Published in final edited form as:

J Proteome Res. 2010 April 5; 9(4): 1995–2005. doi:10.1021/pr9011602.

A multimethodological approach to identification of glycoproteins from the proteome of *Francisella tularensis*, an intracellular microorganism

Lucie Balonova^{1,2,3}, Lenka Hernychova^{1,*}, Benjamin F. Mann⁴, Marek Link¹, Zuzana Bilkova³, Milos V. Novotny⁴, and Jiri Stulik¹

¹ Institute of Molecular Pathology, Faculty of Military Health Sciences, University of Defence, 500 01 Hradec Kralove, Czech Republic

² Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, 530 02 Pardubice, Czech Republic

³ Department of Biological and Biochemical Sciences, Faculty of Chemical Technology, University of Pardubice, 530 02 Pardubice, Czech Republic

⁴ Department of Chemistry, National Center for Glycomics and Glycoproteomics, Indiana University, 47405 Bloomington, IN, USA

Abstract

It appears that most glycoproteins found in pathogenic bacteria are associated with virulence. Despite the recent identification of novel virulence factors, the mechanisms of virulence in *Francisella tularensis* are poorly understood. In spite of its importance, questions about glycosylation of proteins in this bacterium and its potential connection with bacterial virulence have not been answered yet. In the present study, several putative *Francisella tularensis* glycoproteins were characterized through the combination of carbohydrate-specific detection and lectin affinity with highly sensitive mass spectrometry utilizing the bottom-up proteomic approach. The protein PilA that was recently found as being possibly glycosylated, as well as other proteins with designation as novel factors of virulence, were among the proteins identified in this study. The reported data compile the list of potential glycoproteins that may serve as a take-off platform for a further definition of proteins modified by glycans, facilitating a better understanding of the function of protein glycosylation in pathogenicity of *Francisella tularensis*.

Keywords

Francisella tularensis; glycoprotein; glycosylation; hydrazide; lectin affinity; 2-DE; mass spectrometry

*Corresponding author: Lenka Hernychova, Institute of Molecular Pathology, FMHS UO, Trebesska 1575, 500 01 Hradec Kralove, Czech Republic. Tel: ++420973253223, Fax: ++420495513018, hernychova@pmfhk.cz.

Supporting Information Available: Table S1, binding and elution conditions used in lectin affinity chromatography; Table S2, list of identified glycosyltransferases isolated through lectin affinity chromatography. Table S3a, prediction of O-glycosylated sites of FTH_1071; S3b, prediction of O-glycosylated sites of FTH_0414. Spectrum S1, MALDI-MS/MS spectrum of the only PilA tryptic peptide AQLGSDLSALGGAK. Spectrum S2a, MALDI-MS spectrum of membrane protein-enriched fraction of *F. tularensis* FSC200 after β -elimination; S2b, MALDI MS/MS spectrum of selected peak #1411. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Introduction

Francisella tularensis (*F. tularensis*) is a nonmotile, nonsporulating, Gram-negative intracellular pathogen that is capable of causing tularemia, a fatal disease in humans and other mammals. Owing to its high infectivity and potential for airborne transmission, this bacterium has been designated a Category A agent of bioterrorism.^{1,2} It fulfills all requirements for a potential biological weapon: extreme virulence, low infectious dose, ease of aerosol dissemination, and the capacity to cause severe illness and death. Inhalation of as few as 10 colony-forming units is sufficient to cause disease in humans, while 30 – 60% of untreated infections can be fatal.^{3,4} There are four subspecies of *F. tularensis* that are highly conserved in their genomic content⁵ but differ in their virulence: *F. tularensis* subsp. *tularensis* (type A), *novicida*, *mediasiatica*, and *holarctica* (type B).⁶ From these, *F. tularensis* subsp. *tularensis*, *mediasiatica*, and *holarctica* can cause disease in humans, with type A being the most virulent form. In 2004, Nano *et al.* discovered the existence of *Francisella* pathogenicity island (FPI) that is required for intracellular growth and virulence of *F. tularensis* in mice.⁷ Most of the FPI-encoded genes are highly conserved among the strains, which indicates that the presence of FPI alone is important but not sufficient for the high virulence of type A strain. Based on this study, a potential participation of glycosylation in the virulence of this pathogen has been postulated.

Through its involvement in a number of biological processes, such as cell-to-cell recognition, protein folding, and host immune response, glycosylation is undoubtedly among the most biologically important post-translational modifications decorating proteins. The initial presumption that prokaryotes, especially bacteria, lack the cellular machinery needed to glycosylate their proteins has been countered by the growing evidence for the occurrence of glycoproteins in different bacterial species, including numerous important Gram-negative and Gram-positive pathogens such as *Campylobacter jejuni* (*C. jejuni*),⁸ *Pseudomonas aeruginosa* (*P. aeruginosa*),⁹ *Neisseria meningitidis*,¹⁰ *Neisseria gonorrhoeae* (*N. gonorrhoeae*),¹¹ and *Mycobacterium tuberculosis* (*M. tuberculosis*)¹². In addition, the general N-glycosylation system was described in *C. jejuni*,⁸ while an O-glycosylation system was recently reported in *N. gonorrhoeae*.¹³ The presence of glycosylation has been shown to impact the function of bacterial proteins modified by glycans in terms of their implication in adhesiveness and invasion to host cells.^{14,15} Therefore, it is not surprising that cell-surface filamentous appendages, such as pili and flagella, are among the cellular structures with proteins that are heavily glycosylated, as they encounter the first contact with a host cell surface. Although both N- and O-linked structures have been found in bacteria, O-linked glycosylation predominantly occurs in such appendages.¹⁶

A study by Forslund *et al.*¹⁷ found that PilA of *F. tularensis* subsp. *holarctica*, strain FSC200 appears to be post-translationally modified, possibly through glycosylation. This finding was recently supported by the evidence for *Francisella* PilA protein glycosylation in *N. gonorrhoeae* utilizing the extreme promiscuity of PglO oligosaccharyltransferase with regard to protein substrates.¹⁸ In that study, an increase in the PilA relative motilities in *N. gonorrhoeae* protein glycosylation mutants and variants (*PglA*, *PglC*, *pglD*, *pglF* and *pglO*) was observed, when compared with the motility of PilA in the wild-type. In addition, the ability of the antibodies to strain N400 Tfp to react with the PilA-associated appendages was abrogated in the glycosylation null *PglC* mutant.¹⁹ Up to now, PilA is the only reported *F. tularensis* putative glycoprotein.

In the present study, our intent was to confirm the presence of glycosylation in *F. tularensis* PilA and also to search for the presence of other possible N- and O-glycosylated proteins. Consequently, we employed a comprehensive investigation of the *F. tularensis* subsp. *holarctica* FSC200 glycoproteome by combining three fundamentally distinct glycoprotein

detection approaches: (1) hydrazide labelling and (2) lectin blotting, and (3) the widely used glycoprotein enrichment technique of lectin affinity chromatography. The outermost surface of bacteria and their extra- and intracellular membranes are postulated primarily to be glycosylated, as opposed to cellular proteins, although the latter cannot be entirely excluded from consideration. Therefore, the present study was focused on analyzing bacterial fractions enriched in membrane proteins. To our best knowledge, a targeted study of *F. tularensis* glycoproteome using the glycoproteomic tools such as hydrazide chemistry and lectin affinity has previously not been conducted.

Materials and methods

Bacterial strains and culture conditions

The *F. tularensis* ssp. *holarctica* strain FSC200 used in this study was kindly provided by Dr. Åke Forsberg, FOI Swedish Defence Research Agency, Umea, Sweden. Bacteria were grown, harvested, and lysed within a BioSafety Level 2 containment facility. Bacteria were cultured on McLeod agar supplemented with bovine hemoglobin (Becton Dickinson, USA) and IsoVitaleX (Becton Dickinson, USA) at 36.8 °C for 24 – 48 h. Colonies scraped from the plate were inoculated into Chamberlain medium and cultivated for 12 h at 36.8 °C under constant shaking. The 12-h cultures were diluted with fresh Chamberlain medium (OD_{600 nm} 0.1) and grown until the late logarithmic growth phase of bacteria (OD_{600 nm} 0.8). Bacterial cells were collected by centrifugation at 9 000g for 15 min at 4 °C and the pellets were washed three times with cold PBS (pH 7.4). The resulting pellets were resuspended in 50 mM Tris/HCl (pH 8.0). Protease inhibitor cocktail (Roche, Mannheim, Germany) was added to a final dilution 1:50.

