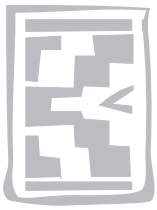


Genetic organisation of the capsule transport gene region from *Haemophilus paragallinarum*

O. DE SMIDT, J. ALBERTYN*, R.R. BRAGG and E. VAN HEERDEN

Department of Microbial, Biochemical and Food Biotechnology, University of the Free State P.O. Box 339, Bloemfontein, 9300 South Africa

ABSTRACT

DE SMIDT, O., ALBERTYN, J., BRAGG, R.R. & VAN HEERDEN, E. 2004. Genetic organisation of the capsule transport gene region from *Haemophilus paragallinarum Onderstepoort Journal of Veterinary Research*, 71:139–152

The region involved in export of the capsule polysaccharides to the cell surface of *Haemophilus paragallinarum* was cloned and the genetic organisation determined. Degenerate primers designed from sequence alignment of the capsule transport genes of *Haemophilus influenzae*, *Pasteurella multocida* and *Actinobacillus pleuropneumoniae* were used to amplify a 2.6 kb fragment containing a segment of the *H. paragallinarum* capsule transport gene locus. This fragment was used as a digoxigenin labelled probe to isolate the complete *H. paragallinarum* capsule transport gene locus from genomic DNA. The sequence of the cloned DNA was determined and analysis revealed the presence of four genes, each showing high homology with known capsule transport genes. The four genes were designated *hctA*, *B*, *C* and *D* (for *H. paragallinarum* capsule transport genes) and the predicted products of these genes likely encode an ATP-dependent export system responsible for transport of the capsule polysaccharides to the cell surface, possibly a member of a super family designated ABC (ATP-binding cassette) transporters.

Keywords: Capsular transport genes, *Haemophilus paragallinarum,* infectious coryza

INTRODUCTION

Haemophilus paragallinarum is a gram-negative, polar staining, non-motile bacterium. In 24-h cultures it appears as short rods, or coccobacilli, 1–3 µm in length and 0.4–0.8 µm in width, with a tendency for filament formation. This organism causes an acute respiratory disease of chickens known as infectious coryza (IC), a disease first recognised as a distinct entity in the late 1920s. Since the disease proved to be infectious and primarily affected the nasal passages, the name "infectious coryza" was adopted (Blackall 1989). The major economic effect of the disease is an increased culling rate in meat chickens and a reduction in egg production (10–40 %) in laying and breeding hens. The disease is limited primarily to chickens and has no public health significance (Yamamoto 1991). All the commercially available bacterins against IC consist of inactivated broth cultures of a combination of two or three different serotypes. Although vaccines against IC have been used in South Africa since 1975, it became apparent in the 1980s that the vaccines were becoming less effective in controlling the disease (Bragg, Coetzee & Verschoor 1996). This may be due to the emergence of previously unknown serovars, serogroups or changes in the population dynamics. Vaccine efficiency is therefore a problem and an alternative to available vaccines is needed.

* Author to whom correspondence is to be directed. E-mail: AlbertynJ.sci@mail.uovs.ac.za

Accepted for publication 11 November 2003—Editor

Capsules are found on the surface of a wide range of bacteria and are often important for virulence. These polysaccharide structures have been the subject of intensive investigation because of their usefulness as vaccines for prevention of bacterial infections (Lee 1987; Boulnois & Roberts 1990). Many researchers sought to understand the role of the capsule in virulence by identifying the genes involved in capsular polysaccharide export and biosynthesis. The genetic organization of the group II capsule gene loci of *Haemophilus influenzae* type b (Kroll, Zamze, Loynd & Moxon 1989; Kroll 1992), *Escherichia coli* K1 and K5 (Boulnois, Roberts, Hodge, Hardy, Jann & Timmis 1987; Jann & Jann 1990), *Pasteurella multocida* M1404 (B:2) (Boyce, Chung & Adler 2000) and *Actinobacillus pleuropneumoniae* serotype 5a (Ward & Inzana 1997) have been determined and are very similar. In each of these species, a central DNA segment necessary for capsular polysaccharide biosynthesis is flanked by DNA encoding proteins for capsule export. Substantial homology exists in the genes required for capsular polysaccharide export among these species, suggesting a common evolutionary origin (Frosch, Edwards, Bousset, Kraube & Weisgerber 1991).

Genetically defined acapsular mutants have been shown to have reduced virulence in a number of organisms (Boyce *et al*. 2000). A mutant defective in the export of the *P. multocida* capsule was constructed by allelic exchange and virulence assays showed the acapsular *P. multocida* to be 10^6 fold less virulent than their encapsulated counterparts (Boyce & Adler 2000). Similar studies have been conducted on the *bexA* gene of *H. influenzae* (Kroll, Hopkins & Moxon 1988). A frame shift mutation engineered at a restriction site within the open reading frame resulted, when introduced into the *cap* locus in the chromosome, in the expression of a mutant phenotype. The noncapsulated mutants of *A. pleuropneumoniae* reported by Inzana, Todd & Veit (1993) showed extreme stability and induced a protective immune response without any symptoms of disease. This not only proves the capsule's involvement in virulence but also offers the opportunity to investigate the possibility of producing live vaccines.

In an attempt to understand the genetic organization of the capsular genes of *H. paragallinarum* degenerate PCR primers, based on the capsule loci of *H. influenzae*, *A. pleuropneumoniae* and *P. multocida,* were used to amplify a section of the capsule transport genes of *H. paragallinarum*. This section was employed as a probe to clone the full-length transport region.

MATERIALS AND METHODS

Bacterial strains

Haemophilus paragallinarum strain 1742, obtained from the Department of Poultry Health, University of Pretoria, South Africa, was grown in TM/SN medium (1 % biosate peptone, 1 % NaCl, 0.5 % glucose, 0.1 % starch and 0.0005 % thiamine solution, oleic acid-albumin complex and chicken serum as supplements) as described by Blackall & Yamamoto (1990), in which 1.5 % agar was used to solidify the medium if required. In liquid culture the organisms were grown without aeration and on solid media in a candle jar at 37 °C. *Escherichia coli* strain Sure®2 (Stratagene) was grown with aeration in Luria-Bertani (LB) broth (Sambrook, Fritsch & Maniatis 1989) under selective pressure with 60 µg/mℓ ampicillin in liquid and solid media when required.

Preparation and analysis of genomic and plasmid DNA

Genomic DNA was prepared from 20, 5 mℓ liquid cultures of *H. paragallinarum* grown for 16 h (Towner 1991). The cells were harvested by centrifugation at 3 000 *g* for 10 min at 4 °C and the mass of the pellet was determined. The pellet was washed in TE-buffer (10 mM Tris-HCl, 1 mM EDTA) pH 8 and centrifuged again at 3 000 *g* for 5 min at 4 °C. The pellet was re-suspended in 40 mℓ/0.5 g cells buffer (50 mM Tris-HCl, pH 8, 0.7 mM sucrose) and lysozyme (20 mg/mℓ) was added before the suspension was incubated on ice for 5 min. Six hundred microlitres EDTA (0.5 M, pH 8) and 500 µℓ 10 % SDS were added for each 0.5 g cells, gently mixed and placed on ice for 5 min. After the addition of 10 mℓ/0.5 g cells digestion buffer (1 % SDS, 50 mM Tris-HCl pH 8, 0.1 M EDTA, 0.2 M NaCl, 0.5 mg/mℓ proteinase K), the suspension was incubated at 55 °C for 3–16 h with mild shaking. One time the volume of pH calibrated phenol (pH 7.8) was added to the lysate and incubated a further 3 h at 25 °C with constant inversion. Cell debris was removed by centrifugation at 4 000 *g* for 10 min and the supernatant mixed with 0.1x the volume 5 M NaCl and placed on ice for 5 min. Genomic DNA was precipitated with 10 mℓ 100 % ethanol, spooled and washed in 1 mℓ 70 % ethanol. After drying, the pellet was suspended in 500 µℓ/ 0.5 g cells TE-buffer and incubated at 50 °C for 1 h or kept at 4 °C overnight before use.

Plasmid DNA was isolated by a rapid alkaline lysis method described by Sambrook *et al.* (1989) and suspended in TE-buffer containing 10 µg/mℓ RNase. Genomic and plasmid DNA were analysed by restriction enzyme digestion. Plasmid DNA was digested with *Eco*RI or *Hin*dIII for 1 h, while genomic DNA was digested using *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I or *Xba*I for 3–16 h. All the enzymes used in these digestions were obtained from Roche Molecular Biochemicals.

