AN ABSTRACT OF THE DISSERTATION OF

Wasna Viratyosin for the degree of Doctor of Philosophy in Microbiology presented on June 6, 2002 Title: Genetic Variation of Chlamydial Inc Proteins

Abstract approved: Daniel D. Rockey

Genomic analysis is a new approach for the characterization and investigation of novel genes, gene clusters, the function of uncharacterized proteins, and genetic diversity in microorganisms. These approaches are important for the study of chlamydiae, a system in which several genomes have been sequenced but in which techniques for genetic manipulation are not available. The objective of this thesis is to combine computer-based analysis of chlamydial inclusion membrane proteins (Incs) with cellular and molecular biological analysis of the bacteria. Three different experimental lines of investigation were examined, focusing on Incs of *C. trachomatis* and *C. pneumoniae*.

Chlamydiae are obligate intracellular bacteria that develop within a non-acidified membrane bound vacuole termed an inclusion. Putative Inc proteins of *C. trachomatis* and *C. pneumoniae* were identified from genomic analysis and a unique structural motif. Selected putative Inc proteins are shown to localize to the inclusion membrane.
*Chlamydia trachomatis* variants with unusual multiple-lobed, non-fusogenic, inclusion were identified from a large scale serotyping study. Fluorescence microscopy showed that IncA, a chlamydial protein localized to the inclusion membrane, was undetectable on non-fusogenic inclusions of these variants. Sequence analysis of *incA* from non-fusogenic variant isolates revealed a defective *incA* in most of the variants. Some variants lack not only IncA on the inclusion membrane but also CT223p, an additional Inc protein. However, no correlation between the absence of CT223p and distinctive inclusion phenotype was identified. Nucleotide sequence analysis revealed sequence variations of *C. trachomatis incA* and CT223 in some variant and wild type isolates.

Comparative analyses of the three recently published *C. pneumoniae* genomes have led to the identification of a novel gene cluster named the CPn1054 gene family. Each member of this family encodes a polypeptide with a hydrophobic domain characteristic of proteins localized to the inclusion membrane. These studies provided evidence that gene variation might occur within this single collection of paralogous genes. Collectively, the variability within this gene family may modulate either phase or antigenic variation, and subsequent physiologic diversity, within a *C. pneumoniae* population.

These studies demonstrate the genetic diversity of Inc proteins and candidate Inc proteins, within and among the different chlamydial species. This work sets the stage for further investigations of the structure and function of this set of proteins that are likely critical to chlamydial intracellular growth.
Genetic Variation of Chlamydial Inc Proteins

By

Wasna Viratyosin

A DISSERTATION

Submitted to

Oregon State University

In Partial Fulfillment
of the Requirements for the
Degree of

Doctor of Philosophy

Presented June 6, 2002
Commencement June 2003
Doctor of Philosophy dissertation of Wasna Viratyosin presented on June 6, 2002

APPROVED:

Redacted for Privacy

Major Professor, representing Microbiology

Redacted for Privacy

Head of the Department of Microbiology

Redacted for Privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for Privacy

Wasna Viratyosin, Author
ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health Award # R29AI42869 and RO1AI48769, and the Oregon State University Department of Microbiology N. L. Tartar Award Program. I would like to thank the Thailand National Center of Genetic Engineering and Biotechnology and the Royal Thai Scholarship program for their funding and support.

On a professional level I have to first and foremost acknowledge my major advisor, Dan Rockey, for mentoring my educational development as a scientist and a professional, and for helping me fulfill my goals. Thanks for all the advice and support.

I would like to thank my committee members Dr. Dennis Hruby, Dr. Lyle Brown, Dr. Jerry Heidel, and Dr. George Bailey for their time and support. Additionally, I would like to thank the rest of the faculty, especially Dr. Jo-Ann Leong, for time, advice, and support.

Thank you to the past and present members of the Rockey lab and my other friends I have made over the years at OSU for being there to talk, share enthusiasm, and have fun. Thanks to everyone for their friendship, cheerful support, and laughs. I would like to thank all my family: Mom, my sisters and brothers who always love and support me.
# TABLE OF CONTENTS

## INTRODUCTION

Page

1

## LITERATURE REVIEW

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular phylogeny of <em>Chlamydia</em></td>
<td>3</td>
</tr>
<tr>
<td>Variants of chlamydiae</td>
<td>4</td>
</tr>
<tr>
<td>Developmental cycle</td>
<td>6</td>
</tr>
<tr>
<td>Elementary bodies and reticulate bodies</td>
<td>8</td>
</tr>
<tr>
<td>Cell surface associated proteins of <em>Chlamydia</em></td>
<td>10</td>
</tr>
<tr>
<td>Major outer membrane protein (MOMP)</td>
<td>11</td>
</tr>
<tr>
<td>Cysteine-rich membrane proteins (CRPs)</td>
<td>14</td>
</tr>
<tr>
<td>PorB</td>
<td>15</td>
</tr>
<tr>
<td>Polymorphic membrane proteins (Pmps)</td>
<td>16</td>
</tr>
<tr>
<td>Lipopolysaccharide (LPS)</td>
<td>19</td>
</tr>
<tr>
<td>Inclusion membrane proteins (Incs)</td>
<td>20</td>
</tr>
<tr>
<td>Genetic variation in chlamydial cell surface proteins</td>
<td>29</td>
</tr>
<tr>
<td>Genetic and phenotypic variation in other bacterial pathogens</td>
<td>32</td>
</tr>
</tbody>
</table>

## MATERIALS AND METHODS

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type and mutant <em>C. trachomatis</em> isolates</td>
<td>35</td>
</tr>
<tr>
<td><em>C. pneumoniae</em> isolates, plasmids and genomes</td>
<td>37</td>
</tr>
<tr>
<td>Chlamydial cell cultures</td>
<td>39</td>
</tr>
<tr>
<td>Purification of elementary bodies</td>
<td>39</td>
</tr>
<tr>
<td>TABLE OF CONTENTS, continued</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>Identification of putative Inc proteins using hydrophilicity analysis</td>
<td>40</td>
</tr>
<tr>
<td>Production of putative Inc proteins</td>
<td>41</td>
</tr>
<tr>
<td>Purification of maltose fusion proteins</td>
<td>44</td>
</tr>
<tr>
<td>Antibody production</td>
<td>46</td>
</tr>
<tr>
<td>Antisera and reagents for non-fusogenic variant studies</td>
<td>46</td>
</tr>
<tr>
<td>Immunofluorescence microscopy</td>
<td>47</td>
</tr>
<tr>
<td>Preparation of Chlamydiae infected cells for protein gel electrophoresis</td>
<td>49</td>
</tr>
<tr>
<td>Preparation of LCR lysates for amplification reaction</td>
<td>49</td>
</tr>
<tr>
<td>Amplification reaction and DNA sequence analysis of <em>incA</em> and CT223</td>
<td>50</td>
</tr>
<tr>
<td>Electrophoresis and immunoblot</td>
<td>52</td>
</tr>
<tr>
<td><em>In silico</em> analysis of the CPn1054 gene family of <em>C. pneumoniae</em> genomes</td>
<td>53</td>
</tr>
<tr>
<td>Phylogenetic analyses</td>
<td>54</td>
</tr>
<tr>
<td>DNA amplification, and sequence analysis of the CPn1054 gene family</td>
<td>54</td>
</tr>
<tr>
<td>Recombinant clones and sequence analyses</td>
<td>55</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS, continued

## RESULTS

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification and bioinformatic analysis of putative Inc proteins</td>
<td>58</td>
</tr>
<tr>
<td>Protein localization using immunofluorescence staining</td>
<td>63</td>
</tr>
<tr>
<td>Fluorescence microscopy of cells infected non-fusogenic variant isolates</td>
<td>66</td>
</tr>
<tr>
<td>Immunoblot analysis of selected non-fusogenic variant isolates</td>
<td>71</td>
</tr>
<tr>
<td>Sequence analysis of (incA) of non-fusogenic variant isolates</td>
<td>73</td>
</tr>
<tr>
<td>DNA sequence analysis of ORF CT223</td>
<td>77</td>
</tr>
<tr>
<td>Upstream sequence analysis of (incA) and CT223</td>
<td>78</td>
</tr>
<tr>
<td>\textit{In silico} analysis of the CPn1054 gene family</td>
<td>81</td>
</tr>
<tr>
<td>The conserved hydrophobic domain of the CPn1054 gene family</td>
<td>85</td>
</tr>
<tr>
<td>Homopolymeric cytosine (poly C) tract and the variation of the length of poly C tract in the CPn1054 gene family</td>
<td>93</td>
</tr>
<tr>
<td>The tandem repeat motif in CPn0007</td>
<td>94</td>
</tr>
<tr>
<td>Framshift mutations and gene conversion in CPn0010</td>
<td>97</td>
</tr>
</tbody>
</table>

## DISCUSSION AND SUMMARY

Page 104

## BIBLIOGRAPHY

Page 124
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The <em>Chlamydia</em> developmental cycle of the infection and replication</td>
<td>7</td>
</tr>
<tr>
<td>2.</td>
<td>Electron microscopy of a chlamydial inclusion of <em>C. psittaci</em> in infected cell</td>
<td>9</td>
</tr>
<tr>
<td>3.</td>
<td>Hydropathy profiles for the secondary structure of known inclusion membrane proteins</td>
<td>25</td>
</tr>
<tr>
<td>4.</td>
<td>Schematic representation of the contig containing the selected <em>inc</em> genes in <em>C. trachomatis</em> genome</td>
<td>42</td>
</tr>
<tr>
<td>5.</td>
<td>Comparison of predicted proteins encoded by <em>C. trachomatis</em> CT058 and <em>C. pneumoniae</em> CPn0107, CPn0367, CPn0369 and CPn0370</td>
<td>62</td>
</tr>
<tr>
<td>6.</td>
<td>Immunofluorescence microscopy of <em>C. trachomatis</em> infected cell HeLa cells fixed with methanol and probed with specific antibody against each tested putative Inc/s</td>
<td>64</td>
</tr>
<tr>
<td>7.</td>
<td>Double-labeled fluorescence microscopy of <em>C. pneumoniae</em> labeled with anti-HSP60 (A) and anti-<em>C. pneumoniae</em> IncA (CPn0186)(B)</td>
<td>65</td>
</tr>
<tr>
<td>8.</td>
<td>The distinct pattern of the localization of CT223 on the inclusion membrane</td>
<td>67</td>
</tr>
<tr>
<td>9.</td>
<td>Fluorescence microscopy of McCoy cells infected with a serovar J(s) variant of <em>Chlamydia trachomatis</em> probed with antibodies against HSP60, IncA, and CT223</td>
<td>69</td>
</tr>
<tr>
<td>10.</td>
<td>Fluorescence microscopy of McCoy cells infected with serovar G(s) 459, an IncA positive variant, of <em>C. trachomatis</em> probed with antibodies against IncA</td>
<td>70</td>
</tr>
<tr>
<td>11.</td>
<td>Immunoblot analysis of non-fusogenic <em>C. trachomatis</em> isolates probed with antibodies against HSP60, IncA, and CT223p</td>
<td>72</td>
</tr>
<tr>
<td>12.</td>
<td>A graphic representation of the results from the DNA sequencing of <em>incA</em> encoded by 10 wild type and 25 non-fusogenic <em>C. trachomatis</em> isolates</td>
<td>74</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES (continued)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Multiple sequence alignment analysis of CT223p of fusogenic and non-fusogenic variant isolates of <em>C. trachomatis</em> isolates</td>
<td>79</td>
</tr>
<tr>
<td>14. A schematic representation of <em>C. pneumoniae</em> CWL029 contigs</td>
<td>82</td>
</tr>
<tr>
<td>15. Phylogenetic tree analysis of the DNA sequence of the CPn1054 gene family obtained from <em>C. pneumoniae</em> CWL029</td>
<td>84</td>
</tr>
<tr>
<td>16. The upstream and 5’ end sequence of the CPn1054 gene family</td>
<td>86</td>
</tr>
<tr>
<td>17. A schematic representation of the structure features and comparison of paralogous genes of CPn1054 gene family in <em>C. pneumoniae</em> CWL029</td>
<td>88</td>
</tr>
<tr>
<td>18. Hydropathy plot analysis of the members of the CPn1054 gene family</td>
<td>89</td>
</tr>
<tr>
<td>19. The conserved hydrophobic domain in the CPn1054 family</td>
<td>91</td>
</tr>
<tr>
<td>20. Phylogenetic tree analysis of amino acid sequence of hydrophobic domain of the CPn1054 gene family obtained from <em>C. pneumoniae</em> CWL029</td>
<td>92</td>
</tr>
<tr>
<td>21. Variation of the length of the poly C tract of the CPn1054 gene family</td>
<td>95</td>
</tr>
<tr>
<td>22. Schematic representation of the open reading frame of CPn0010-0010.1, 1054, 0045 identified in <em>C. pneumoniae</em> CWL029, AR39 and J138</td>
<td>99</td>
</tr>
<tr>
<td>23. Frameshift mutation and gene conversion in CPn0010</td>
<td>100</td>
</tr>
<tr>
<td>24. Gene conversion of CPn0010 in <em>C. pneumoniae</em> isolates</td>
<td>102</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Description of the wild type and non-fusogenic variants of *C. trachomatis* isolates used in this study 36

2. Description of the 12 independent *C. pneumoniae* isolates used in this study 38

3. Oligonucleotide primers used for preparation of putative inc/maleE hybrid recombinant clones for production of fusion proteins 43

4. Oligonucleotide primers used for PCR and nucleotide sequencing of the internal regions of the CPn1054 gene family from *C. pneumoniae* CWL029 57

5. Putative *inc* genes identified by hydrophilicity profile within each sequenced chlamydial genome 59

6. DNA sequence identity of the CPn1054 gene family using Pairwise of BLAST 2 sequence program 83
This thesis is dedicated to my mom and my family
for
love, encouragement and support
Genetic Variation of Chlamydial Inc Proteins

INTRODUCTION

Chlamydiae are obligate intracellular bacteria that cause a variety of human diseases. Two main human pathogens are *C. trachomatis* and *C. pneumoniae*. *Chlamydia trachomatis* is leading cause of preventable blindness and sexually transmitted disease (STD) including pelvic inflammatory disease, chronic pelvic pain, ectopic pregnancy, and epididymitis (Schachter *et al.*, 1990) in the developing and developed countries, respectively. *Chlamydia pneumoniae* is an important cause of community-acquired pneumonia (Saikku *et al.*, 1985) and bronchitis, and has recently been implicated in the etiology of atherosclerosis (Kuo *et al.*, 1993; Mlot *et al.*, 1996). Both pathogens represent a major public health burden.

Among intracellular bacterial pathogens, chlamydiae possess a unique biphasic development cycle. They reside and replicate in non-acidified membrane bound vacuoles termed an inclusion. Proteins that localize to the inclusion membrane are defined as inclusion membrane proteins (Incs). Little is known about the development of chlamydial inclusions and the functions of Inc proteins.

A major area of interest in chlamydial biology is to understand the interaction between the chlamydial inclusion and the host intracellular environment. The role of chlamydial inclusion membrane proteins in the
The chlamydia development cycle and pathogenesis remain to be elucidated. The lack of a method for conducting genetic manipulation in the chlamydiae is the primary limiting factor to attempts to characterize the unknown proteins.

Recently, the availability of several complete Chlamydial genomes provides an alternate approach to elucidate the chlamydial biology. The identification of new potential or functional assignment is extensively facilitated by genomic approaches. Novel gene clusters or families, transport mechanisms, and biochemical pathways are defined and characterized based on the sequence similarity, and structural domain obtained from the genome data.

The focus of this thesis is to identify the putative Inc proteins and/or a new gene family in the chlamydial genome based on structural motifs and the sequence similarity. We first use the unique bi-lobed hydrophobic domain commonly present in Inc proteins as a primarily predictive marker to determine the putative Inc (s) in the C. trachomatis and C. pneumoniae genomes. The second area of investigation is to determine the association of the non-fusogenic inclusion phenotype of C. trachomatis variant isolates and incA sequence. Finally, the comparative genomic analysis was used to identify a new gene family and study the possible role of genetic variations of paralogs encoded within the new gene family named the CPn1054 gene family in C. pneumoniae. Collectively, these studies described variability within the Chlamydiae not previously examined in the literature.
Molecular phylogeny of Chlamydia

All members of the order Chlamydiales are obligate intracellular bacteria that have a unique biphasic developmental cycle. They all have greater than 80% rDNA sequence identity with chlamydial 16S rRNA genes and/or 23S rRNA genes. The order Chlamydiales recently has been revised into four families: Chlamydiaceae, Parachlamydiaceae, Simkaniaceae, and Waddilaceae (Everett et al., 1999; Bush and Everett, 2001). The family Chlamydiaceae, which previously consisted of only a single genus, Chlamydia, was recently divided into two major genera, Chlamydia and Chlamydophila. Members of genus Chlamydia include C. trachomatis, C. muridarum, and C. suis in which the 16S rRNA and 23S rRNA gene sequences are 97% identity. The genus Chlamydophila consists of C. psittaci, C. pneumonae, C. pecorum, C. abortus, C. felis, and C. caviae, which their ribosomal genes are 95% identical. The nine species in these genera were decisively distinguished by analyses of phenotype, antigenicity, associated disease, host range, biological data, genomic endonuclease restriction, DNA-DNA hybridization information, and phylogenetic analysis using the ribosomal operon of 16S rRNA and 23S rRNA (Everett et al., 1999; Everett and Andersen, 1997; Katlenboeck et al., 1993; Pudjiatmoko et al., 1997).
The new proposed reclassification of the order *Chlamydiale*, however, remains a controversial issue (Schachter *et al.*, 2001). The minor sequence difference may not be sufficient to create a new genus or species. The taxonomically significant biological properties for genus or species differentiation in the proposed reclassification were not completely devised. Tanner *et al.* (1999) showed that *C. trachomatis*, *C. psittaci*, *C. pecorum*, and *C. pneumoniae* strains represent a taxonomically and phylogenetically coherent grouping of 16s rRNA into one group. Only 16s rRNA analysis, therefore, may not be a best choice for speciation. Recently, five other genes that encode for major outer membrane protein (MOMP), GroEL chaperonin, KDO-transferase, small cysteine-rich lipoprotein and 60 kDa cysteine-rich protein were used as key determinants for phylogenetic reconstruction (Bush and Everett, 2001). Phylogenetic analysis of those genes demonstrated all genes evolved in concert with the ribosomal genes. However, there are no common biological markers for genus or species differentiation between the organisms within both proposed genera.

**Variants of chlamydiae**

*Chlamydia trachomatis* is currently classified into 19 serovars and categorized into three serogroups or classes, which are: B class (B, Ba, D, Da, E, L1, L2 and L2a), C class (A, C, H, I, Ia, J, Ja, K and L3), and intermediate class (F and G). These groups are determined based on amino acid similarity of variable domains within the major outer membrane proteins (MOMP). The serovar
distribution of B, C, intermediate class and "mix serovars" are 49%, 27.9%, 20.5%, and 2.6%, respectively, in clinical infection (Dean et al., 2000). The classification is consistent with the serological determination of chlamydial strains using polyclonal antisera or monoclonal antibodies against major outer membrane proteins (MOMPs). Serovar A, B, Ba and C are responsible for endemic trachoma. Serovar D- K especially D, E, and F are the major prevalent C. trachomatis strains implicated in sexually transmitted disease (STD) worldwide while significant persistent infections are associated with cervical infection by the C class (Dean et al., 2000). The C class appears less virulent, and requires higher multiplicity of infection and longer incubation period, for replication in the tissue culture cells. A recent longitudinal study provided evidence that C. trachomatis serovar G is associated to the developing cervical cancer, but remains controversial.

