

## MINIREVIEW

### Genome Sequencing and Our Understanding of Chlamydiae†

DANIEL D. ROCKEY,<sup>1\*</sup> JENNIFER LENART,<sup>1</sup> AND RICHARD S. STEPHENS<sup>2</sup>

*Department of Microbiology, Oregon State University, Corvallis, Oregon 97331-3804,<sup>1</sup> and  
Program in Infectious Diseases, School of Public Health, University of California—Berkeley,  
Berkeley, California 94720-7360<sup>2</sup>*

A driving force in the evolution of a microorganism is the ability to colonize a niche. A vertebrate organism represents a unique niche, but to an infecting microbe it is simply an environment to be exploited. Many pathogens explore yet another opportunity—the intracellular environment. The chlamydiae are obligate intracellular pathogens that have established a unique niche within the host cell. As a group their approach has been successful; chlamydiae cause a variety of diseases in animal species at virtually every phylogenetic level. *Chlamydia trachomatis* and *C. pneumoniae* are the most significant human pathogens. *C. trachomatis* is the leading cause of preventable blindness worldwide and the most common sexually transmitted bacterial species (43a, 45). *C. pneumoniae* causes 10% to 20% of community-acquired pneumonia worldwide and has recently been associated with coronary atherosclerosis (29) and possibly other conditions (18). The major determinants of chlamydial pathogenesis are complicated and unclear at present. Apparently, the chlamydiae are not acutely toxigenic, and in many cases infection is asymptomatic and persistent. Serious chlamydial disease is a function of several processes including the ability to attach and grow within target cells, mechanisms of dissemination and persistence, directed remodeling of the intracellular environment, and the nature of the immune response to infection (6).

The chlamydiae undergo a developmental cycle unique among prokaryotes. Following uptake, chlamydiae develop and grow within an intracellular vacuole, called an inclusion, that is distinct from all identified parasitophorous vacuoles. The inclusion membrane is devoid of host cell markers, but lipid markers traffic to the inclusion and suggest a functional interaction with the Golgi apparatus (16). Within the inclusion, chlamydiae undergo a complex developmental cycle composed of functional and structurally distinct forms. The elementary body (EB) is infectious but is metabolically inactive and cannot replicate. This form differentiates upon infection into the non-infectious reticulate body (RB), a larger pleomorphic bacterium that is metabolically active and multiplies. Chlamydiae produce a group-specific lipopolysaccharide (LPS) determinant that is conserved within the genus (7). All chlamydiae encode an abundant protein termed the major outer membrane protein (MOMP, or OmpA) that is surface exposed in *C. psittaci* and *C. trachomatis* (9, 20) and is the major determinant for serologic classification of chlamydial isolates (47). Disulfide-mediated cross-linking of two major developmentally reg-

ulated cysteine-rich proteins results in a rigid lattice, which is apparently necessary for structural integrity (19), as the EB lacks detectable peptidoglycan (PG). Many of the genes involved in these processes were cloned and identified using conventional means: by homology, via expression libraries, and through functional complementation analysis. But in this system still lacking a generally available transformation technique, there are many aspects of the chlamydial infectious process and developmental cycle that remain unclear.

Completion of the sequences of six chlamydial genomes (Table 1), with others soon to follow, has complemented and expanded much of the information acquired in the pregenomic era (46). The primary descriptions of these genome sequences (25, 46) provide thorough and valuable analyses of the different genomes, including much information not covered here. The purpose of this review is to examine the role of the solved genome sequences in our understanding of four model groups of chlamydial proteins. Two of the protein groups, inclusion membrane (Inc) proteins and the polymorphic membrane proteins (Pmp proteins), are, at present, unique to the chlamydiae. Two other groups, the proteins involved in PG synthesis and the chlamydial type III secretion apparatus, represent processes that are present in other organisms. While the four groups of proteins are distinct and unrelated, each represents one aspect of the unique nature of chlamydiae. We address how decoding the genome has expanded our early perspective of these organisms and also point out in each case what questions the genome sequences cannot directly answer.

#### Inc PROTEINS

Proteins localized to the chlamydial inclusion membrane were initially identified in *C. psittaci*, during studies designed to examine differences in serologic responses between infected animals and those immunized with formalin-fixed EB (43). Three gene products were identified in this fashion, IncA, IncB, and IncC (2, 42). Antisera against each Inc protein specifically labels the chlamydial inclusion membrane. Scidmore-Carlson et al. (44) have recently used antisera directed at a membrane fraction from *C. trachomatis*-infected cells to identify a chromosomal fragment that encodes four novel *C. trachomatis* Inc proteins (IncD to IncG). As a group these distinct proteins do not share significant identity with one another or with any genes within the sequence databases. However, each possesses a unique bilobed hydrophobic domain of 50 to 80 amino acids (Fig. 1). This profile is striking, particularly in light of the absence of any significant conservation of primary amino acid sequence. The function of these proteins within the inclusion membrane remains unclear.

The completion of the *C. trachomatis* and *C. pneumoniae* genome sequences allowed a unique approach to the identifi-

\* Corresponding author. Mailing address: Department of Microbiology, Oregon State University, Corvallis, OR 97331-3804. Phone: (541) 737-1848. Fax: (541) 737-0496. E-mail: rockeyd@ucs.orst.edu.

† Oregon State University Agricultural Experiment Station technical paper no. 11639.

TABLE 1. Basic characteristics of sequenced chlamydial genomes<sup>a</sup>

Organism	Strain	Genome size (bp)	Predicted no. of coding sequences	Accession no.	URL address	Reference
<i>C. trachomatis</i>						
Serovar D	D/UW-3/Cx	1,042,519	894	AE001273	http://chlamydia-www.berkeley.edu:4231/	46
Serovar L2	L2/434/Bu	NA <sup>b</sup>	NA	NA	http://chlamydia-www.berkeley.edu:4231/	NA
Biovar MoPn	Nigg	1,069,412	924	AE002160	http://www.tigr.org/tdb/CMR/btc/htmls/SplashPage.html	40
<i>C. pneumoniae</i>						
	CWL029	1,230,230	1,052	AE001363	http://chlamydia-www.berkeley.edu:4231/	25
	AR39	1,229,853	1,052	AE002161	http://www.tigr.org/tdb/CMR/bcp/htmls/SplashPage.html	40
<i>C. psittaci</i>						
	GPIC	NA	NA	NA	http://www.tigr.org/tdb/mdb/mdb.html	NA

<sup>a</sup> The gene designations used within this minireview for unnamed ORFs and their predicted proteins are consistent with those used in the Chlamydia Genome Project website (<http://chlamydia-www.berkeley.edu:4231/>) for *C. trachomatis* serovar D and *C. pneumoniae* CWL029.