PCR analysis and cloning techniques

PCR analysis was performed in a Perkin-Elmer Geneamp 2400 thermocycler. *Haemophilus paragallinarum* genomic DNA (60 ng) was used as template and PCR reactions were carried out in 50 µℓ volumes. The reaction mixtures consisted of a 10x dilution of reaction buffer (100 mM Tris-HCl, 15 mM $MgCl_2$, 500 mM KCl, pH 8.3), 2 pmol of each degenerate primer (Table 1) in different combinations, 0.2 mM dNTP mixture and 5 U of *Taq* polymerase (Roche). The reaction conditions consisted of an initial denaturation cycle of 94 °C for 5 min followed by 25 cycles of 94 °C for 30 s, 45 °C for 30 s, 72 °C for 2 min and a final elongation cycle of 72 °C for 5 min. The same reaction constituents and conditions were used for amplification of the partial *H. paragallinarum* capsule transport gene locus and for production of a DNA probe for screening.

PCR products were purified and DNA fragments were recovered from agarose gels with the GFXä-PCR DNA and gel band purification kit (Amersham Pharmacia Biotech). Purified fragments were cloned into either pGEM-T Easy or pGEM-3Z (Promega).

Escherichia coli strain Sure®2 was grown to early log phase at 18 °C in SOB-media as described by Hanahan (1983). Competent *E. coli* cells were prepared according to the method of Inoue, Nojima & Okayama (1990).

Blotting techniques

Southern hybridisation was used as a method to identify fragments in digested genomic DNA that encode the capsule transport gene locus and colony hybridisation to identify positive clones containing the recombinant plasmids.

Genomic DNA was digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I or *Xba*I and the fragments separated by agarose gel electrophoresis. The DNA was transferred to a Magnacharge nylon membrane (Micron separations, Inc.) by 1 h downward capillary transfer as described by Chomczynski (1992). DNA was linked to the membrane with the GS gene linker™ (BIO-RAD) prior to hybridisation.

Colony blotting to screen for the presence of clones containing the transport gene locus was performed on transformants grown for 16 h on LB plates containing ampicilin. Blotting proceeded as described by the DIG system users' guide for filter hybridisation (Roche Molecular Biochemicals). Colonies were lifted from the growth media and fixed on a magnacharge nylon membrane. The membrane was subjected to lysis in 10 % SDS and denaturation solution (0.5 M NaOH, 1.5 M NaCl) followed by neutralisation (1 M Tris-HCl pH 7.5, 1.5 M NaCl) and washing twice in SSC until all the cell debris was removed.

Hybridisation and colorimetric detection were performed as described in the DIG Nucleic Acid Detection Kit (Roche Molecular Biochemicals).

Probe labelling and screening methods

The 2.6 kb fragment used as a hybridisation probe was amplified from *H. paragallinarum* genomic DNA with primers HctD-1F and HctA-1R (Table 1). This fragment was prepared as the Hct-probe by random prime labelling with digoxigenin using the DIG labelling and detection kit.

TABLE 1 Degenerate and sequence specific oligonucleotides used for the amplification of the capsule transport gene locus from *H. paragallinarum*

Degenerate primers	
HctD-1F	5'- GAT AAA GAT WTW GTH TAT GTR TCR AAT GCA CC -3'
HctC-1F	5'- GCB TCY GAT ATT TAT RTT TCD SAA TCD AG -3'
HctC-1R	5'- CYA AAT AMA RTT GYT GGC GAT C -3'
HctB-1F	5'- ATG ATG TGG CGH AAT GCD TC -3'
HctB-1R	5'- AAC ATT TCY GWR CCR TGA ATC ATY GG -3'
HctA-1R	5'- ATY TTR GTT TCW CAT AGC CCG WVT -3' G
Sequence specific primers	
HctD-1R	5'- GGT GCA TTC GAC ACA TAT AC -3'
HctA-1F	5'- ATT TTA GTT TCT CAT AGT CCA ACC G -3'

The amount of labelled DNA was determined by comparison of the intensity of the spots of a serial dilution of the Hct-probe to that of a labelled control (supplied by the manufacturer).

Sequencing and analysis

Plasmid construct pHctA-D was used as a template for sequencing. Sequencing reactions were performed with the ABI Prism™ Big Dye terminator™ V3.0 cycle sequencing ready reaction kit and data collected on an ABI Prism 377 DNA sequencer (Perkin-Elmer biosystems). Data was analysed using Sequencing analysis V3.3. Sequences were reverse complemented and compared by using Sequence Navigator V 1.0.1 and assembled using Auto-assembler V1.4.0 and DNAssist V2.0.

Sequence submission

Sequence of the transport gene locus was submitted to GenBank, accession number AY116594.

RESULTS

Partial amplification of the *H. paragallinarum* capsule transport gene region

Genomic DNA was isolated from *H. paragallinarum* and used as a template for PCR amplification of the *H. paragallinarum* capsule transport genes. The capsule transport gene sequences of *H. influenzae* (*bexA-D* genes), *A. pleuropneumoniae* (*cpxA-D* genes) and *P. multocida* (*hexA-D* genes) were obtained from GenBank (accession no. X54987, U36397 & AF067175) and submitted to a multiple sequence alignment using DNAssist V 1.0.2. Six degenerate primers were designed (Table 1) from areas in these aligned gene sequences where the sequence was highly conserved.

The PCR performed with different oligonucleotide combinations (Fig. 1A), showed amplification of fragments of expected sizes in lanes 1 (~2.6 kb), 2 (~2.3 kb), 3 (~1.9 kb), 4 (~1.6 kb) and 6 (~1.1 kb). The relative position of each of these fragments in the proposed *H. paragallinarum* transport gene region is indicated in Fig. 1B. Lanes 5 and 7 showed either non-specific or no amplification. More than one band was visible in some lanes due to non-specific priming and a low annealing temperature of 45 °C. The estimated ~2.6 kb fragment amplified by the oligonucleotides HctD-1F and HctA-1R (Fig. 1A, lane 1), was cloned into pGEM-T Easy and designated pHct. The nucleotide sequence of this fragment was determined and analysis revealed considerable homology with the capsule transport genes of related organisms (*H. influenzae*, *A. pleuropneumoniae* and *P. multocida*). This high degree of homology among the four species indicated that the sequenced 2638 bp insert in pHct represented part of the capsule transport gene region of *H. paragallinarum*. By comparison with the capsular transport genes of *P. multocida*, this fragment contained homologues of *hexC* and *hexB* as well as small regions of the 3' end of *hexD* and the 5' region of *hexA.*

Construction of a mini-library to isolate the entire capsule transport gene region

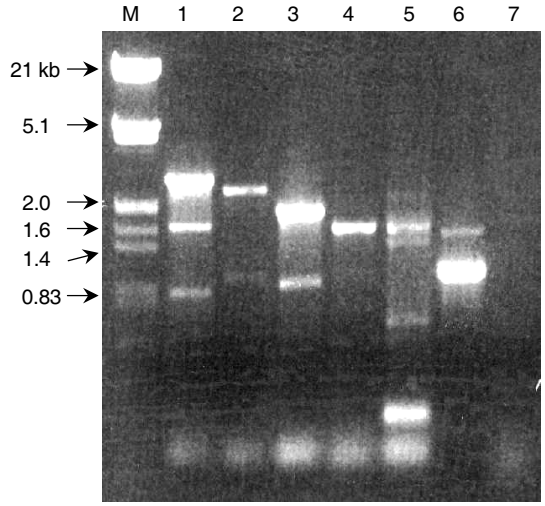
To facilitate the cloning of the full-length capsular transport region, the pHct insert was used as a probe (designated Hct) in southern blotting followed by colony hybridisation. Genomic DNA of *H. paragallinarum* was digested with five different restriction enzymes, transferred to a nylon membrane and hybridised with a digoxigenin labelled Hct-probe under stringent conditions. Southern blotting and hybridisation indicated that a *Hin*dIII fragment of ~6.15 kb (Fig. 2, lane 3) hybridised to the Hct-probe. Hybridisation products visible in lanes 1, 2, 4 and 5 at a position of ~21 kb correspond to the relative position of undigested genomic DNA or unresolved large restriction fragments when using restriction enzymes *BamH*I, *EcoR*I, *Pst*I and *Xba*I.

The *Hin*dIII fragments resolved between 6 kb and 6.5 kb were excised from the gel, purified and cloned into vector pGEM-3Z to construct a mini-library. Colony hybridisation was used as a screening method to identify positive clones containing the transport genes. A total of 93 colonies were visible within 1 day of transformation and two colonies showed hybridisation with the Hct-probe after screening under stringent conditions.

Plasmid DNA, extracted from the above-mentioned colonies and digested with *Hin*dIII, revealed the presence of a ~6.15 kb insert. To confirm that these plasmid constructs did contain the capsule transport region, the 5' and 3' terminal regions were sequenced. Sequencing confirmed that both clones were identical and also gave an indication of the orientation in which the ~6.15 kb fragment was ligated into the vector. Sequence alignment to known capsular genes, using the above-mentioned sequences, indicated that the ~6.15 kb fragment did in fact contain the relevant capsule region.

