Microbiology (2002), **148**, 2171–2179 Printed in Great Britain

Molecular analysis of a haemagglutinin of *Haemophilus paragallinarum*

Rhonda I. Hobb,[1] Hsing-Ju Tseng,[1] John E. Downes,[1] Tamsin D. Terry,[1] Patrick J. Blackall,[2] Masami Takagi[3] and Michael P. Jennings[1]

Author for correspondence: Michael P. Jennings. Tel: +61 7 3365 4879. Fax: +61 7 3365 4620. e-mail: jennings@mailbox.uq.edu.au

[1] Department of Microbiology and Parasitology, School of Molecular and Microbial Sciences, University of Queensland, St Lucia, QLD 4072, Australia

[2] Agency for Food and Fibre Sciences, Animal Research Institute, Department of Primary Industries, Yeerongpilly, Australia

[3] National Veterinary Assay Laboratory, Dobutsu Iyakuhin Kensajo, 1-15-1 Tokura, Kokubunji-shi, Tokyo, Japan

The gene encoding a haemagglutinin of *H. paragallinarum*, *hagA*, has been identified and the full-length nucleotide sequence determined. A ~39 kDa protein, recognized by an anti-haemagglutinin monoclonal antibody, mAb4D, was purified from *H. paragallinarum* strain 0083 and the N-terminal sequence obtained. The full-length nucleotide sequence was obtained by inverse PCR and the deduced amino acid sequence of the protein encoded was shown to be similar to other outer-membrane proteins of closely related organisms in the HAP group (*Haemophilus*, *Actinobacillus*, *Pasteurella*), especially the P5 protein of *Haemophilus influenzae*. The *hagA* gene was cloned into a His-tag expression vector and overexpressed in *Escherichia coli* strain M15(pREP4). The identity of the purified recombinant protein as a *H. paragallinarum* haemagglutinin was confirmed by haemagglutination of chicken red blood cells and reactivity, in a Western blot, with the monoclonal antibody specific for the serovar A haemagglutinin.

Keywords: serotyping antigen, infectious coryza, outer-membrane protein

INTRODUCTION

Infectious coryza, an acute upper respiratory tract disease of chickens, is caused by the bacterium *Haemophilus paragallinarum* (Blackall & Yamamoto, 1997). This disease is of worldwide economic significance and affects both broiler and layer flocks, manifesting primarily as a drop in egg production (10–40%) in layer flocks and retardation of growth due to decreased feed and water consumption in breeder and broiler flocks (Blackall & Yamamoto, 1997). The most common clinical signs include nasal discharge, facial oedema, lacrimation, anorexia and diarrhoea (Blackall & Yamamoto, 1997).

The haemagglutinin antigen of *H. paragallinarum* plays a key role in serotyping, immunity and pathogenicity. The two serotyping schemes, the Page (Page, 1962) and Kume (Kume *et al*., 1983a) schemes, are both performed using haemagglutination inhibition (HI) tests (Blackall & Yamamoto, 1990). The most widely used serotyping scheme, that of Page, groups *H. paragallinarum* isolates into three serovars, A, B and C. Considerable attention has also been paid to the role of haemagglutinin antigens in pathogenicity (Blackall & Yamamoto, 1997) and as protective antigens. For Page serovar A organisms, there is a close correlation between HI titre and both protection (Kume *et al*., 1980; Otsuki & Iritani, 1974) and clearance of the organism from the nostrils of vaccinated chickens (Kume *et al*., 1984). Purified haemagglutinin antigen from a serovar A organism has been shown to induce protective immunity (Iritani *et al*., 1980). For both serovar A and serovar C, the assessment of mutants lacking haemagglutination activity has shown that the haemagglutinin antigen plays a key role in colonization (Sawata & Kume, 1983; Yamaguchi *et al*., 1993). However, neither the protein sequence nor the gene encoding the haemagglutinin antigen has been identified. In this paper we report the isolation, identification and full-length sequence of a gene encoding a haemagglutinin antigen (HagA) of *H. paragallinarum*, as well as the overexpression and purification of recombinant haemagglutinin (rHagA) from *E. coli*.

Abbreviations: AP, alkaline phosphatase; HA, haemagglutination activity; HI, haemagglutination inhibition.

The GenBank accession numbers for the sequences determined in this work are AF491817–AF491827.

METHODS

Bacterial strains. The *H. paragallinarum* strains used in this study are listed in Table 1. They are serotyping reference strains obtained from the culture collection at the Animal

Table 1. *H. paragallinarum* strains used in the sequencing of the *hagA* gene

(P) indicates a reference strain for the Page serotyping scheme; (K) indicates a reference strain for the Kume serotyping scheme. Primer sets required for amplification of the *hagA* gene are indicated in the far right column. Primer set 1 consists of primers HA8/HA10; primer set 2 consists of primers HA8/HA11.

Strain	Source country	Page serovar	Kume serovar	PCR primer set
0083 (P)	USA	A	A-1	2
221 (K)	Japan	A	A-1	2
2403 (K)	Germany	A	A-2	1
E-3C (K)	Brazil	A	A-3	1
HP14 (K)	Australia	A	A-4	2
0222 (P)	USA	B	B-1	1
2671 (K)	Germany	B	B-1	1
Modesto (K, P)	USA	C	C-2	2
H-18 (K)	Japan	C	C-1	1
SA-3 (K)	South Africa	C	C-3	1
HP60 (K)	Australia	C	C-4	1

Research Institute, Brisbane, Australia. *Escherichia coli* strain M15(pREP4) was obtained from Qiagen.

Monoclonal antibody. Monoclonal antibody 4D (mAb4D), originally raised in mice against whole cells of *H. paragallinarum* serovar A strain 221 and shown to be specific for a serovar A haemagglutinin, was used in both HI assays and immunoblots in this study (Takagi *et al*., 1991a).