<sup>b</sup> NA, not available.

cation of additional Inc proteins. First, the sequences were quickly used to locate, by similarity, open reading frames (ORFs) encoding IncA, IncB, and IncC. After their identification, antisera were produced against the predicted *C. trachomatis* and *C. pneumoniae* IncA proteins, and these antisera were used to demonstrate that IncA from each species is localized to the inclusion membrane (1, 3). These data also demonstrated that Inc proteins are not unique to *C. psittaci*.

Further searching for novel Inc proteins was complicated by their general lack of primary sequence identity but facilitated by the unique bilobed hydrophobic domain. Bannantine et al. (1) used this motif as a marker for the identification of candidate Inc (C-Inc) proteins in *C. trachomatis*. The entire *C.*

*trachomatis* genome was searched for ORFs encoding proteins with this secondary structural pattern. Surprisingly, over 40 candidate Inc sequences that contained a similar profile were identified. This finding was in contrast to a parallel screen of the *Mycobacterium tuberculosis* genome, in which none of 587 unidentified ORFs encoded a protein with the characteristic pattern. While each candidate Inc protein shares the hydrophobic domain, as a group they do not share sequence identity nor do they have apparent orthologs in the sequence databases, outside of the chlamydiae. Within *C. trachomatis* there exists some evidence of gene duplication (ORFs CT228 and CT229), but in general the rule is lack of identity between different Incs, even those clustered tightly in the genome. In

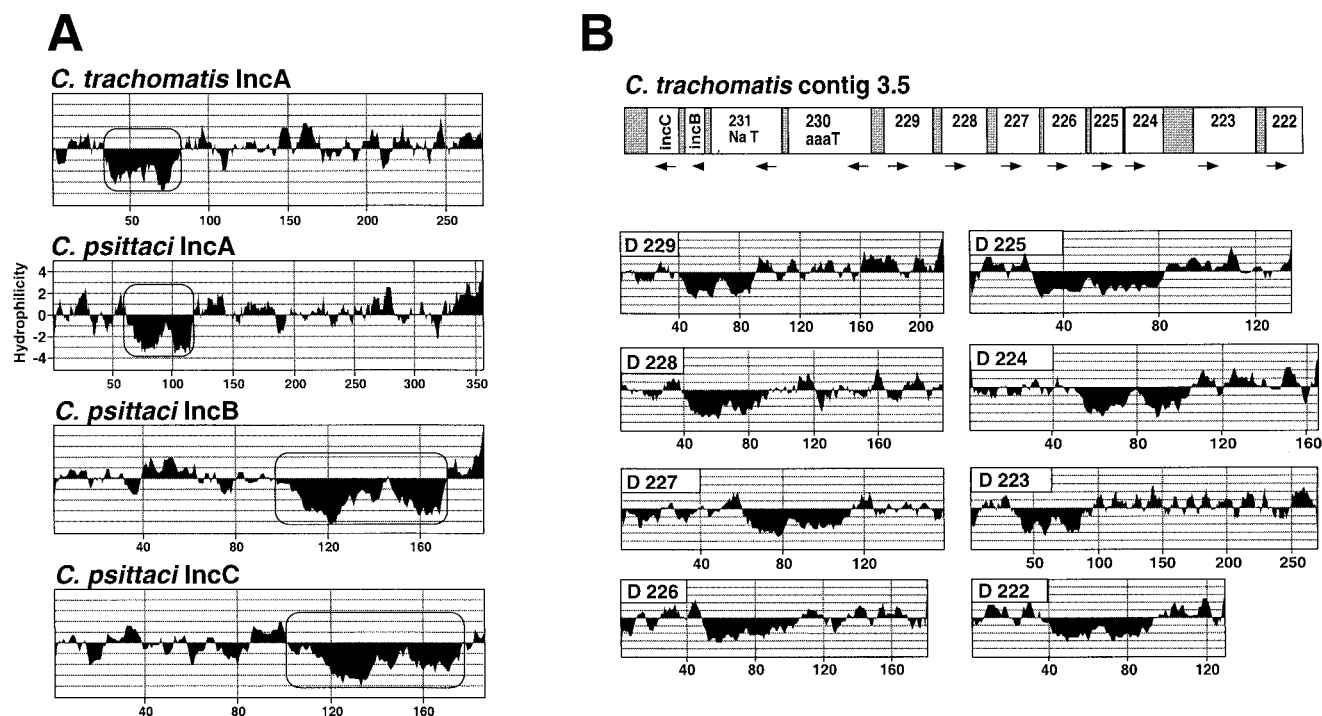


FIG. 1. Analysis of predicted hydrophilicity profiles for several known and candidate Inc proteins. In each profile the horizontal axis shows amino acid position and the y axis is relative hydrophilicity. These profiles were produced using a Kyte-Doolittle algorithm with a window of 7 (30). Scale for hydrophilicity is shown for *C. psittaci* IncA (panel A) and applies to all profiles. (A) Predicted hydrophilicity of four chlamydial Inc proteins. Note that in each there is a large bilobed hydrophobic domain (boxed regions). This domain is found in each known Inc protein. (B) ORFs present on contig 3.5 of the *C. trachomatis* genome. Within this contig are 10 genes encoding candidate Inc proteins. Three of these proteins (IncC, CT223p, CT229p) have been shown to reside in the chlamydial inclusion membrane. Note that eight of these sequences (CT222 to CT229) are contiguous and may represent an operon. The hydrophilicity profiles of the eight tandem candidate Inc proteins are shown below the contig. Only two of these sequences, CT228 and CT229, encode proteins that share significant identity at the amino acid level.

some cases these clusters involve several *inc* genes. An example of this is found on contig 3.5 of the *C. trachomatis* genome, where 10 different *inc* genes are located (Fig. 1B). Eight of these (CT222 to CT229) are sequential. Two of these candidate Inc proteins (CT223p and CT229p) have been localized to the chlamydial inclusion membrane (1). These *inc* clusters are suggestive of an operon structure, and there is evidence with two different clusters that tandem *inc* genes are cotranscribed (2, 44).