*Chlamydia trachomatis* can also be subdivided into three biovars: trachoma, lymphogranuloma venereum (LGV), and mouse pneumonitis. The classification is based on sources, distinct differences in clinical disease, and differences in their *in vitro* cell culture growth characteristics. The trachoma biovar consists of human serovars A to K; plus Ba, Da, la. Serovar A-C are primarily associated with endemic trachoma, and serovar D-K are involved in urogenital tract infection. The LGV biovar contains human serovars L1, L2, L2a, and L3, which are invasive lymphotrophic, sexually transmitted agents. This biovar is more invasive *in vivo* and more readily infects cultured cells *in vitro* (Moulder et al., 1988). A mouse biovar causes pneumonitis in an experimental murine model but is not known to be
pathogenic for humans or other animals. Classification of *C. trachomatis* into different biovars at the molecular level remains questionable. Genotyping of *C. trachomatis* did not correspond to biovar designation. Sequence analysis of *ompA*, which encodes for major outer membrane proteins (MOMPs), could not differentiate among the serological complexes. Analysis showed that A-, C- and H-trachoma are closely related to the L3 serovar while B- and E- trachoma are related to the L1 and L2 serovar. However, the 16S rRNA analysis revealed that the trachoma biovariants are more closely related to each other than they are to LGV isolates.

*Chlamydia pneumoniae* is classified into three distinct biovars: TWAR, Koala, and Equine (Everett et al., 1999) Different *C. pneumoniae* isolates have 94-100% homology with each other but less than 10% DNA homology with *C. psittaci* and less than 5% DNA homology with *C. trachomatis* (Kuo et al., 1995).

**Developmental cycle**

Chlamydiae are obligate intracellular pathogens that possess a unique developmental cycle. The cycle consists of two distinctive forms: elementary bodies (EBs) and reticulate bodies (RBs) (Figure 1). Elementary bodies are infectious, metabolically inactive forms while reticulate bodies are metabolically active but noninfectious forms. The nature of the development cycle is the conversion of EB to RB. This cycle begins when EBs come to attach to susceptible
Figure 1. The *Chlamydia* development cycle of the infection and replication. This cycle typically take 48-72 h although some species have longer cycles.
eukaryotic cells leading to internalization. Within the first 2 h following internalization, EBs, which entirely reside in the non-acidified membrane bound vacuoles known as an inclusion, differentiate into RBs. RBs undergo multiple round of binary fission. As bacterial multiplication, the inclusion expands. After 24-48 hours post infection (hpi), RBs differentiate back into infectious EBs which are subsequently released through host cell lysis. In most chlamydiae, this cycle is complete within 48-72 hpi.

**Elementary bodies and reticulate bodies**

Elementary bodies (EBs) and reticulate bodies (RBs) are primarily defined by morphological criteria (Ward, 1983)(Figure 2). EBs are small structures which are from 0.2 to 0.4 μm in diameter and possess an electron-dense nucleoid. RBs are much larger pleomorphic forms (1.0 to 1.5 μm in diameter). They are less dense, and possess an evenly dispersed, reticulate cytoplasm. Intermediate bodies are present at the transitional point in the developmental cycle. Aberrant RB forms also are present in either gamma interferon or ampicillin treatment and tryptophane depletion condition. EBs are denser and more resistant to mechanical and osmotic stress than are RBs. (Tamura et al. 1967 a, b; Caldwell et al., 1981). The DNA of *C. trachomatis* EB is packed into a condensed nucleoid structure by two histone like proteins, Hc1 and Hc2 (Christiansen et al. 1993; Hackstadt et al., 1991).
Figure 2. Electron microscopy of a chlamydial inclusion of *C. psittaci* in infected cell. Morphological differences between EBs and RBs multiplying binary fission were evident.
Elementary bodies were previously described as dormant like bodies. Recently, whether EBs really are metabolically inactive bodies became a question. It has been shown that EBs carry a pool of ATP. Several active transcripts were identified in EBs (Douglas et al., 2000). The fact was further strengthened by proteomic analysis of the *C. pneumoniae* EBs (Vandahl et al., 2001). The findings suggested that EBs are capable of metabolizing energy and thereby providing fuel for protein expression and active transport of proteins at the very beginning of the developmental cycle. The study also showed that EBs carry a large number of proteins involved in transcription and translation that were functionally active. These findings suggest that EBs are nondividing but metabolically active, at least at a minimal level.

**Cell surface associated proteins of chlamydiae**

*Chlamydiae* are phylogenetically and structurally classified as Gram-negative bacteria. The chlamydial envelope is similar to the outer envelope of gram-negative bacteria, consisting of an outer membrane, an inner membrane and a periplasmic space. However, unlike typical gram-negative bacteria, chlamydiae possess a unique cell envelope containing a cysteine-rich outer membrane, an inner membrane and a periplasmic space, and deficient or novel peptidoglycan structure. Chlamydial outer membranes consist largely of disulfide-cross-linked major outer membrane proteins (MOMP), two cysteine-rich proteins, polymorphic outer membrane proteins (Pmp) and a unique truncated lipopolysaccharide (LPS).
Presumably, the cross-linked supra-macromolecular complex may be the functional equivalent of peptidoglycan (Hatch and Everett, 1995).

**Major outer membrane protein (MOMP)**

MOMP is the predominant surface-exposed chlamydial protein. It is present in both RB(s) and EB(s) (Stephen *et al.*, 1986) and constitutes 60% of the total outer membrane protein complex (Caldwell *et al.*, 1981). Although MOMP is present throughout the developmental cycle, it is structurally different in the two developmental forms (Hatch *et al.*, 1984, 1986; Hackstadt *et al.*, 1985; Newhall *et al.*, 1987). In the RB, MOMP is predominately a monomer whereas in the EB, MOMP is present as a dimer, trimer, and multimeric complex in association with truncated LPS (Birkelund *et al.*, 1988).

MOMP is a 40kDa outer membrane protein with a pI of 4.9 that is synthesized most actively at 20 hpi. It is encoded by *ompA* (Stephen *et al.*, 1986), formerly referred as *omp1*, which is transcribed from early in development through the entire cycle. It is a chlamydial outer membrane protein that exhibits structural heterogeneity between different members of the chlamydiae species. Sequence variation in MOMP is confined mainly to four hypervariable domains (VDs) that are flanked and interspaced by five conserved domains (Yuan *et al.*, 1989).

The antigenic determinants in the hypervariable domains of *ompA* elicit the serovar- subspecies-, serogroup-, and species-specific antibodies (Caldwell and Schachter, 1982; Newhall *et al.*, 1987; Stephens *et al.*, 1982). *C. trachomatis*
serovar variants were analyzed and defined as deduced amino acid sequences of OmpA, that varied by one or more than 1 amino acid from the prototype serovars for variable domains. Therefore, MOMP represents the primary chlamydial serotyping antigen of *C. trachomatis* isolates (Yuan *et al.*, 1989; Zhang *et al.*, 1989).

The significance of variable domains was studied and showed that VD2 and VD4, which are surface-exposed epitopes of MOMP, play a role in attachment since treatment of *C. trachomatis* with specific antibodies to these epitopes blocks attachment and infectivity. VD3 is the smallest and least variable domain, and responsible for potential immunogenicity. Studies of protective or neutralizing antibody (Allen *et al.*, 1991), and immune specificity of murine T-cell against MOMP (Ishizaki *et al.*, 1992) reveals that an important antigenic T-cell epitope of *C. trachomatis* is located at the upstream and within VD3.

Several studies strongly suggest that MOMP plays an important role in pathogenesis. Treatment of infectious EBs with trypsin is correlated with reduced attachment of chlamydiae to HeLa cells (Hackstadt *et al.*, 1985; Su *et al.*, 1988, 1990). Monoclonal antibodies specific to conformation-dependent determinants of *C. pneumonia* MOMP are also potent in the neutralization of infectivity *in vitro* (Wolf *et al.*, 2001). Collectively, MOMP is such an immunodominant antigen that it can be potentially a candidate for a subunit vaccine against chlamydial infection.
Analyses of two *C. psittaci* strains, guinea pig inclusion conjunctivitis (GPIC) and meningopneumonitis (Mn) strain, and *C. trachomatis* revealed the structural conserved region of MOMP, which seven of eight cysteine residues and the four variable domains are identified at precisely the same positions (Zhang et al., 1989). In all *C. pneumoniae* isolates, protein profiles are identical, with a prominent 39.5 kDa band homologue to the MOMPs of other chlamydiae (Campbell et al., 1990; Iijima et al., 1994). All seven cysteine residues involved in disulfide-linked structure of other chlamydial MOMPs are similarly conserved in *C. pneumoniae* (Melgosa et al., 1994). Comparative genomic analyses also revealed that *ompA* of *C. trachomatis* and *C. pneumoniae* have striking sequential and structural similarities (Stephen et al., 1998). Such conserved structure suggests the importance of structure and conformation of the MOMP protein.

Sequence analyses of VD4, the largest hypervariable region of MOMP, were shown to be identical in 13 different *C. pneumoniae* isolates (Keltenboeck et al., 1993). It is likely that *C. pneumoniae* MOMP was conserved among different isolates. In previous studies (Campbell et al., 1990; Christiansen et al., 1999) immunoblot and immunofluorescence analysis of *C. pneumoniae* infected sera failed to detect MOMP, which suggested that *C. pneumoniae* MOMP may be immunorecessive and not surface accessible. However, recent work by Wolf et al. (2001) demonstrated that the *C. pneumoniae* MOMP contains nonlinear and surface-exposed epitopes. Their study also suggested that these epitopes are conformation-dependent and immunodominant.
Several studies directly or indirectly suggest that the MOMP has a role in the adhesion of *C. trachomatis* and *C. psittaci* to host cells. The attachment process to the host ligand was demonstrated in recombinant *E. coli* expressed maltose binding protein (MBP)–MOMP fusion protein and visualized by scanning electron microscopy (Su *et al.*, 1996). MOMP has also been shown to bind heparin in a concentration-dependent manner. Excess heparin or heparan sulfate inhibits binding of MBP-MOMP to eukaryotic cells. Mutant eukaryotic cells defective in glycosaminoglycan elicit the reduction of binding of MBP-MOMP and native EBs.

**Cysteine-rich outer membrane proteins (CRPs)**

Major cysteine-rich outer membrane proteins (CRPs) were the first outer membrane complex proteins identified in the sarkosyl-insoluble fraction of purified outer membrane protein of EBs. They are encoded by *OmcA* and *OmcB* (Lanmden *et al.*, 1990), and *envA* and *envB*, in *C. trachomatis* and *C. psittaci* respectively (Everett *et al.* 1994). *OmcA* is the smaller cysteine rich membrane proteins, which is annotated as 9 Kda-CRP. In contrast to *ompA*, *omcB* is extremely conserved and the gene product is present only in EB (Hatch *et al.*, 1984, 1986) in the ratio of 5 MOMP: 2 OmcA:1OmcB (Everett *et al.*, 1991) and does not appear to be surface exposed (Stephens *et al.*, 2002). These lipoproteins are structurally similar to murein lipoproteins present in gram- negative bacteria (Everett *et al.*, 1994).
The cysteine-rich outer membrane is considered to be primarily responsible for maintaining the structural integrity of the outer membrane in the absence of peptidoglycan (Newhall and Jones, 1983; Hatch et al., 1984; Hackstadt et al., 1985). The cysteine residues within CRP are highly conserved in all chlamydia species, suggesting function in formation of disulfide-like complex structure (Everett et al., 1991). Proteolytic treatment or heat denaturation blocked the binding of OmcB to the host cells. It indirectly suggested OmcB may play a role in the attachment process.

PorB

PorB is a relatively cysteine-rich protein predicted to be in the outer membrane. Like CRPs, it is a highly conserved gene present in the Chlamydiales. Its MW is 39 kDa and a predicted pI is 4.9, which are similar to those of MOMP. PorB possesses a predicted leader sequence with an amino acid sequence that ends in phenylalanine, a typical characteristic of outer membrane proteins of Gram-negative bacteria. It is present in the outer membrane complex of EB and is surface accessible in C. trachomatis (Kubo et al., 2000). PorB is present in the outer membrane in much smaller amounts than MOMP. Its transcription and translation were found throughout the chlamydial development cycle. Functional studies using in vitro liposome-swelling assay revealed that PorB is a substrate-specific porin that preferentially facilitates diffusion of dicarboxylates through the outer membrane (Kubo et al., 2001). It may be responsible for the diffusion of some
metabolites required for chlamydial growth. Furthermore, immunological studies in tissue culture showed that it is a target for antibody-mediated neutralization.

**Polymorphic membrane proteins (Pmps)**

The Pmp(s) were identified as a 90 kDa antigenic protein family and initially named putative outer membrane proteins (POMP) in *C. psittaci*. POMPs were renamed the polymorphic membrane proteins (Pmp) because they may or may not all be in the outer membrane (Grimwood et al., 1999). At least six putative Pmp(s) were identified in *C. psittaci* on the basis of their size (90-110 kDa) (Longbottom et al., 1996). Four of these Pmp(s) are very similar to each other. They are located in pairs at the two different loci on the chromosome. It is likely that these Pmps resulted from duplication. The first locus encodes the *omp91A* and *omp90A* gene, which are absolutely identical. Another locus encodes *omp91B* and *omp90B*, which share 89% similarity and are 86% similar to Omp90. The 90 kDa proteins are major immunogens recognized by post-abortion antisera of sheep infected with *C. psittaci* (Lombottom et al., 1998). Similarly, in *C. pneumonias*, a homologous 98 kDa protein was shown to be an immunogen in a natural infection (Campbell et al., 1990). Immunelectron microscopy and immunogold studies of *C. psittaci* and *C. pneumonias* showed that at least one member of the Pmp family was exposed on the surface of both EBs and RBs (Longbottom et al., 1998; Knudsen et al., 1998).
Genomic analysis of complete *C. trachomatis* and *C. pneumoniae* genomes has also revealed the existence of a complex gene family named the polymorphic membrane protein (Pmp) family, that consisted of predicted outer membrane proteins (Kalman *et al.*, 1999; Grimwood *et al.*, 1998; Stephens and Lammel, 2001). The Pmp family shares similarity to the 90-kDa gene family (POMP) of *C. psittaci*. Surprisingly, it is a large family of 9 related *pmp* genes in *C. trachomatis*, identified and orderly named PmpA to PmpI. A larger homologous family present in *C. pneumoniae* consists of 21 related *pmp* genes (Pmp1 to Pmp21) (Kalman *et al.*, 1999). Unlike *C. psittaci* Pmps, which have high sequence similarity (Tanzer *et al.*, 2001), *C. trachomatis* and *C. pneumoniae* Pmps are remarkably polymorphic in amino acid sequence, molecular weight and predicted pI. These differences lead to difficulty in determining the accurate amino acid identities between proteins. Most proteins show identities below 25%, which is not significant for implying structural and functional similarity between proteins. Phylogenetic analysis of the Pmp family of *C. pneumoniae* and *C. trachomatis*, however, revealed six related groups containing at least one Pmp from each species (Grimwood *et al.*, 1999). The six related families also suggested the possibility of multiple specific roles for the Pmp family.

All Pmps contain the common signature tetrapeptide repeat, glycine-glycine-alanine-isoleucine/leucine/valine (GGAI/L/V). This motif is repeated numerous times (2-12 times) throughout the NH3-terminal half of each protein. Other repeat motif FXXN also occurs multiple times (4-23 times) within the NH3-
terminal half of all Pmps. The C-terminal half of all Pmps, except one truncated protein, CPn452, are also characterized by conserved tryptophans and a carboxyl terminal phenylalanine, a common feature of the outer membrane proteins of gram-negative bacteria (Struyve et al., 1991). Like many integral outer membrane proteins in prokaryotes, most Pmps possess predicted signal peptide leaders for crossing the inner membrane, which suggests a localization of the expressed proteins in the chlamydial outer membrane (Stephens et al., 1991). Christiansen et al. (1999) demonstrated that at least three of C. pneumoniae Pmps are present on the surface of EBs and differentially expressed.

Comparative genomic analysis of pmps of two strains, CWL029 and J138, revealed a nucleotide sequence identity of 89.6-100% and deduced amino acid sequence identity of 71.1-100%. (Shirai et al., 2000). In addition, some pmps may not be expressed as functional proteins, as five contain predicted frameshifts, and one contains a nonsense mutation leading to stop codon. Mass spectrometry analysis shows that major constituents of C. trachomatis L2 outer membrane complex consist of PmpE, PmpG, and PmpH (Mygind et al., 2000; Tanzer et al., 2001). All of the C. trachomatis and C. pneumoniae pmp however, are transcribed during infection but differentially expressed at the surface of the EBs (Grimwood et al., 2001). Expression of Pmps, in both C. trachomatis and C. psittaci, are identified late (72 hpi) in the developmental cycle (Grimwood et al., 2001). By two-dimension gel electrophoresis and mass spectrometry analysis, Pmp2, 6, 7, 8,
10, 13, 14, 20 and 21 are found in EB, but there is no evidence of expression of truncated genes responsible for Pmp5 and Pmp12.

Recently, Pmps were predicted to be transported over the inner membrane by the autotransport system based on the conserved amino acid stretch, GG[A/L/V/I][I/L/V/Y] and FXXN repeat. Henderson et al. (2001) suggested that a carboxyl terminal of the Pmp forms a predictive C-terminal β-barrel traslocation unit, which functions an anchor of the type IV autotransporters. The repeat motifs in Pmp(s) resemble several autotransporter passenger motifs including rOmpA and rOmpB of the Rickettsia. Collectively, homology search and structural motif approaches suggested that the Pmp are predicted autotransporters.

**Lipopolysaccharide (LPS)**

The outer membrane of the chlamydial cell wall consists of a truncated lipopolysaccharide, which harbors a group-specific epitope. Unlike MOMP, LPS is conserved among diverse Chlamydiae strains (Dhir et al., 1971). According to the taxonomic reclassification of the order Chlamydiales, the formally described genus-specific antigen of chlamydial LPS is now named a family-specific antigen (Everett et al., 1999). Loosely associated with both EB and RB cell surface, chlamydial LPS, is an immunodominant antigen that elicits humoral immune response in all types of chlamydial infection. The group specific epitope of LPS, therefore, has been useful in diagnosis of chlamydial infection.
The surface exposure of LPS on RBs and EBs was characterized using immunogold staining with two monoclonal antibodies. These findings also suggested that the chemical structure of LPS may differ during the developmental cycle (Kuo et al., 1987). Unlike other gram-negative bacteria, chlamydia harbors a minimal LPS form containing lipid A and two Kdo residues that chemically resemble the rough (Re type) LPS of *Salmonella enterica* serovar Minnesota Re 595. The predominant structure of chlamydial LPS is a trisaccharide of 3-deoxy-D-manno-oct-2ulosonic (Kdo) residues of the sequence alpha-Kdo-(2→8)-alpha-Kdo-(2→4)-alpha-Kdo (Kosma et al., 1999; Rund et al., 2001). The chlamydiae LPS exhibits low endotoxicity relative to LPS from other species. Antibodies against LPS do not prevent infection of host cells. The low endotoxicity of chlamydial LPS is related to the unique structural features of the lipid A, which is highly hydrophobic due to the presence of unusual long fatty acids.

**Inclusion membrane proteins (Incs)**

*Chlamydiae* are obligate intracellular bacteria that have a unique developmental cycle. Unlike other intracellular bacteria that reside in the host cytoplasm, once endocytosed by host epithelial cells or non-professional phagocytes, *Chlamydiae* replicate and sequester themselves within a non-acidified membrane bound vacuole, termed an inclusion, throughout their intracellular developmental cycle. The inclusion rapidly dissociates from endosomes or lysosomes and establishes interactions with exocytic vesicles by
fusion with trans-Golgi network (TGN)-derived sphingolipid at an early stage of development. Both fusion of exocytic vesicles and avoidance of lysosomal fusion require early chlamydial protein synthesis (Scidmore et al., 1996). Both chlamydial proteins and host factors that govern chlamydial inclusion trafficking, however, remain to be characterized.

The trafficking of Chlamydiae in the host cells is directly shown by the absence or the presence of specific markers of endosomes and lysosomes on the inclusion membrane. It is clearly demonstrated that neither early or late endosome markers such as transferrin receptor and the cation-independent mannose 6-phosphate receptor, or late endosomal/lysosomal markers including the lysosomal glycoprotein LAMP1, LAMP 2, cathepsin D and the vacuolar H-ATPase, are present on the inclusion membrane (Heinzen et al., 1996; Scidmore et al., 1996; Hackstadt et al., 1995; van Ooij et al., 1997) In addition, fluid-phase markers including fluorescein isothiocyanate-conjugated dextran and lucifer yellow are not trafficked to the chlamydial inclusion (Heinzen et al., 1996).