PCR reactions were performed to determine which part or parts of the Hct-probe features in the

A



B

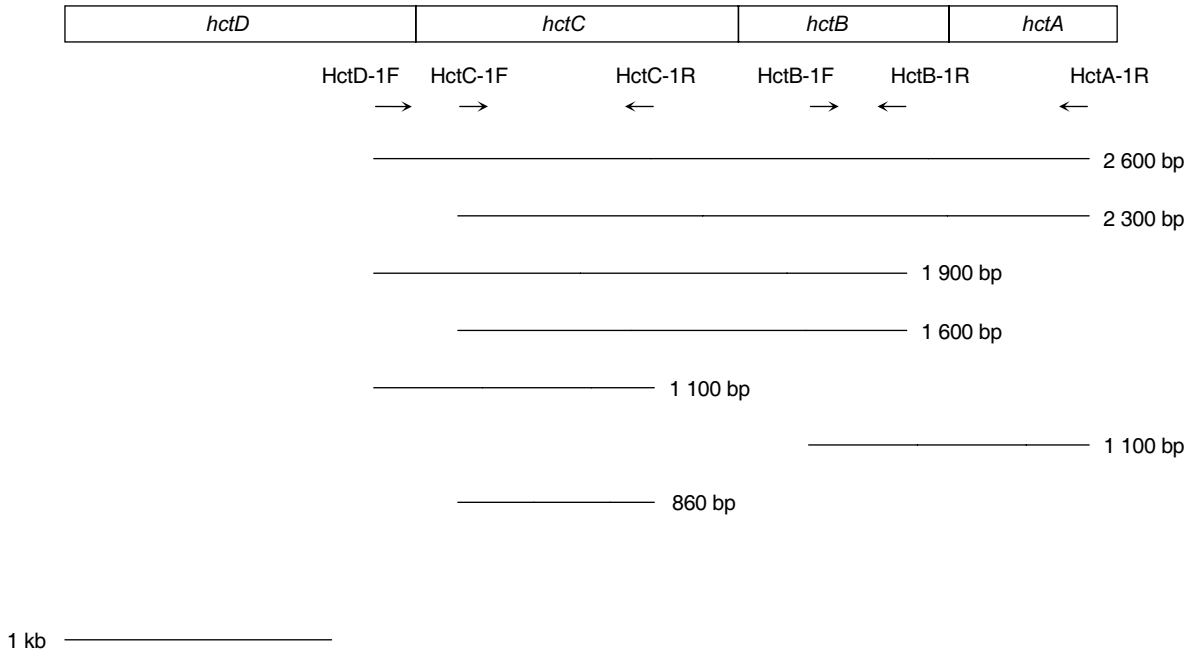


FIG. 1 A Amplification of segments of the *H. paragallinarum* capsule transport gene region

The different degenerate oligonucleotides were used in the following combinations:

HctD-1F & HctA-1R (lane 1), HctC-1F & HctA-1R (lane 2), HctD-1F & HctB-1R (lane 3), HctC-1F & HctB-1R (lane 4), HctD-1F & HctC-1R (lane 5), HctB-1F & HctA-1R (lane 6) and HctC-1F & HctC-1R (lane 7)

B Schematic representation of the proposed *H. paragallinarum* transport gene region indicating the relative positions of the degenerate oligonucleotides used in Fig. 1A

The PCR fragments expected were as follows:

2 600 bp (lane 1, Fig. 1A), 2 300 bp (lane 2, Fig. 1A), 1 900 bp (lane 3, Fig. 1A), 1 600 bp (lane 4, Fig. 1A), 1 100 bp (lane 5, Fig. 1A), 1 100 bp (lane 6, Fig. 1A), 860 bp (lane 7, Fig. 1A)

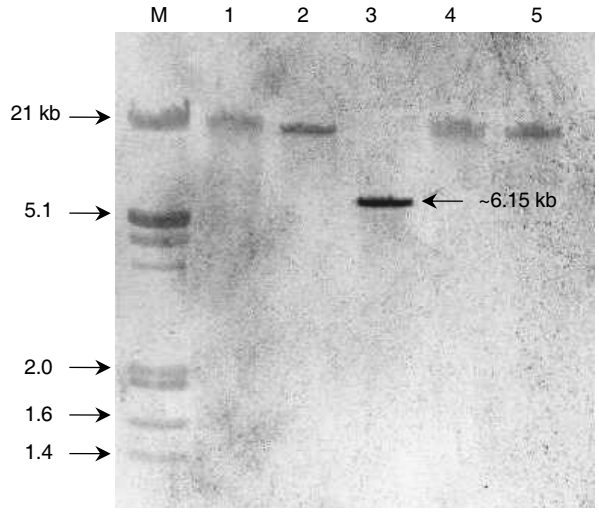


FIG. 2 Southern blot analysis of digested genomic DNA hybridised with the Hct-probe under stringent conditions. Genomic DNA was digested with *Bam*HI (lane 1), *Eco*RI (lane 2), *Hind*III (lane 3), *Pst*I (lane 4) and *Xba*I (lane 5) for 3 h. Arrow indicates positive hybridisation with a fragment ~6.15 kb in size in the *Hind*III digestion

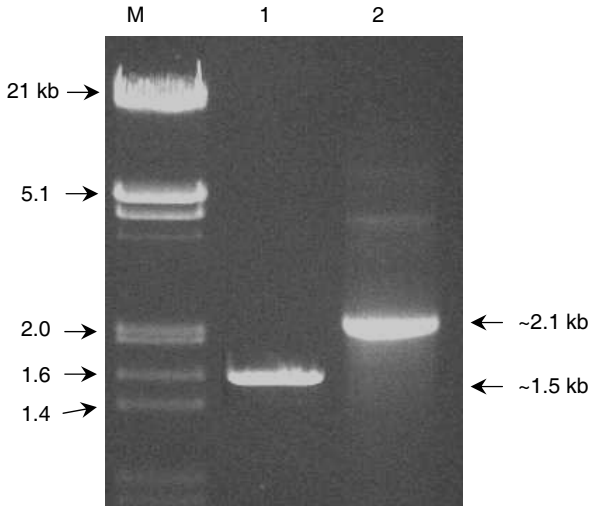


FIG. 3 Amplification of the regions up- and downstream from the Hct-probe sequence within the ~6.15 kb clone. A ~1.5 kb fragment was amplified when oligonucleotide HctD-1R was used in combination with T7 (lane 1). A ~2.1 kb fragment was amplified by oligonucleotides HctA-1F and SP6 (lane 2)

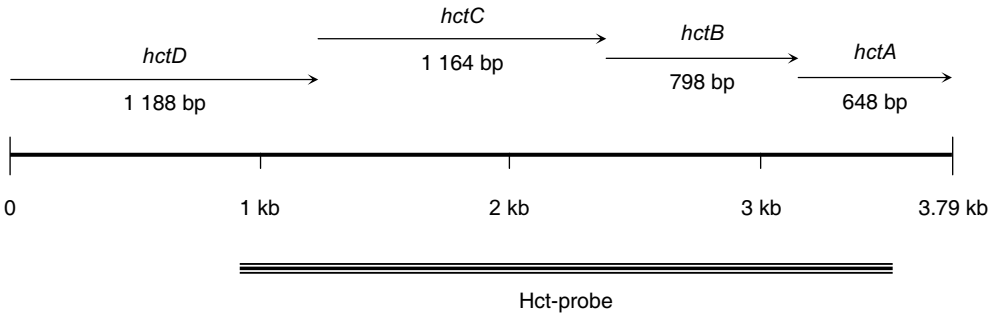


FIG. 4 Genetic map of the capsule transport gene region of *H. paragallinarum*. The locations and directions of transcription of the four open reading frames *hctDCBA* are indicated. The 2.6 kb fragment (Hct) used as the DNA probe in Fig. 2 is also indicated

~6.15 kb fragment. These PCR reactions were performed using sequence specific oligonucleotides designed according to the sequence obtained from the 2.6 kb pHct fragment. Sequence specific oligonucleotide HctD-1R was used in combination with T7 (Fig. 3, lane 1) and HctA-1F in combination with SP6 (Fig. 3, lane 2) (SP6 and T7 have binding sites on opposite sides of the multiple cloning site of pGEM-3Z). Amplification of two bands were visible, a ~1.5 kb band in lane 1 representing the segment upstream and a ~2.1 kb band in lane 2 indicating the segment downstream from the Hct-probe sequence. These results and the high degree of sequence homology with the transport genes of related organisms, verified that the ~6.15 kb fragment represents the entire *H. paragallinarum* capsule transport region and was designated pHctA-D.