Although the data from known Inc proteins suggested the hydrophobicity motif was a flag for localization to the inclusion membrane, this remained to be formally tested. This problem was approached by producing specific antibody to six purified recombinant candidate Incs selected from the *C. trachomatis* genome database. One of these genes (CT233) shared identity with *C. psittaci* IncC, but the remainder (CT223, CT229, CT288, CT442, and CT484) were chosen solely based on the presence of the hydrophobicity motif. These antisera were subsequently tested in immunofluorescence microscopy to determine if the antisera labeled the inclusion membrane within infected cells. Five of the six proteins tested were shown to localize to the inclusion membrane (1). Therefore, the presence of this characteristic hydrophobicity domain appears to be a marker for inclusion membrane localization, although this seems to not be universal. The localization of a protein to the inclusion membrane therefore needs to be confirmed experimentally for each candidate protein. The role of the hydrophobic domains in the biology of the Inc proteins remains elusive. The motif may facilitate transport of these proteins to the inclusion membrane or possibly stabilize the proteins once they are delivered there. Perhaps this domain serves as an internal chaperone for targeting through the type III secretion system, as discussed below.

The completion of the *C. pneumoniae* CWL029 genome allowed a search of this database and, surprisingly, an even larger collection of C-Incs are encoded by this species. In several cases there are clear orthologs between *C. trachomatis* and *C. pneumoniae* (25), but over half of the C-Inc sequences from each species do not share significant similarity with sequences in the other. The *C. pneumoniae* genome also provides evidence of expanded use of candidate *inc* duplication relative to that seen in *C. trachomatis*. An example of this apparent duplication is *C. trachomatis* candidate *inc* CT358 which is homologous to several closely linked ORFs in *C. pneumoniae* (CPn367, CPn369, and CPn370; contig 5.1) (1). These three ORFs encode predicted proteins that contain the characteristic hydrophobicity motif. However, one additional ORF, *C. pneumoniae* CPn107, encodes a protein that shares identity with each of these C-Incs but lacks the hydrophobic domain. Experiments are in progress to determine the location of each of these proteins within infected cells.

Functional data that complement the genomic analyses of Inc proteins are only beginning to accumulate. Currently, IncA, IncF, and IncG from *C. trachomatis* (17) and IncA from *C. psittaci* (41) are known to be exposed at the cytoplasmic surface of the inclusion. Very recent data demonstrate that IncA from *C. trachomatis* participates in homotypic vesicle fusion events between different *C. trachomatis* inclusions (17). There are a myriad of possible functions for the diverse collection of Inc proteins, such as participation in inclusion development, avoidance of lysosomal targeting, nutrient acquisition, and possibly signaling associated with EB-to-RB-to-EB transitions, to name a few. This is an emerging and challenging area of research that may provide answers to important questions regarding chlamydia-host cell vesicular interactions.

Perhaps the most significant concept that has been intro-

duced by these genomic analyses is that the inclusion membrane, an environment in which no host cell protein marker has been identified, is apparently decorated with a large array of distinct, largely unrelated, chlamydial proteins. Eleven distinct proteins have been localized to the inclusion membrane and it is likely there are more. The lack of sequence identity among the known Incs, as well as their lack of identity with known proteins in the databases, slows progress in the analysis of Inc protein function. However, it is intriguing to envision a collection of Inc proteins that are all in contact with the cytoplasm of the infected cell, each serving a unique role in the interaction between the inclusion and the host cytosol.

The diverse Inc proteins encoded by each genome also introduce questions relating to evolutionary biology. How were these unique gene products selected for localization to the inclusion membrane? What is the basis for their sequence diversity? There may be a gene duplication/mutagenesis mechanism that adds new *inc* genes within a genome or possibly a mechanism for random insertion of a (presently cryptic) signature nucleotide sequence encoding the characteristic hydrophobicity profile. Proteins that acquire this tag may subsequently be transported to the inclusion membrane. In either case, if the localization of this protein to the inclusion membrane environment offers a selective advantage, the mutant would survive and the mutation would be propagated. Unfortunately, the very low level of primary sequence identity among most C-Incs does not provide support for any model and this remains an open question.

## Pmp PROTEINS

The completed chlamydial genomes were also important for providing insight to the large family of Pmp proteins. Like the Inc proteins, these large, antigenic molecules were initially identified in *C. psittaci* using an antibody-based screening process. Two sets of two tandem genes each were identified by Longbottom et al. (33) from an expression library of an ovine abortion strain of *C. psittaci*. These genes encode proteins with sizes of 90 or 91 kDa, and the overall similarity of the predicted proteins is above 82%. Two of the predicted proteins are identical.

As the *C. trachomatis* and *C. pneumoniae* genomes were completed it became clear that the *pmp* genes comprised a complex family. *C. trachomatis* encodes nine different Pmp proteins and the family is expanded in *C. pneumoniae* to 21 candidate *pmp* genes or gene fragments (15, 46). Like the Inc proteins, the Pmp family as a whole is grouped not on the basis of primary sequence identity. In this family the common thread is a pair of signature tetrapeptide motifs repeated multiple times in each protein in the family. These tetrapeptide sequences tend to alternate and are clustered toward the amino half of the Pmp proteins. The two sequences GGA(L, V, I) and FXXN are repeated more frequently in the Pmp sequences than in the rest of the genome and, with few notable exceptions, far higher than that seen in the protein sequence databases. For example, PmpA of *C. trachomatis* is a 106-kDa protein with 6 copies of GGAI and 14 copies of FXXN. Overall, the Pmp proteins average 7 GGA(L, V, I) motifs and approximately 12 FXXN motifs. In contrast, only three other proteins have been identified in which the GGAI sequence is repeated more than three times, and in these proteins the FXXN motif is also repeated. Each of the nonchlamydial proteins that possess multiple copies of these tetrapeptide repeats are thought to be involved in adherence to mammalian tissues. As some Pmps have been localized to the chlamydial cell surface (26, 32), roles in adhesion, molecular transport, signal-

