In silico Prediction and Experimental Confirmation of B-cell Epitopes for Molecular Serology of *Chlamydia* spp.

by

Kh Shamsur Rahman

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 8, 2016

Keywords:
Chlamydia species-specific Serology, Immunodominant B-cell Epitopes, B-cell Epitope *in silico* Prediction, Length of Peptide B-cell Epitopes, Protein Disorder Tendency

Approved by

Bernhard Kaltenboeck, Chair, Professor, Pathobiology
Calvin M Johnson, Dean and Professor, Veterinary Medicine Administration
Richard Curtis Bird, Professor, Pathobiology
Frederik W van Ginkel, Associate Professor, Pathobiology

Abstract

Urgently needed species-specific ELISAs for detection of antibodies against *Chlamydia (C.)* spp. have been elusive due to high cross-reactivity of chlamydial antigens. To identify *Chlamydia* species-specific B-cell epitopes, we ranked potential epitopes of immunodominant chlamydial proteins that are polymorphic among all *Chlamydia* species. High-scoring peptides were synthesized with N-terminal biotin followed by a serine-glycine-serine-glycine spacer, immobilized onto streptavidin-coated microtiter plates, and tested with mono-specific mouse hyperimmune sera against each *Chlamydia* species in chemiluminescent ELISAs. For each of nine *Chlamydia* species, 3-9 dominant polymorphic B-cell epitope regions were identified on OmpA, CT618, PmpD, IncA, CT529, CT442, IncG, Omp2, TarP, and IncE proteins. Sixteen-40 amino acid-long peptides of these epitopes reacted highly and with absolute specificity only with homologous, *Chlamydia* mono-species-specific sera. The probability of cross-reactivity of closely related peptide antigens correlated with percent sequence identity, and declined to zero at less than 50% sequence identity.

In the course of this investigation, a failure to accurately predict B-cell epitopes was observed for such prediction algorithms. Using our database of confirmed chlamydial B-cell epitopes and non-epitopes, we sought to understand the reasons for this failure. In our investigation, short 7-12aa peptides of B-cell epitopes bound antibodies poorly, thus epitope mapping with short peptide antigens would have falsely classified many of these epitopes as non-epitopes. We also show in published datasets of confirmed epitopes and non-epitopes a direct correlation between length of peptide antigens and antibody binding. Elimination of short, ≤11aa epitope/non-epitope sequences improved public datasets for evaluation of *in silico* B-cell epitope prediction. Following evaluation of a comprehensive set of algorithms for prediction of protein properties, we show that protein disorder tendency best describes B-cell epitopes. For B-cell epitope identification from a protein with 86% accuracy, we recommend using the 25aa moving average plot of the IUPred-L disorder score, and selecting 16-30aa peptides of peak regions for laboratory testing.

In conclusion, we have developed an accurate approach for B-cell epitope prediction and have applied it to identification of highly specific peptide antigens for molecular serology of *Chlamydia* spp. The combined approach also lends itself to identification of relevant epitopes of other microbial pathogens.

Acknowledgments

I am indebted to my major advisor, Dr. Bernhard Kaltenboeck for his guidance and support throughout my doctoral studies. My particular reminiscence are stirring inspirations of Dr. Kaltenboeck, which have encouraged me to carry out this project to the very end. Just to name a few areas of his immense support, I want to mention research planning, experimental design, laboratory techniques, bioinformatics, data analysis, data presentation, and manuscript writing. Whenever I was in despair about overcoming difficulties in completing this project successfully, Dr. Kaltenboeck gave me both solicited and unsolicited guidance with remarkable encouraging words - "Don't worry! Just do it and figure it out as you go. Don't be dogmatic. Because our dogmas don't cover everything that nature does …".

For guidance, I would like to express my deep appreciation to all members of my advisory committee, Dr. Calvin Johnson, Dr. Richard Curtis Bird, and Dr. Frederik van Ginkel.

I would like to thank our collaborator Dr. Konrad Sachse and his lab members for providing yolk-sac stocks of several *Chlamydia* species for inoculation of mice for raising hyper-immune sera. In addition, I thank Dr. Garry SA Myers for sharing the draft genome sequence of *C. suis*.

I also want to extend my gratitude to my lab members working at Molecular Diagnostic Laboratory, Dr. Erfan Chowdhury, Mrs. Dongya Gao, Dr. Anil Poudel, Dr. Chengming Wang, Dr. Yihang Li, and Dr. Yen-Chen Juan, for being wonderful colleagues with invaluable support in academic work and in daily life. In particular, I want to mention Mrs Gao's help in DNA sequencing; Dr. Chowdhury's support in identification of unique chlamydial proteins *in silico* and standardization of ELISA method; Dr. Poudel's hard work of collecting precious bovine sera from cattle naturally infected with *C. pecorum*; Dr. Juan's support in DNA preparation; and Dr. Wang's and Dr. Li's support by teaching me mouse husbandry.

Lastly, I want to thank my parents and my siblings for encouraging me to come to the USA in the first place, and continuing their support with great inspirations though they are living in another part of the world. I thank my wife Farhana Sultana for her unwavering support by tolerating my time in this research.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

μm	Micrometer; one-millionth of a meter; 10^{-6} meter
AA	Amino acid (residue)
aa	Amino acid position or number of amino acid
ACC	Accuracy in epitope discrimination
ARR	Antibody-reactive region
ASA	Accessible surface area; solvent accessibility
ATP	Adenosine-5'-triphosphate
AUC	Area under the curve; Area under the ROC curve
C.	*Chlamydia*
Cab	*C. abortus*
Cca	*C. caviae*
Cfe	*C. felis*
Cmu	*C. muridarum*
Cpe	*C. pecorum*
Cpn	*C. pneumoniae*
Cps	*C. psittaci*
Csu	*C. suis*
Ctr	*C. trachomatis*
CD4	Cluster of differentiation 4; CD4 T helper cells
CDR	Complementary-determining regions; antibody paratope
Cpaf	Chlamydial protease– or proteasome–like activity factor
CrpA	Cysteine rich outer membrane protein A
CT	Gene 'locus tag' in the genome of '*C. trachomatis*' strain D/UW-3/CX
CV	Coefficient of variation

D/D1/D2	Protein disorder tendency, or IUPred-L, or VSL2B disorder scale
DMSO	Dimethyl sulfoxide
DNA	Deoxy-ribonucleic acid
EB	Elementary body of *Chlamydia* spp.
ELISA	Enzyme-linked immunosorbent assay
*eno*A	Enolase A gene
*gat*A	Aspartyl/glutamyl-tRNA amidotransferase subunit A gene
*gid*A	t-RNA uridine 5-carboxy-methyl-amino-methyl modification protein gene
GTP	Guanosine-5'-triphosphate
H/H1/H2	Protein hydrophilicity, or Miyazawa or Parker scale
*hfl*X	GTPase HflX gene
HRP	Horseradish peroxidase
HSP	Heat shock protein
IEDB	Immune Epitope Database and Analysis Resource
IgG	Immunoglobulin G
IL	Interleukin; cytokine signaling molecule
Inc	Inclusion membrane protein
IncA	Inclusion membrane protein A
IncAf	Inclusion membrane protein A family protein
IncE	Inclusion membrane protein E
IncG	Inclusion membrane protein G
LGV	Lymphogranuloma venereum
LPS	Lipopolysaccharide
MHC	Major histocompatibility complex
MIF	Microimmunofluorescence test
MOMP	Major outer membrane protein; or, OmpA
N	Number of population; in statistics
n	Number of sample; in statistics

Neg	Negative; experimentally confirmed non-epitope (B-cell)
NT	Not tested; hence, status of epitope/non-epitope unknown
OmcB	Outer membrane cysteine-rich protein B; or, Omp2
Omp2	Outer membrane protein 2; or, OmpB
OmpA	Outer membrane protein A; or, MOMP
ORF	Open reading frame
p/P	*P*-value; the probability of incorrectly rejecting the null hypothesis in statistics
PCR	Polymerase chain reaction
PdhC	Pyruvate dehydrogenase C
PID	Percentage identity
pkn5	Serine/threonine-protein kinase 5
Pmp	Polymorphic outer membrane protein
PmpD	Polymorphic outer membrane protein D
PorB	Outer membrane protein B
Pos	Positive; experimentally confirmed B-cell epitope epitope
r^2	R squared; fraction of data explained
RB	Reticulate body of *Chlamydia* spp.
RecA	Recombination protein A
RLU	Relative light unit/sec
RNA	Ribonucleic acid
ROC	Receiver operating characteristic (ROC), or ROC curve
RpsA	30S ribosomal protein S1; ribosomal protein synthesis A
RpsB	30S ribosomal protein S2; ribosomal protein synthesis B
rRNA	Ribosomal RNA (ribonucleic acid)
S/S1/S2	Protein solvent accessibility, or Spine-X, or PaleAle 4.0 scale
SD	Standard deviation; amount of data variation in statistics
SGSG	Serine-glycine-serine-glycine
SPG	Sucrose phosphate glutamate

Tarp	Translocated actin-recruiting phosphoprotein
TNF-α	Tumor necrosis factor (TNF) alpha
UPGMA	Unweighted pair group method with arithmetic mean
UTP	Uridine-5'-triphosphate
VD	(Protein) Variable domain
Ysc	Type III secretion system protein

CHAPTER 1 LITERATURE REVIEW

1.1. BIOLOGY OF CHLAMYDIAE

Chlamydiae are 0.2-1.5 μm in diameter, coccoid, nonmotile, Gram-negative, obligate intracellular bacteria that multiply in eukaryotic cells, with a peculiar developmental life cycle within a membrane bound vacuole (inclusion), residing within the cytoplasm of host cells (1,2). Cell walls contain lipopolysaccharide, functional peptidoglycan, and hexagonally arranged projections with type III secretion system-like structures, analogous in structure and composition to walls of Gram-negative bacteria (1-6).

Chlamydiae replicate by binary fission with two principal alternative cell forms in their life cycles - (i) elementary bodies (EBs) that are infectious but essentially inert in physiological metabolism and (ii) reticulate bodies (RBs) that are metabolically active but non-infectious (1,7,8). When the development cycle is not synchronized, three chlamydial forms – (i) reticulate, (ii) intermediate, and (iii) elementary bodies are present (1,9-11). The alternative forms of the life cycle possess distinctive physiological changes in the outer membrane and nucleoid structure. Elementary bodies are smaller in size (0.2-0.4 μm in diameter) with electron-dense DNA condensed with proteins and few ribosomes surrounded by rigid cell walls that are resistant to sonic lysis (1,9,10). Reticulate bodies are variable in size, typically larger in diameter (0.6-1.5 μm), have less dense structure of nuclear material with more ribosomes surrounded by cell walls that are sensitive to sonic lysis (12).

The development cycle of chlamydiae is completed in three phases - (i) attachment and entry of elementary bodies into host cells and their reorganization into

reticulate bodies, (ii) multiplication of reticulate bodies, and (iii) conversion of a large fraction of the reticulate body population into a new generation of elementary bodies (1). The initial interactions of chlamydiae and host cells are probably by electrostatic forces (13). The uptake of chlamydiae is further secured by *Chlamydia*-specific proteins, such as heat shock protein 70 (hsp70), outer membrane cysteine-rich protein B (OmcB), major outer membrane protein/outer membrane protein A (MOMP/OmpA), or polymorphic outer membrane proteins (Pmp) (14-19). Reduction of disulfide bonds between outer membrane proteins of elementary bodies, by host molecules such as glutathione, renders the outer membrane permeable to host ATP. Subsequently, the intrinsic ATPase activity of elementary bodies generates energy for chlamydial metabolism immediately after entry into the host cells (20). Chlamydiae facilitate their survival in the inclusion by preventing fusion to lysosomes. While the details are not yet known, they achieve this evidently by remodeling the inclusion membrane with Inc proteins (inclusion membrane proteins) produced by reticulate bodies (1,21,23). The mature elementary bodies in inclusion do not divide; rather they are released from the cell to infect new cells of the same host or to infect new hosts if transmission is successful (8,23;24). Cellular exit may occur by two independent modes – (i) extrusion of the inclusion from the infected cell, and (ii) lysis of the intracellular inclusion followed by lysis of the cell. The yield of infectious chlamydiae per host cell varies from 10-1,000 depending on chlamydial strain, the host cell, and growth conditions (24,25).

Chlamydiae cannot be cultivated in cell-free synthetic medium yet, but can be propagated in laboratory animals such as rodents, the yolk sac of chicken embryos, or in cultured cells such as Hela or McCoy cells supplemented with amino acids (1,26-29).

Chlamydiae cannot utilize glucose catabolism for generating energy, but acquire ATP, GTP, and UTP via nucleotide transport proteins from the host cytoplasm, and these exogenous nucleotide triphosphates meet their needs (30). In addition, they encode enzymes for energy producing pathways and substrate-level phosphorylation (31,32). Interestingly, chlamydiae competitively acquire biosynthetic intermediates from metabolic pools of their host cells. From both host-derived and self-synthesized intermediates, chlamydiae synthesize protein, lipids, and glycogen to complete their parasitic life cycle in the host cells (33-38).

1.2. TAXONOMY OF CHLAMYDIAE

The taxonomy of chlamydiae has continuously changed over many decades since the discovery that its members cause trachoma (39). The name Chlamydozoa (meaning cloak and animal) was originally proposed to describe how infectious elementary bodies were cloaked within host cytoplasmic inclusions (40), followed by the proposal of *Chlamydia* in 1945 (1,41,42). Interestingly, accurate historical reports of the blinding disease trachoma exist already in the ancient world, and reliable descriptions date back to between 1553 and 1550 BC (43). Trachoma-causing agents were also observed in conjunctival cells from infants with ophthalmia and in cervical cells from their mothers as early as the 1930s and 1940s (40). Outbreaks of pneumonia associated with psittacine birds in 1930s led to the discovery of the causative agent of psittacosis (40). Because of biological and antigenic similarities among causative organisms of trachoma, psittacosis, and the agent that caused lymphogranuloma venereum (LGV), their close biological relationship was recognized. Since these infectious agents could be passed through bacterial filters and shared many properties with viruses, they were for a long time considered viruses (1,40). In this time of taxonomic uncertainty, many terms such as Bedsonia were proposed for chlamydiae, which created confusion rather than serving the purpose of taxonomy. In 1957, the isolation of the trachoma agent in embryonated hens' eggs in China led to detailed structural and morphological studies, and since 1960 chlamydiae have been universally accepted as bacteria. In 1980, the 'Approved Lists of Bacterial Names' listed just two species to include all chlamydia-like bacteria, and separated the species based on chemical characteristics, morphology, and developmental life cycle (44). *Chlamydia trachomatis* strains were identified with their ability to

accumulate glycogen in inclusions and their sensitivity to sulfadiazine while *Chlamydia psittaci* strains did not accumulate glycogen and were resistant to sulfadiazine.

The development of nucleic acid based taxonomy and isolation of new strains led to two additional species, *Chlamydia pneumoniae* (7) and *Chlamydia pecorum* (45). In 1997 (46), DNA sequence analysis identified chlamydiae that are closely related to *Chlamydia trachomatis*, but isolated from swine, and yet cannot be classified as *Chlamydia trachomatis* since these isolates are resistant to sulfadiazine, supporting the idea that there is a need for molecular based classification of chlamydiae. Subsequently Everett et al., 1999 proposed to separate the family *Chlamydiaceae* into two genera composed of nine species (*Chlamydia* and *Chlamydophila*, 47). Based on the Everett classification, the genus *Chlamydia* included three species (*Chlamydia trachomatis*, *Chlamydia suis* and *Chlamydia muridarum*) and the second genus *Chlamydophila* included the remaining 6 species (*Chlamydophila psittaci*, *Chlamydophil pneumoniae*, *Chlamydophila pecorum*, *Chlamydophila abortus*, *Chlamydophila caviae*, and *Chlamydophila felis*). This nine species classification of *Chlamydiaceae* has been widely accepted. In contrast, the separation of the family *Chlamydiaceae* into two genera has not been accepted by most scientists.

The phylogeny based on *16S rRNA* and *23S rRNA* genes, shown in Fig. 1.1A and 1.1B with three new proposed species (*Chlamydia avium*, *Chlamydia gallinacea* and *Chlamydia ibidis*), shows that *Chlamydiaceae* are well-separated from the related families *Parachlamydiaceae* (*Parachlamydia acanthamoebae*, *Protochlamydia amoebophila*), *Simkaniaceae* (*Simkania negevengis*) and *Waddliaceae* (*Waddlia chondrophila*). The *16S rRNA* or *23S rRNA* phylogenies appear to support the Everett

split of *Chlamydiaceae* into the *Chlamydia* and *Chlamydophila* genera, but comprehensive analysis using the major outer membrane protein or concatenated housekeeping proteins EnoA, GidA, GatA, HflX indicates that all currently recognized species in *Chlamydiaceae* are indeed very closely related (Fig. 1.1C and Fig 1.2). Therefore, recently Sachse et al., 2015 proposed a single genus *Chlamydia* in the family *Chlamydiaceae* to include all currently recognized species (48). Including the recently identified species *C. ibidis*, the single genus *Chlamydia* comprises now 12 species (Table 1.1) based on broadened criteria that encompass biological as well as molecular properties. From now, if not mentioned otherwise, we will use the chlamydial taxonomy proposed by Sachse et al., 2015. These 12 species of the genus *Chlamydia* are (i) *C. abortus* (Cab), (ii) *C. psittaci* (Cps), (iii) *C. caviae* (Cca), (iv) *C. felis* (Cfe), (v) *C. avium*, (vi) *C. gallinacea*, (vii) *C. ibidis*, (viii) *C. pecorum* (Cpe), (ix) *C. pneumoniae* (Cpn), (x) *C. muridarum* (Cmu), (xi) *C. suis* (Csu), and (xii) *C. trachomatis* (Ctr).

Fig. 1.2 shows that the species of the *C. psittaci* clade, comprised of *C. psittaci*, *C. abortus, C. caviae,* and *C. felis*, are closely related and have more than 90% amino acid (AA) sequence identity. This clade contains strains from birds (*C. psittaci*), ruminants (*C. abortus*), guinea pigs (*C. caviae*), and cats (*C. felis*) (Table 1.1). In Fig. 1.2A, the *C. trachomatis* clade consists of three closely related species, *C. trachomatis* in humans, *C. muridarum* in rodents (murine biovar causing mouse pneumonitis, MoPn; 46) and *C. suis* in swine (49), and the species of this clade also have more than 90% sequence identity. The divergent clade of *C. pneumoniae* (mainly reported

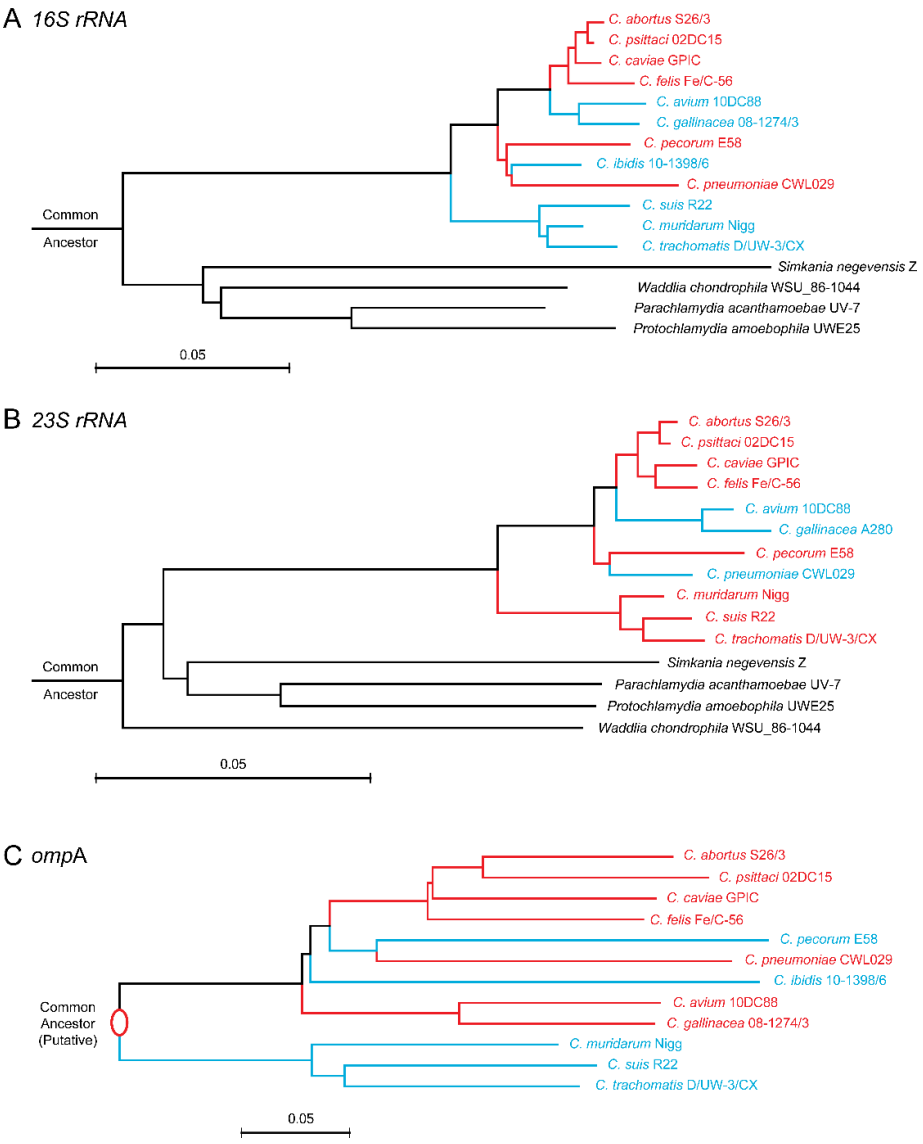
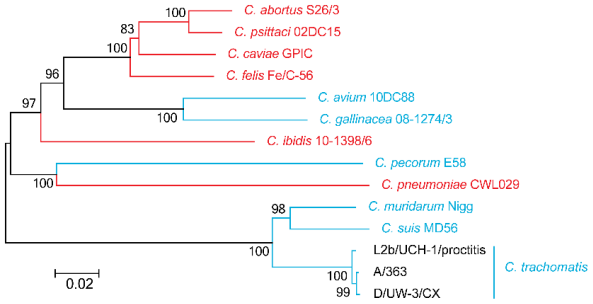


Fig. 1.1. Evolutionary relationships of *Chlamydia* spp. based on the nucleotide sequence of chlamydial (A) *16S rRNA*, (B) *23S rRNA* and (C) *omp*A genes. Bar 0.05 indicates substitution per nucleotide site. Alternating colors for major clades of *Chlamydia* spp. are used for aid in visualization. The putative common ancestor of *Chlamydia* spp., shown as red oval shape, is calculated based on *16S* and *23S rRNA* relationship of *Chlamydia* spp. with external members, *Chlamydiales*.

The evolutionary history was inferred using the Neighbor-Joining method (50). The tree with the optimal sum of branch lengths is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (51). All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA6 (52).

A **Phylogeny of *Chlamydia* spp.**
Multi-Protein Sequence Analysis of *eno*A, *gid*A, *gat*A & *hfl*X encoded concatenated protein sequences



B **Average Sequence Identity (%)** in *eno*A, *gid*A, *gat*A & *hfl*X encoded concatenated proteins

	C. abortus S26/3	*C. psittaci* 02DC15	*C. caviae* GPIC	*C. felis* Fe/C-56	*C. avium* 10DC88	*C. gallinacea* 08-1274/3	*C. ibidis* 10-1398/6	*C. pecorum* E58	*C. pneumoniae* CWL029	*C. muridarum* Nigg	*C. suis* MD56	*C. trachomatis* A/363	*C. trachomatis* D/UW-3/CX	*C. trachomatis* L2b/UCH.
C. abortus S26/3														
C. psittaci 02DC15	97													
C. caviae GPIC	92	92												
C. felis Fe/C-56	91	92	92											
C. avium 10DC88	81	81	82	82										
C. gallinacea 08-1274/3	82	82	81	82	89									
C. ibidis 10-1398/6	81	82	83	82	78	77								
C. pecorum E58	74	74	74	74	70	71	72							
C. pneumoniae CWL029	74	74	74	74	70	70	72	72						
C. muridarum Nigg	73	74	74	73	71	70	73	67	66					
C. suis MD56	73	74	73	73	70	69	73	68	66	93				
C. trachomatis A/363	73	74	74	73	71	70	73	68	67	92	91			
C. trachomatis D/UW-3/CX	73	74	74	73	71	70	73	68	67	92	91	100		
C. trachomatis L2b/UCH-1/proctitis	73	74	74	73	71	70	73	68	67	93	91	100	99	

Fig. 1.2. Evolutionary relationships of *Chlamydia* spp. based on concatenated housekeeping protein sequences. (A) Phylogeny of *Chlamydia* spp. Bar 0.02 indicates amino acid substitutions per site. Alternating colors are used for aid in visualization. (B) Protein sequence identity. Numerical values indicate protein sequence identity. Pairs shown in red have ≥ 90 % sequence identity and pairs in blue have ≥ 99 % sequence identity. Housekeeping protein sequences encoded by *eno*A (enolase), *gid*A (tRNA uridine 5-carboxy-methyl-amino-methyl modification protein GidA), *gat*A (aspartyl/glutamyl-tRNA amidotransferase subunit A) and *hfl*X (GTPase HflX) were aligned, followed by linking together of these four protein sequences for each strain of *Chlamydia* into a long sequence (concatenated protein sequence).

The evolutionary history was inferred using the Neighbor-Joining method (50). The tree with the optimal sum of branch lengths is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (53). The evolutionary distances were computed using the JTT matrix-based method (54). Evolutionary analyses were conducted in MEGA6 (52).

Table 1.1: Current major members of *Chlamydiales*.

Family	Genus	Species	Natural Host	Other Host	Site of Infection	Reviewed from
Chlamydiaceae	***Chlamydia***	(i) *C. abortus*	Sheep, goat	Cattle, swine	Genital, respiratory	48
		(ii) *C. psittaci*	Birds	Mammals	Respiratory	48
		(iii) *C. caviae*	Guinea pig	Horse	Ocular, genital	48
		(iv) *C. felis*	Cat		Ocular, respiratory	48
		(v) *C. avium*	Pigeon, parrots		Respiratory	48
		(vi) *C. gallinacea*	Chicken	Other poultry	Respiratory	48
		(vii) *C. ibidis*	African ibises		Isolated from cloacal swabs	55
		(viii) *C. pecorum*	Cattle, koala	Sheep, goat, swine	Gastro-intestinal, respiratory, urogenital	48
		(ix) *C. pneumoniae*	Human, horse, koala	Amphibians, reptiles	Respiratory, cardiovascular	48
		(x) *C. muridarum*	Rodents		Gastro-intestinal	48
		(xi) *C. suis*	Swine	Ruminants	Gastro-intestinal, genital	48
		(xii) *C. trachomatis*	Human		Ocular (Trachoma), urogenital (STD), lymph node (LGV)	48
Parachlamydiaceae	***Parachlamydia***	*P. acanthamoebae*	Amoeba	Human	Respiratory in humans	56,57
Simkaniaceae	***Simkania***	*S. negevensis*	Amoeba	Human	Respiratory in humans	56
Waddliaceae	***Waddlia***	*W. chondrophila*	Amoeba	Ruminants	Urogenital (abortive) in ruminants	56,58

in humans & marsupials; 7) and *C. pecorum* (isolated from ruminants, swine, and koala; 45) is well separated from the *C. psittaci* and *C. trachomatis* clades. The recently accepted new species *C. avium* (pigeons, parrots; 48,59) and *C. gallinacea* (poultry; 48,59) as well as the candidate species *C. ibidis* (African ibises, 55) are well separated in the phylogenic tree from the previously described *Chlamydia* spp. This suggests that many more species likely exist in the family *Chlamydiaceae* in as of yet unexplored host populations.

1.3. IMPORTANCE OF *CHLAMYDIA* SPP.

The outcome of *Chlamydia* spp. infections is determined by (i) infectious dose, (ii) host genetic susceptibility, and (iii) host immune status (60-63). *Chlamydia* spp. normally infect mucosal epithelium, the single cell columnar layer of the epithelium, where they undergo their biphasic developmental cycle (64). Chlamydial lipopolysaccharide plays an essential role in determining disease conditions such as chronic granulomatous lesions of mononuclear cell aggregates and fibrosis. $CD4^+$ lymphocytes are protective for the host by restricting chlamydial replication via a Th1 immune response (62,63,65). In contrast, TNF-α, IL-1α and IL-6 are involved in pathology associated with *Chlamydia* spp. infection (66). In epithelial cells, *Chlamydia* spp. infection initiates production of various pro-inflammatory mediators, however, despite initiating local inflammation, most infected animals remain asymptomatic.

The human pathogen *C. trachomatis* causes ocular infection often leading to blindness (67), and genitourinary tract infection associated with chronic salpingitis and tubal infertility or extra-uterine pregnancy, and pelvic inflammatory disease (1,68), as well as lymphogranuloma venereum. *C. pneumoniae* infection results in respiratory disease manifestations such as pharyngitis, bronchitis, and pneumonia in humans (7). *C. pneumoniae* was also linked to human atherosclerosis and coronary heart disease (69). In sheep, goats, and cattle, *C. abortus* causes abortion, seminal vesiculitis, and mastitis, and *C. pecorum* causes vaginitis, endometritis, polyarthritis, pneumonia, enteritis, encephalomyelitis and several other latent infections (70,71). *C. pecorum* also causes similar infections in swine and koalas (1). *C. psittaci* primarily infects birds, widely known as psittacosis, causing severe systemic infections or latent respiratory and enteric

infections. *C. psittaci* can be transmitted to humans and cause mild to severe interstitial pneumonia (1,72). *C. suis* causes conjunctivitis, pneumonia, enteritis and polyarthritis, mostly reported in swine (1). *Chlamydia felis* causes conjunctivitis in cats, which is often accompanied by rhinitis. In mice, *C. muridarum* infection causes pneumonitis and less frequently interstitial keratitis. Guinea pigs are infected with *C. caviae* which causes follicular conjunctivitis and interstitial keratitis (1).

Trachoma is the world's leading cause of preventable blindness affecting tens of millions of people in developing countries (73,74). Lymphogranuloma venereum is also common in developing countries leading to severe lymphoadenopathy, genital ulceration, proctitis, strictures and formation of fistulae (73-75). Urogenital *C. trachomatis* infections associated with urethritis, epididymitis, cervicitis, salpingitis and pelvic inflammatory disease are considered the most common bacterial sexually transmitted disease worldwide (73,74). Also in humans, *C. pneumoniae* have been associated with atherosclerosis and coronary heart diseases, lung cancer, cerebral infarction/cerebrovascular disease, and the total human health impact from these diseases has yet to be fully determined (76-80). In animals, infections with typical clinical symptoms have been reported widely for *C. pecorum*, *C. suis* and *C. abortus* (1). Due to the most common asymptomatic animal infections by ubiquitous *Chlamydia* spp., the full scope of the economic loss is highly underestimated. This loss may be as high as 10-15% reduction in gross income, an order of magnitude higher than the reported loss from typical diseases with clear clinical manifestation (81,82).

1.4. GENETIC-, BIO- AND SERO-VARIANTS OF *CHLAMYDIA* SPP.

Compared to most free living prokaryotes, *Chlamydia* spp. possess a very small genome of 1.03×10^6 bp for *C. trachomatis* to 1.23×10^6 bp for *C. pneumoniae*, which are less than half of the *Parachlamydiaceae* genome of $2.4\text{-}3.0\times10^6$ bp (57). This comparison suggests that *Chlamydia* spp. lost significant parts of their genome during evolution. *Chlamydia* spp. are thought to be separated from each other by over a 100 million years of evolution in the presumptive molecular clock. For comparison, lineages of *Homo sapiens* are thought to be separated by no more than 1 million years. Despite very early separation of the chlamydial species from one another during evolution, they have 90% gene content identity compared to the prototype strain *C. trachomatis* D/UW-3/CX. In addition, they have a very high conservation of genome synteny (order of genes in genomes), which indicates that the genomes of *Chlamydia* spp. are highly stable compared to most prokaryotic genomes (1). However, recent genome bioinformatics analyses suggest that recombination between strains and species does occur; and genetic crosses between variant *C. trachomatis* strains can easily be made experimentally by co-infecting host cells (40). Most *Chlamydia* spp. have a conserved 7500 bp plasmid (83,84). However, this plasmid is not required for growth *in vitro* (85), drug-resistance encoded by the plasmid has not been reported, and clinical isolates of *C. trachomatis* may lack this plasmid and yet show persistence *in vivo* (85,86). Stable drug-resistant mutants of *C. pneumoniae* and *C. trachomatis* have not been isolated, but *C. suis* in American pigs became resistant to tetracycline by acquiring a tet(R) transposon-like genetic element from another bacterium (87).

Chlamydia spp. show remarkably high serovariation, despite the fact that they have small genomes with very high synteny, high genome content identity, high nucleotide and amino acid sequence identity, and relatively stable genomes due to minimum influence of plasmids, mobile elements, or phages. Major antigenic variations of this bacterial group are due to a porin protein known as major outer membrane protein (MOMP) that is the most immunodominant protein in *Chlamydia* spp. The allelic differences in the *omp*A gene, which encodes MOMP, are particularly high in *C. suis* followed by *C. pecorum*, *C. trachomatis*, and *C. psittaci*. The strains of *C. trachomatis*, the best studied species of this genus, are clustered into 12 serotypes A to L, or 17 sub-serotypes with subtypes B and Ba, J and Ja, I and Ia, and L1, L2, and L3 (88,89). Human infections by *C. trachomatis* follow three typical patterns, (i) ocular blinding disease (trachoma) caused by A-C serotypes, (ii) urogenital sexually transmitted diseases caused by D-K serotypes, and (iii) lymphogranuloma venereum (LGV) diseases caused by L serotypes. Polyclonal mouse antibodies against MOMP (OmpA) have been introduced for serotyping of *Chlamydia* species in 1960, and subsequently specific monoclonal antibodies were successfully used for serotyping of *C. trachomatis* (90).

The phylogenetic tree, based on the OmpA sequences of 92 strains representing the A-L *C. trachomatis* serovars, shows that these strains are clustered into distinct clades for the three biovars of *C. trachomatis* (Fig. 1.3) that are genetically and serologically separated from each other and from the remaining *Chlamydia* species. These three biovars are distinguishable by their tissue-tropism, pathogenesis in humans and laboratory animals, and biological properties in cell cultures (1). Serotypes for other species are not well-characterized or not characterized at all, but isolates and PCR

products have been directly genotyped by sequencing the *omp*A gene (91-93) and it is apparent that *C. suis* and *C. pecorum* have highly divergent, numerous serotypes (93). For *C. pneumoniae*, three biovars, (i) human, (ii) koala, and (iii) equine are reported with each one having OmpA sequence differences. Six serotypes, A to F, of *C. psittaci* isolates of avian isolates have been characterized (94,95). Analysis of *omp*A sequences from the NCBI database shows that *C. abortus*, *C. caviae*, *C felis*, and *C. muridarum* are mono-serotype chlamydial species (1). *C. avium*, *C. gallinacea* and *C. ibidis* have recently reported, and subsequent studies are required to determine how many serovars these species may have but it is already apparent that *omp*A is highly polymorphic in *C. gallinacea*.

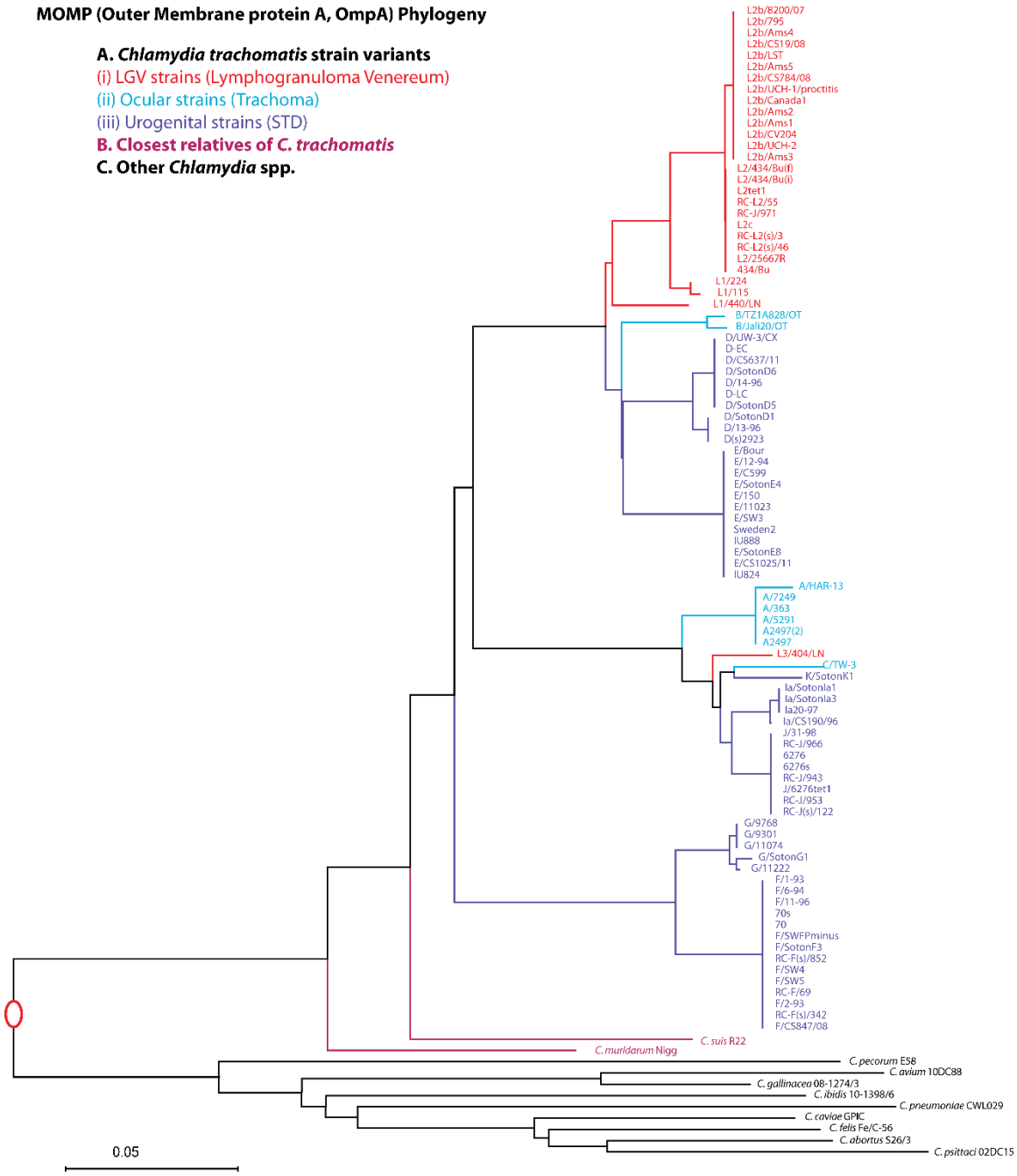


Fig. 1.3. Evolutionary relationships of 90 strains of *C. trachomatis* with the remaining 11 *Chlamydia* spp. based on OmpA amino acid sequences. Bar 0.05 indicates substitution/site. Alternating colors are used as visualization aid. The optimal tree with the sum of branch length = 1 is shown and reconstruction method is same as in Fig. 1.2.

1.5. NEED FOR SPECIFIC SEROLOGICAL ASSAYS

Chlamydiae are ubiquitous bacteria, and they have been detected in, and isolated from, a wide range of hosts including humans, ruminants, mammals, reptiles, birds, fishes, and amoebae. In spite of such broad host-spectrum for *Chlamydia* spp., a high degree of host specificity and tissue-tropism exists for human chlamydial strains/types. Most of the current data accumulated on serovariation for chlamydiae is based on *C. trachomatis* strains. However, the less studied animal chlamydiae have a broader host spectrum, and in herd infections represent a high burden on animal hosts, resulting in asymptomatic infections as well as occasional typical clinical manifestation. In addition, animal chlamydiae represent the epidemiological reservoir from which human strains and species originated, and studying their evolution and epidemiological behavior will increase our understanding of optimal prophylactic and therapeutic strategies.

Detection and species-/type-specific differentiation of chlamydial agents (antigens/nucleic acid) or of anti-chlamydial antibodies are both equally important. With the advent of PCR and automated nucleotide sequencing technology, sequences of the *omp*A allele (genotype) are now routinely used for typing *C. trachomatis* strains. Multiplex real-time PCR, other nucleic acid amplification tests, and microarray differentiation of multiple genes have also been introduced for specific detection of *Chlamydia* spp. of human or veterinary interest (96-98). Such nucleic acid based techniques are easily reproducible or commencially available based on specific need.

The principal drawback of direct detection of chlamydial agents is the stochastic hit/miss nature of this approach, which renders single tests less valuable than serial testing. In contrast, serological assays indicate the history of exposure to an infectious

agent, and paired serum samples indicate the host response to the infections. Thus, for epidemiological or retrospective analyses, serological assays are vastly preferable over antigen detection that is by necessity a prospective methodology. In addition, sampling for antigen detection may be difficult for various reasons, for example because of inaccessible anatomical infections associated with atherosclerosis, coronary heart diseases, lung cancer, or cerebral infarction. Thus, absence of specimens or difficulties in obtaining species leaves serology as an alternative or only option. Obtaining serum or plasma is in contrast a routine procedure for any animal species including humans.

The need for *Chlamydia* species-/type-specific detection of anti-*Chlamydia* antibodies is widely known, particularly for animal chlamydial infections (99). Such specific detection of anti-*Chlamydia* antibodies is perceived as essential for many reasons such as (i) to determine presence or absence of *Chlamydia*-induced diseases, (ii) to estimate the prevalence of the infections, (iii) to confirm clinical diseases, (iv) to determine serovars of importance for vaccination, and (v) to determine immune status after vaccination. In human infections caused by *C. trachomatis* or by *C. pneumoniae*, species-specific serological assays are essential for both patient care and basic understanding of chlamydial diseases. Moreover, several epidemiological and pathogenetic questions require accurate determination of serovar reactivity of anti-*C. trachomatis* antibodies such as (i) establishment of transmission patterns, (ii) association with clinical manifestations and pathogenicity, (iii) determination of tissue tropism or organ affinity, (iv) discrimination of persistent from acute infections, and surveillance for specific *C. trachomatis* serotypes (73).

1.6. PROBLEMS WITH CURRENT SEROLOGICAL ASSAYS

With the successful isolation and propagation of chlamydiae from clinical samples, various techniques have been introduced for species/type-specific detection of chlamydiae. In the 1960s, Wang and Grayston developed the mouse toxicity prevention test and classified 80 strains of *C. trachomatis* into six immunological groups by an effort of 7 years (100,101). Subsequently, they developed a microimmunofluorescent (MIF) test based on elementary body (EB) antigens grown in the yolk sac of developing chicken embryos, and on antisera produced in mice against these antigens. Using cross-testing of the antiserum reactivities against all EB antigens, they reported 13 different *Chlamydia* serovars A, ba, B, C, D, E, F, G, H, I, L1, L2 and L3 (1,88,89,102). Later, 4 additional types (Ia, J, Ja and K) were identified, resulting in 17 *C. trachomatis* serovars. Due to cross-reactivity in microimmunofluorescent test, subsequent efforts were made to use more-specific monoclonal antibodies for serovar determination (101,103). Monoclonal antibodies were further refined to increase the sensitivity and specificity in the assay (104). However, all efforts at immunological approaches towards typing of chlamydial isolates have become moot with the ease and accuracy of genetic methods for typing of isolates or amplification products of *Chlamydia* spp.

In contract, for species- or type-specific detection of antibodies against *Chlamydia* spp., the microimmunofluorescent test has remained the gold standard since its introduction in the 1960s. The reason for this is the high species-cross reactivity of all suitable purified chlamydial antigens. Table 1.2 shows that whole chlamydial immunodominant proteins, the main targets of anti-chlamydial antibodies, have high sequence homology among *Chlamydia* species and serovars. All immunodominant

proteins, including the serovar-determining MOMP, are highly conserved, thus highly cross-reactive in serological assays, and therefore cannot be used for chlamydial species-specific serology. Lipopolysaccharide (LPS) of *Chlamydia* also cannot be used for species-specific serology, because chlamydial LPS is genus-specific such that, e.g., *C. pneumoniae* LPS cross-reacts with anti-*C. trachomatis* antibodies (105), and even LPS of the family *Piscichlamydiaceae* cross-reacts with *Chlamydiaceae* LPS.

However, the MIF has poor sensitivity, poor reproducibility, cross-reactivity, interlaboratory variation, poor correlation with nucleic acid amplification tests, and many technical difficulties (106-115). MIF is traditionally performed as an indirect fluorescent antibody technique that enables observation of the antibody captured onto fixed chlamydial antigens (88,89,116,117). The bound antibodies are detected by fluorescein-conjugated secondary antibodies. The antigens used in the test are usually whole *Chlamydia* elementary bodies that are grown in cell culture, purified, and treated with formalin. The antigen is mixed with normal yolk sac suspension to firmly attach the elementary bodies to the microscope slide as small spots for each serovar, essentially as a macroarray. To measure an antibody titer, two fold serially diluted serum samples were loaded onto spotted slides. The slides are incubated and are washed several times, before adding the fluorescent antibody conjugate, followed by another round of incubation and washes. The prepared slides require painstaking expert microscopic examination for specific fluorescence of elementary bodies. Compared to a microplate ELISA, this technique is inherently tedious and subjective. Additionaly, the MIF for chlamydial serology is complicated due to the very nature of chlamydial antigen preparation. The

Table 1.2. Highly conserved chlamydial immunodominant proteins[a].

Protein	Compared Chlamydial species-specific proteins	*C. abortus* S26/3	*C. psittaci* 02DC15	*C. pecorum* E58	*C. pneumoniae* CWL029	*C. suis* MD56	*C. murirarum* Nigg	*C. trachomatis* D/UW-3/CX	*C. trachomatis* L2b/UCH-2	*C. trachomatis* A/HAR-13
OmpA/ MOMP	*C. abortus* S26/3	100	85	72	76	67	67	66	65	66
	C. pneumoniae CWL029	76	71	71	100	65	66	65	64	65
	C. trachomatis D/UW-3/CX	66	65	65	65	86	83	100	95	86
HSP/ GrpE	*C. abortus* S26/3	100	97	77	78	75	76	76	75	75
	C. pneumoniae CWL029	77	77	72	100	71	70	71	70	71
	C. trachomatis D/UW-3/CX	76	75	72	71	96	93	100	99	98
OmcB/ OmpB	*C. abortus* S26/3	100	98	79	84	72	71	71	70	71
	C. pneumoniae CWL029	84	85	80	100	72	73	72	71	72
	C. trachomatis D/UW-3/CX	71	71	71	72	93	92	100	98	99
OmcA/ PorB	*C. abortus* S26/3	100	98	63	68	65	65	63	63	63
	C. pneumoniae CWL029	68	69	67	100	61	61	59	60	59
	C. trachomatis D/UW-3/CX	63	64	60	59	89	90	100	98	100

[a] **Protein identities (amino acid sequence identity in percentage) were calculated using blast search in NCBI database.**

few reference laboratories can offer this test due to its requirement of expert staff and demanding facilities for cultivation of *Chlamydia*, and standardization of antigens and microscopic observation. Given the fact that *Chlamydia* has now 12 identified species and some of the species have numerous serovars (Fig. 1.4), cultivation of chlamydiae and standardization of antigens are a cumbersome and daunting task for most laboratories. Additionally, this test has also cross-reactivity due to *Chlamydia* genus wide conserved LPS, conserved immunodominant surface exposed antigens such as OmpA/MOMP (which also contains serovar-determinant peptide sequence of *Chlamydia*), Outer membrane protein B, and heat shock proteins (Table 1.2). Disagreements between MIF results from different refence laboratories have also reported (113). Therefore, despite being an essential tool in Chlamydiology, a simple and specific serological assay for antibodies is still lacking, and development of such methodology is essential for future progress in understanding infections by *Chlamydia* spp.

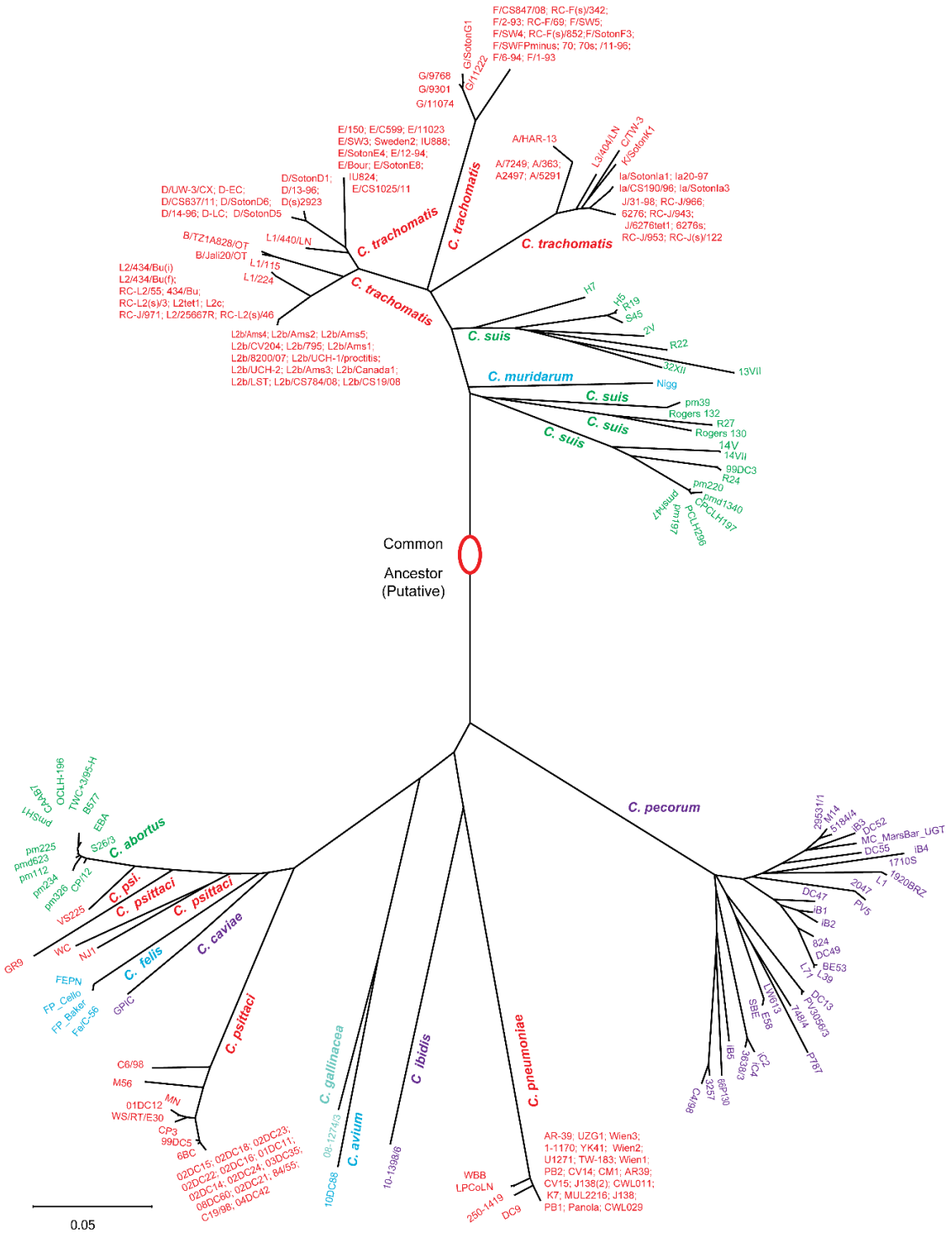


Fig. 1.4. Evolutionary relationships of 11 *Chlamydia* spp. and their subtypes based on OmpA amino acid sequences. Bar 0.05 indicates amino acid substitution/site. For visual aid, alternating colors are used. The strain name is indicated at the tip of branches. OmpA sequences, nearly full-length available in NCBI database, were aligned using muscle algorithms for tree reconstruction in Mega 6 (52).

1.7. RESEARCH RATIONALE AND OBJECTIVES

Current *Chlamydia* species/type-specific serological assays perform poorly or are not available at all. The impediments are (i) time-consuming and cumbersome propagation of the obligately intracellular *Chlamydia* spp. cell culture, particularly for the required large antigen amounts, (ii) time-consuming purification and standardization of chlamydial antigens, (iii) requirement for antigen production of numerous species and serovars of antigens, and (iv) chlamydial antigens that, despite all these efforts for production, are unsatisfactory because of their cross-reactivity.

We hypothesized that if (i) *Chlamydia* species-specific immunodominant B-cell epitopes could be identified, the specific peptide antigens can (ii) be used in a simple ELISA format for specific and sensitive serological assays that avoid cross-reactivity. Therefore, the overall purpose of this study was to develop a simple method for species-specific detection of anti-*Chlamydia* antibodies using peptide epitope antigens.

The specific goals of this investigation were:

1. Prepare an exhaustive list of immunodominant proteins of *Chlamydia* spp.
2. Predict B-cell epitopes *in silico* from polymorphic regions of immunodominant proteins of Chlamydia spp. for species/serovar-specific peptide antigens, and from conserved regions for genus-specific peptide antigens.
3. Raise high-titered *Chlamydia* mono-species-specific mouse antisera by multiple inoculations of viable chlamydial type strains.
4. Evaluate current *in silico* methods for B-cell epitope prediction, and improve them if needed by iterative comparison of predicted with actual antigen seroreactivity.
5. Predict potential for cross-reactivity from relative phylogenetic positions of peptide antigen sequences, and recommend genus-, species-, and serovar-specific antigens.
6. Validate *Chlamydia* spp. peptide serology with confirmed *Chlamydia*-reactive and non-reactive natural host sera.

REFERENCES

1. **Kuo CC, Stephens RS, Bavoil PM and Kaltenboeck B. 2011.** Genus I. *Chlamydia* Jones, Rake and Stearns 1945, 55^{AL}. *In* Bergey's Manual of Systematic Bacteriology, 2nd edition, vol 4, pp 846-865. Springer, New York.

2. **Liechti GW, Kuru E, Hall E, Kalinda A, Brun YV, Van Nieuwenhze M and Maurelli AT. 2014.** A new metabolic cell-wall labelling method reveals peptidoglycan in *Chlamydia trachomatis*. *Nature* **506**: 507-510.

3. **Matsumoto A. 1981.** Electron microscopic observations of surface projections and related intracellular structures of *Chlamydia* organisms. *J Electron Microsc* (Tokyo) **30**: 315-320.

4. **Matsumoto A. 1981.** Isolation and electron microscopic observations of intracytoplasmic inclusions containing *Chlamydia psittaci*. *J Bacteriol* **145**: 605-612.

5. **Hsia RC, Pannekoek Y, Ingerowski E and Bavoil PM. 1997.** Type III secretion genes identify a putative virulence locus of *Chlamydia*. *Mol Microbiol* **25**: 351-359.

6. **Bavoil PM and Hsia RC. 1998.** Type III secretion in *Chlamydia*: a case of déjà vu? *Mol Microbiol* **28**: 860-862.

7. **Grayston JT, Kuo CC, Campbell LA and Wang SP. 1989.** *Chlamydia pneumoniae* sp. nov. for *Chlamydia* sp. strain TWAR. *Int J Syst Bacteriol* **39**: 88-90.

8. **Yang ZP, Cummings PK, Patton DL and Kuo CC. 1994.** Ultrastructural lung pathology of experimental *Chlamydia pneumonia*e pneumonitis in mice. *J Infect Dis* **170**:464-467.

9. **Hackstadt T, Baehr W and Ying Y. 1991.** *Chlamydia trachomatis* developmentally regulated protein is homologous to eukaryotic histone H1. *Proc Natl Acad Sci USA* **88**: 3937-3941.

10. **Wagar EA and Stephens RS. 1988.** Developmental-form-specific DNA-binding proteins in *Chlamydia* spp. *Infect Immun* **56**: 1678-1684.

11. **Tribby II, Friis RR and Moulder JW. 1973.** Effect of chloramphenicol, rifampicin, and nalidixic acid on *Chlamydia psittaci* growing in L cells. *J Infect Dis* **127**: 155-163.

12. **Chi EY, Kuo CC and Grayston JT. 1987.** Unique ultrastructure in the elementary body of *Chlamydia* sp. strain TWAR. *J Bacteriol* **169**: 3757-3763.

13. **Hybiske K and Stephens RS. 2007.** Mechanisms of *Chlamydia trachomatis* entry into nonphagocytic cells. *Infect Immun* **75**: 3925-3934.

14. **Kuo CC, Wang SP and Grayston JT. 1973.** Effect of polycations, polyanions, and neuraminidase on the infectivity of trachoma-inclusion conjunctivitis and lymphogranuloma venereum organisms in HeLa cells: sialic acid residues as possible receptors for trachoma-inclusion conjunctivitis. *Infect Immun* **8**: 74-79.