Preparation of whole-cell lysates

The cells were disrupted using a French press twice at 16 000 psi, while the resulting cell debris along with intact microbes were removed by centrifugation at 12 600g for 30 min at 4 °C. Benzonase nuclease (250 U/μl, Sigma, St. Louis, USA) was added to the supernatant, resulting in a final concentration of 0.5 U/ml of lysate.

Preparation of membrane protein-enriched fraction

Fractions enriched in the membrane proteins were prepared by sodium carbonate extraction according to the method described by Molloy *et al.*²⁰ Briefly, the supernatant was diluted with ice-cold 0.1 M sodium carbonate (pH 11.0) and was gently stirred on ice for 1 h. Carbonate-treated membranes were collected by ultracentrifugation at 115 000g for 1 h at 4 °C. The supernatant was discarded and the membrane pellet was resuspended in ice-cold 50 mM Tris/HCl (pH 8.0), and then collected by centrifugation at 115 000g for 30 min at 4 °C. The final membrane protein-containing pellet was solubilized in various lysis buffers containing protease inhibitor cocktail. The compositions of lysis buffers were designed to be compatible with the downstream methods. For example, for the lectin affinity chromatography, Nonidet P-40 (Roche, Mannheim, Germany) was added to have a final concentration of 0.5%. Samples were then sonicated for 2 min in 1-s pulses with 15-s cooling periods after each pulse. Proteins were quantified by either Bicinchoninic acid or Bradford assays (Sigma, St. Louis, USA) and stored at – 80 °C.

Mini two-dimensional gel electrophoresis and Semi-dry Western blot

For solubilization of sparingly-soluble membrane proteins, a rehydration buffer containing 7 M urea, 2 M thiourea, 1% (w/v) ASB-14, 4% (w/v) CHAPS, 1% (w/v) dithiotreitol (DTT), 1% Ampholytes pH 3 – 10 (Bio-Rad, Hercules, CA), and 0.5% Pharmalytes pH 8 – 10.5 (Amersham Biosciences, Uppsala, Sweden) was used. Typically, proteins were loaded by in-

gel rehydration onto polyacrylamide gel strips with a nonlinear immobilized pH gradient (IPG) from 3 – 10 (GE Healthcare, Uppsala, Sweden) and separated according to their different pI values by isoelectric focusing (IEF). The linear basic strips (pH 6 – 11) were swollen in rehydration buffer containing 0.5% (v/v) IPG buffer and DeStreak overnight, while the samples were cup-loaded at the anodic side. Following IEF, the IPG strips were treated in equilibration buffer containing 2% (w/v) sodium dodecyl sulfate (SDS), 50 mM Tris/HCl (pH 8.8), 6 M urea, 30% (v/v) glycerol, and 1% (w/v) DTT. This was immediately followed by a second equilibration of strip in the same solution containing 4% (w/v) iodoacetamide in place of DTT. In the second dimension, the IPG strips were embedded onto 12% homogeneous SDS polyacrylamide gels, and after electrophoresis, separated proteins were transferred onto BioTrace NT 0.45 μ m nitrocellulose membranes (Gelman Sciences Inc., Ann Harbor, MI). Glycoproteins from the membranes were detected using the DIG Glycan Differentiation kit.

Glycoprotein detection using DIG Glycan Differentiation Kit

DIG Glycan staining (Roche, Mannheim, Germany) was employed following the manufacturer's protocol with slight modifications. Briefly, the membranes were incubated in a tris-buffered saline (TBS) overnight, in order to avoid nonspecific binding. After washing, the membranes were incubated with 1 – 10 μ g/ml of digoxigenin-labeled lectins for 1 h. Unbound lectins were removed by repeated washing in TBS. The membranes were then incubated with 0.75 U/ml of alkaline phosphatase-conjugated anti-digoxigenin for 1 h. Following repeated washes with TBS, a staining solution containing the substrate NBT/BCIP was used to visualize glycoproteins. The reaction was stopped by rinsing the membranes with doubly-distilled water. Transferrin, asialofetuin, and fetuin were used as the positive-control (model) glycoproteins for SNA, PNA, and DSA and MAA lectins, respectively. As a negative control, recombinant FTT Igl C protein from *E. coli* was used.

Glycoprotein detection using Pro-Q^(R) Emerald 300 Glycoprotein Stain Kit

Pro-Q Emerald staining (Invitrogen, Eugene, OR) was performed according to the manufacturer's protocol with slight modifications. Briefly, gels were oxidized with periodic acid for 30 min. After washing with 3% glacial acetic acid to remove residual periodate, the gels were incubated in Pro-Q Emerald 300 staining solution (diluted 25-fold into staining buffer) for 2 h and subsequently washed. Stained gels were visualized by illumination using CCD camera Image station 2000R (Eastman Kodak, Rochester, NY). After detection of glycoproteins, gels were stained with SYPRO Ruby protein gel stain to detect all proteins as a control.

In-gel tryptic digestion of proteins

Protein spots detected on-gel by Pro-Q Emerald staining or on-blot by DIG Glycan staining were excised from the representative gels and subjected to in-gel tryptic digestion. Briefly, gel pieces were destained with 100 mM Tris/HCl (pH 8.5) in 50% acetonitrile for 20 min at 30 °C, followed by equilibration with 50 mM ammonium bicarbonate (pH 7.8) in 5% acetonitrile. After vacuum drying, the gel pieces were swollen in 2.5 μ l of trypsin solution (40 ng/ μ l) for 20 min at 4 °C. Finally, 15 – 30 μ l of equilibration buffer was added just to cover the gel. The samples were incubated at 37°C for 18 h. Resulting peptides were mixed with matrix solution (5 mg/ml of α -cyano-4-hydroxycinnaminic acid in 50% acetonitrile, 0.1% trifluoroacetic acid) and spotted onto a MALDI plate.

Mass spectrometry and database searching

Mass spectra were recorded in positive reflectron mode on a 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Framingham, MA) equipped with an Nd:YAG laser (355

nm) and operated in delayed extraction mode. Internal calibration of mass spectra was conducted utilizing the tryptic peptides as a result of its autolysis. The fragmentation analysis of six most intensive peaks was performed without applying CID. Acquired data were evaluated using GPS Explorer™ Software version 3.6 (Applied Biosystems, Framingham, MA) that integrates the Mascot search algorithm against *F. tularensis* OSU18 genome database. Trypsin was selected as the proteolytic enzyme, and one missed cleavage was allowed. Fixed modifications were set as carbamidomethyl for cysteine residues, while oxidation of methionine was set as a variable modification. Proteins were considered identified with confidence when protein score confidence interval (%) was greater than 95 (p-value < 0.05) and a minimum of two peptide sequences per protein were identified (note, protein confidence interval 95% is equal to Mowse score significance level for the search).

An exception was made for the protein PilA (FTH_0384), as the *in silico* analysis revealed the presence of the only tryptic peptide (Supporting data, Spectrum S1).

Lipopolysaccharide removal prior to lectin affinity chromatography

Lipopolysaccharide (LPS) was eliminated from the bacterial lysate using the Detoxi-gel Endotoxin removing gel, following the manufacturer's instructions (Thermo Scientific, Rockford, IL). Briefly, the gel resin was regenerated by washing with 1% sodium deoxycholate, followed by pyrogen-free buffer to remove the detergent. After the equilibration of the resin with pyrogen-free buffer, the sample was loaded onto the column and incubated with a resin for 1 h. LPS-depleted sample was then collected as a flow-through. Protein quantification was performed using either Bicinchoninic acid or Bradford assay.