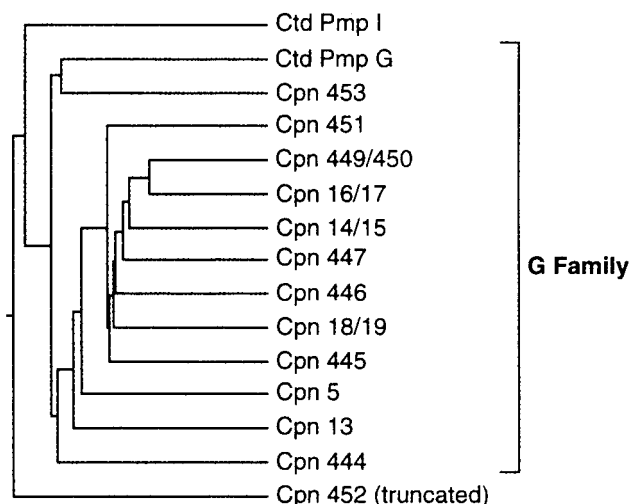


FIG. 2. Evidence for *pmp* expansion in *C. pneumoniae*. Twelve *C. pneumoniae* Pmp proteins share identity with a single *C. trachomatis* Pmp sequence, PmpG. Three of these (CPn449/450p, CPn16/17p, and CPn18/19p) are hypothetical proteins that result from elimination of internal frameshifts in the nucleotide sequence. *C. trachomatis* PmpI and *C. pneumoniae* CPn452 are included as outliers. Grimwood and Stephens constructed this tree using a CLUSTALW alignment with distances generated using Jukes-Cantor for correction (15). Six families of Pmp proteins can be identified using this method, and two of these show significant expansion in *C. pneumoniae*. This figure is reproduced with permission of the publisher (15).

ing, or some other cell wall associated function are likely. But, proposed functions for Pmps are clearly speculative and the completed genomes have opened more questions about these proteins than they have answered. As with the Inc proteins, any models of Pmp function require validation in the laboratory.

There are several general parallels between Inc and Pmp proteins. Earlier, we discussed the possibility that the inclusion membrane is laden with a collection of proteins in contact with the host cell cytoplasm. The discovery of the variety of Pmp sequences allows a similarly modified view of the chlamydial cell surface. It is likely that collections of different Pmps, and possibly other membrane proteins, are present at the cell surface, interspersed among large amounts of MOMP and embedded in LPS. In both cases a unique group of proteins has been expanded and represents a significant fraction of chlamydia-specific sequence. On a nucleotide basis, candidate *inc* genes occupy 12.3% of chlamydia-specific coding capacity in *C. trachomatis* and 18.9% percent in *C. pneumoniae*. The *pmp* genes represent 13.6% and 17.5% of the chlamydia-specific coding capacity in *C. trachomatis* and *C. pneumoniae*, respectively. These percentages are especially compelling in a genus showing a clear pattern of reductive genomic evolution, i.e., the elimination of genes that are apparently unnecessary for obligate intracellular existence (52). In both groups of genes, heterogeneity is the rule. This suggests a requirement for diversity of Inc and Pmp proteins in their respective environments. There are also examples of both *inc* and *pmp* genes that have duplicated to a much higher level in *C. pneumoniae* relative to *C. trachomatis* (Fig. 2). The significance of this is not clear. Finally, there exists the possibility, within both the *inc* and *pmp* genes, that some level of gene modification may allow different subgroups of Pmp or Inc proteins to be differentially produced by variants within a strain or species. While this concept is speculative, some supporting data have been presented. Suchland et al. (49) identified a collection of clinical isolates that lack detectable IncA, and this deficiency is associated with a

phenotypic alteration in inclusion development. Within *C. pneumoniae*, *pmp* genes represented by ORFs CPn14 to CPn19, CPn449/450, CPn452, and CPn468 to CPn470 represent possible examples of frameshift mutations or deletions within *pmp* sequences (15). Additionally, the completed sequence of *C. pneumoniae* strain AR39 has shown that a large, in-frame deletion is present in *pmp6* relative to the CWL029 strain (40). This likely changes the structure of this protein, which may, in turn, alter function.

## PG SYNTHESIS

It has been known for decades that chlamydial development is highly sensitive to beta-lactams and D-cycloserine, inhibitors that block different steps of the PG synthesis pathway in other bacteria (51, 38). It is also well established that PG or PG synthetic precursors cannot be detected in purified chlamydial EB (4, 12), although chlamydial PG may be present in RB and may have a primary role in RB biology (48). These facts constitute the chlamydial peptidoglycan paradox, and detailed, genome-based analyses of the data have been presented (10, 14). One clear result of the chlamydial genome project is that both *C. pneumoniae* and *C. trachomatis* encode proteins forming a nearly complete PG synthesis pathway (Fig. 3). A comparison of known PG synthesis genes with those apparently produced by the chlamydiae shows that the chlamydiae lack homologs to three major components (21). First, the chlamydiae do not encode penicillin-binding protein (PBP1), an enzyme that possesses both transglycosidase and transpeptidase activity. While the chlamydiae encode PBP2 and PBP3, both of which are reported to be similarly bifunctional enzymes (35), there are no in vitro data demonstrating the transglycosidase activity of any PBP2 or PBP3. As proposed by Ghuyssen and Goffin (14), the lack of PBP1 is a key distinction between chlamydial PG and other PG. Their model suggests that the cross-linking of short glycan chains is the critical missing aspect of chlamydial PG synthesis, resulting in an unusual truncated PG. *Chlamydia* also does not encode the diverse collection of small penicillin-binding proteins found in *Escherichia coli* and other bacteria, but the implication of this is not clear (4). Finally, the genomes do not encode D-amino acid racemases, but the host cell can serve as a source of low levels of D-amino acids (27). The chlamydiae have retained three amidases that function in the breakdown of PG in other systems. The retention of multiple amidases suggests that degradation of this cryptic PG may be a critical aspect of growth.