The chlamydial inclusion has been described as an aberrant Golgi-derived vesicle. Although it does not display any markers of the endocytic pathway, it is labeled with a vital fluorescent dye analog of ceramide, 6-[N-(7-nitrobenzo-2-oxa-1, 3-diazol-4yl) aminocaproyl sphingosine] (C₆-NBD ceramide), which has been used extensively to study sphingolipid trafficking in viable cells. Incorporation of this vital dye into the inclusion was suggested as the result of its fusion with a subset of exocytic vesicles (Hackstadt et al., 1995).
C6-NBD ceramide, like endogenous ceramide, is processed to sphingomyelin or glucosylceramide within the Golgi apparatus at the cis—or medial Golgi compartment before transport to the plasma membrane via a vesicle-mediated process (Lipsky and Pagano, 1985). It becomes clear that the acquired sphingomyelin from C6–NBD ceramide is incorporated and can be detected in the chlamydial inclusion within 2 hpi (Hackstadt et al., 1996). Approximately, 50% of the sphingomyelin endogenously synthesized from the added fluorescent ceramide analog is diverted to the chlamydial inclusions.

The interaction of the inclusions with sphingomyelin–containing exocytic vesicles was demonstrated as a common character of all Chlamydia (Hackstadt et al., 1996; Rockey et al., 1996; Wolf et al., 2001). The transiently direct trafficking of fluorescent sphingomyelin from the Golgi apparatus to the inclusion is an active process that is energy-and temperature-dependent and is sensitive to inhibitors of bacterial RNA and protein synthesis (Hackstadt et al., 1996), and sensitive to BrefeldinA (Wylie et al., 1997). This trafficking leading to the acquisition of sphingomyelin is initiated before differentiation of EBs to RBs. Unlike the Toxoplasma gondii vacuole, vesicle interaction with the exocytic pathway and the chlamydial inclusions is not dependent upon route of entry of EBs but determined by the chlamydial early gene expression (Scidmore et al., 1996). The interaction of inclusions with the exocytic pathway has been proposed as a potential mechanism enabling chlamydiae to escape from lysosomal fusion (Hackstadt et al., 1997).
The origin, development and function of the chlamydial inclusion remain to be characterized. Only a set of chlamydial proteins termed inclusion membrane protein (Inc) has been identified on the inclusion membrane of all *Chlamydiae* except *C. pecorum*. The development of the inclusion is presumably accomplished by the integration of multiple inclusion membrane proteins into the host vesicle membrane. During the study of differences in serological response between infected animals and those immunized with formalin-treated EBs, the three Inc proteins, IncA (Rockey and Rosquist, 1994; Rockey *et al.*, 1995; Bannantine *et al.*, 1998a), IncB and IncC (Bannantine *et al.*, 1998b), were initially identified and characterized in *C. psittaci*. An alternative approach using antisera against total purified membrane fractions of *C. trachomatis* infected cells and directly staining the inclusion membrane has accomplished identification of other Inc proteins. Immunofluorescent staining revealed a new set of four chlamydial inclusion membrane proteins, Inc D-G (CT116-119) localized to the *C. trachomatis* inclusions (Scidmore *et al.*, 1998). The operon containing *incD-G* was located downstream of *C. trachomatis incA*. Genome data analysis revealed that IncD-G are *C. trachomatis* specific Inc proteins and are absent in *C. psittaci* and *C. pneumoniae* genomes. Genes encoding for IncD-G are expressed as an operon with transcription detectable in the first 2 h following internalization while *incA* is expressed in the 18 hpi. The early expression of the *incD-G* operon suggests that they are required for the modification of the nascent inclusion.
Genomic analysis of the complete *C. trachomatis* serovar D and *C. pneumoniae* CWL029 genome led to the identification of homologous IncA, IncB and IncC. IncA of *C. trachomatis* was demonstrated to localize in the inclusion membrane (Bannantine *et al.*, 1998). To date, there are 10 identified Inc proteins: IncA, B, C of *C. psittaci*, IncA-G of *C. trachomatis* and *C. pneumoniae* IncA. Surprisingly, these Inc proteins do not share significant sequence similarity (less than 20 %), nor are they similar to any genes within the sequence database. However, each identified Inc protein has the unusually large bi-lobed hydrophobic motif of 60-80 amino acid residues (Figure 3). It was demonstrated the C-terminal hydrophobic domains of IncA, IncG and IncF of *C. trachomatis* are exposed to the cytoplasm (Hackstadt *et al.*, 1999). IncA of *C. psittaci*, is also exposed on the cytoplasmic side of vacuoles and phosphorylated by host cell kinase on the serine and threonine residues (Rockey *et al.*, 1997). Therefore, Inc proteins may play a critical role as mediators of host-parasite interaction via signal transduction pathways in host cells.

The function of the Inc proteins in the chlamydial developmental cycle and host–pathogen interaction remain to be elucidated. Recently, the first eukaryotic protein that interacts with chlamydial inclusion membranes was identified. The specific association of *C. trachomatis* inclusion membrane and mammalian 14-3-3β, a phosphoserine binding protein, via its interaction with IncG has been revealed using a yeast two-hybrid system and immunofluorescent microscopy (Scidmore *et al.*, 2001). The specific localization of 14-3-3β to the inclusion membrane was identified only in
Figure 3. Hydropathy profiles for the secondary structure of known inclusion membrane proteins. Profiles were determined using the algorithm developed by Kyte and Doolittle with a window size of seven amino acids. The vertical axis displays relative hydrophilicity with negative scores indicating relative hydrophobicity. The horizontal axis indicates amino acid number. Note that the size of each graph is similar while the length of each protein is distinct. Hydrophobic region is similar size and shapes are circle in each plot. Each of these proteins has been shown to localize to the chlamydial inclusion membrane using fluorescence microscopy with specific antibody.
C. trachomatis. This is consistent with the absence of an IncG homologue, as
determined by genome data analysis, in either C. psittaci or C. pneumoniae. However,
the role of 14-3-3β in the C. trachomatis infections remains unclear. The species-
specific interaction of 14-3-3β and the inclusion membrane may mediate the signal
transduction and/or other ligand proteins in infected cells. Several studies of
chlamydial development have shown the similarities and differences of the
development cycles of various Chlamydia spp. Each species and strain possesses
distinct characters associated with inclusion development. The significant differences
include inclusion morphology, content of the inclusions, kinetics of the inclusion
development in different eukaryotic tissue cells and interaction of cellular organelles.
In C. trachomatis, C. pneumoniae and some strains of C. psittaci and C. pecorum,
multiple lobed inclusions formed initially are eventually fused to generate a large
inclusion even when the cells are multiply infected (Campbell et al., 1989; Matsumoto
et al., 1991; Patton et al., 1990; Ridderhoff et al., 1989; Rockey et al., 1996). Unlike
C. trachomatis inclusions, inclusions of many strains of C. psittaci including strain
GPIC, do not fuse but appear to divide within infected cells.

Little is known about the function, regulation, and mechanism of the unique
homotypic fusion property of C. trachomatis. Ridderhof and Barnes (1989)
demonstrated that inclusions harboring different serovars of C. trachomatis can fuse
and eventually form a single large inclusion. Treatment of infected cells with
cytochalsin D, a microfilament–disrupting drug, led to the delay of perinuclear
transport and inclusion fusion of C. trachomatis serovar L2 but not serovar E
(Schramm et al., 1995). A reduced effect of cytochalasin D on the fusion of inclusions was also identified in HeLa cells co-infected with C. trachomatis serovar E and F (Ridderhof and Barnes, 1989). Van Ooij et al. (1998) showed that fusion of inclusions is also inhibited at temperature 32°C in several serovars of C. trachomatis. The multiple inclusions formed at 32°C are derived from a high multiplicity of infection as opposed to from the dividing of an inclusion. The inclusion fusion is temperature dependent and required bacterial protein synthesis.

Recently, Hackstadt et al. (1999) demonstrated that the homotypic vesicle fusion of C. trachomatis was inhibited by microinjection of antisera against IncA whereas the fusion of inclusion membranes with sphingomyelin containing vesicles was not disrupted. In addition, antibodies against other Inc proteins such as anti-IncF, IncG were unable to inhibit the homotypic fusion of C. trachomatis inclusions. The detection of the inclusion fusion of C. trachomatis at about 10-12 hpi is correlated to the temporal expression of IncA. In addition, the yeast two-hybrid analysis showed that carboxy-termini of IncA form homodimers interacting with the other IncA. Collectively, these data showed that IncA plays a role in homotypic fusion of C. trachomatis inclusions.

The parasitophorous vacuole is likely a barrier that separates chlamydiae from not only the hostile environment but also nutrient-rich environment of the host cytoplasm. Chlamydiae obtain nutrients such as amino acids and nucleotides, from host pools (Iliffe-Lee et al., 1999). Although the interaction of the inclusion membrane and trans-Golgi-derived vesicles supplies the sphingomyelin for
chlamydiae, the delivery mechanism of other essential precursors remains unclear. It was demonstrated by transfection and microinjection studies of fluorescent tracer that the inclusion membranes do not contain channels or pores, which would allow passive diffusion of nutrients and biosynthetic precursors as small as 520 Da (Heizen et al., 1997). Neither specific transporter nor non-specific diffusion channels, however, have been identified in the chlamydial inclusion membranes. It was suggested that chlamydia may possess a novel mechanism to deliver the nutrient precursors from the nutrient-rich environment of host cytoplasm across the inclusion membrane.

All known chlamydial inclusion membrane proteins have no apparent classical signal sequences at the their amino termini (Rockey et al., 1995, 1997; Scidmore et al., 1998). Studies by Masumoto et al. (1982) using electron microscopy demonstrated that chlamydiae possess the surface projection on the cell surface that is similar to the type III secretion machinery in gram-negative bacteria. The genome data analysis of *C. trachomatis* and *C. pneumoniae* has revealed that chlamydiae possess genes that may encode a putative type III secretion apparatus (Bavoil et al., 1998). This secretion system has been described in several bacterial pathogens including *Yersinia, Salmonella, Shigella*, and *Pseudomonas*. Unlike type III secretion system of other pathogens, genes encoding the type III secretion apparatus in chlamydiae genome are dispersed. The expression of the putative type III apparatus of *C. trachomatis* was demonstrated using RT-PCR. With the absence of the classical signal sequence in the peptide sequence and the presence of genes coding for type III apparatus, Inc proteins are logically considered as a candidate proteins secreted by
type III apparatus. According to BLAST search, CopN which shares sequence similarity to the type III apparatus, is a chlamydial protein observed to be secreted by chlamydiae to the inclusion membrane (Fields et al., 2000). Additionally, Subtil et al. (2001) demonstrated that the IncA, IncB and IncC of both C. trachomatis and C. pneumoniae were secreted by the heterologous type III machinery of Shigella flexneri. These findings provided evidences suggesting that chlamydiae encode and use a type III secretion system.

Genetic variation in chlamydial cell surface proteins

Antigenic polymorphism within C. trachomatis by allelic variation of the single copy of ompA encoding MOMP was identified and considered the genetic basis of antigenic change within C. trachomatis (Barnes et al., 1987; Brunham and Peeling, 1994; Hayes et al., 1994; Lampe et al. 1993; Stephen et al., 1987; Yang et al., 1993). Although the mechanism of genetic variation for ompA remains unclear, comparative analysis of its sequences revealed that mutation and recombination events between serovars may play a role in antigenic variation (Brunham and Peeling, 1994; Hayes et al., 1994) Alteration of MOMP epitopes is considered to play a critical role in the immune evasion in the host cells. Identification of new variants of C. trachomatis isolates revealed a mosaic structure of ompA containing the nucleotide sequence with different ancestries. These variants include Ia (I/H), LGV (L1/L2). The polymorphism of ompA sequence suggested that the genetic variation may result from the inter- or intra-chromosomal recombination within or
between different serovars. Recently, the intraspecies and interspecies recombinations have been identified in *C. trachomatis ompA*. Comprehensive statistical analysis of recombination of 40 *ompA* and *omcB* sequences from *C. trachomatis, C. psittaci, and C. pneumoniae* support intragenic recombination of *C. trachomatis ompA* (Millman et al., 2001). The higher degree of recombination was in the downstream half of the gene. Investigations of mosaic sequences of *ompA* of known recombinant strains including D/B 120, G/UW-57, E/Bour and LGV-98 revealed significant breakpoints for recombination just before and within VD3, a region responsible for elicited T-helper cell responses. The high recombination at the upstream region and within VD3, may be associated with the molecular adaptation. Analyses of intragenic recombination demonstrated that the rate was higher for C class than for B class isolates. Analyses also revealed interspecies recombination of *ompA* between the rodent *C. trachomatis* mouse pneumonitis/ NiggII strain and the *C. pneumoniae* horse N16 strain. Unlike the *C. trachomatis*, *C. pneumoniae* isolates cannot be classified into different serovars because MOMP is highly conserved. However, the strain diversity of *C. pneumoniae* remains unknown.

Among the most interesting findings from the genome was the existence of a large group of polymorphic outer membrane proteins (Pmps). Antigenic variation of the Pmp family has been investigated. Although the biological role of Pmps remains to be described, the unknown role of the Pmp family is of specific interest focusing on their surface exposure and immunogenicity in natural infection.
Hatch et al. (1999) described the variation of Pmp expression in *C. trachomatis* L2 and in *C. psittaci* 6BC. Recently, Christiansen et al. (1999) have also proposed and revealed the differential expression *in vivo* of Pmps in mice infected with *C. pneumoniae*. Comparative analysis of the outer membrane protein genes *ompA* and *pmp* of *C. pneumoniae* CWLO29 and J138 strain revealed the strain diversity of *pmp* genes (Shirai et al., 2000).

Furthermore, Grimwood et al. (2001) revealed the evidence for interstrain variation of some Pmps of *C. pneumoniae* CWLO29, TW-183 and AR39. Interstrain variation of *pmp* gene expression was detected in *pmpG10* (CPn0449/0450), which contains a homopolymeric guanine tracts (poly G) and a predicted frameshift mutation. The variation of string of G residues in a *pmpG10* nucleotide sequence can lead to a change in reading frame. Based upon the genome sequence for strain CWLO29, *pmpG10* contain a string of 13 guanine residues and is out of reading frame. The PmpG10 expression was detected only in AR39 (14 G) and TW- 183 (11 G) but not CWLO29 (13 G). Interestingly, CWLO29 with a difference passage history containing 14 guanine residues, which are in frame, produces a protein product (Knudsen et al., 1999). These findings revealed the interstrain and intrastrain variation in *C. pneumoniae* PmpG10 expression, which are apparently modulated by slip-stranded mispairing mechanism. *C. trachomatis* *pmpG* (CT871) also encoded the string of G which suggested the phase variation is a common trait in PmpG. The differential expression of *pmpG10* resulting from gene switching has been identified and explained by the variation of the length of poly
guanine (poly G) tracts (Pedersen et al., 2000). In addition, Grimwood et al. (2000) revealed the variation in size of PmpG6 in comparative studies of CWL029 and AR39. Genomic analysis showed that PmpG6 in CWL029 contains the three internal tandem repeats of 131 amino acids whereas of AR39 contains only 2 repeats. It is likely that differentiate expression and gene activation or inactivation of Pmps will contribute to the surface variability and altered phenotype. These findings also suggested the interstrain variation of Pmp family was modulated by a novel molecular mechanism. The function of Pmps, however, in chlamydiae growth and development remains uncharacterized.

**Genetic and phenotypic variation in other bacterial pathogens**

Mutation is considered as a mechanism of gene variation and adaptation, often exploited in response to changes in the environment and to promote evolutionary flexibilities. The mutation rate of each gene in the genome is likely variable and it is found to be higher in some particular genes. Obviously, unlike the housekeeping genes with universally conserved sequences, genes involved in the interaction with the environment frequently have a high mutation rate in unpredictable fashion. Such genes are referred to as contingency or mutable genes (Moxon et al., 1994, 1998). Mostly, these hypermutable genes are involved in production or modification of surface structures of microbial pathogens - lipopolysaccharides, lipoproteins, capsular polysaccharides, flagella, pili, and
excreted enzymes, and the modification system enzymes (van Belkum et al., 1998, 1999; Deitsch et al., 1997; Moxon et al., 1998; 2000)

In microbial pathogens, the variations of surface structures are roughly classified into two types: phase variation and antigenic variation (Robertson and Mayer, 1992; Henderson et al., 1999). Phase variation is the reversible change in the presence or absence of a particular surface structure. In contrast, antigenic variation is the changes leading to different variants of a surface structure. Gene conversion and recombination, which lead to gene cassette switching, intragenomic and intergenomic gene transfer, are the major mechanisms of antigenic variation. The variation of the surface structure can be mediated by either homologous or non-homologous recombination. Both phase and antigenic variations are reversible genetic mechanisms facilitating microbial evasion of host immune response and microenvironmental adaptation. Genetic variation of surface structures plays a critical role in microbial pathogenesis - adherence, colonization, invasion, virulence, and persistent infection.

Phase variation involves the regulation of gene expression of surface structures under different circumstances. The variation results from the reversible mutation in the simple short DNA repeat sequence (van Belkum et al., 1998). The short sequence repeat (SSR) or repetitive DNA, is a stretch of homopolymeric or heteropolymeric DNA sequence. SSR can be found in either a noncoding region including promoter or a coding or internal DNA sequence of the genome. Therefore, the phase variable genes have been identified by the presence and the
length of DNA repeats interpreted on their sequence context. The variation of the 
short DNA repeat is considered to be a result of a recA-independent slipped-
stranded mispairing mechanism (Levinson and Gutman, 1987). Variation of the 
short DNA sequence repeat located within the open reading frame, which can alter 
the translational reading frame thereby determining whether or not the protein is 
translated or within the promotor regions where they affect the strength of 
transcription.

