Recent progress in glycomics and proteomics of the Q fever bacterium

**Coxiella burnetii**

R. TOMAN, L. ŠKULTÉTY, K. PALKOVIČOVÁ, G. FLOREZ-RAMIREZ, P. VADOVIČ

Department of Rickettsiology, Institute of Virology, Slovak Academy of Sciences, Dúbravská cesta 9, 845 05 Bratislava, Slovak Republic

**Summary.** – *Coxiella burnetii* is an intracellular, Gram-negative bacterium and causative agent of Q fever. In humans, the disease ranges mostly from a flu-like illness and self-recovering mild pneumonia to severe meningoencephalitis, myocarditis or endocarditis. Recent molecular and biochemical/immunological advances, along with improved instrumentation, have provided unique insight into the host-parasite interrelationship and revealed previously unknown virulence strategies of *C. burnetii*. Noticeable progress has also been achieved in gaining a better understanding of the role of two major outer membrane components – lipopolysaccharide (LPS) and proteins in the life and immunopathobiology of the bacterium. Detailed glycomic studies have brought indispensable structural and functional information on LPS and its role in pathogenesis and immunity of Q fever. Recent proteomic studies have brought a deeper insight into the pathogen`s physiology, virulence and development and offered new possibilities in the investigation of inter/intra-species variation. This review will focus on advances in glycomics and proteomics of *C. burnetii* providing information on unique glycan and protein species, which together with other findings in the field, might lead to both a better understanding of this unusual pathogen and improvements in Q fever diagnosis and therapy.

**Keywords:** Coxiella burnetii; glycomics; proteomics; detection; diagnosis; Q fever

**Contents:**

1. The bacterium
   1.1 Phase variation
2. Outer membrane components
   2.1 Lipopolysaccharide
   2.2 Proteins
3. Diagnosis of Q fever and detection of the pathogen

---

**I. The bacterium**

*Coxiella burnetii* is an intracellular, highly infectious Gram-negative bacterium that causes Q fever, a zoonotic disease capable of being transmitted from animals to humans (Williams, 1991). Based on the 16S rRNA-encoding gene sequence, *C. burnetii* was reclassified from the order *Rickettsiales* to the phylum *Proteobacteria*, class *γ-Proteobacteria*, order *Legionellales*, family *Coxiellaceae* (Weisburg *et al.*, 1989; Waag and Thompson, 2005). *C. burnetii* has a genomic 42.2% guanosine (G)-plus-cytosine (C) content, which is closer to that of members of the order *Legionellales* than to the order *Rickettsiales* (~29% G+C).

Currently, the genome sequences of six *C. burnetii* isolates (CBuG Q212, C BuK Q154, Dugway 5J108-111, RSA331, RSA493, and MSU Goat Q177) are available (Beare *et al.*, 2005; Seshadri *et al.*, 2003). The genomes of all isolates are about 2 Mb in size with approximately 90% coding capacity (~2300 ORFs). Most isolates harbour 1 of 4 autonomously-replicating plasmids termed QpH1, QpRS, QpDV, and QpDG (Beare *et al.*, 2006). These plasmids range from 32 to 42 kb in size and share a common 25-kb “core” region.
along with unique regions. Isolates which lack plasmids contain plasmid sequences integrated into their chromosomes. A large portion (~40%) of genes in *C. burnetii* is of unknown function and approximately 10% of genes in all isolates are pseudogenes, probably reflecting the pathogen’s adaptation to its unique cellular compartment. The chromosome of the bacterium lacks bacteriophage or conjugation genes, and the genes associated with natural competence are pseudogenized in all isolates, indicating that the obligate intracellular bacterium has limited capacity for a horizontal gene transfer. Nevertheless, an intact recombination machinery is available that is most probably important in the insertion sequence element-mediated genome rearrangements discovered in the isolates. Although the metabolic capability of *C. burnetii* is fairly complex, deficiencies in a number of biochemical pathways are present and reflected in its dependence on a eukaryotic host cell or a specialized medium for replication (Beare et al., 2009). *C. burnetii* has a small Gram-negative pleomorphic coc-cocobacilli shape and produces two morphologically distinct cell types that comprise a bi-phasic developmental cycle. A small cell variant (SCV), with its characteristic condensed chromatin, is thought to be an extracellular survival form with enhanced resistance to environmental stressors such as desiccation and heat (Waag, 2007). When SCV invades the host, it develops into a large cell variant (LCV) that is metabolically and divisionally active. Differentiation of LCV to SCV occurs during the stationary phase of an organism’s growth cycle. It involves changes in the surface proteins (Waag, 2007; Ihnatko et al., 2012), but changes in LPS have not been studied thus far.