The image provided by the genome sequences is a PG synthesis pathway that is apparently operative but may produce a novel or truncated PG structure. In most bacterial systems, PG serves functions other than providing a structural sacculus in bacteria, such as development of a scaffold for division within bacterial cells (21) or as an anchor for cell surface structures within the membrane. The role of PG in chlamydiae may be associated with these processes or possibly other unique aspects of chlamydial development. Septation of RB, for example, is complicated by the lack of FtsZ in chlamydiae, and therefore standard paradigms of bacterial division likely do not apply. Su et al. (48) present data suggesting that chlamydial RB contains PG but this has not been supported by others (4, 12). Perhaps the production of trace quantities of PG, localized at the septum, is required for binary fission in this system. Brown and Rockey (8) recently identified an antigenic structure (termed SEP) that appears to localize to the chlamydial division septum and whose distribution is altered by exposure of infected cells to beta-lactam antibiotics. This structure appears analogous to the FtsZ ring in many microorganisms, and we

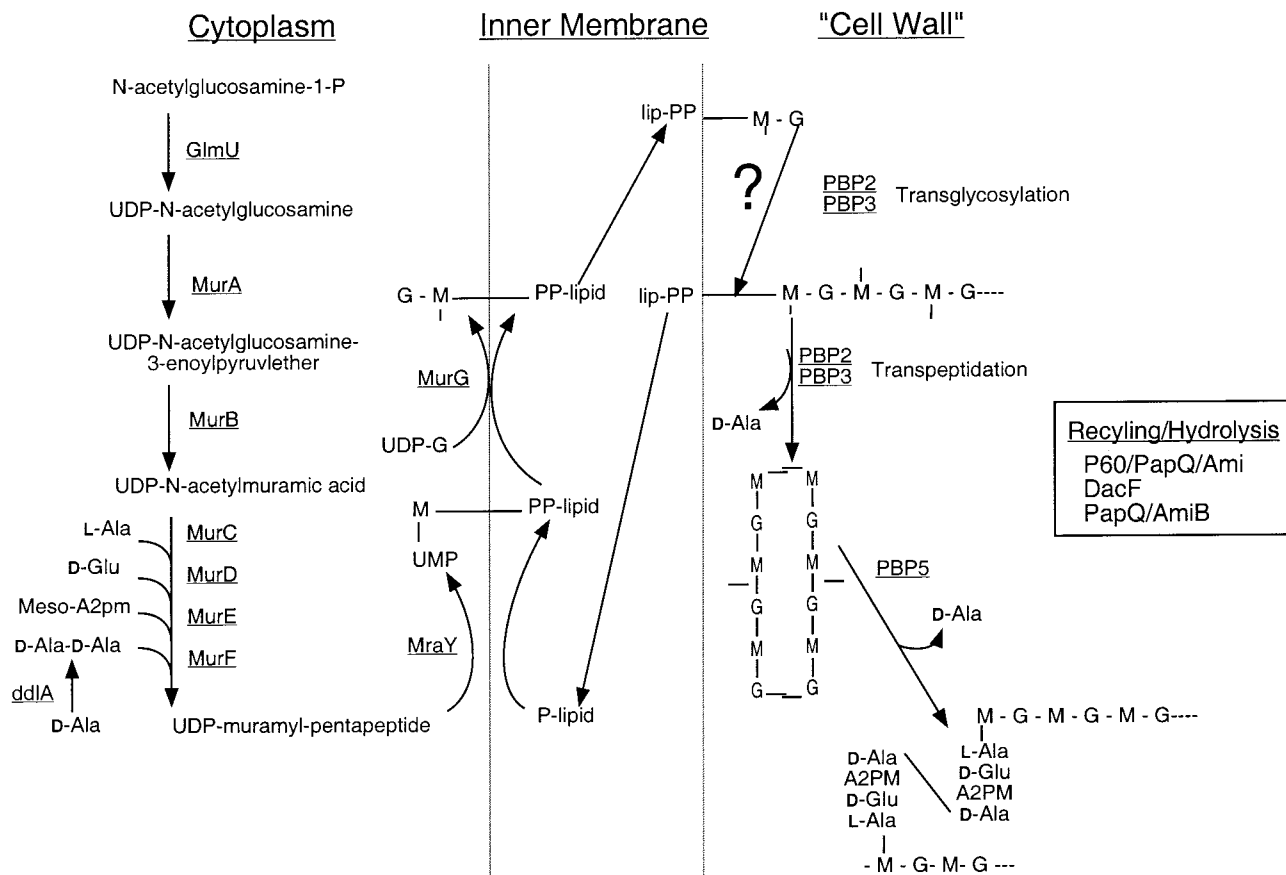


FIG. 3. Enzymatic steps showing potential PG synthesis pathway in chlamydiae. The genomes of both *C. pneumoniae* and *C. trachomatis* have genes encoding each of the enzymes indicated. The major step that may be missing is the transglycosidase required to link (question mark). All structures are hypothetical and are based on other bacterial systems (22, 39). Enzymes involved in PG hydrolysis are listed in the box. Abbreviations: P, phosphate; UDP, uridine diphosphate; A2PM, diaminopimelic acid; G, N-acetylglucosamine; M, N-acetylmuramic acid; PBP, penicillin-binding protein.

are currently examining the possibility that PG is localized to the plane of chlamydial cell division.

A final issue regarding chlamydial PG involves the functional implications of the EB cell wall which lacks PG. The structural support lost through the absence of a PG sacculus has been complemented by a unique mechanism for EB stability, as discussed above. Why does this organism require such a unique cell wall structure, even in the light of carrying an apparently functional PG synthetic pathway? Such questions point out the strengths and limitations of genome sequence analysis. Data from the genome strongly suggest that PG is produced by chlamydiae, but it is left to laboratory investigation to examine expression and function of the genes and proteins involved in the process. The structure, localization, and function of chlamydial PG also remains to elucidated.