The nucleotide sequence of the full-length capsular transport region was determined through primer walking using the ~6.15 kb *Hin*dIII restriction fragment of pHctA-D. Analysis of the complete sequence revealed that the *H. paragallinarum* capsule transport gene region is 3 792 bp in length with a GC content of 37 %, comprising four open reading frames representing the four capsule transport genes designated *hctDCBA* (Fig. 4 and 5). *hctD* contains 1 188 nucleotides and terminates at a TGA stop codon, encoding a putative protein of 395 amino acids. The next open reading frame, *hctC*, starts at the third base of the *hctD* stop codon and encodes a putative protein of 387 amino acids. The third base of the stop codon at the 3'- end of *hctC* is the first base of the ATG at the start of *hctB*, 798 nucleotides in length and coding for a putative protein of 265 amino

–167			…GA	TAA	GTG	TTG	ATA	TAA	ATA	AAA	TTT	CCC	GAG	TCT	TTA	–130
–129	AAA	AAT	TGG	AAT	TAT	TTT	TAT	AAA	AAA	GTT	TTC	TAC	AGG	AAA	TTG	–85
–84	AGC	AAA	AAT	TAA	TAA	TTA	TCT	ATG	ATA	ATT	ACT	CAC	TTT	TAA	TAG	–40
													hctD			
–39	AAA	AAT	CAT	GAT	CAA	AAA	CAA	AAT	AAT	TAA	GGT	AAA	ACT	ATG	CGT	6
													Met	Arg	2	
7	AAA	TCG	CTG	ATT	GCA	GTA	AGT	TAC	TGC	TTA	TTA	TTA	ATG	TCT	TGG	51
3	Lys	Ser	Leu	Ile	Ala	Val	Ser	Tyr	Cys	Leu	Leu	Leu	Met	Ser	Trp	17
52	TCT	TAT	TTG	CCA	AAT	TCA	GGA	CCG	AGC	AAA	GGC	AAT	ATT	GAG	GTA	96
18	Ser	Tyr	Leu	Pro	Asn	Ser	Gly	Pro	Ser	Lys	Gly	Asn	Ile	Glu	Val	32
97	GTC	AAT	AAA	CAG	AAA	TCC	AAT	GAG	GAT	TTG	CTT	GCA	GTA	CAG	TTG	141
33	Val	Asn	Lys	Gln	Lys	Ser	Asn	Glu	Asp	Leu	Leu	Ala	Val	Gln	Leu	47
142	ATC	GAG	GTG	AAT	AAT	AAA	GTT	GCG	GAA	AGT	ATG	TTT	AAT	CAA	CAA	186
48	Ile	Glu	Val	Asn	Asn	Lys	Val	Ala	Glu	Ser	Met	Phe	Asn	Gln	Gln	62
187	CAC	CCT	CAA	TCA	TTT	TTG	CAG	TTT	CCT	TCA	TCA	AAA	GCA	CAT	TAT	231
63	His	Pro	Gln	Ser	Phe	Leu	Gln	Phe	Pro	Ser	Ser	Lys	Ala	His	Tyr	77
232	CAT	GGG	GTA	GTT	AAA	TGC	TGG	TGT	TTA	CTT	GAT	ATT	ACT	CTC	TGG	276
78	His	Gly	Val	Val	Lys	Cys	Trp	Cys	Leu	Leu	Asp	Ile	Thr	Leu	Trp	92
277	GAA	GCA	CCC	GCC	AGC	AAC	TTT	GTT	TGG	CAG	TGT	GTT	GAA	TCA	AGC	321
93	Glu	Ala	Pro	Ala	Ser	Asn	Phe	Val	Trp	Gln	Cys	Val	Glu	Ser	Ser	107
322	CGG	TGT	GTC	GGG	CGG	ACA	AAG	CAC	TCA	CTT	ACC	GGA	ACA	GGT	GGT	366
108	Arg	Cys	Val	Gly	Arg	Thr	Lys	His	Ser	Leu	Thr	Gly	Thr	Gly	Gly	122
367	TAT	AGC	AAT	GGA	AGA	ATA	ACC	ATT	CCT	TTT	GTT	GGT	GCA	TTA	AAA	411
123	Tyr	Ser	Asn	Gly	Arg	Ile	Thr	Ile	Pro	Phe	Val	Gly	Ala	Leu	Lys	137
412	GTA	GCA	GGG	AAA	ACA	CCG	GAG	CAG	ATC	CAA	TCT	GAA	ATT	GTT	GGA	456
138	Val	Ala	Gly	Lys	Thr	Pro	Glu	Gln	Ile	Gln	Ser	Glu	Ile	Val	Gly	152
457	CGT	TTA	CAA	GCA	ATT	GCC	AAT	CAA	CCA	CAA	GCA	GTG	GTG	CGA	ATT	501
153	Arg	Leu	Gln	Ala	Ile	Ala	Asn	Gln	Pro	Gln	Ala	Val	Val	Arg	Ile	167
502	GTG	AAG	AAT	AAT	TCT	GCT	AAT	GTG	ACG	GTT	TTA	ACT	AAA	TCG	ACT	546
168	Val	Lys	Asn	Asn	Ser	Ala	Asn	Val	Thr	Val	Leu	Thr	Lys	Ser	Thr	182
547	ACT	ATT	CGA	ATG	GCT	TTA	ACT	GCT	TAC	GGT	GAA	CGA	AGT	GTT	AGA	591
183	Thr	Ile	Arg	Met	Ala	Leu	Thr	Ala	Tyr	Gly	Glu	Arg	Ser	Val	Arg	197
592	TGC	TAT	TGC	GGC	AGC	AGG	TGG	AGC	CGG	TGG	TAT	GTG	CAA	AGA	TGT	636
198	Cys	Tyr	Cys	Gly	Ser	Arg	Trp	Ser	Arg	Trp	Tyr	Val	Gln	Tyr	Cys	212
637	TTC	AGT	GCG	ACT	GAC	TCG	TGG	GAA	ATC	AGG	GTG	CAA	ACG	ATT	TCT	681
213	Phe	Ser	Ala	Thr	Asp	Ser	Trp	Glu	Ile	Arg	Val	Gln	Thr	Ile	Ser	227
682	TTA	GCC	AGG	ATT	AAC	GGA	GGG	AGC	CAC	AGG	CAA	AAT	ATC	CTA	TTA	726
228	Leu	Ala	Arg	Ile	Asn	Gly	Gly	Ser	His	Arg	Gln	Asn	Ile	Leu	Leu	242
727	CGT	TCC	GGC	GAT	GTA	GTA	ACG	TTA	TTA	AAT	AAT	CCA	CTT	TCT	TTC	771
243	Arg	Ser	Gly	Asp	Val	Val	Thr	Leu	Leu	Asn	Asn	Pro	Leu	Ser	Phe	257
772	ACT	GCA	ATG	GGT	GCG	GTA	GGA	AAT	AGT	AAA	GAA	ATT	CGT	TTT	TCG	816
258	Thr	Ala	Met	Gly	Ala	Val	Gly	Asn	Ser	Lys	Glu	Ile	Arg	Phe	Ser	272
817	GCA	GAA	GGT	TTA	ACT	TTA	GCA	GAA	GCA	ATC	GGT	CGT	TTA	GGT	GGA	861
273	Ala	Glu	Gly	Leu	Thr	Leu	Ala	Glu	Ala	Ile	Gly	Arg	Leu	Gly	Gly	287

FIG. 5 Nucleotide sequence of the capsule transport region. Three thousand nine hundred and twenty three nucleotides of the sequence are shown, from arbitrary points 167 bp upstream of *hctD* to 131 bp downstream of *hctA*. The four open reading frames are indicated as *hctD*, *hctC*, *hctB* and *hctA*, in each case from the first ATG, with the translated peptide sequence beneath. The underlined regions are referred to in the text