1.1 Phase variation

*C. burnetii* phase variation has been defined as a non-reversible switch from virulent phase I cells containing a full-length O-specific LPS I to avirulent phase II cells with largely modified, deep-core LPS II (Stoker and Fiset, 1956; Williams and Waag, 1991). A deeper insight into this complex phenomenon has been provided by Ftacek et al. (2000) based on their study of both the composition and structural features of LPS isolated during serial passage of *C. burnetii* in embryonated hen eggs. The authors considered a known fact (Keenleyside and Whitfield, 1999) that LPSs isolated from many smooth (S) Gram-negative bacteria are heterogeneous in size and shape and may contain, besides (S) LPSs, some rough (R) LPSs and variously truncated LPS molecules. They therefore anticipated that *C. burnetii* contains cell populations that may express multiple LPS structures. From the results obtained, Ftacek et al. (2000) suggested that a redistribution of the existing LPS populations took place during the *C. burnetii* phase variation due to an increasing prevalence of those cells in the whole cell population that express LPS molecules with truncated O-polysaccharide chain and those being of (R) type. Thus, in an attempt to optimize the balance between the expenditure of energy for the synthesis of complete LPS I and the need to resist the host’s immune system, *C. burnetii* has evolved genetic mechanisms of the transition from an energy-demanding (phase I) to less energy-demanding (phase II) states. Therefore, the cell populations with gene deletions in the O-polysaccharide biosynthesis region have a growth advantage in an immunoincompetent host.

Investigations of the O-polysaccharide biosynthesis region in the avirulent phase II Nine Mile isolate (NM II, RSA 439) revealed that a large group of the genes was deleted (Hoover et al., 2002; Denison et al., 2007). This fact was also confirmed by recent proteomic studies (Skulley et al., 2011), which have detected proteins associated with the O-polysaccharide chain biosynthesis only in the virulent phase I Nine Mile isolate (NM I, RSA 493). In this isolate, 17 proteins involved in LPS biosynthesis and metabolism were identified. More than half of the proteins detected in NM I, but not in NM II, were products of genes located in the deleted region of the chromosome. Thus, these proteins were proposed as virulence-associated proteins and biomarkers related to the biosynthesis of LPS.

Phase variation of *C. burnetii* causes modifications in phospholipids, too. Phospholipids from the NM I cells were much more complex than those from the NM II cells (Domingues et al., 2002). Moreover, in the latter, the absence of phospholipids of the phosphatidylinositol class was noticeable.

Nevertheless, some phase II isolates contained no apparent deletions in the O-polysaccharide biosynthesis region (Denison et al., 2007). To the best of our knowledge, the phase variation mechanisms have not been elucidated by bioinformatics tools in *C. burnetii* thus far. Similarly, no R→S transition has ever been observed in *C. burnetii* phase II isolates. Thus, molecular mechanisms influencing the LPS modifications during the *C. burnetii* phase variation represent a further challenge in the investigation of this unusual bacterium.

2. Outer membrane components

As in other Gram-negative bacteria, LPS and proteins are major components of the outer membrane of *C. burnetii*. Both biopolymers are of particular biological, immunological and medical significance. Modern methodologies have an immense potential for gaining significantly deeper insight into the functional interaction of *C. burnetii* LPS and proteins and their roles not only in growth and development of the microbe but also in pathogenesis and immunity of Q fever (Narasaki and Toman, 2012; Ihnatko et al., 2012). An attempt
to summarize some of the recent developments in glycomics
and proteomics in the field is given below.