***H. paragallinarum* whole-cell haemagglutinin purification.** *H. paragallinarum* strain 0083 was grown overnight (37 °C/ static) in TMB, a broth medium prepared by omitting agar from Test medium supplemented with chicken serum and reduced nicotinamide adenine dinucleotide (TM/SN) (Reid & Blackall, 1987). Cells were centrifuged, washed twice in phosphate-buffered saline (PBS, pH 7·2), resuspended in 50 mM Tris/HCl, 10 % (v/v) glycerol, pH 8·0, and lysed by sonication. The cell lysate was fractionated using ammonium sulfate to precipitate proteins at 0–20 %, 20–40 % and 40 + % ammonium sulfate saturation. Precipitated proteins were resuspended and analysed by immunoblotting using mAb4D. The 0–20 %, 20–40 % and 40 + % precipitated fractions were run on a 12 % SDS-polyacrylamide gel, according to Laemmli (1970). The proteins were transferred to a nitrocellulose membrane (Protran, Schleicher and Schuell), using semi-dry transfer (Trans-blot semi-dry transfer cell, Bio-Rad) according to the manufacturer's instructions. mAb4D was used at a dilution of 1/1000 and the secondary antibody at 1/7500 (goat anti-mouse IgG–AP, Promega). Activity of the AP (alkaline phosphatase) conjugate secondary antibody was detected by incubation with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Amresco) with a development time of 16–18 h at room temperature.

N-terminal sequencing. The 0–20 % ammonium sulfate fraction was separated by SDS-PAGE and semi-purified by electroelution (Bio-Rad). The eluted ~39 kDa protein band was run on a 12 % Tris-Tricine polyacrylamide gel. The proteins were blotted onto PVDF membrane (Polyscreen PVDF Transfer Membrane, NEN Life Science Products) using CAPS buffer [10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, pH 11) with a Trans-blot semi-dry transfer cell (Bio-Rad) according to the manufacturer's instructions. The PVDF membrane was soaked in Milli Q water (Millipore) for 10 min with shaking and stained with 0·1 % (w/v) Coomassie blue R250, 50 % (v/v) methanol, 10 % (v/v) acetic acid for 5 min. The membrane was destained [50 % (v/v) methanol, 10 % (v/v) acetic acid] and rinsed in Milli Q water. The N-terminal sequence of the ~39 kDa band was obtained using a PE Biosystems 492cLC protein sequencer.

PCR and inverse PCR. Primers HA1 and HA2 (Table 2), based on N-terminal sequence and alignments of the P5/OMP regions of closely related organisms, were used to amplify the 900 bp core region of the putative haemagglutinin coding sequence from strains 0083, 0222 and Modesto. Chromosomal DNA from each strain was digested overnight with the restriction enzymes *Bfa*I or *Hin*dIII (New England Biolabs). Restriction enzymes were heat-inactivated, according to the manufacturer's specifications, and the DNA precipitated with 3 M sodium acetate and ethanol. The digested chromosomal DNA was self-ligated to form circular DNA using T4 DNA ligase (Promega). Internal primers were designed to amplify either the upstream (HA5/HA6) or downstream (HA3/HA7) sequences of the core region (according to the position of the restriction enzyme site within the core region). These 'inverse' PCR amplification products were identified on a 1 % (w/v) agarose gel and were purified and sequenced to obtain the full-length sequence of the *H. paragallinarum* haemagglutinin gene (*hagA*).

PCR primer sets 1 (HA8/HA10; Table 1) and 2 (HA8/HA11; Table 1), generated by inspection of *hagA* flanking sequences, were used to amplify the full-length *hagA* gene using *Taq* polymerase (Promega). Triplicate PCRs were performed and pooled for use as sequencing template in order to minimize the effect of *Taq* errors on sequence accuracy. The *hagA* gene was fully sequenced in 11 *H. paragallinarum* strains. Of the 11 strains, five were Page serovar A, two were Page serovar B and four were Page serovar C (Table 1).

DNA sequencing. ABI Prism Big Dye Primer Cycle Sequencing

Table 2. Sequences of oligonucleotide primers used in inverse PCR, sequencing and cloning of the haemagglutinin genes of serovars A, B and C of *H. paragallinarum*

The core region of the *hagA* gene was amplified using HA1 and HA2. Primers HA3, HA5, HA6 and HA7 were used in inverse PCR to amplify upstream and downstream regions flanking the core region and for sequencing of the gene. HA8, HA10 and HA11 primers anneal to the intergenic regions of the *hagA* gene to amplify the full-length gene (HA8 upstream, HA11 and HA10 downstream). HA12 and HA13 primers were used for cloning. The underlined sequence corresponds to the nucleotide sequence of the *hagA* gene with GenBank accession number AF491827. The GCA (bold) in HA12 encodes the first amino acid in the mature form of the protein after processing of the leader sequence. The TTA (bold) in HA13 is the stop codon.

Primer	Sequence (5′–3′)
HA1	TGTAGCTCAAGCAGCTCCACAAG
HA2	TCAAGCGATAAGTGCTTTACGACC
HA3	AACGCGAGCATAAACATC
HA5	GCTGTTGAGCTAGGTTA
HA6	AGATGCCCAGCCCGCTT
HA7	CGGTTCTGTAACTGCTGG
HA8	AAGCTTTTATTTTAGATTTATTG
HA10	CTGCTTGCACTAAGCCGTTG
HA11	CGCACGGCATTGATATTGTG
HA12	CGCGGATCC**GCA**CCACAAGCAAATACTTTC
HA13	TGCAGACGTCAACGTGTTAAGAAA**TTA**CTCG

Ready Reaction with AmpliTaq DNA Polymerase, FS′ (PE Applied Biosystems) was used for DNA sequencing. Following ethanol precipitation, samples were sent to the Australian Genomic Research Facility (AGRF) for automated sequencing by an ABI 373A automatic sequencer (Applied Biosystems International, Perkin Elmer).

Cloning, overexpression and purification of HagA protein

Cloning. *H. paragallinarum* strain HP14 (Page serovar A) was streaked on TM/SN (Reid & Blackall, 1987) and incubated at 37 °C overnight in the presence of 5% CO_2. A lysate was prepared by harvesting one plate of HP14 into 100 µl sterile PBS and boiling this suspension for 10 min. The *hagA* gene was amplified from strain HP14 using primers HA12 and HA13 (Table 2). The ~1·1 kb PCR product was extracted using QIAQuick Gel Extraction kit (Qiagen) and cloned into pGEM-T Easy (Promega). The *hagA* gene was subcloned from the resulting plasmid into a pQE30 His-tag fusion vector (Qiagen) by digestion with *Bam*HI (from primer) and *Pst*I (from vector), generating pQE30*hagA*. This plasmid was transferred into an expression strain, *E. coli* M15(pREP4), by electroporation followed by selection on Luria–Bertani (LB) agar supplemented with 0·05% (w/v) glucose, 100 µg ampicillin ml^{-1} and 25 µg kanamycin ml^{-1}. A representative clone containing the recombinant plasmid was selected for purification of rHagA. Other recombinant DNA methods used were essentially as described by Maniatis *et al.* (1989).

