DIPLOMARBEIT

Titel der Diplomarbeit

„Identification and Characterization of novel type III secreted proteins in *Chlamydia* Infection."

Verfasserin

Lena Gehre

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt:	A 441
Studienrichtung lt. Studienblatt:	Diplomstudium Genetik und Mikrobiologie
Betreuerin / Betreuer:	Univ.-Prof. Dr. Thomas Rattei

Acknowledgement
Danksagung

I want to thank my two supervisors, Agathe Subtil and Thomas Rattei. Both of them gave me the possibility to make my "Diplomarbeit" (ou stage de M2) here in Paris at the Institut Pasteur. Special thanks to Agathe for the good supervision of my practical work and special thanks to Thomas for all the patience he showed, even with all the administrative things. They helped me so much with this work, putting a lot of energy, time and good advices in it.

I also want to thank Alice Dautry-Varsat for giving me the possibility to work in her unit.

I am grateful for all the nice support I got from the friendly people here in BIC (not to be forgotten the Utrafs), for listening to all my "petites questions" – merci pour toutes les réponses patientes. Un grand merci surtout à Stéphanie, François et Miriam, pour tous les moments marrants et informatifs (avant tout informatifs évidemment).

Ein riesen Dankeschön geht an meine Familie, meine beiden Geschwister Florian und Julia, und vor allem an meine Eltern, die mich bei meinem doch hin und wieder etwas unsesshaften Studienverlauf immer unglaublich unterstützt haben, moralisch sowie umzugstechnisch. Danke, danke, danke.

Danke auch an Thea und Toni, für sehr vieles.

Index

IV

List of figures and tables

List of abbreviations

AB................. Abberant Body

APS............... Ammonium Persulfate

CBD.............. Chaperone-binding Domain

Cya................ Calmodulin-dependent Adenylate Cyclase

EB.................. Elementary Body

EDTA............ Ethylenediaminetetraacetic Acid

fla-TTS.......... Flagellar Type III Secretion

Glucose-1-P... Glucose-1-Phosphate

IM.................. Inner Membrane

LB.................. Luria Broth Base

LGT............... Lateral Gene Transfer

PCR............... Polymerase Chain Reaction

PPi................. Pyrophosphate

PVDF............ Polyvinylidene Fluoride

RB................. Reticulate Body

SDS............... Sodium Dodecyl Sulfate

SDS-PAGE... Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SPI................. *Salmonella* Pathogenicity Island

SS.................. Soluble Starch Synthetase

TAE............... Tris Acetate EDTA Buffer

TEMED......... Tetramethylethylenediamine

Tris................ Tris(hydroxymethyl)aminomethane

TTS............... Type III Secretion

TTSS.............. Type III Secretion System

1. Introduction

1.1 *Chlamydiales*

Chlamydiales are obligate intracellular gram-negative bacteria that all share a unique developmental cycle. Their host range varies from animals (like mammals or insects) to protozoa. Different diseases are caused by *Chlamydiae*, still not a lot is known about this organism. Its obligate intracellular development is challenging for researchers and no techniques are available to easily manipulate *Chlamydiae* genetically, therefore our understanding of *Chlamydiae* biology is still very fragmented.