Several microbial pathogens including *Neisseria gonorrhoeae*, *Neisseria 
meningitides*, *Haemophilus influenzae*, *Helicobacter pylori*, etc. possess 
mechanisms to ensure variability of their surface structures in particular 
environments (van Belkun *et al.*, 1997, 1999; Hood *et al.*, 1996; Moxon *et al.*, 
special status and distinguished from more conventional mutation events. (Dybvig 
MATERIALS AND METHODS

Wild type and mutant C. trachomatis isolates

Different serovar (B, D, D', E, F, G, H, I, Ia, J and K) of 13 wild type and 26 mutant C. trachomatis isolates collected over the course of 12 years were used in this study (Table 1). These isolates were obtained from Dr. Walter E. Stamm, Division of Allergy and Infectious Disease, School of Medicine, University of Washington, Seattle. The mutant isolates with nonfusogenic inclusions of C. trachomatis infected cells were identified from genital swabs of clinical patients by observation of inclusion morphology during routine serotyping of C. trachomatis at the Seattle-King County Department of Public Health Clinic between 1988 and 1999 (Suchland et al., 2000). Wildtype chlamydial reference strain of serovar D (D/UW-3/cx), serovar J (J/UW-36/cx) and serovar G (G/UW-57) were described by Wang et al., 1985. The serotype of each C. trachomatis isolate was determined using antibodies against MOMP (major outer membrane protein) using the low passage microtiter plate format (Suchland et al., 1991). The serovar and a laboratory reference number were indicated for each clinical isolate. The mutant isolates, possessing nonfusogenic inclusion phenotypes, were represented with a "(s)" designation.
Table 1. Description of wild type and non-fusogenic variants of *C. trachomatis* isolates used in this study

<table>
<thead>
<tr>
<th>Isolate</th>
<th>serovar</th>
<th>year isolated</th>
<th>Source of isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild types</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT9227</td>
<td>B</td>
<td>2000</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9291</td>
<td>D'</td>
<td>2000</td>
<td>urethral infection</td>
</tr>
<tr>
<td>MT9332</td>
<td>E</td>
<td>2000</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9334</td>
<td>F</td>
<td>2000</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9301</td>
<td>G</td>
<td>2000</td>
<td>urethral infection</td>
</tr>
<tr>
<td>MT9309</td>
<td>Ia</td>
<td>2000</td>
<td>urethral infection</td>
</tr>
<tr>
<td>MT2364</td>
<td>J</td>
<td>1994</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT3464</td>
<td>J</td>
<td>1994</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9295</td>
<td>J</td>
<td>2000</td>
<td>urethral infection</td>
</tr>
<tr>
<td>MT9325</td>
<td>J</td>
<td>2000</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9329</td>
<td>J</td>
<td>2000</td>
<td>urethral infection</td>
</tr>
<tr>
<td>MT9336</td>
<td>J</td>
<td>2000</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9346</td>
<td>J</td>
<td>2000</td>
<td>rectal infection</td>
</tr>
<tr>
<td>MT9311</td>
<td>K</td>
<td>2000</td>
<td>urethral infection</td>
</tr>
<tr>
<td>Non-fusogenic variants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT1129</td>
<td>B(s)</td>
<td>1987</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT1012</td>
<td>D(s)</td>
<td>1993</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT2923</td>
<td>D(s)</td>
<td>1994</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT4013</td>
<td>D(s)</td>
<td>1998</td>
<td>rectal infection</td>
</tr>
<tr>
<td>MT5058</td>
<td>D(s)</td>
<td>1996</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT3991</td>
<td>D(s)</td>
<td>1995</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT5328</td>
<td>D'(s)</td>
<td>1996</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT8323</td>
<td>D'(s)</td>
<td>1999</td>
<td>rectal infection</td>
</tr>
<tr>
<td>MT1986</td>
<td>E(s)</td>
<td>1986</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT4022</td>
<td>F(s)</td>
<td>1995</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT8039</td>
<td>F(s)</td>
<td>1995</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT459</td>
<td>G(s)</td>
<td>1988</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT5642</td>
<td>H(s)</td>
<td>1996</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT9005</td>
<td>H(s)</td>
<td>1999</td>
<td>urethral infection</td>
</tr>
<tr>
<td>MT2197</td>
<td>Ia(s)</td>
<td>1994</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT3839</td>
<td>J(s)</td>
<td>1995</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT6565</td>
<td>J(s)</td>
<td>1997</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT7713</td>
<td>Ia(s)</td>
<td>1998</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT893</td>
<td>J(s)</td>
<td>1988</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT1980</td>
<td>J(s)</td>
<td>1994</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT5942</td>
<td>J(s)</td>
<td>1996</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT6276</td>
<td>J(s)</td>
<td>1997</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT6462</td>
<td>J(s)</td>
<td>1998</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT6686</td>
<td>J(s)</td>
<td>1997</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT8068</td>
<td>J(s)</td>
<td>1994</td>
<td>cervical infection</td>
</tr>
<tr>
<td>MT2481</td>
<td>K(s)</td>
<td>1992</td>
<td>cervical infection</td>
</tr>
</tbody>
</table>
C. pneumoniae isolates, plasmids and genomes

Purified genomic DNA used in this study (Table 2) was obtained from Dr. Lee Ann Campbell, Department of Pathobiology, University of Washington, Seattle. Chlamydia pneumoniae was grown in HEp 2 cells for 72 h. Elementary bodies were harvested, and purified using Renografin gradient as described below (Caldwell et al., 1981). Genomic DNA of C. pneumoniae were isolated from purified EBs using QiaAmp DNA mini kits (Qiagen Inc., Valencia, CA) according to the manufacture’s instructions. DTT (Dithiothretol) was added to lyse the chlamydial membrane proteins. Extracted DNA was eluted in 100 μl of Qiagen elution buffer, and stored at −20°C.

Genomic data of C. pneumoniae CWLO29 (http://chlamydia-www.berkeley.edu-4231/, GenBank accession number AE001363 and NC_000922), AR39 (http://www.tigr.org/, GenBank accession number AE002161), and J138 (http://w3.grt.kyushu-u.ac.jp/J138/ GenBank accession number AB036071-AB036089 and NC_002491) were analyzed. Open reading frames (ORFs) and annotated were designated based upon the C. pneumoniae CWLO29 genome. The TA and Zero Blunt™ PCR cloning kits (Invitrogen, San Diego, CA) were used to attain clones for genetic variation study.
Table 2. Description of 12 independent *C. pneumoniae* isolates used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>City/Year Isolated</th>
<th>Origin/clinical presentation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR32</td>
<td>Seattle, US/1983</td>
<td>throat swab/pharyngitis</td>
<td></td>
</tr>
<tr>
<td>TW183</td>
<td>Taiwan/1965</td>
<td>conjunctiva</td>
<td></td>
</tr>
<tr>
<td>AR458</td>
<td>1986</td>
<td>acute respiratory infection</td>
<td></td>
</tr>
<tr>
<td>AR388</td>
<td>1985</td>
<td>acute respiratory infection</td>
<td></td>
</tr>
<tr>
<td>AR231</td>
<td>1984</td>
<td>acute respiratory infection</td>
<td></td>
</tr>
<tr>
<td>AR277</td>
<td>1984</td>
<td>pneumonia</td>
<td></td>
</tr>
<tr>
<td>PS32</td>
<td>1995</td>
<td>carotid endarterectomy</td>
<td></td>
</tr>
<tr>
<td>AC43</td>
<td>Japan/1989</td>
<td>nasopharynx/pneumonia</td>
<td>Yamazaki <em>et al.</em>, 1990</td>
</tr>
<tr>
<td>KA5C</td>
<td>Finland/1987</td>
<td>pneumonia</td>
<td></td>
</tr>
<tr>
<td>KA66</td>
<td>Finland/1987</td>
<td>pneumonia</td>
<td>Ekman <em>et al.</em>, 1993</td>
</tr>
</tbody>
</table>

Note: All of AR isolates were isolated during a period of endemic acute respiratory disease in Seattle during 1983-86. KA isolates were isolated from military trainees *isolated from the Providence Medical Center*
Chlamydial cell cultures

Semiconfluent monolayer cultures of HeLa 229 epithelial cells, (CCL 2.1; American Type Culture Collection) and McCoy murine fibroblast cell line (CRL-1696; American Type Culture Collection) were routinely grown in Minimal Essential Medium (MEM, Gilbo BRL) supplemented with 10% fetal bovine serum (FBS) and 10 μg/ml gentamicin (MEM-10; Whittaker Bioproducts) at an atmosphere of 5% CO2 in humidified air at 37°C before infection. For immunofluorescence microscopy, cell cultures were grown on 12-mm-diameter glass coverslips (no 1) in 24 well plates.

Purification of elementary bodies

Purification of EBs from selected C. trachomatis serovars and C. pneumoniae isolates were conducted from infected HeLa cell monolayers and HEP2 cell lines with serial Renografin gradients (Squibb Diagnostics, New Brunswick, N.J.) as previously described (Caldwell et al., 1981). Infected HeLa cells with 90% of cells containing inclusions were harvested at 48 hpi for C. trachomatis and 72 hpi for C. pneumoniae. Chlamydiae were removed from the monolayer using 4-mm glass beads and cold Hanks balanced salt solution (HBSS, Gibco). The cell suspensions were pooled and ruptured using sonication (Braunsonic model 1510). The cell lysates were centrifuged at 500xg for 15 min at 4°C.
The supernatants were then layered over 8 ml of a 35%(vol/vol) Renografin solution (diatrizote meglumine and diatrizate sodium) in 0.01M HEPES (N-2-hydroxyethylpiperazine-N'2 ethanesulfonic acid) and 0.15M NaCl. The Renografin was then centrifuged at 43,000 x g for 1 h at 4\(^0\)C. The pellets were suspended in SPG buffer (0.25 M Sucrose, 5mM L-glutamic acid, 10 mM sodium phosphate, pH 7.2), and layered over discontinuous Renografin gradients (13ml of 40%, 8 ml of 44%, and 5 ml of 52% Renografin, vol/vol). The gradients were centrifuged at 43,000X g for 1 h at 4\(^0\)C. The purified EB fraction located at the 44/52% Renografin interface was collected, diluted with 3 volumes of SPG and then centrifuged at 30,000xg for 30 min. The EB pellets were washed in SPG to remove the residual Renografin. Chlamydiae were stored at \(-80^0\)C in SPG. Prior to inoculation, Chlamydiae were diluted in SPG and kept on ice.

**Identification of putative Inc proteins using hydrophilicity analysis**

Sequence analyses were performed from all ORFs identified from the C. trachomatis and C. pneumoniae genome database. Hydropathy plots from Mac Vector\(^\text{TM}\) 6.0 (Oxford Molecular) using KYTE-DOLITTLE algorithms with a window of seven were conducted to determine the hydrophobic domain in the amino acid sequences (Kyte and Doolittle, 1982). Deduced amino acid sequence of Chlamydiae genes in which secondary structures contain bi-lobed hydrophobic domain were identified as putative Inc proteins. The sequence similarity of putative Inc proteins was determined using the BLAST (Basic Local Alignment Search
Tool) search and BLAST 2 sequences program from http://www.ncbi.nlm.nih.gov/. All putative Inc proteins of *C. trachomatis* and *C. pneumoniae* were subjected to gap BLAST analysis to identify similarities with known proteins in the database.

**Production of putative Inc proteins**

Six putative *inc* genes of *C. trachomatis*: CT119, CT223, CT229, CT233, CT288, CT442, CT484, and CPn0186, a candidate *incA* gene of *C. pneumoniae*, were selected to identify the protein localization. We selected two genes, CP0186 and CT233 which gene products are similar to known Inc proteins. The gene product of CPn0186 has 20-23% similarity to IncA of *C. trachomatis* and *C. psittaci* while the gene product of CT233 has 44% similarity to IncC of *C. psittaci*. The other four putative *inc* genes were randomly selected with an attempt to provide a cross-section of putative Inc. ORF CT223 and CT229 were selected to be investigated because both are encoded in a cluster of seven putative *inc* genes (Figure 4). Gene product of CT228 consists of two independent hydrophobic domains whose structural feature of two tandem IncA-like protein joined tail to head. ORF CT484 encodes a gene product that is significantly similar to CPn0602 and also structurally similar to IncA protein but the predicted protein possesses a more hydrophobic profile over the entire sequence. Six putative *inc* genes of *C. trachomatis* and one of *C. pneumoniae*, were amplified using specific oligonucleotide primers (Table 3). In some case, the 5' end of the gene was deleted due to the full-length constructs of genes apparently toxic to *E. coli*. 
Figure 4. Schematic representation of the contig containing the selected *inc* genes in *C. trachomatis* genome. (A) ORF maps of two 10 kb segments of the *C. trachomatis* genome labeled contig 2.3 and contig 3.5. ORF encoding putative Inc(s) including CT115 to CT118 from contig 2.3 and CT223 to CT229 from contig 3.5. Note the close proximity of IncA, B, and C to each of the respectively putative Inc clusters. (B) Hydrophathy plots of *C. trachomatis* putative Inc proteins used for production of specific antibodies. Note that each protein contains a bi-lobed hydrophobic domain similar to that seen for the known Inc proteins.
Table 3. Oligonucleotide primers used for preparation of putative inc/malE hybrid recombinant clones for production of fusion proteins

<table>
<thead>
<tr>
<th>Target ORF</th>
<th>Forward and reverse oligonucleotides</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. trachomatis</td>
<td>(5'--------3')</td>
<td></td>
</tr>
<tr>
<td>CT119 (incA)</td>
<td>AGCCATAGGATCTGGTTTCAGCGA GCGCGGATCCTAGGGAGCTTTTTGTAGGGTGA</td>
<td>Full length</td>
</tr>
<tr>
<td>CT223</td>
<td>ATGGTGAGTTTAGCATTAGGGA GCGCGGATCCTACACCCGAGGCGTAATTGAA</td>
<td>Full length</td>
</tr>
<tr>
<td>CT229</td>
<td>ATGAGCTGTTCATATATTTAAT GCGCGGATCCTATTATTTACGACGGGATGCCTG</td>
<td>Full length</td>
</tr>
<tr>
<td>CT233 (inc C)</td>
<td>ATGACGTACTCTATATCCGAT GCGCAAGCTTAGCTTACATATAAAGTTTG</td>
<td>Full length</td>
</tr>
<tr>
<td>CT288</td>
<td>CGAAAAACAAAAACAACTAGAAGAG GACAAGCTTAATGCACTCTCCACAATATCTTCTAA</td>
<td>316-1452</td>
</tr>
<tr>
<td>CT442</td>
<td>ATGAGCAGCTGTACCCTGTTGTT GCGCGGATCCTCATTTGGGTCTGATCCACCG</td>
<td>Full length</td>
</tr>
<tr>
<td>CT484</td>
<td>ATAGTGCCCCTCCCTACGACTCTATC GCGCGCATTCTTATTTAGCTACTACCCCGTGGAA</td>
<td>433-1014</td>
</tr>
<tr>
<td>C. pneumoniae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPn0186 (incA)</td>
<td>GCGCGGAATTCTAAGTGAGGAGAAG GCGCGGATCCTACTGACCTCCTG</td>
<td>Full length</td>
</tr>
</tbody>
</table>
Maltose binding protein (MBP) fusions of selected putative Inc proteins were constructed using pMAL-c2 (New England Biolabs). The amplification products of putative inc genes were produced using Pwo polymerase (Boehringer Mannheim) and serovar D genomic DNA as the template and digested with restriction enzyme Xmn I and BamHI for CT119, 223, 229 and CPn0186 and XmnI and HindIII for CT288, CT442 and 484. The vector was digested sequentially with those enzymes. The digested insert was ligated into the pMAL-c2 vector and transformed into competent E.coli strain DH5α. Positive recombinant clones were identified using restriction enzyme analysis, amplification reaction, and sequence analysis and then overexpressed for production and purification of fusion proteins.

**Purification of maltose fusion protein**

Maltose fusion proteins were purified as described in the protocol’s instruction (New England Biolab). Recombinant E. coli clones containing the fusion plasmid were grown in LB medium supplemented with ampicillin at 37°C. A single colony of each recombinant clone was inoculated into 10 ml LB broth containing 100 μl 2 M glucose and 20 μl ampicillin (50 mg/ml), and grown overnight at 37°C with shaking. Ten ml of the overnight culture was inoculated into one liter of LB broth containing 10 ml 2 M glucose and 2 ml ampicillin (50 mg/ml) and grown at 37°C with shaking until cell culture was 2x10⁸ cells/ml (optical density (OD)=0.4-0.6). Three ml of 0.1M isopropyl-β-D-thiogalactoside (IPTG) was added to the culture and then incubated 2 h at 37°C with good aeration. The
cell cultures were then centrifuged 20 min at 4000 x g at 4°C. The supernatant was removed and pellet was resuspended in 50 ml lysis buffer (10 mM sodium phosphate buffer pH 7.2, 0.5 M NaCl, 0.25% Tween 20, 1mM sodium azide, 10 mM 2-mercaptopethanol and 1 mM EGTA [ethylene-glycol tetraacetic acid]). The cell suspension was kept overnight at -20°C and then thawed in cold water. The cell suspension kept cold on the ice container was sonicated using short bursts to avoid heating the extract. The sonicated cells were centrifuged 14,000 x g for 20 min at 4°C. Supernatant or the crude extract containing the fusion proteins was pooled and stored at -20°C.

The fusion protein was subsequently purified using 12 g of amylose resin (New England Biolabs), which was swollen in 50 ml column buffer (10 mM sodium phosphate buffer pH 7.2, 0.5 M NaCl, 1mM sodium azide, 10 mM 2-mercaptopethanol and 1 mM EGTA) for 15 to 30 min. Amylose resin was loaded into a 2.5x10-cm column and washed with 3 column volumes of column buffer. The crude extract was diluted 1: 5 with column buffer and then loaded onto the column at a flow rate of 1 ml/min. The column was washed with 3 to 5 column volumes of column buffer and added with 10 ml of 10 mM maltose diluted in column buffer to elute the fusion protein. The elution was collected into 10 x 1ml of each fraction.
**Antibody production**

The purified fusion proteins of putative Inc CT223, 229, 288, 442, 484 CPn0186, were used as antigen for the production of monospecific antibody. Antibody production was performed in rabbit, guinea pig, or mice as previously described (Rockey et al., 1995). Monoclonal antibodies against putative Inc, CT 223p, were produced from the Monoclonal Antibody Facility of the University of Oregon. Mice immunized with CT223p were sacrificed and splenic lymphocytes collected. These cells were fused to SP/2 myeloma cells, using a modification of standard hybridoma product techniques.

**Antisera and reagents for non-fusogenic mutant studies**

Rabbit polyclonal antibodies against to *C. trachomatis* IncA was produced using described methods (Bannantine et al., 1998). Monoclonal antibody A57B9 (IgG₁κ) is against a 60 kDa heat shock protein (HSP60), a genus common determinant (Yuan et al., 1992). Monoclonal antibody 20F12 (IgG₁κ) and 12E5 (IgG₂a κ), obtained from the Monoclonal Antibody Facility of the University of Oregon, are specific for the CT223p of *C. trachomatis*. Rabbit polyclonal antibodies against IncA were determined with $^{35}$S-labeled staphylococcal protein A (3700 Bq/ml, Amersham, Piscataway, NJ). Secondary antibodies were purchased from Pierce Chemical Co. (Rockville, III).
**Immunofluorescence microscopy**

In putative Inc protein studies, indirect immunofluorescence staining was used to determine protein localization of Incs on the inclusion membranes. *Chlamydiae* infected cells were performed as described above. Semiconfluent HeLa cells in 6 or 24 well culture trays grown on 12-mm-diameter sterile glass coverslips were infected with *Chlamydia trachomatis* strain LGV 434 (serotype L2) and *C. pneumoniae* strain TW-183 diluted in HBSS (Gibco). These cells were grown in the MEM.10, supplemented with 10% fetal bovine serum and 10 μg/ml gentamicin, and incubated for 1 h at room temperature with gentle rocking. Alternatively, cells were inoculated by centrifugation at 2000 rpm (900 x g) for 1 h. A multiplicity of infection (MOI) of 0.1 and 1 were used for immunofluorescence staining.

Inocula were removed, and cultured cells were incubated with MEM-10 plus 0.7 μg/ml cycloheximide at 37°C for 30 and 40 h. *Chlamydia*-infected cells were washed twice with 1% phosphate buffer saline (PBS), and fixed with absolute methanol for 15 min or 4% para-formaldehyde in 25 mM NaPO₄ and 150 mM NaCl, pH 7.4 for 1 h. Para-formaldehyde treated cells were permeabized with 0.1% Triton X-100 (Pierce Chemical Co) and 0.05 % SDS in PBS for 3 min and then washed twice with PBS. The permeablized cells were then incubated in the blocking solution (10% Tween 20 in 50 mM NaPO4, 150 mM NaCl, pH 7.4) for 30 min and then immunostained with each specified polyclonal antibodies against to putative Inc proteins. All polyclonal antibodies were diluted 1:100 with PBS.
containing 2% (w/v) BSA. After primary antibody staining, cultures were washed 3 times in 1% PBS for 5 min and stained with secondary antibody conjugated with fluorescein or rhodamine for 1 h at room temperature. Coverslips were then mounted onto glass slide using Vectashield (Vector Laboratories) mounting medium.

In non-fusogenic mutant studies, *C. trachomatis* infected McCoy cells were cultured and fixed as described above and probed with genus-specific monoclonal antibody (mAb) CF-2 to the lipopolysaccharides (LPS) determinant of *Chlamydiae*. (Washington Research Foundation, Seattle WA). Mutant isolates were propagated through several cycles of growth in McCoy monolayers.

Double immunofluorescence staining was used to identify localization of Inc proteins, IncA or CT223p and HSP60 in the study of mutant isolates with nonfusogenic inclusions. Wildtype and mutant infected cells were probed with antibodies against HSP60 (mAb A57B9) and Inc protein CT223 (mAb 20F12 or12E5) or IncA (rabbit polyclonal antisera). Monoclonal antibodies (mAbs) against CT223p and chlamydial heat shock protein 60 (HSP 60) were used as controls for immunostaining of inclusion membrane proteins and intracellular developmental forms respectively. Nucleus DNA was labeled with 4,6-Diamidino-2-phenylindole (DAPI), which was diluted 2 μg/ml to the mounting medium.