Lectin affinity chromatography

Lectins used in this study were purchased from Vector Laboratories, Inc. (Burlingame, CA). LPS-depleted samples were diluted with an appropriate lectin binding buffer (Supporting information, Table S1) and added to a 500- μ l aliquot of lectin slurry pre-equilibrated with the lectin binding buffer. After a 2-h incubation period, unbound proteins were washed from the lectin with the binding buffer. Next, the bound proteins were eluted from the lectin using an appropriate elution buffer (Supporting information, Table S1). The unbound proteins were quantified by Bradford assay.

Sample clean-up and tryptic digestion

Bound protein fractions were filtered using 0.22 μ m cellulose acetate filters (Agilent Technologies, Palo Alto, CA), desalted on MICROCON 10 kDa cut-off membrane filters (Millipore, Billerica, MA), dried, and resuspended in a 50 mM ammonium bicarbonate solution. Prior to digestion, the protein amount was determined by the Bradford assay. Samples were then reduced with 10 mM DTT at 60 °C for 1 h, followed by alkylation with 20 mM iodoacetamide in the dark at room temperature for 45 min. Next, proteins were digested with 1:50 trypsin:protein (w/w) ratio at 37 °C for 18 h.

LC/ESI-MS/MS analysis and data processing

Tryptic digests were analyzed by C18 nanoscale reversed-phase liquid chromatography coupled on-line to an XCT Ultra mass spectrometer (Agilent Technologies, Palo Alto, CA) or an LTQ ICR-FT mass spectrometer (Thermo Finnigan, San Jose, CA). Samples were desalted and preconcentrated on a micro-precolumn cartridge C18 (300 μ m i.d. $\times$ 5 mm) (LC Packings, Sunnyvale, CA). After loading and washing the peptides for 10 min with mobile phase A (97%/3%/0.1% water/acetonitrile/formic acid), the trapping column was switched in-line with the analytical column. The separation of peptides was conducted with a Zorbax

300SB C18 column (75 μ m i.d. $\times$ 150 mm) (Agilent Technologies, Palo Alto, CA) with a linear gradient, from 3 to 55% phase B (3%/97%/0.1% water/acetonitrile/formic acid) over a period of 45 min and ramped from 55 to 80% acetonitrile over 10 min. The column eluent was electrosprayed into the mass spectrometer using a 1.8 kV spraying voltage. The spectra were acquired in the mass range from 200 to 2200 m/z. Data were processed using DataAnalysis v.3.4 (Bruker Daltonics, Bremen, Germany) and then subjected to MASCOT searching against the *Francisella tularensis* OSU18 database. The searching criteria were set as follows: trypsin/P was used as the protease, up to 2 missed cleavages were allowed, 1+, 2+ and 3+ ions, carbamidomethylation of cysteine as a fixed modification and oxidation of methionine as a variable modification. Data were then filtered with a ProteinParser v.2.1²¹ to reject peptides with a Mowse probability score threshold less than 30 and peptides containing KK, KR, KR or RR motifs. Only peptides containing more than six amino acids and/or a mass greater than 600 Da were accepted. MS/MS data were concurrently searched against the created randomized decoy database using the same parameters that were used for the target database searches. MASCOT results were filtered with a ProteinParser using the same parameters as for the target database searches. Filtered data from target and decoy database searches were mutually compared. Estimated false positive rate was calculated by dividing the number of incorrect identifications (decoy proteins) obtained from all filtered decoy searches by the number of correct identifications from all filtered target searches. FP rate is 1.8%.

Bioinformatic analysis

All the proteins identified in the present work were analyzed using following tools:

1. LipoP (<http://www.cbs.dtu.dk/services/LipoP/>) was used to identify lipoproteins,
2. NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and NetOGlyc (<http://www.cbs.dtu.dk/services/NetOGlyc/>) were used to predict N- and O-glycosylation sites, respectively,
3. EnsembleGly server (<http://turing.cs.iastate.edu/EnsembleGly/>) and GPP Hirst group glycosylation prediction server (<http://comp.chem.nottingham.ac.uk/glyko/>) were used to predict N- and O-linked glycosylation sites,
4. COG algorithm (www.ncbi.nlm.nih.gov/COG/old/xognitor.html) was used to classify the identified proteins into functional categories,
5. PSORTb v.2.0. program (<http://psort.psorb>) and SignalP server (<http://www.cbs.dtu.dk/services/SignalP/>) were used to predict protein localization,
6. Pfam database (<http://pfam.sanger.ac.uk/protein?acc=Q5NFW3>) was used to determine low complexity regions of proteins FTH_1071 and FTH_0414,
7. Conserved Domain Database²² (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) was used to search for the presence of conserved domains on selected proteins.

Results and discussion

Investigation of bacterial protein glycosylation represents a significant challenge to the current glycoproteomic methodologies. This is predominantly due to the presence of unique monosaccharide units within the bacterial glycan, such as bacillosamine, that are resistant to digestion with PNGase F, the enzyme that is employed to release eukaryotic N-linked glycans. In addition, the O-linked glycosylation that is more abundantly present in the bacterial domains, is significantly more difficult to analyze due to the lack of a universal enzyme to liberate these glycans from the protein backbone. Moreover, mammalian

glycosylation uses a conserved core structure, which is attached to the protein, which contrasts to the variable nature of glycans attached by bacteria. Therefore, the indirect identification of this modification has been accomplished by the use of periodate/hydrazide labelling and through binding to various lectins. The present study was performed following our recently published review²³ and the outline of workflow is depicted in Figure 1. The used techniques are aimed at looking at the global glycoproteome rather than characterizing glycosylation.

Glycoprotein Detection using Pro-Q Emerald 300 Glycoprotein Stain Kit

The fraction of enriched membrane proteins from *F. tularensis*, obtained from sodium carbonate extraction of the whole-cell lysates, was resolved by two-dimensional electrophoresis using the wide-range IPG strips pH 3 – 10 and narrow-range IPG strips pH 6 – 11. The presence of proteins modified by glycosylation was tested using a Pro-Q Emerald fluorescent carbohydrate-specific staining. This staining is based on a two-step reaction. First, diol groups of carbohydrates are oxidized to aldehydes. Subsequently, the aldehydes are reacted with a fluorescently-labeled hydrazide to form a stable hydrazone. Several spots were detected as carbohydrate-positive (Figure 2). Three independent experiments were performed in both pH ranges and proteins were considered as carbohydrate-positive if they were observed in at least two of the three experiments (Table 1). Seven of the ten identified proteins were predicted as lipoproteins, and 2 proteins as signal peptidase I-cleaved proteins using the LipoP algorithm.

Use of lectins for the detection of glycoproteins followed by affinity chromatography

The *F. tularensis* membrane-protein enriched fraction resolved by two-dimensional electrophoresis within the pH range 3 – 10 was evaluated for binding with four lectins differing in their specificities towards the glycan moieties of glycoproteins. The DIG Glycan Differentiation kit contains five lectins – *Sambucus nigra* agglutinin (SNA), *Maackia amurensis* agglutinin (MAA), *Peanut agglutinin* (PNA), *Datura stramonium* agglutinin (DSA) and *Galanthus nivalis* agglutinin (GNA). High-mannose glycans that interact with GNA are characteristic feature of yeast glycoproteins. Therefore, this lectin was excluded from the study. Specificity of the used lectins is as follows:

1. SNA recognizes primarily sialic acid linked (2–6) to galactose,
2. MAA distinctly reacts with sialic acid linked (2–3) to galactose,
3. PNA targets the core disaccharide Gal (1–3) GalNAc and is thus suitable for identifying O-glycosidically linked carbohydrate chains,
4. DSA recognizes structures with terminal Gal (1–4) GlcNAc that are associated with both N- and O- glycans, and GlcNAc in O-glycans.

A total of 20 proteins were identified using all lectin-based procedures, as shown in Figure 3. These proteins are listed in Table 2. Among them, 16 proteins were specific to SNA, 6 proteins interacted with MAA, 2 proteins were detected by DSA and 1 protein was recognized by PNA. An overlap of 4 proteins, namely FTH_1293, FTH_1598, FTH_1206, and FTH_1721, isolated by more than one lectin was observed. Moreover, the multiply-charged variants of the proteins FTH_1293, FTH_0159, FTH_0539, FTH_1598, FTH_1112, FTH_0311, FTH_0941, and FTH_1167 were detected, suggesting the existence of glycoforms with a diverse degree of glycosylation. Of the 20 identified proteins, two of them were predicted as lipoproteins and three as signal peptidase I-cleaved proteins using the LipoP algorithm.