1.1.1 The order *Chlamydiales*

So far, the order *Chlamydiales* is the only one in the class *Chlamydiae*. This order seems to be phylogenetically separated from other eubacteria, having proteins that show only a low level of homology with known proteins (Subtil and Dautry-Varsat, 2004). The tree in figure 1 has been proposed by Horn (Horn, 2008). It has only recently been suggested to divide the family *Chlamydiaceae* into two genera, *Chlamydia* and *Chlamydophila* (former *Chlamydia*). Genetic studies support this new distinction as well as biochemical markers. Only bacteria of the genus *Chlamydia* (*C. trachomatis*, *C. suis* and *C. muridarum*) contain detectable amounts of glycogen particles, this has never been seen for *Chlamydophila* (Chiappino et al., 1995; Rogers et al., 1996). However, the proposal for a division of the genera has largely been critisized by the chlamydial community (Stephens et al., 2009). So far, both (the two genera of *Chlamydophila* and *Chlamydia* as well as *Chlamydia* as one genus comprising all the members) are common. In this work the two genera *Chlamydia* and *Chlamydophila* suggested also by Horn (Horn, 2008) will be applied.
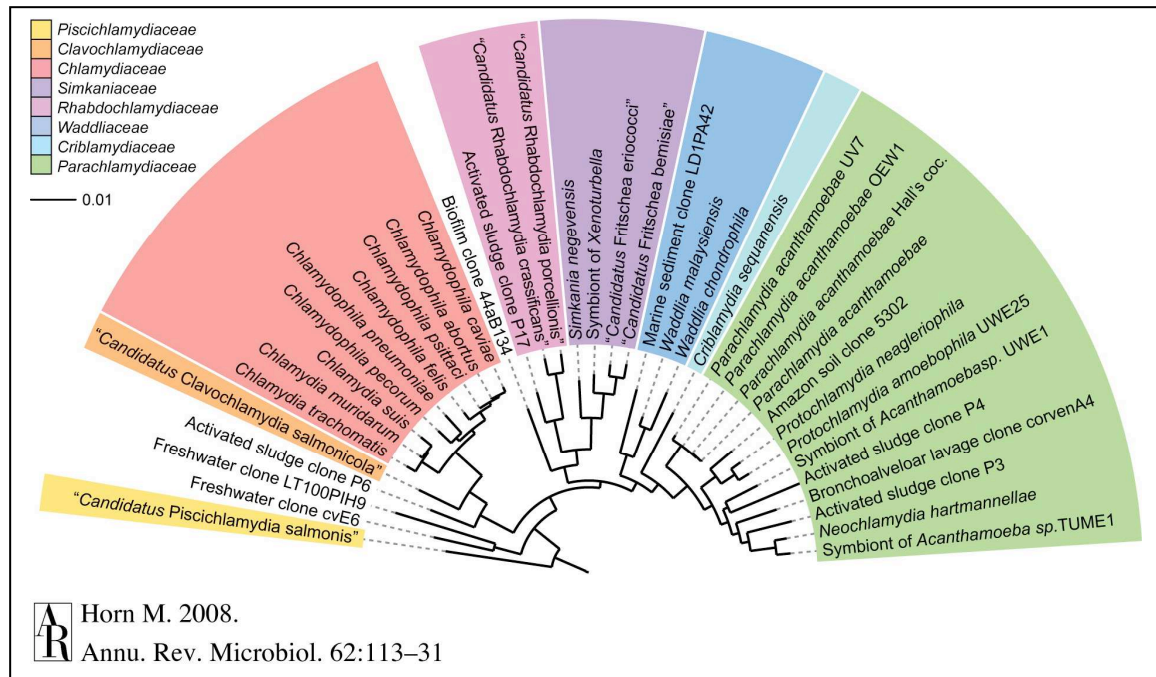
Figure 1: Genetic structure of the phylum *Chlamydiae*. Calculated by 16S rRNA sequence data. Adapted from Horn (2008)

1.1.2 Diseases and Hosts

However, diseases and host range (table 1) do not seem to be linked to the different genera. The main human pathogens comprise *C. trachomatis*, *C. pneumoniae* and *C. psittaci*. *C. trachomatis* is a major cause of blindness and sexually transmitted diseases. Depending on its serovar it leads among others to trachoma, lymphogranuloma venereum and urethritis. A substantial proportion of the infection is asymptomatic (Stamm, 1999), but persistant infections can cause infertility or ectopic pregnancies. *Chlamydiae* can enter a non-infectious but viable state called persistence when exposed to stress. *Chlamydiae* undergoing the persistent state do not ressemble to normal forms of *Chlamydiae*. Their structure seems to be enlarged, irregular and less electron-dense. Although these so-called ABs (abberant bodies) have been demonstrated in vivo, it is still not determined whether *Chlamydiae* undergo a conversion to ABs in order to establish chronic host infections (Schoborg, 2011). The most frequent illnesses linked to *C. pneumoniae* are pneumonia and bronchitis (with an overall share of 10 and 5%, respectively) (Kuo et al., 1995). Connections to artherosclerosis are under investigation. Even though the main host of *C. psittaci* are birds, it is also infective for humans and causes psittacosis, a life-threatening pulmonary infection (Gregory and Schaffner 1997).

