CHARACTERISATION OF THE CAPSULAR POLYSACCHARIDE BIOSYNTHESIS LOCI OF *STREPTOCOCCUS PNEUMONIAE* SEROGROUP 19

by

Judy Kay Morona B. Sc. (Adelaide)

A thesis submitted for the degree of Doctor of Philosophy

Department of Microbiology and Immunology
University of Adelaide

November, 1998

ABSTRACT

In this thesis, the genetic loci encoding capsular polysaccharide synthesis (*cps*) have been characterised for all members of *Streptococcus pneumoniae* serogroup (19F, 19A, 19B and 19C). In each serotype, the *cps* locus is located in the *S. pneumoniae* chromosome between *dexB* and *aliA* and appears to be arranged as a single transcriptional unit. The arrangement of the genes within the *cps19* loci is highly conserved with 13 genes (*cps19A-H*, and *K-O*) common to all four serogroup 19 members. These genes encode functions required for the synthesis of the shared trisaccharide component of the group 19 capsular polysaccharide (CPS) repeat unit structures. Furthermore, the genetic differences between the group 19 *cps* loci identified are consistent with the differences in the CPS structures of individual serotypes. Functions have been assigned to nearly all the *cps19* gene products, based on either gene complementation or similarity to other proteins with known functions. This has enabled biosynthetic pathways for production of all four group 19 CPSs to be proposed.

Nearly all of the common genes from types 19F, 19B and 19C are >95% identical to each other. However, closely related homologues of *cps19fI* and *J*, which encode the type 19F polysaccharide polymerase and repeat unit transporter, respectively, are not found in the type 19B and 19C *cps* loci. In type 19B and 19C this region of the *cps* locus (between *cps19bH* and *cps19bK*) contains five genes which encode a unique polysaccharide polymerase and repeat unit transporter, as well as two additional putative glycosyl transferases and a protein which may be involved in synthesis of an activated ribose precursor. Transformation studies indicated that these five genes encode all of the functions required to convert a type 19F pneumococcus to type 19B. The type 19C *cps*

locus differs from the 19B *cps* locus only in the insertion of a glucosyl transferase gene (*cps19cS*) between *cps19cK* and *cps19cL*. Transformation studies have shown that the presence of this gene accounts for the additional glucose side chain in the otherwise identical repeat unit structures. The type 19C *cps* locus contains 19 genes, and at 21 kb it is the largest pneumococcal capsule gene cluster characterised to date.

Although the *cps19a* and *cps19f* loci are identical in the number and arrangement of the genes present, the similarity between individual genes varies from 70% to 99% identity (for both the nucleotide and the deduced amino acid sequences). This sequence divergence is surprising given that the only difference between their CPS repeat units is the glycosidic linkage which joins the repeat units together ($\alpha(1\rightarrow2)$ for 19F and $\alpha(1\rightarrow3)$ for 19A). Theoretically, only a difference in the *cps19aI* gene, which presumably encodes the polysaccharide polymerase responsible for this linkage, is required to change a type 19F pneumococcus into type 19A. Indeed, this was demonstrated by a transformation event in which the region of the *cps19a* locus encoding Cps19aH and Cps19aI replaced the homologous portion of the *cps19f* locus was sufficient to convert CPS type from 19F to 19A. Given that Cps19fH and Cps19aH are >95% identical, it seems probable that Cps19aI (79% identity) is solely responsible for the observed alteration in CPS type.

The serotype specificity of the *cps19f* genes was investigated by Southern hybridisation analysis of chromosomal DNA from other *S. pneumoniae* serotypes. Large variations in the hybridisation patterns were obtained with the different gene-specific probes. Probes specific for sequences flanking *cps19f* hybridised with all the serotypes tested. However, within the *cps* loci, only *cps19fA* and *cps19fB* were common to all serotypes. Based on the Southern hybridisation analysis a protocol for PCR amplification of *cps* loci was developed and used to amplify the *cps* regions from a variety of pneumococcal serotypes. Direct sequencing of the 5' end of the PCR products was undertaken and identified two classes of *cpsC* gene. Southern hybridisation studies with

cps19aC- and *D*-specific gene probes demonstrated that homologues of the first four genes in the *cps* locus, *cpsA-D,* are present in all serotypes and that all the *cps* loci tested evolved from one of two clonal origins which contained either class I or class II *cpsC* and *D* genes. The *cpsE* gene, which encodes a glucose-1-phosphate transferase, is also conserved (in the two distinct classes) in the *cps* loci of all serotypes tested which contain glucose in their CPS, except type 3.

The sequence analysis of the various *cps* loci presented in this thesis provides further evidence that in nature frequent recombination occurs between different *cps* loci resulting in either complete exchange of the *cps* locus or exchange of only part of the *cps* locus, and could potentially result in the expression of new capsular serotypes.

TABLE OF CONTENTS