Expression and purification. A 10 ml culture of *E. coli* M15(pREP4) containing pQE30*hagA* was grown at 37 °C with shaking overnight in LB broth supplemented with 0·05% (w/v) glucose, 100 µg ampicillin ml^{-1} and 25 µg kanamycin ml^{-1}. The overnight culture was subcultured into 500 ml LB broth supplemented with 0·05% (w/v) glucose, 100 µg ampicillin ml^{-1} and 25 µg kanamycin ml^{-1} and grown at 37 °C, with shaking, to an OD_{600} of 0·3–0·5. Expression of rHagA was induced at 37 °C with 0·5 mM IPTG for 4 h. Cell lysis and purification of the polyhistidine–HagA fusion were as recommended in the manufacturer's instructions (Qiagen QIA-Expressionizt). Briefly, bacteria harbouring pQE30*hagA* were collected and centrifuged at 1000 ***g*** for 10 min at 4 °C. The pellet was washed with PBS and resuspended in 50 ml denaturing lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 8·0) followed by incubation at room temperature for 1 h with agitation. The cell debris was pelleted by centrifugation at 1000 ***g*** for 10 min at 4 °C and the supernatant incubated with pre-equilibrated Ni-NTA resin (Qiagen) for 30 min at room temperature with agitation. The Ni-NTA resin was equilibrated by incubation with 15 ml denaturing lysis buffer containing 20 mM imidazole for 30 min at room temperature with agitation, then washed twice with five bed volumes of wash buffer (100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 8·0, 20 mM imidazole, 500 mM NaCl). It was resuspended in wash buffer and packed into a 10 ml column and washed with a further five bed volumes of wash buffer. The His-tagged protein was eluted in three bed volumes of elution buffer (100 mM NaH_2PO_4, 10 mM Tris, 8 M urea, pH 8·0, 250 mM imidazole) in 2 ml fractions. All eluted fractions were analysed by SDS-PAGE for presence of rHagA. The pooled fractions containing rHagA were dialysed against PBS containing 0·05% (w/v) SDS overnight at 4 °C.

Analysis of purified recombinant HagA protein. The recombinant His-tagged HagA protein was analysed by immunoblotting using mAb4D. Purified rHagA (1 µg) was run on a 12% SDS-polyacrylamide gel, along with whole cells of HP14 as a positive control. The proteins were transferred to nitrocellulose membrane (Protran, Schleicher and Schuell) using semi-dry transfer (Trans-blot semi-dry transfer cell, Bio-Rad) according to the manufacturer's instructions. mAb4D was used at a dilution of 1/50 (due to deterioration of mAb4D) and secondary antibody at 1/2000 (goat anti-mouse IgG–AP conjugate, Promega). Activity of the AP conjugate secondary antibody was detected by incubation with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Amresco) with a development time of 2 h at room temperature.

Haemagglutination assay. The assay for haemagglutination activity (HA) was performed as previously described (Blackall *et al.*, 1990). Briefly, 50 µl diluent was added to the appropriate wells of a U-bottomed microtitre plate. Purified rHagA protein (50 µl) was added to the first well of the row. Doubling dilutions of the purified protein were made across the plate followed by the addition of 50 µl 0·5% (v/v) glutaraldehyde-fixed chicken red blood cells to each well. The plate was incubated at room temperature for 30–60 min. The haemagglutination titre was read as the highest antigen dilution giving at least 50% haemagglutination or one HA unit. One HA unit is that dilution of the antigen that results in a 50% mix of 'shield' and 'button' and is read as the reciprocal of the dilution immediately preceding the first 'button'. Appropriate positive and negative controls were included in the haemagglutination assay. The positive control was a whole-cell suspension of strain 0083 (Page serovar A), prepared as described previously (Blackall *et al.*, 1990). The negative controls consisted of rHagA dialysis buffer (PBS, 0·05% SDS) and a non-related His-tagged purified protein from *H. paragallinarum*. mAb4D and high-titre hyper-immune rabbit reference serotyping antisera to strains 0083 and Modesto (serovars A

and C, respectively) were used in a HI assay as described previously (Blackall *et al.*, 1990).

RESULTS

Partial purification and N-terminal analysis of the haemagglutinin protein

In order to identify the haemagglutinin protein, *H. paragallinarum* proteins were fractionated using ammonium sulfate precipitation so that a well-separated band could be identified by SDS-PAGE and used for N-terminal sequencing. A ~39 kDa protein was present in all three ammonium sulfate precipitation fractions, although it was most highly enriched in the 0–20% fraction as shown in Fig. 1(a). The identity of this band as the *H. paragallinarum* haemagglutinin antigen was confirmed by immunoblot analysis with mAb4D as shown in Fig. 1(b). The N-terminal amino acid sequence of the putative 39 kDa haemagglutinin protein from strain 0083 was determined to be APQANTFYAGAKAG. A BLASTP (Altschul *et al.*, 1997) database search with this sequence revealed homology of this 14 amino acid N-terminal sequence to the P5 protein of *Haemophilus influenzae* (85% identity, 85% similarity). Various other members of the P5 family of outer-membrane proteins of closely related organisms, *Actinobacillus actinomycetemcomitans* Omp29 (Komatsuzawa *et al.*, 1999; GenBank accession no. BAA75215) (100% identity), *Pasteurella* (*Mannheimia*) *haemolytica* PomA (Zeng *et al.*, 1999; GenBank accession no. AAD53408) (100% identity), *Haemophilus ducreyi* OmpA2 (Klesney-Tait *et al.*, 1997; GenBank accession no. AAB49274) (85% identity, 92% similarity), belonging to the HAP group, also shared close similarity with the *H. paragallinarum* haemagglutinin N-terminal sequence shown above. Oligonucleotide primers (HA1 and HA2; see Fig. 2) designed to bind to conserved sequences in the P5 genes of members of the HAP family were used to amplify a 900 bp region from strains 0083, 0222 and Modesto. The nucleotide sequence analysis showed that the deduced N-terminal amino acid sequence of this region was identical in all three strains.