### TYPE III PROTEIN SECRETION

The first genetic clues that chlamydiae encode proteins associated with a chlamydial type III secretion pathway were uncovered by Hsia et al. (22) and provide another example where pregenomic analysis sets the stage for expanded understanding provided by the genome sequence. These investigators identified *C. psittaci* genes that share identity with *yscU*, *lcrD*, *lcrE*, and the *ycyE* chaperone, genes encoding proteins known to participate in type III secretion systems. These chlamydial genes are arranged in an apparent operon that includes

other ORFs. In hindsight, it should also be noted that chlamydial "surface projections," morphologically similar to the type III apparatus identified in *Salmonella typhimurium* by Kubori et al. (28), were characterized by Matsumoto and others over 20 years ago (36, 37, 39). Elegant ultrastructural descriptions of these projections showed that they are found on both EB and RB, appear to localize to one side of the developmental form, and apparently pierce the inclusion membrane within infected cells. It has been speculated that these are the type III secretory apparatus produced by chlamydiae (5).

As the genome sequences became available, a more complete collection of genes encoding proteins associated with type III secretion were identified. The *C. psittaci* gene cluster discussed above was arranged similarly in both sequenced genomes. However, the genome sequence data demonstrate that many aspects of the type III secretory machinery appear different between chlamydiae and other bacteria. First, the genes are not tightly clustered in the genome, such as is seen in the *Yersinia* virulence plasmids or pathogenicity islands found in other species (11, 13). For example, the *yscU* cluster is found on *C. trachomatis* contig 2.0, and *yscC*, *yscJ*, and *yscL*, genes that are closely linked on the yersinia virulence plasmids, are well separated in the *C. trachomatis* genome (contigs 7.3 and 8.7). Chlamydiae encode an apparently minimal collection of proteins involved with the secretory machinery, as shown in the hypothetical structure shown in Fig. 4. For example, chlamydiae encode a single LcrE (YopN) protein involved in the con-

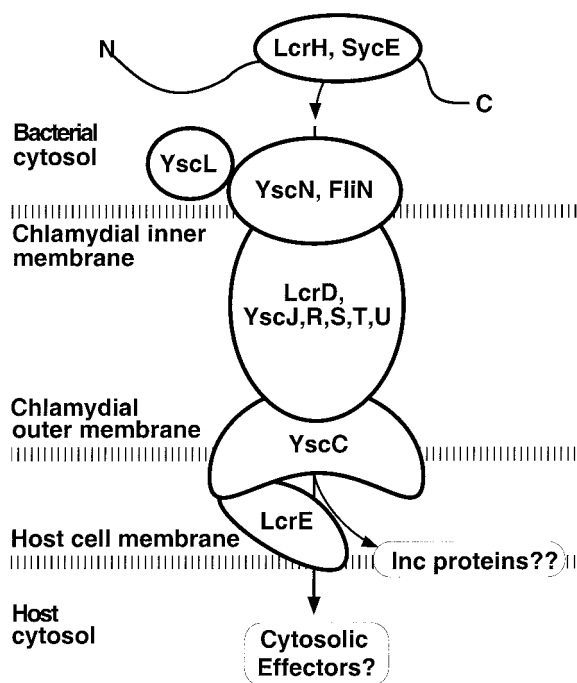


FIG. 4. Drawing of a hypothetical chlamydial type III secretion apparatus present in a simplified chlamydial cell wall. Chlamydial proteins with apparent homologs in other systems are indicated and are placed according to models derived in other gram-negative bacteria (11, 23, 31). Note that orthologs or paralogs to effector proteins have not been identified in the chlamydial system. Inc proteins are indicated as hypothetical effectors secreted and localized to the host-derived inclusion membrane.

trol of effector protein release while yersiniae encode three candidate “plug” proteins: LcrE, TyeA, and LcrG. The genome sequence did not identify apparent homologs to YopB, YopD, or YopK or other known type III proteins that are translocated by yersiniae or other bacteria to the eukaryotic plasma membrane or to any proteins that share identity with known type III effectors delivered to the eukaryotic cytosol. This may not be surprising given the general lack of similarity between effectors from different species. It is logical that Inc proteins are secreted via type III processes as they are localized to the inclusion membrane and as a group they lack classical amino-terminal secretion signals.

Therefore, the genome sequence encodes proteins that form the basic structure of a type III secretory apparatus, although several structural proteins found in other species are apparently absent. It is possible that many proteins involved in chlamydial type III secretion are unique to this system or that single proteins serve multiple functions in the type III secretory process. This will be a fascinating and important area of exploration in the future because type III effector proteins are probably critical to chlamydial interactions with the host cell.

### CONCLUSIONS

The completed chlamydial genomes have added depth to our understanding of many aspects of chlamydiae of which pregenomic analysis has given us a glimpse. Many good examples of this are not discussed in this review. Chlamydiae are auxotrophic for three of four nucleotide triphosphates, but they do encode functional glucose-catabolizing enzymes, which can be used for generation of ATP (24, 50). The general reduction in biosynthetic machinery confirms that chlamydiae

are auxotrophic for many catabolic precursors (52). Insight into chlamydial lipid and glycogen biosynthesis is provided by the genome. While chlamydiae do not encode a large variety of sigma factors as do other bacteria that undergo developmental cycles, there are distinct sigma factors that may be important in developmental genes (34). In some cases completely new avenues of investigation were opened with the genome completion. *C. trachomatis* mouse pneumonitis encodes a putative high-molecular-weight toxin that may participate in the infectious process. The sequenced genomes will allow analysis of global transcription patterns through the use of real-time PCR and DNA chip or microarray technologies, experimental approaches that will provide a fertile field for detailed examination of chlamydial biological processes.

This review addressed two sets of proteins with orthologs in nonchlamydial systems and two sets that appear to be unique to chlamydiae. The genome provides different types of information for each. With regards to PG synthesis and the type III secretion apparatus, the genome provides evidence that these processes are active within chlamydiae and significantly expands our understanding of the mechanisms involved. But, in each case, the data open as many questions as are solved. For example, there remains the need to identify type III effector proteins and to determine the role of PG synthesis in chlamydial biology.