862	TTG	AAT	GAT	GAT	CGT	GCA	GAT	CCA	AGA	GGA	GTA	TTT	ATC	TTT	CGT	906	
288	Leu	Asn	Asp	Asp	Arg	Ala	Asp	Pro	Arg	Gly	Val	Phe	Ile	Phe	Arg	302	
907	TAT	GTT	CCA	TTT	GAA	GAA	ATG	CCC	TTA	AGT	AAA	CAA	AAT	GAA	TGG	951	
303	Tyr	Val	Pro	Phe	Glu	Glu	Met	Pro	Leu	Ser	Lys	Gln	Asn	Glu	Trp	317	
952	CAA	GCC	AAG	GGG	TAT	CAC	AAC	GGA	ATG	AAA	ATT	CCA	ACA	GTA	TAT	996	
318	Gln	Ala	Lys	Gly	Tyr	His	Asn	Gly	Met	Lys	Ile	Pro	Thr	Val	Tyr	332	
997	CAA	GCG	AAT	TTA	CTT	GAA	CCT	CAA	TCA	ATG	TTT	TGG	ATT	CAA	CAA	1041	
333	Gln	Ala	Asn	Leu	Leu	Glu	Pro	Gln	Ser	Met	Phe	Trp	Ile	Gln	Gln	347	
1042	TTT	CCA	ATT	AAA	GAT	AAA	GAT	ATT	GTT	TAT	GTA	TCT	AAT	GCA	CCA	1086	
348	Phe	Pro	Ile	Lys	Asp	Lys	Asp	Ile	Val	Tyr	Val	Ser	Asn	Ala	Pro	362	
1087	TTG	GCT	GAA	TAC	CAA	ATT	TAT	TCG	TAT	GAT	TTA	CGC	CAC	CGT	TGC	1131	
363	Leu	Ala	Glu	Tyr	Gln	Ile	Tyr	Ser	Tyr	Asp	Leu	Arg	His	Arg	Cys	377	
1132	AAC	TAC	ACC	GCC	GGT	TTC	AAC	TGT	AAA	CAA	GTG	TTA	ATA	ATC	TGT	1176	
378	Asn	Tyr	Thr	Ala	Gly	Phe	Asn	Cys	Lys	Gln	Val	Leu	Ile	Ile	Cys	392	
				hctC													
1177	AGG	GGG	AGA	TGA	TG	GAA	CAA	AAT	GTA	GTA	GTT	CAA	TCG	AAA	GAA	1220	
393	Arg	Gly	Arg	***													
					Met	Glu	Gln	Asn	Val	Val	Val	Gln	Ser	Lys	Glu	406	
1221	CAA	CTG	AGA	AAG	TTA	AAA	CAG	TGG	TTG	CGA	AAA	ATT	AAT	CTG	TTA	1265	
407	Gln	Leu	Arg	Lys	Leu	Lys	Gln	Trp	Leu	Arg	Lys	Ile	Asn	Leu	Leu	421	
1266	TTT	TTA	CTG	ACG	GTG	ATT	ATT	CCG	ACT	TTT	TGT	TCG	TTA	TTT	TAT	1310	
422	Phe	Leu	Leu	Thr	Val	Ile	Ile	Pro	Thr	Phe	Cys	Ser	Leu	Phe	Tyr	436	**Region A1 (421–442)**
1311	TTT	TCT	ATT	TGG	GCT	TCC	GAT	GTT	TAT	ATT	TCG	GAG	TCC	AGT	TTT	1355	
437	Phe	Ser	Ile	Trp	Ala	Ser	Asp	Val	Tyr	Ile	Ser	Glu	Ser	Ser	Phe	451	
1356	ATT	GTG	CGT	TCT	TCT	CGT	GCT	CAG	GCA	TCG	CTC	GGA	GGT	ATG	GGG	1400	
452	Ile	Val	Arg	Ser	Ser	Arg	Ala	Gln	Ala	Ser	Leu	Gly	Gly	Met	Gly	466	**Region A2 (458–475)**
1401	GCT	TTA	TTG	CAG	AGT	ATC	GGT	TTT	GCT	CGT	TCG	CAA	GAT	GAT	ACT	1445	
467	Ala	Leu	Leu	Gln	Ser	Ile	Gly	Phe	Ala	Arg	Ser	Gln	Asp	Asp	Thr	481	
1446	TTT	ACG	GTG	CAA	GAA	TTT	ATG	CGT	TCG	CGT	AAT	GCG	TTG	ACA	ACA	1490	
482	Phe	Thr	Val	Gln	Glu	Phe	Met	Arg	Ser	Arg	Asn	Ala	Leu	Thr	Thr	496	
1491	TTG	GAA	AGT	GAG	TTA	CCG	GTG	AGA	AAA	TTT	TAT	GAA	GAT	GAA	GGG	1535	
497	Leu	Glu	Ser	Glu	Leu	Pro	Val	Arg	Lys	Phe	Tyr	Glu	Asp	Glu	Gly	511	
1536	GAT	TTT	TTC	AGC	CCG	TTT	AAT	CCG	TTA	GGT	TTT	TTT	AAT	GAA	CAG	1580	
512	Asp	Phe	Phe	Ser	Pro	Phe	Asn	Pro	Leu	Gly	Phe	Phe	Asn	Glu	Gln	526	
1581	GAA	TTG	TTT	TAT	CAA	TAT	TTT	CGT	AAA	CAT	TTG	ATG	ATT	AAT	ATC	1625	
527	Glu	Leu	Phe	Tyr	Gln	Tyr	Phe	Arg	Lys	His	Leu	Met	Ile	Asn	Ile	541	
1626	GAT	TCT	TTA	TCT	GGG	TAT	TGC	TAC	TTT	ACA	GGT	TCC	GTG	GGT	TTA	1670	
542	Asp	Ser	Leu	Ser	Gly	Tyr	Cys	Tyr	Phe	Thr	Gly	Ser	Val	Gly	Leu	556	
1671	ATG	GCT	GAC	CTC	CGG	CAC	CAA	CAA	GAA	TTA	AAT	GGA	AGC	CAT	TAT	1715	
557	Met	Ala	Asp	Leu	Arg	His	Gln	Gln	Glu	Leu	Asn	Gly	Ser	His	Tyr	571	
1716	TGC	CAT	TTT	GGC	GGG	AAA	CCA	TTT	AGT	GGA	ATA	AAC	TCA	ATG	ATC	1760	
572	Cys	His	Phe	Gly	Gly	Lys	Pro	Phe	Ser	Gly	Ile	Asn	Ser	Met	Ile	586	
1761	GTG	CAC	GTA	AAG	ATA	CAA	TTA	CTT	TGC	GGA	ACA	ATC	GGT	AAT	GAA	1805	
587	Val	His	Val	Lys	Ile	Gln	Leu	Leu	Cys	Gly	Thr	Ile	Gly	Asn	Glu	601	
1806	GCA	GAA	AAA	TAT	TTG	TCT	GAA	ACC	TCG	ACA	GCC	TTA	AGC	CAA	TAT	1850	
602	Ala	Glu	Lys	Tyr	Leu	Ser	Glu	Thr	Ser	Thr	Ala	Leu	Ser	Gln	Tyr	616	

FIG. 5 (*Continued*)

1851	CGT	GTA	AAA	AAT	GGG	ATA	TTT	GAT	ATT	GGG	GCA	CAA	TCT	GAA	TCG	1895
617	Arg	Val	Lys	Asn	Gly	Ile	Phe	Asp	Ile	Gly	Ala	Gln	Ser	Glu	Ser	631
1896	ATT	TTA	ACT	TTA	GTG	CAG	AAG	TTG	CAG	GAT	GAA	CTG	ATT	GCC	ATT	1940
632	Ile	Leu	Thr	Leu	Val	Gln	Lys	Leu	Gln	Asp	Glu	Leu	Ile	Ala	Ile	646
1941	CAG	ACG	CAA	CTT	GAT	CAG	GTG	AGG	GGC	GTT	ATC	TCC	GGA	TAC	CCT	1985
647	Gln	Thr	Gln	Leu	Asp	Gln	Val	Arg	Gly	Val	Ile	Ser	Gly	Tyr	Pro	661
1986	CAG	GTT	AAA	GTG	TTA	AAG	GCA	AGG	CAA	TTT	GAA	AGT	ATT	CGT	GAA	2030
662	Gln	Val	Lys	Val	Leu	Lys	Ala	Arg	Gln	Phe	Glu	Ser	Ile	Arg	Glu	676
2031	AGA	AGT	GGC	ACA	ACA	ATT	GAA	TCC	GGG	GTT	TTT	GAG	GGG	AAA	CCA	2075
677	Arg	Ser	Gly	Thr	Thr	Ile	Glu	Ser	Gly	Val	Phe	Glu	Gly	Lys	Pro	691
2076	TTC	TTT	AAC	AAC	ACA	ATC	AGC	AGA	GTA	CCA	GCC	GTT	AAT	TTA	GAT	2120
692	Phe	Phe	Asn	Asn	Thr	Ile	Ser	Arg	Val	Pro	Ala	Val	Asn	Leu	Asp	706
2121	GAA	ACC	TTG	GCA	AAA	CAG	CAA	TTA	ACA	GCT	GCA	ATG	TCT	TGC	GTT	2165
707	Glu	Thr	Leu	Ala	Lys	Gln	Gln	Leu	Thr	Ala	Ala	Met	Ser	Cys	Val	721
2166	ACA	AGT	GGC	AAA	GAA	GAA	GCT	GGA	AGA	CAA	CAG	CTT	TAT	CTG	GAA	2210
722	Thr	Ser	Gly	Lys	Glu	Glu	Ala	Gly	Arg	Gln	Gln	Leu	Tyr	Leu	Glu	731
2211	ATT	ATT	GCT	AAA	CCT	AGC	CAT	CCA	GAT	TTA	GCA	TTG	GAA	CCG	CAC	2255
737	Ile	Ile	Ala	Lys	Pro	Ser	His	Pro	Asp	Leu	Ala	Leu	Glu	Pro	His	751
2256	CGT	TTG	TAC	AAT	ATT	TTG	GCA	ACT	TTG	ATT	CTT	GGA	TTA	GTT	ATT	2300
752	Arg	Leu	Tyr	Asn	Ile	Leu	Ala	Thr	Leu	Ile	Leu	Gly	Leu	Val	Ile	766
2301	TAT	GGC	GTT	TCA	ACT	TTA	TTA	TTA	GCC	GGT	GTG	AGA	GAG	CAT	AAG	2345
767	Tyr	Gly	Val	Ser	Thr	Leu	Leu	Leu	Ala	Gly	Val	Arg	Glu	His	Lys	781
		hctB														
2346	AAC	TGA	TG	CAG	TAT	GGT	GAA	CAA	ACT	TCG	TTA	AAA	GAT	TCA	TTT	2389
782	Asn	***														
			Met	Gln	Tyr	Gly	Glu	Gln	Thr	Ser	Leu	Lys	Asp	Ser	Phe	795
2390	ACT	ATC	CAA	GGA	CGG	GTG	TTG	AAA	GCG	TTG	TTG	TTG	CGT	GAA	ATT	2434
796	Thr	Ile	Gln	Gly	Arg	Val	Leu	Lys	Ala	Leu	Leu	Leu	Arg	Glu	Ile	810
2435	ATC	ACT	CGT	TAT	GGT	CGT	AAA	AAT	TTA	GGC	TTT	TTG	TGG	GTT	GTT	2479
811	Ile	Thr	Arg	Tyr	Gly	Arg	Lys	Asn	Leu	Gly	Phe	Leu	Trp	Val	Val	825
2480	CGT	GAG	CCA	TTT	TTG	ATG	AGC	CTA	GTT	ATT	GTG	GTA	ATG	TGG	CAT	2524
826	Arg	Glu	Pro	Phe	Leu	Met	Ser	Leu	Val	Ile	Val	Val	Met	Trp	His	840
2525	TTT	TTT	CGT	GCT	GAT	CGC	TTT	TCA	ACA	TTA	AAC	ATT	GTT	GCT	TTT	2569
841	Phe	Phe	Arg	Ala	Asp	Arg	Phe	Ser	Thr	Leu	Asn	Ile	Val	Ala	Phe	855
2570	GCA	ATG	ACG	GTT	ATC	CAT	TAT	TAT	GGA	TGT	GGC	GTA	ATG	CTT	CTA	2614
856	Ala	Met	Thr	Val	Ile	His	Tyr	Tyr	Gly	Cys	Gly	Val	Met	Leu	Leu	870
2615	ACC	GTG	CAA	TTA	GCG	GGA	ATG	GAT	TCC	AAT	ATC	CCA	TTA	CTT	TTA	2659
871	Thr	Val	Gln	Leu	Ala	Gly	Met	Asp	Ser	Asn	Ile	Pro	Leu	Leu	Leu	885
2660	TCA	CGT	AAT	GTA	CGT	CCT	CTT	GAT	ACG	CTT	TTT	TCT	CGT	ATG	ATT	2704
886	Ser	Arg	Asn	Val	Arg	Pro	Leu	Asp	Thr	Leu	Phe	Ser	Arg	Met	Ile	900
2705	TTG	GAG	ATT	GCT	GGT	GCG	ACT	GTA	GCA	CAA	ATT	GTG	ATG	TTA	GTG	2749
901	Leu	Glu	Ile	Ala	Gly	Ala	Thr	Val	Ala	Gln	Ile	Val	Met	Leu	Val	915
2750	ATT	TTA	ATT	GCT	ATT	GAT	TGG	ATC	GGC	TTG	CCA	AAT	GAT	GTG	TTG	2794
916	Ile	Leu	Ile	Ala	Ile	Asp	Trp	Ile	Gly	Leu	Pro	Asn	Asp	Val	Leu	930
2795	TAT	ATG	CTT	TTT	GCT	TGG	TTC	TTA	ATG	GCA	CTG	TTT	GCC	ATT	GGT	2839
931	Tyr	Met	Leu	Phe	Ala	Trp	Phe	Leu	Met	Ala	Leu	Phe	Ala	Ile	Gly	945
2840	TTA	GGT	TTA	ATT	ATT	TGT	GCT	ATT	TCT	TAT	TAT	TTA	GAG	TTT	TTC	2884
946	Leu	Gly	Leu	Ile	Ile	Cys	Ala	Ile	Ser	Tyr	Tyr	Leu	Glu	Phe	Phe	960