It is apparent that the recognition patterns of lectins used in our study are fairly variable, permitting us to visualize the diverse nature of the putative bacterial glycoproteome.

However, the use of DIG Glycan kit allows detection of only highly abundant glycoproteins. Therefore, we performed the lectin affinity chromatography to further increase the probability of identifying less abundant glycoproteins. Based on the results obtained from the DIG glycan kit, the lectins SNA, PNA and DSA were chosen to perform affinity chromatography. In addition, *Conavalia ensiformis agglutinin* (ConA), a lectin with a broad specificity toward the mannose and glucose residues, and *Soybean agglutinin* (SBA), which exhibits affinity for GalNAc, were also used. An important matter that had to be taken into account was the composition of the bacterial sample. The cell envelope of Gram-negative bacteria contains lipopolysaccharide in its outer leaflet of the outer membrane. It was therefore highly desirable to remove this potential contaminant from the sample before performing analysis as it would bind preferentially to different lectins.

There has been considerable discussion among the glycobiologists concerning the issues of lectin non-specificity as the result of commonly occurring protein-protein interactions. In order to minimize the interference of nonspecifically-bound molecules, we used the appropriate sugar competitors for the specific elution of lectin-bound proteins rather than a general eluent such as weak acid. Despite these efforts, the possibility of retaining non-specifically bound proteins together with specifically-bound glycoproteins on the affinity resin still exists. Their presence may arise from their association with glycosylated proteins. Therefore, the proteins identified in this study are designated as “putative glycoproteins”, while their glycosylation status has yet to be confirmed structurally.

Categorization of lectin-isolated putative glycoproteins

The use of lectin affinity chromatography resulted in the identification of 104 proteins, in total, and these proteins are listed in Table 3. Only the proteins that were found in two independent experiments are further considered.

As seen in Figure 4, there was a considerable overlap of proteins among the lectins used in this study. Of the 104 identified proteins, 20.2% were eluted from single lectins, while others were isolated with two lectins (28.8%), three lectins (16.4%), four lectins (18.3%), or all five lectins (16.4%). Of the five lectins employed, the majority of proteins were captured with SBA lectin (86.5%), suggesting the presence of terminal GalNAc residues.

The COGnitor algorithm was used to classify all identified proteins into 17 functional categories. The distribution of proteins, in terms of their function, is illustrated in Figure 5, where a majority of proteins (19%) are involved in energy production and conversion. Of identified proteins, 13% are not related to any functional category, which includes hypothetical and uncharacterized proteins.

A fraction enriched in membrane proteins was prepared in this study. Despite the enrichment technique used here, it is possible that a number of non-membrane proteins could still be present in the processed samples. Therefore, we used PSORTb v.2.0.²⁴, the most precise bacterial localization tool available, with respect to the trilayer composition of the membrane of Gram-negative bacteria. Using this prediction program, the lectin-isolated proteins were categorized into one of the following localization sites: cytoplasm (31.7%), cytoplasmic membrane (24.0%), periplasm (1.0%), outer membrane (2.9%), and extracellular space (1.0%). The remaining proteins (39.4%) comprise those with unknown localization. However, PSORTb is not able to determine lipoproteins, which comprise an important class among the membrane proteins. For this reason, LipoP lipoprotein prediction program was additionally applied. Nevertheless, the exact protein localization remains to be experimentally verified.

Among the 104 identified proteins, 4 lipoproteins with a predicted periplasmic localization and 10 signal peptidase I-cleaved proteins were indicated through LipoP algorithm.

Biological importance of detected putative glycoproteins

There are two groups of enzymes – glycosyltransferases and glycosidases – that are involved in a process of glycosylation. Glycosyltransferases catalyse the transfer of monosaccharide residues from an activated donor to a growing carbohydrate chain, whereas glycosidases catalyse the hydrolysis of glycosidic linkages.²⁵ It has been found that glycosyltransferases in mammalian systems are widely glycosylated.^{26,27} Our results suggest that the bacterial enzymes of this type might be glycosylated as well. This is in accordance with the previous findings of glycosylated bacterial enzymes.²⁸ However, glycosylation does not have to necessarily play a direct role in enzymatic activity. In this study, glycosyltransferases were identified and are listed in Table S2 in Supporting data.

The presence of glycosylation is known to alter immunogenicity of proteins.²⁹ Intriguingly, several of the proteins identified in this study, namely hypothetical protein (FTH_0069), OmpA family protein (FTH_0323), outer membrane protein FopA (FTH_1293), glycerophosphodiester phosphodiesterase (FTH_1463), chaperonin GroEL (FTH_1651), succinate dehydrogenase (FTH_1722), acetyl-CoA carboxylase alpha subunit (FTH_0295), dihydrolipoyllysine-residue succinyltransferase (FTH_1719), DNA-binding protein HU-beta (FTH_0880), 17 kDa lipoprotein TUL4 precursor (FTH_0414), aconitate hydratase (FTH_1708), cell division protein FtsZ (FTH_1830), LemA-like protein (FTH_0357), dihydrolipoamide dehydrogenase (FTH_0312), bacterioferritin (FTH_0620), thioredoxin family protein (FTH_1071), dihydrolipoamide acetyltransferase (FTH_0311), molecular chaperone DnaK (FTH_1167), and elongation factor Tu (FTH_1691) have recently been reported as immunoreactive antigens in the live vaccine strain LVS of *F. tularensis* subsp. *holarctica* reacting with human tularemic sera, as demonstrated by Janovska *et al.*³⁰

It has been found that the outer membrane proteins (OMP) from some bacterial species are glycosylated.^{31–33} Similarly, we found several OMPs³⁴ - OmpA (FTH_0334), OmpA family protein (FTH_0323), FopA (FTH_1293), peptidylprolyl isomerase (FTH_1021), and TUL4 paralogs, lpnA (FTH_0414) and lpnB (FTH_0417) - that are also putative glycoproteins.

Several FPI-encoded proteins were further identified, including proteins PdpB (FTH_0117), hypothetical protein PigF (FTH_0111), hypothetical protein PigG (FTH_0110), and IgIB (FTH_0104). FPI is essential for intracellular growth and virulence of *F. tularensis*, as the disruption of FPI genes resulted in attenuation of the generated mutant bacteria for survival inside macrophages.⁷ Among the proteins that are required for pathogenicity of other Gram-negative bacteria, such as *Pseudomonas* species³⁵, was LemA-like protein (FTH_0357), herein identified by hydrazide labelling and in the eluted fractions of all used lectins.

PilA-encoded type IV pili fiber protein (FTH_0384) was identified by hydrazide labelling and also in the eluted fractions from all five lectins. This protein shares 52.5% similarity with *pilE1*-encoded pilin (FMM1_NEIGO) of *N. gonorrhoeae*, which has a Galβ1–3GlcNAc modification located at Ser70.³⁶ In addition, another *pilB*-encoded type IV pili assembly protein (FTH_0818) was found as potentially modified through glycosylation. These findings support the role of pilin glycosylation in the host-cell adhesion and virulence of *F. tularensis*. Both, *PilA* and *PilB* are components of the type IV pili adhesive structure that plays a pivotal role in virulence for many pathogenic bacteria. Most certainly, in *P. aeruginosa*, *PilA* is the major subunit of the pilin filament, whereas *PilB* is an inner-membrane ATPase involved in extension and retraction of the pilus. In *F. tularensis*, *PilA*

was recently shown to be a potential component of type IV pili that is critical for virulence of *Francisella* via a subcutaneous route of infection.¹⁷

A simultaneous, dual post-translational modification, such as acylation and glycosylation, has been observed in *Mycobacterium* species.³⁷ In our study, the probable thioredoxin family protein FTH_1071, a lipoprotein predicted to be localized in the periplasm, was identified by hydrazide labelling and lectin affinity chromatography. FTH_1071 was predicted to have a DsbA-like thioredoxin domain (Pfam 01323) and was recently found to be essential for virulence of both *F. tularensis* subspecies, *holarctica* and *tularensis*.^{38,39} Interestingly, the Ng1717 periplasmic lipoprotein of *N. gonorrhoeae*, also an isoform of oxidoreductase DsbA, was found to be modified with an O-AcHexDATDH glycan.¹³ The protein FTH_1071 together with another identified 17 kDa lipoprotein TUL4 precursor (FTH_0414) were previously found to interact with TLR2/TLR1 heterodimers of HeLa cell lines, resulting in the induction of a proinflammatory response.⁴⁰ The molecular aspect of the interaction of these lipoproteins with TLR is not clear, but it is probable that it is mediated via acylation, whereas the glycosylation might be involved in the activation of an immune response similar to that described by Sieling *et al.*⁴¹ They demonstrated that glycosylation of mycobacterial lipoglycoprotein LprG is required for the stimulation of innate immune responses via activation of MHC class II-restricted T cells.