An important aspect in the analysis of bacterial genomes involves assigning function to proteins encoded by ORFs with no known functional homologs in other species. As discussed by others, proteins unique to a species may provide the most unique traits of that species. The candidate *inc* and *pmp* genes occupy 12% to 19% of the chlamydia-specific genomic sequences, and in fact, the functions of Inc and Pmp proteins are still a completely open issue. Beyond these families of proteins, a large number of additional ORFs are completely uncharacterized. The many questions raised by the genomic data will support diverse avenues of investigation into mechanisms used by chlamydiae to occupy their unique intracellular niche and can hopefully lead to progress regarding prevention and treatment of chlamydial disease.

### ACKNOWLEDGMENTS

This work was supported by Public Health Service award R29AI42869.

We thank Wendy J. Brown and Wasna Viratyosin for technical and editorial assistance. John Bannantine (NADC, Ames, Iowa) and Robert Heinzen (University of Wyoming) provided critical reviews of the manuscript. We acknowledge Catherine A. Lee (Harvard University School of Medicine) and Kevin D. Young (University of North Dakota School of Medicine) for helpful comments.

### ADDENDUM IN PROOF

At the recent American Society for Microbiology meeting, Ken Fields and Ted Hackstadt presented data indicating that *Chlamydia trachomatis* ORF CT578 and CT579 may be homologs of *yopB* and *yopD*, respectively (K. A. Fields and T. Hackstadt, Abstr. 100th Gen. Meet. Am. Soc. Microbiol., 2000, abstr. D-193, p. 270, 2000). A comprehensive review of chlamydial diseases, phylogeny, biology, and immunology has recently been published (R. S. Stephens (ed.), *Chlamydia: Intracellular Biology, Pathogenesis, and Immunity*, 1999).

### REFERENCES

- Bannantine, J. P., R. S. Griffiths, W. Viratyosin, W. J. Brown, and D. D. Rockey. 1999. A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane. *Cell. Microbiol.* 2:35-48.
- Bannantine, J. P., D. D. Rockey, and T. Hackstadt. 1998. Tandem genes of