Region A3 (753–777)

FIG. 5 (*Continued*)

2885	GGT	AAA	ATT	TGG	GGA	ACA	TTA	TCT	TTT	GTG	ATG	TTT	CCT	ATT	TCC	2929	
961	Gly	Lys	Ile	Trp	Gly	Thr	Leu	Ser	Phe	Val	Met	Phe	Pro	Ile	Ser	975	
2930	GGT	GCA	TTC	TTT	TTA	GTG	AAT	AGT	TTG	CCA	AAC	AAT	CTG	CAA	TCT	2974	
976	Gly	Ala	Phe	Phe	Leu	Val	Asn	Ser	Leu	Pro	Asn	Asn	Leu	Gln	Ser	990	
2975	ATT	TTG	CTT	TGG	TTT	CCA	ATG	GTT	CAC	GGT	ACG	GAA	ATG	TTT	CGT	3019	
991	Ile	Leu	Leu	Trp	Phe	Pro	Met	Val	His	Gly	Thr	Glu	Met	Phe	Arg	1005	**Region B (990–1009)**
3020	CAC	GGT	TAT	TTT	GGT	TCT	TCA	GTT	ATT	ACA	ATG	GAA	TCA	CCG	AGT	3064	
1006	His	Gly	Tyr	Phe	Gly	Ser	Ser	Val	Ile	Thr	Met	Glu	Ser	Pro	Ser	1020	
3065	TAT	TTA	TTT	ATT	TGT	GAT	TTG	GTG	ATG	TTA	TTA	ATC	GGT	CTA	CTG	3109	
1021	Tyr	Leu	Phe	Ile	Cys	Asp	Leu	Val	Met	Leu	Leu	Ile	Gly	Leu	Leu	1035	
											hctA					
3110	ATG	GTG	GGT	AGT	TTT	AGT	AAT	AGG	ATT	AAT	GCA	AGA	TG	ATT	AGT	3153	
1036	Met	Val	Gly	Ser	Phe	Ser	Asn	Arg	Ile	Asn	Ala	Arg	***				
													Met	Ile	Ser	1050	
3154	GTA	GAC	CAC	GTT	TAT	AAA	AAA	TAT	CAA	ACA	CGG	ACA	GGT	TCG	GTA	3198	
1051	Val	Asp	His	Val	Tyr	Lys	Lys	Tyr	Gln	Thr	Arg	Thr	Gly	Ser	Val	1065	
3299	CCC	GTA	TTA	AAT	GAT	ATT	AAT	TTT	AGC	CTT	ACC	AAA	GAA	GAA	AAA	3243	
1066	Pro	Val	Leu	Asn	Asp	Ile	Asn	Phe	Ser	Leu	Thr	Lys	Glu	Glu	Lys	1080	
3244	ATT	GGT	ATT	TTA	GGT	CGC	AAC	GGA	GCA	GGA	AAA	TCA	CCA	TTA	ATT	3288	
1081	Ile	Gly	Ile	Leu	Gly	Arg	Asn	Gly	Ala	Gly	Lys	Ser	Pro	Leu	Ile	1095	**Region C1 (1080–1092)**
3289	CGT	TTA	ATG	AGT	GGT	GTT	GAA	GCT	CCA	ACT	TCA	GGA	ATA	ATT	CGA	3333	
1096	Arg	Leu	Met	Ser	Gly	Val	Glu	Ala	Pro	Thr	Ser	Gly	Ile	Ile	Arg	1110	
3334	CGA	GAA	ATG	AGC	ATT	TCT	TGG	CCA	TTA	GCC	TTT	AGC	GGT	GCA	TTC	3378	
1111	Arg	Glu	Met	Ser	Ile	Ser	Trp	Pro	Leu	Ala	Phe	Ser	Gly	Ala	Phe	1125	
3379	CAA	GGT	AGC	TTA	ACG	GGA	ATG	GAT	AAT	TTA	CGC	TTC	ATT	TGT	CGT	3423	
1126	Gln	Gly	Ser	Leu	Thr	Gly	Met	Asp	Asn	Leu	Arg	Phe	Ile	Cys	Arg	1140	**Region C2 (1131–1138)**
3424	ATT	TAT	AAT	GCT	GAT	ATT	AAT	TAT	GTT	ACT	GAA	TTT	ACG	GAA	TCC	3468	
1141	Ile	Tyr	Asn	Ala	Asp	Ile	Asn	Tyr	Val	Thr	Glu	Phe	Thr	Glu	Ser	1155	
3469	TTT	TCC	GAA	TTG	GGC	AAT	TAT	TTA	TAT	GAG	CCT	GTA	AAA	AAT	TAT	3513	
1156	Phe	Ser	Glu	Leu	Gly	Asn	Tyr	Leu	Tyr	Glu	Pro	Val	Lys	Asn	Tyr	1170	
3514	TCT	TCA	GGA	ATG	AAA	GCA	CGC	TTA	GCT	TTT	GCA	TTG	TCG	TTA	TCC	3558	
1171	Ser	Ser	Gly	Met	Lys	Ala	Arg	Leu	Ala	Phe	Ala	Leu	Ser	Leu	Ser	1185	
3559	GTT	GAG	TTT	GAT	TGC	TAT	CTC	ATT	GAT	GAA	GTG	ATT	GCC	GTT	GGA	3603	
1186	Val	Glu	Phe	Asp	Cys	Tyr	Leu	Ile	Asp	Glu	Val	Ile	Ala	Val	Gly	1200	
3604	GAT	TCT	CGT	TTT	AGT	GAT	AAA	TGT	CGC	TAT	GAA	CTT	TTT	GAA	AAA	3648	
1201	Asp	Ser	Arg	Phe	Ser	Asp	Lys	Cys	Arg	Tyr	Glu	Leu	Phe	Glu	Lys	1215	
3649	CGC	AAA	GAT	CGT	TCC	ATT	ATT	TTA	GTT	TCT	CAT	AGT	CCA	ACC	GCT	3693	
1216	Arg	Lys	Asp	Arg	Ser	Ile	Ile	Leu	Val	Ser	His	Ser	Pro	Thr	Ala	1230	
3694	ATT	AGA	CAA	TAT	TGT	GAT	AAT	GCA	AAA	GTA	TTA	GAT	AAA	GGA	AAA	3738	
1229	Ile	Arg	Gln	Tyr	Cys	Asp	Asn	Ala	Lys	Val	Leu	Asp	Lys	Gly	Lys	1245	
3739	TTG	TTA	GAT	TTC	TCT	TCT	ATT	GAT	GAG	GCT	TAT	CAA	TAT	TAT	AAT	3783	
1246	Leu	Leu	Asp	Phe	Ser	Ser	Ile	Asp	Glu	Ala	Tyr	Gln	Tyr	Tyr	Asn	1260	
3784	CAG	ACA	TAG	AGG	TTA	GAT	TTT	AAA	ATA	AAA	TAA	CGT	TAC	TTT	CTT	3883	
1261	Gln	Thr	***													1260	
3829	GCT	TTA	TCA	TAA	ATT	TCA	ATG	GCT	ATA	GTT	AAG	TTC	GAA	ATA	AAT	3873	
3874	CAA	GGT	AAC	AAG	CTG	AAT	ACA	GTG	AAA	AAT	AGC	ACT	TTT	TAT	GCC	3918	
3919	AAG	GT…															