Immunostaining was visualized using fluorescence microscopy. Fluorescence images were performed on a Leica DMLB microscope using the 100 x objective, and digitally collected with SPOT camera (Diagnostic Instruments,
Sterling Heights, MI). Images were processed with Photoshop 6.0 (Adobe Software, San Jose, CA) and Canvas 6 (Deneba Software, Miami, FL).

Preparation of chlamydiae infected cells for protein gel electrophoresis

Confluent monolayers of HeLa cells or McCoy cells in six well trays were infected with \( 1 \times 10^6 \) wildtype and \( 1 \times 10^8 \) nonfusogenic mutant isolate of C. trachomatis in HBSS for 30 min on a rocker platform. The HBSS was removed and replaced with RPMI-1640 (Gilbo BRL) supplemented with 10% fetal bovine serum (FBS) and 10 \( \mu \)g/ml gentamicin. Monolayer cells were subsequently incubated at 37\( ^\circ \)C in 5% CO\(_2\). Chlamydia- infected cell and mock infected cells were harvested after 30 h. Cultured cells were washed twice with phosphate buffered saline (PBS; 150 mM NaCl, 10 mM NaPO\(_4\) pH 7.2) and lysed in 600 \( \mu \)l of 1x SDS-PAGE sample loading dye (1% sodium dodecylsulfate, 50mM Tris, pH 6.8, 1% 2-mercaptoethanol, and 10 % glycerol). Lysates were mixed and boiled for 5 min, aliquoted, and stored at \(-20^\circ\)C. These lysates were used for electrophoresis and immunoblot analysis.

Preparation of LCR lysates for amplification reaction

Monolayer culture of McCoy murine fibroblast cells grown in 6 well tissue culture tray were infected with different serovar of nonfusogenic mutant and wildtype strains of C. trachomatis (Table 1) at multiplicities of infection of approximately 3 and incubated at 37 \( ^\circ \)C for 30 h. Culture cells were washed with
HBSS, and lysed with a brief sonication. Sonicated cells were centrifuged at high speed (13000Xg) in a microcentrifuged tubes for min and supernatants were removed. Pellets were then resuspended in a commercial ligase chain reaction sample collect buffer (LCR Resuspension Buffer, Abbott laboratories, Abbott Park, IL) was lysed and consequently heated to 98°C for 15 min. The LCR lysates were used for DNA amplification of wild type and non-fusogenic mutants of C. trachomatis isolates.

Amplification reaction and DNA sequence analysis of incA and CT223

In non-fusogenic mutant study, five microliters of the LCR lysates were used as DNA template for amplification of target sequences. All amplification products were generated with Taq polymerase in 100 µl standard reaction, and then purified using the Qiaquick Columns (Qiagen, Valencia, CA). In case of incA sequence analysis, two independent region of incA were amplified with specific oligonucleotide primers that cover the first 500 bp region and full length of the gene. The first fragment consisted of the 5' half of incA and 429 nucleotide upstream of the coding sequence which was amplified using specific oligonucleotide GTTTTTTCATGGCCTCTTCTCT and GATCTGCTATGATTCTTGCG. The second fragment contained the 45 nucleotides upstream of the predictive start codon and the full length incA coding sequence which was amplified using specific oligonucleotide AGCCATAGGATCTGGTTTCAGCGA and GCGCGGATCCTAGGAGCTTTTTGTA GGGTGA. DNA sequence
analysis was performed on the 5’ half of incA contained in both of two fragments. The DNA sequence of the upstream region of incA was analyzed and sequenced from the amplification products of the first fragment.

Additionally, the entire incA sequence was sequenced, in case incA of C. trachomatis mutants isolates was not altered in the first 450 nucleotides. The amplification product of full length incA gene were generated using Taq polymerase. The PCR products was ligated into pCR2.1 vector, and transformed into competent INVαF’ E. coli cells. Transformation was performed using a standard technique and plated on LB plates containing 50 μg/ml of ampicillin and incubated overnight at 37°C. Positive recombinant clones were identified using restriction enzyme digestion analysis and polymerase chain reaction. Plasmid DNA of positive recombinant clones were isolated, purified by Qiagen column and sequenced.

The amplification of the upstream and full length of CT223, encoded for known Inc protein named CT223p was conducted using specific oligonucleotides. The primer sequences for upstream region of CT223 and full length were GCAGCGACAGCACTAATTATTTAAAA and CTCTGGACAGATTAGTTGATTGGAAA and ATGGTGAGTTAGGGA and GCGCGGATCTACCCCG AGAGCGTTAATTGAGGGA and GCGCGGATCTACCCCG AGAGCGTTAATTGAGGGA respectively.

DNA sequence analysis was performed using an Applied Biosystem 373A or 377 DNA sequencer (Foster City, CA). Sequence editing was performed using assemblylign 1.07 (Oxford Molecular Group, Campbell, CA).
**Electrophoresis and immunoblot**

To determine the production of IncA in non-fusogenic mutant of *Chlamydia trachomatis* isolates, protein profiles of chlamydial infected cell lysates at 18 and 30 hpi were determined using SDS-polyacrylamide gel electrophoresis. Five microliters of each diluted lysate sample was loaded on to 12% polyacrylamide gel and electrotransferred to nitrocellulose membrane. HSP60 was used as a Chlamydial protein to demonstrate that equivalent amounts of total chlamydial antigens were load for each sample. Polyacrylamide gel electrophoresis was conducted as previously described (Rockey and Rosquist, 1994) and immunoblot was performed on a Bio-Rad Trans blot cell (BioRad Laboratories) using 25 mM sodium phosphate blot buffer (pH 6.8). Proteins were transferred at 1.0 A for 1-2 h.

After transfer, the membrane was blocked with PBS plus 0.1% Tween 20 and 2% BSA in for 1 h. The membrane was then incubated with primary antibody diluted in blocking solution for 1h at room temperature. Before used, antisera were incubated with uninfected HeLa cells and commercial *E. coli* extract (Promega) for 6 h at room temperature on a platform rocker to increase antibody specificity. Pellets were removed from treated antisera by centrifugation at 1,200xg for 30 min. Absorbed antisera were used for immunofluorescence microscopy and immunoblotting and were stored at 4°C in 0.02% sodium azide.

Immunostaining with heat shock protein 60 (HSP60) was used as a control for a constitutively expressed gene product and for protein quantitation. Monoclonal antibody 20F12 against to CT223 was used for protein localization on
the inclusion membranes. The unbound antibody was removed by three time of washing in PBS with 0.1% Tween 20 for 5 min. The membrane was immunostained with $^{35}$S-anti- immunoglobulin G for IncA. HSP 60 and CT223p stained with monoclonal antibodies were detected using the peroxidase-conjugated chicken anti-mouse IgG for 1 h. The membrane was twice washed in PBS with 0.1% Tween 20 for 15 min. The nitrocellulose membrane stained with radioactive reagent was incubated at 37°C until completely dried. For the peroxidase reaction, membrane was visualized by incubation in chemiluminescent solution (ECL reagent, Pierce Chemical Co) for 5 min. Autoradiographs were performed and subsequently scanned into images and processed using Photoshop (Adobe system) and Canvas (Deneba Software) graphic system.

**In silico analysis of the CPn1054 gene family of C. pneumoniae genomes**

The following sets of ORFs of *C.pneumoniae* CWL 029: CPn0007, 0008/0009, 0010/0010.1, 0011/0012, 0041/0042, 0043/0044, 0045/0046, 0124/0125, 0126, 1054 and 1055/1056 were aligned using CLUSTALW analysis of Mac Vector™ 6.0. Each gene and each predicted gene product was also subjected to gap BLASTX and BLASTN respectively (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al., 1997) to identify homology to known genes and proteins in GenBank. The similarity between two different DNA sequences was determined using BLAST 2 sequence program from http://www.ncbi.nlm.nih.gov/blast/bl2seq
Phylogenetic analyses

The homology within the selected genes mentioned above was demonstrated using phylogenetic analysis. Multiple sequence alignment of the paralog or gene family was initially performed using CLUSTALW (Mac Vector™ 6.0) and subsequent phylogenetic analyses. Amino acid sequences corresponding to the hydrophobic domain of the selected genes were also aligned. Phylogenic analyses of both DNA and amino acid sequences were performed using PAUP (Swofford, 2001). In this study CPn0186 (IncA), and additional four putative inclusion membrane proteins, CPn0284, CPn0285, CPn0829 and CPn0830, were selected as outgroups. Phylogenic trees were inferred by neighbor–joining to estimate evolutionary distances. Bootstrap values were obtained from a consensus of 1000 neighbor-joining trees.

DNA amplification, and sequence analysis of the CP1054 gene family

Purified genomic DNA of selected isolates (Table 2) was amplified using specific oligonucleotide primers flanking the DNA region of interest in CPn0007, 0010/0010.1, 0041, 0043, 0045, 1054 and 1055 (Table 4). The amplification products were purified using a Qiaquick PCR purification spin column kit (Qiagen,
Valencia CA). The purified PCR product was sequenced using the ABI PRISM 377 (Perkin-Elmer, Norwalk, CT).

**Recombinant clones and sequence analyses**

The variation of the length of the poly C tract within a strain of the CPn0043, CPn1054 and CPn1055 was determined. Genomic DNA of *C. pneumoniae* AR39 was amplified using *Pwo* polymerase and specific oligonucleotide primer for CPn0043, CPn1054 and CPn1055 (Table 4). The *Pwo*-PCR products were directly ligated with vector using the Zeroblunt cloning system. According to the manufacturer’s instructions, the ligation with the pCR II vector was conducted with a 1:3 molar ratio of vector to PCR insert and transformed into competent INVαF’E. coli cells. Transformation was performed using a standard technique and plated on LB plates containing 50 μg/ml of ampicillin and incubated overnight at 37°C. Positive recombinant clones were identified using restriction enzyme digestion analysis and polymerase chain reaction. Plasmid DNA of 8-10 different positive recombinant clones were isolated, purified by Qiagen column and sequenced. Variations of the length of the poly C tract of CPn0043, CPn1054 and CPn1055 within *C. pneumoniae* AR 39 were determined for each recombinant clone.

Comparison analysis of the variation of the length of the poly C tract in the CPn1055 using two different DNA polymerases in the amplification reaction was further studied. Genomic DNA of *C. pneumoniae* AR39, PS32 and AR458 were
amplified using *Taq* and *Pwo* DNA polymerases and specific oligonucleotide primer for CPn1055 (Table 4). The PCR product was then cloned into pCR 2.1 and ZeroBlunt (TOPO) vectors respectively and transformed into *E. coli*. Plasmid DNA of positive clones were isolated, purified by Qiagen column and sequenced.

Additionally, the variation of the length of poly C tracts in PCR products amplified from recombinant plasmids but not genomic DNA was determined. Recombinant plasmid DNA extracted from positive recombinant clones of CPn0043, CPn1055, and CPn1054 derived from pCR2.1 and ZeroBlunt vectors were used as a template to generate a set of four amplification products using *Pwo* polymerase in the amplification reaction. The PCR products were isolated, purified, and sequenced as described above.
Table 4. Oligonucleotide primers used for PCR and sequencing of the internal regions of the CPn 1054 gene family from *C. pneumoniae* CWL029

<table>
<thead>
<tr>
<th>ORF</th>
<th>Region Amplified</th>
<th>Forward Primers (5'----3')</th>
<th>Reverse Primers (5'----3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPn0007</td>
<td>7488-8694</td>
<td>AAGAGTTTTTCGCTAT</td>
<td>TGACGACTCTTAGATGTTCG</td>
</tr>
<tr>
<td>CPn0010</td>
<td>13152-13502</td>
<td>GAAAAACACCTCTTTGGGTCTAG</td>
<td>AGACATACCCACGGAATAAATAAA</td>
</tr>
<tr>
<td>CPn0010</td>
<td>13175-14431</td>
<td>TACTCGCGAGGACCATACCTAGC</td>
<td>AACTCATGGAAACGTATCTCGTCC</td>
</tr>
<tr>
<td>CPn0010.1</td>
<td>15178-16324</td>
<td>GCTTTGGATGATTACGACGAG</td>
<td>CTAAGATTTTTTGGGTGCAGC</td>
</tr>
<tr>
<td>CPn0041</td>
<td>55728-56987</td>
<td>GAAAAATTCTTTTCGCAGAGATGGA</td>
<td>TCCTGTGCTCTCTCATAACGCTCAC</td>
</tr>
<tr>
<td>CPn0043</td>
<td>58069-59173</td>
<td>TGCTTTTCAGGCATTCCAGG</td>
<td>GCGAGCTACGACCTCTTTAGCGA</td>
</tr>
<tr>
<td>CPn0045</td>
<td>60755-61856</td>
<td>ATTTGAATCGATGGAGATTCCAGA</td>
<td>TACCTGCAGCATTCTACATCCTT</td>
</tr>
<tr>
<td>Cpn1054</td>
<td>1206885-1208142</td>
<td>AAAGGAAAGTTTTTCCTGGAAGATTCTT</td>
<td>ACTCATGGAAACGTATCTCGTCC</td>
</tr>
<tr>
<td>CPn1055</td>
<td>1209382-1210530</td>
<td>AGTTCTGTGAGATTCCTGAGGC</td>
<td>TATCTCCTCAACTACTCGAGATAAC</td>
</tr>
</tbody>
</table>
RESULTS

Identification and bioinformatic analysis of putative Inc proteins

Hydropilicity profiles of all open reading frames (ORFs) present in the *C. trachomatis* and *C. pneumoniae* genomes were determined using hydropathy plot analysis. A deduced amino acid sequence containing a bi-lobed hydrophobic motif (60-80 amino acids) in the secondary structure was predicted as a putative Inc protein. Based on the selection criterion, 45 of 894 ORFs of *C. trachomatis*, and 68 of 1073 ORFs of *C. pneumoniae*, were identified as putative Inc proteins (Table 5).

Putative *inc* genes occupy 2.7% of chlamydia-specific coding capacity in *C. trachomatis*, and 5.5% in *C. pneumoniae*. Open reading frames corresponding these putative Inc proteins appeared to be randomly distributed throughout each chlamydial genome. There are, however, some exceptions. Two contigs encoding a cluster of *inc* gene products were identified in the *C. trachomatis* genome. The first cluster contained four *inc* genes: CT115-118 in the contig 2.3, and the second was seven *inc* genes: CT222-229 in the contig 3.5. The four-*inc* gene cluster was located downstream of the known inclusion membrane protein of *C. trachomatis incA* (CT119). The gene products of CT115-118 were recently identified on the inclusion membrane using immunofluorescence staining (Scidmore-Carlson *et al.* 1999) named and recognized as IncD-G of *C. trachomatis*. A seven-*inc* gene cluster, CT 222-229, is located close to genes encoding amino acid—and
Table 5. Putative *inc* genes identified by hydrophilicity profile within each sequenced chlamydial genome

<table>
<thead>
<tr>
<th>C. trachomatis</th>
<th>C. pneumoniae</th>
<th>E-value</th>
<th>C. trachomatis</th>
<th>C. pneumoniae</th>
</tr>
</thead>
<tbody>
<tr>
<td>484^a</td>
<td>0602</td>
<td>5.2 e^-85</td>
<td>036</td>
<td>0007</td>
</tr>
<tr>
<td>728</td>
<td>0869</td>
<td>1.8 e^-50</td>
<td>115</td>
<td>0008</td>
</tr>
<tr>
<td>195</td>
<td>0288</td>
<td>2.7 e^-42</td>
<td>116</td>
<td>0010</td>
</tr>
<tr>
<td>565</td>
<td>0822</td>
<td>2.2 e^-30</td>
<td>117</td>
<td>0011</td>
</tr>
<tr>
<td>005</td>
<td>0443</td>
<td>2.0 e^-26</td>
<td>118</td>
<td>0041</td>
</tr>
<tr>
<td>383</td>
<td>0480</td>
<td>2.5 e^-20</td>
<td>134</td>
<td>0043</td>
</tr>
<tr>
<td>288</td>
<td>0065</td>
<td>1.5 e^-15</td>
<td>135</td>
<td>0045</td>
</tr>
<tr>
<td>058</td>
<td>0369</td>
<td>1.7 e^-13</td>
<td>164</td>
<td>0067</td>
</tr>
<tr>
<td>006</td>
<td>0442</td>
<td>7.4 e^-12</td>
<td>192</td>
<td>0124</td>
</tr>
<tr>
<td>850</td>
<td>1008</td>
<td>2.9 e^-12</td>
<td>196</td>
<td>0126</td>
</tr>
<tr>
<td>101</td>
<td>0312</td>
<td>1.4 e^-12</td>
<td>214</td>
<td>0130</td>
</tr>
<tr>
<td>232</td>
<td>0291</td>
<td>4.2 e^-10</td>
<td>222</td>
<td>0131</td>
</tr>
<tr>
<td>058</td>
<td>0370</td>
<td>2 e^-10</td>
<td>223</td>
<td>0132</td>
</tr>
<tr>
<td>618</td>
<td>0753</td>
<td>2 e^-10</td>
<td>224</td>
<td>0146</td>
</tr>
<tr>
<td>483</td>
<td>0601</td>
<td>4.1 e^-08</td>
<td>225</td>
<td>0147</td>
</tr>
<tr>
<td>233</td>
<td>0292</td>
<td>1.5 e^-07</td>
<td>226</td>
<td>0164</td>
</tr>
<tr>
<td>442</td>
<td>0556</td>
<td>1.2 e^-07</td>
<td>227</td>
<td>0166</td>
</tr>
<tr>
<td>058</td>
<td>0367</td>
<td>2 e^-05</td>
<td>228</td>
<td>0169</td>
</tr>
<tr>
<td>440</td>
<td>0554</td>
<td>1.2 e^-06</td>
<td>229</td>
<td>0174</td>
</tr>
<tr>
<td>449</td>
<td>0565</td>
<td>0.00045</td>
<td>300</td>
<td>0215</td>
</tr>
<tr>
<td>345</td>
<td>0026</td>
<td>0.0021</td>
<td>357</td>
<td>0216</td>
</tr>
<tr>
<td>119 (IncA)</td>
<td>0186</td>
<td>0.002</td>
<td>358</td>
<td>0225</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>789</td>
<td>0240</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>813</td>
<td>0241</td>
</tr>
</tbody>
</table>

^a boldface identifies ORFs examined in this work.
sodium–dependent transporter proteins. BLAST analysis revealed that gene products of CT232 and CT233 have significant sequence similarity to known Inc proteins of \textit{C. psittaci} IncC and IncB. In \textit{C. trachomatis}, clusters of putative inc genes were tightly linked to both the \textit{incA} (CT119) and the \textit{incB/incC} (CT223/CT222) homologues. Such gene order, however, is not present in the \textit{C. pneumoniae} genome. The predicted sequences of IncA of \textit{C. trachomatis} and \textit{C. pneumoniae} were more similar to that of \textit{C. psittaci} than to each other. Additionally, like \textit{C. psittaci incB/incC}, several putative inc genes in \textit{C. trachomatis} and \textit{C. pneumoniae} genome, were arranged in tandem. These included CT134/135, CT232/233, CT357/358, CT483/CT484, CPn0130/0131/0132, CPn0146/0147, CPn0215/0216, CPn0266/0267, CPn0284/285, CPn0354/0355, CPn0365/0366, CPn0371/0372, CPn0431/0432, CPn0439/0440, CPn0829/0830, CPn0601/0602, and CPn1054/1055.

