6.5) isoelectric-focusing gels (Fig. 5), and the theoretical pI was calculated as 4.20 by the computer program of DNASTAR Inc. The pI of Protein G cleaved from the cell walls of *Streptococcus* G148 with papain was determined to be less than 3.5 (Åkerström & Björck, 1986).

### *N*-Terminal sequence

The first 35 *N*-terminal amino acid residues of the Protein G' molecule were sequenced and are indicated by the continuous line in Fig. 2. The sequence starts with an *N*-terminal protein-sequence-identified methionine residue from the TTG codon. TTG is a common start codon in Gram-positive bacteria (Uhlén *et al.*, 1983), but is uncommon in *E. coli*. However, Protein A is well expressed from a TTG initiation codon in *E. coli* (Shuttleworth *et al.*, 1987). Translation was initiated part-way in to one

**Table 1. Purification of Protein G'**

| Step                     | Volume (ml) | Total protein (mg) | Protein G' ( $\mu$ g) | Specific amount ( $\mu$ g/mg of total protein) | Recovery (%) |
|--------------------------|-------------|--------------------|-----------------------|--|--------------|
| IgG-Sepharose 4B         |             |                    |                       |  |              |
| Cell-free extract        | 4500        | 163000             | 655600                | 4  | 100          |
| IgG-Sepharose 4B eluate  | 1100        | 560                | 471000                | 841  | 72           |
| Q-Sepharose FF           |             |                    |                       |  |              |
| IgG-Sepharose 4B eluate* | 45          | 189                | 159000                | 841  | 100          |
| Q-Sepharose FF eluate    | 41          | 102                | 92000                 | 902  | 58           |
| Mono Q HR                |             |                    |                       |  |              |
| IgG-Sepharose 4B eluate* | 0.5         | 6.72               | 5650                  | 841  | 100          |
| Mono Q HR eluate         | 8.0         | 3.52               | 3560                  | 1011   | 63           |

\* The IgG-Sepharose 4B eluate was concentrated before application to Q-Sepharose FF and Mono Q HR.

of the IgG-binding domains; however, Guss *et al.* (1986) have shown that such translated molecules still bind IgG efficiently. Initiation of translation is likely to have occurred at the first TTG codon, in view of the  $M_r$  of the protein expressed by *E. coli*, so the Protein G' probably has three functional IgG-binding domains (Fig. 1c, iii). Although the most likely initiation position of Protein G' is indicated in Fig. 2, giving a protein with three IgG-binding domains and an  $M_r$  of 20000 (apparent  $M_r$  35000 by SDS/PAGE), it is conceivable that a molecule with the same *N*-terminal sequence, with two IgG-binding domains, the same theoretical pI and an  $M_r$  of 12500 could be produced by initiation of translation from a TTG codon starting at nucleotide position 915 instead of 705. The *N*-terminal protein sequence and present data are not adequate to distinguish between these possibilities other than to note that 2-fold differences between real and apparent  $M_r$  on SDS/polyacrylamide gels are known, whereas 3-fold differences are not. Protein G with three IgG-binding domains has been demonstrated to have greater affinity for human polyclonal IgG than Protein G with two IgG-binding regions (Eliasson *et al.*, 1989). Guss *et al.* (1986) reported that *N*-terminal sequence analyses of the two gene products of their Protein G gene suggest two different TTG codons may be recognized in *E. coli* for initiation of translation to yield two proteins. However, the same TTG codon may have been recognized for initiation of translation, and the smaller molecule may have been generated by *C*-terminal proteolytic-enzyme cleavage. In contrast, we found only one major gene product, even though there are three TTG codons available for initiation of translation.

We thank Dr. S. P. Chambers for culture of bacteria and Mr. D. R. Hartwell for oligonucleotide synthesis and *N*-terminal sequence analysis.