Among the proteins detected by hydrazide labelling, the putative uncharacterized protein FTH_0069 is a protein ortholog of FTT1676, which has been recently identified as a novel determinant of *Francisella* virulence. Deletion of the FTT1676 abolished the ability of SchuS4 to survive or proliferate intracellularly and thus cause lethality in mice.⁴²

Bioinformatic studies on detected putative glycoproteins

All identified proteins were examined for the presence of a potential eukaryotic N-glycosylation motif represented by the N-XS/T sequon that is necessary but, in case of bacteria, not sufficient for glycosylation.^{43,44} For that purpose, the NetNGlyc program was used. Glycosylation was predicted for those asparagines that occurred within the N-X-S/T sequon for which the N-glycosylation potential crossed the default threshold of 0.5. Additionally, at least 4 of 9 networks of jury agreement supported this prediction. It is necessary to note that all currently available prediction methods were developed for identification of glycosylation sites in mammalian proteins and thus use the rules that might not be directly applicable to prokaryotic systems. Therefore, N-glycosylation sequons of all identified proteins predicted by NetNGlyc were manually inspected for the occurrence of prokaryotic D/E-X_a-N-X_b-S/T sequon.⁴⁵ Up to 18 of all identified proteins contain this extended glycosylation motif, with proteins dTDP-glucose 4,6-dehydratase (FTH_0592) and ribonuclease E (FTH_0719) having two of these motifs. The proteins with no predicted N-glycosylation site are most likely modified through O-glycosylation, or, possibly, they could be part of the non-glycosylated proteome that was, as a result of protein association, co-isolated together with the glycoproteins.

In the case of the proteins FTH_1071 and FTH_0414, a more comprehensive searching for the presence of O-glycosylation sites was performed, using five different prediction tools. The sites predicted as O-glycosylated by at least two methods are summarized in Tables S3 (a and b) in Supporting information. The amino acids Thr34, Thr45, Thr46, Ser32, Ser37, and Ser41 were predicted by at least four methods, representing the best candidates for O-glycosylation sites in the protein FTH_1071. For FTH_0414, O-glycosylations on Thr45 and Thr46 were predicted by four methods. The prediction of N-glycosylation was not taken into consideration due to the primary absence of D/E-X_a-N-X_b-S/T extended sequon in both proteins.

It has been reported that glycan occupancy sites are often associated with low-complexity regions (LCR) within proteins that are of biased composition and consist of different kinds of repeats such as alanine, serine, and proline residues.¹³ Therefore, we used Pfam protein families database to determine such LCRs and compared the position of predicted O-glycosylation sites with the position of LCRs within FTH_1071 and FTH_0414. Indeed, all six predicted O-glycosylation sites in FTH_1071, both predicted as O-glycosylated sites in FTH_0414, occur within the LCRs (Figure 6).

Glycomic studies

Our efforts to define the glycan structures that would definitively confirm the glycosylation, have been hindered by our limitations to completely remove lipopolysaccharide from the membrane protein-rich fraction prior to β -elimination. Instead, O-antigen of LPS consisting of rare sugars 2-acetamido-2,6-dideoxy-D-glucose (QuiNAc), 4,6-dideoxy-4-formamido-D-glucose (Qui4NFm) and 2-acetamido-2-deoxy-D-galacturonamide (GalNAcAN) was observed as a result of the action of the used reducing conditions (Supporting information, Spectrum S2).⁴⁶ Thus, the nature of glycans modifying *F. tularensis* proteins and their functions at the molecular level remains to be elucidated.

Conclusions

Studying bacterial glycoproteins has gained importance due to the recently revealed role of these proteins in the host-pathogen interactions. The study presented here utilized a bottom-up mass-spectrometric approach, in which the presence of the *F. tularensis* glycoproteome was investigated by the carbohydrate-specific detection method and affinity of various lectins. Up to 20 putative glycoproteins were detected using fluorescently labelled hydrazide and lectin blotting, while the use of lectin affinity chromatography resulted in the identification of 104 putative *F. tularensis* subsp. *holarctica* glycoproteins.

In total, 15 proteins were identified with confidence in at least two of the applied approaches. Protein FTH_0069 was detected with both hydrazide labelling and lectin blotting. In contrast, the proteins FTH_1071, FTH_0414, FTH_0384, and FTH_0357 were identified with both hydrazide labelling and lectin affinity chromatography. Proteins identified by lectin labelling and lectin affinity chromatography were FTH_1830, FTH_0159, FTH_1855, FTH_0539, FTH_1112, FTH_0311, FTH_1167, FTH_1761, and FTH_1721. Finally, FTH_1293 was identified using all methods. These proteins represent the best candidates for *F. tularensis* glycosylation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was financially supported by Ministry of Education No. MSMT0021627502 and No. ME08105, Ministry of Defence, Czech Republic No. FVZ0000604 and No. OVUOFVZ200808, and Czech Science Foundation No. GA203/09/0857. Authors wish to thank Jitka Zakova, Milan Madera, and Iveta Klouckova for their excellent technical support, and also William R. Alley, Jr. for a fruitful discussion during the realization of this project. These studies were facilitated through collaborative research with the National Center for Glycomics and Glycoproteomics at Indiana University, which has been supported by grant No. RR018942 from NCRR, U.S. Department of Health and Human Services.