Parachlamydiaceae comprise intracellular symbionts of free-living amoebae, and are therefore often designated as environmental *Chlamydiae* (Hayashi et al., 2010; Leitsch et al., 2010; Collingro et al., 2005). Interestingly, recent studies showed that a large number of sequences in the genomes of *Parachlamydia acanthamoeba* and *Candidatus Protochlamydia amoebophila* are related to genes in plants (Collingro et al., 2011).

C. abortus	*C. psittaci*	*C. felis*	*C. caviae*	*C. pecorum*	*C. pneumoniae*
mammals	birds	cats	guinea pig	mammals	humans

Chlamydiales	*C. trachomatis*	*C. suis*	*C. muridarum*	*P. acanthamboeba*	*C. P. amoebophila*
Typical Host	humans	swine	mice, hamsters	acanthamoeba	acanthamoeba

Table 1: Typical hosts of *Chlamydiales* species.

1.1.3 Developmental cycle

Despite their different host range, *Parachlamydiaceae* and *Chlamydiaceae* possess a common biphasic developmental cycle (figure 2). Infectious particles called elementary bodies (EBs) enter host cells by a to date still obscure mechanism. It has been suggested, that *Chlamydiaceae* (similar to *Shigella* and *Salmonella*) inject effector proteins into epithelial cells prior to invasion. This culminates in remodeling of the host's actin cytoskeleton at the site of entry and finally leads to the pathogen's uptake (Dunn and Valdivia, 2010). Once in the cytoplasm, they build up parasitophorous compartments called inclusions, where they develop and convert into metabolically active reticulate bodies (RBs). RBs are capable of vegetative growth and of division by binary fission. At the last step of their developmental cycle they reconvert into EBs and are released. Depending on the strain, this can take up to 72 h (Fields and Hackstadt, 2000; Subtil et al., 2001). In most of the literature EBs are referred to as being metabolically inert and often compared to a spore-like stadium. However, it has been shown by Haider et al. (Haider et al., 2010) that EBs of *Parachlamydia acanthamoeba* and *Chlamydia trachomatis* are metabolically active for a restricted timespan after being released from their host cell implying the supply of appropriate nutrients.

Chlamydiae do not only interfere with the host cell prior to or during their internalization, they also hijack the cytoskeleton during infection (Carabeo et al., 2002), cause alterations of the host cell's signal-transduction pathways, repress apoptosis for the duration of the developmental cycle and finally induce cell death to release chlamydial progenies (Byrne and
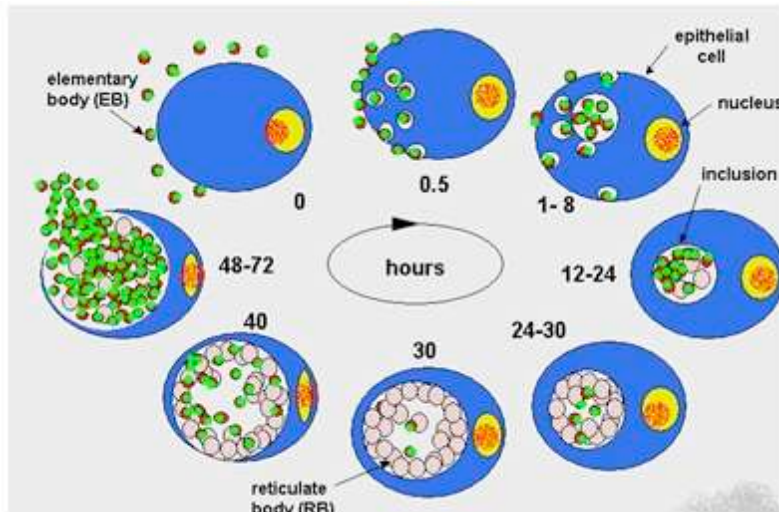
Figure 2: Chlamydial developmental cycle. *Chlamydiaceae* undergo a biphasic developmental cycle which varies between a metabolically inactive form, the elementary body (EB) and a metabolically active form, the reticulate body (RB). EBs invade the epithelial host cell and convert into RBs within a chlamydial-specific compartment, the inclusion. RBs divide by binary fission, reconvert back to EBs and are released at the end of the developmental cycle. Duration depends on the species. (Pandey et al., 2009)