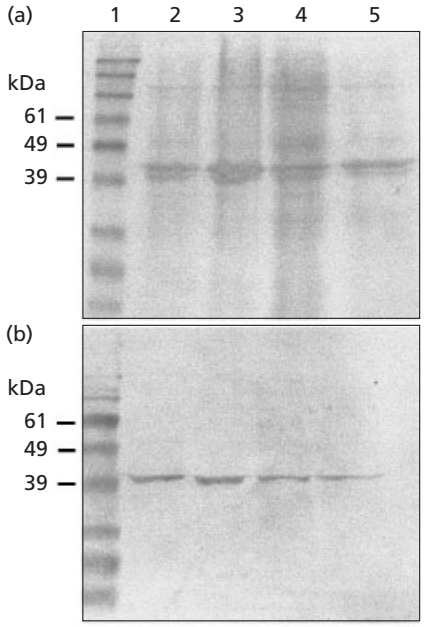

Fig. 1. Partial purification of the haemagglutinin of whole cells of *H. paragallinarum* strain 0083, using ammonium sulfate precipitation. A ~39 kDa protein was recognized by the anti-haemagglutinin monoclonal antibody (mAb4D) in all three fractions. The 0–20% fraction contained the highest concentration of the ~39 kDa protein and was subsequently used for N-terminal sequencing. (a) SDS-PAGE gel stained with Coomassie brilliant blue. (b) Immunoblot with mAb4D. Lanes: 1, molecular mass marker (Benchmark, GibcoBRL); 2, *H. paragallinarum* strain 0083 whole-cell extract; 3, 0–20% ammonium sulfate fraction; 4, 20–40% fraction; 5, 40+ % fraction.

Isolation of the full-length *hagA* gene

In order to obtain the complete sequence of the gene encoding the putative haemagglutinin of *H. paragallinarum* we used inverse PCR as shown in Fig. 2. Amplification of *Hin*dIII-digested strain Modesto DNA using primers HA5 and HA6 resulted in an amplification product of ~300 bp which contained the upstream region of the gene. Use of primers HA3 and HA7 resulted in an amplification product of ~1000 bp from *Bfa*I-digested strain Modesto DNA. This amplification product contained the downstream sequence including the stop codon. Amplification products were purified from a 1% (w/v) agarose gel and sequenced. The inverse PCR products from strain Modesto were used to produce the full-length contig containing the gene encoding the putative haemagglutinin, termed *hagA*. The full-length sequence of *hagA* in strain Modesto was 1026 bp (or 341 aa). A BLASTX (Altschul *et al.*, 1990) database search revealed that this sequence (62% identity, 73% similarity) and the *hagA* genes from other *H. paragallinarum* strains (see Fig. 3) were similar to the *H. influenzae* P5 gene. The sequence data were deposited in GenBank with the accession number AF491827 (strain Modesto). The haemagglutinin gene (*hagA*) of *H. paragallinarum* was slightly smaller than P5 of *H. influenzae* (1059 bp, GenBank accession number L20309) (Fleischmann *et al.*, 1995). The signal peptidase cleavage site in the secretory signal sequence was predicted to be between the amino acid positions 21 and 22, as shown in Fig. 4, using the CBS SignalP program, version 1.1 (Nielsen *et al.*, 1997).

Primer pairs binding outside the *hagA* ORF (HA8/HA10 or HA8/HA11; Table 1) were used to amplify the full-length gene to determine the sequences of all the serovar reference strains. The sequences were deposited in GenBank with the following accession numbers: AF491820 (0083), AF491826 (221), AF491817 (2403), AF491825 (E-3C), AF491821 (HP14), AF491819 (0222), AF491822 (2671), AF491823 (H-18), AF491824 (SA-3) and AF491818 (HP60). The start of the mature form of the HagA protein corresponds with the N-terminal

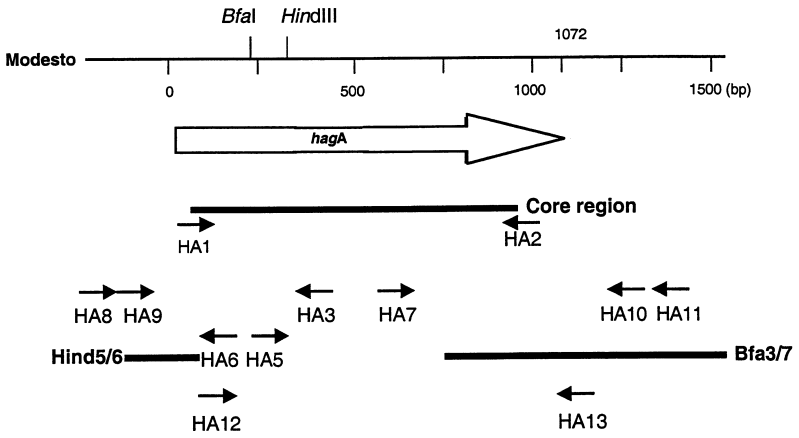

Fig. 2. Schematic representation of inverse-PCR products and primers used to identify the full-length sequence of the *H. paragallinarum hagA* gene. The arrow represents the ORF (*hagA*) and the direction of transcription. The core region is the sequence obtained by HA1/HA2 amplification. Hind5/6 is the inverse PCR product obtained from a *Hin*dIII restriction digest of strain Modesto chromosomal DNA. Primers HA5 and HA6 were used to amplify and sequence the Hind5/6 fragment. Bfa3/7 is the inverse PCR product obtained from a *Bfa*I restriction digest of strain Modesto chromosomal DNA. Primers HA3 and HA7 were used to amplify and sequence the Bfa3/7 fragment.

```
P5-Hinf      1 MKKTAIALVVAGLAAASVAQAAPQENTFYAGVKAGQGSFHDGIKQYLSST  50
HagA-Hpg     1 MKKTAIALAIAGLTAASVAQAAPQANTFYAGAKAGWASFHDGLNQFENSQ  50
                     *   *                                 *   *

P5-Hinf     51 N-YGYRRNTFTYGVFGGYQILNQDNFGLAAELGYDNFGRAKLREAGKPKA  99
HagA-Hpg    51 NAYGTLRNSVTYGVFGGYQITDN----FAVELGYDDFGRAKFRQDGETVI  96
                      *              *                  *

P5-Hinf    100 KHTNHGAHLSLKGSYEVLDGLDVYGKAGVALVRSDYKFYEEANGTRDHKK 149
HagA-Hpg    97 KHTNHGAHLSLKASYPVLEGLDVYARVGAALIRSDYKPTKRAAPNETH-- 144
                           *      *     *           *

P5-Hinf    150 GRHTARASGLFAVGAEYAV--LPELAVRLEYQWLTRVGKYRPQDNPNTAI 197
HagA-Hpg   145 -EHSLKVSPVFAGGLEYNLPSLPELALRVEYQWVNKVGRD------GSRV 187
                     *       *       *      ***  *       ** *

P5-Hinf    198 NYNPWIGSINAGISYRFGQGEAPVVAAPEMVSKTFSLNSDVTFAFGKANL 247
HagA-Hpg   188 DYTPSIGSVTAGLSYRFGQS--APVVEPKVVAKTFALNSDVTFAFGKANL 235
              *     **  *                  * *   *

P5-Hinf    248 KPQAQATLDSVYGEISQVKSAKVAVAGYTDRIGSDAFNVKLSQERADSVA 297
HagA-Hpg   236 RPEAQNVLDGIYGEIAQLKSVQVDLAGYTDRIGSEAANLKLSQRRADTVA 285
               * *       *    * *   *  *         *   *        *

P5-Hinf    298 NYFVAKGVAADAISATGYGEANPVTGATCDQVKGRKALIACLAPDRRVEI 347
HagA-Hpg   286 NYLVSKGVAQEVISSTGYGEANPVTGAKCDAVKGRKALIACLADDRRVEI 335
                   *     *

P5-Hinf    348 A       348
HagA-Hpg   336 SVKGNEN 342
```