15. **Kuo CC and Grayston T. 1976.** Interaction of *Chlamydia trachomatis* organisms and HeLa 229 cells. *Infect Immun* **13**: 1103-1109.

16. **Raulston JE, Davis CH, Schmiel DH, Morgan MW and Wyrick PB. 1993.** Molecular characterization and outer membrane association of a *Chlamydia trachomatis* protein related to the hsp70 family of proteins. *J Biol Chem* **268**: 23139-23147.

17. **Stephens RS, Koshiyama K, Lewis E and Kubo A. 2001.** Heparin-binding outer membrane protein of chlamydiae. *Mol Microbiol* **40**: 691-699.

18. **Ting LM, Hsia RC, Haidaris CG and Bavoil PM. 1995.** Interaction of outer envelope proteins of *Chlamydia psittaci* GPIC with the HeLa cell surface. *Infect Immun* **63**: 3600-3608.

19. **Clifton DR, Dooley CA, Grieshaber SS, Carabeo RA, Fields KA and Hackstadt T. 2005.** Tyrosine phosphorylation of the chlamydial effector protein Tarp is species specific and not required for recruitment of actin. *Infect Immun* **73**: 3860-3868.

20. **Peeling RW, Peeling J and Brunham RC. 1989.** High-resolution 31P nuclear magnetic resonance study of *Chlamydia trachomatis*: induction of ATPase activity in elementary bodies. *Infect Immun* **57**: 3338-3344.

21. **Rockey DD and Rosquist JL. 1994.** Protein antigens of *Chlamydia psittaci* present in infected cells but not detected in the infectious elementary body. *Infect Immun* **62**: 106-112.

22. **Bastidas RJ, Elwell CA, Engel JN and Valdivia RH. 2013.** Chlamydial intracellular survival strategies. *Cold Spring Harbor Perspect Med* **3**: a010256.

23. **Peterson EM and de la Maza LM. 1988.** *Chlamydia* parasitism: ultrastructural characterization of the interaction between the chlamydial cell envelope and the host cell. *J Bacteriol* **170**: 1389-1392.

24. **Neeper ID, Patton DL and Kuo CC. 1990.** Cinematographic observations of growth cycles of *Chlamydia trachomatis* in primary cultures of human amniotic cells. *Infect Immun* **58**: 2042-2047.

25. **Hybiske K and Stephens RS. 2007.** Mechanisms of host cell exit by the intracellular bacterium *Chlamydia*. *Proc Natl Acad Sci USA* **104**: 11430-11435.

26. **Storz J, Shupe JL, Smart RA and Thornley RW.** 1966. Polyarthritis of calves: experimental induction by a psittacosis agent. *Am J Vet Res* **27**: 987.

27. **Allan I and Pearce JH. 1983.** Amino acid requirements of strains of *Chlamydia trachomatis* and *C. psittaci* growing in McCoy cells: relationship with clinical syndrome and host origin. *J Gen Microbiol* **129**: 2001-2007.

28. **Allan I and Pearce JH. 1983.** Differential amino acid utilization by *Chlamydia psittaci* (strain guinea pig inclusion conjunctivitis) and its regulatory effect on chlamydial growth. *J Gen Microbiol* **129**: 1991-2000.

29. **Kuo CC and Grayston JT. 1990.** Amino acid requirements for growth of *Chlamydia pneumoniae* in cell cultures: growth enhancement by lysine or methionine depletion. *J Clin Microbiol* **28**: 1098-1100.

30. **Tipples G and McClarty G. 1993.** The obligate intracellular bacterium *Chlamydia trachomatis* is auxotrophic for three of the four ribonucleoside triphosphates. *Mol Microbiol* **8**: 1105-1114.

31. **Hatch TP, Al-Hossainy E and Silverman JA. 1982.** Adenine nucleotide and lysine transport in *Chlamydia psittaci*. *J Bacteriol* **150**: 662-670.

32. **Iliffe-Lee ER and McClarty G. 1999.** Glucose metabolism in *Chlamydia trachomatis*: the 'energy parasite' hypothesis revisited. *Mol Microbiol* **33**: 177-187.

33. **Alexander JJ. 1968.** Separation of protein synthesis in meningopneumonitis agent from that in L cells by differential susceptibility to cycloheximide. *J Bacteriol* **95**: 327-332.

34. **Plaunt MR and Hatch TP. 1988.** Protein synthesis early in the developmental cycle of *Chlamydia psittaci*. *Infect Immun* **56**: 3021-3025.

35. **Reed SI, Anderson LE and Jenkin HM. 1981.** Use of cycloheximide to study independent lipid metabolism of *Chlamydia trachomatis* cultivated in mouse L cells grown in serum-free medium. *Infect Immun* **31**: 668-673.

36. **Gordon FB and Quan AL. 1965.** Occurrence of glycogen in inclusions of the psittacosis-lymphogranuloma venereum-trachoma agents. *J Infect Dis* **115**: 186-196.

37. **Lin HS and Moulder JW. 1966.** Patterns of response to sulfadiazine, D-cycloserine and D-alanine in members of the psittacosis group. *J Infect Dis* **116**: 372-376.

38. **Scidmore MA, Fischer ER and Hackstadt T. 1996.** Sphingolipids and glycoproteins are differentially trafficked to the *Chlamydia trachomatis* inclusion. *J Cell Biol* **134**: 363-374.

39. **Halberstaedter L and von Prowazek S. 1907.** Untersuchungen über die Malariaparasiten der Affen. *Arb GesundhAmt* **26**: 44–47.

40. **Clarke IN. 2011.** Evolution of *Chlamydia trachomatis*. *Ann N Y Acad Sci* **1230**: E11-E18.

41. **Jones H, Rake G and Stearns B. 1945.** Studies on lymphogranuloma venereum: III. The action of the sulfonamides on the agent of lymphogranuloma venereum. *J Infect Dis* **76**: 55-69.

42. **Page LA. 1968.** Proposal for the recognition of two species in the genus *Chlamydia* Jones, Rake, and Stearns, 1945. *Int J Syst Bacteriol* **18**: 51-66.

43. **Taylor HR. 2008.** Trachoma: a blinding scourge from the Bronze Age to the twenty-first century. *N Engl J Med* **358**: 1872-1873.

44. **Skerman VBD, McGowan V and Sneath PHA. 1980.** Approved lists of bacterial names. *Int J Syst Bacteriol* **30**: 225-420.

45. **Fukushi H and Hirai K. 1992.** Proposal of *Chlamydia pecorum* sp. nov. for *Chlamydia strains* derived from ruminants. *Int J Syst Bacteriol* **42**: 306-308.

46. **Nigg C and Eaton MD. 1944.** Isolation from normal mice of a pneumotropic virus which forms elementary bodies. *J Exp Med* **79**: 497.

47. **Everett KD, Bush RM and Andersen AA. 1999.** Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. *Int J Syst Bacteriol* **49**, 415–440.

48. **Sachse K, Bavoil PM, Kaltenboeck B, Stephens RS, Kuo C-C, Rosselló-Móra R and Horn M. 2015.** Emendation of the family *Chlamydiaceae*: Proposal of a single genus, *Chlamydia*, to include all currently recognized species. *Syst Appl Microbiol* http://dx.doi.org/10.1016/j.syapm.2014.12.004.

49. **Hoelzle LE, Steinhausen G and Wittenbrink MM. 2000.** PCR-based detection of chlamydial infection in swine and subsequent PCR-coupled genotyping of chlamydial omp1-gene amplicons by DNA-hybridization, RFLP-analysis, and nucleotide sequence analysis. *Epidemiol Infect* **125**: 427-439.

50. **Saitou N and Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.

51. **Tamura K, Nei M and Kumar S. 2004.** Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA* **101**: 11030-11035.

52. **Tamura K, Stecher G, Peterson D, Filipski A and Kumar S. 2013.** MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.

53. **Felsenstein J. 1985.** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791.

54. **Jones DT, Taylor WR and Thornton JM. 1992.** The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**: 275-282.

55. **Vorimore F, Hsia RC, Huot-Creasy H, Bastian S, Deruyter L, Passet A, Sachse K, Bavoil P, Myers G, and Laroucau K. 2013.** Isolation of a new *Chlamydia* species from the Feral Sacred Ibis (*Threskiornis aethiopicus*): *Chlamydia ibidis*. *PLoS ONE* **8**: e74823.

56. **Corsaro D and Greub G. 2006.** Pathogenic potential of novel Chlamydiae and diagnostic approaches to infections due to these obligate intracellular bacteria. *Clin Microbiol Rev* **19**: 283-297.

57. **Greub G. 2009.** *Parachlamydia acanthamoebae*, an emerging agent of pneumonia. *Clin Microbiol Infect* **15**: 18-28.

58. **Lamoth F, Pillonel T and Greub G. 2015.** *Waddlia*: An emerging pathogen and a model organism to study the biology of chlamydiae. *Microbes Infect*: S1286-4579. doi: 10.1016/j.micinf.2015.09.021.

59. **Sachse K, Laroucau K, Riege K, Wehner S, Dilcher M, Creasy HH, Weidmann M, Myers G, Vorimore F, Vicari N, Magnino S, Liebler-Tenorio E, Ruettger A, Bavoil PM, Hufert FT, Rosselló-Móra R and Marz M. 2014.** Evidence for the existence of two new members of the family *Chlamydiaceae* and proposal of *Chlamydia avium* sp. nov. and *Chlamydia gallinacea* sp. nov. *Syst Appl Microbiol* **37**: 79-88.

60. **Ramsey KH. 2006.** Alternative mechanisms of pathogenesis. *In Chlamydia*: Genomics and pathogenesis. Editors Bavoil PM and Wyrick PB, Pages 435-475. Horizon Bioscience, Norfolk, UK.

61. **Ward ME. 1999.** Mechanism of *Chlamydia*-induced disease. *In*: *Chlamydia*: Intracellular biology, pathogenesis, immunity. Editor Stephens RS, Pages 29-67. ASM Press, Washington, DC.

62. **Wang C, van Ginkel FW, Kim T, Li D, Li Y, Dennis JC and Kaltenboeck B. 2008.** Temporal delay of peak T-cell immunity determines *Chlamydia pneumoniae* pulmonary disease in mice. *Infect Immun* **76**: 4913-4923.

63. **Wang C, Gao D and Kaltenboeck B. 2009.** Acute *Chlamydia pneumoniae* reinfection accelerates the development of insulin resistance and diabetes in obese C57BL/6 mice. *J Infect Dis* **200**: 279-287.

64. **Brunham RC and Rey-Ladino J. 2005.** Immunology of *Chlamydia* infection: Implications for a *Chlamydia trachomatis* vaccine. *Nat Rev Immun* **5**:149-161.

65. **Morrison RP, Feilzer K and Tumas DB. 1995.** Gene knockout mice establish a primary protective role for major histocompatibility complex class II-restricted responses in *Chlamydia trachomatis* genital tract infection. *Infect Immun* **63**: 4661-4668.

66. **Darville TJ, O'Neill M, Andrews Jr CW, Nagarajan UM, Stahl L and Ojcius DM. 2003.** Toll-like receptor-2 but not Toll-like receptor-4 is essential for development of oviduct pathology in chlamydial genital tract infection. *J Immunol* **171**: 6187-6197.

67. **Grayston JT, Wang SP, Yeh LJ and Kuo CC. 1985.** Importance of reinfection in the pathogenesis of trachoma. *Rev Infect Di* **7**: 717-725.

68. **Grayston JT and Wang SP. 1975.** New knowledge of chlamydiae and the diseases they cause. *J Infect Dis* **132**: 87-105.

69. **Kuo CC, Shor A, Campbell LA, Fukushi H, Patton DL and Grayston JT. 1993.** Demonstration of *Chlamydia pneumoniae* in atherosclerotic lesions of coronary arteries. *J Infect Dis* **167**: 841-849.

70. **Shewen PE. 1980.** Chlamydial infection in animals: A review. *Can Vet J* **21**: 2-11.

71. **Reinhold P, Sachse K and Kaltenboeck B. 2011.** *Chlamydiaceae* in cattle: Commensals, trigger organisms, or pathogens? *Vet J* **189**: 257-267.

72. **Harkinezhad T, Geens T and Vanrompay D. 2009.** *Chlamydophila psittaci* infections in birds: a review with emphasis on zoonotic consequences. *Vet Microbiol* **135**: 68-77.

73. **Pedersen LN, Herrmann B and Møller JK. 2009.** Typing *Chlamydia trachomatis*: from egg yolk to nanotechnology. *FEMS Immunol Med Microbiol* **55**: 120-130.

74. **Menon S, Timms P, Allan JA, Alexander K, Rombauts L, Horner P, Keltz M, Hocking J and Huston WM. 2015.** Human and pathogen factors associated with *Chlamydia trachomatis*-related infertility in women. *Clin Microbiol Rev* **28**: 969-985.

75. **Manavi K. 2006.** A review on infection with *Chlamydia trachomatis*. *Best Pract Res Clin Obstet Gynaecol* **20**: 941-951.

76. **Zhan P, Suo LJ, Qian Q, Shen XK, Qiu LX, Yu LK and Song Y. 2011.** *Chlamydia pneumoniae* infection and lung cancer risk: a meta-analysis. *Eur J Cancer* **47**: 742-747.

77. **Chen J, Zhu M, Ma G, Zhao Z and Sun Z. 2013.** *Chlamydia pneumoniae* infection and cerebrovascular disease: a systematic review and meta-analysis. *BMC Neurol* **13**: 183.

78. **Su X and Chen HL. 2014.** *Chlamydia pneumoniae* infection and cerebral infarction risk: a meta-analysis. *Int J Stroke* **9**: 356-364.

79. **Littman AJ, Jackson LA and Vaughan TL. 2005.** *Chlamydia pneumoniae* and lung cancer: epidemiologic evidence. *Cancer Epidemiol Biomarkers Prev* **14**: 773-778.

80. **Filardo S, Di Pietro M, Farcomeni A, Schiavoni G and Sessa R. 2015.** *Chlamydia pneumoniae*-mediated inflammation in atherosclerosis: A meta-analysis. *Mediators Inflamm:* 378658. doi: 10.1155/2015/378658.

81. **Poudel A, Elsasser TH, Rahman KS, Chowdhury EU and Kaltenboeck B. 2012.** Asymptomatic endemic *Chlamydia pecorum* infections reduce growth rates in calves by up to 48 percent. *PLoS One* **7**: e44961.

82. **Poudel A, Reid E, Newport-Nielsen E, Chowdhury EU, Rahman KS, Sartin J and Kaltenboeck B. 2014.** Endemic asymptomatic *Chlamydia pecorum* genital tract infection reduces fertility in dairy cows. *Proc 13th International Symposium on Human Chlamydial Infections*, pp 258-260. Pacific Grove, CA.

83. **Joseph T, Nano FE, Garon CF and Caldwell HD. 1986.** Molecular characterization of *Chlamydia trachomatis* and *Chlamydia psittaci* plasmids. *Infect Immun* **51**: 699-703.

84. **Palmer L and Falkow S. 1986.** A common plasmid of *Chlamydia trachomatis*. *Plasmid* **16**: 52-62.

85. **Peterson EM, Markoff BA, Schachter J and Luis M. 1990.** The 7.5-kb plasmid present in *Chlamydia trachomatis* is not essential for the growth of this microorganism. *Plasmid* **23**: 144-148.

86. **Stothard DR, Williams JA, Van Der Pol B and Jones RB. 1998.** Identification of a *Chlamydia trachomatis* serovar E urogenital isolate which lacks the cryptic plasmid. *Infect Immun* **66**; 6010-6013.

87. **Dugan J, Rockey DD, Jones L and Andersen AA. 2004.** Tetracycline resistance in *Chlamydia suis* mediated by genomic islands inserted into the chlamydial inv-like gene. *Antimicrob Agents Chemother* **48**: 3989-3995.

88. **Wang SP and Grayston JT. 1970.** Immunologic relationship between genital TRIC, lymphogranuloma venereum, and related organisms in a new microtiter indirect immunofluorescence test. *Am J Ophthalmol* **70**: 367-374.

89. **Wang SP, Kuo CC and Grayston, J. T. 1973.** A simplified method for immunological typing of trachoma-inclusion conjunctivitis-lymphogranuloma venereum organisms. *Infect Immun* **7**: 356-360.

90. Batteiger BE, Newhall WJ, Terho P, Wilde CE and Jones RB. 1986. Antigenic analysis of the major outer membrane protein of *Chlamydia trachomatis* with murine monoclonal antibodies. *Infect Immun* **53**: 530-533.

91. Kaltenboeck B, Kousoulas KG and Storz J. 1993. Structures of and allelic diversity and relationships among the major outer membrane protein (*ompA*) genes of the four chlamydial species. *J Bacteriol* **175**: 487-502.

92. Kaltenboeck B, Schmeer N and Schneider R. 1997. Evidence for numerous *omp*1 alleles of porcine *Chlamydia trachomatis* and novel chlamydial species obtained by PCR. *J Clin Microbiol* **35**: 1835-1841.

93. Kaltenboeck B, Heinen E, Schneider R, Wittenbrink MM and Schmeer N. 2009. OmpA and antigenic diversity of bovine *Chlamydophila pecorum* strains. *Vet Microbiol* **135**: 175-180.

94. Andersen AA. 1991. Serotyping of *Chlamydia psittaci* isolates using serovar-specific monoclonal antibodies with the microimmunofluorescence test. *J Clin Microbiol* **29**: 707-711.

95. Guo W, Li J, Kaltenboeck B, Gong J, Fan W and Wang C. 2015. *Chlamydia gallinacea* is the endemic chlamydial species in chicken (*Gallus gallus*). *Submitted*.

96. Opota O, Jaton K, Branley J, Vanrompay D, Erard V, Borel N, Longbottom D and Greub G. 2015. Improving the molecular diagnosis of *Chlamydia psittaci* and *Chlamydia abortus* infection with a species-specific duplex real-time PCR. *J Med Microbiol* **64**: 1174-1185.

97. de Vries HJ, van der Loeff MFS and Bruisten SM. 2015. High-resolution typing of *Chlamydia trachomatis*: epidemiological and clinical uses. *Curr Opin Infect Dis* **28**: 61-71.

98. Sachse K, Laroucau K, Vorimore F, Magnino S, Feige J, Müller W, Kube S, Hotzel H, Schubert E, Slickers P and Ehricht R. 2009. DNA microarray-based

genotyping of *Chlamydophila psittaci* strains from culture and clinical samples. *Vet Microbiol* **135**: 22-30.

99. Sachse K, Vretou E, Livingstone M, Borel N, Pospischil A and Longbottom D. 2009. Recent developments in the laboratory diagnosis of chlamydial infections. *Vet Microbiol* **135**: 2-21.

100. Wang SP and Grayston JT. 1963. Classification of trachoma virus strains by protection of mice from toxic death. *J Immunol* **90**: 849–856.

101. Wang SP, Kuo CC, Barnes RC, Stephens RS and Grayston JT. 1985. Immunotyping of *Chlamydia trachomatis* with monoclonal antibodies. *J Infect Dis* **152**: 791–800.

102. Wang SP. 2000. The microimmunofluorescence test for *Chlamydia pneumoniae* infection: technique and interpretation. *J Infect Dis* **181**: S421-S425.

103. Stephens RS, Tam MR, Kuo CC and Nowinski RC. 1982. Monoclonal antibodies to *Chlamydia trachomatis*: antibody specificities and antigen characterization. *J Immunol* **128**: 1083-1089.

104. Barnes RC, Rompalo AM and Stamm WE. 1987. Comparison of *Chlamydia trachomatis* serovars causing rectal and cervical infections. *J Infect Dis* **156**: 953-958.

105. Haralambieva I, Iankov I, Petrov D, Ivanova R, Kamarinchev B and Mitov I. 2001. Cross-reaction between the genus-specific lipopolysaccharide antigen of *Chlamydia* spp. and the lipopolysaccharides of *Porphyromonas gingivalis*, *Escherichia coli* O119 and *Salmonella newington*: Implications for diagnosis. *Diagn Microbiol Infect Dis* **41**: 99-106.

106. Bourke SJ, Carrington D, Frew CE, Stevenson RD and Banham SW. 1989. Serological cross-reactivity among chlamydial strains in a family outbreak of psittacosis. *J Infect* **19**: 41-45.

107. Chirgwin K, Roblin PM, Gelling M, Hammerschlag MR and Schachter J. 1991. Infection with *Chlamydia pneumoniae* in Brooklyn. *J Infect Dis* **163**: 757-761.

108. Black CM, Fields PI, Messmer TO and Berdal BP. 1994. Detection of *Chlamydia pneumoniae* in clinical specimens by polymerase chain reaction using nested primers. *Eur J Clin Microbiol Infect Dis* **13**: 752-756.

109. Emre U, Roblin PM, Gelling M, Dumornay W, Rao M, Hammerschlag MR and Schachter J. 1994. The association of *Chlamydia pneumoniae* infection and reactive airway disease in children. *Arch Pediatr Adolesc Med* **48**: 727-732.

110. Persson K and Boman J. 2000. Comparison of five serologic tests for diagnosis of acute infections by *Chlamydia pneumoniae*. *Clin Diagn Lab Immunol* **7**: 739-744.

111. Dowell SF, Peeling RW, Boman J, Carlone GM, Fields BS, Guarner J, Hammerschlag MR, Jackson LA, Kuo CC, Maass M, Messmer TO, Talkington DF, Tondella ML and Zaki, S. R. 2001. Standardizing *Chlamydia pneumoniae* assays: recommendations from the centers for disease control and prevention (USA) and the laboratory centre for disease control (Canada). *Clin Infect Dis* **33**: 492-503.

112. Ossewaarde JM, Tuuminen T, Boersma WG, Sandström M, Palomäki P and Boman J. 2000. A preliminary evaluation of a new enzyme immunoassay to detect *Chlamydia pneumoniae*-specific antibodies. *J Microbiol Methods* **43**: 117-125.

113. Peeling RW, Wang SP, Grayston JT, Blasi F, Boman J, Clad A, Freidank H, Gaydos CA, Gnarpe J, Hagiwara T, Jones RB, Orfila J, Persson K, Puolakkainen M, Saikku P, Schachter J. 2000. *Chlamydia pneumoniae* serology: interlaboratory variation in microimmunofluorescence assay results. *J Infect Dis* **181**: S426-S429.

114. **Tuuminen T, Palomäki P and Paavonen J. 2000.** The use of serologic tests for the diagnosis of chlamydial infections. *J Microbiol Methods* **42**: 265-279.

115. **Boman J and Hammerschlag MR. 2002.** *Chlamydia pneumoniae* and atherosclerosis: critical assessment of diagnostic methods and relevance to treatment studies. *Clin Microbiol Rev* **15**: 1-20.

116. **Wang SP and Grayston JT. 1974.** Human serology in *Chlamydia trachomatis* infection with microimmunofluorescence. *J Infect Dis* **130**: 388-397.

117. **Wang SP. 2000.** The microimmunofluorescence test for *Chlamydia pneumoniae* infection: technique and interpretation. *J Infect Dis* **181**: S421-S425.

CHAPTER 2

2.1 INTRODUCTION

Chlamydiae are obligate intracellular bacteria that replicate in eukaryotic cells within membrane bound vacuoles (1). Infectious, but non-replicating elementary bodies (EBs) and non-infectious, but metabolically active reticulate bodies (RBs) are two unique physiological forms for chlamydiae (1). *Chlamydia (C.)* spp. cause a variety of diseases in humans, other mammals, and birds (1). Until very recently, nine species were recognized including *C. abortus*, *C. psittaci*, *C. caviae*, *C. felis*, *C. pecorum*, *C. pneumoniae*, *C. muridarum*, *C. suis* and *C. trachomatis* (1). In 2014, the new species, *C. avium* and *C. gallinacea* were published (2) which are not part of the present study. In 1999, Everett et al. proposed to subdivide *Chlamydiaceae* into two genera, *Chlamydia* (consisting of *C. muridarum*, *C. suis* and *C. trachomatis*) and *Chlamydophila* (consisting of the remaining six species) (3). This subdivison has now been formally reversed to a single *Chlamydia* genus consisting of 11 *Chlamydia* spp. (4).

Numerous serovars of *C. suis* (5), *C. pecorum* (6-7), *C. trachomatis* (8-9) and *C. psittaci* (10-12) have been reported. Different serovars of a species cause multiple diseases in a single host, e.g. serovars of *C. trachomatis* cause trachoma, infections of reproductive organs, or lymphogranuloma venereum in humans (1). Single hosts can also be infected by multiple *Chlamydia* spp., e.g. humans may be infected by both respiratory transmitted *C. pneumoniae* and sexually transmitted *C. trachomatis* (1), or by *C. trachomatis* and *C. psittaci* in the case of trachoma patients (13). Antibodies produced against one species strongly cross-react with other species, making interpretation of

serological assays difficult (14). For instance, anti-*C. pneumoniae* antibodies may interfere with the diagnosis of sexually transmitted diseases caused by *C. trachomatis* due to cross-reactivity of chlamydial antigens in serological-assays (15-18).

Acute, chronic, or asymptomatic infections with *C. trachomatis* and *C. pneumoniae* have a significant impact on human health (19). Infections with all *Chlamydia* spp. may occur with epidemic to endemic prevalence, with sporadic, subclinical and occasional clinical manifestations in a wide range of animal hosts, resulting in significant economic impact on animal production (20-21). Occasional transmission of *Chlamydia* spp. from animals to humans has been reported (1). Specific serological assays to detect anti-*Chlamydia* antibodies are important to provide differential diagnosis of chlamydial infections for patient care, and to understand chlamydial diseases and epidemiology.

The microimmunofluorescence (MIF) test is the standard serological assay for species-specific detection of antibodies against chlamydiae (22). Detection of specific anti-chlamydial antibodies for nine species and their serovars using the MIF test requires cumbersome production of antigens by growing these *Chlamydia* spp. and their numerous strains in cell culture or developing chicken embryos. Standardization of the MIF test also requires technical expertise in microscopy and antigen preparation that is available only in reference laboratories. Nevertheless, poor sensitivity and cross-reactivity of the MIF test have been reported (15- 18, 23-24). Simple and high-throughput methods are typically unsuitable for species- or serovar-specific anti-chlamydial antibody detection because of high cross-reactivity of standard chlamydial antigens in serological assays,

such as whole *Chlamydia* elementary bodies (EB), lysed EBs, immunodominant proteins, or lipopolysaccharide (23-25).

Chlamydia genus-, species-, subspecies-, and serotype-specific B-cell epitopes have been mapped before to the four variable domains (VD) of the outer membrane protein A (OmpA) by use of monoclonal antibodies (26-27), recombinant protein fragments (28-31), and synthetic peptides (32-36). Based on these epitope mapping studies, synthetic OmpA peptides were tested for *Chlamydia* species-specific serology (23-25, 37-38). However, these studies used peptides as short as 6-10 amino acids and did not use spacers between solid support and peptide to minimize steric hindrance of antibody binding (32-36). Recent computational studies of antigen-antibody complex 3D structures showed that 15-25 amino acid (AA) residues of an epitope are structurally involved in antibody binding (39-42). Short 6-10 AA peptides tend to capture only antibodies binding to linear epitopes composed of adjacent functional residues that comprise less than 10% of all epitopes (42). In contrast, longer peptides capture antibodies binding to conformational epitopes with functionally binding residues spaced apart over 16-30 AA sequences that comprise 55-80% of all epitopes (39). The presence of all functional residues of an epitope also contributes to high-affinity antibody binding (42). Thus, the probability of high-affinity binding is proportional to the length of a peptide antigen. These recent data suggest that previous studies failed to achieve high sensitivity (23-25, 43-44) most likely due to weak antibody binding to the short peptides used (32-36) and due to steric hindrance of antibody binding to these peptides (45).

The long term objectives in the present study have been to identify unique B-cell epitopes among the complete proteomes of nine *Chlamydia* species, and to use these

peptides in specific and sensitive *Chlamydia* species-specific ELISAs. This goal is now facilitated by the availability of complete genome sequences of all chlamydial species and of many serovars of some species (46-57), by proteome-wide mapping of immunodominant proteins in several chlamydial species (58-66), and by advanced computational tools for B-cell epitope discovery (39-42). Using a murine respiratory infection model, we report here peptide based molecular serology for nine *Chlamydia* spp. in a robust and high-throughput ELISA format by identifying immunodominant B-cell epitopes and determining their specific reactivities with mouse hyperimmune sera against these nine *Chlamydia* species. With sera from cattle naturally infected with *C. pecorum* we also confirmed the suitability of these assays for chlamydial sero-epidemiology.

2.2. MATERIALS AND METHODS

***Chlamydia* strains.** Of the nine *Chlamydia* spp. (1), strains *C. abortus* B577 (ATCC: VR 656, DSMZ: DSM 27654), *C. pecorum* E58 (ATCC: VR 628), *C. pneumoniae* CDC/CWL-029 (ATCC: VR-1310), and *C. trachomatis* D UW-3/CX (ATCC VR 885, DSMZ: DSM 19411) were grown on Buffalo Green Monkey Kidney (BGMK) cells, and elementary bodies (EBs) were purified and stored at -80 °C (67). Strains of the remaining five *Chlamydia* spp., i.e. *C. psittaci* 02DC15 (DSMZ: DSM 27008), *C. caviae* GPIC (ATCC: VR 813, DSMZ: DSM 19441), *C. felis* 02DC26 (collection FLI Jena, Germany), *C. muridarum* MoPn/Nigg (ATCC: VR 123, DSMZ: DSM 28544), and *C. suis* 99DC3 (collection FLI), were propagated in developing chicken embryos, and infected yolk sacs were homogenized in sucrose phosphate glutamate (SPG) buffer with disposable tissue grinders and stored at -80°C (68). Chlamydial genome copy numbers in stocks were quantified by *Chlamydia* spp. 23S rRNA PCR (69-70).

***Chlamydia* spp. mono-specific mouse hyper-immune sera.** A/J and BALB/c mice were used to generate sera against each of the 9 chlamydial species. Mice were inoculated intranasally under light isoflurane anesthesia with 20 µl chlamydial stocks (71). High doses of SPG-diluted chlamydiae (10^3 to 10^8 genomes/mouse) were inoculated three times in 4-6-week intervals to produce high-titered, high-affinity IgG (immunoglobulin G) antibodies against chlamydial antigens encountered during natural *Chlamydia* infection. To minimize losses of mice, particularly for highly virulent strains, doses at the low end of the used spectrum were chosen for the first inoculation (1.0×10^7, 4.9×10^4, 1.3×10^5, 1.4×10^5, 3.6×10^8, 1.3×10^6, 9.7×10^2, 2.6×10^4, 5.0×10^7 genomes of *C.*

abortus, C. psittaci, C. caviae, C. felis, C. pecorum, C. pneumoniae, C. muridarum, C. suis and *C. trachomatis,* respectively). To maximize antigenic stimulus, these doses were increased by 2-10-fold in the 2nd (2.0×10^7, 4.9×10^5, 1.3×10^6, 5.6×10^5, 3.6×10^8, 5.0×10^6, 3.9×10^3, 1.3×10^5, and 1.0×10^8 genomes, respectively) and 3rd inoculations (4.0×10^7, 4.9×10^5, 1.3×10^6, 2.8×10^6, 3.6×10^8, 5.0×10^6, 9.7×10^3, 6.5×10^5 and 2.0×10^8 genomes, respectively). Mice were exsanguinated 3 weeks after the last inoculation by axillary cut-down under anesthesia. Heparin plasma samples were collected in microtainer tubes with a gel band (Becton, Dickinson & Company, Franklin Lakes, NJ) by centrifugation at 3,000 rpm and stored at -80°C. Although plasma and serum cannot be interchangeably used in most instances, we termed this heparinized plasma as serum here forth for simplicity. In the final species-specific serum pools, a total of 47 *C. abortus*, 12 *C. psittaci*, 9 *C. caviae*, 48 *C. felis*, 50 *C. pecorum*, 39 *C. pneumoniae*, 28 *C. muridarum*, 28 *C. suis* and 47 *C. trachomatis*-specific sera were combined. Of these, 18 sera for *C. abortus*, 8 for *C. psittaci*, 48 for *C. felis*, 10 for *C. muridarum* and 30 for *C. trachomatis* were from Balb/c mice, and the remaining sera were from A/J mice. All animal experimental protocols were approved by the Institutional Animal Care and Use Committees at Auburn University and Friedrich-Loeffler-Institute.

Bovine anti-*C. pecorum* sera. Sera were collected from cows that had experienced multiple episodes of natural infection with multiple *C. pecorum* strains (72), while calves were sampled between 11-15 weeks of age, after serum colostrum antibodies had disappeared and calves had experienced a first episode of *C. pecorum* infection (20). The *C. pecorum* infection of all calves and cows had been confirmed by detection of chlamydial DNA by real-time PCR for the *Chlamydia* spp. 23S rRNA gene, and

by genotyping of infecting *C. pecorum* strains by *Chlamydia* spp. ompA PCR and DNA sequencing (20, 72).

Chlamydial peptide sequences for epitope discovery. Matching type strains and genome sequences were available and used for raising hyper-immune sera for five chlamydial species: *C. trachomatis* (strain D/UW-3/CX) (46), *C. pneumoniae* (strain CWL029) (47), *C. muridarum* (strain Nigg) (48), *C. pecorum* (strain E58) (49), *C. caviae* (strain GPIC) (50), and peptides were designed from the type strain proteomes. The genome of the non-type *C. psittaci* strain 02DC15 used in this investigation was available and virtually identical to type strain 6BC (51-53), and peptides were designed from the 02DC15 proteome. For *C. abortus* type strain B577 and *C. felis* non-type strain 02DC26 used for raising antisera, the genome sequences were not available. Instead the genomes of *C. abortus* strain S26/3 (54) and of *C. felis* strain Fe/C-56 (55) were used to design peptide antigens. Compared to these strains with known genome, the strains used to raise antisera had an identical OmpA sequence (73) or an OmpA with a single polymorphism (*C. felis* Fe/C-56; GenBank accession number KP165540), and these species are also known to show minimal strain diversity based on available genomes and multilocus sequence typing (NCBI GenBank database). As the last one of the nine chlamydial species, the first complete genome of *C. suis* non-type strain MD56 has only recently become available (56). Initial sequences used in the alignment were deduced from the incomplete genome of type strain S45 (personal communication G. Myers). For raising hyper-immune sera, *C. suis* strain 99DC3 was used. Because of the high diversity of *C. suis* strains (5), peptides of antibody-reactive regions of *C. suis* strain 99DC3 were confirmed by DNA sequencing. The sequences of the gene fragments of *ompA*, *omp2*,

pmpD, *incA*, *incG*, and of *C. suis* 99DC3 homologs to CT529 and CT618 are available at GenBank (accession numbers KP165534- KP165539, KP165542).

Bioinformatic analyses of chlamydial genomes, immunodominant proteins, and B-cell epitopes. To identify species-specific B-cell epitopes, we first identified and ranked 72 immunodominant proteins among all chlamydial proteomes (46-57), based on published data (58-66, 74-82). These protein sequences of all nine chlamydial species, and of serovars among species, were first aligned in the freeware Jalview (83) by use of the MUSCLE algorithm and weights for amino acid substitutions based on the Blosum62 AA substitution matrix (83-85). Alignments were optimized by varying alignment parameters, in particular increasing gap opening and extension penalties, or by manual editing. Polymorphic regions suitable for identification of species-specific epitope candidates were further subjected to *in silico* B-cell epitope analyses. Optimal predictive algorithms based on recent knowledge of B-cell epitope structure and length (39-42) were used to define peptides for testing. These algorithms determined protein intrinsically unstructured/disordered tendency (86), relative solvent accessibility/surface exposed tendency (87), and hydrophilicity (88). A linear combination of these scores was used to rank and select peptides from polymorphic regions for species-specific reactivity, as well as from selected conserved regions for genus-specific reactivity.

For display of aligned peptide sequences, phylogenetic trees were constructed by the UPGMA method (Unweighted Pair Group Method with Arithmetic mean, 83). Within-tree sequence distances were calculated by percentage identity (PID) of amino acids of antibody reactive regions (83).

Determination of peptide antigen seroreactivity by chemiluminescent ELISA. Peptides were chemically synthesized with an N-terminal biotin followed by a serine-glycine-serine-glycine (SGSG) spacer (89-90), the specific sequence, and a carboxyl C-terminus (AAPPTec, Louisville, KY, USA; ProImmune, Oxford, UK; and GenScript, Piscataway, NJ, USA). White flat-bottom microtiter plates coated with covalently linked streptavidin were used for binding of biotinylated peptides (Nunc; Fisher Scientific, Pittsburgh, PA, USA). Peptides were initially dissolved at 2.5 µM/mL in DMSO (dimethyl sulfoxide, Sigma-Aldrich, St. Louis, MO, USA). For coating microtiter plates, each peptide was diluted first to 25 nM/mL in DMSO, then to 0.25 nM/mL in assay diluent (0.15 M NaCl, 20 mM Tris-HCl, pH 7.5 containing 2% rabbit serum, 0.2% bovine serum albumin, 0.2% casein, 0.2% polyethylene glycol, 0.05% Tween-20, 0.004% benzalkonium chloride, Sigma-Aldrich, St. Louis, MO, USA).

Plates were incubated for 15 min at room temperature with 400 µl wash buffer/well (0.25 M NaCl, 20 mM Tris-HCl, pH 7.5, containing 0.1% Tween-20, 0.001% benzalkonium chloride), then rinsed twice with 300 µl wash buffer. Subsequently, 100 µl of peptide solutions in assay diluent (0.25 nM/mL) were added per well (25 picoMoles peptide), and incubated at room temperature for 30 min with agitation. Peptide solutions were aspirated and the plates were washed 5 times with 300 µl wash buffer, followed by 30 min incubation with 300 µl blocking buffer per well (0.15 M NaCl, 20 mM Tris-HCl, pH 7.5 containing 10% rabbit serum, 1% bovine serum albumin, 1% casein, 1% polyethylene glycol, 0.004% benzalkonium chloride). To capture low-affinity antibodies, lower concentrations of NaCl were used in wash buffer (0.15 M), assay diluent (0.1 M) and blocking buffer (0.1 M). Sera were diluted 1:50 or 1:100 in assay diluent, 100 µl of

diluted sera was added per well, and incubated for 1 hour with agitation at room temperature. After 5 washes with wash buffer, 100 µl secondary antibody diluted 1:1,000 in assay diluent was added (horseradish peroxidase conjugated goat anti-mouse IgG-$F(ab')_2$ fragment; Bethyl Laboratories, Inc., Montgomery, TX, USA), and microtiter plates were incubated for 30 min. After 5 washes, 100 µl of freshly prepared HRP (horseradish peroxidase) chemiluminescent substrate (Roche Diagnostics Corporation, Indianapolis, IN, USA), and luminescence was determined after 2 min on a Spectrafluor Plus reader at 500 ms integration time at 100% gain and maximum dynamic range. The signals were indicated as relative light units per second (rlu/sec), and for ease of display divided by 1,000 ($rlu/sec \times 10^{-3}$).

For bovine sera, modified assay conditions were used. The assay diluent was 0.2 M NaCl, 20 mM Tris-HCl, pH 7.5, containing 10% chicken serum, 0.5% polyethylene glycol, 0.1% Tween-20, 0.004% benzalkonium chloride. Blocking buffer was 0.2 M NaCl, 20 mM Tris-HCl, pH 7.5, 10% chicken serum, 1% polyethylene glycol, 0.004% benzalkonium chloride, and wash buffer was 0.3 M NaCl, 20 mM Tris-HCl, pH 7.5, containing 0.1% Tween-20, and 0.001% benzalkonium chloride. Goat anti-bovine IgG-heavy and light chain HRP conjugate (Bethyl Laboratories, Inc., Montgomery, TX, USA) was diluted 1:500.

All sera were analyzed in wells coated with specific peptides and in a non-coated well, and for the final background-corrected results 150% of the background signal (mean + 2 SD) in the non-coated well of each serum was subtracted from its specific peptide signals. For all peptides, we stringently defined species-specific reactivity as any reactivity above this background with the homologous mouse serum pool, but no signal

above the background with any of the 8 heterologous serum pools. Any peptide reactive with a heterologous mouse serum pool was considered cross-reactive. To avoid false-positive results in quantitative evaluation of the reactivity of any peptide with individual mouse sera, and with bovine sera from naturally infected cattle, we used a more stringent cut-off of 10,000 rlu/sec.

Statistical analyses. All statistical analyses were performed by the software package Statistica 7.1 software package (Statsoft, Tulsa, Oklahoma, USA), and p values ≤ 0.05 were considered significant. Differences between means of peptide reactivities and background were analyzed by paired one-tailed Student t test. The probability of cross-reactivity among peptides of a clade was estimated by logistic regression against percent sequence identity. Signal strength of cross-reactive peptides in dependence of sequence identity was calculated in linear regression analysis. For analysis of categorical data, one- or two-tailed Fisher Exact test was used as described in appropriate sections.

2.3 RESULTS

Identification of antibody-reactive regions of chlamydial proteins. To identify protein regions containing strong B-cell epitopes, a total of 72 chlamydial proteins, described in the literature as immunodominant, were catalogued and ranked based on seroreactivity in the published literature, the number of studies describing the reactivity, and the evolutionary polymorphism among homologs of the nine *Chlamydia* spp. All available sequences of each immunodominant protein were aligned, and polymorphic peptide regions in each protein alignment were ranked based on B-cell epitope scores (polymorphism, disorder and surface exposed tendencies). Conserved peptide regions were ranked based predicted scores for disorder and surface-exposed tendencies, and hydrophilicity.

High scoring peptides were synthesized from sequences of *C. trachomatis* (42 proteins), *C. pneumoniae* (24 proteins), *C. abortus* (15 proteins), *C. pecorum* (12 proteins), and *C. muridarum* (11 proteins). Peptides were synthesized chemically with N-terminal biotin, followed by serine-glycine-serine-glycine (SGSG) as spacer/linker and the chlamydial sequence, and attached to streptavidin-coated white microtiter plates. Peptides were tested with homologous or heterologous *Chlamydia* mono-specific hyperimmune sera pooled from 9-50 mice for IgG antibodies in a chemiluminescent ELISA format. For the remaining *Chlamydia* spp., additional peptides that were homologous to the reactive peptides were tested.

TABLE 2.1. *Chlamydia* spp. proteins with antibody-reactive regions suitable for molecular serology

Protein	Locus tag[a]	Sequence identity[b]	Peptides evaluated[c]	ARRs identified[d]	References[e]
OmpA/MOMP Outer membrane protein A	CT_681	65-86	126	6	44, 58, 60-62, 64, 66, 74-77
PmpD Polymorphic outer membrane protein D	CT_812	33-87	129	5	58, 60, 62, 66, 77, 79
CT618 Inclusion membrane protein CT618	CT_618	23-72	41	3	59-61
IncA Inclusion membrane protein A	CT_119	12-80	182	1	59-65, 81
CT529 Inclusion membrane protein CT529	CT_529	31-88	33	1	59-62, 64
CrpA Cysteine rich outer membrane protein A	CT_442	20-88[f]	16	2	59-64
IncG Inclusion membrane protein G	CT_118	41-45[g]	17	1	59-60
OmcB/Omp2 Outer membrane cysteine rich protein B	CT_443	71-99	24	2	44, 58, 60-62, 64, 77-78
TarP Translocated actin-recruiting phosphoprotein	CT_456	24-92	20	1	60, 62-63, 66, 82
IncE Inclusion membrane protein E	CT_116	53-59[g]	5	1	59-62, 64

[a] **Gene locus tag in the genome of *C. trachomatis* strain D/UW-3/CX.**

[b] **Range of pairwise sequence identities among the nine *Chlamydia* spp.**

[c] **Peptides tested with mouse hyper-immune sera. Peptides from the following proteins did not react with mouse sera: CtrCpaf, CtrFtsH, CtrIncB, CtrIncC, CtrIncD, CtrLcrE, CtrIncF, CtrPmpB, CtrPmpC, CtrPkn5, CtrCT058, CtrCT089, CtrCT143, CtrCT147, CtrCT223, CtrCT226, CtrCT228, CtrCT228, CtrCT241, CtrCT381, CtrCT484, CtrCT541, CtrCT561, CtrCT579, CtrCT603, CtrCT619, CtrCT694, CtrCT741, CtrCT795, CtrCT813, CtrCT823, CtrCT875, Cpn0808, CpnPmp2, CpnPmp6, CpnPmp10, CpnPdhC, CpnPorB, CpnRecA, CpnRpsA, CpnRpsB, CpnYscC, CpnYscL, CpnYwbM, CpnCT858, Cpn0525, Cab063, CabCT058, CabCT541, CabIncAf, CabPmp15G, CpeCT143, CpeORF663 and CmuCT228.**

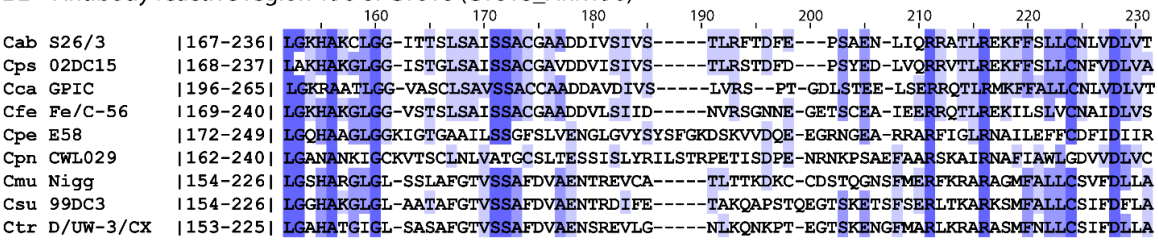
[d] **Antibody-reactive regions identified with mouse sera.**

[e] **Key references are cited.**

[f] **Homolog not found in *C. felis*.**

[g] **Homologs to the *C. trachomatis* protein found only in *C. muridarum* and *C. suis*.**

A Antibody reactive region 190 of CT618 (CT618_ARR190)



B CT618_ARR190 peptide reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_CT618_201-216	LRFTDFEPSAENLIQR	**1**	0	-	-	-	0	-	-	-
Cab_S26/3_CT618_187-226	ACGAADDIVSIVSTLRFTDFEPSAENLIQRRATLREKFFS	**35**	27	0	0	0	0	0	0	0
Cps_02DC15_CT618_189-228	CGAVDDVISIVSTLRSTDFDPSYEDLVQRRVTLREKFFSL	1	**610**	0	0	0	0	0	0	0
Cca_GPIC_CT618_230-245	VRSPTGDLSTEELSER	-	1	**155**	0	-	-	-	-	-
Cca_GPIC_CT618_220-249	ADDAVDIVSLVRSPTGDLSTEELSERRQTL	0	0	**486**	0	0	0	0	0	0
Cfe_Fe/C_CT618_193-222	ADDVLSIIDNVRSGNNEGETSCEAIEERRQ	0	0	0	**2**	-	-	-	-	-
Cpe_E58_CT618_211-226	DSKVVDQEEGRNGEAR	0	0	0	-	**51**	0	0	0	0
Cpe_E58_CT618_194-233	FSLVENGLGVYSYSFGKDSKVVDQEEGRNGEARRARFIGL	0	0	0	0	**115**	0	0	0	0
Cpn_CWL029_CT618_201-216	PETISDPENRNKPSAE	0	0	0	0	0	**167**	0	0	0
Cpn_CWL029_CT618_186-225	LTESSISLYRILSTRPETISDPENRNKPSAEFAARSKAIR	0	0	0	0	0	**1**	0	0	0
Cmu_Nigg_CT618_190-205	TKDKCCDSTQGNSFME	0	0	0	0	0	0	**267**	0	0
Csu_99DC3_CT618_183-212	DIFETAKQAPSTQEGTSKETSFSERLTKAR	-	-	-	-	-	-	7	**0**	6
Ctr_D/UW-3_CT618_186-201	NLKQNKPTEGTSKENG	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_CT618_192-207	PTEGTSKENGFMARLK	-	-	-	-	-	-	0	0	**1**
Ctr_D/UW-3_CT618_181-210	REVLGNLKQNKPTEGTSKENGFMARLKRAR	-	-	-	-	-	-	2	4	**4**
Background subtracted	Heterologous peptides or no peptide	4	12	11	4	4	4	4	4	4

Fig. 2.1. CT618 antibody-reactive region 190, CT618_ARR190.

(A) Complete ARR190 alignment of the putative inclusion membrane protein CT618 (CT618; locus tag CT_618 of C. trachomatis strain D/UW-3/CX) with the homologs of the remaining eight Chlamydia species. **The alignment includes the antibody-reactive polymorphic region bracketed by regions conserved among the nine chlamydial species. Numbers in vertical lines correspond to amino acid residue numbers of each species. Numbers on top of the alignment indicate approximate residue numbers of the Ctr_CT618 protein. The designation antibody-reactive region 190 (ARR190) is derived from the approximate central residue of the species-specific antibody determining region of the Ctr_CT618.**

(B) Antibody reactivity of CT618_ARR190 peptides with Chlamydia mono-species-specific antisera pooled from 9-50 mice. **Peptides are named for *Chlamydia* species-specificity (Cab, Cps, etc.) followed by strain and protein designation, and amino acid positions of the peptide in the respective proteins. Average reactivity of 3 repeats of each peptide at high stringency ELISA (inter-assay coefficient of variation (CV) = 11.2%, intra-assay CV = 8.5%). Bold-lettered peptide signals indicate the specific reactivity with homologous sera. All signals of positive serum pools are significantly above background ($p < 10^{-3}$; one-tailed Student's t-test). Bold-lettered peptides show strong species-specific reactivity with homologous sera. Cross-reactivity with Cps-specific sera is evident for the strongly-reactive Cab peptide Cab_S26/3_CT618_187-226.**

In total, 812 peptides were tested, and 23 antibody-reactive regions (ARRs) in 10 proteins were identified. These proteins, in order of ARR dominance and numbers, are OmpA/MOMP, PmpD, CT618, IncA, CT529, CT442, IncG, OmcB/Omp2, TarP, and IncE (Table 2.1). A total of 593 peptides were synthesized from the 23 ARRs of these 10 proteins for all nine *Chlamydia* spp. The remaining 219 peptides were synthesized for initial screening of all ranked proteins and did not show reactivity with mouse sera (Footnote *c* in Table 2.1). Collectively, we defined the reactivity of 134 genus-, species-, and serovar-specific peptides of B-cell epitopes of these proteins from nine chlamydial species.

Figure 2.1 is an example of an ARR, showing the alignment of the sequences of nine chlamydial species of a polymorphic region of chlamydial protein CT618 (Fig. 2.1A) and the reactivity of peptides from this region with a battery of corresponding nine species-specific anti-chlamydial sera (Fig. 2.1B). Cross-reactivity with heterologous sera was only found for *C. abortus* and *C. psittaci* peptides. All remaining peptides reacted significantly ($p < 10^{-3}$) only with homologous sera, thus provided species-specific reactivity (Fig. 1.1B). Of the total of 23 ARRs identified are shown in Figures 2.S1-S21. The remaining fourteen ARRs can be obtained from authors upon request.

Sequence divergence and cross-reactivity. To derive the relationship between sequence divergence and cross-reactivity, 93 peptides that had strongly reacted with their respective cognate antiserum were identified. Species/strain variant peptides of these 93 peptides were tested with sera raised by immunization with different chlamydial species (heterologous sera), resulting in 700 tests of peptides with heterologous sera (Fig. 2.2A). We defined peptide cross-reactivity as the ability of a peptide to bind antibodies elicited

by a different chlamydial species or strain with at least one AA variation in the peptide sequence (42).

Only peptides with 50-96% AA sequence identity to the respective heterologous chlamydial peptide showed cross-reactivity while those with less than 50% identity did not. As expected, the frequency of cross-reactivity strongly increased with sequence identity, with 11% of peptides with 50-60% identity, but 70% of peptides with 90-96% sequence identity. Logistic regression analysis showed that probability of cross-reactivity at 50%, 65% and 85% sequence identity was 4%, 20% and 80%, respectively (Fig. 2.2A). In linear regression analysis, the signal strength of these cross-reactive peptides relative to the signal of their homologous peptides did not correlate with percent sequence identity (Fig. 2.2B) and cross-reactive variant peptides produced on average 50%, and always less than 100%, of the signal generated by the homologous peptides. The results show that increasing sequence identity of heterologous peptides increases the probability, but not intensity, of cross-reactivity.

Phylogenetic reconstruction of reactive peptides predicts antigenic specificity. Figure 2.3 is an example of an ARR sub-region alignment of reactive peptides, and their strain variations within species (serovars). The phylogram of these CT618_ARR190 peptides enables calculation of the probability of peptide cross-reactivity using the logistic regression between phylogenetic distances and cross-reactivity as shown in Figure 2.2A. The 80% sequence identity between *C. abortus* and *C. psittaci* peptides in Figure 2.3, resulting in ~60% probability of cross-reactivity (Fig. 2.2A), translates into actual cross-reactivity as shown in Figure 2.1B. In contrast, none of the well-separated peptides shown in Figure 2.3 in cross-reactive (Fig. 2.1B).

Identification of mono-species and multi-species-specific reactive peptide antigens. To determine specificity of reactive peptides identified in the screening phase, they were tested with the nine mono-species-specific serum pools. Initial phylogenetic trees were constructed for each reactive peptide from its alignment with the homologous sequences from remaining 8 strains of the *Chlamydia* species used to raise the sera, followed by sub-trees constructed from all serovars of all 9 species (Fig. 2.1 and 2.3; Figures 2.S1-S8). The probability of cross-reactivity among peptides of a clade was estimated by logistic regression against percent sequence identity (Fig. 2.2A). Highly reactive peptides were grouped into *Chlamydia* single species- (Table 2.2) or multi species-specific peptides (Table 2.3) based on empirical and probabilistic cross-reactivity with heterologous sera. All peptide antigens in Table 2.2 react strongly with homologous sera, but not with the remaining eight heterologous sera. These peptides are also evolutionarily well separated from the remaining *Chlamydia* spp.

In addition to polymorphisms at the *Chlamydia* genus level, certain ARRs show polymorphism at the *Chlamydia* species level. For instance, all ARRs of the *C. suis* OmpA protein are highly divergent, followed by *C. pecorum*, *C. trachomatis* and *C. psittaci*. The *C. suis* OmpA peptide of AAs 166-181 (Csu_99DC3_OmpA_166-181) shows 25 sequence variants among just 55 available sequences, including 22 major variants with more than 2 AA difference (Table 2.2). In contrast, 22 sequence variants of this peptide are present among 741 *C. trachomatis* sequences, including 10 major

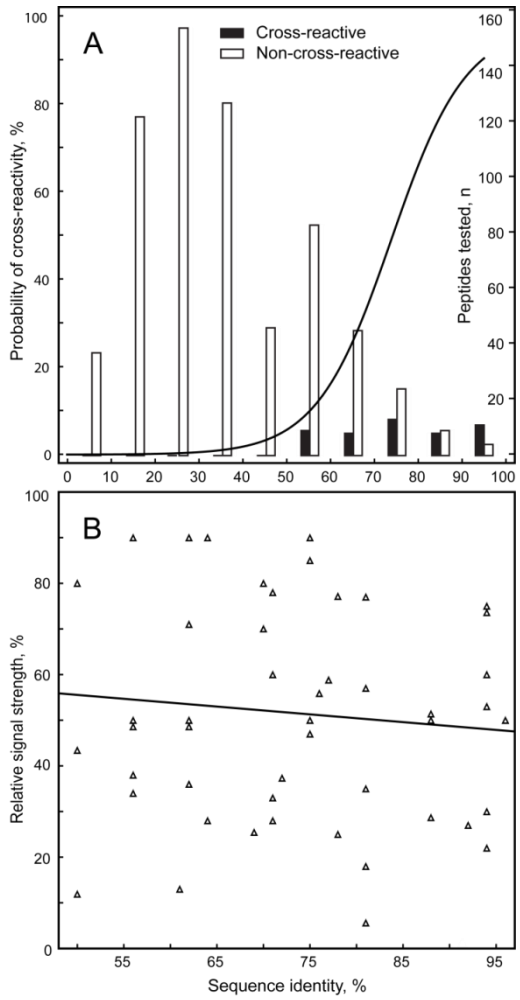


Fig. 2.2. Cross-reactivity of peptides with heterologous sera.