Ojcius, 2004). It has been shown that *Chlamydiae* possess a functional type III secretion system (TTSS) from very early stages on. Clifton et al. (Clifton et al., 2004) discovered the first early TTSS secreted protein, the translocated actin-recruiting phosphoprotein (Tarp). It has been suggested that the TTS machinery is already present and "preloaded" with Tarp on EBs in order to mediate engulfment of the pathogen upon contact with a host cell (Clifton et al., 2004). Secretion of proteins (effectors) can be targeted to the inclusion membrane as well as to the cytoplasm of the host cell. Unpublished data of Subtil (Institut Pasteur, Paris) suggest secretion of effectors into the inclusion lumen, but their function is still not clear. Identification of TTS effectors is not an easy task. Its attempt through direct functional analysis is strongly hampered by the genetical intractability of *Chlamydiae*. Interesting assays have been developed to overcome this problem, and two of them (a bioinformatical approach and a heterologous test of secretion) will be detailed below.

1.2 Type III Secretion System

1.2.1 Background

Protein secretion is indispensible for the bacterial development and survival. Not only it is necessary to build up and modify the bacterial membranes, but it is also a key step of bacterial virulence for pathogens and symbionts. Defects in the ability of secretion can render pathogenic bacteria non-pathogenic. Six different secretion (type I-VI) systems have been identified in gram-negative bacteria so far. All of them differ in their structure as well as in their mode of translocation. For instance, the TTSS is able to inject effectors directly into the host cell, whereas the Sec pathway transports proteins into the periplasmic space or the outer

membrane (Tseng et al., 2009). The TTSS is a very old type of secretion, which seems to have been present in bacteria already over 1 billion years ago (for instance, in *Chlamydiae*) (Cavalier-Smith, 2006; Yoon et al., 2004). Interestingly, in some species the genes encoding structural proteins for the TTS apparatus can be on a dedicated plasmid (eg. *Shigella*) which allows frequent LGT of the plasmid to a new species (Buchrieser et al., 2000). In other bacteria, as for example *Salmonella* and *Chlamydiae*, these genes are located on the bacterial chromosome. The very well investigated *Salmonella* encode the genes on a so-called *Salmonella* pathogenicity island (SPI), also prone to LGT (Shea et al., 1996). However, none of it applies to *Chlamydiales*. Their genes encoding the structural proteins of the TTS apparatus are found in three distinct conserved genomic clusters (Stephens et al., 1998).

Evolutionary studies suggest the conclusion that the contemporary TTSS has either evolved from a bacterial flagellum (termed flagellar Type III Secretion System) (Nguyen et al., 2000) or coevolved with it from a common ancestor (Aizawa, 2001). All *Chlamydiaceae* (not *Parachlamydiaceae*) that have been examined so far still have several of these fla-TTS genes, even though the bacteria are non-motile. What all chlamydial genomes encode are the structural proteins building up the TTS injectisome (a molecular "needle-complex" consisting out of 20-25 proteins), the translocator apparatus (the base and a rod-like structure) and chaperones which are required for the proper secretion of effectors (Peters et al., 2007; Betts-Hampikian and Fields, 2010). This needle-complex can cross three different membranes: the bacteria inner membrane, the bacteria outer membrane and the plasma membrane of the host cell. For given reasons not all of the TTS subunits have been undoubtedly identified in *Chlamydiae* up to date, but comparison to *Yersinia-* or *Salmonella*-machineries can give a good structural and functional idea (figure 3, table 2).