References

1. Oyston PC, Sjøstedt A, Titball RW. Nat Rev Microbiol 2004;2:967–78. [PubMed: 15550942]

2. Dennis DT, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Friedlander AM, Hauer J, Layton M, Lillibridge SR, McDade JE, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, Tonat K. *Jama* 2001;285:2763–73. [PubMed: 11386933]
3. Saslaw S, Eigelsbach HT, Wilson HE, Prior JA, Carhart S. *Arch Intern Med* 1961;107:689–701. [PubMed: 13746668]
4. Saslaw S, Eigelsbach HT, Prior JA, Wilson HE, Carhart S. *Arch Intern Med* 1961;107:702–14. [PubMed: 13746667]
5. Johansson A, Farlow J, Larsson P, Dukerich M, Chambers E, Bystrom M, Fox J, Chu M, Forsman M, Sjostedt A, Keim P. *J Bacteriol* 2004;186:5808–18. [PubMed: 15317786]
6. Svensson K, Larsson P, Johansson D, Bystrom M, Forsman M, Johansson A. *J Bacteriol* 2005;187:3903–8. [PubMed: 15901721]
7. Nano FE, Zhang N, Cowley SC, Klose KE, Cheung KK, Roberts MJ, Ludu JS, Letendre GW, Meierovics AI, Stephens G, Elkins KL. *J Bacteriol* 2004;186:6430–6. [PubMed: 15375123]
8. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. *Mol Microbiol* 1999;32:1022–30. [PubMed: 10361304]
9. Castric P, Cassels FJ, Carlson RW. *J Biol Chem* 2001;276:26479–85. [PubMed: 11342554]
10. Stimson E, Virji M, Makepeace K, Dell A, Morris HR, Payne G, Saunders JR, Jennings MP, Barker S, Panico M, et al. *Mol Microbiol* 1995;17:1201–14. [PubMed: 8594338]
11. Hegge FT, Hitchen PG, Aas FE, Kristiansen H, Lovold C, Egge-Jacobsen W, Panico M, Leong WY, Bull V, Virji M, Morris HR, Dell A, Koomey M. *Proc Natl Acad Sci U S A* 2004;101:10798–803. [PubMed: 15249686]
12. Dobos KM, Khoo KH, Swiderek KM, Brennan PJ, Belisle JT. *J Bacteriol* 1996;178:2498–506. [PubMed: 8626314]
13. Vik A, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, Koomey M. *Proc Natl Acad Sci U S A* 2009;106:4447–52. [PubMed: 19251655]
14. Szymanski CM, Burr DH, Guerry P. *Infect Immun* 2002;70:2242–4. [PubMed: 11895996]
15. Karlyshev AV, Everest P, Linton D, Cawthraw S, Newell DG, Wren BW. *Microbiology* 2004;150:1957–64. [PubMed: 15184581]
16. Aas FE, Egge-Jacobsen W, Winther-Larsen HC, Lovold C, Hitchen PG, Dell A, Koomey M. *J Biol Chem* 2006;281:27712–23. [PubMed: 16825186]
17. Forslund AL, Kuoppa K, Svensson K, Salomonsson E, Johansson A, Bystrom M, Oyston PC, Michell SL, Titball RW, Noppa L, Frithz-Lindsten E, Forsman M, Forsberg A. *Mol Microbiol* 2006;59:1818–30. [PubMed: 16553886]
18. Faridmoayer A, Fentabil MA, Haurat MF, Yi W, Woodward R, Wang PG, Feldman MF. *J Biol Chem* 2008;283:34596–604. [PubMed: 18930921]
19. Salomonsson E, Forsberg A, Roos N, Holz C, Maier B, Koomey M, Winther-Larsen HC. *Microbiology* 2009;155:2546–59. [PubMed: 19423631]
20. Molloy MP, Herbert BR, Slade MB, Rabilloud T, Nouwens AS, Williams KL, Gooley AA. *Eur J Biochem* 2000;267:2871–81. [PubMed: 10806384]
21. Mann B, Madera M, Sheng Q, Tang H, Mechref Y, Novotny MV. *Rapid Commun Mass Spectrom* 2008;22:3823–34. [PubMed: 18985620]
22. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH. *Nucleic Acids Res* 2009;37:D205–10. [PubMed: 18984618]
23. Balonova L, Hernychova L, Bilkova Z. *Expert Rev Proteomics* 2009;6:75–85. [PubMed: 19210128]
24. Gady JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS. *Bioinformatics* 2005;21:617–23. [PubMed: 15501914]
25. Taylor, ME.; Drickamer, K. *Introduction to glycobiology*. 2. Oxford University Press, Inc; New York: 2006. p. 3-16.

26. Chen R, Jiang X, Sun D, Han G, Wang F, Ye M, Wang L, Zou H. *J Proteome Res* 2009;8:651–61. [PubMed: 19159218]
27. D'Agostaro G, Bendiak B, Tropak M. *Eur J Biochem* 1989;183:211–7. [PubMed: 2502398]
28. Plummer TH Jr, Tarentino AL, Hauer CR. *J Biol Chem* 1995;270:13192–6. [PubMed: 7768916]
29. Baumeister W, Lembcke G. *J Bioenerg Biomembr* 1992;24:567–75. [PubMed: 1459988]
30. Janovska S, Pavkova I, Hubalek M, Lenco J, Macela A, Stulik J. *Immunol Lett* 2007;108:151–9. [PubMed: 17241671]
31. Scott NE, Bogema DR, Connolly AM, Falconer L, Djordjevic SP, Cordwell SJ. *J Proteome Res* 2009;8:4654–64. [PubMed: 19689120]
32. Ku SC, Schulz BL, Power PM, Jennings MP. *Biochem Biophys Res Commun* 2009;378:84–9. [PubMed: 19013435]
33. Gonzalez-Zamorano M, Mendoza-Hernandez G, Xolalpa W, Parada C, Vallecillo AJ, Bigi F, Espitia C. *J Proteome Res* 2009;8:721–33. [PubMed: 19196185]
34. Huntley JF, Conley PG, Hagman KE, Norgard MV. *J Bacteriol* 2007;189:561–74. [PubMed: 17114266]
35. Hrabak EM, Willis DK. *J Bacteriol* 1992;174:3011–20. [PubMed: 1314807]
36. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA. *Nature* 1995;378:32–8. [PubMed: 7477282]
37. Belisle, JT.; Braunstein, M.; Rosenkrands, I.; Andersen, P. *Tuberculosis: Pathogenesis, Protection, and Control*. Cole, ST.; Eisenach, KD.; McMurray, DN.; Jacobs, WR., editors. ASM Press; Washington, D.C: 2005. p. 235-260.
38. Qin A, Scott DW, Thompson JA, Mann BJ. *Infect Immun* 2009;77:152–61. [PubMed: 18981253]
39. Straskova A, Pavkova I, Link M, Forslund AL, Kuoppa K, Noppa L, Kroca M, Fucikova A, Klimentova J, Krocova Z, Forsberg A, Stulik J. *J Proteome Res* 2009;8:5336–46. [PubMed: 19799467]
40. Thakran S, Li H, Lavine CL, Miller MA, Bina JE, Bina XR, Re F. *J Biol Chem* 2008;283:3751–60. [PubMed: 18079113]
41. Sieling PA, Hill PJ, Dobos KM, Brookman K, Kuhlman AM, Fabri M, Krutzik SR, Rea TH, Heaslip DG, Belisle JT, Modlin RL. *J Immunol* 2008;180:5833–42. [PubMed: 18424702]
42. Wehrly TD, Chong A, Virtaneva K, Sturdevant DE, Child R, Edwards JA, Brouwer D, Nair V, Fischer ER, Wicke L, Curda AJ, Kupko JJ 3rd, Martens C, Crane DD, Bosio CM, Porcella SF, Celli J. *Cell Microbiol* 2009;11:1128–50. [PubMed: 19388904]
43. Nita-Lazar M, Wacker M, Schegg B, Amber S, Aebi M. *Glycobiology* 2005;15:361–7. [PubMed: 15574802]
44. Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano MA, Whitfield C, Aebi M. *Proc Natl Acad Sci U S A* 2006;103:7088–93. [PubMed: 16641107]
45. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi M. *Embo J* 2006;25:1957–66. [PubMed: 16619027]
46. Vinogradov EV, Shashkov AS, Knirel YA, Kochetkov NK, Tochtamysheva NV, Averin SF, Goncharova OV, Khlebnikov VS. *Carbohydr Res* 1991;214:289–97. [PubMed: 1769021]

Figure 1.
The experimental workflow performed for studying the *F. tularensis* glycoproteome.

Figure 2.

Detection of glycoproteins using the Pro-Q Emerald dye. (A) Mini two-dimensional electrophoresis in a wide pH range from 3 – 10. (B) Mini two-dimensional electrophoresis in a basic pH range from 6 – 11. The molecular weight standard Candy Cane (Invitrogen, Eugene, OR) consists of four glycosylated and four nonglycosylated proteins, α_2 -macroglobulin (180 kDa), phosphorylase b (97 kDa), glucose oxidase (82 kDa), bovine serum albumin (66 kDa), α_1 -acid glycoprotein (42 kDa), carbonic anhydrase (29 kDa), avidin (18 kDa), and lysozyme (14 kDa).

Figure 3.

Detection of glycoproteins with lectins using DIG glycan differentiation kit. (A) Detection of glycoproteins using SNA lectin. (B) Detection of glycoproteins using MAA lectin. (C) Detection of glycoproteins using DSA lectin. (D) Detection of glycoproteins using PNA lectin. The spots marked with asterisks indicate the proteins that were not identified.

Figure 4. Distribution of lectin-isolated *F. tularensis* proteins among five different lectins. The numbers in columns represent the percentage of identified proteins.