(A) **The probability of reactivity of a peptide with heterologous sera raised against the other chlamydial species/strains is shown in dependence of percent sequence identity of the peptide to the respective heterologous chlamydial peptide. Peptides that produced with the heterologous sera more than 5% of the signal of homologous sera were designated as cross-reactive. Probability of cross-reactivity analyzed by logistic regression (N = 700, $p < 10^{-6}$) is shown on the left ordinate, the number of peptides analyzed for each 10% bracket, visualized by bars, is shown on the right ordinate. Cross-reactive peptides are found exclusively at 50% or higher sequence identity.**

(B) **Signal strength of cross-reactive peptides relative to the signal of the homologous peptides in dependence of sequence identity. In linear regression analysis, the signal strength is independent of the degree of sequence identity (n=49, r^2=0.01, p=0.5).**

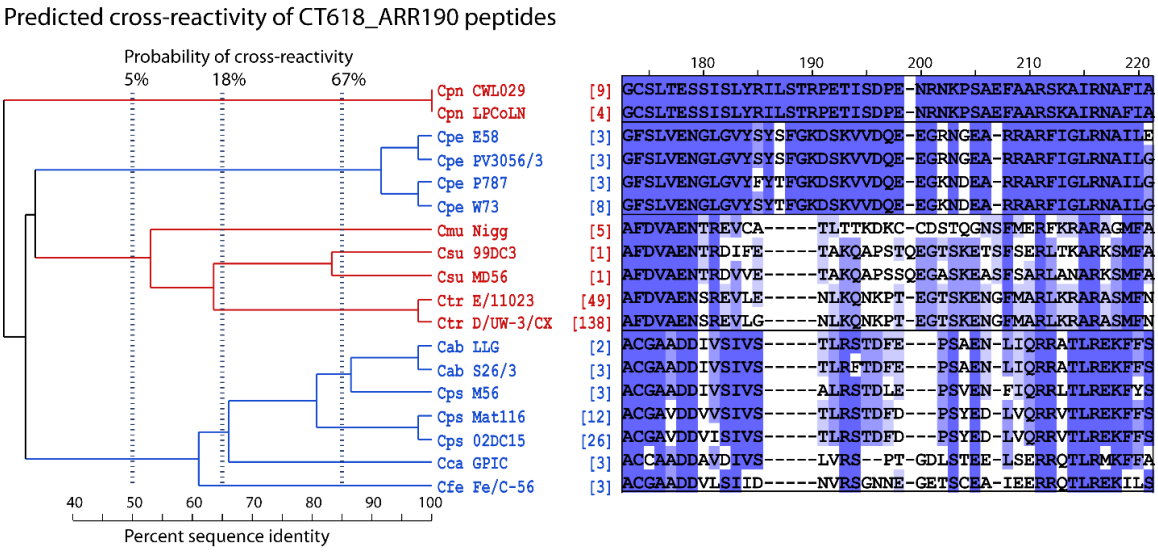


Fig. 2.3. Evolutionary relationship and predicted cross-reactivity among CT618_ARR190 sequences. Numbers on top of the sub-region alignment indicate approximate residue numbers of the Ctr_CT618 protein in the alignment of the overall region (Fig. 2.1). Corresponding residues in mismatched sequences are not always identical to the overall ARR alignment because of additional sequences, different sequence lengths, gap insertions, and higher gap insertion penalty in the sub-region alignments. Numbers in brackets indicate the frequency of the strain-specific sequence available in the NCBI protein database. For clarity, deep clades (sequences of typically less than 50% amino acid identity) are indicated in alternating color, with the sequences separated into boxes. Vertical dashed lines indicate the calculated probabilities of antibody cross-reactivity among peptides with a single common ancestor at these thresholds. This ARR shows high evolutionary divergence, allowing for robust differentiation of a chlamydial species (Cpn or Cpe) or a group of closely-related chlamydial species (clade Cmu, Csu & Ctr and clade Cab, Cps, Cca & Cfe). Cps species/strain-specific peptides are evolutionary closest to Cab followed by Cca and Cfe species/strain. Therefore, Cps peptides have the highest probability of cross-reactivity with Cab. Cpn or Cpe peptide sequences are evolutionary well-separated and these peptides have a very low probability of serological cross-reactivity with the remaining eight chlamydial species.

TABLE 2.2. Highly reactive *Chlamydia* species-specific peptides.

Peptide	Sequence	ELISA signal (rlu/sec ×10^{-3})[a]	Sequence Variants / Accessions[b]	Peptides / Species Detection[c]	ARR cross-reactivity[e]
Cab_S26/3_OmpA_153-176	NLVGLIGVKGSSIAADQLPNVGIT	365	2/39	1	I
Cab_S26/3_PmpD_1060-1089	KIESPTSNVYYSAHESVKQPENKTLADINS	185	2/7	1	II
Cab_S26/3_IncA_324-353	STAVTEHADIPRDPNRDPRGGRGGQSSPSV	109	2/5	1	
Cps_02DC15_OmpA_158-181	LVGLIGFSAASSISTDLPTQLPNV	106	30/288	8[d]	III
Cps_05DC15_OmpA_250-265	ASSNFPLPITAGTTEA	78	17/204	4[d]	IV
Cps_02DC15_OmpA_333-348	STTALPNNSGKDVLSD	299	20/209	5[d]	III
Cps_02DC15_IncA_321-360	SLTSTTETADQGDLRDPSGDRYGGWGAQSSYRLSPSVTMS	328	7/77	4[d]	V
Cps_02DC15_CT618_105-134	YEVDSATGSFKIVTKNIQKPNGEVEIVSSR	249	3/41	2	
Cps_02DC15_CT618_189-228	CGAVDDVISIVSTLRSTDFDPSYEDLVQRRVTLREKFFSL	343	3/41	2	
Cca_GPIC_OmpA_159-168	VTGTDLQGQY	117	1/5	1	
Cca_GPIC_IncA_316-355	LIGVMVQDGAESSTVEEASQDDSAQPQDENQSDAGEHKDS	321	1/4	1	
Cca_GPIC_CT618_134-163	EVDAQTGNFVLQTKTVQLEDGTQRVVPSRV	383	1/3	1	
Cca_GPIC_CT618_220-249	ADDAVDIVSLVRSPTGDLSTEELSERRQTL	180	1/3	1	
Cfe_Fe/C_OmpA_160-175	IGLAGTDFANQRPNVE	46	1/10	1	
Cfe_Fe/C_PmpD_1055-1084	PNVKSVEKIESPSAKSYYSNYEIEKNPIEK	52	1/6	1	
Cfe_Fe/C_CT618_108-137	DSASGNFKIGVKSVKNENGETVLVPCRILK	528	1/3	1	
Cpe_E58_OmpA_090-105	TSPNNAADSSTTAERA	48	24/203	8[d]	
Cpe_E58_OmpA_161-176	ISGSSLEGKYPNANIS	63	27/209	10[d]	
Cpe_E58_OmpA_323-338	LGQATTVDGTNKFADS	198	27/215	11[d]	
Cpe_E58_IncA_281-300	AAPAAPAAPAAPAAPA	110	14/126	9[d]	
Cpe_E58_IncA_311-326	PAPENNDNNNDDNAAS	513	8/89	5	
Cpe_E58_CT529_209-248	IIRERRAYQRCLERLNQKEVGQEESGSAQEVQAMRSSYVK	154	2/15	1	
Cpe_E58_CT442_151-190	DGSNQIFVDSNRDIRRPGSGGSGGVSASGALEQVANIVMN	94	3/19	3	
Cpn_CWL029_PmpD_0147-0187	EKISSDTKENRKDLETEDPSKKSGLKEVSSDLPKSPETAV	80	1/16	1	
Cpn_CWL029_PmpD_1131-1170	NKEETLVSAGVQINMSSPTPNKDKAVDTPVLADIISITVD	116	1/16	1	
Cpn_CWL029_IncA_331-370	QKAESEFIACVRDRTFGRRETPPPTTPVVEGDESQEEDEG	340	2/13	2	
Cpn_CWL029_CT618_201-216	PETISDPENRNKPSAE	215	1/13	1	
Cmu_Nigg_PmpD_724-739	KVETADINSDKQEAEE	20	1/6	1	

Cmu_Nigg_PmpD_1038-1053	EIGDLEDSVNSEKTPS	50	1/6	1	
Cmu_Nigg_CT618_190-205	TKDKCCDSTQGNSFME	41	1/5	1	
Csu_99DC3_OmpA_166-181	FGLTTTSVAAQDLPNV	6	25/55	22[d]	
Csu_99DC3_OmpA_317-332	TISGKGQDAQTLQDTM	20	20/40	12[d]	
Csu_99DC3_IncA_259-293	LANIKEALIKPSRPPLPKNGFPRTMPPCPPRQTPP	76	1/1	1	
Csu_99DC3_CT529_207-236	EERCNRILCGQADEVLGINNTMCEQFVRQR	13	2/2	2	
Ctr_D/UW-3_OmpA_082-105	FQMGAKPTTDTGNSAAPSTLTARE	166	28/601	11[d]	
Ctr_D/UW-3_OmpA_159-174	FGDNENQKTVKAESVP	20	22/741	10[d]	
Ctr_D/UW-3_OmpA_313-328	IFDTTTLNPTIAGAGD	535	11/738	3	VI
Ctr_D/UW-3_OmpA_324-339	AGAGDVKTGAEGQLGD	73	18/725	6[d]	
Ctr_D/UW-3_OmpA_306-345	QPKSATAIFDTTTLNPTIAGAGDVKTGAEGQLGDTMQIVS	587	18/725	8	VI
Ctr_D/UW-3_PmpD_536-565	ARAPQALPTQEEFPLFSKKEGRPLSSGYSG	3	3/201	1	
Ctr_D/UW-3_PmpD_1036-1065	SGTPVQQGHAISKPEAEIESSSEPEGAHSL	6	5/202	3	
Ctr_D/UW-3_CT529_200-239	SAERADCEARCARIAREESLLEVPGEENACEKKVAGEKAK	27	4/182	2	
Ctr_D/UW-3_IncG_108-147	RPSDQQESGGRLSEESASPQASPTSSTFGLESALRSIGDS	40	4/183	2	

[a] **Background corrected average signals of 6 repeats with homologous and heterologous sera at high stringency are shown. The signal for all peptides was highly significantly above background ($p < 10^{-5}$; one-tailed Student's t-test). The method has an inter-assay CV of 11.2%, and an intra-assay CV of 8.5%. All peptides were non-reactive above background with non-homologous mouse sera against the 8 remaining chlamydial species.**

[b] **Sequence variants among all GenBank sequence accessions for the respective *Chlamydia* spp. peptide sequences.**

[c] **Number of variant peptides within the respective *Chlamydia* spp. required to provide *Chlamydia* species-specific antibody binding to ≥ 95% of all GenBank sequence accessions at ≤ 2 amino acid mismatch tolerance.**

[d] **Peptides from this ARR show serovar-specific reactivity.**

[e] **Potential for cross-reactivity of the antibody reactive peptide region (ARR) with other *Chlamydia* species/strains, I) Cps strains VS225 & GR9; II) Cps GR9 & possibly other Cps strains; III) Cab strains with Cps strains VS225& GR9; IV) Cab strains with Cps strain VS225; V) ARR absent in Cps strain VS225, no reactivity; VI) Central residues shared among all chlamydial species; homologs from other species show unpredictable patterns of cross-reactivity.**

TABLE 2.3. Peptides with cross-reactivity among *Chlamydia* species.

Peptide	Sequence	ELISA signal with *Chlamydia* mono-specific antisera (rlu/sec ×10^{-3})[a]									Sequence variants / Accessions of reactive species[b]	Peptides / Species Detection[c]
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr		
Cab_S26/3_OmpA_089-104	PTGTAAANYKTPTDRP	23	0	0	0	0	22	0	0	0	Cab (2/49), Cpn (2/36)	1,1
Cab_S26/3_OmpA_223-262	NVVSSPAQFVVHKPRGYKGTAFPLPLTAGTDQATDTKSAT	0	0	0	0	0	39	1	0	1	Cpn (3/37)[d]	1
Cab_S26/3_OmpA_309-324	AVLNLTTWNPTLLGEA	5	74	6	0	6	61	0	0	113	Cab (1/38), Cps (14/132), Cpn (2/39)	1,4, 1, 3
Cab_S26/3_OmpA_302-341	AQPKLAAAVLNLTTWNPTLLGEATALDTSNKFADFLQIAS	17	89	15	0	6	117	0	0	89	Cab (2/38), Cps (20/209), Cca (1/5), Cpn (2/39)	1,7,1,1,2
Cab_S26/3_OmpA_366-389	KWSITGEARLINERAAHMNAQFRF	0	321	0	0	0	0	35	74	0	Cps (4/287), Csu (1/33), Cmu (1/13)	1
Cab_S26/3_PmpD_1074-1113	ESVKQPENKTLADINSIGIDLASFVSSDDETPVPPQIIVP	136	269	40	134	0	0	0	0	0	Cab (1/8), Cps (3/76), Cca (1/3), Cfe (1/6)	1,1,1,1
Cca_GPIC_PmpD_571-610	ITFSYNKGTILPFPKVAASSEGESAPEAPKESSPVDLGVR	0	3	191	0	1	1	0	0	0	Cca (1/3)[d]	1
Cfe_Fe/C_PmpD_569-598	FSYNSGKFLPLPMPSAEVSEENSSQNAPVE	0	0	182	0	0	0	0	0	0	Cca (1/3)[d]	1
Cfe_Fe/C_PmpD_1083-1112	EKTLADISSIGVDLASFVTNDDGSSPLPPQ	116	214	151	260	0	1	1	0	0	Cab (1/8), Cps (3/76), Cca (1/3), Cfe (1/6),	1,1,1,1
Cpn_CWL029_PmpD_654-693	EKSLNACSHGDHYPPKTVEEEVPPSLLEEHPVVSSTDIRG	0	0	0	2	0	115	0	0	0	Cpn (1/16)[d]	1
Cpe_E58_OmpA_313-328	PIFNLTTWNPTLLGQA	4	11	4	0	5	40	0	0	87	Cpn (2/39), Ctr (11/738)	4
Csu_99DC3_OmpA_345-384	RKSCGLAVGTTIVDADKYAVTVETRLIDERAAHVNAQFRF	0	267	0	0	0	0	36	87	0	Cps (1/287), Csu (1/33), Cmu (1/13)	2,1,1
Csu_99DC3_PmpD_535-564	NARAPQAVPTRDPEEVFSLSAESLNGCSGG	0	0	6	0	8	0	0	16	0	Csu (2/2)[d]	2
Ctr_D/UW-3_OmpA_041-056	EGFGGDPCDPCATWCD	0	27	0	16	1	0	6	0	0	Cps (4/284), Cfe (1/13)	1,1
Ctr_D/UW-3_OmpA_104-119	RENPAYGRHMQDAEMF	0	0	0	0	0	119	2	0	197	Cpn (1/36), Ctr (4/741)	1,2
Ctr_D/UW-3_OmpA_080-119	KEFQMGAKPTTDTGNSAAPSTLTARENPAYGRHMQDAEMF	0	0	0	0	0	294	9	0	376	Cpn (2/36), Ctr (24/763)	1,2,16,9
Ctr_D/UW-3_OmpA_152-181	SFNLVGLFGDNENQKTVKAESVPNMSFDQS	0	5	0	0	0	0	0	0	97	Ctr (28/601)	12
Ctr_D/UW-3_OmpA_233-248	EFTINKPKGYVGKEFP	0	0	0	0	0	0	0	41	100	Csu (25/98), Ctr (8/750)	10,3
Ctr_D/UW-3_OmpA_245-260	KEFPLDLTAGTDAATG	0	0	0	0	0	0	0	0	105	Ctr (9/750)[d]	4
Ctr_D/UW-3_OmpA_226-265	NVLCNAAEFTINKPKGYVGKEFPLDLTAGTDAATGTKDAS	0	0	0	0	0	0	0	24	295	Csu (28/98), Ctr (12/750)	12,4

[a] **Average reactivity of 3 repeats of each peptide at high stringency ELISA with background subtracted ($p < 10^{-3}$; one-tailed Student's t-test).**

[b] **Sequence variants among all GenBank peptide sequence accessions for the reactive chlamydial species.**

[c] **Number of variant peptides within the respective *Chlamydia* spp. required to provide *Chlamydia* species-specific antibody binding to ≥ 95% of all GenBank sequence accessions at ≤ 2 amino acid mismatch tolerance.**

[d] **Despite experimental reactivity with only a single species, potential for cross-reactivity between species exists because of extensive shared peptide sequences.**

variants. Compared to OmpA ARRs, PmpD, CT442, CT529, CT618, IncG or IncA ARRs showed low divergence within a species, yet high genus divergence. Thus, non OmpA ARRs require only single or few peptides to safely allow sero-detection of all strains of each *Chlamydia* species (Table 2.2, footnote *c*).

Multi-species peptides shown in Table 2.3 were cross-reactive with heterologous sera, or have extensive shared AA so that cross-reactivity in original hosts is likely, even if they did not react with the heterologous mouse serum pools. For instance, the Ctr_D/UW-3_OmpA_104-119 peptide reacted strongly with homologous *C. trachomatis*-specific sera as well as heterologous *C. pneumoniae*-specific sera (Table 2.3). In addition, Ctr_D/UW-3_OmpA_245-260 reacted only with homologous sera, yet this peptide cannot be used for *C. trachomatis*-specific seroassays because this region extensively shares AA with *C. suis* (≤ 88% sequence identity). Therefore, reliable detection of antibodies, e.g. against *C. trachomatis*, with single or few peptide variants (Table 2.3, footnote *c*) will be a promising application of such peptides, even at the risk of some loss of species-specificity.

Stochastic antibody responses of individual mouse sera. Unlike MHC-restricted T cell immunity, the presence or absence of a B-cell response to a given epitope is stochastic due to immunoglobulin gene recombination that determines complementarity of the epitope binding regions (91). Once an immunoglobulin has evolved that binds a specific B-cell epitope, affinity maturation under antigenic re-exposure drives dominance of this antibody. To account for this stochasticity of individual seroreactivity with single peptide epitopes, we used serum pools from 9-50 mice to identify highly reactive peptide antigens in the screening process. To evaluate

antibody reactivity of individual animals, we selected 6 *C. pecorum* E58 peptides of low to high reactivity with the homologous *C. pecorum* serum pool, and tested them with the pool-constituent individual mouse sera.

Results in Figure 2.4A attest to the stochastic reactivity to these 6 peptides. They also show that the percentage of animals producing antibodies has a 2.23-fold greater influence on the averaged reactivity of all mice tested (pool reactivity) than the antibody level of the individual reactive sera (Fig. 2.4A).

Results in Figure 2.4B show that the first 5 peptides (excluding immunodominant peptide IncA_311-326 that reacted strongly with 100% of the sera) reacted with only 30%, 53%, 60%, 20% and 53% of the individual sera, respectively. In contrast, combining results of 2, 3, 4 or all 5 of the less immunodominant peptides renders 83, 87, 93, and 97% of the individual sera positive ($p \leq 0.05$; one-tailed Fisher Exact test). Therefore, use of multiple, even weakly-reactive, peptides reduces stochastic variation of antibody responses to peptide epitopes and allows detection of peptide-specific antibodies with higher sensitivity.

Validity of chlamydial B-cell epitope identification in the heterologous murine host confirmed in the homologous bovine host. In the previous experiments, B-cell epitopes had been identified by use of sera raised in the heterologous murine host by 3× high-dose intranasal inoculation of viable chlamydiae. To evaluate the antigenic potential of these same epitopes in the natural host, we analyzed seroreactivity to the 6 *C. pecorum* peptide antigens tested in Figure 2.4A with sera of cattle naturally exposed to high levels of endemic *C. pecorum* infection. All peptides identified in the murine assays exhibited immunodominant reactivity with these bovine sera, albeit in different patterns (Figure

2.5A). Cows that had been exposed to multiple episodes of natural infection showed a higher percentage of positive sera against the 6 peptides than calves that had experienced only a single episode (59.8% vs. 34.2%; p = 0.0002; two-tailed Fisher Exact test). Interestingly, peptide epitope IncA281-300 dominated in calves over IncA311-326, while the order was reversed in cows, suggesting that IncA311-326 outcompetes IncA281-300 during affinity maturation. Similar to mouse sera, combinations of 2-3 strongly reactive peptides produce 100% sensitivity, and combination of less immunodominant peptides also increase the robustness of detection of anti-chlamydial antibodies in bovine sera (Figure 2.5B).

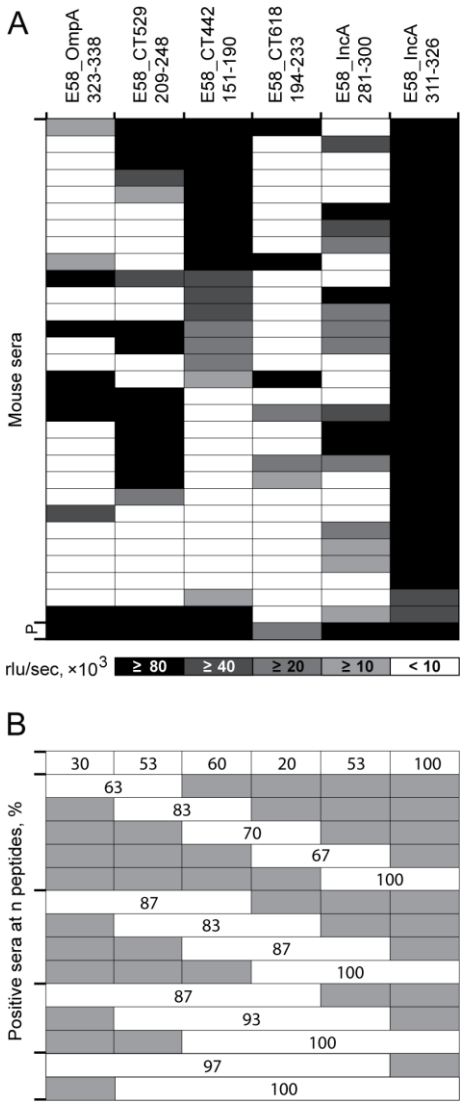


Fig. 2.4. Reactivity of *C. pecorum*-specific individual mouse sera with *C. pecorum* peptides.

(A) **Heat map of reactivity of 30 mouse sera with 6 peptides designed from 5 proteins (OmpA, CT529, CT442, CT618, and IncA) of *C. pecorum* strain E58. Shading intensity is proportional to peptide reactivity with each serum (mean of 2 experiments, interassay CV 8.5%), with a cut-off of 10,000 rlu/sec. P indicates the pool of all mouse sera. The mean of all 15 ratios of the 6 peptide ELISA reactivities of all 30 sera (positive & negative) versus that of only positive sera is 3.57 versus 1.60 (p = 0.046; *t*-test), indicating that the percentage of positive sera has a 2.23-fold higher influence on the level of reactivity of the pool of all 30 sera than the amplitude of reactivity of individual positive sera. In other words, the number of positive sera among all sera of the pool has a 2.23 fold higher influence on the signal amplitude of the serum pool than the signal amplitude of the individual positive sera of the pool.**

(B) **Combined percent seropositivity of these single and multiple *C. pecorum* E58 peptides. Shaded areas indicate peptides not included in the combined reactivity of the evaluated peptides.**

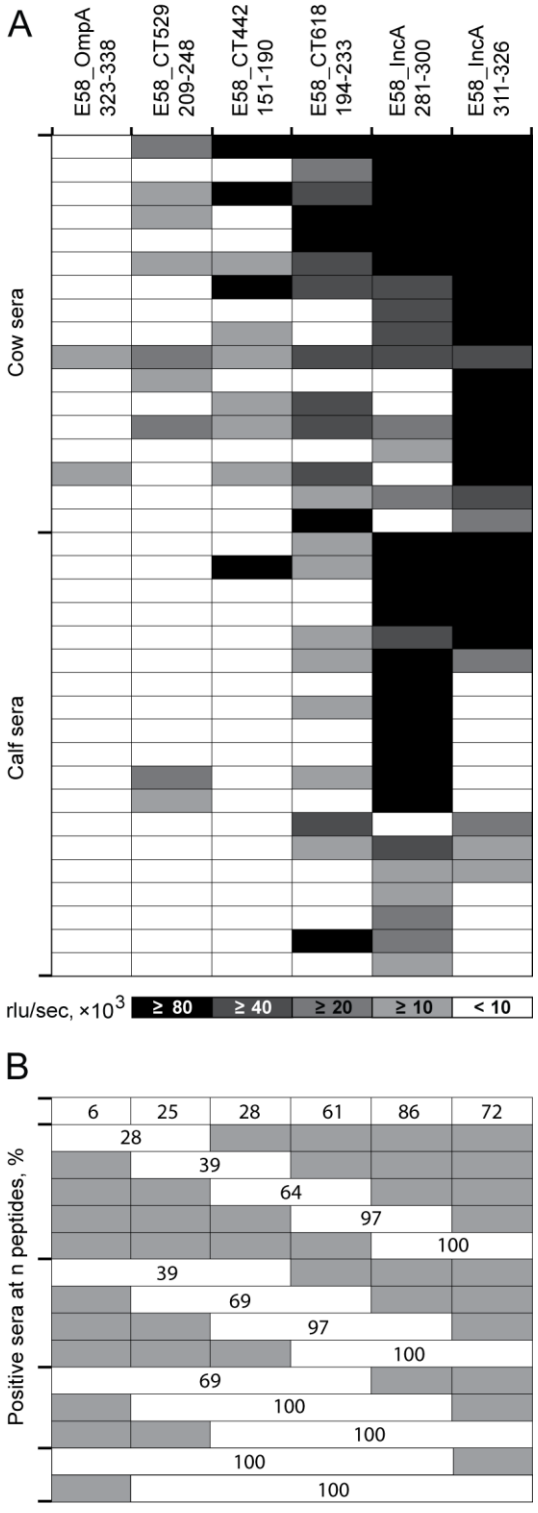


Fig. 2.5. Reactivity of *C. pecorum*-specific individual cow and calf sera with *C. pecorum* peptides.

(A) Heat map of reactivity of 17 cow and 19 calf sera with 6 C. pecorum E58 peptides. **These bovine sera were obtained from animals in 2 herds with endemic *C. pecorum* infections. The cows had**

experienced multiple episodes of natural infection with multiple *C. pecorum* strains (72), while calves were sampled between 11-15 weeks of age, after serum colostrum antibodies had disappeared and calves had experienced a first episode of *C. pecorum* infection (20). Shading intensity is proportional to peptide reactivity with each serum (mean of 2 experiments, interassay CV 7.5%), with a cut-off of 10,000 rlu/sec.

(B) Combined percent seropositivity of these single and multiple C. pecorum E58 peptides. **Shaded areas indicate peptides not included in the combined reactivity of the evaluated peptides.**

2.4. DISCUSSION

For each of the nine *Chlamydia* species, we identified 3-9 immunodominant B-cell epitopes in 23 regions of 10 immunodominant proteins. Peptides corresponding to these epitopes provided highly reactive and specific antigens in robust ELISAs for chlamydial serology, thus achieving the original objective of this study. Immunodominance of the protein was a requirement for identifying strong B-cell epitopes, directing our search to highly polymorphic regions of the immunodominant proteins that were otherwise conserved throughout all chlamydial species. The need for polymorphism within highly conserved proteins, to obtain species-specificity, reduced the number of candidates very substantially to approximately 72 known immunodominant chlamydial proteins, information that had been accessible from classical (74-82), as well as several recent proteome-wide comprehensive studies (58-66). We tested peptides from a total of 64 proteins, mainly at polymorphic regions for species- or strain-specificity, but also at conserved regions for multi-species- or genus-specificity (Table 2.1).

Inadvertently, the search for polymorphic regions directed the investigation to sequences that are enriched for B-cell epitopes (41). As positive results emerged, these data in combination with negative data allowed iterative improvements of B-cell epitope predictive algorithms as well as peptide design for maximum reactivity. Highly reactive peptides strongly accumulated in regions with highest sequence polymorphism, which typically also had high predicted scores for protein disorder tendency, surface exposed tendency, and hydrophilicity (data not shown, manuscript in preparation).

Improvement of peptide reactivity was achieved by use of long, 16-40 AA peptides rather than short 8-11 AA peptides (Fig. 2.1B, panels C of Figures S1-S21), consistent with the results of several recent computational studies on B-cell epitopes (39-42). In our hands, 16 AA-long peptides produced on average a 20-fold higher signal than 8-11 AA-long peptides, and 20-40 AA-long peptides produced another average 3-fold increase over 16 AA peptides (manuscript in preparation).

An important component in maximizing reactivity of the peptide antigens was their accessibility to cognate antibodies. This was maximized by (i) capturing biotinylated peptides on streptavidin that was covalently attached to the solid surface rather than by hydrophobic binding of unmodified peptides or peptide carrier proteins; and by (ii) using the highly flexible hydrophilic SGSG amino acid linker/spacer (89-90) as N-terminal portion of the peptide followed by the specific chlamydial AA sequence. Other commonly used linkers such as the hydrophobic aminohexanoic acid (45) provided inconsistent results, for some peptides with reactivity equal to SGSG, but lower or absent reactivity in others (data not shown). Overall, the methodology used in this investigation resulted in a robust ELISA platform with an inter-assay coefficient of variation of 11.2 % and an intra-assay CV of 8.5 %.

The large data set that was created by testing many species/strain variant peptides with 8 heterologous sera (n = 700) allowed a probabilistic estimation of peptide antigen cross-reactivity (Fig. 2.2). This allows a rational choice of peptide antigens depending on assay objectives (genus, species or serovar-specificity) in chlamydial serology. Comprehensive phylograms of B-cell epitope regions among all *Chlamydia* spp. including the strain-variants provide the critical information on probability of cross-

reactivity (Fig. 2.3, panels C of Figures 2.S1-S8). In addition, the more than 1,600 currently available OmpA sequences of *Chlamydia* spp. allow construction of serovar phylograms, thus pinpointing epitopes for serovar-specific molecular serology of *Chlamydia* spp (Figures 2.S3B, S4B, S8B).

Peptide-based chlamydial serology has been approached before (23-25, 37-38). However, the reactivity of these peptides had not been tested with a comprehensive battery of mono-specific antisera against all *Chlamydia* species. We report here chlamydial proteome-wide and genus-wide discovery of dominant B-cell epitopes by comprehensive testing with improved methodology. Previous studies focused mainly on OmpA peptides, and to avoid cross-reactivity typically used very short peptides and coated those directly to solid surfaces (30-34). Our data show that both short peptides and suboptimal linkers profoundly reduce peptide reactivity in ELISAs. In addition, the high OmpA polymorphism necessitates large numbers of peptide antigens for species-specific detection of anti-chlamydial antibodies. Our investigation has identified dominant B-cell epitopes from other proteins that are highly conserved within chlamydial species, yet well separated among species (PmpD, IncA, CT618, CT529, and CT442), thus facilitating *Chlamydia* species-specific assays by use of few, but highly reactive peptide antigens. In contrast, OmpA as the main serovar determinant of *Chlamydia*, provides optimal epitopes for determination of serovar-specific antibodies, as we demonstrate for one species, *C. pecorum* (Figures 2.S3, 2.S4 and 2.S8).

The strength of the multi-peptide approach for serological determination of the chlamydial exposure status is also that the probability of false positive results is vastly reduced if reactivity not with a single, but with multiple species-specific peptides is used

as stringent requirement for establishing exposure status to a chlamydial species. Importantly, simultaneous use of multiple peptides of each chlamydial species in serological assays also reduces stochastic variation of antibody levels against individual peptides, and thereby increases sensitivity of the assays, as shown in Figures 2.4B and 2.5B.

Antibody responses against many chlamydial proteins such as inclusion membrane proteins (92) are produced only in the context of replication of chlamydiae during infection. Therefore, infectious, rather than inactivated and adjuvanted, chlamydiae were used to reproduce immune responses after natural infection. These inoculations were repeated twice to create high-titered, affinity-matured antibody responses. Thus, these hyperimmune antisera represent the natural antibody response to chlamydial infections as closely as experimentally achievable, with the notable, and possibly consequential, exception that the murine host does not represent the natural host for most chlamydial species. However, in contrast to laboratory mice, for most natural host species it is very difficult to obtain animals that have never been exposed to chlamydial infection. Thus, the murine host offers the certainty of non-exposure which is essential for creation of mono-specific anti-chlamydial sera.

Antibody production against many chlamydial proteins is detectable only in the original host, but not in heterologous hosts such as laboratory rodents (60). This host dependency may be the result of differential protein expression and immune accessibility in the original versus the heterologous host. However, immunodominance of some chlamydial proteins is independent of the host in which the anti-chlamydial antibodies were raised (60). In this investigation, the vast majority of the identified species-specific

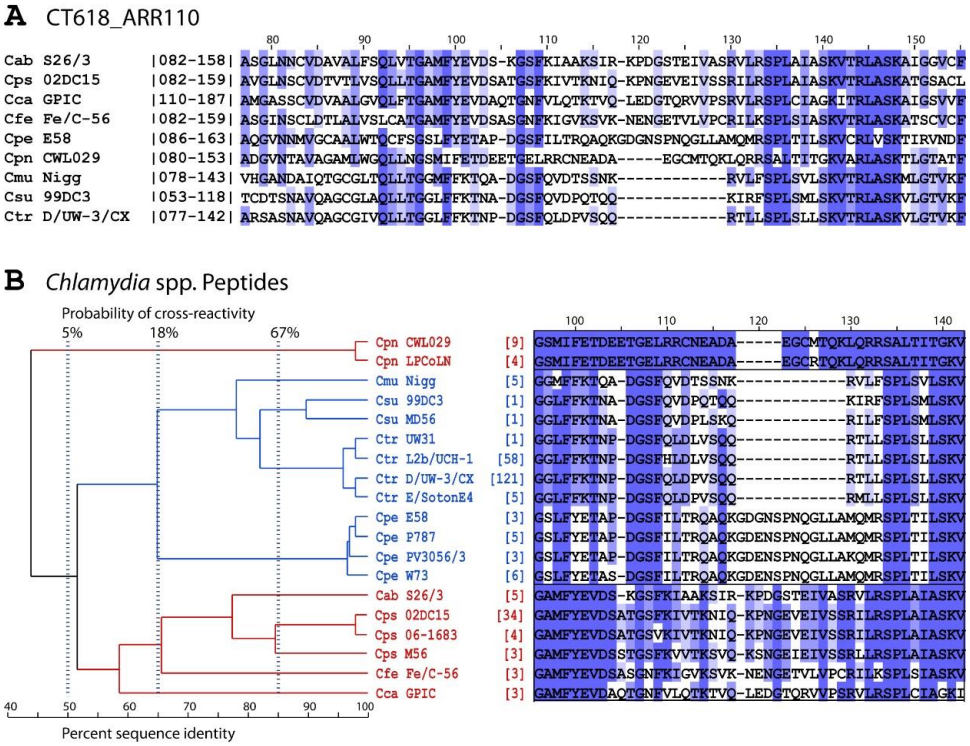
epitopes is located on host-independent proteins such as OmpA, PmpD, IncA, Omp2, CT442, CT529 and CT618 (Table 1, footnote *e*). This indicates that epitopes identified in our murine heterologous model approach (Tables 2 and 3) will also be immunodominant in natural infections of the homologous host. We have confirmed this homologous immunodominance, i.e., the host-independent nature of these reactive peptides, with bovine sera for 6 *C. pecorum* peptides from 5 proteins (Figures 2.4, 2.5), and for peptides shown in Tables 2.2 and 2.3 of *C. abortus*, *C. pneumoniae*, and *C. suis* (data not shown, manuscript in preparation).

On 54 chlamydial proteins that have been reported as immunodominant we could not identify B-cell epitopes (footnote *c* of Table 2.1). While this may be explained by the fact that we tested only few peptides, it may also have to do with poor immunogenicity in the heterologous host. This suggests that additional host-dependent B-cell epitopes may be identified by screening with homologous sera after the chlamydial exposure status has been ascertained by use of the current set of species-specific peptides. In fact, a preliminary screen of non-reactive peptides with bovine sera has demonstrated high and specific reactivity of a number of these peptides (data not shown, manuscript in preparation).

In summary, the peptide antigens found in this investigation produce high and absolutely species-specific signals in a robust ELISA format. In addition to *Chlamydia* species-specific peptides, type-specific peptides for strains of *C. suis*, *C. pecorum*, *C. trachomatis*, and *C. psittaci*, and genus-specific peptides were identified. Because of the simplicity and robustness of these peptide ELISAs, molecular serology of chlamydial infections may now become accessible for non-specialist laboratories. Such serological

assays would also have the added advantage of retrospectively capturing the history of chlamydial infection, rather than stochastically sample a single point in time as PCR detection does. These peptide antigens may also be used in multiplexed assays such as microarrays (89-90) or fluorescent bead assays (91-92). We anticipate that these peptide ELISAs have the potential to vastly improve chlamydial serology, in particular *Chlamydia* species-specific serological diagnosis. By allowing serological dissection of multi-species chlamydial infections, they will further the understanding of chlamydial diseases in retrospective epidemiological investigations of human and animal chlamydial infections. In a wider context, the present methodological approach of epitope identification has a great potential of being useful in both immunological research and laboratory diagnosis of other microbial infections.

SUPPLEMENTARY DATA



C Peptide Reactivities

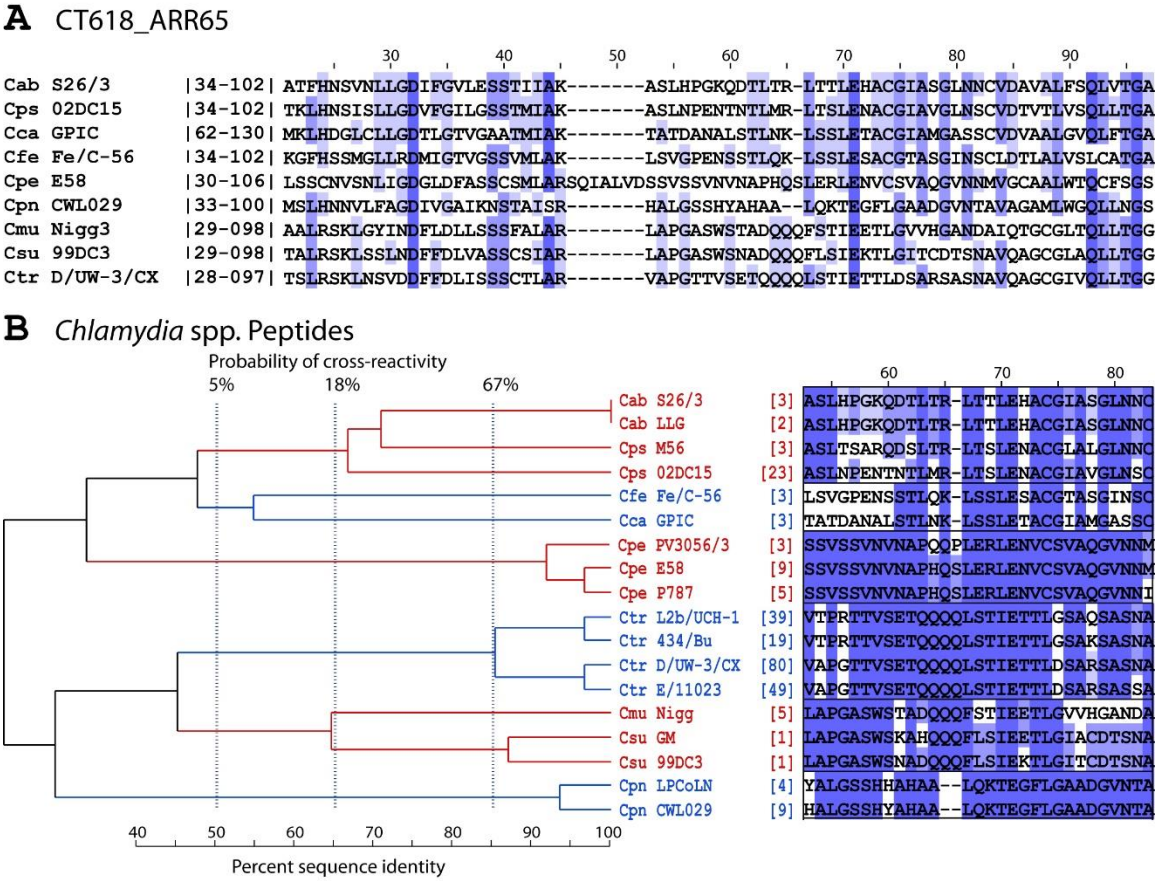
Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_CT618_116-131	AAKSIRKPDGSTEIVA	**0**	0	-	-	-	0	-	-	-
Cab_S26/3_CT618_100-139	TGAMFYEVDSK-GSFKIAAKSIRKPDGSTEIVASRVLRSPL	**0**	2	0	0	0	0	0	0	0
Cps_02DC15_CT618_105-134	YEVDSATGSFKIVTKNIQKPNGEVEIVSSR	0	**539**	0	0	0	0	0	0	0
Cca_GPIC_CT618_145-160	QTKTVQLEDGTQRVVP	-	0	**238**	0	-	-	-	-	-
Cca_GPIC_CT618_134-163	EVDAQTGNFVLQTKTVQLEDGTQRVVPSRV	0	0	**695**	0	0	0	0	0	0
Cfe_Fe/C_CT618_108-137	DSASGNFKIGVKSVKNENGETVLVPCRILK	0	0	0	**750**	0	0	0	0	0
Cpe_E58_CT618_120-135	TRQAQKGDGNSPNQGL	0	0	0	0	**23**	0	0	0	0
Cpe_E58_CT618_106-145	SLFYETAPDGSFILTRQAQKGDGNSPNQGLLAMQMRSPLT	0	0	0	0	**12**	0	0	0	0
Cpn_CWL029_CT618_100-139	SMIFETDEETGELRRCNEADAEGCMTQKLQRRSALTITGK	0	0	0	0	0	**382**	0	0	0
Cmu_Nigg_CT618_111-126	VDTSSNKRVLFSPLSV	-	-	-	-	-	-	**1**	1	0
Csu_99DC3_CT618_099-128	LFFKTNADGSFQVDPQTQQKIRFSPLSMLS	-	-	-	-	-	-	1	**1**	4
Ctr_D/UW-3_CT618_101-116	KTNPDGSFQLDPVSQQ	-	-	-	-	-	-	0	0	**1**
Ctr_D/UW-3_CT618_111-126	DPVSQQRTLLSPLSLL	-	-	-	-	-	-	0	0	**0**

Fig. 2.S1. CT618 antibody-reactive region 110, CT618_ARR110.

(A) Complete ARR110 alignment of the putative inclusion membrane protein CT618 (CT618; locus tag CT_618 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among CT618_ARR110 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of a chlamydial species (Cpn) or a group of closely-related chlamydial species (clade Cab & Cps).

(C) Antibody reactivity of CT618_ARR110 peptides. Peptide sequences of Cpn, Cfe, Cca spp. and clade Cab & Cps are evolutionary separated (B) and these peptides show species/clade-specific reactivity.



Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_CT618_061-090	LHPGKQDTLTRLTTLEHACGIASGLNNCVD	**1**	0	2	0	0	0	0	0	0
Cps_02DC15_CT618_059-098	ASLNPENTNTLMRLTSLENACGIAVGLNSCVDTVTLVSQL	1	**0**	0	0	0	0	0	0	0
Cfe_Fe/C_CT618_057-086	AKLSVGPENSSTLQKLSSLESACGTASGIN	1	1	0	**2**	-	-	-	-	-
Cpe_E58_CT618_059-074	LVDSSVSSVNVNAPHQ	0	0	0	-	**7**	0	0	0	0
Cpe_E58_CT618_062-091	SSVSSVNVNAPHQSLERLENVCSVAQGVNN	0	0	0	0	**25**	7	0	0	0
Csu_99DC3_CT618_055-084	APGASWSNADQQQFLSIEKTLGITCDTSNA	-	-	-	-	-	-	**1**	0	0
Cmu_Nigg3_CT618_049-063	FALARLAPGASWSTAD	-	-	-	-	-	-	1	**6**	0
Ctr_D/UW-3_CT618_048-063	CTLARVAPGTTVSETQ	-	-	-	-	-	-	0	0	**1**
Ctr_D/UW-3_CT618_058-073	TVSETQQQQLSTIETT	0	0	0	0	0	0	0	0	**29**

Fig. 2.S2. CT618 antibody-reactive region 65, CT618_ARR65.

(A) Complete ARR65 alignment of the putative inclusion membrane protein CT618 (CT618; locus tag CT_618 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among CT618_ARR65 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of a chlamydial species (Cpe & Cpn) or a group of closely-related chlamydial species (clade Cab & Cps and clade Cmu & Csu).

(C) Antibody reactivity of CT618_ARR65 peptides. Cpe and Ctr sera do not react with peptides from heterologous species, but show low reactivity with homologous peptides. Cpe and Ctr peptide sequences are evolutionary separated (B); and therefore, these peptides can be used for species-specific reactivity.

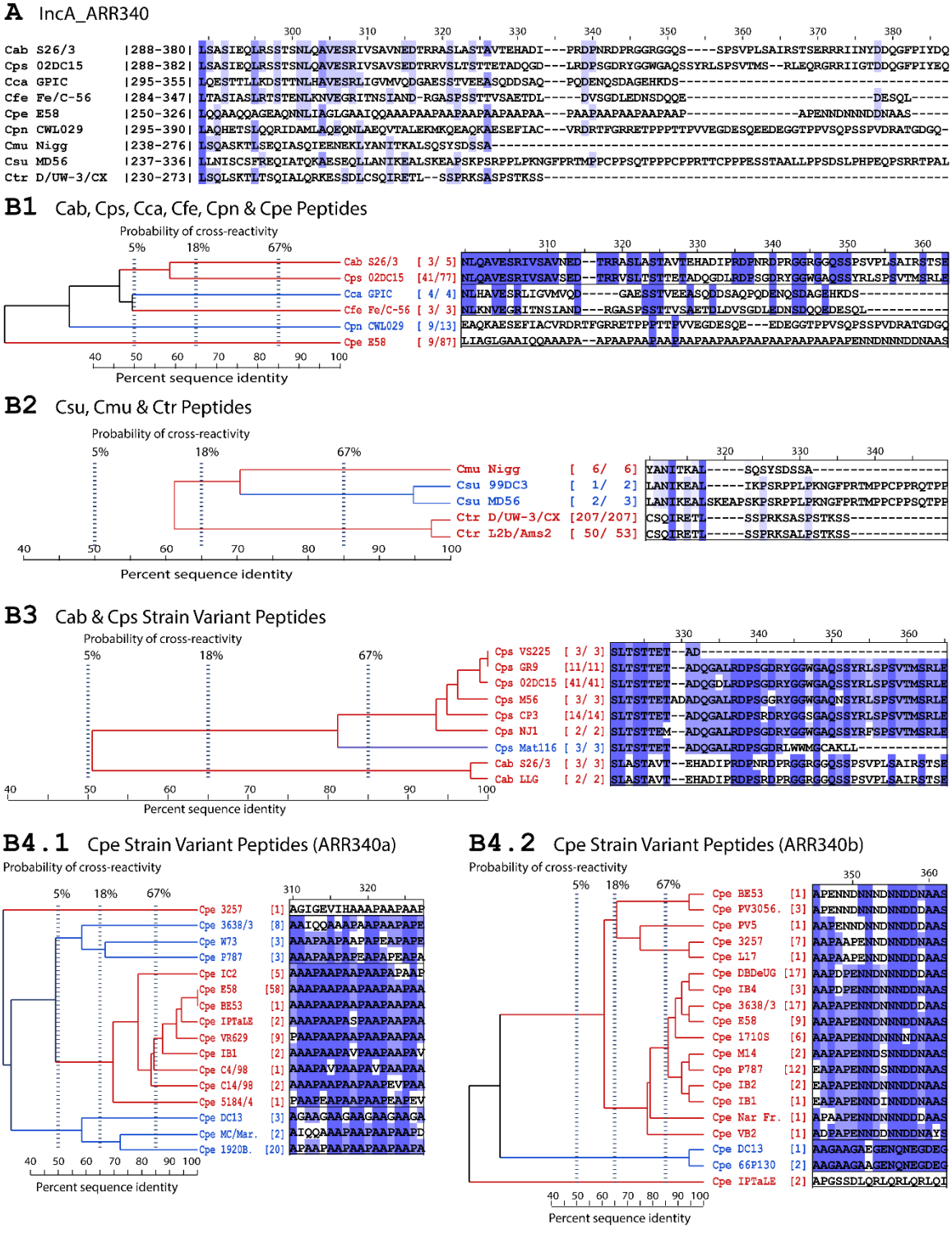
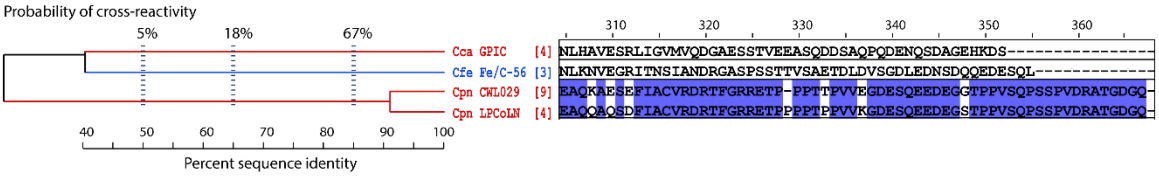


Fig. 2.S3. IncA antibody-reactive region 340, IncA_ARR340.

B5 Cca, Cfe & Cpn Peptides



C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_IncA_325-335	TAVTEHADIPR	**17**	0	0	0	0	0	0	0	0
Cab_S26/3_IncA_337-348	PNRDPRGGRGGQ	**89**	0	0	0	0	0	0	0	0
Cab_S26/3_IncA_351-362	PSVPLSAIRSTS	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_IncA_324-353	STAVTEHADIPRDPNRDPRGGRGGQSSPSV	**266**	0	0	0	0	0	0	0	0
Cps_02DC15_IncA_325-335	TTETADQGDLR	0	**13**	0	0	0	0	0	0	0
Cps_02DC15_IncA_337-348	PSGDRYGGWGAQ	0	**203**	0	0	0	0	0	0	0
Cps_02DC15_IncA_351-366	YRLSPSVTMSRLEQRG	0	**0**	0	0	0	0	0	0	0
Cps_02DC15_IncA_321-360	SLTSTTETADQGDLRDPSGDRYGGWGAQSSYRLSPSVTMS	0	**572**	0	0	0	0	0	0	0
Cps_Mat_IncA_322-351	LTSTTETADQGALRDPSGDRLWWMGCAKLL	0	**97**	0	0	0	0	0	0	0
Cps_NJ1_IncA_334-363	LRDPSGDRYGGWGAQSSYRFSPSVTMSRLE	0	**354**	0	0	0	0	0	0	0
Cps_CP3_IncA_325-354	TTETADQGALRDPSRDRYGGSGAQSSYRLS	0	**8**	0	0	-	-	-	-	-
Cca_GPIC_IncA_323-336	DGAESSTVEEASQD	-	-	**0**	0	-	-	-	-	-
Cca_GPIC_IncA_331-346	EEASQDDSAQPQDENQ	0	0	**225**	0	0	0	0	0	0
Cca_GPIC_IncA_340-355	QPQDENQSDAGEHKDS	0	0	**267**	0	0	0	0	0	0
Cca_GPIC_IncA_316-355	LIGVMVQDGAESSTVEEASQDDSAQPQDENQSDAGEHKDS	0	0	**670**	0	0	0	0	0	0
Cfe_Fe/C_IncA_316-329	SPSSTTVSAETDLD	0	0	8	**0**	0	0	3	1	0
Cfe_Fe/C_IncA_324-339	AETDLDVSGDLEDNSD	0	0	0	**24**	0	0	0	0	0
Cfe_Fe/C_IncA_332-347	GDLEDNSDQQEDESQL	0	5	0	**0**	0	0	0	0	0
Cpe_E58_IncA_281-300	AAPAAPAAPAAPAAPA	0	0	0	0	**350**	0	0	0	0
Cpe_E58_IncA_311-326	PAPENNDNNNDDNAAS	0	0	0	0	**1175**	0	0	0	0
Cpe_DC13_IncA_271-286	VIPAGAAGAAGAAGAA	-	-	-	-	**0**	0	-	-	-
Cpe_IC2_IncA_271-286	EAPAPEAPAPEAPAPE	-	-	-	-	**0**	0	-	-	-
Cpe_C14_IncA_271-286	EVPAAPEVPAAPEVPA	-	-	-	-	**0**	0	-	-	-
Cpe_5184_IncA_271-286	EVPAAPEVPAAPEAPE	-	-	-	-	**0**	0	-	-	-
Cpe_BE53_IncA_311-326	ENNDNNNDNNDDNAAS	-	-	-	-	**424**	0	-	-	-
Cpe_PV5_IncA_311-326	PENNDNNDNNDDDAAS	-	-	-	-	**400**	0	-	-	-
Cpe_DC13_IncA_311-326	GAAGAEGENQNEGDEG	-	-	-	-	**0**	0	-	-	-
Cpn_CWL029_IncA_331-345	QKAESEFIACVRDRT	0	0	0	0	0	**76**	0	0	0
Cpn_CWL029_IncA_336-350	EFIACVRDRTFGRRE	0	0	-	-	-	**281**	-	-	-
Cpn_CWL029_IncA_341-355	VRDRTFGRRETPPPT	-	-	-	-	-	**164**	-	-	-
Cpn_CWL029_IncA_346-360	FGRRETPPPTTPVVE	-	-	-	-	-	**90**	-	-	-
Cpn_CWL029_IncA_356-370	TPVVEGDESQEEDEG	-	-	-	-	-	**102**	-	-	-
Cpn_CWL029_IncA_331-370	QKAESEFIACVRDRTFGRRETPPPTTPVVEGDESQEEDEG	0	0	0	0	0	**698**	0	0	0
Cmu_Nigg_IncA_243-258	KTLSEQIASQIEENEK	-	-	-	-	-	-	**0**	0	0
Cmu_Nigg_IncA_250-265	ASQIEENEKLYANITK	-	-	-	-	-	-	**0**	0	0
Cmu_Nigg_IncA_261-276	ANITKALSQSYSDSSA	-	-	-	-	-	-	**0**	0	0
Csu_99DC3_IncA_259-293	LANIKEALIKPSRPPLPKNGFPRTMPPCPPRQTPP	0	0	0	0	0	0	0	**605**	0
Csu_99DC3_IncA_289-318	RQTPPPCPPRTTCPPPASSTTSLLPPAESL	-	-	-	-	-	-	0	**0**	0
Csu_99DC3_IncA_305-334	ASSTTSLLPPAESLPHQEQPSPQPLSLFSN	-	-	-	-	-	-	0	**0**	0
Csu_99DC3_IncA_320-349	HQEQPSPQPLSLFSNWTLPNLLARNKYTKL	-	-	-	-	-	-	0	**0**	0
Csu_MD56_IncA_271-286	PSKPSRPPLPKNGFPR	-	-	-	-	-	-	0	**75**	0
Csu_MD56_IncA_259-298	LANIKEALSKEAPSKPSRPPLPKNGFPRTMPPCPPSQTPP	0	0	0	0	0	0	0	**165**	0
Ctr_D/UW-3_IncA_243-257	LQRKESSDLCSQIRE	0	-	-	-	0	0	-	-	**0**
Ctr_D/UW-3_IncA_258-273	TLSSPRKSASPSTKSS	0	-	-	-	0	0	-	-	**0**
Ctr_D/UW-3_IncA_234-273	SKTLTSQIALQRKESSDLCSQIRETLSSPRKSASPSTKSS	0	0	0	0	0	0	0	0	**0**

Fig. 2.S3. IncA antibody-reactive region 340, IncA_ARR340. (Continued)

(A) Complete ARR340 alignment of the IncA protein (Inclusion membrane protein A, IncA; locus tag CT_119 of C. trachomatis strain D/UW-3/CX). Because of the absence of a homologous region in *C. trachomatis*, the *C. abortus* IncA protein is used as numbering reference.

(B1) IncA_ARR340 peptide phylogeny of the Cab-Cpe clade. Numbers in brackets indicate the frequency of the predominant strain-specific sequence shown among the total accessions for this sequence cluster available in the NCBI protein database. The IncA C-terminal region shows high evolutionary divergence, allowing for robust species differentiation of Cab, Cps, Cca, Cfe, Cpn and Cpe.

(B2) IncA_ARR340 peptide phylogeny of the Cmu-Csu-Ctr clade. The IncA C-terminal region of *C. suis* shows high evolutionary divergence from closely related *C. muridarum* and *C. trachomatis* (these two species have a truncated ARR), allowing for robust Csu differentiation.

(B3) Contrast phylogram of strain-variant peptides of C. psittaci and C. abortus. C. psittaci and *C. abortus* are the most closely related species in phylogram B1; however, in the most divergent region of the ARR these two species segregate into deeply separate clades comprising non-cross-reactive peptides. Within *C. abortus*, the peptide sequences are highly conserved and the serovariant peptides have a high probability of producing cross-reactivity. Similarly, within *C. psittaci*, the strain variant peptides are highly conserved (with the exception of Mat116 and VS225) and the serovariant peptides have a high probability of cross-reactivity (the truncated ARR of strain VS225 is most likely non-reactive).

(B4) Phylograms of strain-variant IncA_ARR340 sub-region a & b (ARR340a&b) peptides of C. pecorum. C. pecorum is one of the most distinct species in phylogram B1; this region shows also strain variations in sub-region a (B4.1) and sub-region b (B4.2) peptide sequences.