Fig. 3. Alignment of the deduced amino acid sequence of *H. paragallinarum* strain HP14 *hagA* (GenBank accession number AF491821) with the P5 protein of *H. influenzae* (GenBank accession number L20309). The sequences were aligned using CLUSTALW (Thompson *et al.*, 1994). Bold residues indicate identical amino acids; asterisks (*) represent similar amino acids.

sequence obtained from strain 0083, and is presumably conserved across the other 10 strains sequenced, as this region of the protein is highly conserved amongst the strains analysed. A single copy of *hagA* was shown to be present in the *H. paragallinarum* strains tested by Southern blot analysis (result not shown). The degree of sequence variation between the strains sequenced is small (Fig. 4). A phylogenetic tree was constructed based on amino acid sequences (Fig. 5). Strains did not cluster according to Page serovar groups. Close examination of amino acid sequence differences at each point of variation in the *H. paragallinarum* haemagglutinin protein revealed that none of these changes correlate with recognized serovar groupings (Fig. 4), except at position 88, where an arginine (charged residue) in serovar B is replaced with a leucine, phenylalanine or serine (non-polar residues) in serovars A or C. However, the significance of this residue change is unclear due to the small number of sequences examined.

Cloning, overexpression and purification of recombinant HagA protein

In order to confirm that the *hagA* gene encoded the *H. paragallinarum* haemagglutination activity, the gene was overexpressed and the protein purified for immunoblot analysis and haemagglutination assays. The gene from strain HP14 was cloned into the expression vector pQE30 (see Methods). Following induction of a mid-exponential-phase culture of *E. coli* M15(pREP4)/pQE30*hagA* with IPTG at 37 °C for 4 h, a ~37 kDa protein was observed on a 12% SDS-PAGE gel. High-

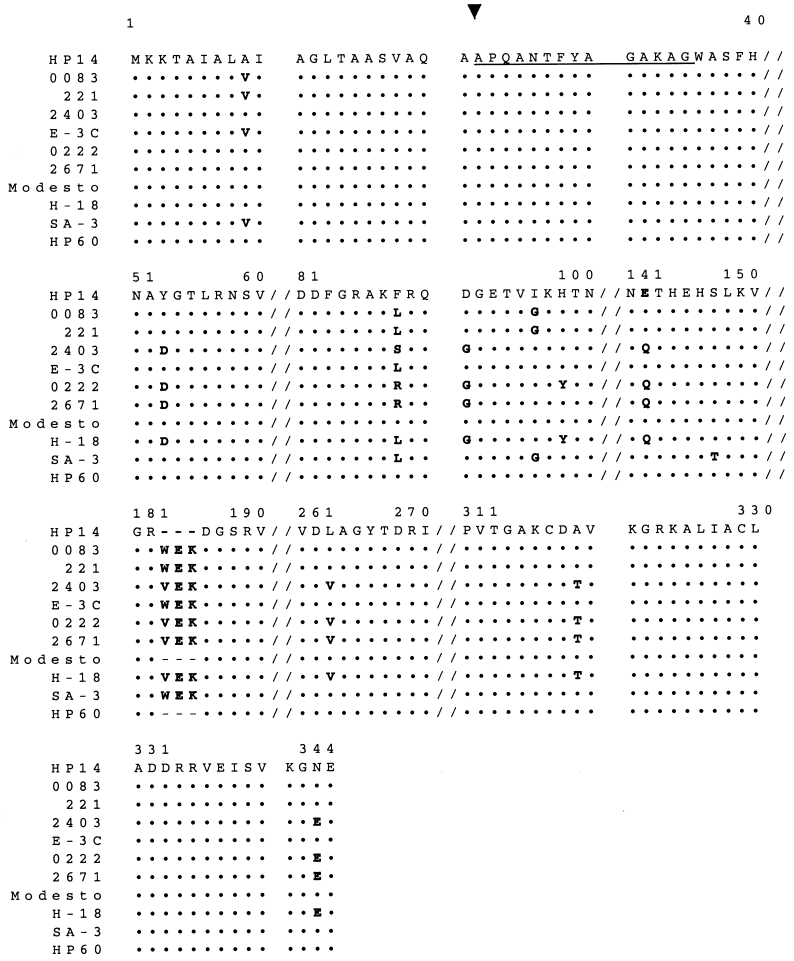

Fig. 4. Alignment of the deduced amino acid sequences of the *hagA* gene of the 11 *H. paragallinarum* strains. The amino acid sequences of *H. paragallinarum* strains 0083 (serovar A), 221 (serovar A), 2403 (serovar A), E-3C (serovar A), HP14 (serovar A), 0222 (serovar B), 2671 (serovar B), Modesto (serovar C), H-18 (serovar C), SA-3 (serovar C) and HP60 (serovar C) were aligned using the Multalin program, version 5.3.3 (Corpet, 1988). Only those regions of the sequences that contain polymorphisms are shown; regions of sequence identical to HP14 (See Fig. 3; accession number AF491821) are indicated by two slashes (//). The amino acid sequence of strain HP14 is shown above the sequence (see Fig. 3 for complete sequence). The predicted cleavage site is indicated by an arrow. The N-terminal sequence obtained from strain 0083 is underlined.