Component	Predicted structure or function
CdsC	Component of outer membrane ring
CdsD	Integral inner membrane ring protein
CdsF	Needle subunit protein
CdsJ	Predicted lipoprotein; spans the periplasmic space
CdsL	ATPase inhibitor
CdsN	ATPase
CdsQ	Basal body protein; required for structural assembly; homolog of motor-switch protein of fla-TTS
CdsR	Integral IM protein with multiple transmembrane domains
CdsS	Integral IM protein with multiple transmembrane domains
CdsT	Integral IM protein with multiple transmembrane domains
CdsU	Integral IM protein; by analogy with fla-TTS, associates with CdsJ, CdsN ATPase, and its putative negative regulator, CdsL
CdsV	Integral IM protein; belongs to the Flagellar/Hr/Invasion Protein Export Pore (FHIPEP) protein family; highly conserved amino terminus has 6–8 predicted transmembrane domains; large, less conserved, hydrophilic C terminus, predicted in cytoplasm where it might interact with other TTS proteins

CopB	Translocator protein	
CopD	Translocator protein	
CT584	Putative needle tip protein	

Table 2: Components of the chlamydial TTS machinery and their function. IM = inner membrane. (adapted from Peters et al., 2007; Betts-Hampikian and Fields, 2010)
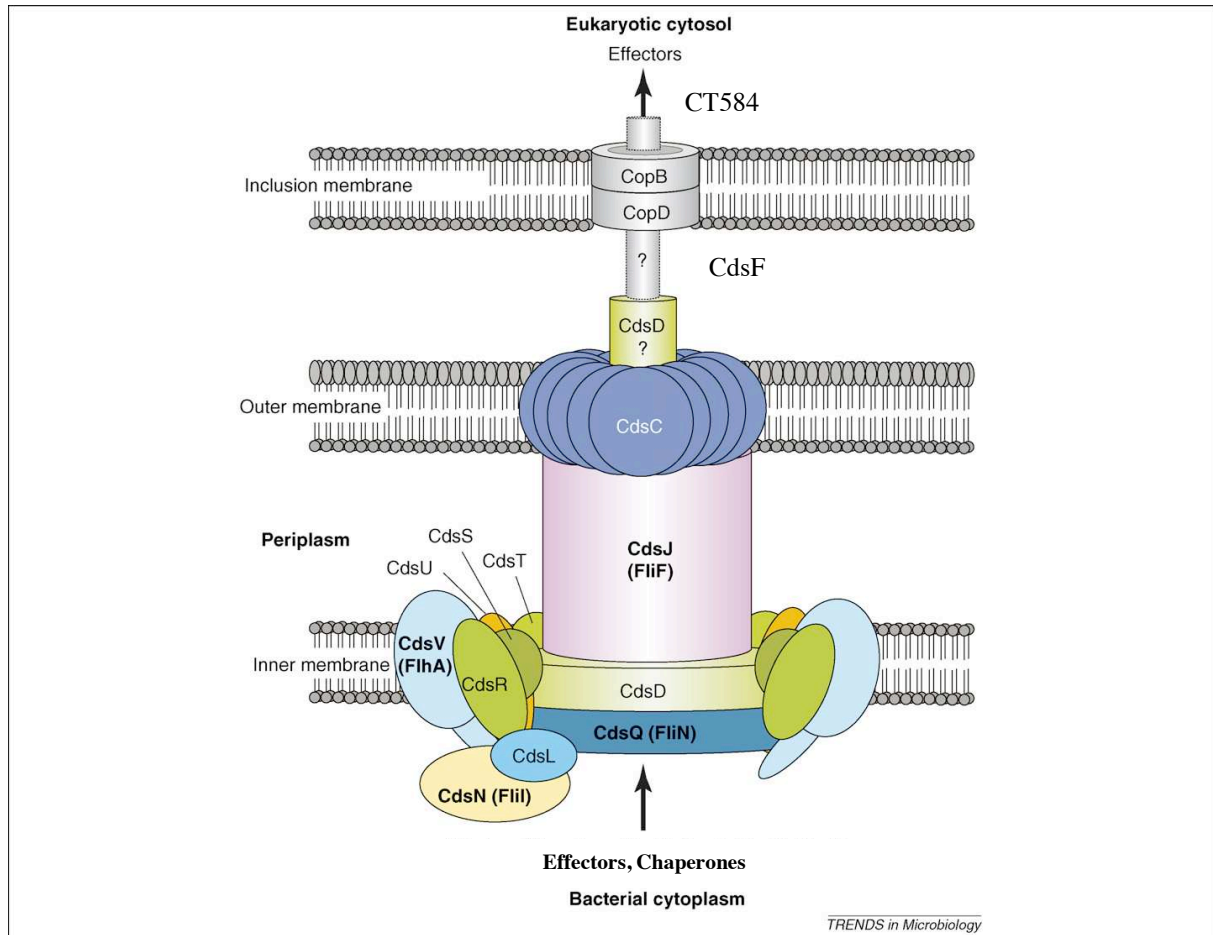
Figure 3: Chlamydial TTS apparatus. Putative structure obtained by comparison with *Yersinia* and *Salmonella*. Depicted in parentheses are flagellar paralogs. (Peters et al., 2007; Betts-Hampikian and Fields, 2010)