(B5) Phylograms of strain-variant IncA_ARR340 peptides of Cpn, Cca and Cfe. C. pneumoniae has two sequence variants, derived from human and koala strains, and Cfe & Cca sequence variation in this ARR has not been reported.

(C) Antibody reactivity of IncA_ARR340 peptides. Reactive peptides were found for all chlamydial species except *C. muridarum* and *C. trachomatis* (homologous ARR in these two species is truncated) and these peptides produce very high signals with absolute species-specificity. Cps strain variant peptides cross-react with sera raised by Cps strain 02DC15 because of high sequence conservation in *C. psittaci*, allowing detection of all serovars of *C. psittaci* using only few peptides. Additionally, none of the Cps variant peptides react with the remaining *Chlamydia* species including the closest *C. abortus*-specific sera. Sero-variant peptides of Cpe_IncA_ARR340a do not react with *C. pecorum* strain E58-specific sera necessitating several peptides for detection of all serovars; in contrast, sero-variant peptides of Cpe_IncA_ARR340b react strongly allowing detection of the majority of serovars by use of few peptides.

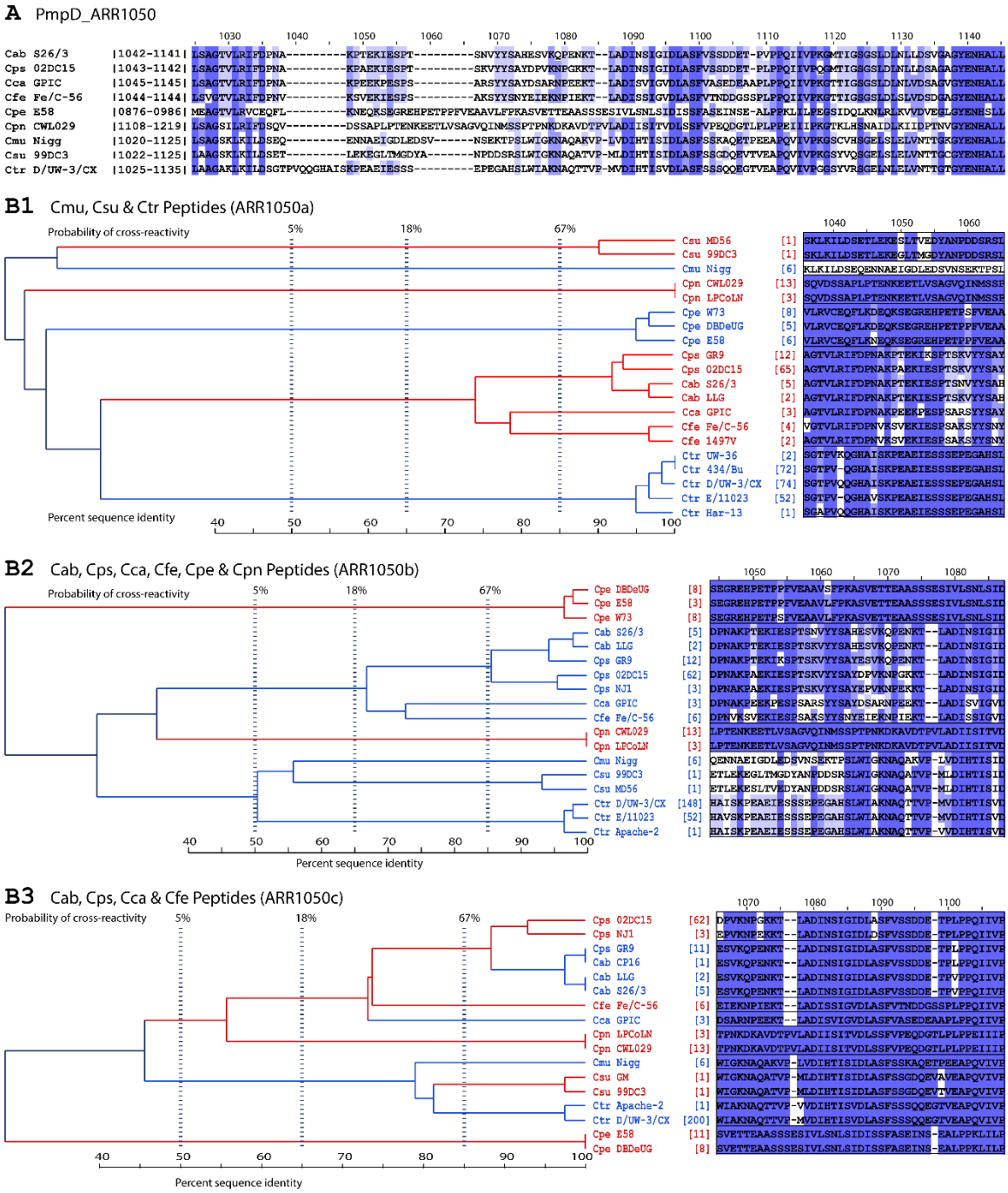


Fig. 2.S4. PmpD antibody-reactive region 1050, PmpD_ARR1050.

(A) Complete ARR1050 alignment of the polymorphic outer membrane protein D (PmpD; locus tag CT_812 of *C. trachomatis* strain D/UW-3/CX).

PmpD_ARR1050 (ctd.)

C Peptide Reactivities

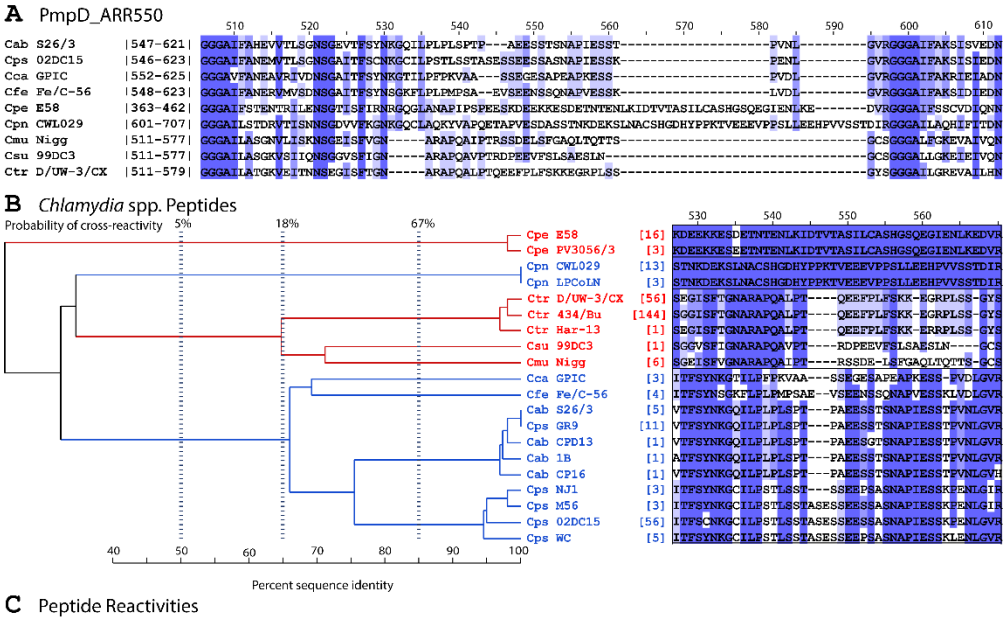
Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_PmpD_1052-1067	DPNAKPTEKIESPTSN	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_1060-1075	KIESPTSNVYYSAHES	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_1068-1083	VYYSAHESVKQPENKT	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_1060-1089	KIESPTSNVYYSAHESVKQPENKTLADINS	**381**	0	0	0	0	0	0	0	0
Cab_S26/3_PmpD_1074-1113	ESVKQPENKTLADINSIGIDLASFVSSDDETPVPPQIIVP	**217**	342	152	256	0	0	0	0	0
Cps_02DC15_PmpD_1053-1092	DPNAKPAEKIESPTSKVYYSAYDPVKNPGKKTLADINSIG	0	**26**	0	0	0	0	0	0	0
Cca_GPIC_PmpD_1056-1095	PNAKPEEKPESPSARSYYSAYDSARNPEEKTLADISVIGV	0	0	**40**	0	0	0	0	0	0
Cfe_Fe/C_PmpD_1055-1084	PNVKSVEKIESPSAKSYYSNYEIEKNPIEK	0	0	0	**95**	0	0	0	0	0
Cfe_Fe/C_PmpD_1083-1112	EKTLADISSIGVDLASFVTNDDGSSPLPPQ	165	247	260	**442**	0	0	0	0	0
Cpe_E58_PmpD_0886-0901	EQFLKNEQKSEGREHP	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_0896-0911	EGREHPETPPFVEAAV	0	0	0	-	**3**	0	0	0	0
Cpe_E58_PmpD_0905-0920	PFVEAAVLFPKASVET	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_0913-0928	FPKASVETTEAASSSE	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_0921-0936	TEAASSSESIVLSNLS	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_0889-0918	LKNEQKSEGREHPETPPFVEAAVLFPKASV	-	-	-	-	**4**	0	-	-	-
Cpe_E58_PmpD_0916-0955	ASVETTEAASSSESIVLSNLSIDISSFASEINSEALPPKL	-	-	-	-	**0**	0	-	-	-
Cpn_CWL029_PmpD_1119-1134	SQVDSSAPLPTENKEE	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_1121-1136	VDSSAPLPTENKEETL	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_1126-1141	PLPTENKEETLVSAGV	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_1138-1153	SAGVQINMSSPTPNKD	0	0	0	0	0	**3**	0	0	0
Cpn_CWL029_PmpD_1146-1161	SSPTPNKDKAVDTPVL	0	0	0	0	0	**31**	0	0	0
Cpn_CWL029_PmpD_1131-1170	NKEETLVSAGVQINMSSPTPNKDKAVDTPVLADIISITVD	0	0	0	0	0	**404**	0	0	0
Cmu_Nigg_PmpD_1029-1044	LDSEQENNAEIGDLED	-	-	-	-	-	-	**0**	0	0
Cmu_Nigg_PmpD_1038-1153	EIGDLEDSVNSEKTPS	0	0	0	0	0	0	**264**	0	0
Cmu_Nigg_PmpD_1044-1059	DSVNSEKTPSLWIGKN	-	-	-	-	-	-	**2**	4	3
Csu_99DC3_PmpD_1029-1058	KILDSETLEKEGLTMGDYANPDDSRSLWIG	-	-	-	-	-	-	2	**1**	8
Csu_99DC3_PmpD_1077-1106	LASFSSGDQEVTVEAPQVIVPKGSYVQSGE	-	-	-	-	-	-	0	**1**	3
Ctr_D/UW-3_PmpD_1035-1050	DSGTPVQQGHAISKPE	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_PmpD_1045-1060	AISKPEAEIESSSEPE	-	-	-	-	-	-	0	0	**10**
Ctr_D/UW-3_PmpD_1094-1109	QQEGTVEAPQVIVPGG	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_PmpD_1036-1065	SGTPVQQGHAISKPEAEIESSSEPEGAHSL	0	0	0	0	0	0	0	0	**41**
Ctr_D/UW-3_PmpD_1044-1083	HAISKPEAEIESSSEPEGAHSLWIAKNAQTTVPMVDIHTI	0	0	0	0	0	0	0	0	**0**
Ctr_D/UW-3_PmpD_1069-1108	KNAQTTVPMVDIHTISVDLASFSSSQQEGTVEAPQVIVPG	0	0	0	5	0	0	0	0	**0**

Fig. 2.S4. PmpD antibody-reactive region 1050, PmpD_ARR1050. (Continued)

(B1) PmpD_ARR1050 upstream sub-region a (ARR1050a) peptide phylogeny of Chlamydia spp. This ARR allows for robust differentiation of Cmu/Csu/Ctr from the remaining chlamydial species. (B2) PmpD_ARR1050 center sub-region b (ARR1050b) peptide phylogeny of *Chlamydia* spp. This ARR shows high evolutionary divergence, allowing separation of Cpe, Cpn, Cmu, Csu & Ctr or of a group of closely-related chlamydial species (clade Cab & Cps).

(B3) PmpD_ARR1050 downstream sub-region c (ARR1050c) peptide phylogeny of Chlamydia spp. This ARR shows low evolutionary divergence in clade Cab, Cps, Cca & Cfe, but is well-separated from the remaining *Chlamydia* spp.

(C) Antibody reactivity of PmpD_ARR1050 peptides. PmpD_ARR1050c peptides (Cab_S26/3_PmpD_1074-1113 and its partial homologue Cfe_Fe/C_PmpD_1083-1112) cross-react with Cab, Cps, Cca and Cfe-specific sera because of extensive shared amino acid residues. However, high conservation within the clade, but deep separation from the other *Chlamydia* spp., allows for detection of all serovars of clade Cab, Cps, Cca & Cfe using a single peptide. All other reactive peptides of PmpD_ARR1050a&b react only with homologous sera, and therefore are suitable for species differentiation.



C Peptide Reactivities

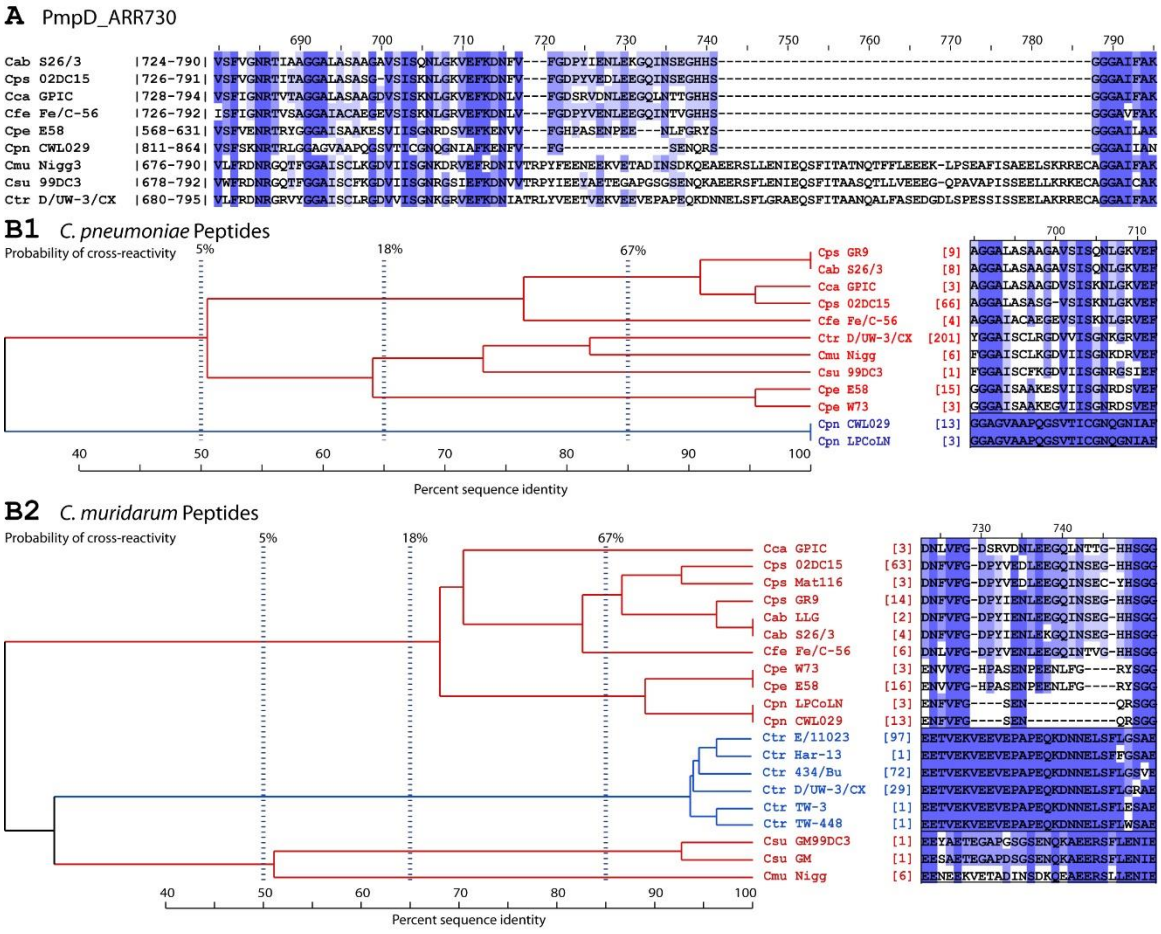
Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_PmpD_565-580	EVTFSYNKGQILPLPL	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_571-586	NKGQILPLPLSPTPAE	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_577-592	PLPLSPTPAEESSTSN	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_583-598	TPAEESSTSNAPIESS	**2**	0	0	0	0	0	0	0	0
Cab_S26/3_PmpD_589-604	STSNAPIESSTPVNLG	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_576-605	LPLPLSPTPAEESSTSNAPIESSTPVNLGV	**15**	0	7	0	0	0	0	0	0
Cps_02DC15_PmpD_564-603	AITFSCNKGCILPSTLSSTASESSEESSASNAPIESSKPE	0	**0**	95	0	0	0	0	0	0
Cca_GPIC_PmpD_571-610	ITFSYNKGTILPFPKVAASSEGESAPEAPKESSPVDLGVR	2	0	**335**	1	2	3	0	1	3
Cfe_Fe/C_PmpD_569-598	FSYNSGKFLPLPMPSAEVSEENSSQNAPVE	0	0	349	**0**	0	0	0	0	0
Cpe_E58_PmpD_387-402	NRGQGLANAPIPSPEE	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_397-412	IPSPEESKDEEKKESD	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_407-422	EKKESDETNTENLKID	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_417-432	ENLKIDTVTASILCAS	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_425-440	TASILCASHGSQEGIE	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_433-448	HGSQEGIENLKEDVRG	-	-	-	-	**0**	0	-	-	-
Cpe_E58_PmpD_398-427	PSPEESKDEEKKESDETNTENLKIDTVTAS	-	-	-	-	**0**	0	-	-	-
Cpn_CWL029_PmpD_631-646	AQKYVAPQETAPVESD	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_637-652	PQETAPVESDASSTNK	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_640-655	TAPVESDASSTNKDEK	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_643-658	VESDASSTNKDEKSLN	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_647-662	ASSTNKDEKSLNACSH	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_656-671	SLNACSHGDHYPPKTV	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_662-677	HGDHYPPKTVEEEVPP	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_667-682	PPKTVEEEVPPSLLEE	0	0	0	0	0	**114**	0	0	0
Cpn_CWL029_PmpD_672-687	EEEVPPSLLEEHPVVS	0	0	0	0	0	**170**	0	0	0
Cpn_CWL029_PmpD_677-692	PSLLEEHPVVSSTDIR	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_654-693	EKSLNACSHGDHYPPKTVEEEVPPSLLEEHPVVSSTDIRG	5	0	7	0	2	**245**	2	1	2
Cmu_Nigg_PmpD_537-552	RAPQAIPTRSSDELSF	-	-	-	-	-	-	**0**	0	0
Cmu_Nigg_PmpD_545-560	RSSDELSFGAQLTQTT	-	-	-	-	-	-	**5**	0	0
Csu_99DC3_PmpD_517-546	ASGKVSIIQNSGGVSFIGNARAPQAVPTRD	-	-	-	-	-	-	0	**0**	0
Csu_99DC3_PmpD_535-564	NARAPQAVPTRDPEEVFSLSAESLNGCSGG	0	0	16	0	17	0	0	**42**	1
Ctr_D/UW-3_PmpD_534-549	GNARAPQALPTQEEFP	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_PmpD_545-560	QEEFPLFSKKEGRPLS	0	0	0	0	0	0	0	0	**20**
Ctr_D/UW-3_PmpD_536-565	ARAPQALPTQEEFPLFSKKEGRPLSSGYSG	0	0	0	0	0	0	0	0	**29**

Fig. 2.S5. PmpD antibody-reactive region 550, PmpD_ARR550.

(A) Complete ARR550 alignment of the polymorphic outer membrane protein D (PmpD; locus tag CT_812 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among PmpD_ARR550 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of chlamydial species (Cpe & Cpn), or of a group of closely-related chlamydial species (clade Ctr, Csu & Cmu and clade Cca, Cfe, Cab & Cps).

(C) Antibody reactivity of PmpD_ARR550 peptides. Because of partial sharing of AA residues (B), Cab, Cps, Cca and Cfe peptides cross-react with heterologous sera. All other peptides (Cpn, Csu & Ctr) react only with homologous sera.



C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_PmpD_737-752	ALASAAGAVSISQNLG	**0**	-	-	-	0	0	-	-	0
Cpe_E58_PmpD_607-722	GHPASENPEENLFGRY	0	-	-	-	**0**	0	-	-	0
Cpn_CWL029_PmpD_824-839	GVAAPQGSVTICGNQG	0	0	0	0	0	**37**	0	0	0
Cpn_CWL029_PmpD_845-860	ENFVFGSENQRSGGGA	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_818-857	TRLGGAGVAAPQGSVTICGNQGNIAFKENFVFGSENQRSG	0	0	0	0	0	**0**	0	0	0
Cmu_Nigg_PmpD_724-739	KVETADINSDKQEAEE	0	0	0	0	0	0	**152**	0	0
Cmu_Nigg_PmpD_755-770	NQTFFLEEEKLPSEAF	-	-	-	-	-	-	**0**	0	0
Csu_99DC3_PmpD_718-747	PYIEEYAETEGAPGSGSENQKAEERSFLEN	0	0	0	0	0	0	0	**0**	3
Csu_99DC3_PmpD_749-778	EQSFITAASQTLLVEEEGQPAVAPISSEEL	-	-	-	-	-	-	0	**0**	0
Ctr_D/UW-3_PmpD_717-732	ATRLYVEETVEKVEEV	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_PmpD_727-742	EKVEEVEPAPEQKDNN	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_PmpD_760-775	QALFASEDGDLSPESS	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_PmpD_713-752	KDNIATRLYVEETVEKVEEVEPAPEQKDNNELSFLGRAEQ	0	0	0	0	0	0	0	0	**0**

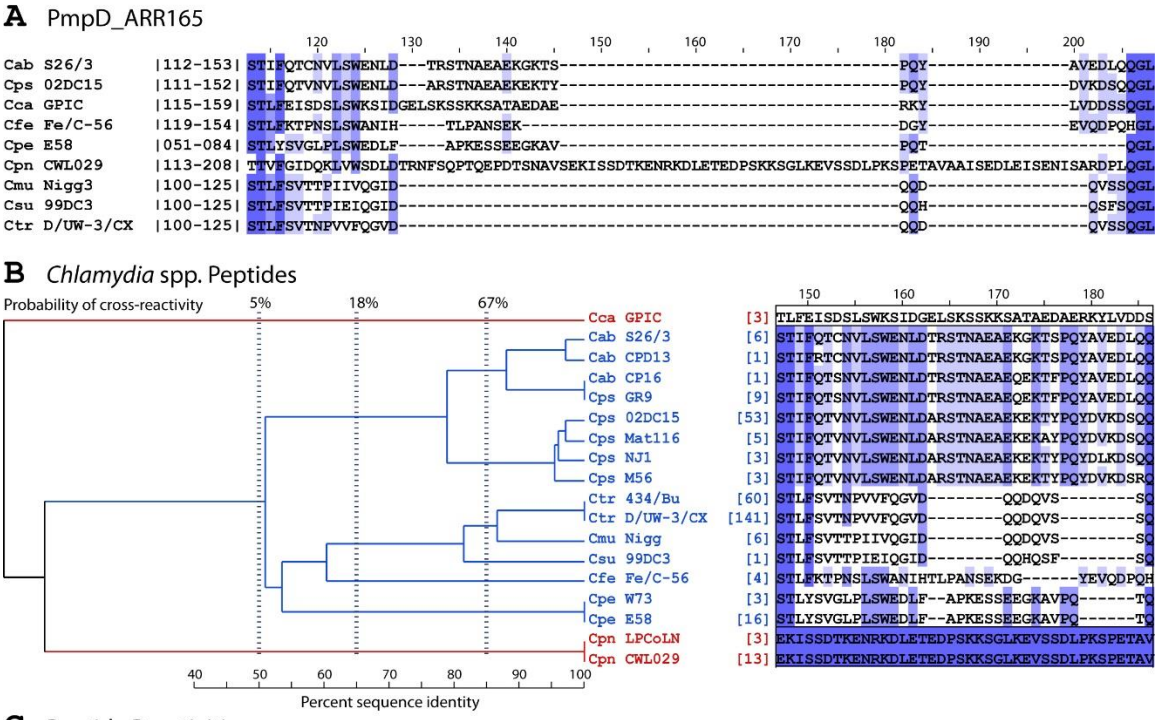
Fig. 2.S6. PmpD antibody-reactive region 730, PmpD_ARR730.

(A) Complete ARR730 alignment of the polymorphic outer membrane protein D (PmpD; locus tag CT_812 of *C. trachomatis* strain D/UW-3/CX).

(B1) Evolutionary relationships among PmpD_ARR730 sequences. In this ARR, Cpn shows high evolutionary divergence, allowing for robust differentiation of *C. pneumoniae* from the remaining C*hlamydia* spp.

(B2) Evolutionary relationships among PmpD_ARR730 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of chlamydial species (Cpn, Cpe, Cmu, Csu, & Ctr) or of a group of closely-related chlamydial species (clade Cca, Cfe, Cab & Cps).

(C) Antibody reactivity of PmpD_ARR730 peptides. Because of evolutionary separation of Cpn (B1) and Cmu (B2), Cpn and Cmu peptides react with homologous sera, and all other peptides (Cab, Cpe, Csu & Ctr) do not cross-react with Cpn or Cmu specific sera.



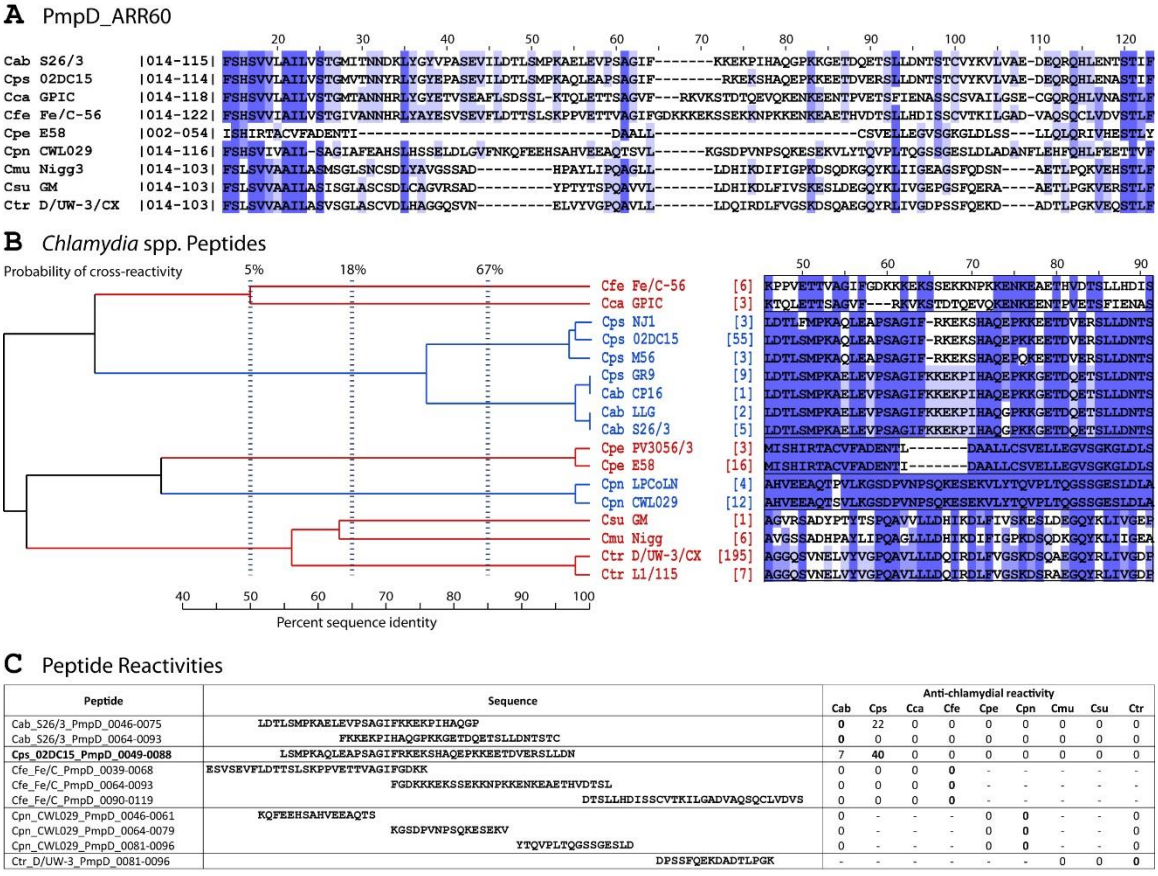
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_PmpD_0121-0136	LSWENLDTRSTNAEAE	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_0129-0144	RSTNAEAEKGKTSPQY	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_0137-0152	KGKTSPQYAVEDLQQG	**0**	-	-	-	0	0	-	-	0
Cab_S26/3_PmpD_0122-0151	SWENLDTRSTNAEAEKGKTSPQYAVEDLQQ	**0**	0	0	0	0	0	0	0	0
Cfe_Fe/C_PmpD_0123-0152	KTPNSLSWANIHTLPANSEKDGYEVQDPQH	3	1	0	**1**	-	-	-	-	-
Cpe_E58_PmpD_0062-0077	WEDLFAPKESSEEGKA	-	-	-	-	**1**	2	-	-	-
Cpe_E58_PmpD_0072-0087	SEEGKAVPQTQGLGFL	-	-	-	-	**0**	0	-	-	-
Cpn_CWL029_PmpD_0125-0140	SDLDTRNFSQPTQEPD	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_0133-0148	SQPTQEPDTSNAVSEK	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_0140-0155	DTSNAVSEKISSDTKE	0	-	-	-	0	**1**	-	-	0
Cpn_CWL029_PmpD_0147-0162	KISSDTKENRKDLETE	0	-	-	-	0	**53**	-	-	0
Cpn_CWL029_PmpD_0155-0170	ENRKDLETEDPSKKSG	0	-	-	-	0	**3**	-	-	0
Cpn_CWL029_PmpD_0161-0176	ETEDPSKKSGLKEVSS	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_0171-0186	LKEVSSDLPKSPETAV	0	-	-	-	0	**50**	-	-	0
Cpn_CWL029_PmpD_0191-0206	EDLEISENISARDPLQ	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_0147-0186	EKISSDTKENRKDLETEDPSKKSGLKEVSSDLPKSPETAV	0	0	0	0	0	**121**	0	0	0
Ctr_D/UW-3_PmpD_0104-0119	SVTNPVVFQGVDQQDQ	-	-	-	-	-	-	0	0	**0**

Fig. 2.S7: PmpD antibody-reactive region 165, PmpD_ARR165.

(A) Complete ARR165 alignment of the polymorphic outer membrane protein D (PmpD; locus tag CT_812 of *C. trachomatis* strain D/UW-3/CX). The designation of ARR165 is derived from *C. pneumoniae* because this species has long inserted sequence and the homologous regions of this ARR are absent in *Chlamydia* spp.

(B) Evolutionary relationships among PmpD_ARR165 sequences. In this ARR, Cpn shows high evolutionary divergence, allowing for robust differentiation of *C. pneumoniae* from the remaining C*hlamydia* spp.

(C) Antibody reactivity of PmpD_ARR165 peptides. Because of evolutionary separation of Cpn (B), Cpn peptides react with homologous sera, and all other peptides (Cab, Cfe, Cpe, & Ctr) do not cross-react with *C. pneumoniae*-specific sera.



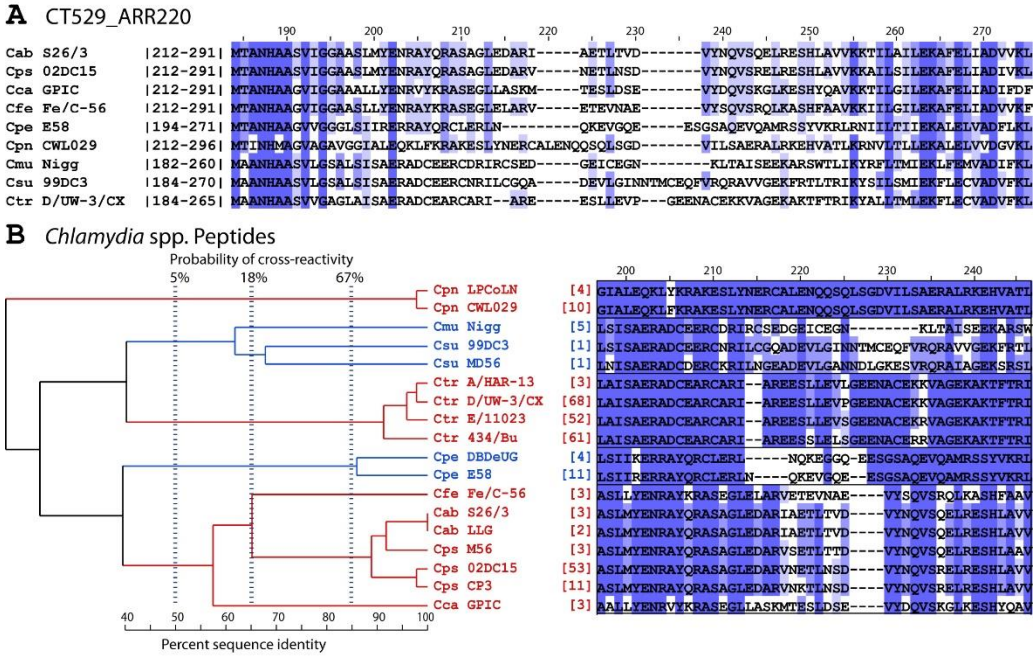
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_PmpD_0046-0075	LDTLSMPKAELEVPSAGIFKKEKPIHAQGP	**0**	22	0	0	0	0	0	0	0
Cab_S26/3_PmpD_0064-0093	FKKEKPIHAQGPKKGETDQETSLLDNTSTC	**0**	0	0	0	0	0	0	0	0
Cps_02DC15_PmpD_0049-0088	LSMPKAQLEAPSAGIFRKEKSHAQEPKKEETDVERSLLDN	7	**40**	0	0	0	0	0	0	0
Cfe_Fe/C_PmpD_0039-0068	ESVSEVFLDTTSLSKPPVETTVAGIFGDKK	0	0	0	**0**	-	-	-	-	-
Cfe_Fe/C_PmpD_0064-0093	FGDKKKEKSSEKKNPKKENKEAETHVDTSL	0	0	0	**0**	-	-	-	-	-
Cfe_Fe/C_PmpD_0090-0119	DTSLLHDISSCVTKILGADVAQSQCLVDVS	0	0	0	**0**	-	-	-	-	-
Cpn_CWL029_PmpD_0046-0061	KQFEEHSAHVEEAQTS	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_0064-0079	KGSDPVNPSQKESEKV	0	-	-	-	0	**0**	-	-	0
Cpn_CWL029_PmpD_0081-0096	YTQVPLTQGSSGESLD	0	-	-	-	0	**0**	-	-	0
Ctr_D/UW-3_PmpD_0081-0096	DPSSFQEKDADTLPGK	-	-	-	-	-	-	0	0	**0**

Fig. 2.S8: PmpD antibody-reactive region 60, PmpD_ARR60.

(A) Complete ARR60 alignment of the polymorphic outer membrane protein D (PmpD; locus tag CT_812 of *C. trachomatis* strain D/UW-3/CX). *C. abortus* PmpD is used for numbering on the top of alignments.

(B) Evolutionary relationships among PmpD_ARR60 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of chlamydial species (Cfe, Cca, Cpe & Cpn), or of a group of closely-related chlamydial species (clade Cab & Cps and clade Ctr, Csu & Cmu).

(C) Antibody reactivity of PmpD_ARR60 peptides. Cps as well as Cab peptides react with *C. psittaci*-specific sera because of sequence conservation; Cfe, Cpn & Ctr peptides do not cross-react with Cps-specific sera because of evolutionary separation of clade Cab & Cps from the remaining species (B).



C Peptide Reactivities

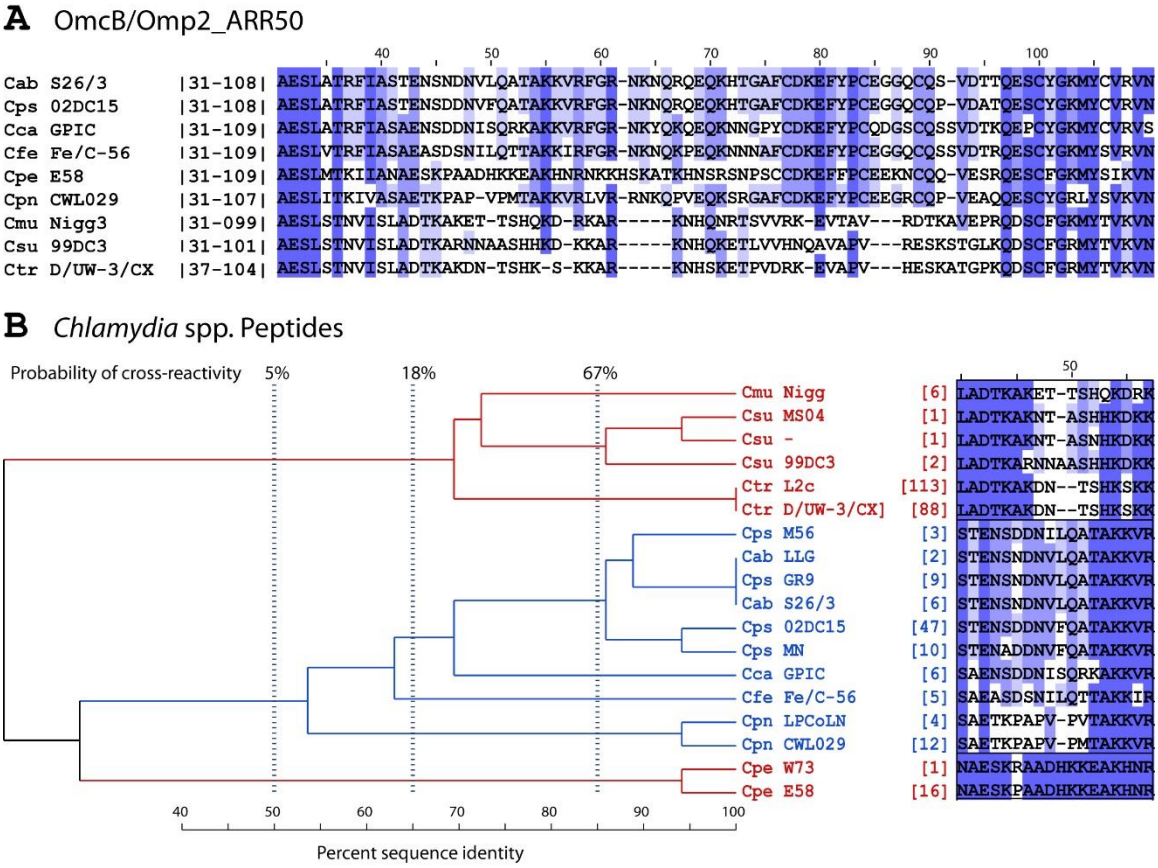
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_CT529_235-250	QRASAGLEDARIAETL	**0**	0	-	-	-	-	-	-	-
Cab_S26/3_CT529_227-266	LMYENRAYQRASAGLEDARIAETLTVDVYNQVSQELRESH	**55**	0	0	0	0	0	0	0	0
Cps_02DC15_CT529_227-266	LMYENRAYQRASAGLEDARVNETLNSDVYNQVSRELRESH	0	**0**	0	0	0	0	0	0	0
Cca_GPIC_CT529_235-250	KRASEGLLASKMTESL	-	0	**0**	0	-	-	-	-	-
Cfe_Fe/C_CT529_231-260	NRAYKRASEGLELARVETEVNAEVYSQVSR	5	0	0	**0**	-	-	-	-	-
Cpe_E58_CT529_217-232	QRCLERLNQKEVGQEE	0	0	0	-	**98**	0	0	0	0
Cpe_E58_CT529_225-240	QKEVGQEESGSAQEVQ	0	0	0	-	**41**	0	0	0	0
Cpe_E58_CT529_235-250	SAQEVQAMRSSYVKRL	-	-	-	-	**0**	0	-	-	-
Cpe_E58_CT529_209-248	IIRERRAYQRCLERLNQKEVGQEESGSAQEVQAMRSSYVK	0	0	0	0	**277**	**0**	0	0	0
Cpn_CWL029_CT529_234-249	FKRAKESLYNERCALE	0	-	-	-	0	**0**	-	-	-
Cpn_CWL029_CT529_244-259	ERCALENQQSQLSGDV	0	-	-	-	0	**0**	-	-	-
Cpn_CWL029_CT529_254-269	QLSGDVILSAERALRK	0	0	0	-	0	**5**	0	0	0
Cpn_CWL029_CT529_236-275	RAKESLYNERCALENQQSQLSGDVILSAERALRKEHVATL	0	0	0	0	0	**59**	0	0	0
Cmu_Nigg_CT529_207-222	RCDRIRCSEDGEICEG	0	0	0	-	0	0	**67**	0	0
Cmu_Nigg_CT529_213-228	CSEDGEICEGNKLTAI	0	0	0	-	0	0	**63**	0	0
Cmu_Nigg_CT529_219-234	ICEGNKLTAISEEKAR	0	0	0	-	0	0	**5**	0	0
Csu_99DC3_CT529_207-236	EERCNRILCGQADEVLGINNTMCEQFVRQR	0	0	0	0	0	0	0	**32**	0
Csu_99DC3_CT529_219-248	DEVLGINNTMCEQFVRQRAVVGEKFRTLTR	-	-	-	-	-	-	0	**0**	0
Ctr_D/UW-3_CT529_208-223	ARCARIAREESLLEVP	-	-	-	-	-	-	0	0	**20**
Ctr_D/UW-3_CT529_215-230	REESLLEVPGEENACE	-	-	-	-	-	-	0	0	**50**
Ctr_D/UW-3_CT529_222-227	VPGEENACEKKVAGEK	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_CT529_200-239	SAERADCEARCARIAREESLLEVPGEENACEKKVAGEKAK	0	0	0	0	0	0	0	0	**132**
Ctr_434/Bu_CT529_215-230	REESLELSGEENACE	-	-	-	-	-	-	0	0	**0**

Fig. 2.S9. CT529 antibody-reactive region 220, CT529_ARR220.

(A) Complete ARR220 alignment of the putative inclusion membrane protein CT529 (CT529; locus tag CT_529 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among CT529_ARR220 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of a chlamydial species (Cpn, Ctr & Cpe) or a group of closely-related chlamydial species (clade Cab & Cps).

(C) Antibody reactivity of CT529_ARR220 peptides. Peptide sequences of Cpe, Cpn, Cmu, Csu and Ctr species are evolutionary well-separated (B) and these peptides show species-specific reactivity. The peptide Ctr_434/Bu_CT529_215-230, which differs from Ctr_D/UW-3/CX_CT529_215-230 by only 3 AA shown in red, does not react with sera raised by Ctr strain D/UW-3/CX.



C Peptide Reactivities

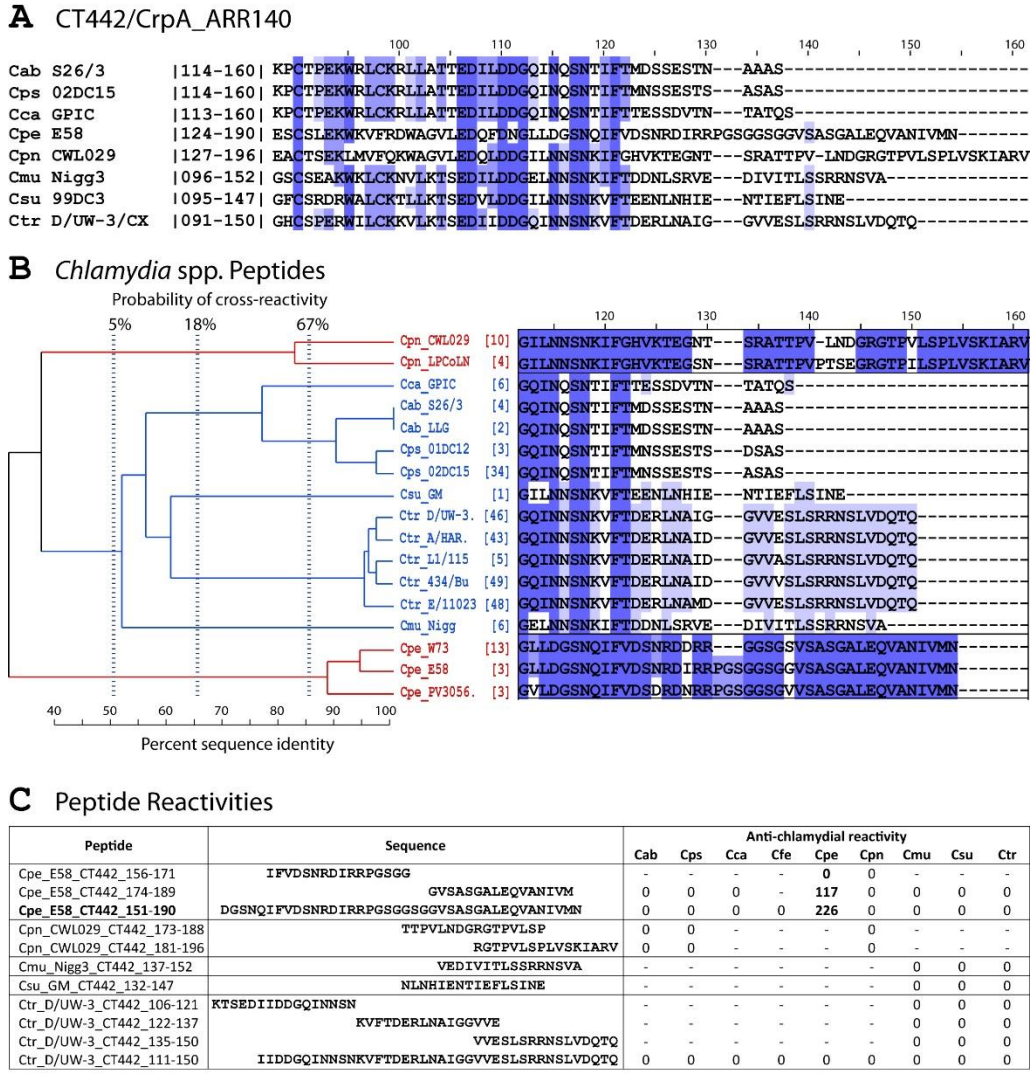
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_OmcB_040-069	ASTENSNDNVLQATAKKVRFGRNKNQRQEQ	**0**	**0**	0	0	0	10	0	0	0
Cpe_E58_OmcB_040-069	ANAESKPAADHKKEAKHNRNKKHSKATKHN	-	-	-	-	**0**	0	-	-	-
Ctr_D/UW-3_OmcB_047-062	LADTKAKDNTSHKSKK	0	0	0	0	0	0	0	0	**36**
Ctr_D/UW-3_OmcB_058-073	HKSKKARKNHSKETPV	0	0	0	-	0	0	0	0	**0**
Ctr_D/UW-3_OmcB_069-084	KETPVDRKEVAPVHES	-	-	-	-	-	-	0	1	**0**
Ctr_D/UW-3_OmcB_080-095	PVHESKATGPKQDSCF	-	-	-	-	-	-	0	0	**2**
Ctr_D/UW-3_OmcB_043-072	NVISLADTKAKDNTSHKSKKARKNHSKETP	0	0	0	0	0	0	0	0	**21**
Csu_99DC3_OmcB_037-066	NVISLADTKARNNAASHHKDKKARKNHQKE	-	-	-	-	-	-	0	0	0
Csu_99DC3_OmcB_060-089	RKNHQKETLVVHNQAVAPVRESKSTGLKQD	-	-	-	-	-	-	0	0	0

Fig. 2.S10: OmcB/Omp2 antibody-reactive region 50, OmcB_ARR50.

(A) Complete ARR50 alignment of the outer membrane cysteine rich protein B/outer membrane protein 2 (OmcB/Omp2; locus tag CT_443 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among OmcB_ARR50 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of *C. pecorum* or a group of closely related species (clade Cmu, Csu & Ctr or clade Cab, Cps, Cca, Cfe & Cpn). This ARR shows very low evolutionary divergence at species level such as the sequences of Ctr strains are fully conserved, allowing for robust detection of all serovars in *C. trachomatis*.

(C) Antibody reactivity of OmcB_ARR50 peptides. Ctr sera do not cross-react with peptides from heterologous species (Cab, Cpe or Csu) because of low sequence conservation. This ARR may be used to detect all serovars of Ctr using a single peptide.



C Peptide Reactivities

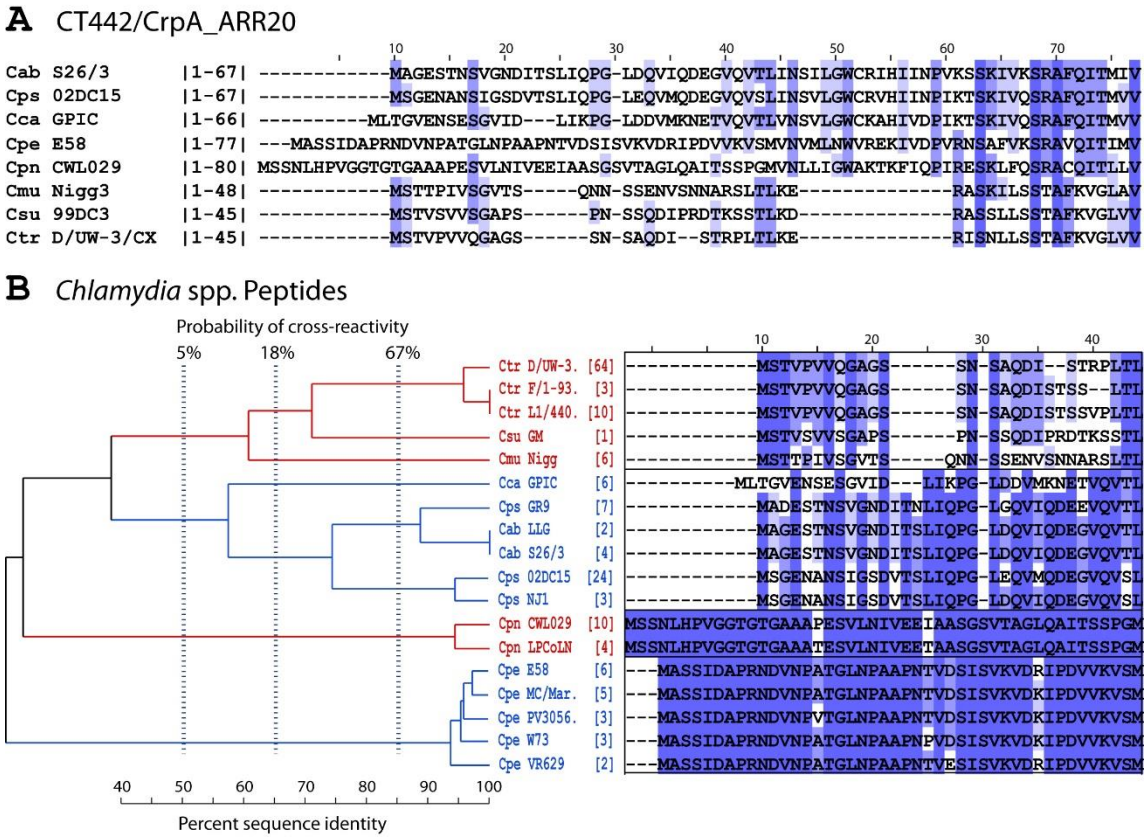
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cpe_E58_CT442_156-171	IFVDSNRDIRRPGSGG	-	-	-	-	**0**	0	-	-	-
Cpe_E58_CT442_174-189	GVSASGALEQVANIVM	0	0	0	-	**117**	0	0	0	0
Cpe_E58_CT442_151-190	DGSNQIFVDSNRDIRRPGSGGSGGVSASGALEQVANIVMN	0	0	0	0	**226**	0	0	0	0
Cpn_CWL029_CT442_173-188	TTPVLNDGRGTPVLSP	0	0	-	-	-	0	-	-	-
Cpn_CWL029_CT442_181-196	RGTPVLSPLVSKIARV	0	0	-	-	-	0	-	-	-
Cmu_Nigg3_CT442_137-152	VEDIVITLSSRRNSVA	-	-	-	-	-	-	0	0	0
Csu_GM_CT442_132-147	NLNHIENTIEFLSINE	-	-	-	-	-	-	0	0	0
Ctr_D/UW-3_CT442_106-121	KTSEDIIDDGQINNSN	-	-	-	-	-	-	0	0	0
Ctr_D/UW-3_CT442_122-137	KVFTDERLNAIGGVVE	-	-	-	-	-	-	0	0	0
Ctr_D/UW-3_CT442_135-150	VVESLSRRNSLVDQTQ	-	-	-	-	-	-	0	0	0
Ctr_D/UW-3_CT442_111-150	IIDDGQINNSNKVFTDERLNAIGGVVESLSRRNSLVDQTQ	0	0	0	0	0	0	0	0	0

Fig. 2.S11. CT442/CrpA antibody-reactive region 140, CT442_ARR140.

(A) Complete ARR140 alignment of CT442/ Chlamydia 15 kDa cysteine-rich outer membrane protein/ cysteine rich outer membrane protein A (CT442/CrpA; locus tag CT_442 of *C. trachomatis* strain D/UW-3/CX). This ARR alignment is from the N-terminus of CT442 proteins of eight *Chlamydia* spp. *C. felis* does not have a homolog of the CT442 protein.

(B) Evolutionary relationships among CT442_ARR140 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of chlamydial species (Cpn, Cmu & Cpe), or of groups of closely related species (clade Csu & Ctr and clade Cab, Cps & Cca). This ARR shows some evolutionary divergence at species level, such as Cpe that has three sequence variants.

(C) Antibody reactivity of CT442_ARR140 peptides. Cpe peptides react with homologous sera. Since Cpn, Cmu, Csu or Ctr peptides do not react with homologous or heterologous sera, Cpe peptides allow species-specific detection of Cpe antibodies.



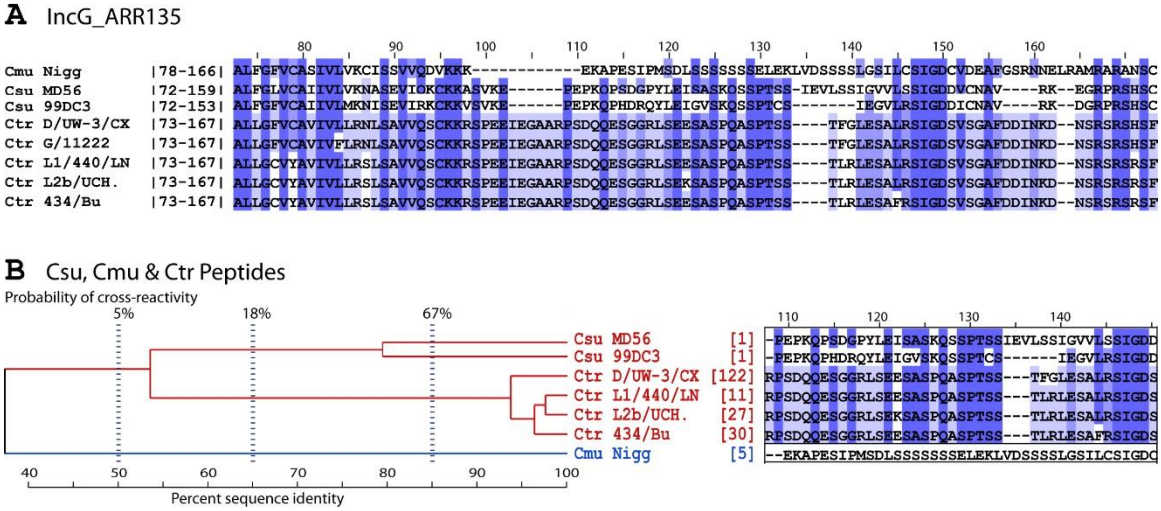
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cpe_E58_CT442_002-041	ASSIDAPRNDVNPATGLNPAAPNTVDSISVKVDRIPDVVK	0	0	3	0	**64**	1	0	0	0
Cpe_E58_CT442_040-055	VKVSMVNVMLNWVREK	-	-	-	-	**0**	0	-	-	-
Cpe_E58_CT442_050-065	NWVREKIVDPVRNSAF	-	-	-	-	**0**	0	-	-	-
Ctr_D/UW-3_CT442_001-040	MSTVPVVQGAGSSNSAQDISTRPLTLKERISNLLSSTAFK	0	0	0	0	0	0	0	0	**0**

Fig. 2.S12. CT442/CrpA antibody-reactive region 20, CT442_ARR20.