FIG. 5 (*Continued*)

HctB	(207)	S	I	L	L	W	F	P	M	V	H	G	T	E	M	F	R	H	G	Y	F
BexB	(208)	S	I	A	L	W	F	P	M	I	H	G	T	E	M	F	R	H	G	Y	F
OppB	(209)	R	T	A	R	A	K	G	L	P	M	R	R	I	I	F	R	H	A	L	K

FIG. 6 Alignment of the relatively hydrophilic portions of HctB, BexB and OppB. The number in brackets is the position of the first amino acid in each sequence. Identical amino acids in all three genes are boxed, and the matches of the OppB sequence to the Dassa/Hofnung consensus are underlined

TABLE 2 Comparison of capsular transport gene and protein sizes and % identity and similarity between proteins from *H. paragallinarum* 1742 with those of related bacterial species

H. paragallinarum		Related bacterial species	Comparison on protein level		
ORF[a]	Protein[b]	ORF[a]	Accession no.[c]	% Identity	% Similarity
hctA (648)	HctA (215)	*bexA* (654) (*H. influenzae*)	P10640	77.2	85.1
		*cpx*A (615) (*A. pleuropneumoniae*)	U36397	78.6	85.6
		ctrD (651) (*N. meningitidis*)	M57677	78.6	87.4
		*hex*A (660) (*P. multocida*)	AF067175	75.3	85.6
hctB (798)	HctB (265)	*bex*B (798) (*H. influenzae*)	P19391	57.7	86.0
		*cpx*B (798) (*A. pleuropneumoniae*	U36397	57.7	84.9
		ctrC (798) (*N. meningitidis*)	M57677	58.1	86.7
		*hex*B (798) (*P. multocida*)	AF067175	58.5	82.3
hctC (1164)	HctC (387)	*bex*C (1134) (*H. influenzae*)	P22930	43.7	67.2
		*cpx*C (1167) (*A. pleuropneumoniae*)	U36397	43.9	65.6
		ctrB (1164) (*N. meningitidis*)	M57677	38.7	58.1
		*hex*C (1137) (*P. multocida*)	AF067175	49.9	71.8
hctD (1188)	HctD (395)	*bex*D (1182) (*H. influenzae*)	P22236	42.5	63.0
		*cpx*D (1212) (*A. pleuropneumoniae*)	U36397	42.5	65.0
		ctrA (1176) (*N. meningitidis*)	M57677	41.5	62.5
		*hex*D (1182) (*P. multocida*)	AF067175	43.0	68.1

[a] Open reading frame of each capsule transport gene and corresponding nucleotide size in base pairs
[b] Predicted proteins for each capsule transport gene and protein size in amino acids
[c] GenBank accession numbers of capsular transport sequences from related bacterial species

acids. *hctB* terminates with a TGA stop codon where it overlaps with the *hctA* start codon. *hctA* contains 648 nucleotides, encodes a putative protein of 215 amino acids and terminates at a TAG stop codon. Downstream of *hctA* all reading frames in both directions are closed with multiple stop codons. Part of an open reading frame is present upstream of *hctD*, which showed considerable homology with the *P. multocida hyaA* biosynthesis gene. The overlapping stop and start codons in the *hct* genes (Fig. 4 and 5) indicate that these four genes are probably transcriptionally coupled.

DISCUSSION

Analysis of the *H. paragallinarum* hctDCBA gene cluster revealed a clear bias toward codons rich in nucleotides A and T (37 % GC content) consistent with the 39 % GC content of the *H. influenzae* capsule gene cluster (Kroll, Loynds, Brophy & Moxon 1990) and 37 % GC content of the *H. influenzae* genome overall (Roy & Smith 1973). It also correlates with the calculated GC contents of *A. pleuropneumoniae* (40 %) and *P. multocida* (37 %). The gene lengths and region size correlate well with those of related organisms, all belonging to the family Pasteurellaceae (Table 2). Blast searches of the combined, non-redundant nucleotide and protein databases at the National Centre for Biotechnology Information (NCBI) indicated that *H. paragallinarum hctDCBA* were highly homologous at both the nucleotide and amino acid levels to *H. influenzae bexDCBA* (Kroll *et al.* 1990), *A. pleuropneumoniae cpxDCBA* (Ward & Inzana 1997), *P. multocida hexDCBA* (Chung, Zhang & Adler 1998) and *Neis-*

seria meningitidis ctrABCD (Frosch, Muller, Bousset & Muller 1992) (Table 2).

The predicted amino acid sequences of the *hct* genes showed significant identity with the capsule transport genes of related organisms. The predicted HctA protein showed on average 77 % identity and 85.4 % similarity with the A proteins of *H. influenzae*, *A. pleuropneumoniae* and *P. multocida*. HctA contains the ATP-binding domains A (GXLGRXGXGKS) and B (XXDNLRFI) (Walker, Sarste, Runswick & Gay 1982) at amino acids 1080–1092 and 1131–1138 respectively (Fig. 5, regions C1 and C2), which are conserved in the BexA and CpxA homologues (Kroll *et al.* 1990; Fath & Kolter 1993; Ward & Inzana 1997). The nucleotide homology as well as the high degree of similarity between homologous proteins, supports the speculation that *hct*A might encode an ATP-binding protein component of a polysaccharide export apparatus.

HctB protein showed an average of 58 % identity and 84.4 % similarity with its corresponding homologues and is predicted to be a hydrophobic protein over most of its length, containing at least six potential membrane-spanning a-helical domains (Kyte & Doolittle 1982; Kroll *et al.* 1990). A short relatively hydrophilic region starting at amino acid 990 (Fig. 5, region B) aligned with a similar region in OppB of *Salmonella typhimurium* (Hiles, Gallagher, Jamieson & Higgins 1987) and BexB from *H. influenzae* (Kroll *et al.* 1990). Furthermore, each showed a marginal sequence similarity to a consensus thought to be involved in intermolecular interactions in the oligopeptide transporter (Dassa & Hofnung 1985). Fig. 5 (region B) shows the position of this sequence on HctB and Fig. 6 shows an alignment of the relatively hydrophilic portions of HctB, BexB and OppB. HctB is therefore a candidate for an integral inner-membrane component of the putative polysaccharide exporter.