2.1 Lipopolysaccharide

As mentioned earlier, virulent C. burnetii cells contain
a full-length O-specific LPS I. The polymer contains (Mayer
et al., 1988; Toman et al., 2009; Narasaki and Toman, 2012)
two unusual sugars in its O-poly saccharide chain, namely
virenose (Vir, 6-deoxy-3-C-methyl-D-gulopyranose) and
dihydrohydroxy streptose [Strep, 3-C-(hydroxymethyl)-L-
lyxofuranose]. Both sugars have not been found in other
bacterial LPSs and are considered unique biomarkers of
C. burnetii. The enantiomeric forms and ring conforma-
tions of both saccharides were established (Toman
et al., 1998) and earlier methylation analysis showed that Strep
was present in the parent LPS I in a furanose form (Toman,
1991). It was therefore anticipated that a furanose to pyra-
none tautomerization took place in the course of the isolation
procedure (Toman et al., 1998; Fig. 1). Genes involved in
the synthesis of both sugars were suggested (Hoover et al.,
2002; Thompson et al., 2003) and their protein products
NDP-hexose 3-C-methyltransferase TylCIII (CBU0691) and
methyltransferase FkbM family (CBU0683) were
identified by the proteomic analysis (Skultety et al., 2005).
Most recently, a homology-based biosynthetic pathway for
Vir has been proposed based on the results of preceding
genomic, glycomic, and proteomic analyses (Narasaki et
al., 2011; Narasaki and Toman, 2012; Florez-Ramirez et
al., 2012). However several enzymatic steps taken from
the bioinformatic approach must be verified by laboratory
experiments in the future.

From the methylation linkage analysis it is anticipated
that mainly Vir, Strep and D-mannose (Man) are located in
terminal positions of the O-polysaccharide chain of LPS I.
Vir, Man, and probably D-glucosamine (GlcN) also consti-
tute its backbone (Vadovic et al., 2005; Toman et al., 2009;
Narasaki and Toman, 2012). Two terminal Man and 2,3- and
3,4-disubstituted D,D-hexaose-D,D-manno-heptoses (D,D-Hep)
were from the core region of LPS I (Toman and Skultety,
1996). Current knowledge on a detailed chemical structure
of LPS I is lacking and therefore only a tentative structural
model for LPS I can be depicted (Fig. 2).

Both compositional and structural studies of LPS II
have produced a lot of controversy in the past (Schramek
and Mayer, 1982; Amano and Williams, 1984; Mayer et
al., 1988). More detailed studies (Toman et al., 1993;
Toman and Skultety, 1994) confirmed the presence of enteric
3-deoxy-D-manno-oct-2-uloseonic acid (Kdo) in LPS II and
subsequently demonstrated the presence of three Kdo resi-
dues with a structural arrangement similar to that of many
bacterial LPSs. Linkage analysis of the carbohydrate region
revealed two terminal Man, 2- and 3,4-linked D,D-Hep.
Fast atom bombardment (FAB)- and electrospray ionization
(ESI)-mass spectrometries (MS) of the lipid A deprived LPS
II established the sequence of its sugar residues (Toman and
Skultety, 1996) as shown in Fig. 3. The subsequent matrix
assisted laser desorption ionization (MALDI)-MS measure-
ments confirmed the previous results and established the
molecular mass of LPS II to 2841.58 (Toman et al., 2003b).
Schematic presentation of structural arrangement of the sugar residues in LPS I from *C. burnetii*

GalNAc, N-acetyl-D-galactosamine; GlcNAc, N-acetyl-D-glucosamine; Glc, D-glucose; Hep, D-glycero-D-manno-heptose; Kdo, 3-deoxy-D-manno-oct-2-ulosonic acid; Man, D-mannose; Strep, dihydrohydroxystreptose [3-C-(hydroxymethyl)-L-lyxose]; Vir, virenose (6-deoxy-3-C-methyl-D-gulose). GlcN I and GlcN II, the backbone D-glucosamines of lipid A substituted with acyl chains. The underlined sugars prevail in the O-specific chain.

Basic chemical structure of LPS II from *C. burnetii* in avirulent phase II

GlcN I and GlcN II, the backbone D-glucosamines of lipid A substituted with acyl chains. Hep, D-glycero-D-manno-heptose; Kdo, 3-deoxy-D-manno-oct-2-ulosonic acid; Man, D-mannose; X = length of the acyl chain.