(A) Complete ARR20 alignment of the CT442/ cysteine rich outer membrane protein A / *Chlamydia* 15 kDa cysteine-rich outer membrane protein/ cysteine rich outer membrane protein A (CT442/CrpA; locus tag CT_442 of *C. trachomatis* strain D/UW- 3/CX). This ARR alignment is from the C-terminus of CT442 proteins of eight *Chlamydia* spp. *C. felis* does not have a homolog of the CT442 protein.

(B) Evolutionary relationships among CT442_ARR20 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of chlamydial species (Cpn & Cpe), or of groups of closely related species (clade Cmu, Csu & Ctr and clade Cab, Cps & Cca). This ARR shows minimal evolutionary divergence at species level, such as Cpe that has five minor sequence variants that each differ only by 2 out of 41 AA from one another.

(C) Antibody reactivity of CT442_ARR140 peptides. Cpe peptides react with homologous sera, and a single peptide will likely be sufficient for detection of all Cpe serovariants. Peptides from heterologous species (Ctr) do not react with homologous or heterologous sera.



C Peptide Reactivities

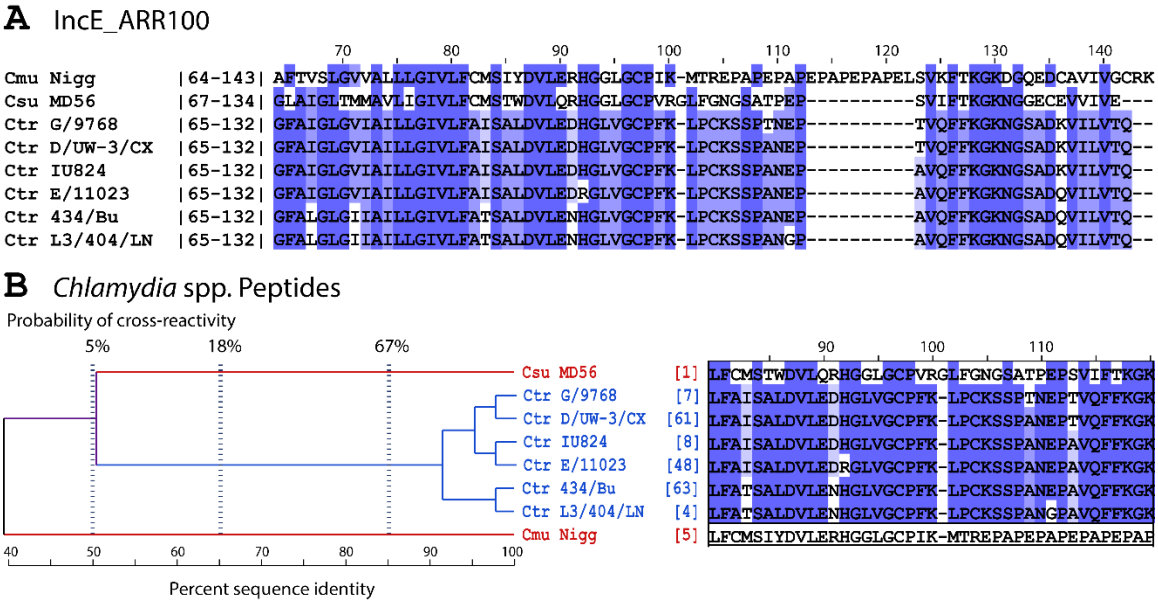
Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cmu_Nigg3_IncG_103-118	KEKAPESIPMSDLSSS	-	-	-	-	-	-	**6**	0	0
Cmu_Nigg3_IncG_118-133	SSSSSELEKLVDSSSS	-	-	-	-	-	-	**1**	0	0
Cmu_Nigg3_IncG_124-129	LEKLVDSSSSLGSILC	-	-	-	-	-	-	**27**	0	0
Csu_99DC3_IncG_085-114	KNISEVIRKCKKVSVKEPEPKQPHDRQYLE	-	-	-	-	-	-	0	**0**	2
Csu_99DC3_IncG_107-136	PHDRQYLEIGVSKQSSPTCSIEGVLRSIGD	-	-	-	-	-	-	1	**0**	1
Csu_99DC3_IncG_124-153	TCSIEGVLRSIGDDICNAVRKDGRPRSHSC	-	-	-	-	-	-	0	**0**	0
Ctr_D/UW-3_IncG_097-112	KRSPEEIEGAARPSDQ	-	-	-	-	-	-	0	0	**2**
Ctr_D/UW-3_IncG_117-132	GRLSEESASPQASPTS	-	-	-	-	-	-	0	0	**3**
Ctr_D/UW-3_IncG_122-137	ESASPQASPTSSTFGL	-	-	-	-	-	-	0	0	**7**
Ctr_D/UW-3_IncG_127-142	QASPTSSTFGLESALR	-	-	-	-	-	-	0	0	**59**
Ctr_D/UW-3_IncG_133-148	STFGLESALRSIGDSV	-	-	-	-	-	-	0	0	**1**
Ctr_D/UW-3_IncG_149-164	SGAFDDINKDNSRSRS	-	-	-	-	-	-	0	0	**1**
Ctr_D/UW-3_IncG_108-147	RPSDQQESGGRLSEESASPQASPTSSTFGLESALRSIGDS	0	0	0	0	0	0	0	0	**170**
Ctr_434/Bu_IncG_127-142	QASPTSSTLRLESAFR	-	-	-	-	-	-	0	0	**0**

Fig. 2.S13. IncG antibody-reactive region 135, IncG_ARR135.

(A) Complete ARR135 alignment of the inclusion membrane protein G (IncG; locus tag CT_118 of *C. trachomatis* strain D/UW- 3/CX). This ARR alignment is from the N-terminus of CT118 protein of three *Chlamydia* spp. (Cmu, Csu & Ctr). All other *Chlamydia* spp. lack homologs of this CT118 protein.

(B) Evolutionary relationships among IncG_ARR135 sequences. This ARR of N-terminus IncG shows high evolutionary divergence, yet Cmu, Csu, and Ctr share some of the central residues. This allows for robust differentiation of the clade Cmu, Csu & Ctr, and within the clade. This ARR shows four minor variant sequences in *C. trachomatis*, but major divergence for the two known Csu sequences.

(C) Antibody reactivity of IncG_ARR135 peptides. Cmu & Ctr peptides react with homologous sera, but not Csu peptides. Ctr and Cmu peptides do not cross-react with heterologous sera, and allow species-specific detection. The peptide Ctr_434/Bu_IncG_127-142, which differs from Ctr_D/UW_3/CX_IncG_127-142 by only 3 AA shown in red, does not react with sera against Ctr strain D/UW-3/CX.



C Peptide Reactivities

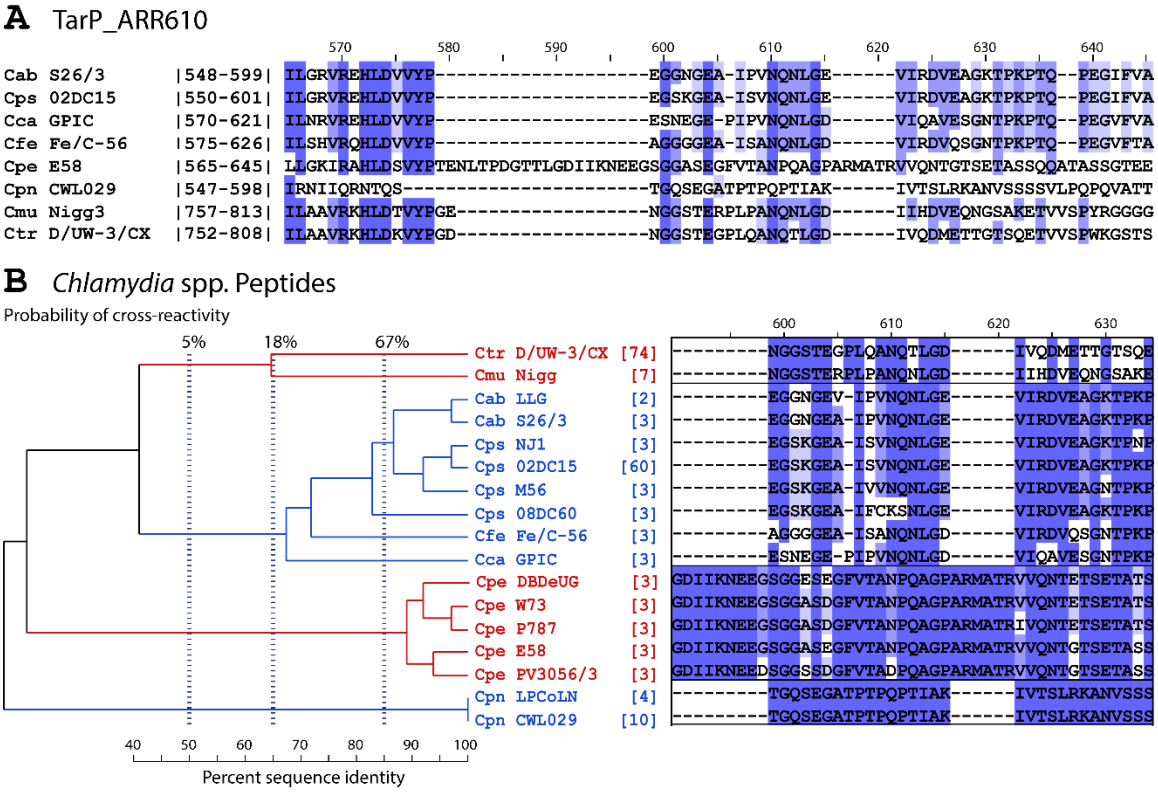
Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Ctr_D/UW-3_IncE_087-102	LDVLEDHGLVGCPFKL	-	-	-	-	-	-	0	0	**2**
Ctr_D/UW-3_IncE_097-112	GCPFKLPCKSSPANEP	-	-	-	-	-	-	0	0	**6**
Ctr_D/UW-3_IncE_107-123	SPANEPTVQFFKGKNG	-	-	-	-	-	-	0	0	**0**
Ctr_D/UW-3_IncE_081-120	LFAISALDVLEDHGLVGCPFKLPCKSSPANEPTVQFFKGK	0	0	0	0	0	0	0	0	**10**

Fig. 2.S14. IncE antibody-reactive region 100, IncE_ARR100.

(A) Complete ARR100 alignment of the inclusion membrane protein E (IncE; locus tag CT_116 of *C. trachomatis* strain D/UW-3/CX). This ARR alignment is from the N-terminus of CT116 protein of three *Chlamydia* spp. (Cmu, Csu & Ctr). All other *Chlamydia* spp. lack homologs of this CT116 protein.

(B) Evolutionary relationships among IncE_ARR100 sequences. This ARR of N-terminus IncG shows high evolutionary divergence between Cmu, Csu, and Ctr. allowing for robust differentiation of the clade Cmu, Csu & Ctr. This ARR also shows evolutionary divergence at species level, where Ctr has six minor sequence variants.

(C) Antibody reactivity of IncE_ARR100 peptides. Ctr peptides react with homologous sera and do not cross-react with heterologous sera, and therefore allow species-specific detection of Ctr.



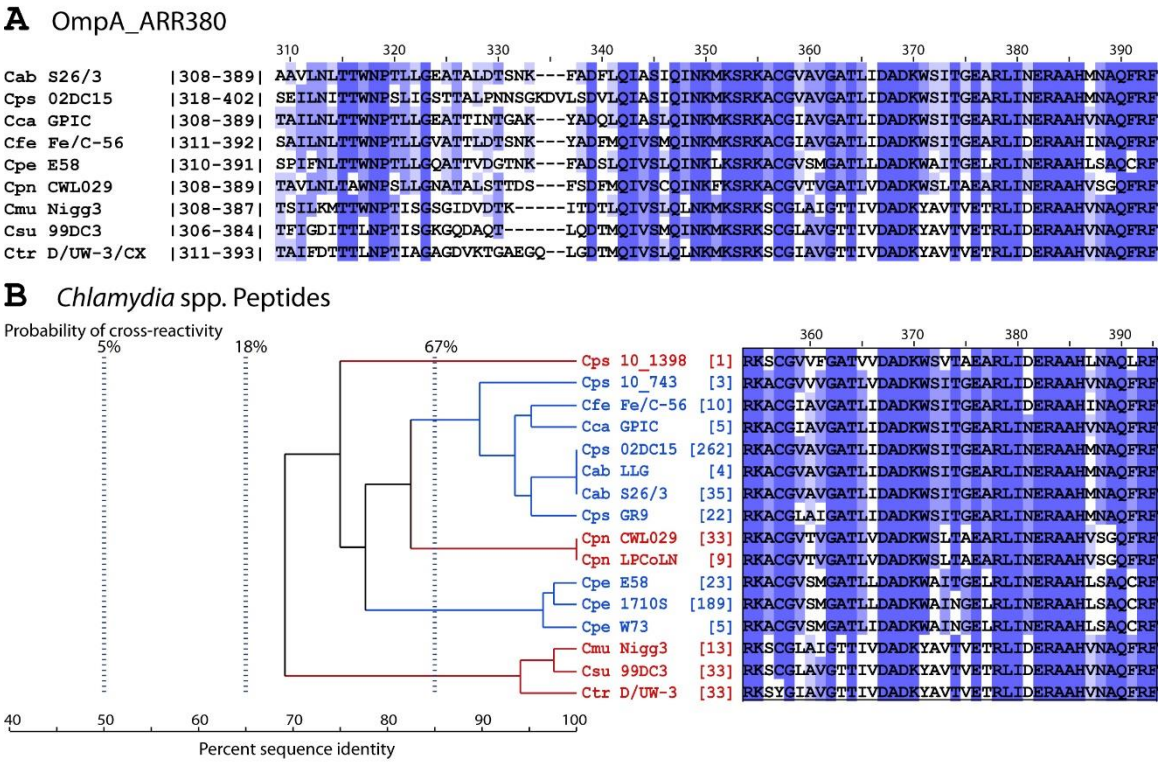
Peptide	Sequence	Anti-chlamydial reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cpe_E58_Tarp_582-597	LTPDGTTLGDIIKNEE	0	5	0	-	**0**	11	0	0	0
Cpe_E58_Tarp_601-616	GASEGFVTANPQAGPA	0	0	0	-	**16**	0	0	0	0
Cpe_E58_Tarp_611-626	PQAGPARMATRVVQNT	-	-	-	-	**0**	0	-	-	-
Cpe_E58_Tarp_593-632	IKNEEGSGGASEGFVTANPQAGPARMATRVVQNTGTSETA	0	0	0	0	**5**	0	0	0	0

Fig. 2.S15. TarP antibody-reactive region 610, TarP_ARR610.

(A) Complete ARR610 alignment of the translocated actin-recruiting phosphoprotein (TarP; locus tag CT_456 of *C. trachomatis* strain D/UW-3/CX). In this ARR, Cpe aligns with the homologous protein regions of the remaining seven *Chlamydia* spp., but has a major insertion The Cpe sequence of this ARR is repeated downstream, and the downstream sequence aligns differently (not shown here). TarP of Csu is not shown because the sequence is very divergent from other *Chlamydia* spp. and sequences of this ARR do not align at all with the available sequence of Csu strain MD56.

(B) Evolutionary relationships among TarP_ARR610 sequences. This ARR shows high evolutionary divergence, allowing for robust differentiation of species, or of groups of closely related species. This ARR also shows evolutionary divergence at the species level such that five minor Cpe variant sequences are known.

(C) Antibody reactivity of TarP_ARR610 peptides. Cpe peptides react only with homologous sera and therefore allow species-specific detection of *C. pecorum*.



C Peptide Reactivities

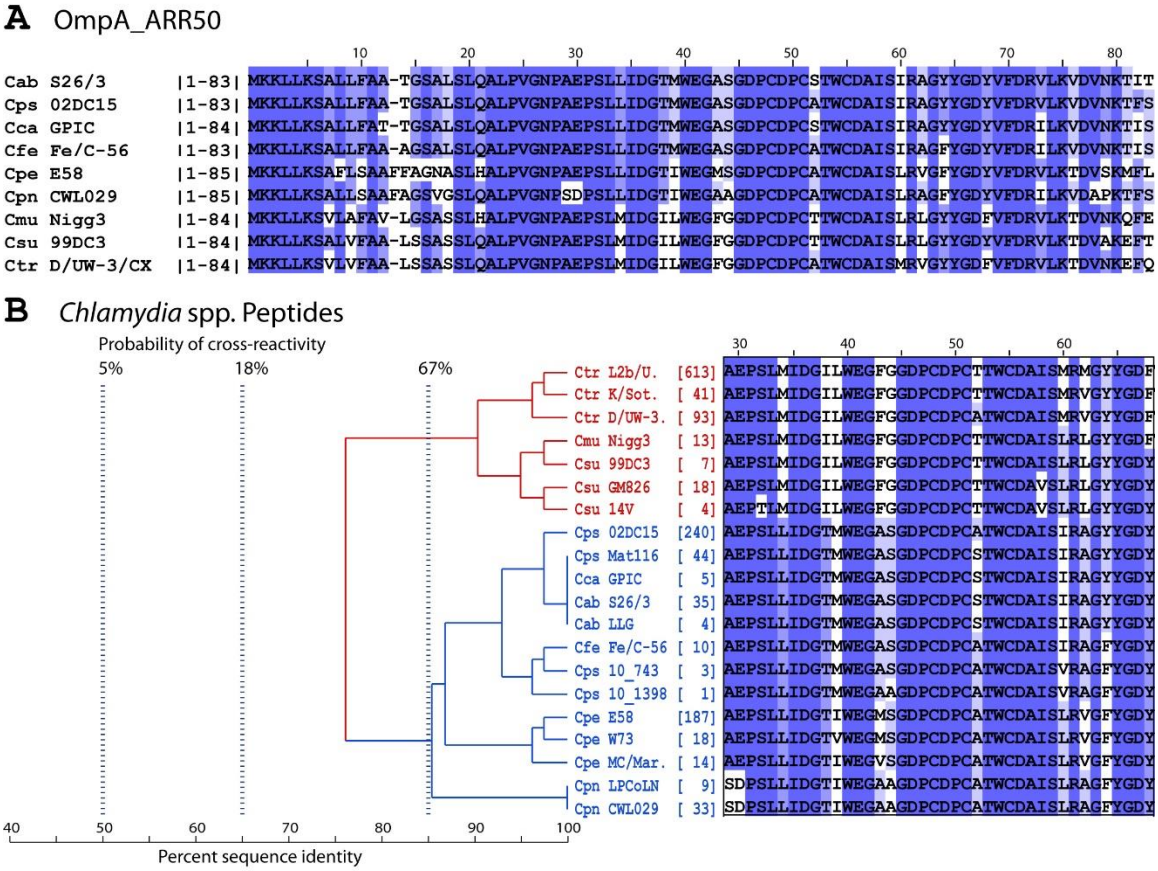
Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_OmpA_367-382	WSITGEARLINERAAH	0	**8**	0	0	0	0	0	0	0
Cab_S26/3_OmpA_366-389	KWSITGEARLINERAAHMNAQFRF	**17**	**370**	**17**	**5**	0	0	35	74	0
Cpe_E58_OmpA_376-391	RLINERAAHLSAQCRF	0	**14**	0	0	0	0	**12**	**14**	0
Csu_99DC3_OmpA_345-384	RKSCGLAVGTTIVDADKYAVTVETRLIDERAAHVNAQFRF	0	**500**	0	**10**	0	0	**99**	**248**	0
Ctr_D/UW-3_OmpA_371-386	YAVTVETRLIDERAAH	**5**	**2**	0	0	0	**2**	0	0	0
Ctr_D/UW-3_OmpA_378-393	RLIDERAAHVNAQFRF	0	**222**	0	**5**	0	0	**77**	**83**	0

Fig. 2.S16. OmpA antibody-reactive region 380, OmpA_ARR380.

(A) Complete ARR380 alignment of the major outer membrane protein A (OmpA; locus tag CT_681 of *C. trachomatis* strain D/UW-3/CX). This ARR is from the N-terminus of *Chlamydia* OmpA protein.

(B) Evolutionary relationships among OmpA_ARR380 sequences. This ARR shows very low evolutionary divergence, allowing for robust detection of all serovars in *Chlamydia* spp. The main clade of Cps strains and the clade of Cpe strains have ≤2AA variations, and the sequence is fully conserved in Csu & Ctr species.

(C) Antibody reactivity of OmpA_ARR380 peptides. Cps, Cmu & Csu sera react with peptides from homologous *Chlamydia* species, and these sera also cross-react with peptides from heterologous species because of sequence conservation. This ARR can be used to detect all serovars of Cps and Csu using a single peptide.



C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cpe_E58_OmpA_042-057	EGMSGDPCDPCATWCD	0	**5**	0	**9**	0	0	**3**	**2**	**1**
Cmu_Nigg3_OmpA_049-064	DPCTTWCDAISLRLGY	0	0	0	0	0	0	0	0	0
Ctr_D/UW-3_OmpA_041-056	EGFGGDPCDPCATWCD	0	**27**	0	**5**	0	0	**20**	**1**	**1**

Fig. 2.S17. OmpA antibody-reactive region 50, OmpA_ARR50.

(A) Complete ARR50 alignment of the major outer membrane protein A (OmpA; locus tag CT_681 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among OmpA_ARR50 sequences. This ARR shows very low evolutionary, allowing for robust detection of all serovars in *Chlamydia* spp., albeit at low signal intensity.

(C) Antibody reactivity of OmpA_ARR50 peptides. Cps, Cfe & Cmu sera cross-react with peptides from heterologous species because of sequence conservation. This ARR can be used to detect all serovars of Cps using a single peptide.

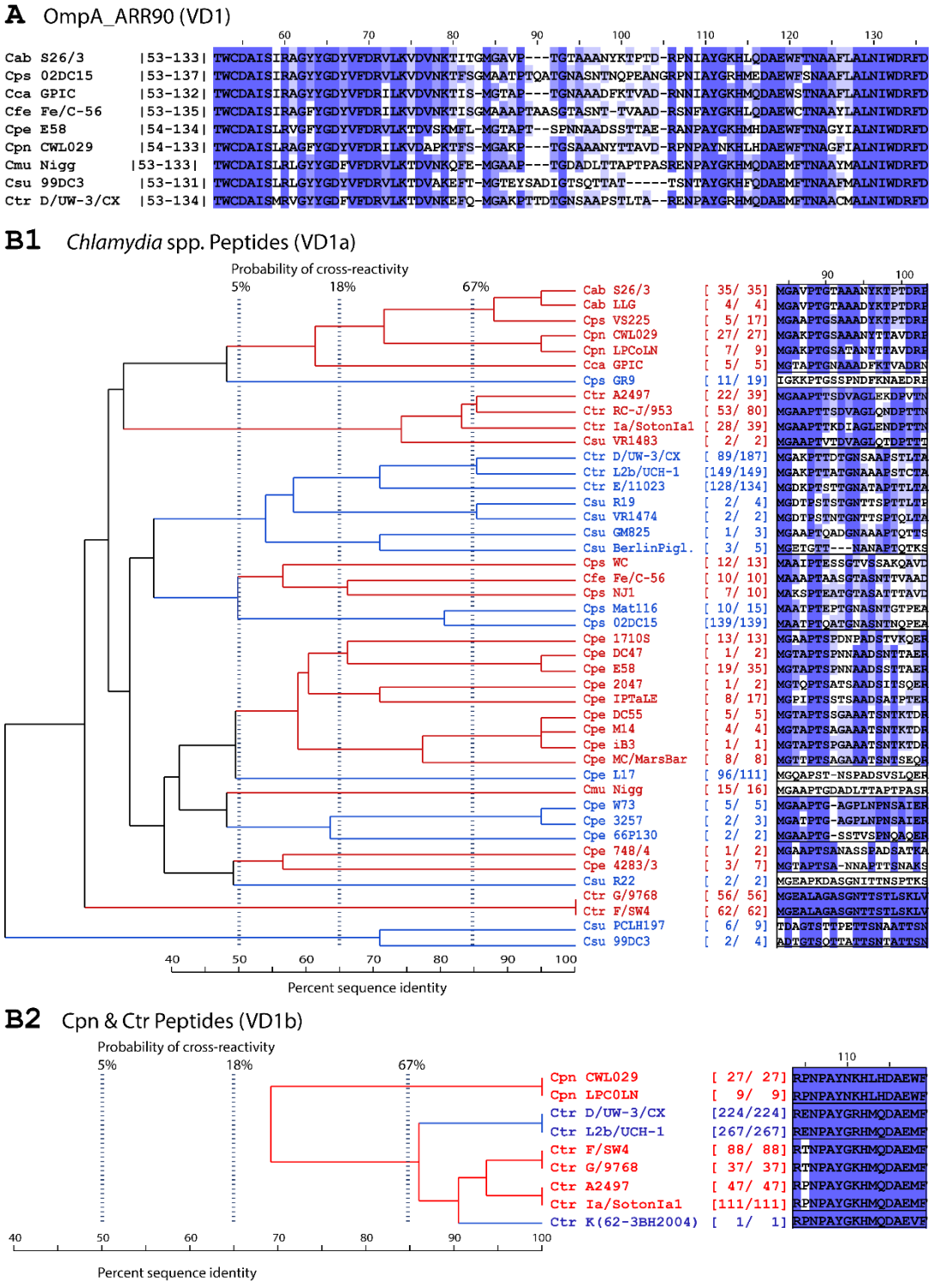


Fig. 2.S18. OmpA antibody-reactive region 90, OmpA_ARR90 (VD1).

(A) Complete ARR90 (VD1) alignment of the major outer membrane protein A (OmpA; locus tag CT_681 of *C. trachomatis* strain D/UW-3/CX).

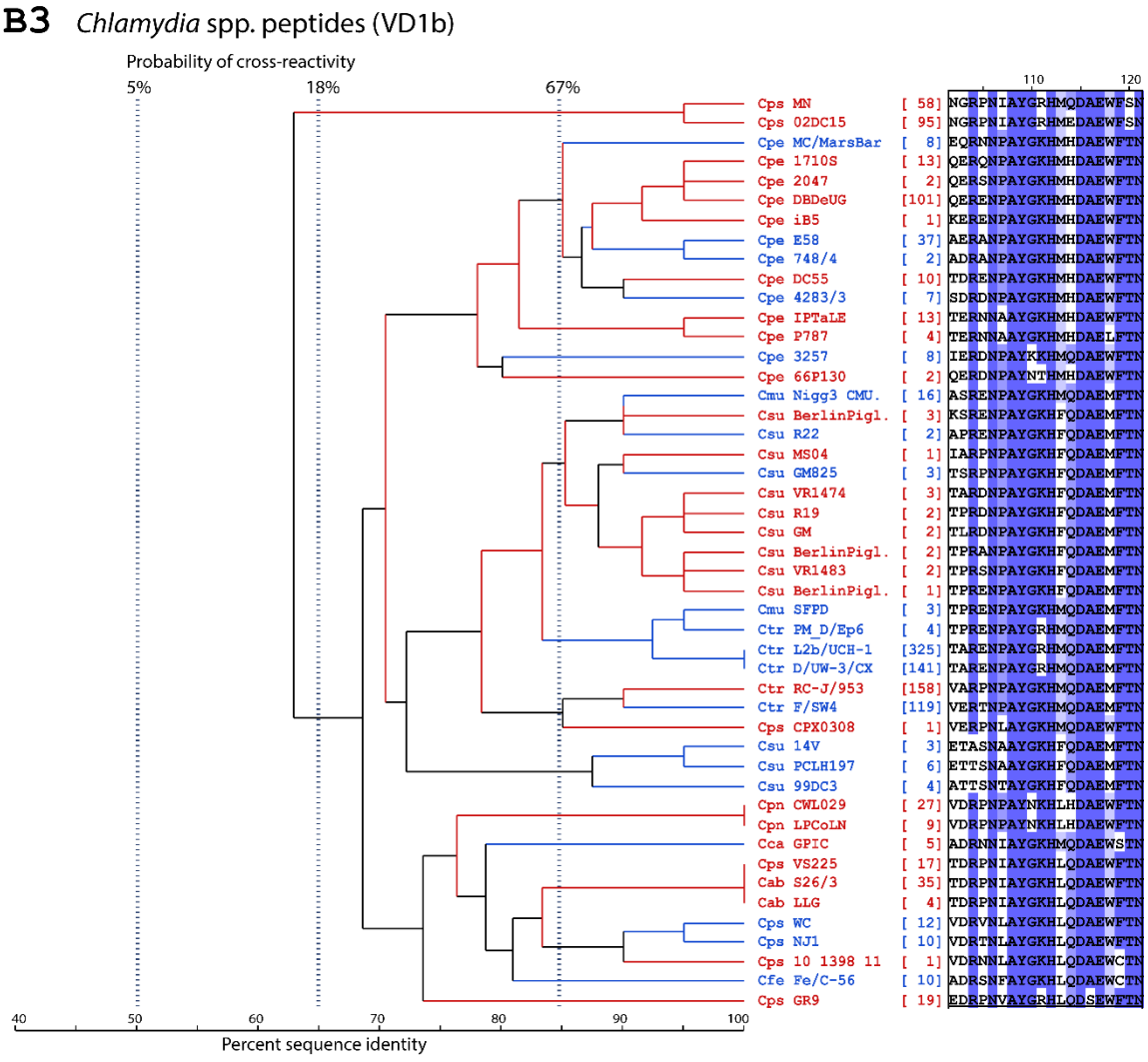


Fig. 2.S18. OmpA antibody-reactive region 90, OmpA_ARR90 (VD1). (Continued)

(B1) Evolutionary relationships among upstream OmpA_VD1 (VD1a) sequences of Chlamydia *spp*. Numbers in brackets indicate the frequency of the predominant strain-specific sequence shown among the total accessions for this sequence cluster available in the NCBI protein database. Deep clades (<50% sequence identity) are indicated in alternating color, with the sequences separated into boxes. VD1a shows high evolutionary divergence, allowing for robust species-, and even serovar differentiation. Deep clades cluster strain-variant peptide sequences of single species with the following exceptions: Cab, Cps strain VS225, Cpn strains occupy the same clade with Cca; Csu strain VR1483 with Ctr strains; and Cps strain NJ1 with Cfe.

(B2) Evolutionary relationships among downstream OmpA_VD1 (VD1b) sequences of C. trachomatis *and* C. pneumoniae. Clades (<92% sequence identity) are indicated in alternating color. VD1b shows high conservation and allows detection of all serovars of *C. pneumoniae* using a single peptide, or of *C. trachomatis* with few peptides because of the lower evolutionary divergence in VD1b than in VD1a.

C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_OmpA_095-102	ANYKTPTD	**0**	0	0	0	0	**0**	0	0	0
Cab_S26/3_OmpA_089-104	PTGTAAANYKTPTDRP	**70**	0	0	-	0	**42**	0	0	0
Cab_S26/3_OmpA_084-107	GMGAVPTGTAAANYKTPTDRPNIA	**46**	0	0	0	0	**112**	0	0	0
Cps_02DC15_OmpA_091-106	QATGNASNTNQPEANG	0	**0**	0	0	0	0	0	0	0
Cps_02DC15_OmpA_086-109	AATPTQATGNASNTNQPEANGRPN	0	**12**	0	0	0	0	0	0	0
Cca_GPIC_OmpA_094-101	ADFKTVAD	0	0	**0**	0	0	0	0	0	0
Cca_GPIC_OmpA_089-101	TGNAAADFKTVAD	0	0	**0**	0	0	0	0	0	0
Cfe_Fe/C_OmpA_090-104	TAASGTASNTTVAAD	0	0	0	**0**	0	0	0	0	0
Cpe_E58_OmpA_096-103	ADSSTTAE	0	0	0	0	**0**	0	0	0	0
Cpe_E58_OmpA_090-105	TSPNNAADSSTTAERA	0	0	0	0	**56**	0	0	0	0
Cpe_3257_OmpA_090-105	PTGAGPLNPNSAIERD	0	0	0	0	**0**	0	0	0	0
Cpe_L17_OmpA_090-105	PSTNSPADSVSLQERE	0	0	0	0	**0**	0	0	0	0
Cpe_1710S_OmpA_090-105	TSPDNPADSTVKQERQ	0	0	0	0	**0**	0	0	0	0
Cpe_iB3_OmpA_090-105	TSPGAAATSNTKTDRE	0	0	0	0	**0**	0	0	0	0
Cpe_2047_OmpA_090-105	TSATSAADSITSQERS	0	0	0	0	**0**	0	0	0	0
Cpe_4283/3_OmpA_090-105	SANNAPTTSNAKSDRD	0	0	0	0	**0**	0	0	0	0
Cpe_748/4_OmpA_090-105	ANASSPADSATKADRA	0	0	0	0	**0**	0	0	0	0
Cpn_CWL029_OmpA_089-104	PTGSAAANYTTAVDRP	0	0	0	-	0	**248**	0	0	0
Cmu_Nigg_OmpA_091-104	DADLTTAPTPASRE	0	0	0	0	0	0	**25**	0	0
Csu_99DC3_OmpA_090-105	ADTGTSQTTATTSNTA	-	-	-	-	-	-	0	**0**	0
Ctr_D/UW-3_OmpA_070-085	FDRVLKTDVNKEFQMG	0	0	0	-	0	0	0	2	**2**
Ctr_D/UW-3_OmpA_071-086	DRVLKTDVNKEFQMGA	0	0	0	0	0	0	0	0	**12**
Ctr_D/UW-3_OmpA_083-098	QMGAKPTTDTGNSAAP	0	0	0	-	0	0	0	0	**425**
Ctr_D/UW-3_OmpA_092-107	TGNSAAPSTLTARENP	0	0	0	-	0	0	0	0	**263**
Ctr_D/UW-3_OmpA_104-119	RENPAYGRHMQDAEMF	0	0	0	0	0	53	8	0	**453**
Ctr_D/UW-3_OmpA_082-105	FQMGAKPTTDTGNSAAPSTLTARE	0	0	0	0	0	0	0	0	**462**
Ctr_D/UW-3_OmpA_080-119	KEFQMGAKPTTDTGNSAAPSTLTARENPAYGRHMQDAEMF	1	4	2	0	4	633	31	5	**712**

Fig. 2.S18. OmpA antibody-reactive region 90, OmpA_ARR90 (VD1). (Continued)

(B3) Evolutionary relationships among downstream OmpA_VD1 (VD1b) sequences of Chlamydia *spp*. Deep clades (<50% sequence identity) are separated into boxes, and more closely related clades (<90% sequence identity) are indicated by alternating color. In addition to Ctr & Cpn spp. (B2), homologous VD1b peptide sequences are also highly conserved in all remaining *Chlamydia* spp.

(C) Antibody reactivity of OmpA_VD1 peptides. The Cab VD1a peptide is evolutionary close to Cpn, and therefore crossreactive with Cpn-specific sera. However, this Cab peptide can be used to differentiate Cab-specific from Cps-speci_c antibodies. Cps strain VS225 occupies the same clade as Cab, but the majority of Cps strains occupy different positions well separated from Cab (B). Cpe strain variant peptides are not reacting with any heterologous sera. Peptide Ctr_D/UW-3/CX_OmpA_104-119 reacts with homologous Ctr-specific sera as well as with heterologous Cpn-specific sera because of extensive sequence conservation.

(D) OmpA_VD1: Sequence variations and major clades in Chlamydia *spp*. Please find Figure 2.11D on the next page. Numbers in brackets indicate the frequency of the predominant strain-specific sequence shown among the total accessions for this sequence cluster available in the NCBI protein database. The sequences of *Chlamydia* spp./serovars are grouped into boxes based on the percent identity among 50 AA long VD1 sequences, and 80% sequence identity between common ancestors is used as the threshold level (shown in green vertical line). Peptide sequences within a box differ by less than 10 AA excluding inserted gaps, while they di_er between boxes by more than 10 AA. VD1 shows high evolutionary divergence with many serovars, particularly in Cpe, Csu, Cps and Ctr.

D OmpA_VD1: Major Clades of *Chlamydia* spp.

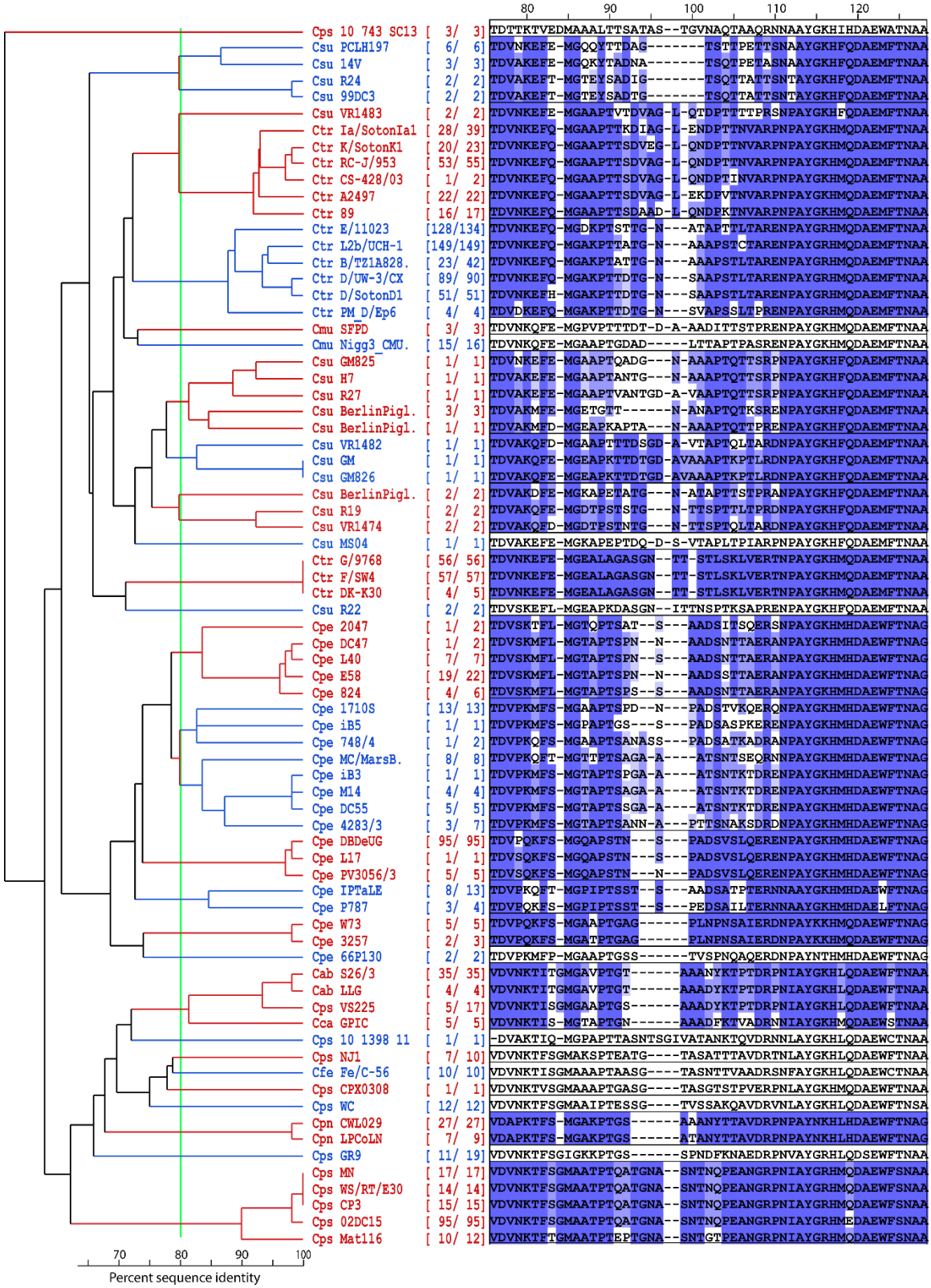


Fig. 2.S18. OmpA antibody-reactive region 90, OmpA_ARR90 (VD1). (Continued)

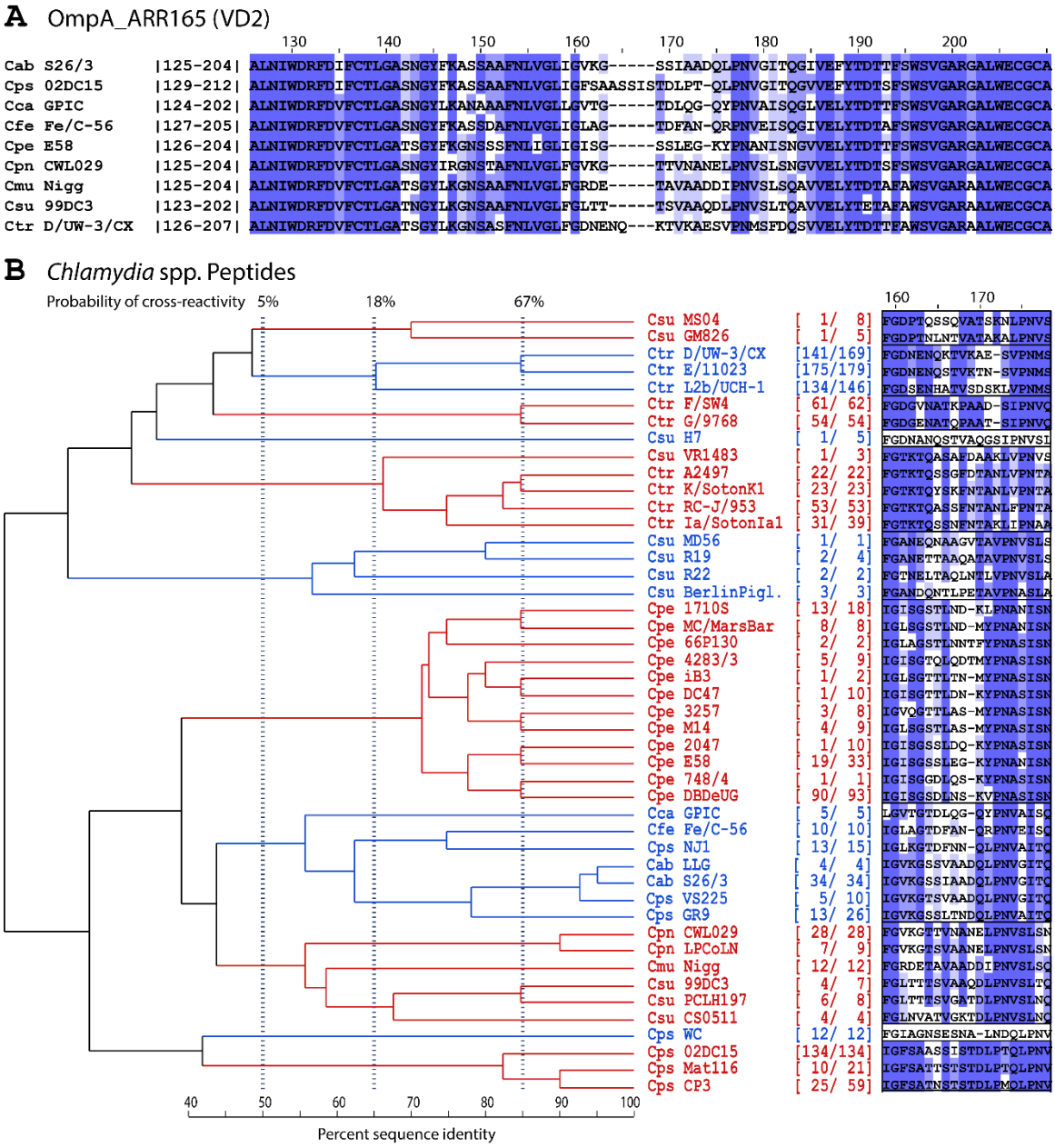


Fig. 2.S19. OmpA antibody-reactive region 165, OmpA_ARR165 (VD2).

(A) Complete ARR165 (VD2) alignment of the major outer membrane protein A (OmpA; locus tag CT_681 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among OmpA_VD2 sequences. VD2 shows high evolutionary divergence, allowing for robust species-, and even serovar di_erentiation. Deep clades cluster strain-variant peptide sequences of single species with the following exceptions: Cps strain NJ1 occupies the same clade with Cfe; and Cps strain VS225 & GR9 with Cab strains.

C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_OmpA_163-170	SSIAADQL	**5**	0	0	0	0	0	0	0	0
Cab_S26/3_OmpA_158-173	IGVKGSSIAADQLPNV	**225**	0	-	-	-	0	-	-	-
Cab_S26/3_OmpA_153-176	NLVGLIGVKGSSIAADQLPNVGIT	**632**	0	0	0	0	0	0	0	0
Cps_02DC15_OmpA_163-178	GFSAASSISTDLPTQL	0	**12**	0	0	0	0	0	0	0
Cps_02DC15_OmpA_158-181	LVGLIGFSAASSISTDLPTQLPNV	2	**234**	0	0	0	0	0	0	0
Cca_GPIC_OmpA_162-169	TDLQGQYP	0	0	**0**	0	0	0	0	0	0
Cca_GPIC_OmpA_159-168	VTGTDLQGQY	0	0	**158**	0	0	0	0	0	0
Cfe_Fe/C_OmpA_165-171	TDFANQR	0	0	0	**0**	0	0	0	0	0
Cfe_Fe/C_OmpA_160-175	IGLAGTDFANQRPNVE	0	0	0	**107**	0	0	0	0	0
Cpe_E58_OmpA_164-170	SSLEGKY	0	0	0	0	**0**	0	0	0	0
Cpe_E58_OmpA_164-174	SSLEGKYPNAN	-	-	-	-	**12**	0	-	-	0
Cpe_E58_OmpA_161-176	ISGSSLEGKYPNANIS	0	0	0	0	**33**	0	0	0	0
Cpe_1710S_OmpA_161-176	ISGSTLNDKLPNANIS	0	0	0	0	0	0	0	0	0
Cpe_DC47_OmpA_161-176	ISGTTLDNKYPNASIS	0	0	0	0	0	0	0	0	0
Cpe_iB3_OmpA_161-176	LSGTTLTNMYPNASIS	0	0	0	0	0	0	0	0	0
Cpe_4283/3_OmpA_161-176	SGTQLQDTMYPNASIS	0	0	0	0	0	0	0	0	0
Cpn_CWL029_OmpA_163-170	TTVNANEL	0	0	0	0	0	**0**	0	0	0
Cpn_CWL029_OmpA_158-173	FGVKGTTVNANELPNV	0	0	0	0	0	**30**	0	0	0
Cmu_Nigg3_OmpA_160-170	RDETAVAADDI	0	0	0	0	0	0	**58**	0	0
Csu_99DC3_OmpA_166-181	FGLTTTSVAAQDLPNV	0	0	0	0	0	0	0	**42**	0
Csu_99DC3_OmpA_179-194	PNVSLTQAVVELYTET	0	0	0	0	0	0	0	**1**	2
Ctr_D/UW-3_OmpA_126-141	ALNIWDRFDVFCTLGA	0	0	0	0	0	0	0	0	**0**
Ctr_D/UW-3_OmpA_149-164	NSASFNLVGLFGDNEN	0	0	0	0	0	0	0	0	**0**
Ctr_D/UW-3_OmpA_159-174	FGDNENQKTVKAESVP	0	0	0	0	0	0	0	0	**31**
Ctr_D/UW-3_OmpA_152-181	SFNLVGLFGDNENQKTVKAESVPNMSFDQS	0	5	0	0	0	0	0	0	**97**

Fig. 2.S19. OmpA antibody-reactive region 165, OmpA_ARR165 (VD2). (Continued)

(C) Antibody reactivity of OmpA_VD2 peptides. The VD2 peptides do not cross-react with heterologous sera because of their evolutionary divergence (B), and therefore they are suitable for di_erentiation of Cab, Cca, Cfe, Cpn and Cmu species or serovars of Csu, Ctr, Cpe and Cps.

(D) OmpA_VD2: Sequence variations and major clades in Chlamydia *spp*. *Please find Figure 2.12D on the next page.* The sequences of *Chlamydia* spp./serovars are grouped into boxes based on the percent identity among 40 AA long VD2 sequences, and 80% sequence identity between common ancestors is used as the threshold level (shown in green vertical line). VD2 shows high evolutionary divergence resulting in many serovars, particularly in Cpe, Csu, Cps and Ctr.

D OmpA_VD2: Major Clades of *Chlamydia* spp.

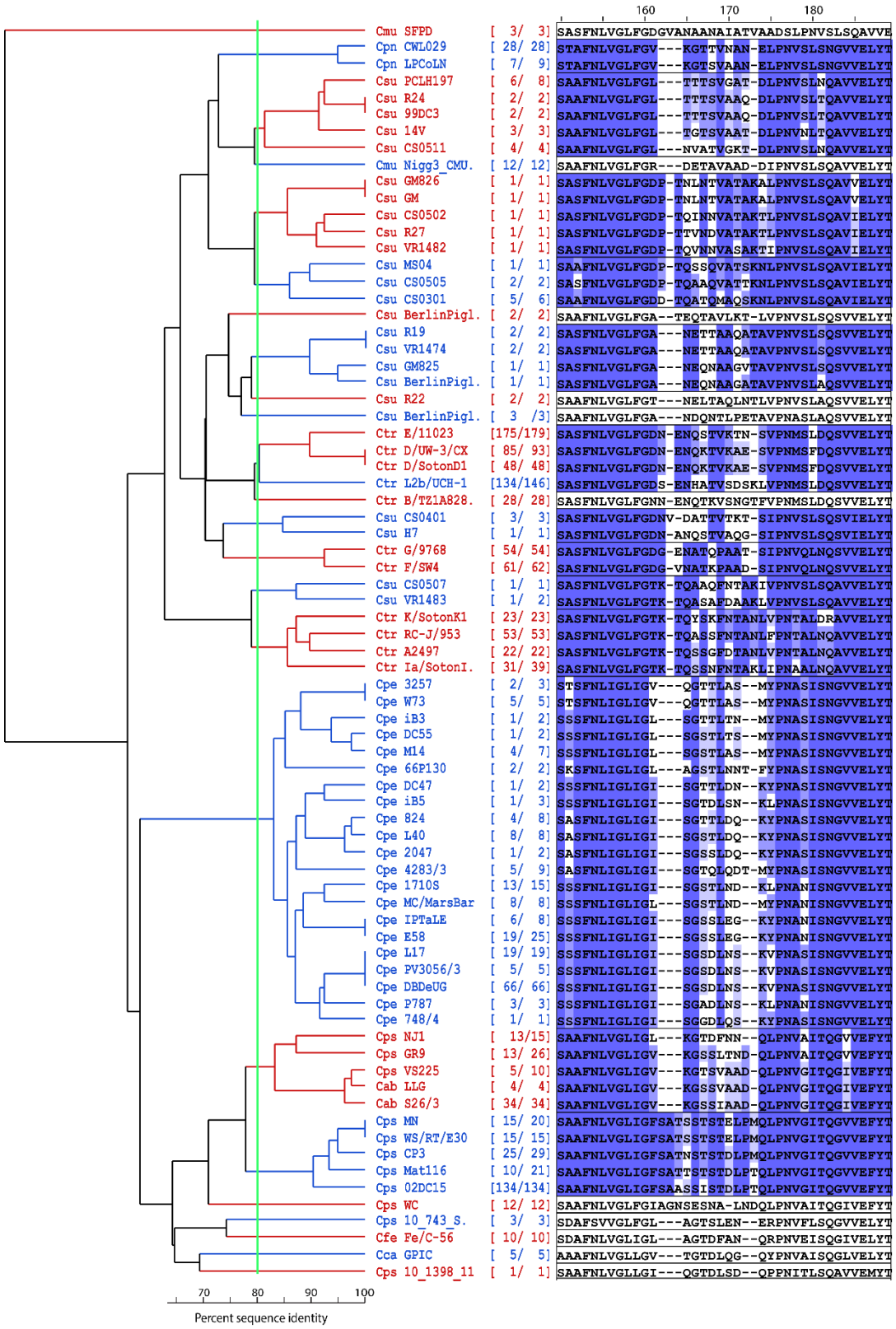


Fig. 2.S19. OmpA antibody-reactive region 165, OmpA_ARR165 (VD2). (Continued)

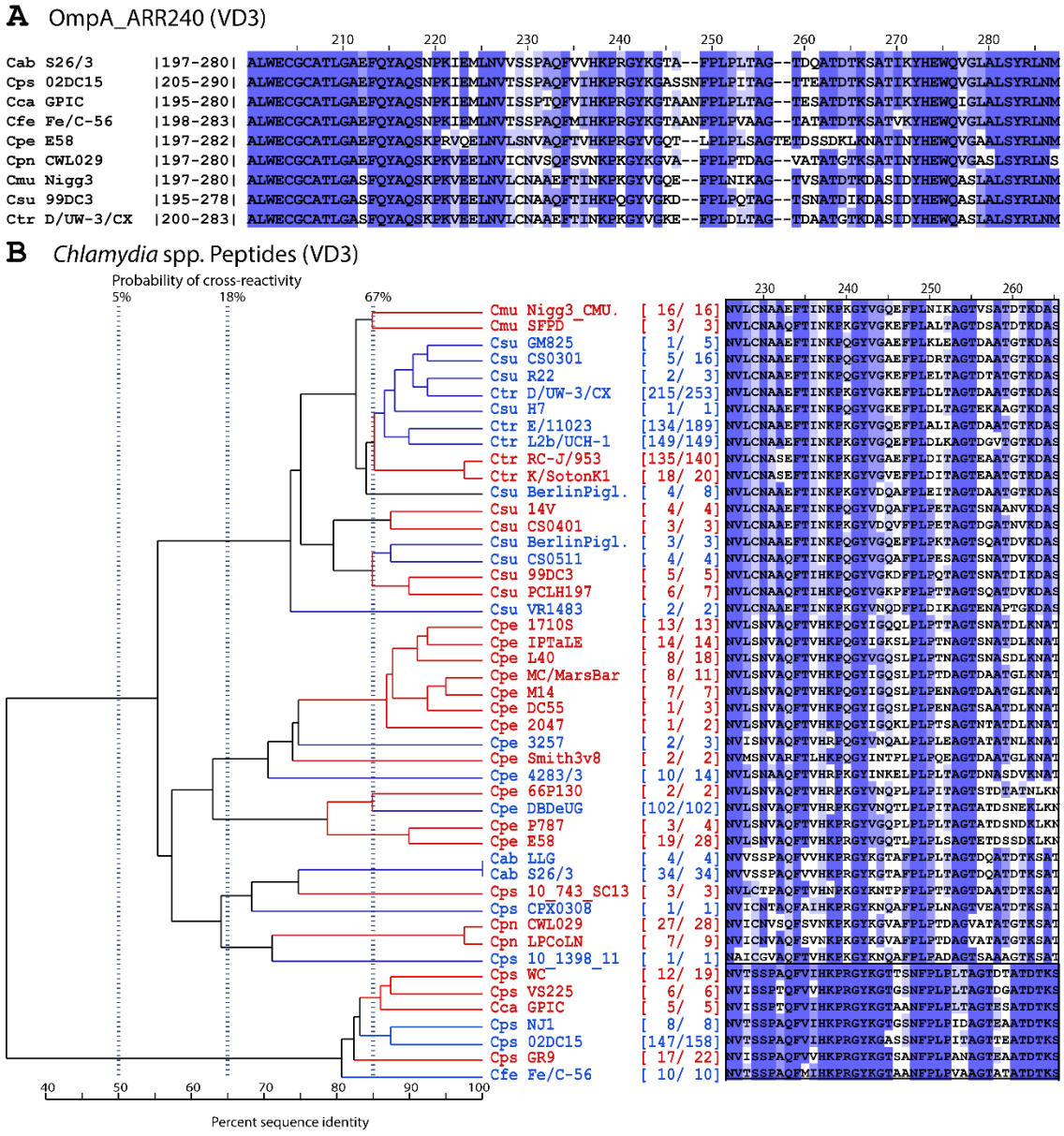


Fig. 2.S20. OmpA antibody-reactive region 240, OmpA_ARR240 (VD3)

(A) Complete ARR240 (VD3) alignment of the major outer membrane protein A (OmpA; locus tag CT_681 of *C. trachomatis* strain D/UW-3/CX).

(B) Evolutionary relationships among OmpA_VD3 (VD3) sequences of Chlamydia *spp*. Deep clades (<50% sequence identity) are separated into boxes, and more closely related clades (<85% sequence identity) are indicated by alternating color. VD3 of Ctr and Cps shows lower evolutionary divergence than VD1, VD2 & VD4b (Figure 2.11, 2.12 & 2.14), and therefore allows detection of the majority of strains by use of few peptides. In contrast, Cpe and Csu show high divergence requiring several peptides to cover the majority of strains of these two species.