The multiple protein sequence alignment of HctC with the respective C proteins of *H. influenzae*, *A. pleuropneumoniae* and *P. multocida* showed a lower homology (average of 45.8 % identity and 68.2 % similarity) in comparison to HctA and B with their corresponding homologues. Transposon mutagenesis of *bexC* (Kroll *et al.* 1990) suggested that this gene might be a periplasmic protein. Prediction of protein subcellular localisation of the HexC protein performed with PSORT (Nakai & Kahehisa 1991), suggested an inner membrane protein, possibly with a periplasmic domain, concurring with the transposon mutagenesis data on BexC (Chung *et al.* 1998). The N-terminus of BexC containing phosphatase activity suggests that the protein is either excreted into the periplasm with cleavage of an N-terminal leader peptide or anchored in the bacterial inner membrane by an uncleaved N-terminal domain to protrude into the periplasm. It is therefore a candidate for a periplasmically orientated component of a capsular polysaccharide exporter. Ward & Inzana (1997) predicted the CpxC protein of *A. pleuropneumoniae* to be relatively hydrophilic with hydrophobic domains near the N and C-termini that may serve as membrane anchors. Three long hydrophobic stretches of amino acid sequence with membrane-spanning potential allowing the possibility of anchoring at more than one site have been identified in BexC (Kroll *et al.* 1990). Similar stretches of sequence are present in HctC at amino acids 421–442, 458–475 (Fig. 5, regions A1 and A2) at the proposed N-terminal and 753–777 at the C-terminal (Fig. 5, region A3). Considering this information and the facts known about the HctC homologues, it is proposed that this protein serves as the second component of a protein complex involved in polysaccharide export across the cytoplasmic membrane (Reizer, Reizer & Saver 1992).

HctD showed an average of 42.6 % identity and 65.3 % similarity with the predicted D proteins of *H. influenzae*, *A. pleuropneumoniae* and *P. multocida*. HctD showed similarity of 63 % with BexD and 62.5 % with CrtA from *H. influenzae* and *N. meningitidis* respectively. CtrA from *Neisseria meningitidis* is believed to be an outer membrane protein with porin properties (Frosch *et al.* 1992). In addition, BexD and its homologues is believed to be outer membrane associated (Kroll *et al.* 1990; Rosenow, Esumah, Roberts & Jann 1995), mutations in the *bexD* gene coding for this corresponding protein accumulated polysaccharides in the periplasmic space (Bronner, Clarke & Whitfield 1994). Based on these similarities with CtrA and BexD, HctD it is probably an outer membrane protein involved in capsular polysaccharide transport across the outer membrane, possibly with porin properties.

These data are therefore consistent with the hypothesis that the *hctABCD* gene cluster encodes proteins that form an export complex for capsule polysaccharides. The findings will greatly facilitate the investigation at molecular level of the role of the *H. paragallinarum* capsule in pathogenesis. However, confirmation of the importance of each gene product and elucidation of the function of each protein will require characterization of the phenotypic impact of in-frame deletions or other mutations in the respective genes. In-frame deletions might lead

to reduced virulence with the possible use as a live vaccine.

ACKNOWLEDGEMENTS

This study was supported financially by the National Research Foundation (NRF) as a student scholarship to O. de Smidt and core funding to J. Albertyn. Financial contributions from the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State are gratefully acknowledged.

REFERENCES

BLACKALL, P.J. 1989. The avian *haemophili*. *Clinical Microbiology Reviews*, 2:270–277.

BLACKALL, P.J. & YAMAMOTO, R. 1990. Infectious coryza, in *A laboratory manual for the isolation and identification of avian pathogens* (3rd ed.), edited by H.G. Prachase, L.H. Arp, C.H. Domermuth & J.E. Pearson. Ames, Iowa: American Association of Avian Pathologists.

BOULNOIS, G.J., & ROBERTS, I.S. 1990. Genetics of capsular polysaccharide production in bacteria. *Current Topics in Microbiology and Immunology*, 150:1–18.

BOULNOIS, G.J., ROBERTS, I.S., HODGE, R., HARDY, K.R., JANN, K.B. & TIMMIS, K.N. 1987. Analysis of the K1 capsule biosynthesis genes of *Escherichia coli:* definition of three functional regions for capsule production. *Molecular and General Genetics*, 208:242–246.

BOYCE, J.D. & ADLER, B. 2000. The capsule is a virulence determinant in the pathogenesis of *Pasteurella multocida* M1404 (B:2). *Infection and Immunity*, 68:3463–3468.

BOYCE, J. D., CHUNG, J. Y. & ADLER, B. 2000. Genetic organization of the capsule biosynthetic locus of *Pasteurella multocida* M1404 (B:2). *Veterinary Microbiology*, 72:121–134.

BRAGG, R.R., COETZEE, L. & VERSCHOOR, J.A. 1996. Changes in incidence of different serovars of *Haemophilus paragallinarum* in South Africa: a possible explanation for vaccination failures. *Onderstepoort Journal of Veterinary Research*, 63:217–226.

BRONNER, D., CLARKE, B.R. & WHITFIELD, C. 1994. Identification of an ATP-binding cassette transport system required for the translocation of lipopolisaccharide O-antigen side-chains across the cytoplasmic membrane of *Klebsiella pneumoniae* serotype O1. *Molecular Microbiology*, 14:505–519.

CHOMCZYNSKI, P. 1992. One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Analytical Biochemistry*, 201:134–139.

CHUNG, J. Y., ZHANG, Y. & ADLER, B. 1998. The capsule biosynthetic locus of *Pasteurella multocida* A:1. *FEMS Microbiology Letters*, 166:289–296.

DASSA, E. & HOFNUNG, M. 1985. Sequence of gene *malG* in *E.coli* K12: homologies between integral membrane components form binding protein-dependent transport systems. *EMBO Journal*, 4:2287–2293.

FATH, M.J. & KOLTER, R. 1993. ABC transporters: bacterial exporters. *Microbiology Reviews*, 57:995–1017.

FROSCH, M., EDWARDS, U., BOUSSET, K., KRAUBE, B. & WEISGERBER, C. 1991. Evidence for a common molecular origin of the capsule gene loci in gram-negative bacteria expressing group II capsular polysaccharides. *Microbiology*, 5:1251–1263.

FROSCH, M., MULLER, D., BOUSSET, K. & MULLER, A. 1992. Conserved outer membrane protein of *Neisseria meningitidis* involved in capsule expression. *Infection and Immunity*, 60:789–803.

HANAHAN, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, 166:557–580.

HILES, I.C., GALLAGHER, M.P., JAMIESON, D.J. & HIGGINS, C.F. 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium. Journal of Molecular Biology*, 195:125–142.

INOUE, H., NOJIMA, H. & OKAYAMA, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96: 23–28.

INZANA, T.J., TODD, J. & VEIT, H. P. 1993. Safety, stability and efficiency of noncapsulated mutants of *Actinobacillus pleuropneumoniae* for the use in live vaccines. *Infection and Immunity*, 61:1682–1686.

JANN, B. & JANN, K. 1990. Structure and biosynthesis of the capsular antigens of *Escherichia coli*. *Current Topics in Microbiology and Immunology*, 150:19–42.

KROLL, J.S. 1992. The genetics of encapsulation in *Haemophilus influenzae*. *Journal of Infectious Diseases*, 165:3343–3347.

KROLL, J.S., HOPKINS, I. & MOXON, E.R. 1988. Capsule loss in *H. influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. *Cell*, 53:347–356.

KROLL, J.S., ZAMZE, S., LOYND, B. & MOXON, E.R. 1989. Common organization of chromosomal loci for the production of different capsular polysaccharides in *Haemophilus influenzae*. *Journal of Bacteriology*, 171:3343–3347.

KROLL, J.S., LOYNDS, B., BROPHY, L.N. & MOXON, E.R. 1990. The *bex* locus in encapsulated *Haemophilus influenzae*. A chromosomal region involved in capsule polysaccharide export. *Molecular Microbiology*, 4:1853–1862.

KYTE, J. & DOOLITTLE, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157:105–132.

LEE, C.J. 1987. Bacterial capsular polysaccharides biochemistry, immunity and vaccine. *Molecular Immunology*, 24: 1005–1019.

NAKAI, K, & KAHEHISA, M. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins*, 11:95–110.

REIZER, J., REIZER, A. & SAVER JR., M.H. 1992. A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. *Protein Science,* 1: 1326–1332.

ROSENOW, C., ESUMAH, F., ROBERTS, I.S. & JANN, K. 1995. Characterization and localization of the KpsE protein of *Escherichia coli* K5, which is involved in polysaccharide export. *Journal of Bacteriology*, 177:1337–1143.

ROY, P.H. & SMITH, H.O. 1973. DNA methylases of *H. influenzae* Rd. *Journal of Molecular Biology*, 81:427–444.

SAMBROOK, J., FRITSCH, E.F. & MANIATIS, T. 1989. Molecular cloning, in *A laboratory manual* (2nd ed.). New York: Cold Spring Harbour Laboratory Press.

TOWNER, P. 1991. A practical approach, in *Essential Molecular Biology*, edited by T.A. Brown. Oxford: IRL Press.

WALKER, J.E., SARSTE, M., RUNSWICK, M.J. & GAY, N.J. 1982. Distantly related sequences in the a- and b-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO Journal*, 1:945–951.

WARD, C.K. & INZANA, T.J. 1997. Identification and characterization of a DNA region involved in the export of capsular polysaccharides by *Actinobacillus pleuropneumoniae* serotype 5a. *American Society of Microbiology*, 65:2491–2496.

YAMAMOTO, R. 1991. Infectious coryza, in *Diseases of Poultry* (9th ed.), edited by M.S. Hofstad, H.J. Barnes, B.W. Calnek, W.M. Reid & H.E. Yonder Jr. Iowa State University Press.