## REFERENCES

- Åkerström, B. & Björck, L. (1986) *J. Biol. Chem.* **261**, 10240–10247
- Åkerström, B., Brodin, T., Reis, K., & Björck, L. (1985) *J. Immunol.* **135**, 2589–2592
- Åkerström, B., Nielson, E. & Björck, L. (1987) *J. Biol. Chem.* **262**, 13388–13391
- Björck, L. & Kronvall, G. (1984) *J. Immunol.* **133**, 969–974
- Björck, L., Kastern, W., Lindahl, G. & Wideback, K. (1987) *Mol. Immunol.* **24**, 1113–1122
- Carter, P., Bedouelle, H. & Winter, G. (1985) *Nucleic Acids Res.* **13**, 4431–4443
- Clewell, D. B. & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. U.S.A.* **62**, 1159–1166
- Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2110–2114
- Deininger, P. L. (1983) *Anal. Biochem.* **129**, 216–223
- Eliasson, M., Olsson, A., Palmcrantz, E., Wiberg, K., Inganäs, M., Guss, B., Lindberg, M. & Uhlén, M. (1988) *J. Biol. Chem.* **263**, 4323–4327
- Eliasson, M., Andersson, R., Olsson, A., Wigzell, H. & Uhlén, M. (1989) *J. Immunol.* **142**, 575–581
- Erntell, M., Myhre, E. B., Sjöbring, U. & Björck, L. (1988) *Mol. Immunol.* **25**, 121–126
- Fahnestock, S. R. (1987) *Trends Biotechnol.* **5**, 79–83
- Fahnestock, S. R., Alexander, P., Nagle, J. & Filpula, D. (1986) *J. Bacteriol.* **167**, 870–880
- Falkenberg, C., Björck, L., Åkerström, B. & Nilsson, S. (1988) *Biomed. Chromatogr.* **2**, 221–225
- Filpula, D., Alexander, P. & Fahnestock, S. R. (1987) *Nucleic Acids Res.* **15**, 7210
- Goding, J. W. (1978) *J. Immunol. Methods.* **20**, 241–253
- Goward, C. R. & Barstow, D. A. (1989) *Microbiol. Immunol.* **33**, 123–127
- Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961–3965
- Guss, B., Leander, K., Hellman, U., Uhlén, M., Sjöquist, J. & Lindberg, M. (1985) *Eur. J. Biochem.* **153**, 579–585
- Guss, B., Eliasson, M., Olsson, A., Uhlén, M., Frej, A.-K., Jornvall, H., Flock, J.-I. & Lindberg, M. (1986) *EMBO J.* **5**, 1567–1575
- Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114**, 193–197
- Hunkapiller, M. W. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 486–494
- Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 399–413
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Langone, I. I. (1982) *Adv. Immunol.* **32**, 157–252
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- McDonnell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146
- Nygren, P.-Å., Eliasson, M., Abrahamsén, L., Uhlén, M. & Palmcrantz, E. (1988) *J. Mol. Recog.* **1**, 69–74
- Olsson, A., Eliasson, M., Guss, B., Nilsson, B., Hellman, U., Lindberg, M. & Uhlén, M. (1987) *Eur. J. Biochem.* **168**, 319–324
- Radloff, R., Bauer, W. & Vinograd, J. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **57**, 1514–1521
- Reis, K. J., Hansson, H. F. & Björck, L. (1986) *Mol. Immunol.* **23**, 425–431
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Shuttleworth, H. L., Duggleby, C. J., Jones, S. A., Atkinson, T. & Minton, N. P. (1987) *Gene* **58**, 283–295
- Sjödahl, J. (1977) *Eur. J. Biochem.* **73**, 343–351
- Sjöbring, U., Falkenberg, C., Nielson, E., Åkerström, B. & Björck, L. (1988) *J. Immunol.* **140**, 1595–1599
- Sjöbring, U., Björck, L. & Kastern, W. (1989) *Mol. Microbiol.* **3**, 319–327
- Southern, E. M. (1975) *J. Mol. Biol.* **143**, 503–517
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- Uhlén, M., Nilsson, B., Guss, B., Lindberg, M., Gatenback, S. & Philipson, L. (1983) *Gene* **23**, 369–378
- Uhlén, M., Guss, B., Nilsson, B., Gatenbeck, S., Philipson, L. & Lindberg, M. (1984) *J. Biol. Chem.* **259**, 1695–1702
- Warnes, A., Walkland, A. & Stephenson, J. R. (1986) *J. Immunol. Methods.* **93**, 63–70