Sequence similarity analyses were conducted to determine the similarity of putative Incs to known proteins in GenBank. BLAST search analysis showed that there was no significant similarity of these putative Inc proteins to any known protein in the database. However, the similarity analysis of putative Inc proteins of \textit{C. trachomatis} and \textit{C. pneumoniae} using the BLAST search and BLAST 2 sequences program demonstrated that some ORFs are orthologous or paralogous genes which have significantly sequence homology between and/or within the species (Table 5). For example, ORF CT484 and CPn0602 are orthologs considerably derived from gene speciation, which share significantly sequence
similarity each other between species. The similarity data are consistent with BLAST data present in the Chlamydia Genome Project (Stephens et al., 1998). Genomic analysis also revealed that several putative inc(s) were defined as paralogs likely derived by gene duplication in C. pneumoniae genome. These included CPn0007, 0008, 0010, 0041, 0043, 0045, 1054 and 1056, and CPn0367, 0369, 0370 and 0524. CPn0367, 0369, 0370, and 0524 have sequence similarity to CT058, a putative C. trachomatis inc. These putative inc(s) were quite similar each other and to CPn0107, which is smaller than the others and lacks a bilobed hydrophobic motif, the predictive marker for localization to the inclusion membrane. Except CPn0524, sequence similarity alignment of these genes, especially at the carboxy termini was shown (Figure 5). The similarity of paralogous genes, CPn0007, 0008, 0010, 0041, 0043, 0045, 1054 and 1056 will be further studied using phylogenetic analysis and comparative genomic approaches later in the this chapter.

Phylogenetic analysis of the primary sequence of putative inc using the PAUP method failed to identify the genetic relationship within the Inc group. This was true for analyses of either complete protein sequence, the predicted hydrophobic domain, or analyses of the genes remaining after removal of the bilobed hydrophobic region. Collectively, among putative Inc proteins, a bilobed hydrophobic domain is the only common structural motif present in the secondary structure of the chlamydial Inc protein.
Figure 5. Comparison of predicted proteins encoded by *C. trachomatis* CT058 and *C. pneumoniae* CPn0107, CPn0367, CPn0369 and CPn0370.

A) Hydrophilicity profiles for three of these predicted proteins. (Note the scale differences for each plot). B) Similarity of these predicted proteins. Sequence alignment of the carboxy termini of each protein using CLUSTALW (Mac VectorTM 6.0). Amino acid similarity and position are indicated to the left of each line.
**Protein localization using immunofluorescence staining**

Immunofluorescence staining was used to test the hypothesis that a bi-lobed hydrophobic domain can be used as predictive marker for chlamydial inclusion membrane proteins. Six putative *inc* genes of *C. trachomatis*: CT119, CT223, CT229, CT233, CT288, CT442, CT484, and CPn0186, a candidate *incA* gene of *C. pneumoniae*, were selected to identify the protein localization (Table 5).

Fluorescence staining revealed that all selected putative Inc proteins were localized onto the inclusion membrane except CT484 p (Figure 6). The localization of gene product of CPn0186, the *C. pneumoniae* IncA homology, was also shown on the inclusion membrane protein of *C. pneumoniae* infected cells (Figure 7). Although this ORF encodes the sequence that is about 20 % similar to *C. psittaci* and *C. trachomatis incA*, the secondary structure of the gene product contains the common bi-lobed hydrophobic domain identified in known Inc proteins. Five of six selected putative Incs were localized onto the inclusion membrane, including the CT233 homologue, *C. psittaci* IncC, CT442 (a cysteine rich protein), and three previously unexamined proteins encoded by open reading frames CT223, CT229 and CT288. The localization of CT233 and CT288 cannot be detected on the chlamydial inclusion after methanol fixation. Fixation with paraformldehyde and subsequent permeabilization with Triton X-100 was used to demonstrate the localization of these proteins on the inclusion the *C. trachomatis* infected cells. Both proteins were shown to localize to discrete patches within the inclusion (Figure 6). Anti- CT223p antibodies also labeled *C. trachomatis* L2.
Figure 6. Immunofluorescence microscopy of *C. trachomatis* infected HeLa cells fixed with methanol and probed with specific antibody against each tested putative Inc. Panels A and B represent a cell doubly labeled with anti-IncA (A) and HSP60 (B) and demonstrate the characteristic labeling evident for known Inc proteins. Panel C-H each represents cells singly labeled with each specific antisera against CT229(C), CT223(D), CT228(E), CT233(F), CT442(G) and CT484(H). Arrows indicate the inclusion membrane in E,F and G. Bar in H represents 10 microns for each image.
Figure 7. Double-labeled fluorescence microscopy of *C. pneumoniae* labeled with anti–HSP60 (A) and anti-*C. pneumoniae* IncA (CPn0186)(B). Infected cells were fixed with methanol at 44 h. and labelled with anti–HSP60 (A) and anti-*C. pneumoniae* IncA (CPn0186)(B). Note the IncA is distributed the margins of the chlamydial inclusion, similarly to as seen with *C. psittaci* and *C. trachomatis* (Figure 6A).
inclusions with a unique pattern (Figure 8). The distinctive pattern between IncA and CT223p was shown using double immunostaining with anti-IncA and CT223p.

A gene product of ORF CT484, one of selected putative inc genes was not found on the inclusion membrane. The localization of gene product CT484 was absolutely undetectable from either early or late inclusions fixed with either methanol or paraformaldehyde. The gene product of this putative inc has extensive sequence similarity with a corresponding gene product of CPn0602 in C. pneumoniae genome. Cross reactivity of antisera against CT484p, therefore, also was determined. A weak immunostaining of C. pneumoniae-infected cells was observed with anti-CT484p antibodies (data not shown).

**Fluorescence microscopy of cells infected non-fusogenic variant isolates**

A previous study by Suchland *et al.* (2000) revealed that 1.5 % (176 of 11,440) of independent clinical isolates of *C. trachomatis* infected patients collected over the 12 years showed a variant inclusion phenotype. Variant isolates form multiple-lobed inclusions within single cells infected at MOI greater than 1 whereas wild type isolates form a typical single inclusion, following infection of cells at high multiplicities (Figure 6). In this study, 13 wild type and 26 non-fusogenic variant isolates were subjected to examine the localization of IncA protein on the *C. trachomatis* infected cells using immunofluorescence staining.
Figure 8. The distinct pattern of the localization of CT223 on the inclusion membrane. *Chlamydia trachomatis* serovar L2 infected HeLa cells fixed with methanol and doubly labeled with anti-IncA (A) and anti-CT223p antisera.
The localization of IncA to the inclusion membrane is always demonstrated in wild type clinical isolates. In contrast, multiple-lobed inclusions form in infected cells under similar condition by variant isolates. Two distinct groups of nonfusogenic variant isolates were identified with respect to the presence or absence of IncA in fluorescence profile. Most non-fusogenic variant isolates lack IncA protein localized on the inclusion membrane. Twenty four of 26 non-fusogenic mutant isolates were negative for immunofluorescence staining of IncA protein (Figure 9). Two non-fusogenic variant isolates {G(s)459 and J(s)5942} were positive for IncA protein immunofluorescence staining (Figure 9,10). The IncA-positive variants were morphologically distinct from the IncA-negative nonfusogenic variants. A high degree of variability in the number, size and shape of individual vacuoles was observed within the infected cells. The fluorescence staining of IncA and CT223p in these isolates was very strong, suggesting possibly high levels of each protein.

Additionally, the localization of CT223p on the inclusion membrane was determined. Initially, identification of CT223p was used as the positive control of the localization of Inc of C. trachomatis infected cells. In these clinical isolates, four variant isolates showed the absence of the CT223p on the multiple-lobed inclusion. One of the 13 wild type isolate (J3464) and 3 of the 26 variants {J(s)1980, J(s)6276, J(s)6686} were shown to lack CT223 as determined by fluorescence staining. All identified CT223-negative isolates were of serovar J or J(s). There was no correlation between fusogenicity and non-fusogenicity, or any
Figure 9. Fluorescence microscopy of McCoy cells infected with serovar J(s) variant of Chlamydia trachomatis probed with antibodies against HSP60, IncA and CT223. Cells were infected and incubated 30 h prior to methanol fixation and labeling with antibodies specific for IncA (left column), or with a combination of anti-HSP60 (green) and anti-CT223 (red). The isolates number is indicated to the left of each pair of pictures. Note that the images in the left column do not show the same cells presented in the right column. All images were magnified and photographed under identical conditions.
Figure 10. Fluorescence microscopy of McCoy cells infected with serovar G(s) 459, an Inc A positive variant, of *C. trachomatis* probed with antibodies against IncA. Cells were infected and incubated 30 h prior to methanol fixation and labeling with antibodies specific for IncA (A) Wildtype serovar G; (B) servar G(s)459. In these cells, the varied size and shape of the IncA-positive inclusion are evident. As with figure 8, note that these images do not represent the exact same cells.
other observable aspect of inclusion development, and the presence of absence of CT223p.

**Immunoblotting analysis of selected non-fusogenic variant isolates**

The protein profile of IncA and CT223p of wild type and variant infected cells was determined by immunoblotting. It was used to confirm the result of immunofluorescence staining of IncA and CT223p and to verify that those proteins were not present on the inclusions at detectable levels within infected cells. Chlamydial HSP60 was included as a control to demonstrate that similar amount of total chlamydial antigen was loaded for each sample. The immunoblots of representatives of serovar D, J, H, of wild type and variant isolates was shown (Figure 11). The results of immunoblot were consistent with those of immunofluorescence microscopy. Isolates that lacked detectable IncA and CT223p in the inclusion membrane were also negative by immunoblot.

According to the presence of CT223p in immunofluorescence microscopy and immunoblotting determined by monoclonal antibody, the possibility exists that CT223p negative variants are actually producing this protein, but the alteration in sequence eliminated the epitopes recognized by mAb. To address this, the variant CT223 sequence was cloned into E.coli using a similar approach to that used for production of the wild type CT223p. The mutant protein was recognized by the anti-CT223p-mAb in immunoblot analyses (data not shown).
Figure 11. Immunoblot analysis of non-fusogenic *C. trachomatis* isolates probed with antibodies against HSP60, IncA, and CT223p. Lysates of HeLa cells infected with selected non-fusogenic and wild type *C. trachomatis* isolates were electrophoresed, transferred to nitrocellulose membrane, and probed with either anti-IncA, anti-CT223p or anti-HSP60. Lane 1+10, Mock-infected lysate; 2+11, wild type serovar D; 3, D(s) 2923; 4, D(s) 8093; 5+12, wild type serovar J; 6, J(s) 893; 7, J(s) 1980; 8, B(s) 1129; 9, H(s) 5942; 13, J(s) 1980; 14, G(s) 459
Sequence analysis of incA of non-fusogenic variant isolates

The incA sequence from 26 independent nonfusogenic variants and 13 wild type isolates representing ten serovars or sub-serovars were examined (Table 1). We studied a single isolate for serovar B(s), G(s), H(s), E(s) and K(s), two isolates of serovars F(s), and Ia(s) and three or more isolates for serovars D(s), D-(s) and J(s). Wild type isolates used for control were of serovar B, D-, F, G, Ia, and E and seven for serovar J. The variation of incA sequence in the population of wild type C. trachomatis isolates was also examined. Sequence data and Genbank accession number for variants are presented (Figure 12).

In these studies, 5 of 13 (38%) wild type isolates (E9332, J2364, J3464, J9311 and J 9325) contained incA identical to prototype incA sequence, and 8 of 13 (62%) wild type isolates showed nucleotide variation of incA sequence (Figure 12). Three of 13 (39%) wild type isolates (D-9291, F9334, and J9336) were incA variants with the substitution of isoleucine with threonine (I47T) at codon 47 and glutamic acid with lysine (E116K) at codon 116 in the gene product. Such variant sequences were also identified in nonfusogenic variant isolates. These sequences were slightly different from the prototype incA sequence of C. trachomatis serovar D{(D/ UW-3/Cx), Genbank Accession #: AE001273} present in the genome project. However, there was no change resulting in alteration of a reading frame or introduction of a stop codon of incA identified in those wild type isolates.
Figure 12. A graphic representation of the results from the DNA sequencing of incA encoded by 10 wild type and 25 non-fusogenic C. trachomatis isolates. To the left of the isolate designation are the accession numbers and the result from fluorescence microscopy of the isolates with antibodies to Inc. The complete IncA sequence was determined for the isolates shown in bold while the first 450-550 nucleotides was analyzed for all other isolates. The scale of the amino acids in the hydrophobic domain is used to position in the symbols representing the mutations. In the legend the symbols are defined—note that a variety of distinct lesions within the incA coding sequence are associated with the lack of IncA within the inclusion membrane. Isolates G(s) and J(s)5942 are nonfusers encoding IncA that is properly localized to the inclusion membrane. Q, silent base changes in incA; *, a base change that leads to an amino-acid substitution; ▼, a single base deletions or insertions that lead to premature truncation of the protein; ●, single base changes that leads directly to a stop codon. Solid bars indicate larger deletions and are drawn to scale.
<table>
<thead>
<tr>
<th>Wild type</th>
<th>IncA IF</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>B9227</td>
<td>+</td>
<td>AF273326</td>
</tr>
<tr>
<td>D-9291</td>
<td>+</td>
<td>AF273328</td>
</tr>
<tr>
<td>F9334</td>
<td>+</td>
<td>AF327327</td>
</tr>
<tr>
<td>G9301</td>
<td>+</td>
<td>AF273329</td>
</tr>
<tr>
<td>Ia9309</td>
<td>+</td>
<td>AF273330</td>
</tr>
<tr>
<td>J9329</td>
<td>+</td>
<td>AF273331</td>
</tr>
<tr>
<td>J9336</td>
<td>+</td>
<td>AF273332</td>
</tr>
<tr>
<td>J9346</td>
<td>+</td>
<td>AF273333</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variant</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>B(s)1129</td>
<td>AF276501</td>
</tr>
<tr>
<td>D(s)2923</td>
<td>AF163773</td>
</tr>
<tr>
<td>D(s)4013</td>
<td>AF279340</td>
</tr>
<tr>
<td>D(s)8039</td>
<td>AF276502</td>
</tr>
<tr>
<td>D(s)1012</td>
<td>AF276503</td>
</tr>
<tr>
<td>D(s)5058</td>
<td>AF279341</td>
</tr>
<tr>
<td>D(s)5328</td>
<td>AF279342</td>
</tr>
<tr>
<td>D(s)3991</td>
<td>AF279359</td>
</tr>
<tr>
<td>D(s)1986</td>
<td>AF279360</td>
</tr>
<tr>
<td>E(s)1986</td>
<td>AF279343</td>
</tr>
<tr>
<td>F(s)4022</td>
<td>AF279344</td>
</tr>
<tr>
<td>G(s)459</td>
<td>AF279345</td>
</tr>
<tr>
<td>H(s)5642</td>
<td>AF279346</td>
</tr>
<tr>
<td>H(s)9005</td>
<td>AF273334</td>
</tr>
<tr>
<td>Ia(s)2197</td>
<td>AF279347</td>
</tr>
<tr>
<td>Ia(s)7713</td>
<td>AF279348</td>
</tr>
<tr>
<td>J(s)893</td>
<td>AF279349</td>
</tr>
<tr>
<td>J(s)6462</td>
<td>AF279350</td>
</tr>
<tr>
<td>J(s)6276</td>
<td>AF279351</td>
</tr>
<tr>
<td>J(s)1960</td>
<td>AF279352</td>
</tr>
<tr>
<td>J(s)3839</td>
<td>AF279353</td>
</tr>
<tr>
<td>J(s)6656</td>
<td>AF279358</td>
</tr>
<tr>
<td>J(s)5942</td>
<td>AF279354</td>
</tr>
<tr>
<td>J(s)6886</td>
<td>AF279355</td>
</tr>
<tr>
<td>J(s)6276</td>
<td>AF279356</td>
</tr>
<tr>
<td>K(s)2481</td>
<td>AF279357</td>
</tr>
</tbody>
</table>

Figure 12.
On the other hand, these analyses identified a wide variety of distinct incA sequence associated with the non-fusogenic phenotype (Figure 12). The sequence analysis of 18 of 26 nonfusogenic mutant isolates revealed the alteration of the 5’ half of incA, which led to the truncated products. The incA sequence of most non-fusogenic variant isolates (12 of 26) contained a frameshift mutation resulting from single base change, which either altered the reading frame or caused directly to a stop codon. Most of these are within the 5’ half of the gene, with the two exceptions being D(s)1012 and F(s)8068. Seven of 26 mutant isolates: B(s)1129, D(s)1012, D(s)8323, D(-s)5328, D(-s)3991, E(s)1986 and J(s)1980 contained nonsense mutations leading to the truncated gene product of incA.

Large deletions in incA were also identified from 4 non-fusogenic variant isolates: J(s)893, J(s)6462, J(s)6686, and H(s)5642. An identical 51 nucleotide (position 152-202) in-frame deletion of incA which resulted in the internally truncated protein product was identified in J(s)893 and J(s) 6462 isolates. In addition, J(s) 6686 and H(s)5642 have a 14 nucleotide (position 18-31) and a 671 nucleotide out-of-frame deletion respectively, which leads to a change in reading frame and a short, truncated, predicted products. A large deletion (671 nucleotides) identified from H(s)5642 isolate extended from position 366 of incA into the 3’ untranslated region. The sequence analysis of several selected isolates of serovar D(s), D(-s) and J(s) were performed. Three different isolates of serovar D(-s): 8323, 5328 and 3991 have an identical nucleotide change that leads to the introduction of a stop codon.
The complete \textit{incA} sequence was also determined from an IncA-negative non-fusogenic isolates, D(s) 2923, and two IncA-positive non-fusogenic isolates G(s)459 and J(s)5942. In D(s)2923, \textit{incA} was altered at 3 nucleotides, leading to the amino acid substitutions with the substitution of isoleucine with threonine (I47T) at codon 47 and glutamic acid with lysine (E116K) at codon 116. Such these changes in amino acid sequence were also identified in wild type isolates. In IncA-positive non-fusogenic isolates, G(s)459 and J(s)5942 showed the variation of \textit{incA} but they still encode intact, complete full-length \textit{incA} sequences. Therefore, there were no mutations in \textit{incA} that were uniquely associated with the IncA-positive nonfusogenic isolates.

\textbf{DNA sequence analysis of ORF CT223}

The complete sequence of ORF CT223 was conducted from six wild type and six non-fusogenic variant isolates. Included in these analyses were 3 serovar J or J(s) isolates that were CT223p-negative by fluorescence microscopy and immunoblot. Each DNA sequence were translated into predicted amino acid sequence and compared to those of \textit{C. trachomatis} serovar D and L2 provided through the Chlamydia Genome Project. In contrast to \textit{incA}, there is considerable variation in the CT223 sequence of the serovar D and L2. The predicted sequences of CT223 for serovar D and L2 have shared 91 \% similarity. While \textit{incA} varies at only a few nucleotides, CT223 present in the database vary at 34 nucleotides leading to 26 differences in amino acid sequence. The CT223p of the variant
sequences consisted of mosaic of the serovar D and L2 genomic sequence. Sequence variation resulting from nucleotide substitution in CT223 was identified in both wild type and variant isolates. In the three CT223p-negative isolates, J3464, J(s)1980, J(s)6686, the deduced amino acid sequences of CT223 were identical to each other but distinct from the either prototype CT223 sequence serovar D or L2 {Genbank accession number’s for J3464 (AF279363); J(s)1980 (AF279362); J(s)6686 (AF279364)}. Each sequence had two unique nucleotide substitutions that resulted in serine at position 22 changed to leucine, and glutamine at position 120 changed to arginine (Figure 13). Neither of these mutations was identified in any of six CT223 positive isolates or within the CT223 published sequence from serovar D or L2. There is no association of changes in amino acid and the absence of CT223p determined by immunofluorescence and immunoblot studies.