- Chlamydia psittaci* that encode proteins localized to the inclusion membrane. *Mol. Microbiol.* **28**:1017–1026.
3. Bannantine, J. P., W. E. Stamm, R. J. Suchland, and D. D. Rockey. 1998. *Chlamydia trachomatis* IncA is localized to the inclusion membrane and is recognized by antisera from infected humans and primates. *Infect. Immun.* **66**:6017–6021.
  4. Barbour, A. G., K.-I. Amano, T. Hackstadt, L. Perry, and H. D. Caldwell. 1982. *Chlamydia trachomatis* has penicillin-binding proteins but not detectable muramic acid. *J. Bacteriol.* **151**:420–428.
  5. Bavoil, P., and R. Hsia. 1998. Type III secretion in *Chlamydia*: a case of deja vu? *Mol. Microbiol.* **28**:860–862.
  6. Bavoil, P. M., R. C. Hsia, and R. G. Rank. 1996. Prospects for a vaccine against *Chlamydia* genital disease. I. Microbiology and pathogenesis. *Bull. Inst. Pasteur* **94**:5–54.
  7. Brade, L., O. Holst, P. Kosma, Y.-X. Zhang, H. Paulsen, R. Krausse, and H. Brade. 1990. Characterization of murine monoclonal and murine, rabbit, and human polyclonal antibodies against chlamydial lipopolysaccharide. *Infect. Immun.* **58**:205–213.
  8. Brown, W. J., and D. D. Rockey. 2000. Identification of an antigen localized to an apparent septum within dividing chlamydiae. *Infect. Immun.* **68**:708–715.
  9. Caldwell, H. D., J. Kromhout, and J. Schachter. 1981. Purification and partial characterization of the major outer membrane protein of *Chlamydia trachomatis*. *Infect. Immun.* **31**:1161–1176.
  10. Chopra, I., C. Storey, T. J. Falla, and J. H. Pearce. 1998. Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. *Microbiology* **144**:2673–2678.
  11. Cornelis, G. R., A. Boland, A. P. Boyd, C. Geuijen, M. Iriarte, C. Neyt, M. P. Sory, and I. Stainier. 1998. The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol. Mol. Biol. Rev.* **62**:1315–1352.
  12. Fox, A., J. C. Rogers, J. Gilbert, S. Morgan, C. H. Davis, S. Knight, and P. B. Wyrick. 1990. Muramic acid is not detectable in *Chlamydia psittaci* or *Chlamydia trachomatis* by gas chromatography-mass spectrometry. *Infect. Immun.* **58**:835–837.
  13. Galan, J. E. 1999. Interaction of *Salmonella* with host cells through the centisome 63 type III secretion system. *Curr. Opin. Microbiol.* **2**:46–50.
  14. Ghuysen, J. M., and C. Goffin. 1999. Lack of cell wall peptidoglycan versus penicillin sensitivity: new insights into the chlamydial anomaly. *Antimicrob. Agents Chemother.* **43**:2339–2344.
  15. Grimwood, J., and R. S. Stephens. 1999. Computational analysis of the polymorphic membrane protein superfamily of *Chlamydia trachomatis* and *Chlamydia pneumoniae*. *Microb. Comp. Genomics* **4**:187–201.
  16. Hackstadt, T., E. R. Fischer, M. A. Scidmore, D. D. Rockey, and R. A. Heinzen. 1997. Origins and functions of the chlamydial inclusion. *Trends Microbiol.* **5**:288–293.
  17. Hackstadt, T., M. A. Scidmore-Carlson, E. I. Shaw, and E. R. Fischer. 1999. The *Chlamydia trachomatis* IncA protein is required for homotypic vesicle fusion. *Cell. Microbiol.* **1**:119–130.
  18. Hahn, D. L. 1999. *Chlamydia pneumoniae*, asthma, and COPD: what is the evidence? *Ann. Allergy Asthma Immunol.* **83**:271–288, 291.
  19. Hatch, T. P. 1996. Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae? *J. Bacteriol.* **178**:1–5.
  20. Hatch, T. P., I. Allan, and J. H. Pearce. 1984. Structural and polypeptide differences between envelopes of infective and reproductive life cycle forms of *Chlamydia* spp. *J. Bacteriol.* **157**:13–20.
  21. Holtje, J. V. 1998. Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **62**:181–203.
  22. Hsia, R. C., Y. Pannekoek, E. Ingerowski, and P. M. Bavoil. 1997. Type III secretion genes identify a putative virulence locus of *Chlamydia*. *Mol. Microbiol.* **25**:351–359.
  23. Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**:379–433.
  24. Iliffe-Lee, E. R., and G. McClarty. 1999. Glucose metabolism in *Chlamydia trachomatis*: the 'energy parasite' hypothesis revisited. *Mol. Microbiol.* **33**:177–187.
  25. Kalman, S., W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis, and R. S. Stephens. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nat. Genet.* **21**:385–389.
  26. Knudsen, K., A. S. Madsen, P. Mygind, G. Christiansen, and S. Birkelund. 1999. Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of *Chlamydia pneumoniae*. *Infect. Immun.* **67**:375–383.
  27. Kreil, G. 1997. D-amino acids in animal peptides. *Annu. Rev. Biochem.* **66**:337–345.
  28. Kubori, T., Y. Matsushima, D. Nakamura, J. Uralil, M. Lara-Tejero, A. Sukhan, J. E. Galan, and S. Aizawa. 1998. Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**:602–605.
  29. Kuo, C., and L. A. Campbell. 1998. Is infection with *Chlamydia pneumoniae* a causative agent in atherosclerosis? *Mol. Med. Today* **4**:426–430.
  30. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**:105–132.
  31. Lee, C. A. 1997. Type III secretion systems: machines to deliver bacterial proteins into eukaryotic cells? *Trends Microbiol.* **5**:148–156.
  32. Longbottom, D., J. Findlay, E. Vretou, and S. M. Dunbar. 1998. Immunoelectron microscopic localisation of the OMP90 family on the outer membrane surface of *Chlamydia psittaci*. *FEMS Microbiol. Lett.* **164**:111–117.
  33. Longbottom, D., M. Russell, S. M. Dunbar, G. E. Jones, and A. J. Herring. 1998. Molecular cloning and characterization of the genes coding for the highly immunogenic cluster of 90-kilodalton envelope proteins from the *Chlamydia psittaci* subtype that causes abortion in sheep. *Infect. Immun.* **66**:1317–1324.
  34. Mathews, S. A., K. M. Volp, and P. Timms. 1999. Development of a quantitative gene expression assay for *Chlamydia trachomatis* identified temporal expression of sigma factors. *FEBS Lett.* **458**:354–358.
  35. Matsuhashi, M., M. Wachi, and F. Ishino. 1990. Machinery for cell growth and division: penicillin-binding proteins and other proteins. *Res. Microbiol.* **141**:89–103.
  36. Matsumoto, A. 1981. Electron microscopic observations of surface projections and related intracellular structures of *Chlamydia* organisms. *J. Electron Microsc.* **30**:315–320.
  37. Matsumoto, A., E. Fujiwara, and N. Higashi. 1976. Observations of the surface projections of infectious small cell of *Chlamydia psittaci* in thin sections. *J. Electron Microsc.* **25**:169–170.
  38. Moulder, J. W. 1993. Why is *Chlamydia* sensitive to penicillin in the absence of peptidoglycan? *Infect. Agents Dis.* **2**:87–99.
  39. Nichols, B. A., P. Y. Setzer, F. Pang, and C. R. Dawson. 1985. New view of the surface projections of *Chlamydia trachomatis*. *J. Bacteriol.* **164**:344–349.
  40. Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28**:1397–1406.
  41. Rockey, D. D., D. Grosenbach, D. E. Hruba, M. G. Peacock, R. A. Heinzen, and T. Hackstadt. 1997. *Chlamydia psittaci* IncA is phosphorylated by the host cell and is exposed on the cytoplasmic face of the developing inclusion. *Mol. Microbiol.* **24**:217–228.
  42. Rockey, D. D., R. A. Heinzen, and T. Hackstadt. 1995. Cloning and characterization of a *Chlamydia psittaci* gene coding for a protein localized in the inclusion membrane of infected cells. *Mol. Microbiol.* **15**:617–626.
  43. Rockey, D. D., and J. L. Rosquist. 1994. Protein antigens of *Chlamydia psittaci* present in infected cells but not detected in the infectious elementary body. *Infect. Immun.* **62**:106–112.
  - 43a. Schachter, J., and C. R. Dawson. 1990. The epidemiology of trachoma predicts more blindness in the future. *Scand. J. Infect. Dis. Suppl.* **69**:55–62.
  44. Scidmore-Carlson, M. A., E. I. Shaw, C. A. Dooley, E. R. Fischer, and T. Hackstadt. 1999. Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins. *Mol. Microbiol.* **33**:753–765.
  45. Stamm, W. E. 1999. *Chlamydia trachomatis* infections: progress and problems. *J. Infect. Dis.* **179**(Suppl. 2):S380–S383.
  46. Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**:754–759.
  47. Stephens, R. S., E. A. Wagar, and G. K. Schoolnik. 1988. High-resolution mapping of serovar-specific and common antigenic determinants of the major outer membrane protein of *Chlamydia trachomatis*. *J. Exp. Med.* **167**:817–831.
  48. Su, H., Y. X. Zhang, and R. Li. 1985. Presence of muramic acid in *Chlamydia trachomatis* proved by liquid chromatography-mass spectrometry. *Kexue Tongbao* **30**:695–699.
  49. Suchland, R. J., D. D. Rockey, J. P. Bannantine, and W. E. Stamm. 2000. Isolates of *Chlamydia trachomatis* that occupy nonfusogenic inclusions lack IncA, a protein localized to the inclusion membrane. *Infect. Immun.* **68**:360–367.
  50. Tipples, G., and G. McClarty. 1993. The obligate intracellular bacterium *Chlamydia trachomatis* is auxotrophic for three of the four ribonucleoside triphosphates. *Mol. Microbiol.* **8**:1105–1114.
  51. Weiss, E. 1950. The effect of antibiotics on agents of the psittacosis-lymphogranuloma group. I. The effect of penicillin. *J. Infect. Dis.* **87**:249–263.
  52. Zomorodipour, A., and S. G. Andersson. 1999. Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. *FEBS Lett.* **452**:11–15.