Figure 5.
Distribution of lectin-isolated *F. tularensis* proteins to their functional categories.
Information was resourced from COGnitor algorithm
(www.ncbi.nlm.nih.gov/COG/old/xognitor.html).

Figure 6. Domain organization and structural architecture of (A) FTH_1071 and (B) FTH_0414. Predicted O-glycosylated sites are highlighted red. Sig.p = signal peptide, LCR = low-complexity region.

Table 1

List of putative glycoproteins detected by Pro-Q Emerald staining

Spot no.	Gene locus ^a	MW [kDa]	pI	pH range	GlycoSite ^b	GlycoSite ^c	PSORTb ^d	LipoP ^e
1,2	FTH_0069	37.17	7.62	3–10	0	-	?	SPII
3	FTH_0323	46.95	6.27	3–10	5	-	?	SPII
4,5	FTH_1071	39.69	4.89	3–10	3	-	?	SPII
6	not identified							
7	FTH_0414	15.77	9.67	3–10	0	-	?	SPII
8	FTH_0417	17.40	4.91	3–10	2	-	?	SPII
8	FTH_0317	12.32	4.96	3–10	2	-	?	SPII
9	FTH_0384	13.59	9.06	3–10	2	-	?	SPI
11	FTH_1293	41.45	5.58	3–10	4	DSN ₃₀₉ IS	OM	SPI
10	FTH_0646	14.93	8.92	3–10	3	-	?	SPII
12	FTH_0357	21.97	5.63	6–11	2	-	cyt	-
13	FTH_0572	22.49	9.14	6–11	2	-	?	SPII

^a the accession number in the genome sequence of *F. tularensis* subsp. *holarctica* OSU18

^b the number of eukaryotic N-glycosylation motifs obtained from NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

^c prokaryotic N-glycosylation motif obtained by manual inspection of protein sequence (the index indicates the position of asparagine within the protein sequence)

^d prediction of the protein localization using PSORTb program (<http://psort.org/psortb>), cyt – cytoplasmic, CM – cytoplasmic membrane, OM – outer membrane, PP – periplasm, EC – extracellular space, ? – unknown localization

^e prediction of lipoproteins (SPII cleavage site ID) and SPI (cleavage site I) using LipoP algorithm (<http://www.cbs.dtu.dk/services/LipoP/>)

The accession numbers written in bold represent immunoreacted antigens (from study by Janovska *et al*(30)).

Table 2

List of DIG Glycan detected putative glycoproteins

Spot no.	Gene locus ^a	Lectin	MW [kDa]	pI	GlycoSite ^b	GlycoSite ^c	PSORTb ^d	Lipo ^e
1	FTH_1830	SNA	39.8	4.76	4	DFN ₃ DS	cyt	-
2,3,5	FTH_1293	SNA MAA	41.3	5.59	4	DSN ₃₀₉ IS	OM	SPI
9	FTH_0927	SNA	35.2	5.68	2	-	cyt	-
11,12	FTH_0159	SNA	30.4	5.78	2	-	cyt	-
13	FTH_0738	SNA	24.6	5.61	0	-	?	-
14	FTH_1206	SNA	27.4	5.80	1	-	?	-
15,16	FTH_1598	SNA MAA DSA	36.1	6.48	2	DRN ₁₉₃ NT	?	-
17	FTH_1855	SNA	33.7	5.46	2	-	?	SPI
18,29	FTH_0539	SNA	35.8	5.80	4	-	cyt	-
19	FTH_0069	SNA	37.2	7.62	0	-	?	SPII
20	FTH_0611	SNA	47.2	5.70	1	-	cyt	-
21,30,31	FTH_0941	SNA	51.3	5.88	2	-	cyt	-
22,23	FTH_1112	SNA	44.3	5.62	1	-	cyt	-
24,25	FTH_0311	SNA	56.9	5.08	0	-	CM	-
26,27,28	FTH_1167	SNA	69.4	4.88	3	ERN ₄₁₇ TT	PP	-
32	FTH_1761	SNA	46.8	5.61	2	-	cyt	-
33	FTH_0516	MAA	30.8	6.85	0	-	cyt	-
34	FTH_1021	MAA	29.6	9.04	4	-	OM	SPII
35	FTH_1721	MAA PNA	27.2	8.42	1	-	cyt	-
36	FTH_1463	MAA	39.1	5.39	1	DIN ₆₁ MT	?	SPI

^a the accession number in the genome sequence of *F. tularensis* subsp. *holarctica* OSU18

^b the number of eukaryotic N-glycosylation motifs obtained from NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

^c prokaryotic N-glycosylation motif obtained by manual inspection of protein sequence (the index indicates the position of asparagine within the protein sequence)

^d prediction of the protein localization using PSORTb program (<http://psort.org/psortb>), cyt – cytoplasmic, CM – outer membrane, PP – periplasm, EC – extracellular space, ? – unknown localization

^e prediction of lipoproteins (SPII cleavage site II) and SPII (cleavage site I) using LipoP algorithm (<http://www.cbs.dtu.dk/services/LipoP/>)

The accession numbers written in bold represent immunoreacted antigens (from study by Janovska *et al*30).

Table 3

List of identified proteins isolated using lectin affinity chromatography

Gene locus ^a	Lectin	MW [kDa]	pI	ΣGlycoSite ^b	GlycoSite ^c	PSORTb ^d	Lipo ^e
FTH_1651	ConA, DSA, PNA, SBA, SNA	57.40	4.72	3	-	cyt	-
FTH_1722	ConA, DSA, PNA, SBA, SNA	65.86	6.14	1	DVN ₅₈₁ MS	?	-
FTH_0310	ConA, DSA, PNA, SBA, SNA	100.27	5.65	1	-	cyt	-
FTH_0295	ConA, DSA, PNA, SBA, SNA	35.44	8.07	1	-	?	-
FTH_1719	ConA, DSA, PNA, SBA, SNA	52.75	4.89	2	-	cyt	-
FTH_0384	ConA, DSA, PNA, SBA, SNA	13.48	9.22	2	-	?	SPI
FTH_1732	ConA, DSA, PNA, SBA, SNA	49.87	4.89	0	-	cyt	-
FTH_0159	ConA, DSA, PNA, SBA, SNA	30.22	6.10	2	-	cyt	-
FTH_0880	ConA, DSA, PNA, SBA, SNA	9.47	10.59	1	-	?	-
FTH_0414	ConA, DSA, PNA, SBA, SNA	15.77	9.67	0	-	?	SPII
FTH_1708	ConA, DSA, PNA, SBA, SNA	102.70	5.30	4	-	cyt	-
FTH_1721	ConA, DSA, PNA, SBA, SNA	26.57	8.18	1	-	cyt	-
FTH_1503	ConA, DSA, PNA, SBA, SNA	69.85	10.14	4	-	CM	-
FTH_0570	ConA, DSA, PNA, SBA, SNA	19.79	9.74	1	-	?	SPI
FTH_1830	ConA, DSA, PNA, SBA, SNA	39.75	4.49	4	DFN ₃ DS	cyt	-
FTH_0357	ConA, DSA, PNA, SBA, SNA	21.99	5.43	2	-	cyt	-
FTH_0039	ConA, DSA, PNA, SBA, SNA	15.30	5.16	1	-	?	SPII
FTH_0219	ConA, DSA, PNA, SBA	26.42	8.94	0	-	?	-
FTH_1113	ConA, DSA, PNA, SNA	12.89	4.28	1	EKN ₁₈ MS	cyt	-
FTH_0228	ConA, DSA, PNA, SNA	17.82	10.82	0	-	cyt	-
FTH_1617	ConA, DSA, PNA, SNA	38.46	9.50	5	-	CM	SPI
FTH_0334	ConA, DSA, PNA, SNA	23.29	4.95	0	-	OM	SPII
FTH_0312	ConA, DSA, PNA, SNA	50.53	5.87	0	-	cyt	-
FTH_0620	ConA, DSA, PNA, SNA	18.46	5.83	0	-	cyt	-
FTH_1293	ConA, DSA, PNA, SNA	41.36	5.43	4	DSN ₃₀₉ IS	OM	SPI
FTH_1760	ConA, DSA, SBA, SNA	87.36	5.09	4	-	?	-
FTH_0592	ConA, DSA, SBA, SNA	65.71	8.17	4	DDN ₁₅₈ ET EEN ₃₅₁ IS	CM	-