**Upstream sequence analysis of incA and CT223**

The upstream region of *incA* of 6 of 13 wild type and 18 of 26 variant isolates were amplified and sequenced in order to determine the variation of predicted promotor regions of each isolate. These included D(s)2923 and J(s)893 isolates with intact coding sequences. Changes in the promotor sequence might be responsible for the lack of the IncA in these isolates. However, the *incA* upstream region of both of wild type and variant isolates was identical to that of the *C.* responsible for the lack of the IncA in these isolates. However, the *incA* upstream
Figure 13. Multiple sequence alignment analysis of CT223p of fusogenic and non-fusogenic variants of C. trachomatis isolates. Deduce amino acid sequence from 6 wild type and 6 non-fusogenic variants were aligned and analyzed and compared to those of C. trachomatis serovar D and L2 provided through the Chlamydia Genome Project. The variant sequences consisted of mosaic of the serovar D and L2 genomic sequence. In the three CT223 negative isolates, J 3464, J(s)1980, J(s) 6686, the deduced amino acid sequences of CT223 were identical but distinct from the either prototype CT223 sequence (serovar D or L2) {accession #’s for J3464(AF279363); J(s)(AF279362); J(s)6686(AF279364)}. Each sequence had two unique nucleotide substitutions, which resulted in serine at position 22 changed to leucine and glutamine at position 120 changed to arginine.
Figure 13.
region of both of wild type and variant isolates was identical to that of the C. trachomatis prototype strain. There is no difference in the predicted promotor region of incA in any selected isolates. On the other hand, the upstream sequence of CT223 of variant isolates that lack the protein CT223p showed sequence variation. The nucleotide substitution of thymine (T) with cytosine (C) of the upstream region (at −111 from the start site) was identified in all CT223-negative variants but not any wild type and CT223-positive variant isolates (data not shown) {Genbank accession number's for J3464 (AF279363); J(s)1980 (AF279362); J(s)6686 (AF279364)}.

In silico analysis of the CPn1054 gene family

A novel gene family sharing extensive sequence similarity within the C. pneumoniae genome was identified using Gap BLASTX and BLASTN search and multiple sequence alignment analysis of CLUSTALW (Vector™ 6.0, Oxford Molecular). This gene family, designated as the CPn1054 gene family, consisted of 19 C. pneumoniae-specific genes: CPn0007, 0008, 0009, 0010, 0010.1, 0011, 0012, 0041, 0042, 0043, 0044, 0045, 0046, 0124, 0125, 0126,1054, 1055, and 1056. These genes are located in four separate regions on the chromosome: contig 1.0-1.1 (CPn0007-CPn0012), contig 1.5-1.6 (CPn0041-CPn0046), contig 2.5 (CPn0124-126) and contig 13.0-13.1 (CPn1054-CPn1056) (Figure 14). Similarity of DNA sequences of paralogous genes using BLAST 2 Sequence program ranged between
*Chlamydia pneumoniae* CWL029 genome

**Contig 1.1**

**Contigs 1.5-1.6**

**Contig 2.5**

**Contigs 13.0-13.1**

Figure 14. A schematic representation of *C. pneumoniae* CWL029 contigs. The CPn1054 gene family is encoded in the contig 1.0-1.1 (CPn0007-0012), contig 1.5-1.6 (CPn0041-0046), contig 2.5 (CPn0124-0126), and contig 13.0-13.1 (CPn1054-1056).
Table 6. DNA sequence identity of paralogous genes in the CPn1054 family using Pairwise of BLAST 2 sequence program

<table>
<thead>
<tr>
<th>ORF</th>
<th>% identity for paralogous genes of the DNA sequence&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0008/9</td>
</tr>
<tr>
<td>0008/9</td>
<td>81</td>
</tr>
<tr>
<td>0010/10.1</td>
<td>81</td>
</tr>
<tr>
<td>0041/42</td>
<td>78</td>
</tr>
<tr>
<td>0043/44</td>
<td>76</td>
</tr>
<tr>
<td>0045/46</td>
<td>85</td>
</tr>
<tr>
<td>1054</td>
<td>81</td>
</tr>
<tr>
<td>1055/56</td>
<td>82</td>
</tr>
</tbody>
</table>

<sup>a</sup> Pairwise of BLAST 2 sequence program was performed to determine the identity between two different DNA sequences (Tatiana A, 1999). Each value was rounded to the nearest integer, and values of >85% are presented in boldface. No analysis was possible for CPn0011/0012, CPn0124/0125, and CPn0126 because there was no significantly specified sequence due to alignment of paralogs less than 20%.
Figure 15. Phylogenetic tree analysis of DNA sequences of the CPn1054 gene family obtained from *C. pneumoniae* CWL029. Multiple sequence alignments were performed and analyzed using CLUSTALW and PAUP program. Distance matrices were inferred by neighbor–joining to estimate evolutionary distances. Bootstrap values (based on 1000 replicates) are represented at each node.
20-99% (Table 6). The existence of the paralogous genes and a gene family was further strengthened using phylogenetic analysis performed by PAUP (Figure 15). This also suggested that this gene family has extensive genetic relationship distinct from the other putative Inc proteins. A classical signal peptide and conserved Shine-Dalgarno sequence were not identified in the predicted products of this gene family. In CPn0008, 0010, 0041, 0045 and 1055 contained the unusual putative start codon GTG at the different site in different strains (Figure 16). Genomics analyses revealed that the DNA sequence of CPn0008, 0010, 0011, 0041, 0043, 0045, 0124 and 1055 were similar to the 5' end of the CP1054 DNA sequence; while CP0009, 0010.1, 0012, 0042, 0044, 0046, 0125 and 1056 were similar to the 3' end of the CP1054 (Figure 17).

**The conserved hydrophobic domain of the CPn1054 gene family**

In a previous study, we demonstrated that a bi-lobed (50-60 amino acid) hydrophobic domain is a predictive marker for localization into the inclusion membrane. Consequently, 68 *C. pneumoniae* proteins, including several members of the CPn1054 gene family, have been reported as putative inclusion membrane-associated proteins (Inc proteins; Figure 18). In this study, high amino acid sequence similarity was identified in the bi-lobed hydrophobic domains of the CPn1054 gene family. Such high similarity of the bi-lobed hydrophobic domain in the secondary structure in this gene family was unique which was not present in any known Inc proteins (Figure 19). Additionally, the phylogenetic tree analyses
Figure 16. The upstream and 5’ end sequence of the CPn1054 gene family. Multiple sequence alignments of the upstream region of the CPn1054 gene family obtained from the CWL029 genome. Stop codon of CPn00009, 0040, 0042, 0044, 1054 (▼) were indicated. The unusual start codon GTG (▲) was identified in CPn0045, 0043,1055. Comparative genomics revealed that the start site of CPn0008, 0010,1054 (★) varied in different strains. Homopolymeric cytosine (poly C) tracts encoded in either the predicted upstream or at the 5’ end of coding region of the CPn1054 gene family.
Figure 16.
Figure 17. A schematic representation of the structure features and comparison of paralogous genes of the CPn1054 gene family in *C. pneumoniae* CWL029. The 5' end of CPn1054 was similar to CPn0008, 0010, 0011, 0043,0045, and 1055 while the 3' end was similar to the CPn0009, 0010.1, 0012, 0042, 0044, 0046, and 1056. Range of similarity is 20-99 % Positions of the DNA sequence in the contig included in the alignment were indicated with numbers on the top of each allele. Deletion
Figure 18. Hydropathy plot analysis of the members of the CPn1054 gene family. The analysis was performed using the Kyte-Doolittle method with a sliding window of 7. These proteins are classified as putative inclusion membrane proteins containing the bilobed hydrophobic domain (boxed), a predictive marker for localization to the inclusion membrane (Inc). Note that scale on the horizontal axis is different for each protein.
Figure 18.
Figure 19. The conserved hydrophobic domain in the CPn1054 gene family. Multiple sequence alignments of the hydrophobic domains of the CPn1054 gene family using CLUSTALW analysis. This analysis revealed the remarkably high amino acid sequence similarity of the hydrophobic domain, which was found only in this putative Inc protein family.
Figure 20. Phylogenetic tree analysis of amino acid sequence of hydrophobic domains of the CPn1054 gene family obtained from *C. pneumoniae* CWL029. Multiple sequence alignments were performed and analyzed using CLUSTALW and PAUP program. Distance matrices were inferred by neighbor-joining to estimate evolutionary distances. Bootstrap values (based on 1000 replicates) are represented at each node.
also demonstrated the homology of the hydrophobic domain of the CPn1054 gene family (Figure 20).

**Homopolymeric cytosine (poly C) tract and the variations of the length of poly C tracts in the CPn1054 gene family**

Genomic analysis of the CPn1054 gene family revealed a short DNA repeat of 6-15 homopolymeric cytosine (poly C) residues, positioned either upstream or at the predicted 5'end of CPn0008, 0010, 0041, 0043, 0045, 1054, and 1055 (Figure 16). Comparative genomic analysis of CWL029, AR39 and J138 also revealed variation of the length of poly C tracts between strains in CPn0008, 0010, 1054 and 1055, but not in CPn0041 (CCCCCCTCCCCC), 0043(CCCCCCCCCCCCCC), and 0045(CCCCCC) (Figure 21 A). We further investigated the variation between strains of the length of poly C tract for these genes from the 12 different *C. pneumoniae* isolates. The variation of the length of poly C tracts between strains was identified in CPn0010, CPn0043, 1054, and 1055. No variation between strains of the length of poly C tracts of CPn0041 and CPn0045 was identified in any isolate.

The variation within a strain of the length of the poly C tracts of CPn0043, CPn1054, and CPn1055 was determined from 8-10 recombinant clones generated by either *Taq* or *Pwo* amplified PCR products. Within *C. pneumoniae* AR 39, the length of poly C tract of these genes in each recombinant clone was shown (Figure 21 B). Additionally, the variation within a strain and between strains of the length of poly C tract of CPn1055 was identified in *C. pneumoniae* AR 39, AR 458 and
PS 32 isolates (Figure 21 C). Whether or not the amplification process using either Taq or Pwo polymerase had an effect on the variation of the length of poly C tracts was determined. The intrastrain variation of the length of the poly C tract of CPn1055 in C. pneumoniae PS 32 was identified using either Taq or Pwo polymerase in amplification reaction. (Figure 21 D). In contrast, no variation of the length of poly C tract of CPn0043, 1054, and 1055 was identified in PCR products amplified from plasmid DNA of recombinant clones derived from either TA or ZeroBlunt vector.

The tandem repeat motif in CPn0007

Hydropathy plot and multiple sequence alignment analysis revealed that CPn0007 contained three identical copies of a hydrophobic domain (Figure 18) which corresponded to “tandem repeat motif” while the others genes of the interest have only a single hydrophobic domain. Comparative genomics of the complete genome sequence of CWL029, AR39, and J183 indicated that the three tandem repeat motif in CPn0007 was not identical in all three strains (Shirai et al., 2000). The CPn0007 of J138 was a truncated variant with the deletion of 110 nucleotides that led to lack of the second domain in the tandem repeats.
Figure 21. Variation of the length of the poly C tract in the CPn1054 gene family.

(A) Interstrain genetic variation of the length of poly C tract in the CPn1054 family identified using comparative genome analysis of the complete C. pneumoniae CWL029, AR 39 and J 138 obtained from the genomes.

(B) Intrastrain variation of the length of poly C tract of the CPn0043, 1054,1055 in C. pneumoniae AR 39.

(C) Intrastrain variation of the length of poly C tracts of CPn1055 identified in C. pneumoniae AR 39, AR458, and PS32.

(D) Comparison of two different DNA polymerase enzymes. Taq and Pwo, were used for polymerase chain reactions. CPn1055 of C. pneumoniae PS32 was amplified by Taq and Pwo and directly cloned into pCR2.1, TA kit and Zeroblunt respectively.
Figure 21.
In this study the size variation of CPn0007 of 12 different *C. pneumoniae* isolates generated using polymerase chain reaction was further determined. It revealed that the PCR product corresponding to CPn0007 was identified in all *C. pneumoniae* isolates tested. There was no evidence of deletion or insertion of a repeat motif of CPn0007 in any isolates (data not shown).

**Frameshift and Gene conversion in the CPn1054 gene family**

Comparative genomics analysis of the three genomes showed that the sequence variation in CPn0010, CPn0043, and CPn1054 was caused by deletion, nucleotide substitution and gene conversion. Frameshift mutations were found which lead to a truncated sequence of CPn0010-0010.1 and CPn0043-0044 in CWL029 but not in AR 39 and J138. Frameshift mutations and nucleotide substitution were also identified that led to a truncated sequence of CPn1054 in AR39 and J138 (Figure 22). In CWL029, CPn0010-0010.1 and CPn1054 shared 99% sequence similarity with a frameshift mutation and small difference identified at the 3’ end of each gene, while in AR39 and J138, CPn0010 were identical to CPn1054 with one gap and gene conversion (Figure 23).

In this study, the frameshift mutation and gene conversion of CPn0010 were further investigated from 12 different *C. pneumoniae* isolates. The frameshift mutation leading to the truncated sequence of CPn0010 was identified in CWL029 and 4 other isolates, AR388, AR231, KA5C and KA66. Evidence for gene conversion, in which CPn0010 was converted to CPn1054, was identified in 2 of 12
isolates, AR 39 and AC43 (Figure 24). The nucleotide sequences of the frameshift sequence and the gene conversion of CPn0010 at the 3’ end of the gene identified from 12 selected samples were deposited in GenBank under following accession numbers: AF474017-474026, and AF461543-461552 respectively.
Figure 22. Schematic representation of the open reading frame of CPn0010-0010.1, 1054, 0045 identified in *C. pneumoniae* CWL029, AR39 and J138. Comparative genomics analysis demonstrated gene fusion or gene fission found in the CPn0010 I in CWL-029 strain, CPn1054 in AR 39 and CPN0045 in AR39 and J138. Truncated genes caused by frameshift mutation were identified as shown.
Figure 23. Frameshift mutation and gene conversion in CPn0010.

A) Comparative genomics revealed the frameshift mutation (⁎), which lead to the truncated CPn0010 (894 bp) in CWL029 but not in AR39 and J138 (2322 bp). Stop codons (▼) in truncated CPn0010 (CWL029) resulting from the frameshift were shown.

B) Gene conversion in CPn0010 identified in AR39 and J138. Multiple sequence alignment of the 3’ end of CPn0010-0010.1(CWL029), CPn0010 (AR 39, J138) and CPn1054 (CWL029) were constructed using CLUSTALW analysis. In CWL029, CPn0010-0010.1 is similar to CPn1054 but not identical. The difference between CP0010 and CP1054 is found at the 3’ end of the gene. In contrast, in J138 and AR39, CPn0010 is converted to be identical to CPn1054 explained by gene conversion.
Figure 23.
Figure 24. Gene conversion of the CPn0010 in *C. pneumoniae* isolates. The 3’ end of the CPn0010 identified from 12 *C. pneumoniae* isolates were aligned and compared with that of CWL029. Two of 12 *C. pneumoniae* isolates, AR 39 and AC 43, CPn0010 are converted to be identical to CPn1054.
Figure 24.
Discussion and Summary

A major area of interest in the chlamydial biology is to understand the interaction between the chlamydial inclusion and the host intracellular environment. The role of chlamydial proteins in the inclusion membrane in the development cycle and pathogenesis remain to be elucidated. However, the lack of genetic manipulation techniques in the chlamydiae is the primary limiting factor to blocking characterization of unknown functional proteins, including inclusion membrane proteins. A new approach is to analyze the microbial genome and determine the function of proteins predicted from the sequence. This includes similarity searching, structural motif analysis, and comparative genomics. Attributes of function and structure are commonly inherited in a protein family.

The structural motif of putative chlamydiae Inc proteins

A protein family is commonly defined based on high sequence similarity. However, a protein family can also be classified using a common structural motif as a signature. The motif, small sequence units of protein families, is identified as a highly similar region in alignments of protein sequence. Motifs are widely used to predict functional regions of proteins (Henikoff et al., 1997). Motifs may reflect either common ancestry or convergence from independent origins. Identification of a signature motif can be important for the exploration of structural and functional inferences within unknown genes or gene products.
The putative Inc family is mainly classified by a unique structural motif rather than sequence similarity. Although they share less sequence similarity between each other, they retain a common structural motif. This study demonstrated that the structural motif, a bi-lobed hydrophobic domain in the secondary structure of Inc proteins, is a motif for targeting protein to the inclusion. The mechanisms exploited by chlamydiae to target the Inc proteins to the inclusion are unknown. It appears that the bi-lobed hydrophobic motif is not the only structural domain required for protein localization to the inclusion membrane. Recently, some chlamydial proteins lacking the signature hydrophobicity domain have also been identified on the inclusion membrane. These include Cap1, a protective antigen recognized by cytotoxic T cells (Fling et al., 2001) and CopN, a predicted target of the type III secretion apparatus (Fields et al., 2000). It is suggested that there are other, uncharacterized, structural motifs required for protein localization to inclusion membrane.

Genomic analysis is a valuable approach for chlamydial researchers. Two C. trachomatis and three C. pneumoniae genomes have been completely sequenced, yielding many otherwise uninvestigated avenues of research. Gene families and novel functional genes were also identified, including the polymorphic membrane proteins family (Pmp) (Grimwood et al., 1999) and a chlamydiae protein that interacts with an apoptotic death receptor (Stenner-Liewen et al., 2002). The Pmp family was identified using the structural domain containing the common signature tetrapetptide repeat, GGAI/L/V followed by the FXXN motif.
Recently, the predicted function of the Pmp family has also been determined using a structural motif by Henderson and Lam (2001). They showed that the Pmp family resembles other gram-negative bacterial proteins secreted via an autotransporter system. The Pmp proteins harbor the structural motif for autotransporters including a predicted signal sequence, a passenger domain \{GG [A/I/V/I] [I/L/V/Y] and FXXN\}, and a $\beta$-barrel structure at the carboxyl terminal region within the outer membrane.

Recently, Stenner-Liewen et al (2002) has identified a hypothetical chlamydial protein that shared significant identity and predicted structural homology to the death domains (DDs) of members of the mammalian TNF-receptor family. The chlamydial protein, which has homology to human DDs of human DR4 and DR5 are referred to as Chlamydia protein Associating with Death Domains (CADD). In vitro experiment showed that CADD is capable of inducing apoptosis and interacting with Fas domain.

**Gene homology and Gene order**

Gene arrangement of putative *inc*(s) is mostly dispersed but some are clustered. Clusters of *inc*(s) in *C. trachomatis* were tightly linked to both the *incA* and the *incB/incC* homologs, but no such gene arrangement was seen in *C. pneumoniae* genome. These findings suggest that gene order of *inc* group is not generally conserved among the *Chlamydia*. 
There are some examples of conserved gene order and gene cluster. The cluster of \textit{incB/incC} and genes encoding amino acid-and sodium-dependent transporter gene are conserved in not only \textit{C. trachomatis} and \textit{C. psittaci} (Bannantine \textit{et al.}, 1998) but also \textit{C. pneumoniae}. Typically, such a conserved gene cluster conveys functional coupling between the genes product present in them (Overbeek \textit{et al.}, 1999). Whether functional significance is related to gene arrangement on the microbial chromosome remains to be elucidated. However, these analyses demonstrated that the family of putative Inc protein is randomly expanded in both \textit{C. trachomatis} and \textit{C. pneumoniae}.

Genomic comparison revealed that about 80\% of all predicted coding sequences in \textit{C. trachomatis} and \textit{C. pneumoniae} are orthologous. The average of identity between orthologs is 62\% at the amino acid level (Kalman \textit{et al.}, 1999). Orthologs are defined as related genes in different genomes that have been created by the splitting of a taxonomic lineage, while paralogs are related genes in the same genome created by gene duplication events (Fitch, 1970). In putative \textit{inc} analysis, 20 of 60 ORFs were defined as orthologs. The range of identity is about 20-70\% for the \textit{inc} orthologs. The presence of both orthologs and paralogs in putative Inc(s) suggested that some Inc proteins are conserved and required to function in biological development in both \textit{C. trachomatis} and \textit{C. pneumoniae}, while other Inc groups are species specific determinants that play a role in specific functions.
Possible functions of the putative Inc protein

Evidence of a large number of putative Inc proteins suggested that possibly, the chlamydiae have an extraordinary repertoire of unique proteins in contact with the cytosol of infected host cells although their functions remain uncharacterized. They may be important in some aspects of communication with the host cell cytoplasmic environment. This may include aspects of intracellular vesicular trafficking, establishment or maintenance of the inclusion, and/or nutrient acquisition.

The structural motif of a bi-lobed hydrophobic domain in all Inc proteins leads to the speculation that one lobe of the motif may be embedded in the chlamydial cell wall whereas the other may be in the inclusion membrane (Hackstadt et al., 1999). These proteins may then serve to anchor the pathogen to the inclusion, facilitating direct contact between the RB and the inclusion as well as providing molecular contact between the RB and the cytosol. Microinjection studies demonstrated that three *C. trachomatis* Inc proteins, including IncA, IncD and IncG, are cytosol-exposed at the carboxyl terminus of the protein (Hackstadt et al., 1999).