The lipid A portion of LPS is linked to the core oligosaccharide mostly via Kdo and serves as the hydrophobic anchor of LPS in the outer membrane (Alexander and Rietschel, 2001). It is well known that lipid A, as the principal endotoxic
component of LPS, plays a major role in the pathogenesis of bacterial infections and is an important contributor to massive inflammation, sepsis, and septic shock leading to fatalities in Gram-negative bacteria infections (Alexander and Rietschel, 2001). Toman et al. (2003a) investigated lipids A from the LPSs I of Henzerling and S isolates and showed that both lipids A were tetraacylated with two amide-linked 3-hydroxy and two ester-linked nonhydroxylated fatty acids attached to the GlcN disaccharide. Differences were found only in the small structural details. In the isolate Priscilla, two major tetraacylated molecular species were also found (Toman et al., 2004) as the major components despite a considerable microheterogeneity of the lipid A. From these and our latest results the structural model for the lipid A was proposed (Fig. 3) as an integral part of the structural features of LPS II. However, one should bear in mind that differences in nature, length and possibly in distribution of fatty acids along the GlcN disaccharide are apparent. These fatty acid variations contribute to a remarkable diversity of the C. burnetii lipid A although its biochemical synthesis is a highly conserved process like in other Gram-negative bacteria.

The composition and structure of lipid A of C. burnetii differ considerably from those published (Alexander and Rietschel, 2001) for the classical form of enterobacterial lipid A with high endotoxicity found, e.g. in Escherichia coli or Neisseria meningitidis. Thus, distinct structural features of the C. burnetii lipid A could be the reason for its reduced endotoxic potency in comparison with the enterobacterial lipids A.

Most recently, biological properties and functional characteristics of LPS I and LPS II and their roles in both the diagnosis and immunobiology of Q fever have been reviewed (Toman et al., 2009; Narasaki and Toman, 2012) and are not featured in this paper. Nevertheless, the latest finding in the field (Barry et al., 2012) is worth mentioning here. The authors found that LPS I did not stimulate virulent phase I C. burnetii trafficking to phagolysosomes. The defect in targeting the cells to degradative compartments involves an antagonistic engagement of Toll-like receptor 4 by LPS I and a lack of p38α-mitogen-activated protein kinase (p38α-MAPK)-driven phosphorylation and recruitment of vacuolar protein-sorting 41 (Vps41)-homotypic fusion and protein sorting (HOPS) complex to LPS I-containing vesicles. Thus, virulent C. burnetii can evade the immune response via variations in LPS composition, and p38α-MAPK and its cross-talk with Vps41 play a central role in the trafficking of bacteria to phagolysosomes.

### 2.2 Proteins

Proteins are functional molecules that operate through all metabolic and regulatory pathways of organisms and cells. They are often organized into multiprotein complexes that function as molecular drivers of many processes. The world of proteins is very complex. Due to alternative transcription initiation and/or alternative splicing, the expression of single gene might produce several transcripts, which may be translated to several proteins due to alternative translation initiation. These proteins might be further post-translationally modified and furthermore, the magnitude of the protein species abundance within bacterial cell may differ by 7–10 in order of magnitude. With such a high level of complexity, large scale protein analysis based on MS becomes a necessary tool to advance C. burnetii research.

The complete sequencing of genomes of six C. burnetii isolates (see above) has brought new opportunities to obtain more information about the proteome of the bacterium, and also provided a shift towards better understanding of the mechanisms involved in its pathogenesis. Analyses of C. burnetii genomes revealed the presence of many mobile elements, pseudogenes and hypothetical proteins, suggesting ongoing genome reduction (Seshadri et al., 2003; Beare et al., 2005). During reduction, genes accumulate mutations, lose functions and eventually disappear. It seems probable that these processes are also associated with phase variation at which time a large group of genes is lost (Vodkin et al., 1986; Hoover et al., 2002; Thompson et al., 2003; Denison et al., 2007).