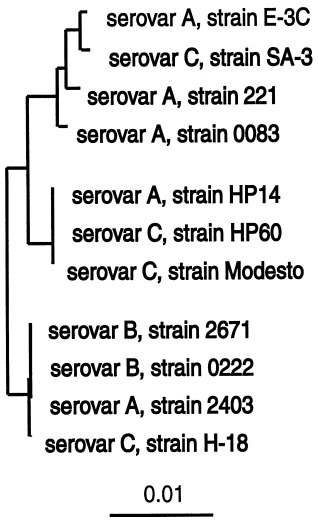

Fig. 5. Phylogenetic tree showing the relationship between the full-length *hagA* gene sequences of the 11 serotyping reference strains of *H. paragallinarum*. Strains 0083, 221, 2403, E-3C and HP14 belong to Page serovar A; strains 0222 and 2671 belong to Page serovar B, and strains Modesto, H-18, SA-3 and HP60 belong to Page serovar C. The evolutionary distance tree was constructed using the PAM–Dayhoff matrix for amino acid similarity and neighbour joining (ARB software package).

level expression of this protein was observed by comparing whole-cell lysates before and after induction. Analysis by SDS-PAGE of a culture of M15(pREP4)/pQE30*hagA* induced for 4 h and lysed with a French press revealed that the expressed His-tagged recombinant HagA (rHagA) was produced as an insoluble form, most probably in inclusion bodies (data not shown). As a result, denaturing conditions using urea-based buffers were chosen to purify the rHagA.

The rHagA protein purified from the Ni-NTA column was free of contaminant proteins as determined by SDS-PAGE analysis (Fig. 6a, lane 3). From a 500 ml culture, approximately 23 mg rHagA protein was purified at a concentration of 0·58 mg ml^{-1} as determined using a bicinchoninic acid protein estimation kit (Pierce).

Confirmation of the purified protein as a *H. paragallinarum* haemagglutinin

To confirm the identity of the protein encoded by the cloned gene with the previously identified 39 kDa haemagglutinin protein (Iritani *et al*., 1980), we examined the rHagA protein in both immunoblots and HI assays. Purified rHagA protein and *H. paragallinarum* whole-cell extracts were analysed for reactivity with the

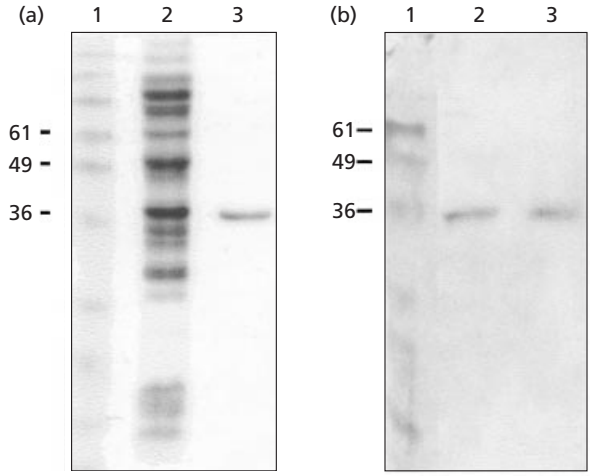

Fig. 6. Characterization of recombinant His-tagged HagA protein. (a) Coomassie-stained SDS-PAGE gel of purified protein. Lanes: 1, molecular mass ladder (Benchmark, GibcoBRL); 2, *H. paragallinarum* strain HP14 whole-cell lysate (~1 mg total protein); 3, purified recombinant His-tagged HagA protein (1 μg). (b) Western blot analysis of purified protein using the anti-haemagglutinin monoclonal antibody (mAb4D). Lanes: 1, molecular mass ladder (Benchmark, GibcoBRL); 2, *H. paragallinarum* strain HP14 whole-cell preparation (~1 mg total protein); 3, purified recombinant HagA protein (1 μg).

serovar A specific mAb4D (Takagi *et al*., 1991a) by immunoblotting. The result (Fig. 6b) revealed reactivity with both the *H. paragallinarum* whole cells and the rHagA, with a single band of the expected molecular mass.

The purified rHagA was tested for activity in a HA assay (Blackall *et al*., 1990). A titre of 2200 HA units per mg rHagA protein was obtained. The negative controls (buffer and unrelated His-tagged protein; see Methods) did not display haemagglutination activity, confirming that neither the His-tag motif nor the buffer formulation was responsible for the haemagglutination activity.

The haemagglutination activity of rHagA, in conjunction with the recognition of this protein by monoclonal antibody mAb4D in immunoblots, confirms the identity of the recombinant protein with the haemagglutinin of *H. paragallinarum* characterized by previous workers (Iritani *et al*., 1980; Takagi *et al*., 1991b). HI assays did not demonstrate inhibition of haemagglutination of rHagA protein using mAb4D or reference polyclonal serotyping antisera.

DISCUSSION

Outer-membrane proteins of pathogenic bacteria are of particular interest in terms of their potential as vaccine candidates and for their roles as virulence determinants. The haemagglutinin antigens of *H. paragallinarum* have been suggested to be potential candidates for vaccines against infectious coryza (Iritani *et al*., 1980). In addition, the *H. paragallinarum* haemagglutinin is a key component of the two serotyping schemes used for this organism, the Kume and the Page schemes, which are based on inhibition of haemagglutination activity (Kume *et al*., 1983a; Page, 1962).

In the present study, a haemagglutinin antigen of serovar A *H. paragallinarum*, strain 0083, was identified and partially purified. A single protein band of the expected molecular mass corresponding to the haemagglutinin was identified by a haemagglutination-inhibiting monoclonal antibody, mAb4D (Takagi *et al*., 1991a), in a Western blot. The band was isolated and the N-terminal sequence determined. The resultant N-terminal sequence identified this haemagglutinin of *H. paragallinarum* as a member of a family of outer-membrane proteins (the OmpA family) including *P.* (*M.*) *haemolytica* PomA and *A. actinomycetemcomitans* Omp29, as well as *H. influenzae* P5, which functions as an adhesin (Webb & Cripps, 1998). The full-length sequence of the gene encoding this haemagglutinin (*hagA*) was obtained using inverse PCR technology. The *hagA* gene was sequenced and the deduced amino acid sequence contained a sequence identical to that originally obtained by N-terminal sequencing of the partially purified single band identified by mAb4D (Fig. 1), confirming that the cloned gene encoded the same protein that was originally identified by mAb4D and N-terminal sequencing (Fig. 2).

The *hagA* gene of *H. paragallinarum* strain HP14 was cloned and overexpressed in *E. coli*. The ~37 kDa purified recombinant protein (rHagA) was recognized, as a single band, by the serovar A anti-haemagglutinin monoclonal antibody, mAb4D. The size of rHagA is consistent with the estimated molecular mass deduced by Iritani *et al*. (1980). A small size difference was observed between the strain 0083 haemagglutinin (this study; Fig. 1, lane 2 and Iritani *et al*., 1980) and the haemagglutinin protein of strain HP14 (Fig. 6a, lane 2; 6b, lane 2). This molecular mass difference is presumably due to either the amino acid sequence differences in the *hagA* genes between the two strains or perhaps post-translational modifications (see below).