Non-fusogenic variant of *C. trachomatis*

*Chlamydia trachomatis* typically produces a single large inclusion during development. Infection of an epithelial cell with the multiplicity of infection greater than one of *C. trachomatis* results in multiple inclusions within the cell, which
eventually fuse to form a single large inclusion. Previous studies conducted by Matsumoto et al. (1991) and van Ooij et al. (1998) suggested that the fusion of *C. trachomatis* inclusion is a relatively late event in the developmental cycle. These authors showed that the fusion of the *C. trachomatis* inclusion did not appear to be initiated before 10 hpi. The developing inclusion was not uniformly completed until between 18 and 24 hpi. Additionally, *C. trachomatis* can form atypical nonfusogenic inclusions in low temperature culture (van Ooij et al., 1998). The mechanism of homotypic fusion of *C. trachomatis*, however, remains to be characterized.

In this detailed study, we showed that the non-fusogenic variants were apparently associated with the absence of the IncA in the nonfusogenic inclusion membrane. Previously, Suchland et al. (2000) have identified the variant isolates with the nonfusogenic inclusion phenotype that was cell line- and temperature-independent in the clinical studies. They demonstrated that the variants are stable upon repeated passages, and lack IncA on the inclusion. In our subsequent incA sequence analyses, we demonstrated diversity in incA within the nonfusogenic clinical *C. trachomatis* isolates. A variety of mutations of incA were identified, including nucleotide substitutions that directly introduce stop codons, single base frameshift, and large deletions. In most cases (24/26 IncA-negative isolates), undetectable IncA in the inclusion membrane was correlated to the interruption of the proper reading frame. In the remaining two isolates, the reason for the absence of detectable IncA remains unclear.
The role of IncA in homotypic fusion of *C. trachomatis* inclusion was supported by electron microscopy and microinjection studies of Hackstadt *et al.* (1999). They demonstrated that the fusion of *C. trachomatis* inclusions can be detected by electron microscopy by about 10 hpi. This is consistent with the temporal expression of *C. trachomatis* IncA and detection of IncA on the developing inclusion. Microinjection of antibody specific to IncA can block the fusion of inclusions. Through two-hybrid system analysis, they also showed that the IncA can interact with the homologous IncA via the carboxyl-terminus of protein. These data suggested the model in which the interactions between *C. trachomatis* IncA monomer may modulate inclusion structure, which possibly facilitate the association of opposing membrane faces to favor membrane fusion.

In this study, sequence analysis lead to the detection of *incA* and CT223 variant isolates in the clinical samples. Three of 13 (23%) wild type and one of nonfusogenic variant isolates, D(s)2923, possess IncA variants with the substitution of isoleucine with threonine at codon 47 and glutamic acid with lysine at codon 116 (IncA 147T, E16K) while 5 of 13 (38%) wild type isolates encoded prototype IncA presented in the *C. trachomatis* serovar D genome. These results were similar to the studies of clinical *C. trachomatis* isolates in Netherlands reported recently by Pannekoek *et al.* (2001). They demonstrated that 9 of 25 (36%) wild type isolates of clinical *C. trachomatis* encoded prototype *incA* and 7 of 25 (28%) were IncA variants (IncA 147T and E116K). However, there is no association with the
nonfusogenic inclusion phenotype of *C. trachomatis* and variant IncA (I47T, E116K).

Though the non-fusogenic D(s)2923 isolate encodes an intact *incA* sequence, expression remain unclear. Sequence analysis of the 300 nucleotide upstream of *incA* in the isolates {D(s)2923, J(s)893, and J(s)6462} encoding apparently intact polypeptides were identical to that of the serovar D genomic sequence. Analysis of upstream region of *incA* for selected wild type and variant isolates showed no variation. There are, therefore, no differences in the putative regulatory region to lead to the absence of IncA in these isolates that have apparently intact coding sequence. Therefore, the mechanism used by these isolates to block the synthesis or accumulation of the IncA remains to be characterized.

In contrast to *incA*, there is considerable variation in the serovar D and L2 CT223 sequence. While *incA* varies at only a few nucleotides, CT223 varies at 34 nucleotides leading to 26 differences in amino acid sequence. We also identified the variation of CT223 sequence which contains identical two-amino acid substitution in one wild type (J3464) and two nonfusogenic variant isolates {J(s)1980; J(s)6686}. In these CT223-negative isolates, serine changes to leucine at position 22; and glutamine changes to arginine at position 120. The CT223 sequence variant, however, was not directly associated with the absence of this protein in either immunofluorescence or immunoblot analysis. Sequence analyses of the upstream region of the CT223 for CT223-negative isolates reveal a single
nucleotide substitution of thymine for the cytosine. Combination of the evidence of variant sequence within coding sequence and the putative regulatory region of CT223p will lead to future studies that will focus on gene expression at the transcription level. Serine is a candidate target for phosphorylation site in signal transduction. Therefore, the effect of changes in serine to leucine at position 22 and glutamine to arginine at position 120 on gene expression and protein product remains to be characterized. The absence of CT223 in the inclusion membrane may be associated with the requirement of the posttranslational modification or another uncharacterized mechanism.

These findings have also allowed a phenotypic comparison study of strains that produce IncA or do not produce IncA, which may lead to differences in the biology of chlamydial infection and disease. Giesler et al. (2001) demonstrated that the difference patterns of disease are associated with the non-fusogenic phenotypes. The non-fusogenic phenotype might offer selective advantage to the variant strains to cause persistent infections.

The CPn1054 gene family and genetic variation

Comparative genomics analysis is an invaluable approach for the investigations of novel gene families, species or strain specific pathogenesis, and genetic diversity in the population. Recently, genomic comparison studies demonstrated that the overall genome organization of C. pneumoniae is highly conserved among various isolates. From detailed pairwise comparison of both
genome and proteome sequences, *C. pneumoniae* genome AR39, and J183 are greater than 99.9% identical to CWL029 (Read et al., 2000; Shirai et al., 2000). Although genomic organization and gene order of these independent strains are very similar, there is evidence for genetic variation, which may contribute to strain-specific phenotypes.

Relatively little is known about molecular pathogenesis, genetic diversity and the adaptive strategy of *C. pneumoniae*. A key adaptation strategy used in several microbial pathogens is to change biological characteristics to gain "fitness". A repertoire of variations is required for pathogens to generate the advantageous variants or heterogeneous progeny within an isolate. Variability enables a single bacterial clone to adapt without the acquisition of additional genes. Such phenotypic variation can be achieved by regulating gene expression or through genetic change mechanisms. Point mutations, homologous recombination, and the action of transposon-like elements are key mechanisms by which gene flexibility is achieved. Those mobile elements enable the DNA rearrangement and mutations, which contribute to the new variants in the population. Except the bacteriophage present in *C. pneumoniae* AR39, there are no functional inserted sequence (IS) elements, transposons, retrotrans or prophages in the whole genome (Read et al. 2000). This unusual feature suggests a novel mechanism for genetic variation in *C. pneumoniae*.

In genomic analysis, it is remarkably clear that coding regions of the genome are organized hierarchically into gene families and superfamilies. A group
of genes in which all of the members have > 50% pairwise amino acid similarity is defined as a gene family and an alignable groups of genes with similarity below this threshold is a superfamily (Gruar and Li, 1999; Thronton and DeSalle, 2000). In this study, we have identified a new cluster genes designated as the CPn1054 gene family based on the single quantitative measure of pairwise similarity. The significant similarity of these genes suggested that they are paralogous genes that can be classified in a new gene family. Our previous study has identified some gene products of these genes as putative inclusion membrane proteins (Incs). Seven genes in the gene family, CPn0007, 0008, 0009, 0010, 0010.1, 0011 and 0012 are encoded between the polymorphic membrane protein genes, *pmp1* and *pmp2*, which encode cell surface proteins of *C. pneumoniae*. Members of the CPn1054 gene family were categorized as paralogs, which apparently arose through gene duplication, followed by diversification. Their protein products may thus serve similar but not identical functions. The unusual start codon GTG was identified in the CPn0043, CPn0045 and CPn1055. The members of this family appeared to be variably interrupted by frameshifts. These may be mediated by gene fusion or gene fission, which involves the combination of two genes into one or the spitting of one gene into two proteins. Such events could occur by deletion or insertion mutation leading to removal of a start codon. These analyses also suggested that there were substantial 5’ and 3’ conservation of these paralogs, with the greatest variation in the midregion. These genes are found only in *C. pneumoniae* and have yet to be functionally characterized. They, therefore, are determined as unknown,
hypothetical, and *C. pneumoniae* specific genes, which may be required for specific attributes.

Comparative analysis within and among the complete genome sequence has provided evidence that genetic variation within this gene family may be modulated through multiple mechanisms. Several paralogous genes, including CPn0008, 0010, 0043, 1054, and 1055 in the CPn1054 gene family, contain poly C repeats either upstream or at the predicted 5’end of the coding region. The length of the poly C stretches in the CPn1054 gene family varied between paralogous genes in each genome, and within single genes in the three genomes. The length of the poly C tract also varied between different isolates and within an isolate, which suggested the presence of the heterogeneous population in each *C. pneumoniae* isolate. The variation of the length of this repeat appeared to be involved in the determination of the predictive start site of these genes. The variation in the length of the poly C tract moved upstream-located ATG/GTG codon in or out of frame, which may affect either transcription or translation. The repertoire of variants in the family and the presence of variability of the length of the short sequence repeat suggests that this family may also play a possible role in either functional or antigenic variation, which can contribute to the genetic diversity or unknown virulence traits of *C. pneumoniae*.

Short sequence repeats such as homopolymeric nucleotides, which usually are identified in contingency or hypermutable loci, are involved in a molecular mechanism of genetic variation in numerous pathogenic bacteria (Brian *et al.* 1992;
Moxon et al. 1998; Deitsch et al. 1997). The presence of contingency loci provides a repertoire of variation, and therefore genomic flexibility for the bacterium. In several pathogens, the deletion or insertion of a homopolymeric repeat can alter the level of gene expression or the translational frame of gene encoding cell surface molecules, thereby leading to the new variants in the population. In C. pneumoniae, variation of the short repeat of homopolymeric nucleotides was first identified in the Pmp family (Christen et al., 1999; Stephens and Lammel, 2001). Comparative genomic analysis and cloning expression showed that the length of the poly G tract of pmpG 10 varies between strains and within an isolate (Pedersen et al., 2001). Furthermore, variation of the length of poly G has been demonstrated that it plays a role in the differential expression of PmpG10 (Pedersen et al., 2001).

Variation in size via recombination within the tandem repeats was identified in the Pmp family and CPn0007, a member of the CPn1054 gene family (Shirai et al. 2000 a, b). Comparative genomics analysis reveals the shorter sequence of PmpG 6 in AR 39 and J138. In CWL.029, PmpG 6 contains three tandem repeats of 131 amino acids, whereas that of AR39 and J138 contains only two repeats (Grimwood et al., 2001; Shirai et al., 2000a, b). Like PmpG6, the size variation of CPn0007 is caused by the deletion of a tandem repeat. In J138 genome, CPn0007 encodes a truncated sequence with a gene product that lacks 110 amino acid residues in the middle region (Shirai et al., 2000a) leading to the absence of the second tandem repeat in this protein. This variation mechanism may play a function in genetic variation in both families. We have further studied the variation of
CPn0007 in 12 clinical isolates. However, there was no evidence for insertion or deletion of the repeat motif of CPn0007 in any isolates.

In the CPn1054 gene family, each gene is derived from gene duplication. About 50% of all gene duplication will lead to functional divergence (Wagner et al., 1998). Each duplicated gene can elicit a differential expression pattern. They might also be selected for subfunctionalization (Massingham et al., 2001). On the other hand, not every gene duplication result in a new function. Some duplicate genes have lost function resulting from mutation and become pseudogenes (Wagner, 1998). Comparative genomic analysis revealed frameshift mutations and nucleotide substitution in CPn0010, 0045, and 1054, which resulted in prematurely terminated gene products in the genome. These genes might be unfunctional or pseudogenes that were transitionally inert in some isolates.

Recently, in the comparative analysis of C. pneumoniae strain CWL029 and AR 39 Jordan et al. (2001) identified and described an independent conversion between paralogs Cp0764 (CPn 10540) and Cp0797 (CPn0010-0010.1) in AR-39 but not CWL-029. The nucleotide sequence of CP764 and CP797 are 100% identical except for one deletion. In contrast, there is no gene conversion in CWL029 strain. It encodes the two paralogs, CPn0010 and CPn1054, which lack 100% identity, but contains the 28 nucleotide difference at the 3' end of the coding region. Gene conversion is an intragenomic, nonreciprocal recombination event that results in identical (homogenized) gene sequence. It is likely that gene conversion between related genes or paralogs may result in antigenic variation.
Comparative genomics reveal the genetic variation within this family that may be mediated by gene conversion between paralogs (CPn0010 and CPn1054) (Jordan et al., 2001). Apparently, there are two copies of either CPn0010 or CPn1054 in the genome. In CWL029, CPn0010, is very similar but not identical to CPn1054. The significant difference between these two paralogs is present at the 3' end of the gene in CWL029. In contrast, CPn0010 of AR39 and J138 are identical to CPn1054 of CWL029 but different from CPn0010 of CWL029 (Figure 23). In AR 39 and J138, CPn0010 sequence is identical to the CPn1054 sequence, which can presumably be explained by gene conversion. In this present study, we have further investigated the 3' end of CPn10010 sequence in selected C. pneumonia isolates. Two of 12 isolates (AR39, AC43) were identified in which CPn0010 was converted to CPn1054. Mutations leading to the pseudogene of CPn0010 are identified in some isolates. The result also showed that gene conversion of CPn0010 of AR39 was consistent to that found in the AR39 genome data. With the difference of the 3' end sequence the two copies of CPn10 or CPn1054 in chlamydiae genome may have altered function, lose function, or gain totally new function.

Gene conversion is a mechanism contributing to genetic and, subsequently, antigenic variation of a repetitive gene family. The mechanism of gene conversion is plausibly explained by a nonreciprocal recombination of two separate chromosomes within a cell or two branches of replication intermediates. This recombination results in the excision of all or part of the nucleotide sequence of
one gene and its replacement by a replica of the nucleotide sequence from the other
gene. In pathogenic *Neisseriae*, pilin variation is caused by not only reciprocal but
also nonreciprocal recombination. The asymmetric recombination of *pilE*
(expressed loci) and *pilS* (silent loci) leads to the transfer of the variable sequence
of unexpressed genes to the expression locus (Haas *et al.*, 1986). In the Hop-family
member of *H. pylori*, the conversion mechanism generates the chimeric sequence
consisting of the divergence of 5’ region and identical 3’ region (Jordan *et al.*, 2001).

Recently, the coding region of the CPn1054 gene family has been proposed
to be the hypervariable region in the genome of *C. pneumoniae* (Daugaard *et al.*, 2001). The extensive variation of this DNA region between isolates was identified
using restriction fragment length polymorphisms analysis. The frameshift mutation
in CPn0010 resulting in gene fusion or gene fission of CWL029 and AR39
identified in these studies was consistent with those of the genome sequences. The
CPn0010 sequence of TW183, 2023 and two Finnish isolates (KA5C and KA66)
identified in these studies were consistent with those of *C. pneumoniae* isolates
studied by Daugaard *et al.* (2001). In addition, the importance of the poly C tracts
as sites of variation for *C. pneumoniae* has also been discussed. The difference in
the length of the poly C tract within a strain and between strains was identified in
CPn0008 and CPn1054 respectively. These data supported the event of intrastrain
and interstrain variation of these genes.
The analysis presented here supports multiple mechanisms including; slipped strand mispairing within the repeats, frameshift mutations, homologous recombination and gene conversion for antigenic variation, and adaptive evolution in *C. pneumoniae* biology. Although the proteins in the CPn1054 gene family have previously been classified as putative inclusion membrane proteins, their subcellular locations and functions remain to be identified. It is also not yet known whether the members of CPn1054 gene family are expressed individually or coordinately, or to what extent each gene is expressed during the course of an infection. However, the variation, both within and between strains, is a potential requisite for this gene family that may contribute to the unique biology of *C. pneumoniae*.

**Significance and summary of the research**

These studies have used genomic analysis as an approach for identification of uncharacterized gene clusters and variants from the Chlamydia genome projects. The information derived from the genome sequences has provided an approach to identify putative Inc proteins, a novel gene family and genetic variation.

A unique structural motif of a bi-lobed hydrophobic domain was shown to be a common character for chlamydial inclusion membrane proteins Incs. Virtually all Inc proteins lack significant similarity to known proteins within the database. A large group of putative Inc proteins encoded in the *C. trachomatis* and
C. pneumoniae genomes can be defined in two groups: orthologs and paralogs. These findings suggested that some Inc proteins are conserved and required for fundamental development and growth and some are species-specific determinants that may have specific attributes.

Chlamydiae develop and multiply in non-acidified vacuoles, which can be thought of as safe niches to separate them from the hostile intracellular environment (Sinai and Joiner, 1997). The mechanism of chlamydial inclusion development, however, remains unknown. Without the available genetic manipulation technique to conduct variant strains in chlamydiae research, a discovery of clinical collection of C. trachomatis variants with the multiple lobed inclusions become a research project of interest. Studies of these variants may elucidate a mechanism of the inclusion development and new insight of chlamydial biology. In our study, most of these non-fusogenic variants uniformly lacked IncA, a chlamydial Inc protein, on the inclusion membrane. Sequence analyses of incA with a variety of mutations are consistent with the absence of IncA in most of non-fusogenic variants. Our findings suggested that IncA is a chlamydial protein involving in homotypic vesicle fusion. However, a few nonfusogenic variants have IncA localized on the inclusion membrane. Therefore, in some isolates, additional, yet unidentified factor, also play a role in the efficiency or rate of fusion of C. trachomatis inclusion. Additionally, the absence of IncA and CT223p in the inclusion suggested that these proteins may be not critically required for pathogenesis and survival in the eukaryotic environment. Sequence analyses of
incA and CT223 from clinical isolates revealed the first evidence of genetic variation of C. trachomatis Inc proteins.

Analysis of the three recently published C. pneumoniae genomes has led to the identification of a new gene family named the CPn1054 gene family which consists of 19 predicted genes and gene fragments. In this family, paralogous genes share not only high sequence similarity but also common structural motifs. They contain a bi-lobed hydrophobic domain in the secondary structure, which is a common signature for protein localization in the inclusion. This family is a gene cluster present in only C. pneumoniae. The CPn1054 gene family, therefore, is defined as a species-specific cluster that will provide significantly biological attributes, particularly genetic and phenotypic variation in C. pneumoniae.

Comparative analysis of this gene family within and among the published genome sequences has provided evidence that gene variation through multiple mechanisms might occur within this single collection of paralogous genes. Frameshift mutations are found that result in both truncated gene products and pseudogenes that vary both within and among the different genomes. Variation via recombination of tandem repeats is also observed in a single distantly related member of the CPn1054 gene family. Finally, several genes in this family contain poly C tracts either upstream or within the predicted 5' end of the coding region. The length of the poly C tracts varies between paralogous genes in each genome, and within single genes in the three genomes. Sequence analysis of genomic DNA from a collection of 12 C. pneumoniae clinical isolates was used to analyze the
extent of the variation in the CPn1054 gene family. These studies demonstrated that frameshifts, gene conversions, and changes in the length of the poly C sequences are evident both between strains and within strains at several of the different loci. In particular, changes in the length of the poly C tract associated with different the CPn1054 gene family members are common in each tested C. pneumoniae isolate. Collectively, the variability identified within this newly described gene family may modulate either phase or antigenic variation and subsequent physiologic diversity within a C. pneumoniae population.

These studies have opened new questions of whether variation of inclusion membrane proteins may involve in pathogenesis and genetic diversity. These remain challenges and much further investigation will be required to uncover the chlamydial infection and host-pathogens interaction.
BIBLIOGRAPHY