The initial proteomic studies (Skultety et al., 2005; Samoilis et al., 2007) were focused on the identification of proteins from C. burnetii isolates NM I and NM II that were resolved by 2-dimensional electrophoresis (2-DE) or liquid chromatography (LC). These works resulted in the production of proteome reference maps for both phase I and phase II C. burnetii. An example of such a reference map for the NM I isolate is given in Fig. 4. As many as 555 and 600 distinct proteins are covered by 2-DE and this number increases up to 800 proteins covered by liquid chromatography (LC).

![Fig. 4](image)

Colloidal Coomassie-stained 2-DE proteome reference map of C. burnetii NM I isolate.
protein spots were detected in the 2-DE gels of phase I and II isolates, respectively. Using various MALDI and ESI MS/MS approaches, more than 200 ORF products were identified in the whole cell lysates of both phases of the bacterium (Skultety et al., 2005, 2011; Samoilis et al., 2007). A typical workflow of protein identification using LC MS/MS analysis is depicted in Fig. 5. The function of each identified protein was then predicted and extensive bioinformatic analyses were performed. Among the proteins identified, 10 proteins involved in pathogenesis and pathogenicity were detected including the OmpH (CBU0612), a previously identified marker for Q fever endocarditis (Sekeyova et al., 2009).

In recent years, several comparative proteomic studies based on MS/MS techniques have been performed with C. burnetii to look for differences between the SCV and LCV cells (Coleman et al., 2007), NM vs. Q212 isolates (Papadioti et al., 2011, 2012), persistent vs. short-term infection (Vranakis et al., 2011a) or antibiotic resistant isolates (Vranakis et al., 2011b, 2012) of C. burnetii.

It is well known that surface associated proteins (SAPs) and effector proteins that functionally mimic the activities of eukaryotic proteins play several important roles in the host-parasite interactions of Gram-negative bacteria related to both pathogenesis and virulence. Due to the role in invasion and intracellular survival of pathogen in the host, SAPs from C. burnetii could be the crucial antigens for diagnosis and vaccine development. We recently applied various bioinformatic tools and databases for in silico prediction of the outer membrane proteins (OMPs) and lipoproteins (LPs) of C. burnetii (Flores-Ramirez et al., 2009). Furthermore, three different proteomic methodologies based on detergent extraction, basic protein isolation and speroplast formation followed by ESI MS/MS coupled to nanoscale ultra-performance LC were employed to investigate these proteins (Flores-Ramirez, 2012). Recent studies on C. burnetii (Voth et al., 2009; Chen et al., 2010; Samoilis et al., 2010) have confirmed the existence of or discovered proteins with predicted cell envelope association including the integral membrane proteins and LPs anchored to the outer membrane with an N-terminal lipid tail. Furthermore, potential membrane associated proteins that may form complexes with anchored proteins and proteins associated with the Type IV secretion system (Seshadri et al., 2003) were detected together with effector proteins that are transported through this system.

In summary, as many as 822 uniquely identified proteins which represent 45% of all ORFs of C. burnetii were identified in the proteomic studies mentioned here. Given the fact that no organism expresses its entire proteome potential at a given time, the identification of almost half of the bacterium’s proteome can be considered as a success (Vranakis et al., 2013). So far, proteomic studies have generated a vast amount of infor-
mation on several aspects of the bacterium such as virulence factors, detection/diagnostic and immunogenic biomarkers, inter/intra-species variation, resistance to antibiotics, SAPs and secreted effector proteins with significant clinical impact. All of the proteins identified were assigned to 20 functional categories based on the criteria of Samoilis et al. (2007). The largest group includes 239 proteins with unknown predicted function which is a reminder of the extra efforts required on understanding *C. burnetii*. The next largest group includes 98 proteins associated with protein synthesis and the third largest group contains 74 proteins associated with energy metabolism or electron transport. Sixty of the proteins are cell envelope related and 52 proteins with post-translational modification, degradation, or protein turnover. Due to the resistant nature of the bacterium, other functional categories of particular interest comprise proteins involved in pathogenesis (17), toxin production and resistance (13), adaptation to atypical conditions (11), and detoxification (6) (Vranakis et al., 2013). Thus, the identified proteins provide an additional insight into the pathogenesis and immunity of Q fever.