The sequencing of the 11 serotyping reference strains revealed a surprisingly small degree of sequence variation amongst the strains, given that the haemagglutinin is presumed to be the major *H. paragallinarum* serotyping antigen. We had expected to find amino acid sequence variations that correlated with the serological differences, but only limited variation was observed and, apart from a single conserved residue in serovar B sequences (Arg88), none of these sequence variations correlated with the serological groupings of the strains. The serotyping reactions have two components: the haemagglutinin that aggregates the chicken red blood cells, and the antisera which are added to the reaction to inhibit the haemagglutination activity. The sequence variation within the HagA protein did not explain the phenotypic differences observed among the strains in the HI assay. Thus, it is clear that alternative explanations are required to explain the antibody binding that differentiates the serovars in a HI assay; for example (a)

post-translational modifications of the HagA protein may enable expression of particular phenotypes to allow serotypic variation amongst strains to develop and (b) another surface protein(s) may be involved in the serotypic differences observed rather than the haemagglutinin protein identified, i.e. steric hindrance of the haemagglutinin function may occur if another membrane protein interferes with the interaction between serotyping antibodies and the haemagglutinin protein. If this is the case, then the difference between the serovars may result from variations in the expression or sequence of these other proteins. Alternatively, (c) there are multiple haemagglutinins (Kume *et al*., 1983b).

The idea that a post-translational modification may occur has some support in the literature. Recent reports of post-translational modifications of prokaryotic surface proteins have opened up a new aspect of microbial pathogenesis. There are a growing number of reports suggesting not only that many bacteria glycosylate their surface proteins but also that this process can be critical to pathogenicity (Tuomanen, 1996). Examples include significant pathogens such as *Neisseria* spp. (Power *et al*., 2000; Stimson *et al*., 1995), mycobacteria (Dobos *et al*., 1996) and streptococci (Erickson & Herzberg, 1993). A post-translational modification of *Campylobacter coli* flagellin, which involves a terminal sialic acid moiety, has also been identified (Doig *et al*., 1996). Variation in the glycosylated structures between strains is implicated in the discrimination of serotype-specific epitopes of *C. coli* (Doig *et al*., 1996). The haemagglutination activity of *Myxococcus xanthus* fimbriae has also been shown to be inhibited by the addition of specific sugars, which indicates a function for glycosylation in agglutination by this organism (Dobson *et al*., 1979). Carbohydrate analysis has previously revealed the presence of sialic acid in the purified haemagglutinin from whole cells of *H. paragallinarum* (Iritani *et al*., 1980). In addition, Iritani *et al*. (1980) found that treatment of purified haemagglutinin with glycosidase inhibited haemagglutination activity. This suggests that the haemagglutinin protein may be a glycoprotein, and if the region of the protein involved in agglutination or haemagglutination inhibition carries the carbohydrate moiety, then the differences in glycosylation could give rise to serovar specificity.

Takagi *et al*. (1991b) have previously reported the cloning of a genomic fragment containing a serovar A haemagglutinin of *H. paragallinarum* strain 221 in the vector pBR322 and expression of recombinant haemagglutinin in *E. coli* strain C600. This *E. coli* strain was shown to possess haemagglutination activity and protected chickens against infectious coryza upon challenge (Takagi *et al*., 1991b). However, neither the sequence of the gene nor the resultant protein was obtained. Takagi *et al*. (1991b) also did not purify the recombinant protein, but instead used *E. coli* expressing the protein as the immunogen in vaccination trials and were able to induce HI antibody in the chickens. In the case of our clone, the rHagA was overexpressed by *E. coli* M15(pREP4) in an insoluble form, most probably as inclusion bodies (data not shown) due to the lack of the *H. paragallinarum* signal sequence, which prevents the protein from being secreted and presented on the outer membrane of *E. coli*. This is consistent with our observations that the *E. coli* M15(pREP4) cells expressing rHagA are unable to directly haemagglutinate chicken red blood cells (data not shown) unlike the clone of Takagi *et al*. (1991b).

The deduced amino acid sequence of HagA is closely related to that of the *H. influenzae* P5 protein, as shown in Fig. 3. It is believed that the P5 outer-membrane protein of non-typable *H. influenzae* (NTHi) may play a role in NTHi pathogenesis by acting as an adhesin that binds to respiratory mucin (Webb & Cripps, 1998). Due to the significant amino acid sequence similarity between the *H. paragallinarum* HagA and *H. influenzae* P5 proteins, it is possible that a similar mechanism may play a role in infection by *H. paragallinarum*. A role for *H. influenzae* P5-mediated attachment to host structures is suggested by the observation that sialic-acid-containing oligosaccharides of respiratory mucin bind P5 (Webb & Cripps, 1998). Although the host ranges are quite different for these organisms, it is possible that such highly conserved proteins may share similar functions.

In conclusion, the *hagA* gene has been cloned and overexpressed in *E. coli* and the recombinant protein has been shown to be a functional haemagglutinin. Physicochemical and immunological analyses are consistent with this gene encoding the ~39 kDa haemagglutinin previously described by other workers (Takagi *et al*., 1991b; Iritani *et al*., 1980). The full-length DNA sequence has been determined in 11 *H. paragallinarum* serotyping strains. There is no correlation between sequence variation in this gene and the serovar of the typing strain sequenced, suggesting that the immunological differences that underlie the Page and Kume serotyping schemes do not rely solely on antibodies directed at this amino acid sequence. Serovar differences may be due to other haemagglutinins in *H. paragallinarum*, three of which have been previously described in serovar A organisms (Kume *et al*., 1983b), or blocking antibodies directed at alternative proteins or post-translational modifications. Investigations into the use of the rHagA protein as a vaccine against infectious coryza are under way.

ACKNOWLEDGEMENTS

Rhonda Hobb is supported by a scholarship from the Agency for Food and Fibre Science within the Department of Primary Industries and a University of Queensland Graduate School Award Scholarship. Hsing-Ju Tseng is supported by a University of Queensland Postgraduate Research Scholarship. Tamsin Terry's position was supported by an ARC SPIRT Grant C098050013 and Bioproperties (Australia) Pty Ltd.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Altschul, S. F., Madden, T. L., Schaffer, A. J., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Blackall, P. J. & Yamamoto, R. (1990). Infectious coryza. In *A Laboratory Manual for the Isolation and Identification of Avian Pathogens*, pp. 27–31. Edited by H. G. Pearson, L. H. Arp, C. H. Domermuth & J. E. Pearson. Ames, IA: American Association of Avian Pathologists.