Gene locus ^d	Lectin	MW [kDa]	pI	ΣGlycoSite ^b	GlycoSite ^c	PSORTb ^d	Lipo ^e
FTH_1377	ConA, DSA, SBA, SNA	44.26	9.77	1	-	?	-
FTH_0836	ConA, DSA, SBA, SNA	12.87	10.55	1	-	?	-
FTH_1071	ConA, DSA, SBA, SNA	39.55	4.67	3	-	?	SPI
FTH_1686	DSA, PNA, SBA, SNA	18.73	9.54	0	-	?	-
FTH_1734	DSA, PNA, SBA, SNA	55.54	4.68	0	-	?	-
FTH_0236	DSA, PNA, SBA, SNA	12.20	10.80	0	-	?	-
FTH_1837	DSA, PNA, SBA, SNA	21.08	7.07	1	ESN ₂ LS	CM	-
FTH_0098	DSA, PNA, SBA, SNA	13.73	9.02	3	-	?	SPI
FTH_0311	DSA, PNA, SBA, SNA	56.79	4.79	1	-	CM	-
FTH_0827	ConA, DSA, SBA	39.19	9.02	2	DEN ₆₈ IT	cyt	-
FTH_0257	ConA, DSA, SBA	16.78	10.72	1	-	?	-
FTH_1761	ConA, DSA, SBA	46.28	5.55	2	-	cyt	-
FTH_0604	ConA, DSA, SBA	32.98	6.51	4	-	?	-
FTH_1855	ConA, DSA, SBA	33.82	5.74	2	-	?	SPI
FTH_0172	ConA, DSA, SBA	61.95	8.74	5	ETN ₇₂ FS	CM	SPI
FTH_1612	ConA, DSA, SBA	50.09	10.12	3	-	CM	-
FTH_0593	ConA, DSA, SBA	23.78	10.21	0	-	CM	SPI
FTH_1167	ConA, DSA, SNA	69.18	4.62	3	ERN ₄₁₇ TT	PP	-
FTH_0253	ConA, DSA, SNA	13.38	11.54	0	-	cyt	-
FTH_1764	ConA, DSA, SNA	25.20	6.95	0	-	cyt	-
FTH_0719	DSA, PNA, SBA	95.95	7.12	6	DVN ₃₀₂ SS ETN ₇₉₁ QT	cyt	-
FTH_0232	DSA, PNA, SBA	22.55	10.29	0	-	?	-
FTH_1691	DSA, SBA, SNA	43.39	4.87	1	-	cyt	-
FTH_1216	DSA, SBA, SNA	50.31	10.31	0	-	?	-
FTH_1763	DSA, SBA, SNA	47.59	7.06	2	-	cyt	-
FTH_0187	PNA, SBA, SNA	76.20	8.58	1	-	CM	-
FTH_0104	DSA, PNA	58.90	4.48	3	-	?	-
FTH_0234	DSA, PNA	30.40	11.54	1	-	?	-
FTH_0585	DSA, SBA	84.19	8.45	3	-	?	-
FTH_1599	DSA, SBA	49.32	4.33	2	-	?	SPI

Gene locus ^d	Lectin	MW [kDa]	pI	ΣGlycoSite ^b	GlycoSite ^c	PSORTb ^d	Lipo ^e
FTH_1733	DSA, SBA	33.19	8.87	2	-	?	-
FTH_1424	DSA, SBA	70.74	5.42	2	DTN ₃₆ GS	CM	-
FTH_1696	DSA, SBA	57.78	8.36	1	-	?	-
FTH_0447	DSA, SBA	32.13	9.72	0	-	?	-
FTH_0886	DSA, SBA	40.42	8.97	3	DDN ₄₁ ES	?	-
FTH_1736	DSA, SBA	17.38	7.54	1	DIN ₄ IT	cyt	-
FTH_0826	DSA, SBA	53.44	9.24	4	-	cyt	-
FTH_0139	DSA, SBA	49.45	4.72	1	-	cyt	-
FTH_1609	DSA, SBA	66.68	8.69	6	DGN ₃₇₁ VT	CM	SPI
FTH_1536	DSA, SBA	25.35	6.96	4	-	?	-
FTH_0184	DSA, SBA	64.28	7.63	3	-	CM	-
FTH_1735	DSA, SBA	19.20	6.11	1	-	cyt	-
FTH_1558	DSA, SBA	36.08	8.69	1	-	CM	-
FTH_1379	DSA, SBA	44.90	7.41	2	-	?	-
FTH_0887	DSA, SBA	34.59	9.97	1	-	?	-
FTH_1662	DSA, SBA	23.83	9.93	3	-	?	-
FTH_0543	DSA, SBA	22.44	5.34	0	-	?	-
FTH_0589	DSA, SBA	33.26	9.53	0	-	?	-
FTH_0111	DSA, SBA	24.59	4.86	2	-	?	-
FTH_0186	DSA, SBA	34.46	7.02	2	-	CM	-
FTH_0628	DSA, SBA	35.58	7.16	0	-	?	-
FTH_0838	DSA, SBA	34.48	6.28	2	-	CM	-
FTH_0837	DSA, SBA	69.67	9.52	2	-	CM	-
FTH_0110	DSA, SBA	44.64	4.55	5	-	?	-
FTH_1117	DSA, SBA	37.84	9.62	2	-	?	-
FTH_0423	SBA, SNA	33.44	7.53	0	-	EC	-
FTH_0151	DSA	25.39	6.06	2	EDN ₄₄ LT	cyt	-
FTH_1112	DSA	44.02	5.72	1	-	cyt	-
FTH_1078	SBA	16.87	7.40	0	-	?	-
FTH_0327	SBA	30.12	7.48	1	-	CM	-
FTH_0853	SBA	50.85	9.46	0	-	CM	-

Gene locus ^a	Lectin	MW [kDa]	pI	Σ GlycoSite ^b	GlycoSite ^c	PSORTb ^d	LipoP ^e
FTH_1329	SBA	26.43	8.75	1	-	cyt	-
FTH_1726	SBA	47.16	9.78	1	-	CM	-
FTH_0539	SBA	35.44	6.05	4	-	cyt	-
FTH_0828	SBA	21.78	10.43	1	-	CM	-
FTH_1462	SBA	48.02	8.51	2	-	CM	-
FTH_0541	SBA	28.13	8.37	0	-	cyt	-
FTH_0251	SBA	48.46	10.01	0	-	CM	-
FTH_1168	SBA	43.99	8.35	4	-	cyt	-
FTH_1245	SBA	34.91	6.97	2	-	cyt	-
FTH_0799	SBA	150.39	7.97	4	ESN ₁₁₄₉ IS	cyt	-
FTH_1373	SBA	40.83	10.24	2	-	CM	-
FTH_0174	SBA	36.14	9.83	0	-	CM	-
FTH_1478	SBA	45.43	8.25	2	-	CM	-
FTH_1047	SBA	67.39	7.25	1	-	cyt	-
FTH_0117	SBA	127.47	9.63	8	-	OM	-
FTH_0397	SBA	28.00	8.33	2	-	?	-

^a the accession number in the genome sequence of *F. tularensis* subsp. *holarctica* OSU18

^b the number of eukaryotic N-glycosylation motifs obtained from NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>)

^c prokaryotic N-glycosylation motif obtained by manual inspection of protein sequence (the index indicates the position of asparagine within the protein sequence)

^d prediction of the protein localization using PSORTb program (<http://psort.org/psortb>), cyt – cytoplasmic, CM cytoplasmic membrane, OM – outer membrane, PP – periplasm, EC – extracellular space, ? – unknown localization

^e prediction of lipoproteins (SPII cleavage site ID) and SPI (cleavage site D) using LipoP algorithm (<http://www.cbs.dtu.dk/services/LipoP/>)

The accession numbers written in bold represent immunoreacted antigens (from study by Janovska *et al*(30)).