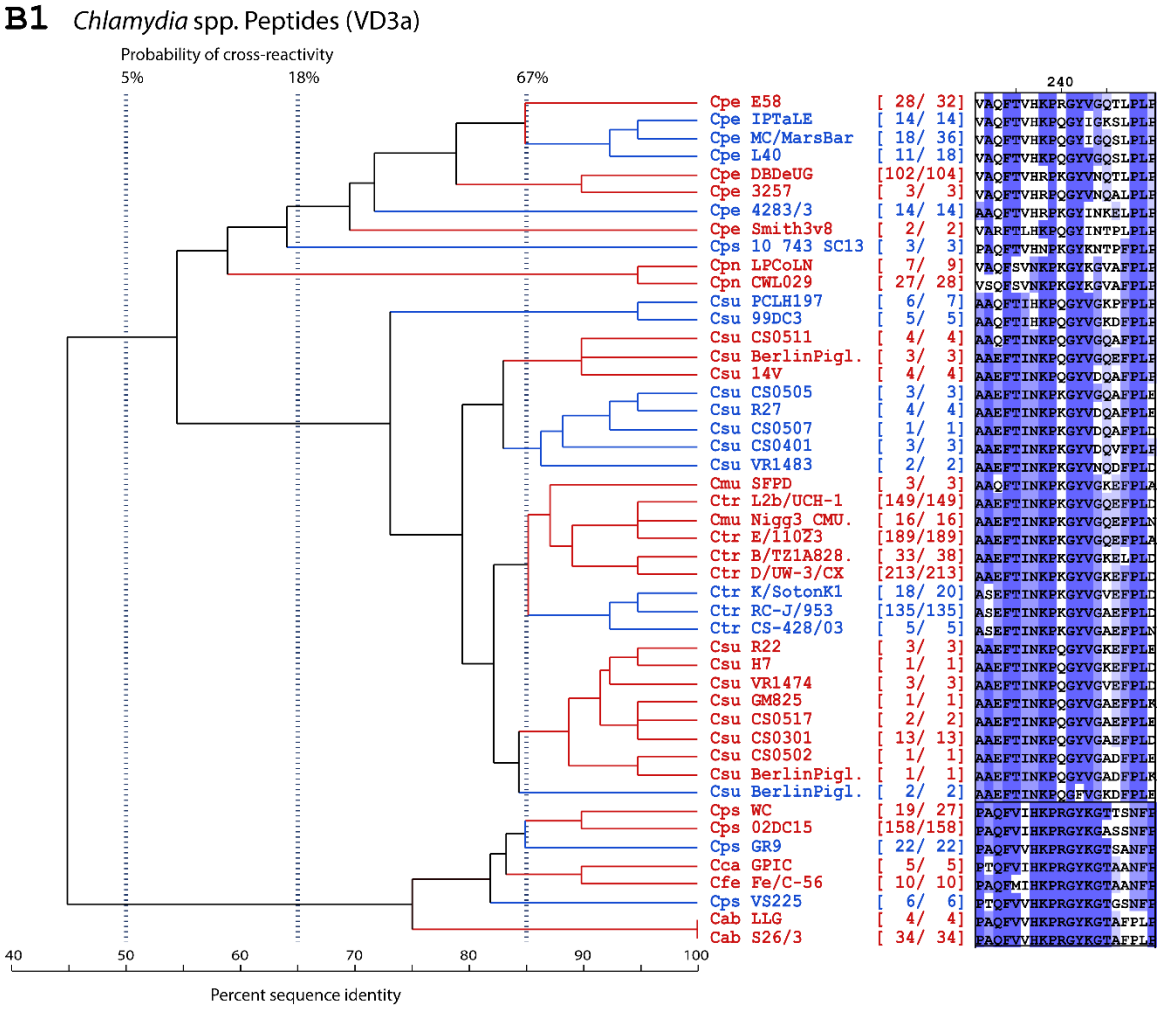


Fig. 2.S20. OmpA antibody-reactive region 240, OmpA_ARR240 (VD3). (Continued)

(B1) Evolutionary relationships among upstream OmpA_VD3 (VD3a) sequences of Chlamydia *spp*. The sequences of VD3a cluster into two deep clades of closely related *Chlamydia* spp./serovars (clade Cpe, Cpn, Csu, Cmu & Ctr and clade Cps, Cca, Cfe & Cab). Extensive sharing of amino acids occurs within each clade of VD3a.

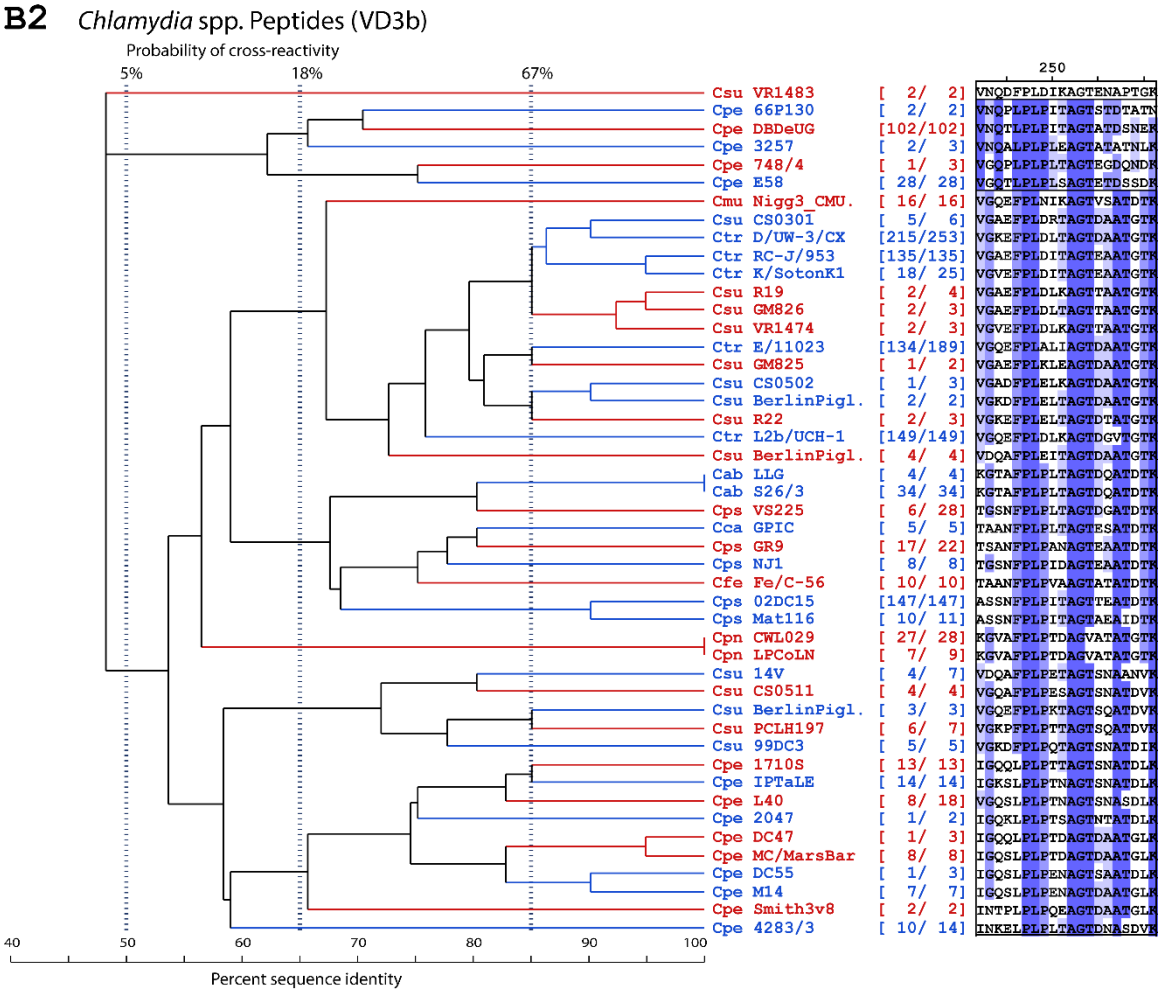


Fig. 2.S20. OmpA antibody-reactive region 240, OmpA_ARR240 (VD3). (Continued)

(B2) Evolutionary relationships among downstream OmpA_VD3 (VD3b) sequences of Chlamydia *spp*. Similar to VD3a, VD3b show extensive sequence homology among strains, yet also extensive sequence variations, particularly in Cpe and Csu. The sequences of VD3a cluster into three deep clades (clades Csu VR1483, Cpe E58, and all remaining *Chlamydia* spp. serovars).

C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_OmpA_244-259	FPLPLTAGTDQATDTK	**0**	0	-	-	-	0	-	-	-
Cab_S26/3_OmpA_223-262	NVVSSPAQFVVHKPRGYKGTAFPLPLTAGTDQATDTKSAT	**6**	2	0	4	0	142	0	0	0
Cps_05DC15_OmpA_250-265	ASSNFPLPITAGTTEA	0	**55**	0	0	0	0	0	0	0
Cca_GPIC_OmpA_240-255	TAANFPLPLTAGTESA	-	0	**0**	0	-	-	-	-	-
Cfe_Fe/C_OmpA_243-258	TAANFPLPVAAGTATA	-	0	0	**0**	-	-	-	-	-
Cpe_E58_OmpA_242-255	QTLPLPLSAGTETD	0	0	0	0	**0**	0	0	0	0
Cpe_E58_OmpA_242-257	QTLPLPLSAGTETDSS	-	-	-	0	**0**	-	-	0	-
Cpe_E58_OmpA_249-264	SAGTETDSSDKLKN	0	0	0	0	**0**	0	0	0	0
Cpe_E58_OmpA_260-275	LKNATINYHEWQVGAA	0	0	0	0	**0**	0	0	0	0
Cpe_DC55_OmpA_242-257	QSLPLPENAGTSAATD	-	-	-	6	0	**-**	-	0	-
Cpe_3v8_OmpA_242-257	TPLPLPQEAGTDAATG	-	-	-	0	0	**-**	-	0	-
Cpe_1710S_OmpA_242-257	QQLPLPTTAGTSNATD	-	-	-	0	0	**-**	-	0	-
Cpe_4283/3_OmpA_242-257	KELPLPLTAGTDNASD	-	-	-	0	0	**-**	-	0	-
Cpn_CWL029_OmpA_242-257	VAFPLPTDAGVATATG	0	0	0	0	0	**32**	0	0	0
Cmu_Nigg3_OmpA_242-257	QEFPLNIKAGTVSATD	0	0	0	0	0	0	**22**	0	0
Csu_99DC3_OmpA_223-238	KDFPLPQTAGTSNATD	-	-	-	-	-	-	0	**1**	4
Csu_99DC3_OmpA_239-254	IKDASIDYHEWQASLA	0	0	0	0	0	0	0	**2**	0
Ctr_D/UW-3_OmpA_233-248	EFTINKPKGYVGKEFP	0	0	0	0	0	0	0	99	**259**
Ctr_D/UW-3_OmpA_245-260	KEFPLDLTAGTDAATG	0	0	0	0	0	0	0	0	**381**
Ctr_D/UW-3_OmpA_226-265	NVLCNAAEFTINKPKGYVGKEFPLDLTAGTDAATGTKDAS	0	0	0	0	0	0	0	83	**671**
Ctr_D/UW-3_OmpA_259-274	TGTKDASIDYHEWQAS	0	0	0	-	0	0	0	0	**0**

Fig. 2.S20. OmpA antibody-reactive region 240, OmpA_ARR240 (VD3). (Continued)

(C) Antibody reactivity of OmpA_VD3 peptides. Please find Figure 2.13C on the preceding page. Because of extensive shared amino acid residues (B), the Cab_S26/3_OmpA_223-262 peptide reacts with heterologous Cpn-sera, and Ctr_D/UW-3_OmpA_233-248 & Ctr_D/UW-3_OmpA_226-265 peptides react with heterologous Csu-sera. The Ctr_D/UW-3_OmpA_245-260 and Cps_05DC15_OmpA_250-265 peptides react only with homologous sera; however, the central residues are shared with other *Chlamydia* spp., therefore these peptides have a high probability of cross-reactivity (B2). Cross-reactivity of peptides do not allow conclusive species-specific detection of anti-chlamydial antibodies, however, this cross-reactivity may be advantageous for detection of the majority of strains by use of only few peptides.

(D) OmpA_VD3: Sequence variations and major clades in Chlamydia *spp*. *Please find Figure 2.13D on the next page.* The sequences of *Chlamydia* spp./serovars are grouped into boxes based on the percent identity among 50 AA long VD3 sequences, and 80% sequence identity between common ancestors is used as the threshold level (shown in green vertical line). VD3 shows lowest evolutionary divergence of the OmpA VDs, particularly in *C. trachomatis and C. psittaci*. The dominant serovars of Ctr and Cps can be detected by use of only few peptides. However, Cpe & Csu are very divergent in VD3 with many variations of peptide sequences.

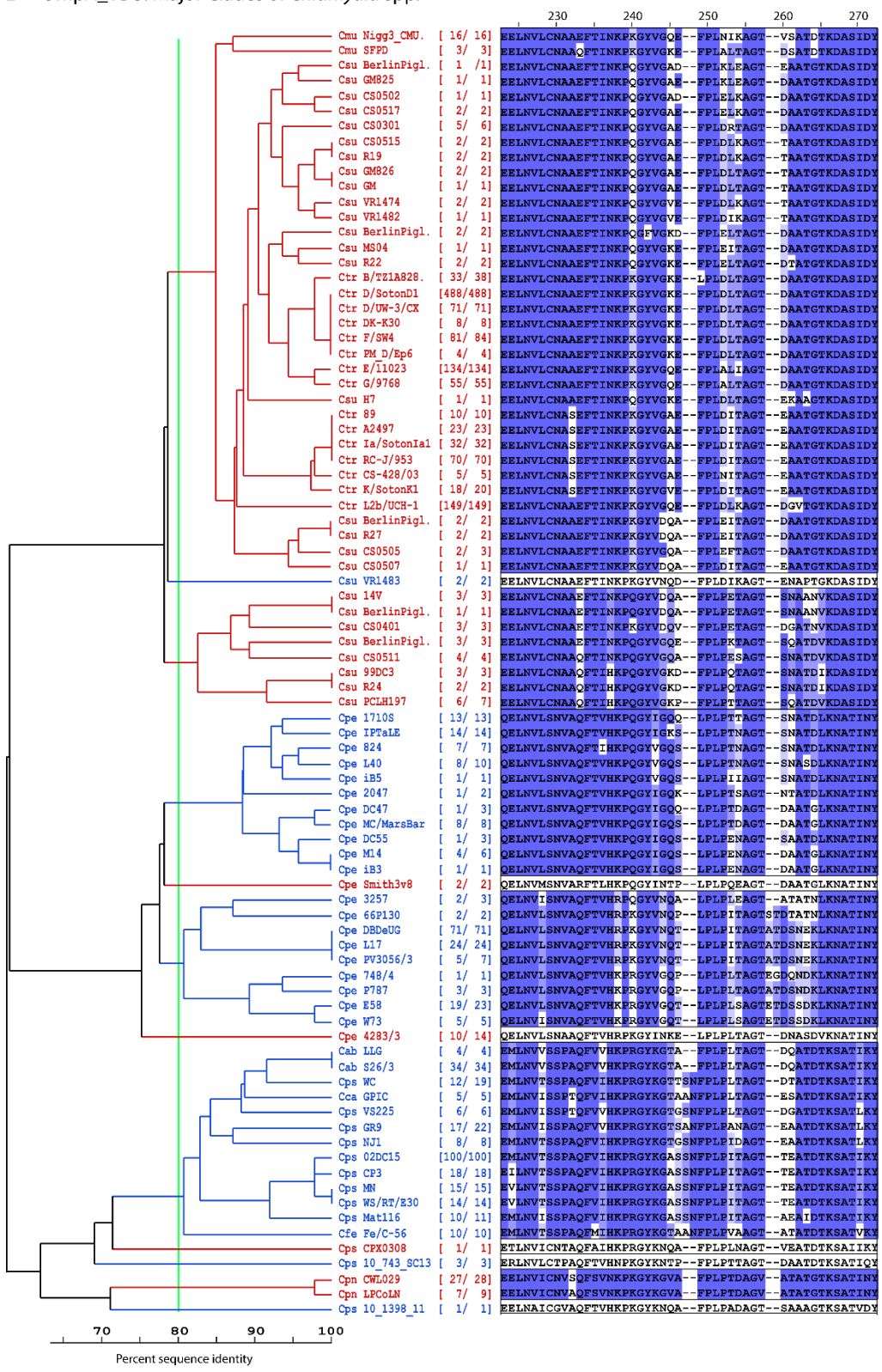


Fig. 2.S20. OmpA antibody-reactive region 240, OmpA_ARR240 (VD3). (Continued)

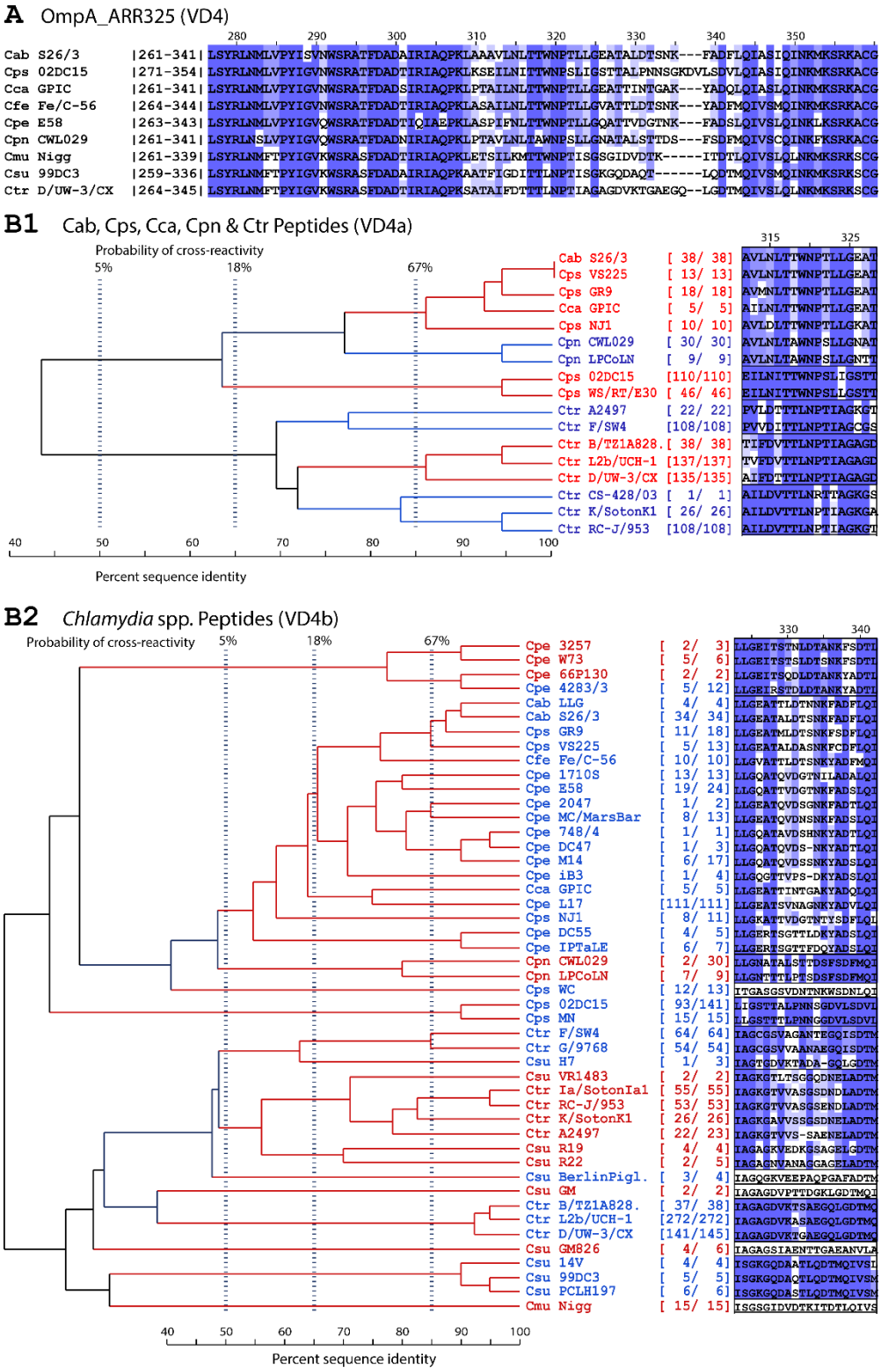


Fig. 2.S21: OmpA antibody-reactive region 325, OmpA_ARR325 (VD4).

(A) Complete ARR325 (VD4) alignment of the major outer membrane protein A (OmpA; locus tag CT_681 of *C. trachomatis* strain D/UW-3/CX).

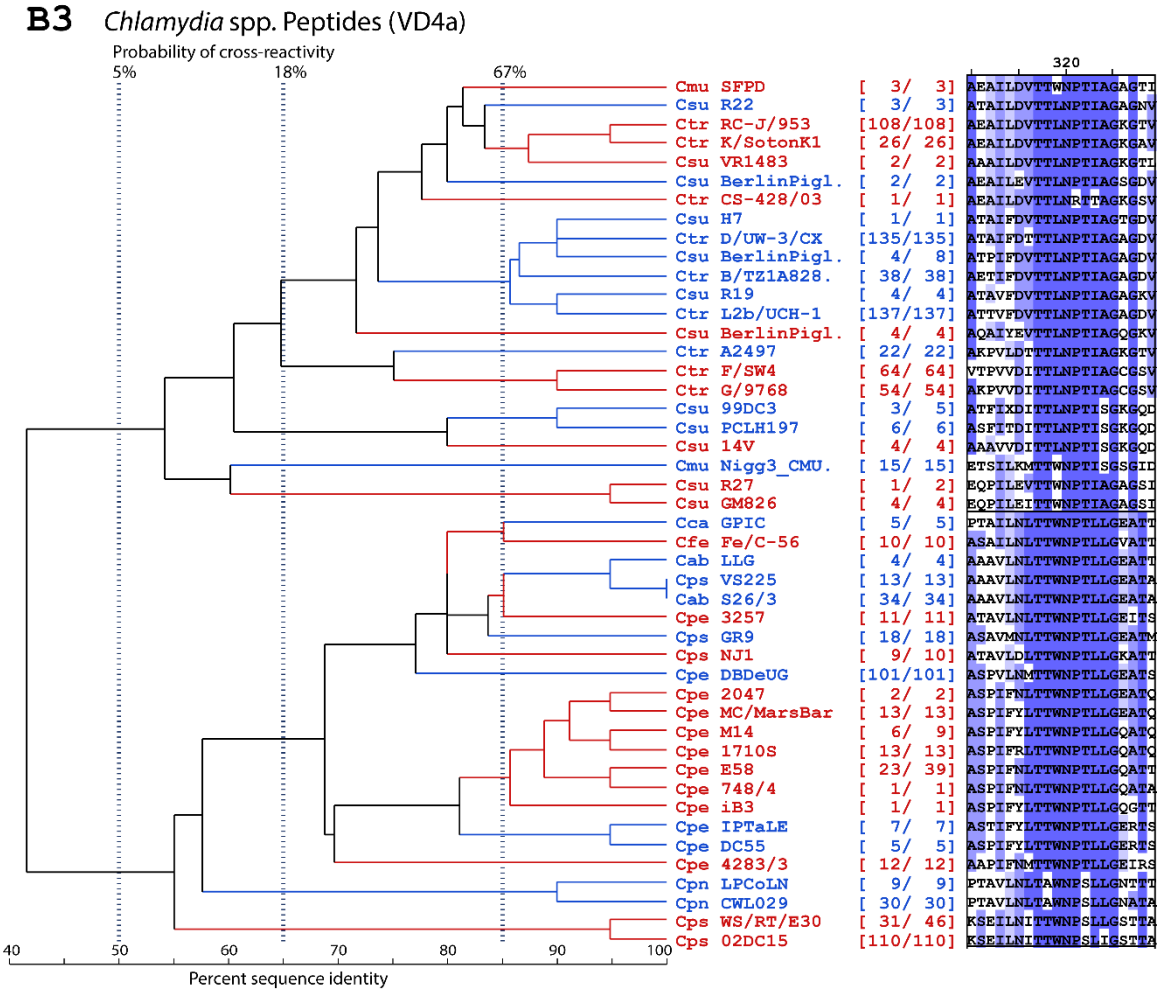


Fig. 2.S21: OmpA antibody-reactive region 325, OmpA_ARR325 (VD4). (Continued)

(B1) Evolutionary relationships among upstream OmpA_VD4 (VD4a) sequences of Cab, Cps, Cca, Cpn & Ctr spp. Clades (<75% sequence identity) are separated into boxes. VD4a shows low evolutionary divergence, allowing detection of all known serovars of *Chlamydia* spp. with a single (Cab, Cca & Cpn) or only few peptides (Cps and Ctr).

(B2) Evolutionary relationships among downstream OmpA_VD4 (VD4b) sequences of *Chlamydia* spp. Sequences from deep clades (<50% sequence identity) are separated into boxes. VD4b shows high evolutionary divergence, allowing for robust species-, and even serovar differentiation. Deep clades cluster strain-variant peptides of single species with the following exceptions: Cab, Cps strains GR9 & VS225 and Cfe occupy the same deep clade together with numerous Cpe strains; Cpe strain L17 with Cca; and Csu strain VR1483 with Ctr strains. However, VD4b sequences of Cpn, Cmu or major clades of Cps, Ccu & Ctr strains are well-separated and therefore suitable for differentiation of these species or serovars.

C Peptide Reactivities

Peptide	Sequence	Anti-chlamydial sero-reactivity								
		Cab	Cps	Cca	Cfe	Cpe	Cpn	Cmu	Csu	Ctr
Cab_S26/3_OmpA_326-333	ALDTSNKF	**0**	0	0	0	0	0	0	0	0
Cab_S26/3_OmpA_324-339	ATALDTSNKFADFLQI	**4**	0	0	0	0	6	0	0	0
Cab_S26/3_OmpA_309-324	AVLNLTTWNPTLLGEA	**7**	47	28	0	4	31	2	0	402
Cab_S26/3_OmpA_302-341	AQPKLAAAVLNLTTWNPTLLGEATALDTSNKFADFLQIAS	**47**	174	49	5	18	271	315	2	1
Cps_02DC15_OmpA_333-348	STTALPNNSGKDVLSD	0	**433**	0	0	0	0	0	0	0
Cca_GPIC_OmpA_309-324	AILNLTTWNPTLLGEA	7	43	**30**	0	4	27	2	0	398
Cca_GPIC_OmpA_326-333	TINTGAKY	0	0	**0**	0	0	0	0	0	0
Cca_GPIC_OmpA_321-336	LGEATTINTGAKYADQ	-	1	**37**	0	-	-	-	-	-
Cfe_Fe/C_OmpA_329-336	TLDTSNKY	0	0	0	**0**	0	0	0	0	0
Cfe_Fe/C_OmpA_324-339	LGVATTLDTSNKYADF	-	1	0	**2**	-	-	-	-	-
Cpe_E58_OmpA_313-328	PIFNLTTWNPTLLGQA	13	33	24	0	**7**	41	0	0	109
Cpe_E58_OmpA_323-338	LGQATTVDGTNKFADS	0	0	0	0	**170**	0	0	0	0
Cpe_2047_OmpA_323-338	LGEATQVDSGNKFADT	0	0	0	0	**0**	0	0	0	0
Cpe_M14_OmpA_323-338	LGQATQVDSSNKYADS	0	0	0	0	**0**	0	0	0	0
Cpe_748/4_OmpA_323-338	LGQATAVDSHNKYADT	0	0	0	0	**0**	0	0	0	0
Cpe_L17_OmpA_323-338	LGEATSVNAGNKYADV	0	0	0	0	**0**	0	0	0	0
Cpe_1710S_OmpA_323-338	LGQATQVDGTNILADA	0	0	0	0	**0**	0	0	0	0
Cpe_DC55_OmpA_323-338	LGERTSGTTLDKYADS	0	0	0	0	**0**	0	0	0	0
Cpe_iB3_OmpA_323-338	LGQGTTVPSDKYADSL	0	0	0	0	**0**	0	0	0	0
Cpe_3257_OmpA_323-338	EITSTNLDTANKFSDT	0	0	0	0	**0**	0	0	0	0
Cpe_4283/3_OmpA_323-338	EIRSTDLDTANKYADT	0	0	0	0	**0**	0	0	0	0
Cpn_CWL029_OmpA_309-324	AVLNLTAWNPSLLGNA	3	0	2	0	0	**57**	**1**	0	38
Cpn_CWL029_OmpA_326-335	ALSTTDSFSD	0	0	0	0	0	**3**	0	0	0
Cpn_CWL029_OmpA_324-339	ATALSTTDSFSDFMQI	0	0	-	-	-	**6**	-	-	-
Cmu_Nigg_OmpA_311-326	ETSILDMTTWNPTISG	0	1	1	0	0	1	**0**	0	233
Cmu_Nigg_OmpA_325-333	IDVDTKITD	0	0	0	0	0	0	**6**	0	0
Cmu_Nigg_OmpA_320-335	ISGSGIDVDTKITDTL	0	0	0	0	0	0	**43**	0	0
Csu_99DC3_OmpA_313-328	FIGDITTLNPTISGKG	0	1	0	0	0	0	0	**0**	293
Csu_99DC3_OmpA_317-332	TISGKGQDAQTLQDTM	0	0	0	0	0	0	0	**94**	0
Ctr_D/UW-3_OmpA_292-307	WSRASFDADTIRIAQP	0	0	0	0	0	0	0	0	**0**
Ctr_D/UW-3_OmpA_313-328	IFDTTTLNPTIAGAGD	0	0	0	0	0	0	0	0	**944**
Ctr_D/UW-3_OmpA_324-339	AGAGDVKTGAEGQLGD	0	0	0	0	0	0	0	0	**391**
Ctr_D/UW-3_OmpA_306-345	QPKSATAIFDTTTLNPTIAGAGDVKTGAEGQLGDTMQIVS	0	0	0	0	0	0	0	0	**996**
Ctr_D/UW-3_OmpA_339-354	DTMQIVSLQLNKMKSR	0	0	0	0	0	0	0	0	**0**

Fig. 2.S21: OmpA antibody-reactive region 325, OmpA_ARR325 (VD4). (Continued)

(B3) Evolutionary relationships among upstream OmpA_VD4 (VD4a) sequences of *Chlamydia* spp. Sequences from deep clades (<50% sequence identity) are separated into boxes and more closely related clades (<85% sequence identity) are indicated by alternating color. VD4a shows low evolutionary divergence, particularly for highly conserved central residues, allowing detection of all known serovars of a chlamydial species with a single (Cab, Cca & Cpn) or only few peptides (Cps and Ctr).

(C) Antibody reactivity of OmpA_VD4 peptides. The Cab VD4a peptides are evolutionary close, and therefore cross-reactive with heterologous sera. However, these VD4a peptides can be used to detect antibodies produced against all serovars of *Chlamydia* spp. using only few peptides. The VD4b peptides are evolutionary divergent, therefore do not cross-react with heterologous sera, and are suitable for differentiation of *C. muridarum* or serovars of Cps, Ctr and Csu.

(D) OmpA_VD4: Sequence variations and major clades in *Chlamydia* spp. *Please _nd Figure 2.13D on the next page.* The sequences of *Chlamydia* spp./serovars are grouped into boxes based on the percent identity among 50 AA long VD4 sequences, and 80% sequence identity between common ancestors is used as the threshold level (shown in green vertical line). VD4 shows high evolutionary divergence including many serovars, particularly for Cpe, Csu, Cps and Ctr.

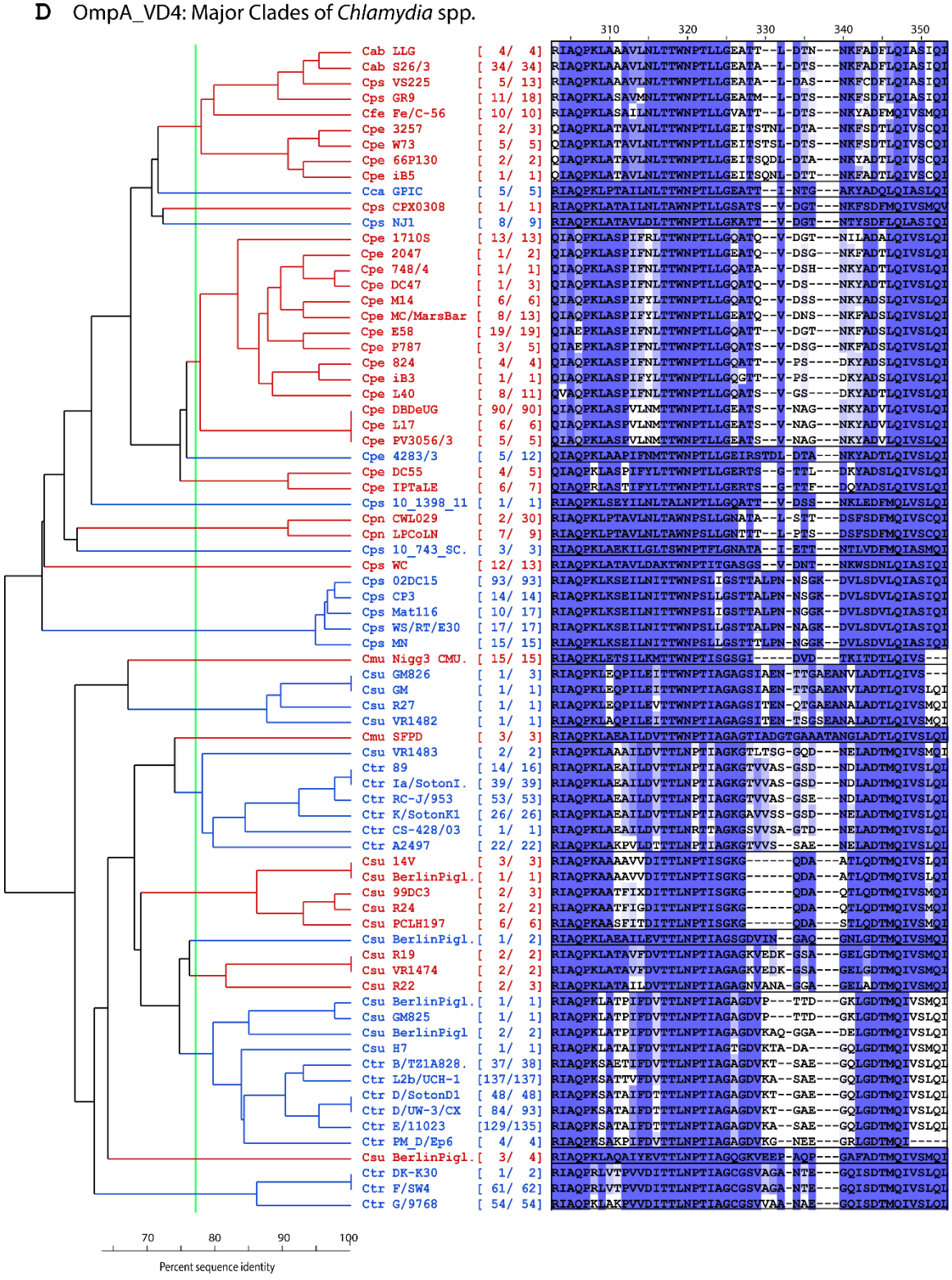


Fig. 2.S21: OmpA antibody-reactive region 325, OmpA_ARR325 (VD4). (Continued)

REFERENCES

1. **Kuo CC, Stephens RS, Bavoil PM and Kaltenboeck B. 2011.** Genus I. *Chlamydia* Jones, Rake and Stearns 1945, 55^{AL}. *In* Bergey's Manual of Systematic Bacteriology, 2nd edition, vol 4, pp 846-865. Springer, New York.

2. **Sachse K, Laroucau K, Riege K, Wehner S, Dilcher M, Creasy HH, Weidmann M, Myers G, Vorimore F, Vicari N, Magnino S, Liebler-Tenorio E, Ruettger A, Bavoil PM, Hufert FT, Rosselló-Móra R and Marz M. 2014.** Evidence for the existence of two new members of the family *Chlamydiaceae* and proposal of *Chlamydia avium* sp. nov. and *Chlamydia gallinacea* sp. nov. *Syst Appl Microbiol* **37**: 79-88.

3. **Everett KD, Bush RM and Andersen AA. 1999.** Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. *Int J Syst Bacteriol* **49**, 415–440.

4. **Sachse K, Bavoil PM, Kaltenboeck B, Stephens RS, Kuo C-C, Rosselló-Móra R and Horn M. 2015.** Emendation of the family *Chlamydiaceae*: Proposal of a single genus, *Chlamydia*, to include all currently recognized species. *Syst Appl Microbiol* http://dx.doi.org/10.1016/j.syapm.2014.12.004.

5. **Kaltenboeck B, Schmeer N and Schneider R. 1997.** Evidence for numerous *omp*1 alleles of porcine *Chlamydia trachomatis* and novel chlamydial species obtained by PCR. *J Clin Microbiol* **35**: 1835-1841.

6. **Mohamad KY, Roche SM, Myers GS, Bavoil PM, Laroucau K, Magnino S, Laurent S, Rasschaert D and Rodolakis A. 2008.** Preliminary phylogenetic identification of virulent *Chlamydophila pecorum* strains. *Infect Genet Evol* **8**: 764-771.

7. **Mohamad KY, Kaltenboeck B, Rahman KS, Magnino S, Sachse K and Rodolakis A. 2014.** Host adaptation of *Chlamydia pecorum* towards low virulence evident in co-evolution of the *ompA*, *incA*, and ORF663 loci. *PLoS One* **9**: e103615.

8. **Wang SP and Grayston JT. 1974.** Human serology in *Chlamydia trachomatis* infection with microimmunofluorescence. *J Infect Dis* **130**: 388-397.

9. **Nunes A, Borrego MJ, Nunes B, Florindo C, Gomes JP and 2009.** Evolutionary dynamics of *ompA*, the gene encoding the *Chlamydia trachomatis* key antigen. *J Bacteriol* **191**: 7182-7192.

10. **Vanrompay D, Butaye P, Sayada C, Ducatelle R and Haesebrouck F. 1997.** Characterization of avian *Chlamydia psittaci* strains using *omp1* restriction mapping and serovar-specific monoclonal antibodies. *Res Microbiol* **148**: 327-333.

11. **Andersen AA. 2005.** Serotyping of US isolates of *Chlamydophila psittaci* from domestic and wild birds. *J Vet Diagn Invest* **17**: 479-482.

12. **Sachse K, Laroucau K, Vorimore F, Magnino S, Feige J, Müller W, Kube S, Hotzel H, Schubert E, Slickers P and Ehricht R. 2009.** DNA microarray-based genotyping of *Chlamydophila psittaci* strains from culture and clinical samples. *Vet Microbiol* **135**: 22-30.

13. **Dean D, Rothschild J, Ruettger A, Kandel RP and Sachse K. 2013.** Zoonotic *Chlamydiaceae* species associated with trachoma, Nepal. Emerg. *Infect Dis* **19**: 1948.

14. **Sachse, K, Vretou E, Livingstone M, Borel N, Pospischil A and Longbottom D. 2009.** Recent developments in the laboratory diagnosis of chlamydial infections. *Vet Microbiol* **135**: 2-21.

15. **Ozanne G and Lefebvre J. 1992.** Specificity of the microimmunofluorescence assay for the serodiagnosis of *Chlamydia pneumoniae* infections. *Can J Microbiol* **38**: 1185-1189.

16. **Kern DG, Neill MA and Schachter J. 1993.** A seroepidemiologic study of *Chlamydia pneumoniae* in Rhode Island. Evidence of serologic cross-reactivity. *Chest J* **104**: 208-213.

17. **Wagenvoort JHT, Koumans D and Van de Cruijs M. 1999.** How useful is the *Chlamydia* micro-immunofluorescence (MIF) test for the gynaecologist? *Eur J Obstet Gynecol Reprod Biol* **84**: 13-15.

18. **Wong YK, Sueur JM, Fall CH, Orfila J and Ward ME. 1999.** The species specificity of the microimmunofluorescence antibody test and comparisons with a time resolved fluoroscopic immunoassay for measuring IgG antibodies against *Chlamydia pneumoniae*. *J Clin Pathol* **52**: 99-102.

19. **Leonard CA and Borel N. 2014.** Chronic chlamydial diseases: from atherosclerosis to urogenital infections. *Curr Clin Microbiol Reports* **1**: 1-12.

20. **Poudel A, Elsasser TH, Rahman KS, Chowdhury EU and Kaltenboeck B. 2012.** Asymptomatic endemic *Chlamydia pecorum* infections reduce growth rates in calves by up to 48 percent. *PLoS One* **7**: e44961.

21. **Reinhold P, Sachse K and Kaltenboeck B. 2011.** *Chlamydiaceae* in cattle: Commensals, trigger organisms, or pathogens? *Vet J* **189**: 257-267.

22. **Wang SP and Grayston JT. 1970.** Immunologic relationship between genital TRIC, lymphogranuloma venereum, and related organisms in a new microtiter indirect immunofluorescence test. *Am J Ophthalmol* **70**: 367-374.

23. **Bas S, Muzzin P, Ninet B, Bornand JE, Scieux C and Vischer TL. 2001.** Chlamydial serology: comparative diagnostic value of immunoblotting, microimmunofluorescence test, and immunoassays using different recombinant proteins as antigens. *J Clin Microbiol* **39**: 1368-1377.

24. **Bas S, S Genevay, Schenkel MC and Vischer TL. 2002.** Importance of species-specific antigens in the serodiagnosis of *Chlamydia trachomatis* reactive arthritis. *Rheumatol* **41**: 1017-1020.

25. **Bas S, Muzzin P and Vischer TL. 2001.** *Chlamydia trachomatis* serology: diagnostic value of outer membrane protein 2 compared with that of other antigens. *J Clin Microbiol* **39**: 4082-4085.

26. **Stephens RS, Tam MR, Kuo CC and Nowinski RC. 1982.** Monoclonal antibodies to *Chlamydia trachomatis*: antibody specificities and antigen characterization. *J Immunol* **128**: 1083-1089.

27. **Batteiger BE, Newhall WJ, Terho P, Wilde CE and Jones RB. 1986.** Antigenic analysis of the major outer membrane protein of *Chlamydia trachomatis* with murine monoclonal antibodies. *Infect Immun* **53**: 530-533.

28. **Baehr W, Zhang YX, Joseph T, Su HUA, Nano FE, Everett KD and Caldwell HD. 1988.** Mapping antigenic domains expressed by *Chlamydia trachomatis* major outer membrane protein genes. *Proc Natl Acad Sci USA* **85**: 4000-4004.

29. **Mygind P, Christiansen G, Persson K and Birkelund S. 2000.** Detection of *Chlamydia trachomatis*-specific antibodies in human sera by recombinant major outer-membrane protein polyantigens. *J Med Microbiol* **49**: 457-465.

30. **Klein M, Kötz A, Bernardo K and Krönke M. 2003.** Detection of *Chlamydia pneumoniae*-specific antibodies binding to the VD2 and VD3 regions of the major outer membrane protein. *J Clin Microbiol* **41**: 1957-1962.

31. **Livingstone M, Entrican G, Wattegedera S, Buxton D, McKendrick IJ and Longbottom D. 2005.** Antibody responses to recombinant protein fragments of the major outer membrane protein and polymorphic outer membrane protein POMP90 in *Chlamydophila abortus*-infected pregnant sheep. *Clin Diagn Lab Immunol* **12**: 770-777.

32. **Zhong GM, Reid RE and Brunham RC. 1990.** Mapping antigenic sites on the major outer membrane protein of *Chlamydia trachomatis* with synthetic peptides. *Infect Immun* **58**: 1450-1455.

33. **Conlan JW, Clarke IN and Ward ME. 1988.** Epitope mapping with solid-phase peptides: identification of type-, subspecies-, species-and genus-reactive antibody binding domains on the major outer membrane protein of *Chlamydia trachomatis*. *Mol Microbiol* **2**: 673-679.

34. **Pal S, Cheng X, Peterson EM and de la Maza LM. 1993.** Mapping of a surface-exposed B cell epitope to the variable sequent 3 of the major outer-membrane protein of *Chlamydia trachomatis*. *J Gen Microbiol* **139**: 1565-1570.

35. **Batteiger BE. 1996.** The major outer membrane protein of a single *Chlamydia trachomatis* serovar can possess more than one serovar-specific epitope. *Infect Immun* **64**: 542-547.

36. **Villeneuve A, Brossay L, Paradis G and Hébert J. 1994.** Determination of neutralizing epitopes in variable domains I and IV of the major outer-membrane protein from *Chlamydia trachomatis* serovar K. *Microbiol* **140**: 2481-2487.

37. **Närvänen A, Puolakkainen M, Hao W, Kino K and Suni J. 1997.** Detection of antibodies to *Chlamydia trachomatis* with peptide-based species-specific enzyme immunoassay. *Infect Dis Obstet Gynecol* **5**: 349-354.

38. **Baud D, Regan L and Greub G. 2010.** Comparison of five commercial serological tests for the detection of anti-*Chlamydia trachomatis* antibodies. *Eur J Clin Microbiol Infect Dis* **29**: 669-675.

39. **Sivalingam GN and Shepherd AJ. 2012.** An analysis of B cell epitope discontinuity. *Mol Immunol* **51**: 304-309.

40. **Kringelum JV, Nielsen M, Padkjær SB and Lund O. 2013.** Structural analysis of B cell epitopes in antibody: protein complexes. *Mol Immunol* **53**: 24-34.

41. **Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM and Pupko T. 2008.** Computational characterization of B cell epitopes. *Mol Immunol* **45**: 3477-3489.

42. **Van Regenmortel and MHV. 2009.** What is a B cell epitope? *In* Epitope Mapping Protocols, *Methods Mol Biol* vol **524**, Humana Press: pp 3-20.

43. **Wills GS, Horner PJ, Reynolds R, Johnson AM, Muir DA, Brown DW, Winston A, Broadbent AJ, Parker D and McClure MO. 2009.** Pgp3 antibody enzyme-linked immunosorbent assay, a sensitive and specific assay for seroepidemiological analysis of *Chlamydia trachomatis* infection. *Clin Vaccine Immunol* **16**: 835-843.

44. **Cunningham AF and Ward ME. 2003.** Characterization of human humoral responses to the major outer membrane protein and OMP2 of *Chlamydophila pneumoniae*. *FEMS Microbiol Lett* **227**: 73-79.

45. **Andresen H, Grötzinger C, Zarse K, Kreuzer OJ, Ehrentreich-Förster E and Bier FF. 2006.** Functional peptide microarrays for specific and sensitive antibody diagnostics. *Proteomics* **6**: 1376-1384.

46. **Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV and Davis RW. 1998.** Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**: 754-759.

47. **Kalman S, Mitchell W, Marathe R, Lammel C, Fan J, Hyman RW, Olinger L, Grimwood J, Davis RW and Stephens R. 1999.** Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nat Genet* **21**: 385-389.

48. **Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J and Fraser CM. 2000.** Genome

sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res* **28**: 1397-1406.

49. Mojica S, Creasy HH, Daugherty S, Read TD, Kim T, Kaltenboeck B, Bavoil P and Myers GS. 2011. Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58. *J Bacteriol* **193**: 3690.

50. Read TD, Myers GS, Brunham RC, Nelson WC, Paulsen IT, Heidelberg J, Holtzapple E, Khouri H, Federova NB, Carty HA, Umayam LA, Haft DH, Peterson J, Beanan MJ, White O, Salzberg SL, Hsia RC, McClarty G, Rank RG, Bavoil PM and Fraser CM. 2003. Genome sequence of *Chlamydophila caviae* (*Chlamydia psittaci* GPIC): examining the role of niche-specific genes in the evolution of the *Chlamydiaceae*. *Nucleic Acids Res* **31**: 2134-2147.

51. Schöfl G, Voigt A, Litsche K, Sachse K and Saluz HP. 2011. Complete genome sequences of four mammalian isolates of *Chlamydophila psittaci*. *J Bacteriol* **193**: 4258.

52. Voigt A, Schöfl G, Heidrich A, Sachse K and Saluz HP. 2011. Full-length de novo sequence of the *Chlamydophila psittaci* type strain, 6BC. *J Bacteriol* **193**: 2662-2663.

53. Read TD, Joseph SJ, Didelot X, Liang B, Patel L and Dean D. 2013. Comparative analysis of *Chlamydia psittaci* genomes reveals the recent emergence of a pathogenic lineage with a broad host range. *MBio* **4**: e00604-12.

54. Thomson NR, Yeats C, Bell K, Holden MT, Bentley SD, Livingstone M, Cerdeño-Tárraga AM, Harris B, Doggett J, Ormond D, Mungall K, Clarke K, Feltwell T, Hance Z, Sanders M, Quail MA, Price C, Barrell BG, Parkhill J and Longbottom D. 2005. The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. *Genome Res* **15**: 629-640.

55. **Azuma Y, Hirakawa H, Yamashita A, Cai Y, Rahman MA, Suzuki H, Mitaku S, Toh H, Goto S, Murakami T, Sugi K, Hayashi H, Fukushi H, Hattori M, Kuhara S and Shirai M. 2006.** Genome sequence of the cat pathogen, *Chlamydophila felis*. *DNA Res* **13**: 15-23.

56. **Donati M, Huot-Creasy H, Humphrys M, Di Paolo M, Di Francesco A and Myers GS. 2014.** Genome sequence of *Chlamydia suis* MD56, isolated from the conjunctiva of a weaned piglet. *Genome Announc* **2**: e00425-14.

57. **Van Lent S, Piet JR, Beeckman D, van der Ende A, van Nieuwerburgh F, Bavoil P, Myers G, Vanrompay D and Pannekoek Y. 2012.** Full genome sequences of all nine *Chlamydia psittaci* genotype reference strains. *J Bacteriol* **194**: 6930-6931.

58. **Bunk S, Susnea I, Rupp J, Summersgill JT, Maass M, Stegmann W, Schrattenholz A, Wendel A, Przybylski M and Hermann C. 2008.** Immunoproteomic identification and serological responses to novel *Chlamydia pneumoniae* antigens that are associated with persistent *C. pneumoniae* infections. *J Immunol* **180**: 5490-5498.

59. **Li Z, Chen C, Chen D, Wu Y, Zhong Y and Zhong G. 2008.** Characterization of fifty putative inclusion membrane proteins encoded in the *Chlamydia trachomatis* genome. *Infect Immun* **76**: 2746-2757.

60. **Wang J, Zhang Y, Lu C, Lei L, Yu P and Zhong G. 2010.** A genome-wide profiling of the humoral immune response to *Chlamydia trachomatis* infection reveals vaccine candidate antigens expressed in humans. *J Immunol* **185**: 1670-1680.

61. **Cruz-Fisher MI, Cheng C, Sun G, Pal S, Teng A, Molina DM, Kayala MA, Vigil A, Baldi P, Felgner PL, Liang X and de la Maza LM. 2011.** Identification of immunodominant antigens by probing a whole *Chlamydia trachomatis* open reading frame proteome microarray using sera from immunized mice. *Infect Immun* **79**: 246-257.

62. **Rodgers AK, Budrys NM, Gong S, Wang J, Holden A, Schenken RS and Zhong G. 2011.** Genome-wide identification of *Chlamydia trachomatis* antigens associated with tubal factor infertility. *Fertil Steril* **96**: 715-721.

63. **Lu C, Holland MJ, Gong S, Peng B, Bailey RL, Mabey DW, Wu Y and Zhong G. 2012.** Genome-wide identification of *Chlamydia trachomatis* antigens associated with trachomatous trichiasis. *Invest Ophthalmol Vis Sci* **53**: 2551-2559.

64. **Teng A, Cruz-Fisher MI, Cheng C, Pal S, Sun G, Ralli-Jain P, Molina DM, Felgner PL, Liang X and de la Maza LM. 2012.** Proteomic identification of immunodominant chlamydial antigens in a mouse model. *J Proteomics* **77**: 176-186.

65. **Yasui Y, Yanatori I, Kawai Y, Miura K, Suminami Y, Hirota T, Tamari M, Ouchi K and Kishi F. 2012.** Genomic screening for *Chlamydophila pneumoniae*-specific antigens using serum samples from patients with primary infection. *FEMS Microbiol Lett* **329**: 168-176.

66. **Forsbach-Birk V, Foddis C, Simnacher U, Wilkat M, Longbottom D, Walder G, Benesch C, Ganter M, Sachse K and Essig A. 2013.** Profiling antibody responses to infections by *Chlamydia abortus* enables identification of potential virulence factors and candidates for serodiagnosis. *PLoS One* **8**: e80310.

67. **Li D, Vaglenov A, Kim T, Wang C, Gao D and Kaltenboeck B. 2005.** High-yield culture and purification of *Chlamydiaceae* bacteria. *J Microbiol Methods* **61**: 17-24.

68. **Storz J, Shupe JL, Smart RA and Thornley RW.** 1966. Polyarthritis of calves: experimental induction by a psittacosis agent. *Am J Vet Res* **27**: 987.

69. **DeGraves FJ, Gao D, Hehnen HR, Schlapp T and Kaltenboeck B. 2003.** Quantitative detection of *Chlamydia psittaci* and *C. pecorum* by high-sensitivity real-time PCR reveals high prevalence of vaginal infection in cattle. *J Clin Microbiol* **41**: 1726-1729.

70. Ehricht R, Slickers P, Goellner S, Hotzel H and Sachse K. 2006. Optimized DNA microarray assay allows detection and genotyping of single PCR-amplifiable target copies. *Mol Cell Probes* **20**: 60-63.

71. Wang C, van Ginkel FW, Kim T, Li D, Li Y, Dennis JC and Kaltenboeck B. 2008. Temporal delay of peak T-cell immunity determines *Chlamydia pneumoniae* pulmonary disease in mice. *Infect Immun* **76**: 4913-4923.

72. Poudel A, Reid E, Newport-Nielsen E, Chowdhury EU, Rahman KS, Sartin J and Kaltenboeck B. 2014. Endemic asymptomatic *Chlamydia pecorum* genital tract infection reduces fertility in dairy cows. *Proc 13th International Symposium on Human Chlamydial Infections*, pp 258-260. Pacific Grove, CA.

73. Kaltenboeck B, Kousoulas KG and Storz J. 1993. Structures of and allelic diversity and relationships among the major outer membrane protein (*ompA*) genes of the four chlamydial species. *J Bacteriol* **175**: 487-502.

74. Caldwell HD, Stewart S, Johnson S and Taylor H. 1987. Tear and serum antibody response to *Chlamydia trachomatis* antigens during acute chlamydial conjunctivitis in monkeys as determined by immunoblotting. *Infect Immun* **55**: 93-98.

75. 75. Stephens RS, Wagar EA and Schoolnik GK. 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.

76. Campbell LA, Kuo C, Wang S and Grayston JT. 1990. Serological response to *Chlamydia pneumoniae* infection. *J Clin Microbiol* **28**: 1261-1264.

77. Essig A, Simnacher U, Susa M and Marre R. 1999. Analysis of the humoral immune response to *Chlamydia pneumoniae* by immunoblotting and immunoprecipitation. *Clin Diagn Lab Immunol* **6**: 819-825.

78. **Portig I, Goodall JC, Bailey RL and Gaston JSH. 2003.** Characterization of the humoral immune response to *Chlamydia* outer membrane protein 2 in chlamydial infection. *Clin Diagn Lab Immunol* **10**: 103-107.

79. **Crane DD, Carlson JH, Fischer ER, Bavoil P, Hsia R, Tan C, Kuo C and Caldwell HD. 2006.** *Chlamydia trachomatis* polymorphic membrane protein D is a species-common pan-neutralizing antigen. *Proc Natl Acad Sci USA* **103**: 1894-1899.

80. **Hongliang C, Zhou Z, Zhan H, Yanhua Z, Zhongyu L, Yingbiao L, Guozhi D and Yimou W. 2010.** Serodiagnosis of *Chlamydia pneumoniae* infection using three inclusion membrane proteins. *J Clin Lab Anal* **24**: 55-61.

81. **Mohamad YK, Rekiki A, Berri M and Rodolakis A. 2010.** Recombinant 35-kDa inclusion membrane protein IncA as a candidate antigen for serodiagnosis of *Chlamydophila pecorum*. *Vet Microbiol* **143**: 424-428.

82. **Wang J, Zhang Y, Yu P and Zhong G. 2010.** Immunodominant regions of a *Chlamydia trachomatis* type III secretion effector protein, Tarp. Clin. *Vaccine Immunol* **17**: 1371-1376.

83. **Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ. 2009.** Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

84. **Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

85. **Henikoff S and Henikoff JG. 1992.** Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* **89**: 10915-10919.

86. **Dosztányi Z, Csizmok V, Tompa P and Simon I. 2005.** IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**: 3433-3434.

87. **Mirabello C and Pollastri G. 2013.** Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics* **29**: 2056-2058.

88. **Parker JMR, Guo D and Hodges RS. 1986.** New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **25**: 5425-5432.

89. **Geysen HM, Rodda SJ, Mason TJ, Tribbick G and Schoofs PG. 1987.** Strategies for epitope analysis using peptide synthesis. *J Immunol Methods* **102**: 259-274.