Blackall, P. J. & Yamamoto, R. (1997). Infectious coryza. In *Diseases of Poultry*, pp. 179–190. Edited by B. W. B. Calnek, H. J. Barnes, C. W. Beard, L. R. McDougald & Y. M. Saif. Ames, Iowa: Iowa State University Press.

Blackall, P. J., Eaves, L. E. & Aus, G. (1990). Serotyping of *Haemophilus paragallinarum* by the Page scheme: comparison of the use of agglutination and hemagglutination-inhibition tests. *Avian Dis* **34**, 643–645.

Corpet, F. (1988). Multiple sequence alignment with hierarchial clustering. *Nucleic Acids Res* **16**, 10881–10890.

Dobos, K. M., Khoo, K. H., Swiderek, K. M, Brennan, P. J. & Belisle, J. T. (1996). Definition of the full extent of glycosylation of the 45-kDa glycoprotein of *Mycobacterium tuberculosis*. *J Bacteriol* **178**, 2498–2506.

Dobson, W. J., McCurdy, H. D. & McRae, H. J. (1979). The function of fimbriae in *Myxococcus xanthus*. 2. The role of fimbriae in cell–cell interactions. *Can J Microbiol* **25**, 1359–1372.

Doig, P. K., Kinsella, N., Guerry, P. & Trust, T. J. (1996). Characterisation of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety. *Mol Microbiol* **19**, 379–387.

Erickson, P. & Herzberg, M. C. (1993). Evidence for covalent linkage of a carbohydrate polymer to a glycoprotein of *Streptococcus sanguis*. *J Biol Chem* **268**, 23780–23783.

Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995). Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Iritani, Y., Katagiri, K. & Arita, H. (1980). Purification and properties of *Haemophilus paragallinarum* haemagglutinin. *Am J Vet Res* **41**, 2114–2118.

Klesney-Tait, J., Hiltke, T. J., Maciver, I., Spinola, S. M., Radolf, J. D. & Hansen, E. J. (1997). The major outer membrane protein of *Haemophilus ducreyi* consists of two OmpA homologs. *J Bacteriol* **179**, 1764–1773.

Komatsuzawa, H., Kawai, T., Wilson, M. E., Taubman, M. A., Sugai, M. & Suginaka, H. (1999). Cloning of the gene encoding the *Actinobacillus actinomycetemcomitans* serotype b OmpA-like outer membrane protein. *Infect Immun* **67**, 942–945.

Kume, K., Sawata, A. & Nakase, Y. (1980). Relationship between protective activity and antigen structure of *Haemophilus paragallinarum* serotypes 1 and 2. *Am J Vet Res* **41**, 97–100.

Kume, K., Sawata, A., Nakai, T. & Matsumoto, M. (1983a). Serological classification of *Haemophilus paragallinarum* with a hemagglutinin system. *J Clin Microbiol* **17**, 958–964.

Kume, K., Sawata, A. & Nakai, T. (1983b). Serologic and immunologic studies on three types of haemagglutinin of *Haemophilus paragallinarum* serotype 1 organisms. *Jpn J Vet Sci* **45**, 783–792.

Kume, K., Sawata, A. & Nakai, T. (1984). Clearance of the challenge organisms from the upper respiratory tract of chickens injected with an inactivated *Haemophilus paragallinarum* vaccine. *Jpn J Vet Sci* **46**, 843–850.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Maniatis, T. F., Fritsch, E. F. & Sambrook, J. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Otsuki, K. & Iritani, Y. (1974). Preparation and immunological response to a new mixed vaccine composed of inactivated Newcastle Disease virus, inactivated infectious bronchitis virus, and inactivated *Hemophilus gallinarum*. *Avian Dis* **18**, 297–304.

Page, L. A. (1962). *Haemophilus* infections in chickens: characteristics of 12 *Haemophilus* isolates recovered from diseased chickens. *Am J Vet Res* **23**, 85–95.

Power, P. M., Roddam, L. F., Dieckelmann, M., Srikhanta, Y. N., Tan, Y.-C. & Jennings, M. P. (2000). Genetic characterisation of pilin glycosylation in *Neisseria meningititis*. *Microbiology* **146**, 967–979.

Reid, G. G. & Blackall, P. J. (1987). Comparison of adjuvants for an inactivated infectious coryza vaccine. *Avian Dis* **31**, 59–63.

Sawata, A. & Kume, K. (1983). Relationships between virulence and morphological or serological properties of variants dissociated from serotype 1 *Haemophilus paragallinarum* strains. *J Clin Microbiol* **18**, 49–55.

Stimson, E., Virji, M., Makepeace, K. & 9 other authors (1995). Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetoamido-2,4,6-trideoxyhexose. *Mol Microbiol* **17**, 1201–1214.

Takagi, M., Hirayama, N., Makie, H. & Ohta, S. (1991a). Production, characterization and protective effect of monoclonal antibodies to *Haemophilus paragallinarum* serotype A. *Vet Microbiol* **27**, 327–338.

Takagi, M., Ohmae, K., Hirayama, N. & Ohta, S. (1991b). Expression of hemagglutinin of *Haemophilus paragallinarum* serotype A in *Escherichia coli*. *J Vet Med Sci* **53**, 917–920.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tuomanen, E. I. (1996). Surprise? Bacteria glycosylate proteins too. *J Clin Invest* **98**, 2659–2660.

Webb, D. C. & Cripps, A. W. (1998). Secondary structure and molecular analysis of interstrain variability in the P5 outer-membrane protein on non-typable *Haemophilus influenzae* isolated from diverse anatomical sites. *J Med Microbiol* **47**, 1059–1067.

Yamaguchi, T., Kobayashi, M., Masaki, S. & Iritani, Y. (1993). Isolation and characterization of a *Haemophilus paragallinarum* mutant that lacks a haemagglutinating antigen. *Avian Dis* **37**, 970–976.

Zeng, H., Pandher, K. & Murphy, G. L. (1999). Molecular cloning of the *Pasteurella haemolytica pom*A gene and identification of bovine antibodies against PomA surface domains. *Infect Immun* **67**, 4968–4973.

Received 5 July 2001; revised 21 January 2002; accepted 20 February 2002.