3. Diagnosis of Q fever and detection of the pathogen

The signs and symptoms of human Q fever are frequently non-specific, which can complicate and delay clinical diagnosis. The incubation period may vary from a few days to several weeks, probably depending in part on the infectious dose and the immune status of the host. Typical presentations of acute disease include a flu-like symptoms (fever, headache, chills, and fatigue), pneumonia and granulomatous hepatitis. Acute disease is usually self-limited, even without antibiotic therapy (Maurin and Raoult, 1999). It is estimated that ~ 5% of asymptomatic or acute Q fever infections give rise to more serious chronic Q fever. The most frequent and serious chronic presentation of Q fever is a culture-negative endocarditis. Without prompt treatment, chronic Q fever can result from the bacterium that migrated to the bone marrow and subsequently migrated to other tissues, e.g. the endocardium (Waag, 2007).

Various serological methods are currently being used for the rapid and sensitive diagnosis of the disease (Maurin and Raoult 1999; Slaba et al., 2005) but ambiguous results have been obtained in several cases. Detection of the infectious agent by PCR is very useful, especially in cases where serological tests have brought ambiguous results (Waag and Thompson, 2005). The insertion sequence IS1111 is the target of choice in PCR detection of *C. burnetii* as the repetitive element has multiple genomic copies, thereby increasing assay sensitivity. However, PCR-based technology is sensitive mainly in the early disease state. It was reported that *C. burnetii* DNA becomes undetectable in serum 17 days after infection (Schneeberger et al., 2010). Protein microarrays have recently been employed to systematically evaluate the human immune response to *C. burnetii* infection (Beare et al., 2008; Vigil et al., 2010, 2011). For example, Vigil et al. (2010) used an array containing 1901 *C. burnetii* ORFs (84% of the entire proteome) to test Q-fever patient sera and the respective controls. Threeteen antigens specifically reacted with patient sera, and nine of these were validated using an immunostrip platform applicable to a clinical laboratory. The results are promising for further research and development of a new generation of diagnostic and vaccine antigens based on recombinant protein.

Unique biomarkers Vir and Strep in LPS I provide a good opportunity to develop a new generation of monoclonal antibodies (MAbs) against them which would enable rapid, sensitive and specific detection of virulent *C. burnetii*. A MAb (IgG2b subclass) has been generated that was shown to be highly specific for the presence of Vir in the pathogen and LPS I as only Vir-containing *C. burnetii* isolates/variants and their LPSs reacted with the MAb (Palkovicova et al., 2009). In addition, no cross-reaction was observed with the cells and LPSs from the selected species of the families Rickettsiaceae, Piscirickettsiaceae, Chlamydiaceae, and Enterobacteriaceae. For the field applications, the MAb is currently tested in the lateral flow assay format. The preliminary results indicate high application potential of the assay for rapid detection of the infectious agent in various environmental samples when compared with the current methods used in the field (Frangoulidis and Toman, unpublished data).

Implementation of various MS techniques may also substantially contribute to progress, especially in discrimination of *C. burnetii* isolates. Currently, few works have been focused on the determination and identification of specific *C. burnetii* biomarkers (Shaw et al., 2004; Pierce et al., 2007; Skultety et al., 2007; Hernychova et al., 2008). It has been suggested that proteomic approaches including MALDI, time-of-flight (TOF) MS and ESI MS/MS might be suitable for detection, identification, and typing of the bacterium in complex environment. In this connection, Hernychova et al., 2008 concluded that MALDI TOF MS will not be sufficient for obtaining relevant proteomic data. The authors suggested that application of a shotgun strategy based on the LC MS/MS techniques operated in a multiple reaction monitoring scanning mode may confirm the identity of the *C. burnetii* pathogen according to its specific proteins or peptides.

Acknowledgement. The authors wish to thank Mrs. M. Benkovičová and E. Hasíková for growing and preparation of the highly purified *C. burnetii* cells and some LPS and protein samples. This work was supported in part by the grants No. 2/0026/12 and No. 2/0156/11 of the Scientific Grant Agency of Ministry of Education of Slovak Republic and the Slovak Academy of Sciences.
References