90. **Kaltenboeck B, Heard D, DeGraves FJ and Schmeer N. 1997.** Use of synthetic antigens improves detection by enzyme-linked immunosorbent assay of antibodies against abortigenic *Chlamydia psittaci* in ruminants. *J Clin Microbiol* **35**: 2293-2298.

91. **Brack C, Hirama M, Schuller RL and Tonegawa S. 1978.** A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**: 1-14.

92. **Rockey DD and Rosquist JL. 1994.** Protein antigens of *Chlamydia psittaci* present in infected cells but not detected in the infectious elementary body. *Infect Immun* **62**: 106-112.

93. **Andresen H, Zarse K, Grötzinger C, Hollidt JM, Ehrentreich-Förster E, Bier FF and Kreuzer OJ. 2006.** Development of peptide microarrays for epitope mapping of antibodies against the human TSH receptor. *J Immunol Methods* **315**: 11-18.

94. **Cosandey V, Debrot FN, Kaeser J, Marti R, Passeraub P, Petremand J, Prim D and Pfeifer ME. 2012.** Construction of a peptide microarray for auto-anti-body detection. *Chimia Intl J Chemistry* **66**: 803-806.

95. **Elshal MF and McCoy JP. 2006.** Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. *PMC Methods* **38**: 317-323.

96. **Chowdhury F, Williams A and Johnson P. 2009.** Validation and comparison of two multiplex technologies, Luminex® and Mesoscale Discovery, for human cytokine profiling. *J Immunol Methods* **340**: 55-64.

CHAPTER 3

3.1. INTRODUCTION

Knowledge of B-cell epitopes of proteins is essential in many fields of applied biomedical research, such as antibody diagnostics and therapeutics, vaccines, as well basic research. Laboratory methods for identification of such epitopes are time-consuming and labor-intensive. Hence, any reduction in the need for discovery and confirmatory wet-lab research by epitope prediction algorithms is highly desirable. Among *in silico* predictive methods from primary sequence information, epitope prediction algorithms are distinguished for their lack of reliability (1). This underperformance prompted us to examine current approaches to B-cell epitope prediction by use of extensive data on epitopes and confirmed non-epitope regions of the *Chlamydia* spp. proteome, accumulated in research on chlamydial molecular serology (2).

Recent 3D antibody-antigen complex studies (3-7) show that about 15-22aa antigen peptide residues are structurally involved in binding of epitopes to ~17aa residues in antibody complementarity-determining regions (CDRs; paratopes). Among these 15-22 structural epitope residues, about 2-5aa, termed functional residues, contribute most of the total binding energy to antibodies (6). These functional residues lie only in a very small fraction of B-cell epitopes closely spaced to each other embedded among structural residues, representing the classical concept of continuous B-cell epitopes. In the vast majority of B-cell epitopes, functional as well as structural residues are randomly distributed in 15-150aa antigen sequences, essentially representing linear but discontinuous epitopes. Thus, a peptide antigen can effectively bind an antibody only if it

contains the majority of the functional residues, and only a small fraction of short peptides of 4-11aa will contain sufficient functional residues for high affinity binding (6). Therefore, short peptide targets in B-cell epitope mapping and prediction may represent an inherent, unsolvable conundrum when by definition most of these short peptides, even from proven dominant epitopes, will fail to bind antibodies strongly, with many false negative (non-epitope) results.

Mammalian immune systems can be forced to generate antibodies against virtually any molecule, regardless of antigen origin, by using excessive amounts of adjuvants and antigens. However, the antibody response did evolve in response to infections that generate much lower antigen exposure, thus antibodies may be preferentially directed towards proteins and peptide regions with optimal epitope properties. Antibody formation during an immune response to any given epitope is inherently stochastic due to the random availability of a cognate B-cell receptor within the large pool of circulating B-cells, all with different B-cell receptors generated by recombination of the immunoglobulin gene (8). Another level of stochasticity in the antibody response to any given protein is the exposure of a protein to the immune system. Wang *et al*. (2010) report that only 4.2% of about 900 *Chlamydia trachomatis* (Ctr) proteins induce natural antibody responses in ≥ 40% of human hosts (9). Therefore, any peptide of the remaining 95.8 % non-immunodominant proteins is unlikely to elicit antibodies, regardless of its B-cell epitope properties. Hence, for accurate evaluation of epitope prediction methods, epitope/non-epitope data should be derived from testing of known immunodominant proteins, with multiple rather than single sera to account for the stochasticity of the antibody response.

B-cell epitope prediction has been first based on various properties of individual amino acids (AA) such as hydrophilicity, hydrophobicity, solvent accessibility, flexibility, or beta-turn propensity, and combinations thereof (10-16). However, even the best combinations of AA propensity scales performed only marginally better than random sequence selection (1). With the availability of B-cell epitope databases, antigenicity scales (17,18) and machine learning approaches (19-24) have been attempted, and improved prediction accuracy has been reported. Nevertheless, due to epitope redundancy (20), the predictive power may have been overestimated because these algorithms performed poorly on independent data (24). Therefore, B-cell epitope prediction algorithms must be evaluated on independent datasets that had not been used to train/develop the algorithms/scales.

The ever increasing number of solved 3D protein structures has allowed development and testing of numerous complex algorithms for prediction of physicochemical and structural properties of proteins directly from the AA sequences. Among these properties (scales), disorder tendency which describes protein regions without defined 3D structure that are inherently flexible, hydrophilic, solvent accessible, and thermally mobile (high B-factor) (25,26). Incidentally, all of these properties are shared with B-cell epitopes (16); thus, protein disorder tendency is a prime candidate scale for B-cell epitope prediction due to its multifaceted properties (27).

This investigation is an extension of a comprehensive study that identified immunodominant B-cell epitopes of *Chlamydia* spp. (2). After encountering numerous failures of *in silico* B-cell epitope prediction, we used the first principles established above to analyze the shortcomings of B-cell prediction algorithms. Using pools of hyper-

immune mouse sera we determined epitope/non-epitope regions of immunodominant *Chlamydia* spp. proteins by use of long, 16-40aa peptide antigens. These data created epitope/non-epitope datasets for accurate testing of numerous B-cell epitope prediction or AA/protein property algorithms/scales (henceforth termed scales). Subsequent testing revealed that public datasets were biased towards short epitope/non-epitope antigens, and removal of these short antigens dramatically increased prediction accuracy of most scales. We show that in general machine learning methods cannot predict epitopes with high accuracy; rather, many scales designed for prediction of protein properties, particularly disorder tendency, identify B-cell epitopes with better accuracy.

3.2 METHODS

B-cell epitope peptide reactivity with anti-chlamydial hyper-immune sera. Hyper-immune sera were raised in mice as described (2). Briefly, 9-50 mice were 3× challenged with high but non-lethal intranasal chlamydial inocula, to mimic antibody production following natural infections. Bovine sera used were obtained from animals with PCR-confirmed natural *Chlamydia* spp. infection (2). Peptide antigens were chemically synthesized with N-terminal biotin, captured onto streptavidin coated microtiter plates, and incubated with hyperimmune sera. Primary antibodies were detected with horseradish peroxidase-conjugated secondary antibodies in chemiluminescent ELISA (2).

B-cell epitope/non-epitope datasets. These datasets are described briefly below, a detailed description is in Table 3.S2, and sequences are provided in the Supplementary B-cell epitope Dataset S10.

Concatenated epitope/non-epitope virtual proteins. Epitopes and non-epitopes of the Lbtope_Fixed_non_redundant and Lbtope_Confirm datasets (24) were grouped based on sequence length. Concatenated polyproteins of the sequences of each length group were constructed by randomly combining all sequences. A similar polyprotein was constructed of the epitopes of the fbcpred.pos.nr80 dataset (21). This dataset does not contain experimentally validated non-epitopes, instead random Swiss-Prot peptides of the BCP12, BCP14 or BCP18 datasets were used (20).

Concatenated virtual proteins of 50aa-extended epitopes/non-epitopes embedded in random sequences. All 16-20aa epitopes of Lbtope_Confirm and fbcpred.pos.nr80 datasets were extended symmetrically to 50aa with source protein sequences. These fragments were interspersed with random 150aa Swiss-Prot sequences into a concatenated virtual polyprotein. Similar polyproteins were assembled from all 16-33aa non-epitopes of the Lbtope_Confirm dataset, and all epitopes/non-epitopes of the Chl_18Prot and Chl_43Prot datasets. For assembly of additional non-epitope datasets, random 50aa peptides of the Bcpreds epitope source proteins (Bcpreds_Prot), Swiss-Prot proteins, and the *C. trachomatis* proteome were similarly linked with 150aa interspacing sequences into concatenated polyproteins.

B-cell epitope/non-epitope annotation of individual Chlamydia *spp. proteins*. All residues of 18 immunodominant proteins of *Chlamydia* spp. were annotated in the Chl_18Prot dataset as Pos (positive, epitope), Neg (negative, non-epitope) or NT (not tested, unknown epitope status). The annotation is based on the reactivity of 16-20aa peptide antigens with murine and/or bovine sera. For peptide datasets, 10aa-spaced peptides of these proteins were used.

B-cell epitope prediction with protein property algorithms. *Computation of AA residue scores for physicochemical, structural, and evolutionary protein properties*. Website-based freeware algorithms/scales (Appendices S8, S9; 10-15,17-20,22-24,28-31,33-55) for protein properties were used to calculate individual residue scores for AA sequences of individual or polyproteins. Moving window scores were assigned to the center residue of the particular window. When required, missing scores for N- and C-

terminal residues were inserted using scores of the adjacent residues. If algorithms/scales did not provide an output score for internal residues/windows, the minimum score of this protein was inserted. The polymorphism score was calculated by inverting the multiple sequence conservation score of the AACon algorithm in the Jalview freeware (35).

Comparison of receiver operating characteristic (ROC) curves of protein property scales for B-cell epitope prediction. Bimodal epitope/non-epitope classification was achieved by F test classification based on the linear predictor variable in discriminant analysis with the software package JMP Pro 11 (SAS Institute Inc., Cary, North Carolina, USA). This software was also used to construct ROC curves and calculation of area under the ROC curve (AUC) for ranking of protein property scales for B-cell epitope prediction. Data formatting was performed in Microsoft Office Excel 2013, and all additional statistical analyses were performed by the Statistica 7.1 software package (Statsoft, Tulsa, Oklahoma, USA). Differences between means of peptide reactivities and/or background were analyzed by one-tailed paired Student t-test, and p values ≤ 0.05 were considered significant. The significance of differences between B-cell epitope prediction scales was tested by one-tailed paired Student t-tests of (i) the AUC values obtained for multiple datasets; or of (ii) the AUC values obtained for individual proteins; or, for direct comparison of ROC curves, of (iii) the sensitivity values sampled in 0.05 increments for specificities between 0.05-0.85, and in 0.02 increments for specificities from 0.90 to 0.98.

3.4 RESULTS

Antibody binding increases with length of peptide antigens. Within a peptide B-cell epitope, 15-22 residues are typically structurally involved in antibody-binding (3-7). Sivalingam and Shepherd (2012) reasoned that clustering or random distribution of the structural residues would determine the length of peptide antigens required for antibody binding (6). In this study, we tested length-dependent peptide antigen reactivity for previously identified epitope regions of the chlamydial outer membrane protein A (OmpA) (2). Seven-12 aa peptide antigens invariably produced lower ELISA signals than longer ones (Fig. 3.1A). Occasionally, extensive elongation of peptide antigens may mask structural residues and reduce the signal relative to a slightly shorter peptide antigen of optimum length (Fig. 3.1A; 32aa vs. 24aa Cpe peptides).

To quantify the effect of peptide length on antibody binding, peptide antigens of different lengths from 17 epitope regions of OmpA and inclusion membrane A (IncA) proteins of *Chlamydia* spp. were tested (Fig. 3.1B). Compared to short 7-12aa peptides, intermediate 16aa peptides produced 1.8-fold ELISA signal intensity, and long 24-40aa peptides produced 4.1-fold signal intensity ($P<10^{-2}$, one tailed Student *t*-test, relative log-transformed signal). Importantly, the main 14.3-fold reactivity increase was achieved by elongation of the 13 lowest-reactive short peptides from 7-12aa to 24-40aa ($P<10^{-4}$), while elongation of the 13 lowest-reactive intermediate 16aa peptides to 24-40aa produced a moderate 3.3-fold increase ($P<10^{-4}$).

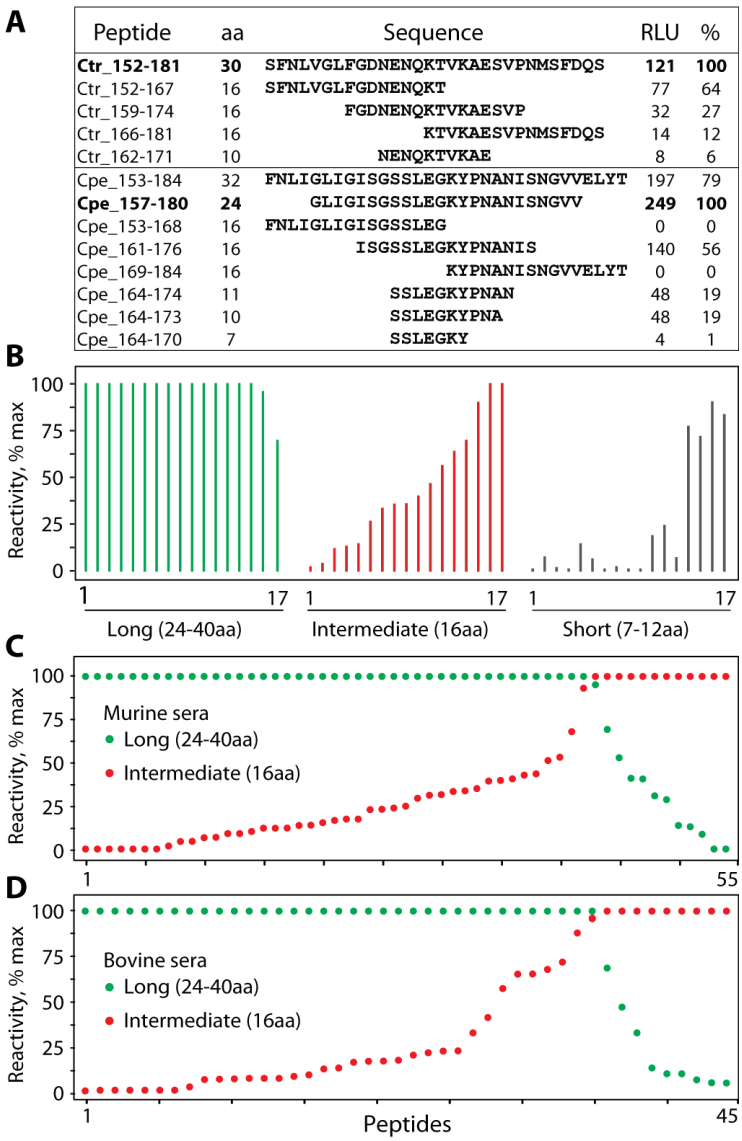


Fig. 3.1. Peptide reactivity increases with length.

***(A) Elongation of peptides around the center of two epitopes increases the ELISA signal*. aa, amino acids; RLU, relative light units/sec, mean of 6 repeats; %, percent signal of maximally reactive peptide; Ctr, *C. trachomatis*; numbers indicate peptide position on the OmpA protein; Cpe; *C. pecorum*.**

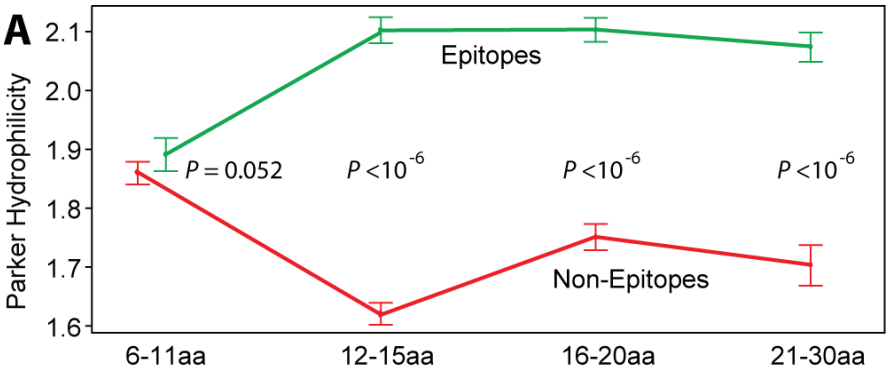
***(B) Relative signal from 17 epitopes in dependence on peptide antigen length.* Epitope signals are represented by vertical lines in the same order for long, intermediate, and short peptides.**

(C) Relative reactivity with murine antisera of corresponding long and intermediate peptides of 55 epitopes.

(D) Relative reactivity with bovine antisera of corresponding long and intermediate peptides of 45 epitopes.

To identify optimum peptide antigen lengths, we tested the central 16aa and 24-40aa peptide antigens of 55 unique epitopes on 28 *Chlamydia* spp. proteins with pooled mouse sera (Fig. 3.1C). As observed before, ~20% of elongated peptides produced a reduced signal, presumably due to epitope masking. However, long 24-40aa peptides produced on average a 2.1-fold higher signal than the corresponding 16aa peptides ($P<10^{-4}$). The reactivity of the 28 lowest-reactive 16aa peptides increased 9.1-fold for the respective long 24-40aa peptides ($P<10^{-4}$). To confirm the host-independence of length-dependent peptide reactivity, another set of chlamydial peptides yielded equivalent results with bovine sera (Fig. 3.1D).

Evaluation of B-cell epitope prediction algorithms confounded by overrepresentation of short false-negative epitopes in public datasets. Many investigators typically use short peptide antigens of 4-11aa for epitope mapping, with results added to public reference databases that serve as generally accepted test platforms for B-cell epitope prediction algorithms. These datasets may therefore be biased towards short epitopes and many false-negative epitope determinations due to the marginal antibody binding of short peptides. This may explain the poor, close to random, epitope prediction accuracy (1) that most epitope prediction scales show in practical application, even if they scored highly in evaluation with public datasets. We hypothesized that removing short epitope/non-epitope sequences from public datasets would allow correct performance ranking of B-cell epitope prediction scales. To test this hypothesis, we used the "Lbtope_Variable_non_redundant" dataset with 8,011 B-cell epitopes and 10,868 non-epitopes, retrieved by Singh *et al*., (2013) from experimentally



B

Length (aa)	AUC			
	Hydrophilicity (Parker)	Hydrophobicity (Miyazawa)	Bepipred	Disorder (IUPred-L)
6 - 11***	0.50	0.50	0.53	0.50
12 - 15	0.59	0.61	0.60	0.65
16 - 20	0.58	0.59	0.58	0.64
21 - 30	0.58	0.60	0.60	0.66

Fig. 3.2. B-cell epitope prediction score and performance (AUC in ROC analysis) in dependence on the length of epitope peptides. Epitopes and non-epitopes are grouped based on length in the Lbtope_Variable_non_redundant dataset.

(A) Epitope length-dependent hydrophilicity (± 95% CI). **Hydrophilicity (Parker), hydrophilicity (11) scores obtained by using default settings in the ProtScale tools of the ExPASy server (28).**

(B) Epitope length-dependent prediction performance of different prediction scales. *****, $P < 10^{-6}$ for comparison of any scale for 6-11aa to any other scale for longer epitopes. Hydrophobicity (Miyazawa (29), a ProtScale (28) for hydrophobicity; Bepipred, a hidden Markov model combined with the Parker hydrophilicity scale (19); IUPred-L, an algorithm for protein disorder tendency (30).**

validated epitopes as well as non-epitopes from the Immune Epitope Database, IEDB (24). Among the 6-11aa, 12-15aa, 16-20aa, and 21-30aa sequences of the Lbtope_Variable_non_redundant epitope/non-epitope dataset, epitope/non-epitope ratios are 0.47, 0.67, 1.08, and 1.86, respectively. Importantly, 5-10aa non-epitopes comprise ~50%, and 5-16aa non-epitopes ~80% of all non-epitopes deposited in the parent IEDB (24). For analysis shown in Fig. 3.2 epitopes and non-epitopes were grouped by length, and all sequences of each length group were randomly concatenated into a single virtual protein. Hydrophilicity of all consecutive non-overlapping 20aa peptide windows of each concatenate was predicted by use of the Parker ProtScale in ExPASy (11,28), a generally accepted parameter used in B-cell epitope prediction. Results in Fig 3.2A show that in the 12-15aa, 16-20aa, and 21-30aa length concatenates, the hydrophilicity scores of epitope and non-epitope virtual proteins differ highly significantly ($P<10^{-6}$, Student's *t*-test). In contrast, the hydrophilicity of epitope and non-epitope virtual proteins is not different for the 6-11aa length concatenates ($P = 0.052$). Thus, for peptides longer than 11aa, the hydrophilicity scale discriminates between epitopes and non-epitopes, but not for shorter peptides.

Similarly, with ~0.50 AUC in ROC analyses, additional scales in Fig. 3.2B show random distribution of epitope vs. non-epitope prediction for the 6-11aa concatenates, but highly significantly increased prediction accuracy for concatenates of peptides longer than 11 amino acids, indicating significant discrimination between epitopes and non-epitopes (Fig. 3.2B). Analysis of the "Lbtope_Confirm" subset, composed of epitopes/non-epitopes that were at least twice independently experimentally validated, confirmed the result of the Lbtope_Variable_non_redundant dataset (Fig. 3.S1).

Prediction accuracy of B-cell epitope algorithms depends on epitope/non-epitope discrimination of test datasets. Ideal algorithms/scales for prediction of B-cell epitopes should discriminate known epitopes from experimentally validated non-epitopes, and identify epitopes within complete source proteins and proteomes. Since prediction accuracy should ideally be validated with multiple datasets, we evaluated the prediction performance by use of 4 positive datasets of experimentally validated B-cell epitopes and 5 negative datasets of experimentally validated non-epitopes or of random peptides from proteomes (Tables 3.1, 3.S2). All 16-20aa epitopes/non-epitopes sequences of the datasets were centered within their 50aa source protein sequences, and these 50aa sequences were randomly concatenated into a single virtual protein, interspersed with random 150aa sequences from the Swiss-Prot database (Table 3.S2, Dataset 3.S10). For evaluation of B-cell epitope prediction, we used the original score of each algorithm with default settings for each amino acid residue, thus each 16-20aa epitope/non-epitope sequence received individual scores for the 20 central residues. Correct or incorrect epitope prediction of all residues was evaluated by AUC in ROC analysis.

In Table 3.1, the column for each scale indicates epitope versus non-epitope discrimination (AUC in ROC analysis) of the scale for the each compared combination of positive and negative datasets. The average AUC column, next to the rightmost column, indicates epitope/non-epitope discrimination averaged over all tested scales, and therefore ranks the combined discrimination in both positive and negative datasets. The 4 positive datasets can be ranked by their AUC in comparison to the negative Swiss-Prot or Ctr-

Table 3.1. B-cell epitope prediction accuracy (AUC of ROC curves) in dependence on the evaluation dataset.

Datasets[a]		AUC[b]											
Positive	Negative	Antigenicity[c]	Lbtope[d]	Accessibility[e]	Beta-turn[f]	Bcepred[g]	Flexibility[h]	Hydrophilicity (Parker)	Hydrophobicity (Miyazawa)	Bepipred	Disorder (IUPred-L)	**Average AUC**	**Rank[i]**
1. fBcpreds	1. Bcpreds_Prot	0.47	0.51	0.52	0.52	0.54	0.54	0.55	0.56	0.57	0.58[k]	**0.55**	**19**
	2. Swiss-Prot	0.44	0.56	0.59	0.55	0.60	0.58	0.62	0.63	0.63	0.67[k]	**0.61**	**13**
	3. Ctr_Proteome	0.38	0.50	0.59	0.57	0.62	0.59	0.66	0.66	0.66	0.72[k]	**0.63**	**11**
2. Lbtope_ Confirm	4. Lbtope_Confirm	0.47	0.97[j]	0.59	0.57	0.62	0.59	0.61	0.61	0.62	0.66[k]	**0.61**	**13**
	2. Swiss-Prot	0.44	0.84[j]	0.64	0.58	0.66	0.62	0.66	0.68	0.67	0.71[k]	**0.65**	**9**
	3. Ctr_Proteome	0.38	0.81[j]	0.63	0.61	0.67	0.63	0.69	0.70	0.70	0.75[k]	**0.67**	**7**
3. Chl_43Prot (50aa apart)	5. Chl_18Prot	0.39	0.57	0.60	0.64	0.62	0.64	0.67	0.67	0.71	0.73[k]	**0.66**	**8**
	2. Swiss-Prot	0.43	0.58	0.60	0.65	0.64	0.68	0.73	0.73	0.77	0.80[k]	**0.70**	**4**
	3. Ctr_Proteome	0.36	0.55	0.60	0.68	0.66	0.69	0.77	0.75	0.79	0.83[k]	**0.72**	**2**
4. Chl_18Prot (≥10aa apart)	5. Chl_18Prot	0.39	0.63	0.63	0.64	0.64	0.63	0.66	0.69	0.72	0.80[k]	**0.68**	**6**
	2. Swiss-Prot	0.43	0.64	0.62	0.65	0.66	0.68	0.73	0.74	0.78	0.87[k]	**0.71**	**3**
	3. Ctr_Proteome	0.37	0.60	0.62	0.67	0.67	0.68	0.76	0.76	0.80	0.88[k]	**0.73**	**1**
Average AUC		**0.41**	**0.57**	**0.60**	**0.61**	**0.63**	**0.63**	**0.68**	**0.68**	**0.70**	**0.75**[k]		
Rank				**16**	**15**	**13**	**13**	**8**	**8**	**6**	**1**		

[a], datasets with experimentally identified epitopes/non-epitopes or random peptides (Table 3.S2).

[b], AUC data are shown for the best performing scale of each category as defined for all scales tested (Appendices S8, S9).

[c], Antigenicity scale (17), IEDB tool for antibody epitope prediction.

[d], Support vector machine model (SVM) trained on the Lbtope_Confirm dataset (24).

[e], Surface accessibility scale (13), IEDB tool for epitope prediction.

[f], Beta-turn scale (31), IEDB tool for epitope prediction.

[g], Average of 7 propensity scales for epitope prediction, (15).

[h], Flexibility scale (12), IEDB tool for epitope prediction.

[i], quality of scales or datasets was ranked by AUC, with rank number determined by 1 for the highest AUC and addition of 1 for each 0.01 AUC reduction; Antigenicity and Lbtope scales were excluded from ranking because Antigenicity is a negative predictor and Lbtope was trained on the analysis dataset.

[j]; not applicable for quality ranking because the datasets served to train the Lbtope SVM.

[k], highest AUC in the compared dataset.

Proteome datasets, clearly showing significantly higher discrimination for both chlamydial datasets than for the fBcpreds and Lbtope_Confirm datasets ($P \leq 0.02$, one tailed paired Student's t-test).

The average AUC row, next to the bottom row in Table 3.1, indicates epitope/non-epitope discrimination averaged over the 12 tested pairs of datasets, and shows that disorder tendency discriminated best of all algorithms tested (AUC = 0.75 for IUPred-L (30), highly significantly better than Bepipred (AUC = 0.70), the next-best algorithm in the 12 AUC comparisons of positive and negative datasets ($P < 10^{-3}$, one tailed paired Student's t-test). Since the machine learning Lbtope algorithm was trained on the Lbtope_Confirm dataset, it performed extremely well in this dataset (AUC = 0.97, 0.84, and 0.81), but very poorly, close to randomization, in all other datasets (average AUC = 0.57). In contrast, the protein disorder scale IUPred-L consistently discriminated best (Table 3.1), with AUCs depending on the datasets (0.58-0.88).

Protein disorder most accurately predicts B-cell epitopes. Most B-cell prediction algorithms/scales score the context-dependent epitope likelihood for each amino acid residue by averaging the adjacent 2-4 residues. To simulate the wider sequence context in which B-cell prediction operates, we asked the question if scoring complete peptides improved prediction accuracy, and if so, what the peptide length dependency of such an improvement was. Results in Tables 3.S3 and 3.S4 show that prediction for hydrophobicity, hydrophilicity, flexibility, Bcpreds, or Bepipred improves substantially by peptide scoring rather than individual AA scoring, and the optimum peptide length is 25aa.

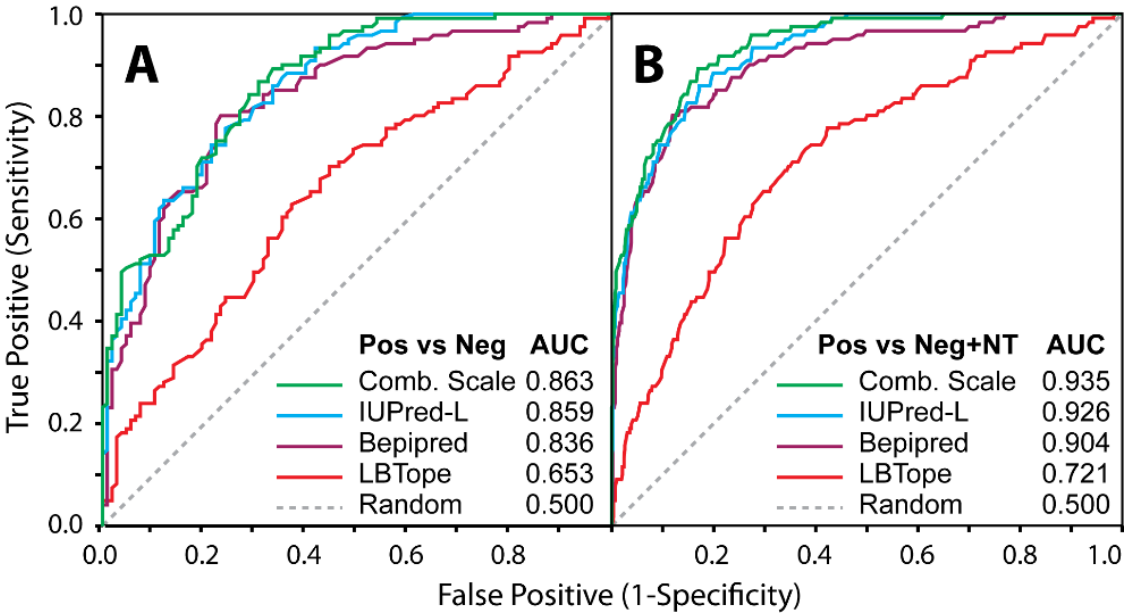


Fig. 3.3. Comparison of ROC curves for prediction of 25aa epitopes. Plots of epitope-positive rate versus false positive rate for the 18 chlamydial protein dataset are shown.

(A) Prediction of epitopes from confirmed non-epitopes (25aa epitopes/non-epitopes spaced 10aa). The combined scale represents the arithmetic mean of two disorder scales, IUPred-L (30) & VSL2B (33), and one solvent accessibility scale, ASA, Spine-X (34).

(B) Prediction of epitopes from the total remaining proteins (non-epitope plus non-tested regions). In both datasets (Fig 3.4A and 3.4B), the combined scale and the single disorder (IUPred-L) scale performed best (highest sensitivity at given specificity or vice versa), significantly better than Bepipred or LBTope (one tailed paired Student's *t*-test, $P < 10^{-4}$).

In analyses shown in Tables 3.1 and 3.S3, most epitope/non-epitope sequences were derived from public datasets of variable and largely unknown discrimination accuracy. For maximum accuracy, we therefore selected the 18 chlamydial protein dataset (*Chl*_18Prot; S4-S9) with extensively validated epitopes as well as non-epitopes on each protein, all identified in a single investigation (2). To account for overall differences of the proteins, all scores of each protein were standardized to a mean = 0 and SD = 1. Using this optimized *Chl*_18Prot dataset, 151 primary scales for B-cell epitope prediction, and 126 combinations of the primary scales were evaluated (Appendices S8, S9).

For epitope/non-epitope discrimination in the ROC curve in Fig 3.3A, IUPred-L sensitivity at given specificity (or specificity at given sensitivity) is higher than that of any other primary scale ($P<10^{-4}$, one tailed paired Student's t-test). Similarly, when epitopes were discriminated from the complete remaining proteins in Fig. 3.3B, IUPred-L performed significantly better than any other primary scale, with 87% specificity at 80% sensitivity and 86% accuracy (Fig. 3.S5). Interestingly, in Fig 3.3B all scales performed at higher level as compared to Fig. 3.3A. We ascribe this increased prediction accuracy to a pre-selection bias in the non-epitope dataset because non-epitopes had scored high in epitope prediction and failed to react in testing, but in this process created a high-scoring negative dataset that confounded evaluation (Fig. 3.S6). In final testing of the overall prediction approach applied to the 18 individual proteins of the *Chl*_18Prot dataset, IUPred-L also best discriminated individual epitope residues from the whole remaining protein (average AUC of IUPred-L = 0.91, minimum = 0.74, maximum = 1.00, SD = 0.08; Table 3.2).

Table 3.2. Epitope prediction accuracy (AUC) averaged for individual proteins of the 18-chlamydial protein dataset[a].

Scale[b]	Lbtope	Coils	Accessibility	Polymorphism	Miyazawa	Bepipred	VSL2B	IUPred-L	Combined	AUC ± SD	Rank
Lbtope		A[c]	C	C	C	C	C	C	C	0.75 ± 0.13	18
Coils (Spine-X)			-	-	-	A	A	B	B	0.82 ± 0.15	11
Accessibility (Spine-X)				-	-	-	-	A	A	0.86 ± 0.08	7
Polymorphism					-	-	-	A	B	0.87 ± 0.09	6
Hydrophobicity (Miyazawa)						-	-	B	C	0.88 ± 0.07	5
Bepipred							-	A	A	0.89 ± 0.08	4
Disorder (VSL2B)								-	A	0.89 ± 0.12	4
Disorder (IUPred-L)									-	0.91 ± 0.08	2
Combined										0.92 ± 0.07	1

[a], original scores obtained with default options for the algorithm/scale were smoothed by a sliding window method in which the score for each residue was averaged for the adjacent ± 12 residues (25 aa window). Smoothed scores of residues were standardized (mean = 0 and SD = 1) for each of the 18 chlamydial proteins and discrimination of epitope residues from remaining total residues was tested for each of the 18 proteins individually.

[b], Polymorphism, sequence divergence in multiple sequence alignment, calculated by inverting the conservation score of AACon in the Jalview freeware (35); Coils (Spine-X), coils predicted in secondary structure (36).

[c], A, $0.05 \geq P > 0.01$; B, $0.01 \geq P > 0.001$; C, $0.001 > P$.

Proposed B-cell epitope prediction. To improve B-cell epitope prediction, investigators frequently combine scales (16). To test this concept, we evaluated 126 combined scales that were derived by linear combination of 2 to 14 primary scales. Compared to IUPred-L (AUC = 0.91, Table 3.2), the best combined scale (AUC = 0.92, IUPred-L & VSL2B disorder scales and Spine-X relative solvent accessibility scale) provided only a statistically not significant 0.8% AUC increase and 26% variance reduction, due to the multi-collinearity of epitope prediction scales.

Fig. 3.S7 demonstrates the application of the previous findings for B-cell epitope prediction in an actual example for which we generated epitope scanning data of the complete chlamydial protein IncA. First, default scores for the complete *C. pecorum* IncA protein sequence were obtained from freeware webservers for the 4 primary scales IUPred-L, VSL2B, ASA Spine-X, and Miyazawa hydrophobicity (Fig 3.S7 A, B, C). The 25aa moving average scores were standardized, and hydrophobicity was inverted (Fig 3.S7 D), and the combined score of the 4 primary scales was plotted above the actual reactivities obtained in the peptide epitope scan (Fig 3.S7 E). These plots clearly demonstrate that for practical B-cell prediction, the 25aa moving average of a default IUPred-L protein disorder plot is sufficient and optimal for identification of B-cell epitope regions from which the best suited 16-30aa peptides should be selected for actual testing.

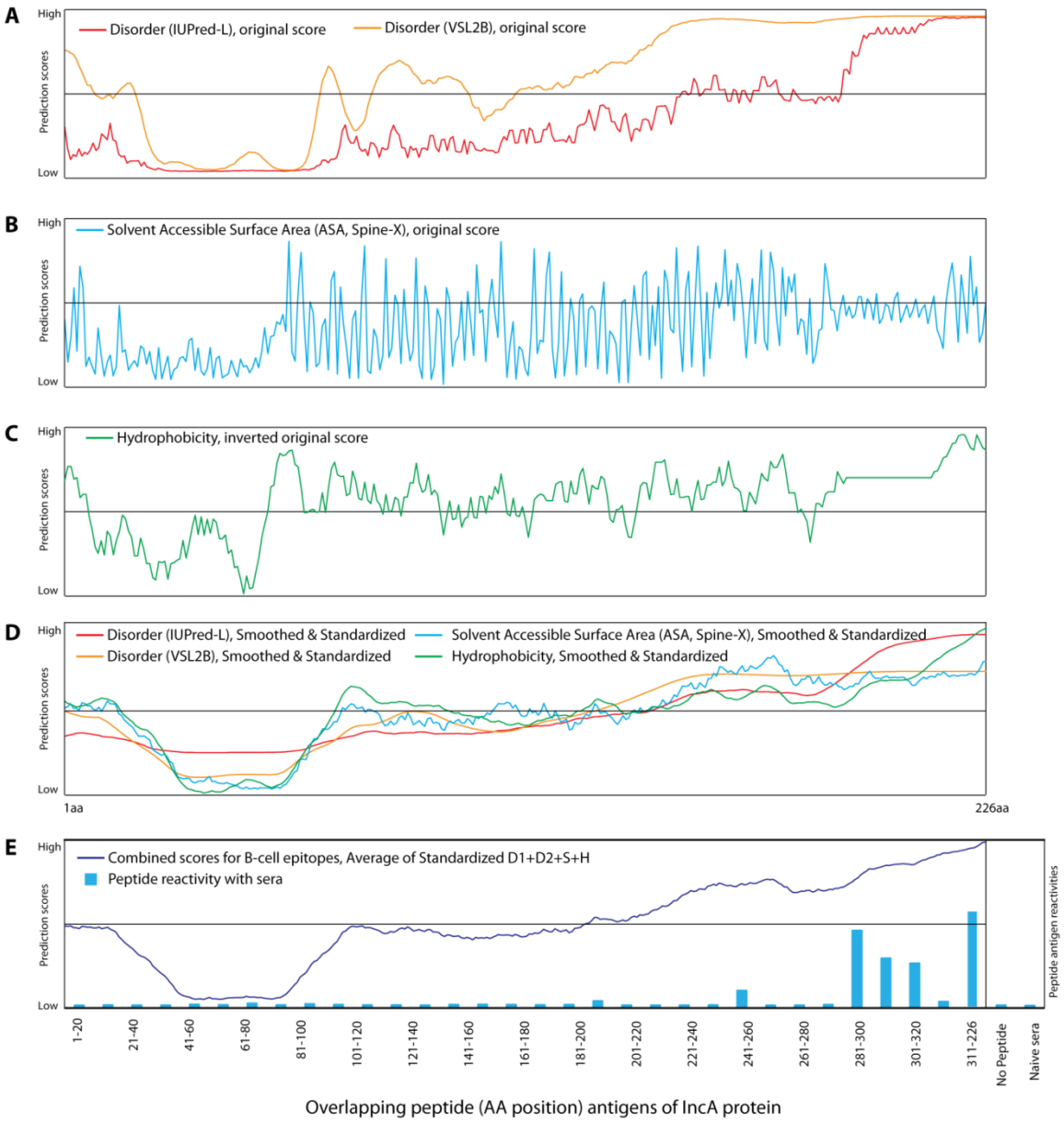


Fig. 3.4. Optimal B-cell epitope prediction. (A) Original disorder scores plotted against IncA protein residues. (B) Original solvent accessibility scores. (C) Original hydrophilicity scores. (D) Moving average score. (E) Combined score.

(*A*) *Original disorder scores plotted against IncA protein residues.* The web-servers of disorder scales IUPred-L (D1, http://iupred.enzim.hu/; (30) and VSL2B (D2, http://www.disprot.org/metapredictor.php, (33) were used for obtaining disorder scores using the IncA protein sequence of *C. pecorum*. The scores were transferred to an Excel Spreadsheet and plotted against IncA residue number (Fig 3.S7A).

(*B*) *Original solvent accessibility scores.* Scores were obtained using protein sequence as input in the ASA_Spine-X webserver (http://sparks-lab.org/SPINE-X/, (34), which predicts solvent

accessible surface area/solvent accessibility (S) of protein residues. Alternative scales for solvent accessibility, as such PaleAle 4.0: (http://distillf.ucd.ie/porterpaleale/, (37), can also be used (Fig 3.S5), but the output score of Spine-X is most convenient because it can most easily be transferred to an Excel Spreadsheet.

(*C*) *Original hydrophilicity scores.* Hydrophobicity (28,29) scores were obtained for protein sequences with default options in the ExPASy web-server (http://web.expasy.org/protscale/; (28). The default 9aa windows scores of Protscale were assigned to the central AA residues of the window, and the missing scores for C- and N-terminal 4 residues were inserted as the 4 immediately adjacent scores. To convert this hydrophobicity score into a hydrophilicity (H) score, original scores were multiplied by -1.

Alternatively, the Parker hydrophilicity scales at Bcepred (http://www.imtech.res.in/raghava/bcepred/bcepred_submission.html), ExPASy (http://web.expasy.org/protscale/), or IEDB (http://tools.immuneepitope.org/bcell/) can also be used (Appendices S8 and S9), which performed equally well when combined with IUPred-L disorder. Since the Parker scale is a hydrophilicity scale, score inversion is not required.

(*D*) *Moving average score*. A moving 25aa window average was determined for the original scores obtained with predictor default settings for window size. The smoothing scores are the arithmetic mean scores over ±12aa adjacent regions for each residue, thus the smoothing window size is 25aa (12aa+1aa+12aa). Scores of each scale were standardized with mean 0 and standard deviation 1. The standardized scores of 4 scales were plotted against AA residues of the IncA protein.

(*E*) *Combined score.* Combined scores for each IncA residue were calculated as the mean of score of 4 different scales (D1, D2, S and H). A total of 32 peptides, 20aa long and 10aa overlapping, were tested with mouse sera. Reactivity of these peptides with mouse sera are shown as blue bars. Peptides from the IncA C-terminus showed strong reactivity, but peptides from the remaining IncA regions produced only background level signal. This C-terminal region has high prediction scores for all 4 scales shown in Fig 3.S7D. The combined score was also highest in the C-terminal region (Fig 3.S7E). For illustration, we show here the combined score of 4 scales. Sequence polymorphism also marginally improves B-cell epitope prediction accuracy, but requires extensive effort to derive the polymorphism scores from multiple sequence alignment. Based on our analyses of 126 weighted or unweighted combinations of primary scales, we recommend to use the best combination as shown in Fig. 3.S5, or Appendices S8 and S9. The best combined scales typically are disorder IUPred-L together with relative solvent accessibility such as Spine-X or PaleAle4.0.

3.4 DISCUSSION

The results of this study suggest strategies for B-cell epitope identification that deviate from current approaches that many investigators use. In the typical initial *in silico* prediction step, it is more important to identify protein regions that harbor B-cell epitopes with high probability than to immediately focus on peptide antigens of specified length. Reactive regions can be identified with very high predictive accuracy using the 25aa moving average of the IUPred-L disorder score of a protein (87% specificity at 80% sensitivity, 86% accuracy; Fig. 3.S5). Next, these high probability epitope regions should be confirmed with 20-30aa long peptide antigens using pooled antisera. Fine mapping of highly reactive regions with overlapping 16aa peptides using the individually reactive antisera of the pool identifies regions with several functional AA residues embedded in structural epitope residues (6). Further reduction in peptide antigen length entails mapping with very short 6-12aa peptides. Success at this stage relies on stochastic identification (Figs. 3.2, 3.S1) of closely-spaced randomly-distributed functional residues that maximally contribute to antibody binding. Antibody binding of such short peptides is, however typically low (Fig. 3.1) because antigens of less than 16aa are too small for binding to the complete CDR of an antibody (3-7).

This approach is derived from the conclusive evidence that short 7-12 peptide antigens of confirmed *Chlamydia* spp. epitopes bind antibodies poorly (Fig. 3.1), and therefore many of these epitopes would be falsely classified as non-epitopes if they were identified by short-peptide mapping. The poor reactivity of short peptide antigens combined with data in Figs. 3.2 and 3.S1 strongly suggest that many of the short non-epitopes in public B-cell epitope datasets are likely to be actual epitopes. An over-

representation of false non-epitopes in the public IEBD points towards a fundamental shortcoming that may impede research in B-cell epitope prediction. The target sequences, and thus B-cell prediction algorithms developed, may inherently be suboptimal since most investigators draw training and test datasets from the public databases such as IEBD for development of B-cell epitope prediction algorithms.

C. trachomatis OmpA serovar-specific peptide serology has used 6-10aa peptides (56-60), with inconsistent results. In our study, we observed strong, but completely serovar-specific antibody reactivity by use of ≥16aa peptide antigens. However, in addition to the 3-6 central polymorphic serovar-determinant AA residues, inclusion of conserved adjacent residues shared among chlamydial species was required. Since the inclusion of these shared residues did not induce cross-reactivity, we conclude that these N- and C-terminal AA stabilize peptide-antibody binding while the central polymorphic AA are the actual functional epitope residues. In some instances, extensive elongation of peptide antigens reduces reactivity (Fig. 3.1), presumably by masking functional residues.

In this study, we also defined a peptide antigen length of 16-30 AA residues as optimally suited for B-cell epitope assays, and created high quality B-cell epitope datasets. For actual B-cell epitope prediction, we show that the IUPred-L protein disorder scale consistently performs best, in chlamydial as well as public datasets. However, if a 25aa moving average score is used, several other protein property scales such as hydrophilicity (Parker), hydrophobicity (Miyazawa), solvent accessibility (Spine-X), or Bepipred perform close to the IUPred-L disorder scale, and these scales may also provide improvement in a combined scale (92% specificity at 80% sensitivity, 90% accuracy in Fig 3.S5). In fact, 20-30aa moving window scores of these scales actually reflect disorder

tendency (Globplot-2, (38). Therefore, our data show that scores of 20-30aa regions are optimal for B-cell epitope prediction (Tables 3.S3, 3.S4), consistent with the higher antibody binding of 16-30aa peptide antigens (Fig 3.1).

B-cell epitopes have historically been recognized as hydrophilic (10,11), flexible (12), mobile (high B factor (61,62), surface exposed or solvent accessible (3,5,13,63), enriched with beta-turns (14) or coils/loops (3-4,7), and highly sequence-polymorphic (3-5). Our exhaustive testing of protein property scales in multiple datasets has shown that these properties remain excellent predictors of B-cell epitopes if used as moving averages of sufficiently long sequence windows. However, the critical property that in our hands best characterizes B-cell epitopes is protein disorder tendency (Appendices S8, S9). This quantitative characteristic of protein sequences has emerged over the last two decades, and it synthesizes these properties into a single descriptor that represents protein regions without defined 3D structure (25). Compared to highly structured protein regions, disordered regions may have several functional advantages for efficient interactions with partner molecules (26,27), such as the capacity of initiating binding by long range electrostatic interactions, high flexibility, binding plasticity and speed, minimal steric restrictions in binding, and the ability to form very stable intertwined complexes (26-27,64-69). Hence, our investigation merges theoretical advances in protein biophysics with very practical aspects of protein interaction, the identification of peptide sequences best suited for recognition by CDRs of antibodies.

SUPPLEMENTARY DATA

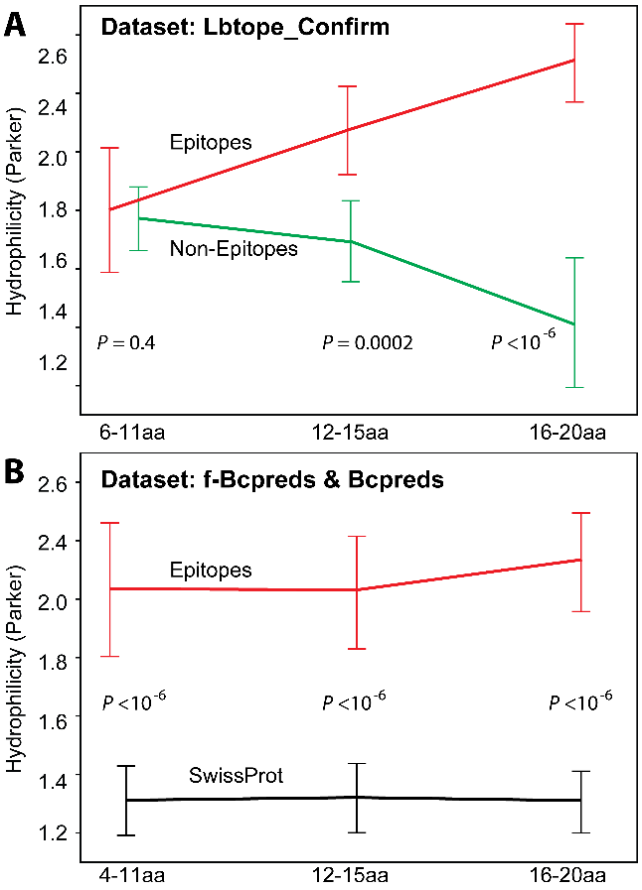


Fig. 3.S1. B-cell epitope prediction in dependence on epitope/non-epitope sequence length.

(A) Epitope length-dependent hydrophilicity of epitopes and non-epitopes. All 6-20aa epitopes and non-epitopes of the Lbtope_Confirm dataset (24) were grouped into 6-11aa, 12-15aa and 16-20aa sequences. Comparisons of the 11-15aa and 16-20aa epitopes with the corresponding non-epitopes shows highly significant differences in hydrophilicity scores (±95% CI). In contrast, 6-11aa epitopes and non-epitopes do not differ.

(B) Epitope length-dependent hydrophilicity of fBcpreds_Epitopes (21) and random Swiss-Prot peptides. The 4-11aa, 12-15aa and 16-20aa epitopes of the fBcpreds dataset were compared to all random Swiss-Prot 12aa, 14aa, and 18aa peptides of the Bcpreds BCP12, BCP14 and BCP18 datasets (20). The hydrophilicity scores of both datasets are length-independent, and those of epitopes are significantly higher than those of the corresponding random Swiss-Prot peptides. An additional finding is that the 6-11aa and 12-15aa non-epitopes of the Lbtope_Confirm dastaset have higher hydrophilicity scores than the random Swiss-Prot peptides. The length dependent epitope/non-epitope discrimination in the Lbtope_Confirm dastaset may be due to a high frequency of incorrect identification of short 4-15aa peptides as non-epitopes.

Table 3.S2. Brief description of the datasets.

Concatenated Virtual Polyprotein Datasets: Extended 50aa B-cell Epitope/Non-Epitope embedded into 150aa random sequences.
1.1 Epitopes: Positive datasets **1.1.1 fBcpreds (Epitopes):** All 16-20 amino acid (aa) epitopes were selected from 934 unique B-cell epitopes of the fBcpreds_pos_nr80 dataset (http://ailab.ist.psu.edu/bcpred/data.html; (21). These 16-20aa epitopes (201 in total) were extended to 50aa from the source proteins retrieved by BLAST search from the NCBI database. Sequences are shown in the Supplementary Dataset S10. **1.1.2 Lbtope_Confirm (Epitopes):** All 16-20 amino acid (aa) epitopes were selected from 1042 unique B-cell epitopes of the Lbtope_Confirm dataset (http://crdd.osdd.net/raghava/lbtope/data.php; (24). Epitopes of Lbtope_Confirm have been reported as epitopes in at least two studies (24). These 16-20aa epitopes (240 in total) were extended to 50aa as described for fBcpreds (Epitopes). Sequences are shown in the Supplementary Dataset S10. **1.1.3 Chl_43Prot (Epitopes):** Epitopes from 43 proteins of *Chlamydia* spp.; sequences are non-overlapping and ≥50aa apart; epitopes were confirmed by ELISA with murine hyper-immune sera; each of these peptide antigens in this dataset showed strong reactivity with individual or pooled sera (2). These sequences are highly divergent and no cross-reactivity was found in more than 95% of the peptides tested (2). These 16-40aa epitopes (70 in total) were extended to 50aa as described for fBcpreds (Epitopes). Sequences are shown in the Supplementary Dataset S10. **1.1.4 Chl_18Prot (Epitopes):** Epitopes from 18 proteins of *Chlamydia* spp. were confirmed by ELISA with murine hyper-immune sera or bovine sera of animals with PCR-confirmed natural *Chlamydia* spp. infection (2). A total of 119 peptide sequences were included in this dataset, some of the sequences are overlapping, but ≥10aa apart. These 16-20aa epitopes were extended to 50aa as described for fBcpreds (Epitopes). Sequences are shown in the Supplementary Dataset S10.
1.2 Random Peptides or Non-Epitopes: Negative datasets **1.2.1 fBcpreds_Prot (50aa random peptides):** Random 50aa peptides of fBcpreds epitope (21) source proteins were extracted. The number of random peptides drawn from a protein was equal to the number of epitopes present on this protein, and these 201 random peptides were used as a negative dataset. Sequences are shown in the Supplementary Dataset S10. **1.2.2 Swiss-Prot (50aa random peptides):** A total of 260 peptides, 50aa in length, were randomly extracted from Swiss-Prot with the RandSeq Tool (http://web.expasy.org/randseq/) and used as negative (non-epitopes) dataset. The amino acid (AA) composition of these random peptides were based on average AA composition of the Swiss-Prot database. Sequences are shown in the Supplementary Dataset S10.

1.2.3 Ctr_Proteome (random peptides): Peptide dataset of the *C. trachomatis* proteome of ~900 proteins (32). From every 1,000aa region of the proteome of *C. trachomatis*, 50aa peptide sequences were randomly selected and the total number of these random peptides was 310. Antibody reactivity of these peptides is unknown. Sequences are shown in the Supplementary Dataset S10.

1.2.4 Lbtope_Confirm (Non-Epitopes): All 16-33aa non-epitopes were selected from 1795 unique B-cell non-epitopes of Lbtope_Confirm dataset (http://crdd.osdd.net/raghava/lbtope/data.php; (24). These non-epitopes have been reported as non-epitopes in at least two studies (24). These 16-33aa non-epitopes (135 in total) were extended to 50aa as described for fBcpreds (Epitopes). Sequences are shown in the Supplementary Dataset S10.

1.2.5 Chl_18Prot (Non-Epitopes): Non-reactivity was confirmed by ELISA with mouse or bovine sera as described for Chl_18Prot_Epitopes. A total of 107 peptide sequences was included in this dataset, some of the sequences are overlapping, but ≥10aa apart. These 16-20aa non-epitopes were extended to 50aa as described for fBcpreds (Epitopes). Sequences are shown in the Supplementary Dataset S10.

1.3 Random Swiss-Prot Sequences: Inter-Peptide spacing sequences flanking each sequence of the positive/negative datasets

1.3.1 Swiss-Prot (150aa sequences/inter-peptide spacing sequences): A total of 25 sequences, 150aa in length, were selected from randomly extracted Swiss-Prot sequences (http://web.expasy.org/randseq/). The AA composition of these random peptides was based on the average AA composition of the Swiss-Prot database. These Swiss-Prot sequences were randomly selected to flank peptide sequence of all datasets into a single concatenated virtual polyprotein. Sequences are shown in the Supplementary Dataset S10.

Individual Protein Sequence Datasets: B-cell Epitope/Non-Epitope/Random Peptide Annotated.

2.1.1 Chl_18Prot (Pos, Neg and NT residues/peptides): All residues of 18 immunodominant proteins of *Chlamydia* spp. were annotated as Pos (positive, epitope), Neg (negative, non-epitope) or NT residues (not tested, hence not known as epitope or non-epitope residue). The annotation is based on the reactivity of peptide antigens with murine and/or bovine sera as described for Chl_18Prot_Epitopes. All Pos and Neg regions are based on reactivity tested with 16-20aa peptides (2). The number of residues annotated as Pos, Neg, NT and Neg+NT are 1536, 1458, 5537 and 6995, respectively. The epitope density was 18% of total residues on these 18 proteins. In addition, 5-49aa peptides were drawn 10aa apart from these annotated regions that are designed as Pos, Neg, and Neg+NT peptides and the total number of peptides in these three datasets are 119, 107 and 632 respectively. Protein sequences are shown in the Supplementary Dataset S10.

Table 3.S3. Optimal B-cell epitope prediction for virtual concatenated polyproteins by 20-30aa Peptide scores.

Scales	Residue Scoring	Peptide Scoring						
	Central 1aa[a]	**10aa**[b]	**15aa**[b]	**20aa**[b]	**25aa**[b]	**30aa**[b]	**Δ AUC**[c]	**ΔAUC**[d]
Hydrophilicity, ProtScale (Parker, 1986; ref-11)	0.691	0.735	0.757	0.774	0.781	**0.782**	**0.044**[E]	**0.046**[F]
Hydrophobicity, ProtScale (Miyazawa, 1985; ref-29)	0.697	0.741	0.763	0.776	**0.779**	0.776	**0.044**[E]	**0.038**[F]
Flexibility, IEDB (Karplus, 1985; ref-12)	0.645	0.690	0.719	0.732	**0.733**	0.723	**0.045**[E]	**0.043**[F]
Beta-turn, IEDB (Chou, 1978; ref-31)	0.616	0.653	0.670	0.683	**0.689**	0.685	**0.037**[E]	**0.036**[f]
Bepipred, IEDB (Larsen, 2006; ref-19)	0.720	0.748	0.772	0.785	0.789	**0.790**	**0.028**[E]	**0.041**[F]
Bcepred, Avg7 (Saha, 2004; ref-15)	0.649	0.690	0.714	0.724	**0.725**	0.719	**0.041**[E]	**0.035**[f]
Lbtope_Confirm (Singh, 2013; ref-24)	0.565	0.570	0.579	0.582	**0.612**	0.573	**0.005**[e]	**0.042**
Disorder, VSL2B (Peng, 2006; ref-33)	0.762	0.768	0.774	0.775	**0.776**	0.775	**0.006**[e]	**0.008**
Disorder, IUPred-L (Dosztányi, 2005; ref-30)	0.769	0.776	**0.777**	0.776	0.776	0.774	**0.007**[e]	**0**

[a] A single score of the central residue was considered for each peptide in the datasets shown in Table 3.1.

[b] Scores of the central 10aa/15aa/20aa/25aa/30aa residues were averaged to a single peptide score.

[c] Comparison of 1aa versus 10aa peptide scoring.

[d] Comparison of 10aa versus 25aa peptide scoring.

[E/e] AUC significantly higher for 10aa peptide scoring than for 1aa scoring ([E], $10^{-6} \leq P < 10^{-3}$; e, $10^{-3} \leq P \leq 10^{-2}$; one tailed paired Student's t-test with 12 AUC values).

[F/f] AUC significantly higher for 25aa peptide scoring than for 10aa peptide scoring ([F], $10^{-4} \leq P < 10^{-3}$; f, $10^{-3} \leq P \leq 0.023$).

In contrast to Table 3.1, where scores for each of the central 20 AA residues were considered, we used here an averaged score of the central AA residues of epitopes/non-epitopes and indicate the average AUC value of all 12 comparisons that were shown in Table 3.1. We asked the question if scoring complete peptides improved prediction accuracy, and if so, what the peptide length dependency of such an improvement was. Here, we determined the optimal window for the moving average by comparing windows of 1aa, 10aa, 15aa, 20aa, 25aa, or 30aa peptides of the datasets shown in Table 3.1. Average AUC values that differ by <0.01 values from the maximum (bold red font) are shown in red font. The data again

confirm that B-cell epitope prediction performance is optimum when scores for 20aa-30aa peptide regions were used. Scores for long peptide particularly improve prediction for individual AA propensity scales such as hydrophilicity, hydrophobicity, flexibility, or Bepipred. Disorder tendency scales such as IUPRed-L and VSL2B are only marginally improved because they already analyze individual residues in the context of their neighboring residues.

Table 3.S4. Standardization of individual protein scores improves B-cell epitope prediction.

Datasets[a]	Scales[b]	Effect of Standardization (Residue Scores)			Effect of Long Peptide Scores (Standardized)[g]							
		(Non-Standard)[c]	(Standard)[d]	**Δ AUC[e]**	5 aa	9 aa	17 aa	25 aa	33 aa	41 aa	49 aa	**Δ AUC[h]**
Pos vs Neg	Hydrophilicity (Parker, 1986; ref-11)	0.63	0.62	**-0.01**	0.67	0.69	0.74	0.75	0.78	**0.79**	0.78	**0.06[i]**
	Hydrophobicity (Miyazawa, 1985; ref-29)	0.65	0.65	**0.00**	0.70	0.73	0.78	0.80	0.82	**0.84**	0.84	**0.07[i]**
	Bepipred, IEDB (Larsen, 2006; ref-19)	0.74	0.74	**0.00**	0.75	0.77	0.81	0.84	0.85	**0.86**	0.86	**0.07[i]**
	Lbtope_Confirm (Singh, 2013; ref-24)	0.61	0.60	**-0.01**	0.60	0.61	0.63	0.65	**0.66**	0.65	0.65	**0.04[i]**
	Disorder, VSL2B (Peng, 2006; ref-33)	0.73	0.82	**0.09[f]**	0.82	0.83	0.84	**0.84**	0.84	0.84	0.82	**0.01**
	Disorder, IUPred-L (Dosztányi, 2005; ref-30)	0.79	0.84	**0.05[f]**	0.84	0.85	0.85	**0.86**	0.86	0.86	0.85	**0.01**
	Solvent Accessibility (ASA_Spine-X; ref-34)	0.50	0.61	**0.11[f]**	0.67	0.74	0.80	0.81	**0.83**	0.81	0.81	**0.07[i]**
	Polymorphism (Waterhouse, 2009; ref-35)	0.70	0.73	**0.03[f]**	0.82	0.84	0.84	**0.85**	0.84	0.83	0.83	**0.01**
Pos vs Neg+NT	Hydrophilicity (Parker, 1986; ref-11)	0.71	0.72	**0.01**	0.77	0.79	0.83	0.84	0.85	**0.86**	0.86	**0.05[i]**
	Hydrophobicity (Miyazawa, 1985; ref-29)	0.74	0.75	**0.01**	0.80	0.83	0.87	0.88	0.89	**0.90**	0.89	**0.05[i]**
	Bepipred, IEDB (Larsen, 2006; ref-19)	0.82	0.83	**0.01**	0.84	0.86	0.89	0.90	**0.91**	0.91	0.90	**0.04[i]**
	Lbtope_Confirm (Singh, 2013; ref-24)	0.65	0.66	**0.01**	0.68	0.68	0.71	0.72	**0.73**	0.72	0.72	**0.04[i]**
	Disorder, VSL2B (Peng, 2006; ref-33)	0.82	0.88	**0.06[f]**	0.88	0.89	0.90	**0.90**	0.90	0.89	0.88	**0.01**
	Disorder, IUPred-L (Dosztányi, 2005; ref-30)	0.85	0.91	**0.06[f]**	0.91	0.92	0.92	**0.93**	0.92	0.91	0.90	**0.01**
	Solvent Accessibility (ASA_Spine-X; ref-34)	0.67	0.67	**0.00**	0.76	0.82	0.88	0.89	**0.90**	0.89	0.88	**0.07[i]**
	Polymorphism (Waterhouse, 2009; ref-35)	0.75	0.82	**0.07[f]**	0.87	0.88	0.89	**0.89**	0.88	0.88	0.88	**0.01**

[a] The Chl-18Prot dataset was analyzed. Pos, Positive (Epitopes); Neg, Negative (Non-epitope); NT, Not Tested (hence epitope or non-epitope is not known). Thus, Pos vs Neg indicates epitopes were compared with non-epitopes; and Pos vs Neg+NT indicates epitopes were compared with remaining total protein.

[b] Solvent Accessibility (ASA_Spine-X), residue solvent accessibility (34); Polymorphism, sequence divergence in multiple sequence alignment, calculated by inverting the conservation score of AACon in the Jalview freeware (35).

[c] Original, non-standardized score for central 1aa residue in the peptides. These scores were obtained with individual protein sequence as input.

[d] Standardized score for central 1aa residue in the peptides. Original scores were standardized for each of the 18 chlamydial proteins.

[e] Difference in AUC values between standardized scoring and non-standardized scoring.

[f] Sensitivity at given specificity is significantly higher in ROC curves for standardized scoring than non-standardized scoring ([f], $10^{-6} \leq P \leq 0.01$; one tailed paired Student's t-test).

[g] Peptide scores were calculated using the average of standardized scores for the central 5aa, 9aa, 17aa, 25aa, 33aa, 41aa or 49aa residues.

[h] Difference in AUC values between standardized scoring of 25aa and 9aa peptides.

[i] Sensitivity at given specificity is significantly higher in ROC curves for 25aa standardized scoring than 9aa standardized scoring ([i], $10^{-6} \leq P \leq 0.01$; one tailed paired Student's t-test).

In Table 3.S3, the requirement for scores of long peptides for optimum prediction is shown by using concatenated polyproteins. In this approach, scores cannot be standardized for each epitope/non-epitope source proteins. Here, we asked the same question but using 18 individual chlamydial proteins. Importantly, we standardized scores for each protein to a mean of 0 and an SD of 1. Experimentally validated epitopes were compared with experimentally validated non-epitopes (Pos vs Neg) or the total remaining protein (Pos vs Neg+NT). Standardization significantly improved prediction performance of disorder, sequence polymorphism and solvent accessibility scales, but not of individual AA propensity scales such as hydrophilicity or hydrophobicity. Similar to Table 3.S3, scores averaged for long sequence windows provided significantly higher sensitivity at given specificity for AA propensity scales as well solvent accessibility scales, but not for polymorphism or for disorder scales. Average AUC values that differ by <0.01 values from the maximum (bold red font) are shown in red font.

A

Prediction Methods	40% Sensitivity		60% Sensitivity		80% Sensitivity		90% Sensitivity		95% Sensitivity		Average		
	Specificity	ACC	Specificity	ACC	Specificity	ACC	Specificity	ACC	Specificity	ACC	AUC	ACC	Rank
D1, Disorder (IUPred-L)	99	89	96	90	87	86	75	77	67	72	**0.925**	**83**	**5**
D2, Disorder (VSL2B)	98	89	92	87	81	81	71	74	65	70	**0.900**	**80**	**8**
S1, Solvent Accessibility (Spine-X)	97	88	94	89	83	83	70	73	44	52	**0.884**	**77**	**11**
S2, Solvent Accessibility (PaleAle4.0)	96	87	91	86	84	83	70	73	53	60	**0.889**	**78**	**10**
H1, Hydrophilicity (Miyazawa)	98	89	92	87	79	79	62	67	52	59	**0.880**	**76**	**12**
H2, Hydrophilicity (Parker)	96	87	87	83	70	72	49	56	44	52	**0.837**	**70**	**18**
2*D1+S2, Combined Scale	99	89	97	91	90	88	83	84	78	81	**0.938**	**87**	**1**
2*D1+S1, Combined Scale	99	89	96	90	92	90	81	82	75	78	**0.934**	**86**	**2**
D1+S2, Combined Scale	98	89	96	90	90	88	86	87	77	80	**0.937**	**87**	**1**
D1+S1, Combined Scale	99	89	96	90	92	90	82	83	73	77	**0.932**	**86**	**2**
D1+D2+S2, Combined Scale	99	89	97	91	91	89	83	84	75	78	**0.937**	**86**	**2**
D1+D2+S1, Combined Scale	99	89	97	91	90	88	82	83	73	77	**0.935**	**86**	**2**
Bepipred, IEDB	98	89	96	90	87	86	73	76	55	61	**0.904**	**80**	**8**

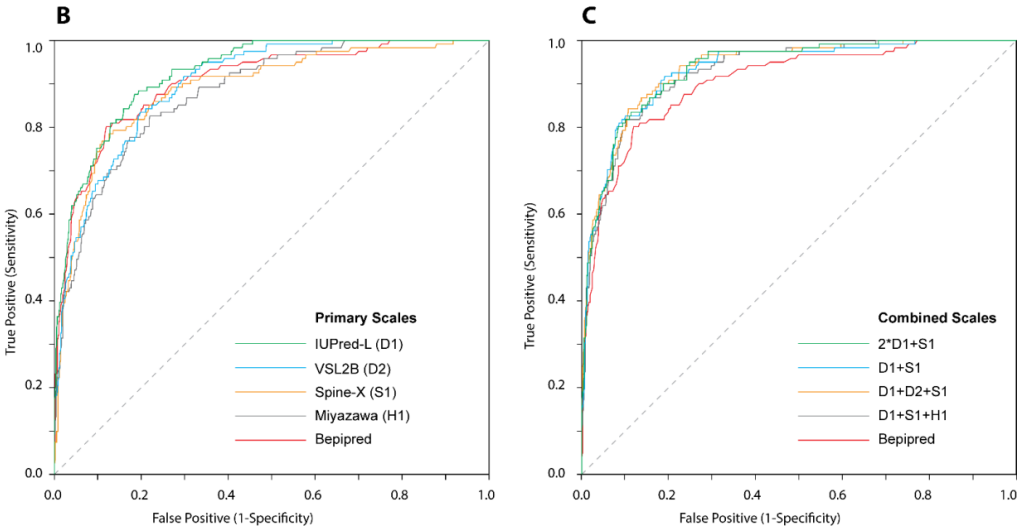


Fig. 3.S5. Combined scales provide only marginal improvement for B-cell epitope prediction.

(A) Prediction performance with optimal 25aa moving average scores of the Chl-18Prot dataset. At five specified sensitivities, B-cell epitope prediction specificities and the corresponding accuracies (ACC) were shown. Scale performance were ranked based on the average of five ACC values, and the top scale were ranked 1 and subsequently remaining scales were ranked with addition of 1 to rank for 1% reduction of ACC. 2*D1, D1 score weighted 2×.

(B) Prediction by use of primary scales or (C) combined scales. Plots of true positive versus false positive (ROC curve) are shown.

Here we asked, for the 18 chlamydial proteins, the question whether combined scales will at moving 25aa averages improve B-cell epitope prediction. Data confirm the previous result that the combination of scales provides only marginally improved B-cell epitope prediction (Fig. 3.3, Table 3.2). The best combination of the primary scales provides only ~4% improvement of prediction accuracy (ACC) in five tests at 40, 60, 80, 90 and 95% sensitivities. The dominant conclusion is that, as shown in Tables 3.S3 and 3.S4, the main improvement for B-cell epitope prediction comes from using scores of long peptides obtained by an optimal primary scale (IUPred-L) and standardized for each protein (required for comparison only).

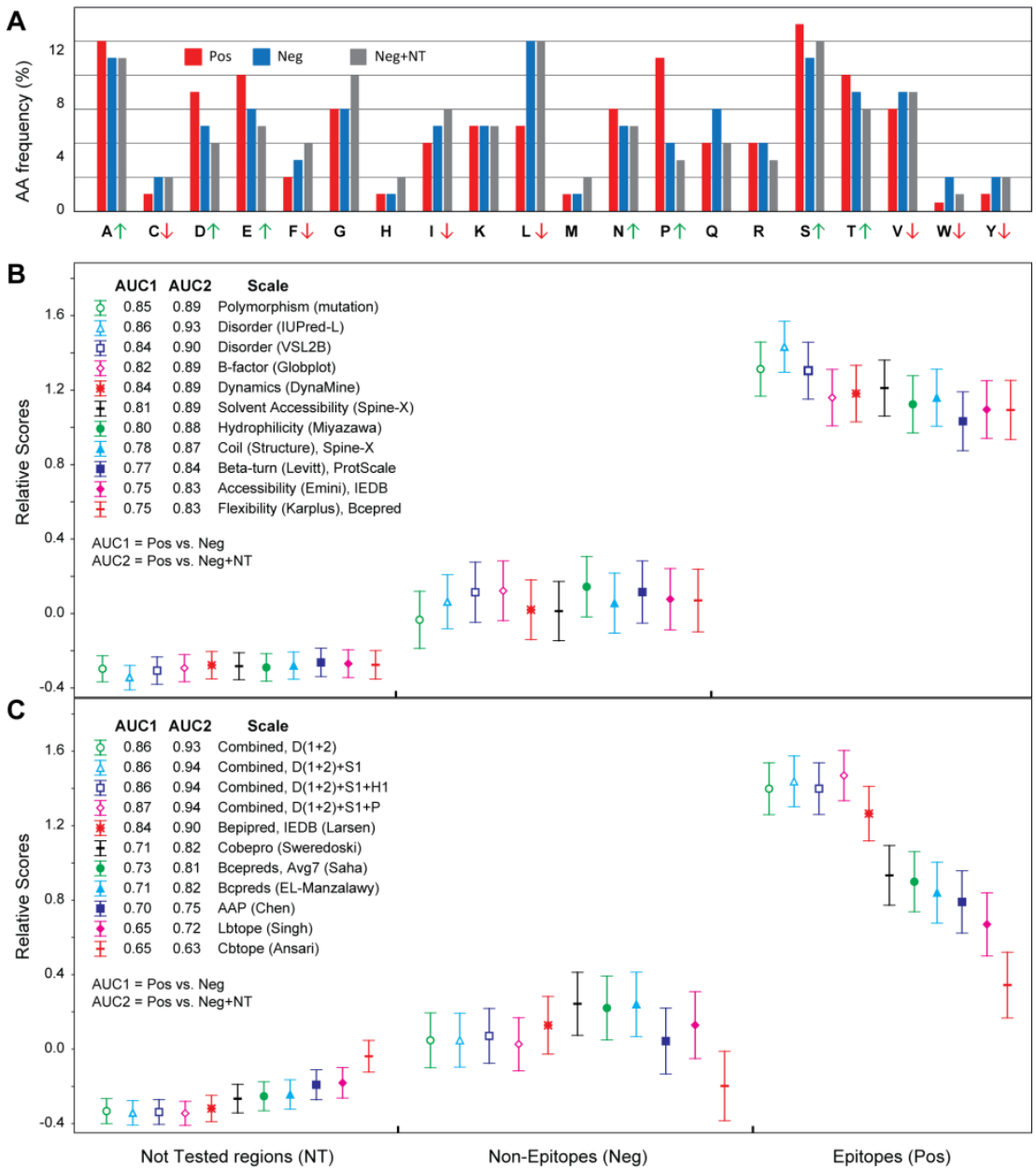


Fig. 3.S6. Dominant properties of B-cell epitopes.

(A) Comparative amino acid frequencies. Pos, Epitopes; Neg, Non-Epitopes; NT, Not Tested in the 18 chlamydial protein dataset. (B) Primary scales and (C) Combined scales and machine learning algorithms.

Fig. 3.S6A shows that B-cell epitopes are enriched for proline, followed by glutamic and aspartic acids, and alanine, asparagine, threonine, and serine. Epitopes are relatively depleted of leucine, isoleucine, tryptophan, phenylalanine, tyrosine, and cysteine. Fig 3.S6B and Table 3.4 show that scores for virtually all physicochemical, structural, and evolutionary protein properties (Appendices S8, S9) are consistently higher when epitopes were tested against the remaining protein than against experimentally validated non-epitopes. As a consequence, the prediction performance (AUC in ROC curves) against remaining total protein sequences is also consistently higher for all scales. An explanation for this counterintuitive observation is that non-epitopes had initially been selected as candidate epitopes by high scores in

prediction scales, but failed to react with antibodies in wet-lab screening. The higher scores for tested non-epitopes induced a pre-selection bias that makes evaluation of B-cell epitope prediction scales more difficult. Fig 3.S6C compares B-cell epitope prediction scales that were the best combinations of the best primary scales in this study, and published combination scales. Together with Figs. 3.2 and 3.S1, Fig 3.S6 illustrates the poor discriminatory power of machine learning algorithms, resulting from training in suboptimal datasets that used short peptide antigens for antibody assays. For example, Lbtope was trained on confirmed epitopes and short 6-16aa (80%) non-epitopes. In contrast, Bcpreds was trained on confirmed epitopes but random Swiss-Prot peptides as non-epitopes, and performed significantly better than Lbtope (0.06-0.19 AUC value difference between Bcpreds & Lbtope). Of the published combined B-cell epitope prediction scales, only Bepipred showed acceptable performance, marginally better than the accurate Parker hydrophilicity scale (0.008 Δ AUC compared to Parker hydrophilicity, Table S3). Bepipred, however, combines the Parker scale with a hidden Markov model. Finally, we show that the single IUPred-L protein disorder scale, or optimal combined scales, are significantly more sensitive at any given specificity than Bepipred (Student's t-test, $P < 10^{-4}$, Fig. 3.3). Protein propensity scales such as disorder tendency, relative solvent accessibility, hydrophobicity, or hydrophilicity do not require B-cell epitope datasets for training, but provide an absolute score. We show in this study that all these scores perform well in all datasets when applied to optimal length 20-30aa peptides (Appendices S8, S9), and hence these scales should be more reliable for B-cell epitope prediction from various source proteins.

Table 3.S7. Prediction performance (AUC in ROC curves) of all scales tested B-cell epitopes for proteins by using 25aa peptide scores.

Designation	Scale	AUC	Rank	Web address for freeware tools
D1	Disorder (IUPred-L)	**0.93**	**2**	http://iupred.enzim.hu/
D2	Disorder (VSL2B)	**0.90**	**5**	http://www.disprot.org/metapredictor.php
D3	Disorder (IUPred-S)	**0.92**	**3**	http://iupred.enzim.hu/
D4	Disorder (PrDOS)	**0.89**	**6**	http://prdos.hgc.jp/cgi-bin/top.cgi
D5	Disorder (B-factors-3.5sd_GlobPlot2)	**0.89**	**6**	http://globplot.embl.de/
D6	Disorder (FoldUnfold)	**0.89**	**6**	http://bioinfo.protres.ru/ogu/
D7	Disorder (DynaMine)	**0.89**	**6**	http://dynamine.ibsquare.be/submission/
D8	Disorder (B-factors-2sd_GlobPlot2)	**0.89**	**6**	http://globplot.embl.de/
D9	Disorder (Spine-D)	**0.88**	**7**	http://sparks-lab.org/SPINE-D/
D10	Disorder (ΔPSI_Spine-X)	**0.88**	**7**	http://sparks-lab.org/SPINE-X/
D11	Disorder (HotLoops_DisEMBL1.5)	**0.88**	**7**	http://dis.embl.de/
D12	Disorder (X-Ray_Espritz)	**0.88**	**7**	http://protein.bio.unipd.it/espritz/
D13	Disorder (Remark465_DisEMBL1.5)	**0.87**	**8**	http://dis.embl.de/
D14	Disorder (Poodle_I+SS+ASA)	**0.87**	**8**	http://mbs.cbrc.jp/poodle/poodle-i.html
D15	Disorder (NMR_Espritz)	**0.86**	**9**	http://protein.bio.unipd.it/espritz/
D16	Disorder (Poodle_I+SS+ASA+SimS)	**0.86**	**9**	http://mbs.cbrc.jp/poodle/poodle-i.html
D17	Disorder (VSL3)	**0.85**	**10**	http://www.disprot.org/metapredictor.php
D18	Disorder (Ponder-Fit1)	**0.84**	**11**	http://www.disprot.org/metapredictor.php
D19	Disorder (Coils_DisEMBL1.5)	**0.84**	**11**	http://dis.embl.de/
D20	Disorder (Disorder to Order_VLXT)	**0.83**	**12**	http://www.disprot.org/metapredictor.php
D21	Disorder (Poodle_I)	**0.83**	**12**	http://mbs.cbrc.jp/poodle/poodle-i.html
D22	Disorder (Kyte-Doolittle_Globplot2)	**0.82**	**13**	http://globplot.embl.de/
D23	Disorder (Hopp-Woods_Globplot2)	**0.82**	**13**	http://globplot.embl.de/
D24	Disorder (Remark-465_Globplot2)	**0.81**	**14**	http://globplot.embl.de/
D25	Disorder (Poodle_Missing)	**0.80**	**15**	http://mbs.cbrc.jp/poodle/poodle.html
D26	Disorder (Russell-Linding_Globplot2)	**0.79**	**16**	http://globplot.embl.de/
D27	Disorder (Disprot_Espritz)	**0.77**	**18**	http://protein.bio.unipd.it/espritz/
D28	Disorder (Poodle_B-factor)	**0.76**	**19**	http://mbs.cbrc.jp/poodle/poodle.html
D29	Disorder (Poodle_Long)	**0.75**	**20**	http://mbs.cbrc.jp/poodle/poodle.html
D30	Disorder (Ponder-Fit2)	**0.73**	**22**	http://www.disprot.org/metapredictor.php
D31	Disorder (ΔPhi_Spine-X)	**0.71**	**24**	http://sparks-lab.org/SPINE-X/
D32	Disorder (Disorder to Order_Anchor)	**0.61**	**34**	http://iupred.enzim.hu/
mD1	Disorder (MFDp)	**0.92**	**3**	http://biomine-ws.ece.ualberta.ca/MFDp.html
mD2	Disorder (M_GeneSilico)	**0.91**	**4**	http://genesilico.pl/metadisorder/
mD3	Disorder (MMD_GeneSilico)	**0.91**	**4**	http://genesilico.pl/metadisorder/
mD4	Disorder (MetaPrDOS)	**0.90**	**5**	http://prdos.hgc.jp/cgi-bin/meta/top.cgi
mD5	Disorder (MMD2_GeneSilico)	**0.90**	**5**	http://genesilico.pl/metadisorder/
B1	Bepipred (IEDB)	**0.90**	**5**	http://tools.immuneepitope.org/bcell/
B2	AVG4_HFAE (Bcepred)	**0.85**	**10**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B3	MIN4_HFAE (Bcepred)	**0.84**	**11**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B4	MAX4_HFAE (Bcepred)	**0.83**	**12**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B5	Bcpreds	**0.82**	**13**	http://ailab.ist.psu.edu/bcpred/predict.html
B6	COBEpro	**0.82**	**13**	http://scratch.proteomics.ics.uci.edu/
B7	AVG7 (Bcepred)	**0.81**	**14**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B8	MAX7 (Bcepred)	**0.79**	**16**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B9	Chen AAP (Bcpreds)	**0.75**	**20**	http://ailab.ist.psu.edu/bcpred/predict.html
B10	Lbtope	**0.72**	**23**	http://crdd.osdd.net/raghava/lbtope/protein.php
B11	MIN7 (Bcepred)	**0.69**	**26**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B12	Polarity (Ponnuswamy, Bcepred)	**0.64**	**31**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B13	Cbtope	**0.63**	**32**	http://www.imtech.res.in/raghava/cbtope/submit.php
B14	Antigenicity (Kolaskar, IEDB)	**0.42**	**53**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
B15	Antigenicity (Kolaskar, Bcepred)	**0.41**	**54**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html

Designation	Scale	AUC	Rank	Web address for freeware tools
S1	Accessible Surface Area (ASA_Spine-X)	**0.89**	**6**	http://sparks-lab.org/SPINE-X/
S2	Relative Solvent Accessibility (PaleAle4.0)	**0.89**	**6**	http://distillf.ucd.ie/porterpaleale/
S3	Relative Solvent Accessibility (RSA_SANN)	**0.87**	**8**	http://lee.kias.re.kr/~newton/sann/
S4	Solvent Accessibility (ACCpro20)	**0.87**	**8**	http://scratch.proteomics.ics.uci.edu/
S5	Relative Solvent Accessibility (RSA_NetSurfP)	**0.83**	**12**	http://www.cbs.dtu.dk/services/NetSurfP/
S6	Absolute Solvent Accessibility (ASA_SANN)	**0.82**	**13**	http://lee.kias.re.kr/~newton/sann/
S7	Absolute Solvent Accessibility (ASA_NetSurfP)	**0.79**	**16**	http://www.cbs.dtu.dk/services/NetSurfP/
S8	bRSA (Exposed/buried, PaleAle4.0)	**0.79**	**16**	http://distillf.ucd.ie/porterpaleale/
P	Polymorphism (Mutation, AACon in Jalview)	**0.89**	**6**	http://www.jalview.org/
H1	Hydrophobicity (Miyazawa, ProtScale)	**0.88**	**7**	http://web.expasy.org/protscale/
H2	Hydrophobicity (Guy, ProtScale)	**0.87**	**8**	http://web.expasy.org/protscale/
H3	Hydrophobicity (Manavalan, ProtScale)	**0.87**	**8**	http://web.expasy.org/protscale/
H4	Hydrophobicity (Sweet, ProtScale)	**0.86**	**9**	http://web.expasy.org/protscale/
H5	Hydrophilicity (Parker, Bcepred)	**0.85**	**10**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
H6	Hydrophilicity (Parker, ProtScale)	**0.84**	**11**	http://web.expasy.org/protscale/
H7	Hydrophilicity (Parker, IEDB)	**0.84**	**11**	http://tools.immuneepitope.org/bcell/
H8	Hydrophobicity (Kyte, ProtScale)	**0.81**	**14**	http://web.expasy.org/protscale/
H9	Hydrophobicity (Hopp, ProtScale)	**0.80**	**15**	http://web.expasy.org/protscale/
H10	Hydrophobicity (Bull, ProtScale)	**0.78**	**17**	http://web.expasy.org/protscale/
C1	Coil (SS, Spine-X)	**0.87**	**8**	http://sparks-lab.org/SPINE-X/
H_SS	Helix (SS, Spine-X)	**0.72**	**23**	http://sparks-lab.org/SPINE-X/
E_SS	Strand (SS, Spine-X)	**0.83**	**12**	http://sparks-lab.org/SPINE-X/
H+E_SS	H+E (SS, Spine-X)	**0.87**	**8**	http://bioinf.cs.ucl.ac.uk/psipred/
C2	Coil (SS, PSIpred)	**0.88**	**7**	http://bioinf.cs.ucl.ac.uk/psipred/
H_SS	Strand (SS, PSIpred)	**0.67**	**28**	http://bioinf.cs.ucl.ac.uk/psipred/
E_SS	Helix (SS, PSIpred)	**0.66**	**29**	http://bioinf.cs.ucl.ac.uk/psipred/
H+E_SS	H+E (SS, PSIpred)	**0.88**	**7**	http://bioinf.cs.ucl.ac.uk/psipred/
C3	Coil (SS, NetSurfP)	**0.76**	**19**	http://www.cbs.dtu.dk/services/NetSurfP/
H_SS	Helix (SS, NetSurfP)	**0.65**	**30**	http://www.cbs.dtu.dk/services/NetSurfP/
E_SS	Strand (SS, NetSurfP)	**0.54**	**41**	http://www.cbs.dtu.dk/services/NetSurfP/
H+E_SS	H+E (SS, NetSurfP)	**0.76**	**19**	http://www.cbs.dtu.dk/services/NetSurfP/
S_SS	bSS (C/H+E, Porter)	**0.77**	**18**	http://distillf.ucd.ie/porterpaleale/
bT1	beta-Turn (Levitt, ProtScale)	**0.84**	**11**	http://web.expasy.org/protscale/
bT2	beta-Turn (Chou, IEDB)	**0.76**	**19**	http://tools.immuneepitope.org/bcell/
bT3	beta-Turn (Deleage, ProtScale)	**0.75**	**20**	http://web.expasy.org/protscale/
bT4	beta-Turn (Pellequer, Bcepred)	**0.70**	**25**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
A1	Accessibility (Emini, IEDB)	**0.83**	**12**	http://tools.immuneepitope.org/bcell/
A2	Accessibility (Emini, Bcepred)	**0.81**	**14**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
A3	Accessibility (Rose, ProtScale)	**0.79**	**16**	http://web.expasy.org/protscale/
A4	Buried (Rose, ProtScale)	**0.79**	**16**	http://web.expasy.org/protscale/
A5	Accessibility (Janin, Bcepred)	**0.78**	**17**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
A6	Accessible (Janin, ProtScale)	**0.70**	**25**	http://web.expasy.org/protscale/
F1	Flexibility (Karplus, Bcepred)	**0.83**	**12**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
F2	Flexibility (Karplus, IEDB)	**0.82**	**13**	http://tools.immuneepitope.org/bcell/
F3	Flexibility (Bhaskaran, ProtScale)	**0.81**	**14**	http://web.expasy.org/protscale/
F4	Flexibility (RMSF, PredyFlexy)	**0.62**	**33**	http://www.dsimb.inserm.fr/dsimb_tools/predyflexy/index.html
F5	Flexibility (B-factor, PredyFlexy)	**0.62**	**33**	http://www.dsimb.inserm.fr/dsimb_tools/predyflexy/index.html
F6	Flexibility (Flexibility, PredyFlexy)	**0.59**	**36**	http://www.dsimb.inserm.fr/dsimb_tools/predyflexy/index.html
M1	D1 + D2 + S1	**0.94**	**1**	See D1, D2 and S1
M2	D1 + D2 + P	**0.94**	**1**	See D1, D2 and P
M3	D1 + D2 + C1	**0.94**	**1**	See D1, D2 and C1
M4	D1 + D2 + S1 + H1	**0.94**	**1**	See D1, D2, S1 and H1
M5	D1 + D2 + S1 + P	**0.94**	**1**	See D1, D2, S1 and P
M6	D1 + D2	**0.93**	**2**	See D1 and D2
M7	D1 + D2 + H1	**0.93**	**2**	See D1, D2 and H1
M8	D1 + D2 + B1	**0.93**	**2**	See D1, D2 and B1
M9	D1 + S1	**0.92**	**3**	See D1 and S1
M10	D1 + H1	**0.90**	**5**	See D1 and H1

Amino acid (aa) sequences of chlamydial proteins (18 in total) were annotated as B-cell epitopes (Pos), non-epitopes (Neg) and Not Tested (NT) based on peptide reactivity with murine and/or bovine sera

(Supplementary B-cell Epitope Datasets). The protein sequences were used as input to obtain prediction scores for 151 scales. These scores were then standardized for each protein to a mean of 0 and standard deviation of 1 for individual proteins before statistical analysis of the combined peptide scores of all proteins. Discrimination of epitope peptides from remaining peptide regions ('Pos' versus 'Neg+NT') were tested in ROC analysis using 25aa peptide scores. Novel 126 combined scales were derived by linearly combining primary scales. The top performing scale was given rank 1, and subsequent scales were ranked by addition of 1 rank number for each 0.01 AUC reduction. The partial list of the 109 best scales is shown below. Top scales of each parameter category are shown in red font.

Table 3.S8. Ranking of epitope prediction scales by residue scores of individually tested chlamydial proteins.

Scale Category	Designation	Scale Implementation	AUC ± SD	Rank	Web address
Disorder Tendency	D1	IUPred-L (Dosztányi, 2005; ref-30)	**0.91 ± 0.08**	**2**	http://iupred.enzim.hu/
	D2	VSL2B (Peng, 2006; ref-33)	**0.89 ± 0.12**	**4**	http://www.disprot.org/metapredictor.php
	D3	B-factors 3.5sd (Linding, 2003a; ref-38)	**0.87 ± 0.10**	**6**	http://globplot.embl.de/
	D4	FoldUnfold (Galzitskaya, 2006; ref-39)	**0.87 ± 0.10**	**6**	http://bioinfo.protres.ru/ogu/
	D5	PrDOS (Ishida, 2007; ref-40)	**0.86 ± 0.09**	**7**	http://prdos.hgc.jp/cgi-bin/top.cgi
	D6	Remark465 (Linding, 2003b; ref-41)	**0.86 ± 0.10**	**7**	http://dis.embl.de/
	D7	DynaMine (Cilia, 2014; ref-42)	**0.85 ± 0.12**	**8**	http://dynamine.ibsquare.be/submission/
	mD1	M_MetaPredictor (Kozlowski, 2012; ref-43)	**0.89 ± 0.09**	**4**	http://genesilico.pl/metadisorder/
	mD2	MFDp_MetaPredictor (Mizianty, 2010; ref-44)	**0.87 ± 0.15**	**6**	http://biomine-ws.ece.ualberta.ca/MFDp.html
	mD3	MetaPrDOS (Ishida, 2008; ref-45)	**0.87 ± 0.10**	**6**	http://genesilico.pl/metadisorder/
Hydrophilicity	H1	ProtScale (Miyazawa, 1985; ref-29)	**0.88 ± 0.07**	**5**	http://web.expasy.org/protscale/
	H2	ProtScale (Guy, 1985; ref-46)	**0.86 ± 0.07**	**7**	http://web.expasy.org/protscale/
	H3	ProtScale (Sweet, 1983; ref-47)	**0.85 ± 0.09**	**8**	http://web.expasy.org/protscale/
	H4	ProtScale (Parker, 1986; ref-11)	**0.85 ± 0.10**	**8**	http://web.expasy.org/protscale/
	H5	ProtScale (Hopp, 1981; ref-10)	**0.79 ± 0.14**	**14**	http://web.expasy.org/protscale/
Solvent Accessibility	S1	ASA, Spine-X (Faraggi, 2009a; ref-34)	**0.86 ± 0.08**	**7**	http://sparks-lab.org/SPINE-X/
	S2	RSA, PaleAle4.0 (Mirabello, 2013; ref-37)	**0.88 ± 0.08**	**5**	http://distillf.ucd.ie/porterpaleale/
	S3	RSA, SANN (Joo, 2012; ref-48)	**0.85 ± 0.09**	**8**	http://lee.kias.re.kr/~newton/sann/
	S4	RSA, NetSurfP (Petersen, 2009; ref-49)	**0.84 ± 0.10**	**9**	http://www.cbs.dtu.dk/services/NetSurfP/
	S5	RSA, ACCpro20 (Magnan, 2014; ref-50)	**0.83 ± 0.09**	**10**	http://scratch.proteomics.ics.uci.edu/
Polymorphism	P	Sequence Alignment, AACon, Jalview, (Waterhouse, 2009; ref-35)	**0.87 ± 0.09**	**6**	http://www.jalview.org/
Flexibility	F1	ProtScale (Bhaskaran, 1988; ref-51)	**0.83 ± 0.08**	**10**	http://web.expasy.org/protscale/
	F2	IEDB (Karplus, 1985; ref-12)	**0.83 ± 0.10**	**10**	http://tools.immuneepitope.org/bcell/
Coils,	C1	C_Spine-X (Faraggi, 2009b; ref-36)	**0.82 ± 0.15**	**11**	http://sparks-lab.org/SPINE-X/
	C2	C_PSIpred (McGuffin, 2000; ref-52)	**0.82 ± 0.16**	**11**	http://bioinf.cs.ucl.ac.uk/psipred/
Accessibility	A1	IEDB (Emini, 1985; ref-13)	**0.82 ± 0.12**	**11**	http://tools.immuneepitope.org/bcell/
	A2	ProtScale (Rose, 1985; ref-53)	**0.79 ± 0.16**	**14**	http://web.expasy.org/protscale/
beta-Turn	bT1	ProtScale (Levitt, 1978; ref-54)	**0.81 ± 0.14**	**12**	http://web.expasy.org/protscale/
	bT2	ProtScale (Deleage, 1987; ref-55)	**0.77 ± 0.14**	**16**	http://web.expasy.org/protscale/
Specific B-cell epitope prediction	B1	Bepipred, IEDB (Larsen, 2006; ref-19)	**0.89 ± 0.08**	**4**	http://tools.immuneepitope.org/bcell/
	B2	Bcepred (Saha, 2004; ref-15)	**0.83 ± 0.06**	**10**	http://www.imtech.res.in/raghava/Bcepred/Bcepred_submission.html
	B3	Bcpreds (El-Manzalawy, 2008b; ref-20)	**0.82 ± 0.12**	**11**	http://ailab.ist.psu.edu/bcpred/predict.html
	B4	COBEpro (Sweredoski, 2009; ref-22)	**0.80 ± 0.10**	**13**	http://scratch.proteomics.ics.uci.edu/
	B5	Lbtope_Confirm (Singh, 2013; ref-24)	**0.75 ± 0.13**	**18**	http://crdd.osdd.net/raghava/lbtope/protein.php
	B6	AAP (Chen, 2007; ref-18)	**0.73 ± 0.15**	**20**	http://ailab.ist.psu.edu/bcpred/predict.html
	B7	Cbtope (Ansari, 2010; ref-23)	**0.63 ± 0.10**	**30**	http://www.imtech.res.in/raghava/cbtope/submit.php
	B8	Antigenicity, IEDB (Kolaskar, 1990; ref-17)	**0.27 ± 0.16**	**66**	http://tools.immuneepitope.org/bcell/
Combined scales	M1	D1 + D2 + S1	**0.92 ± 0.07**	**1**	See D1, D2 & S1
	M2	D1 + D2 + S1 + H1	**0.92 ± 0.07**	**1**	See D1, D2, S1 & H1
	M3	D1 + D2 + S1 + P	**0.92 ± 0.07**	**1**	See D1, D2 S1 & P
	M4	D1 + D2 + P	**0.92 ± 0.08**	**1**	See D1, D2 & P
	M5	D1 + D2 + B1	**0.92 ± 0.08**	**1**	See D1, D2 & B1

The scales were further tested for discriminatory power of individual epitope residues of 18 chlamydial proteins. Prediction scores were obtained with default options, 25aa moving average, and standardized for each protein (to allow for use in combined scales). The partial list of the best scales is shown below. Best performing scales for each parameter category are shown in red font.

REFERENCES

1. **1. Blythe MJ and Flower DR. 2005.** Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci* **14**: 246-248.

2. **2. Rahman KS, Chowdhury EU, Poudel A, Ruettger A, Sachse K and Kaltenboeck B. 2015.** Defining species-specific immunodominant B cell epitopes for molecular serology of *Chlamydia* species. *Clin Vacc Immunol* **22**: 539-552.

3. **Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM and Pupko T. 2008.** Computational characterization of B-cell epitopes. *Mol Immunol* **45:** 3477-3489.

4. **Ofran Y, Schlessinger A and Rost B. 2008.** Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. *J Immunol* **181**: 6230-6235.

5. **Sun J, Xu T, Wang S, Li G, Wu D and Cao Z. 2011.** Does difference exist between epitope and non-epitope residues? *Immunome Res* **201**: 1-11.

6. **Sivalingam GN and Shepherd AJ. 2012.** An analysis of B-cell epitope discontinuity. *Mol Immunol* **51**: 304-309.

7. **Kringelum JV, Nielsen M, Padkjær SB and Lund O. 2013.** Structural analysis of B-cell epitopes in antibody: protein complexes. *Mol Immunol* **53**: 24-34.

8. **Brack C, Hirama M, Lenhard-Schuller R and Tonegawa S. 1978.** A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**: 1-14.

9. **Wang J, Zhang Y, Lu C, Lei L, Yu P and Zhong G. 2010.** A genome-wide profiling of the humoral immune response to *Chlamydia trachomatis* infection reveals vaccine candidate antigens expressed in humans. *J Immunol* **185**: 1670-1680.

10. **Hopp TP and Woods KR. 1981**. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* **78**: 3824-3828.

11. **Parker JM, Guo D and Hodges RS. 1986.** New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **25**: 5425-5432.

12. **Karplus PA and Schulz GE. 1985.** Prediction of chain flexibility in proteins - a tool for the selection of peptide antigens. *Naturwissenschaften* **72**: 212-213.

13. **Emini EA, Hughes JV, Perlow D and Boger J. 1985.** Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* **55**: 836-839.

14. **Pellequer JL, Westhof E and Van Regenmortel MH. 1993**. Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunol lett* **36**: 83-99.

15. **Saha S & Raghava GP. 2004.** BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In *Artificial Immune Systems* (Nicosia G, Cutello V, Bentley PJ and Timmis JI eds), pp 197-204. Springer, Heidelberg, Germany.

16. **Ponomarenko JV and Van Regenmortel MH. 2009.** B cell epitope prediction. In *Structural Bioinformatics 2nd Edition* (Bourne PE and Gu J eds), pp 849–879. John Wiley, Hoboken, NJ, USA.

17. **Kolaskar AS and Prasad CT. 1990.** A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* **276**: 172-174.

18. **Chen J, Liu H, Yang J and Chou KC. 2007.** Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* **33**: 423-428.

19. **Larsen JE, Lund O and Nielsen M. 2006.** Improved method for predicting linear B-cell epitopes. *Immunome Res* **2**: 1-7.

20. **El-Manzalawy Y, Dobbs D and Honavar V. 2008.** Predicting linear B-cell epitopes using string kernels. *J Mol Recognit*, **21**: 243-255.

21. **El-Manzalawy Y, Dobbs D and Honavar V. 2008.** Predicting flexible length linear B-cell epitopes. *Comput Syst Bioinformatics Conf* **7**: 121-132.

22. **Sweredoski MJ and Baldi P. 2009.** COBEpro: a novel system for predicting continuous B-cell epitopes. *Protein Eng Des Sel* **22**: 113-120.

23. **Ansari HR and Raghava GP. 2010.** Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res* **6**: 1-9.

24. **Singh H, Ansari HR and Raghava GP. 2013.** Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS One* **8**: e62216.

25. **Uversky VN. 2013.** Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta (BBA)-Proteins and Proteomics* **1834**: 932-951.

26. **Uversky, VN. 2013.** A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* **22**: 693-724.

27. **Liu Z and Huang Y. 2014.** Advantages of proteins being disordered. *Protein Sci* **23**: 539-550.

28. **Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel R D and Bairoch A. 2005.** Protein identification and analysis tools on the ExPASy server. *Proteomics Protocols Handbook* (Walker JM *ed*), pp 571-607. Humana Press, Totowa, NJ, USA.

29. **Miyazawa S and Jernigan RL. 1985.** Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **18**: 534-552.

30. **Dosztányi Z, Csizmok V, Tompa P and Simon I. 2005.** IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**: 3433-3434.

31. **Chou YE and Fasman RC. 1978.** Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* **47**: 45-148.

32. **Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV and Davis RW. 1998.** Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**: 754-759.

33. **Peng K, Radivojac P, Vucetic S, Dunker AK and Obradovic Z. 2006.** Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**: 208.

34. **Faraggi E, Xue B and Zhou Y. 2009.** Improving the prediction accuracy of residue solvent accessibility and real-value backbone torsion angles of proteins by guided-learning through a two-layer neural network. *Proteins* **74**: 847-856.

35. **Waterhouse AM, Procter JB, Martin DM, Clamp M and Barton GJ. 2009.** Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

36. **Faraggi E, Yang Y, Zhang S and Zhou Y. 2009.** Predicting continuous local structure and the effect of its substitution for secondary structure in fragment-free protein structure prediction. *Structure* **17**: 1515-1527.

37. **Mirabello C and Pollastri G. 2013.** Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics* **29**: 2056-58.

38. **Linding R, Russell RB, Neduva V and Gibson TJ. 2003.** GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**: 3701-3708.

39. Galzitskaya OV, Garbuzynskiy SO and Lobanov MY. 2006. FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics* **22**: 2948-2949.

40. Ishida T and Kinoshita K. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* **35**: W460-W464.

41. Linding R, Jensen L J, Diella F, Bork P, Gibson TJ and Russell RB. 2003. Protein disorder prediction: implications for structural proteomics. *Structure* **11**: 1453-1459.

42. Cilia E, Pancsa R, Tompa P, Lenaerts T and Vranken WF. 2014. The DynaMine webserver: predicting protein dynamics from sequence. *Nucleic Acids Res* **42**: W264-W270.

43. Kozlowski LP and Bujnicki JM. 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**: 111.

44. Mizianty MJ, Stach W, Chen K, Kedarisetti KD, Disfani FM and Kurgan L. 2010. Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics* **26**: i489-i496.

45. Ishida T and Kinoshita K. 2008. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* **24**: 1344-1348.

46. Guy HR. 1985. Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys J* **47**: 61-70.

47. Sweet RM and Eisenberg D. 1983. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J Mol Biol* **171**: 479-488.

48. Joo K, Lee SJ and Lee J. 2012. SANN: Solvent accessibility prediction of proteins by nearest neighbor method. *Proteins* **80**: 1791-1797.

49. **Petersen B, Petersen TN, Andersen P, Nielsen M and Lundegaard C. 2009.** A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol* **9**: 51.

50. **Magnan CN and Baldi P. 2014.** SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* **30**: 2592-2597.

51. **Bhaskaran R and Ponnuswamy PK. 1988.** Positional flexibilities of amino acid residues in globular proteins. *Int J Pept Protein Res* **32**: 241-255.

52. **McGuffin LJ, Bryson K and Jones DT. 2000.** The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404-405.

53. **Rose GD, Geselowitz AR, Lesser GJ, Lee RH and Zehfus MH. 1985**. Hydrophobicity of amino acid residues in globular proteins. *Science* **229**: 834-838.

54. **Levitt M. 1978.** Conformational preferences of amino acids in globular proteins. *Biochemistry* **17**: 4277-4285.

55. **Deleage G and Roux B. 1987.** An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng* **1**: 289-294.

56. **Zhong GM, Reid RE and Brunham RC. 1990.** Mapping antigenic sites on the major outer membrane protein of *Chlamydia trachomatis* with synthetic peptides. *Infect Immun* **58**: 1450-1455.

57. **Conlan JW, Clarke IN and Ward ME. 1988.** Epitope mapping with solid-phase peptides: identification of type-, subspecies-, species-and genus-reactive antibody binding domains on the major outer membrane protein of *Chlamydia trachomatis*. *Mol Microbiol* **2**: 673-679.

58. **Pal S, Cheng X, Peterson EM and de la Maza LM. 1993.** Mapping of a surface-exposed B cell epitope to the variable sequent 3 of the major outer-membrane protein of *Chlamydia trachomatis*. *J Gen Microbiol* **139**: 1565-1570.

59. **Villeneuve A, Brossay L, Paradis G and Hébert J. 1994.** Determination of neutralizing epitopes in variable domains I and IV of the major outer-membrane protein from *Chlamydia trachomatis* serovar K. *Microbiol* **140**: 2481-2487.

60. **Batteiger BE. 1996.** The major outer membrane protein of a single *Chlamydia trachomatis* serovar can possess more than one serovar-specific epitope. *Infect Immun* **64**: 542-547.

61. **Westhof E, Altschuh D, Moras D, Bloomer AC, Mondragon A, Klug A and Van Regenmortel MHV. 1984.** Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature* **311**: 123-126.

62. **62. Tainer JA, Getzoff ED, Alexander H, Houghten RA, Olson AJ, Lerner RA and Hendrickson WA. 1984.** The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature* **312**: 127-134.

63. **Novotný J, Handschumacher M, Haber E, Bruccoleri RE, Carlson WB, Fanning DW, Smith JA and Rose GD. 1986.** Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc Natl Acad Sci USA* **83**: 226-230.

64. **Kriwacki RW, Hengst L, Tennant L, Reed SI and Wright PE. 1996.** Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* **93**: 11504-11509.

65. **Wright PE and Dyson HJ. 1999.** Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**: 321-331.

66. **Shoemaker BA, Portman JJ and Wolynes PG. 2000.** Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci USA* **97**: 8868-8873.

67. **Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC and Obradovic Z. 2001.** Intrinsically disordered protein. *J Mol Graph Model* **9**: 26-59.

68. **Dyson HJ and Wright PE. 2005.** Intrinsically unstructured proteins and their functions. *Nature* **6**: 197-208.

69. **Sugase KJ, Dyson HJ and Wright PE. 2007.** Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**: 1021-1025.

CHAPTER 4. CONCLUSIONS

4.1 SUMMARY AND OUTLOOK

Detection of type-/species-/genus-specific anti-*Chlamydia* antibodies is an essential tool in *Chlamydia* pathogenetic and epidemiological research as well as in everyday diagnostics. Despite poor sensitivity, cross-reactivity, cumbersome production of chlamydial antigens, and the tedious and subjective microscopic readout, the microimmunofluorescent test has been for 50 years the only subtype-specific test in chlamydial serology, and is available only in reference laboratories. The main objective of this study was to develop as modern replacement a simple, yet robust, assay for subtype-specific detection of anti-chlamydial antibodies. In this current study, we accomplished this objective by identifying *Chlamydia* species/type-specific B-cell epitopes. In the course of this investigation, we made several observations that are noteworthy as important criteria for the development of peptide antigens based assays, for other pathogens as well. Dominant B-cell epitopes are a prerequisite for robust assays with high signal to noise ratio, and such dominant epitopes were present only in the immunodominant proteins of *Chlamydia* spp. The outer membrane protein A (OmpA), also known as MOMP (major outer membrane protein), which also determines the serovars in *Chlamydia* spp., is the most immunodominant chlamydial protein. In our hands, OmpA peptides were among those peptides that produced the highest signals in our chemiluminescent ELISA. Interestingly, peptides from several other more conserved proteins showed as strong reactivity as OmpA peptides. Strongly reacting OmpA peptides are suitable only for detection of serovar-specific antibody in several *Chlamydia* spp. In

contrast, strongly reactive peptides from other, more conserved proteins have the advantage of detection of antibodies against most of the serovars of a chlamydial species by using just a single or few peptides, due to high sequence conservation within the species but high polymorphism within the genus *Chlamydia*.

Chlamydia spp. are ubiquitous bacteria which infect a wide range of hosts, but only the natural (homologous) hosts in which a *Chlamydia* spp. establishes endemic infection usually maintain high anti-*Chlamydia* antibodies. However, transient natural infections with different *Chlamydia* spp. are very common, and they create a specificity problem in validation of serological assays by *Chlamydia*-positive natural host sera. Therefore, we raised hyper-immune sera in laboratory mice, heterologous hosts for all *Chlamydia* spp. except for *C. muridarum*, to avoid any potential presence of antibodies against an unknown chlamydial species/strain. Using these mouse sera, we identified suitable peptides for specific detection of anti-*Chlamydia* antibodies. Then, we confirmed with these peptides the suitability of our assay for detection of anti-chlamydial antibodies from natural *Chlamydia* hosts such as calves and cows. Importantly, B-cell epitopes that were immunodominant for *C. pecorum* infections in mice were also dominant in the bovine species, the homologous host, therefore we termed them host-independent epitopes. However, numerous peptides from additional proteins that did not react with the heterologous mouse sera did react strongly with bovine sera. This result suggests that natural homologous hosts produce antibodies against a much wider set of host-dependent epitopes, presumably because a chlamydial species that is optimally adapted provides better antigenic stimulation in the homologous than in a heterologous host. Therefore, host-independent epitopes can be used for specific confirmation of chlamydial infections,

while antibodies against host-dependent epitopes may be effective markers in studies of disease phenotype after chlamydial infection. Such studies, for example on fertility prognosis or prediction of pathological immune responses after chlamydial infections, are important for many host species. We have already tested all 900 peptides of this investigation with pooled sera of 135 patients with *C. trachomatis* infections, and identified ~25 host-dependent epitopes from additional 15 proteins, which we are now testing with individual sera as predictors of differential outcome phenotype.

Currently, 12 species of *Chlamydia* recognized, and several of these have numerous serovars. A high efficacy of *in silico* B-cell epitope prediction was essential for successful completion of this investigation since our objective was to develop assays for all chlamydial species, by scanning whole proteomes of *Chlamydia* spp. We used several algorithms for B-cell epitope prediction, but found them unsatisfactory, and therefore iteratively improved B-cell epitope prediction methodology. The major improvement was the use of long 16-30aa peptide antigens that optimally fit the size of antibody paratopes. Typically used 7-12aa peptide antigens are frequently only weakly reactive or non-reactive, simply because they are too short to occupy the full antibody paratopes. We found a systematic problem that the B-cell epitope prediction research community had neglected, and that compromised evaluation of all B-cell epitope prediction algorithms. Such algorithms were trained, evaluated, or developed with datasets from public databases such as IEDB (Immune Epitope Database). Because of the intense focus on short epitopes, 6-11aa peptide sequences are vastly overrepresented in these datasets. When we removed these short sequences and evaluated more than 100 primary scales for prediction of protein properties as well as standard algorithms for B-cell epitope

prediction, we found that protein disorder tendency is the best predictor. This confirmed the results that we obtained with our own data by testing of chlamydial B-cell epitopes predicted by the IUPred-L, the most efficient algorithm for calculating protein disorder tendency. In addition, other scales such as hydrophilicity, solvent accessibility, beta-turn, coils/loops, and evolutionary sequence polymorphism predicted B-cell epitopes efficiently if a 25aa running average score was used instead of the standard 5-8aa window score. Taken together, these findings clearly suggest that minimum of 16 and up to 30aa B-cell epitope residues may be required for optimum fit with antibody paratopes, resulting in high-affinity binding. This study has obvious implications in designing antibody assays and vaccines by providing optimum length peptide antigens for maximum antibody binding.

In summary, the peptide antigens found in this investigation produce high and absolutely species-specific signals in a robust ELISA format. In addition to *Chlamydia* species-specific peptides, type-specific peptides for strains of *C. suis*, *C. pecorum*, *C. trachomatis*, and *C. psittaci*, and genus-specific peptides were identified. Because of the simplicity and robustness of these peptide ELISAs, molecular serology of chlamydial infections may now become accessible for non-specialist laboratories. Such serological assays would also have the added advantage of retrospectively capturing the history of chlamydial infection, rather than stochastically sample a single point in time as PCR detection does. These peptide antigens may also serve in multiplexed assays such as microarrays or fluorescent bead assays. In fact, collaborators in Germany have already successfully converted the single peptide/well format into a microarray that can accommodate up to 200 different spotted covalently bound peptide antigens. We

anticipate that these peptide ELISAs have the potential to vastly improve chlamydial serology, in particular *Chlamydia* species-specific serological diagnosis. By allowing serological dissection of multi-species chlamydial infections, they will further the understanding of chlamydial diseases in retrospective epidemiological investigations of human and animal chlamydial infections. In a wider context, the present methodological approach of epitope identification has a great potential of being useful in both immunological research and laboratory diagnosis of other microbial infections.