1.2.2 Secretion Signal and Effectors

The molecular recognition of effectors by the TTS machinery is elusive. Still, the identification of a signal that selects for TTS substrates would help to identify effector proteins. The detection of novel effectors is an intriguing attempt, since they constitute putative virulence factors whose characterisation would lead to a better understanding of *Chlamydiae* and their infections.

What could this signal consist out of? In all secretion systems examined so far it has been revealed that the protein sequence is harboring the signal responsible for a functional secretion (Ramamurthi and Schneewind, 2002). If this is also valid for the TTSS is highly controversial. Researches argue if the information for TTS is in the mRNA- (mRNA signal

hypothesis) or in the protein-sequence (peptide signal hypothesis); both theories and their mechanisms are depicted in figure 4. Evidence for both exist but seem to be somehow contradictory. Anderson and Schneewind (Anderson and Schneewind, 1997) observed that frame-shift mutations of amino termini do not abolish transport of effectors, supporting strongly the mRNA signal hypothesis. The control test revealed opposite results: Lloyd et al. (Lloyd et al., 2001) replaced 17 nucleotides within the first 10 codons of a *Yersinia pestis* effector resulting in a polypeptide, which encodes the same amino acid sequence as the wild-type one. Interestingly, the secretion signal remained functional. Similar results have been obtained for a *Salmonella* protein (Rüssman et al., 2002). However, it remains elusive if the signal lies in a specific secondary structure of the mRNA or in specific features of a proteinaceous signal, eg. water accessibility states, amino acid composition, secondary structures as coils, helices or strands.

Both hypothesis could also go hand in hand. It has been suggested by Ramamurthi and Schneewind (Ramamurthi and Schneewind, 2002) that some distinct property of the mRNA recruits ribosomes to the proximity of the TTS machinery, where translation into an amino acid sequence takes place. Other properties of the nascent amino acid sequence then engage the TTSS.

Chaperone dependent processes also play an important role (Cheng et at., 1997). Several effectors associate with specific chaperones at the chaperone-binding domain (CBD) and a lack of its specific chaperone leads to reduced or abolished secretion (Lee and Galan, 2004). Chaperones may prevent the premature association of secreted proteins in the bacterial cytoplasm that would target them for premature degradation (Galàn and Collmer, 1999). Another function is the triggering of effectors through the narrow TTS conduit (Stebbins and Galàn, 2001). In any case, the chaperone seems to remain inside the bacteria during secretion of the effector into the host cell and is released by ATPase activity, which was shown for at least one case (Akeda and Galàn, 2005). Cheng et al. (Cheng et al., 1997) reported a signal, the CBD, located within amino terminal amino acids 15-100 being important (but not indispensible) for functional secretion. This CBD-signal seems to be conserved for groups of effectors sharing the same chaperone.

An example of a unique group of effectors is the family of Inc proteins in *Chlamydiae*. These proteins were shown to be secreted by a TTSS and to insert with their hydrophobic domain into the inclusion membrane (Subtil et al., 2001). The translocated actin-recruiting phosphoprotein (Tarp) is another substrate of the TTSS playing a role in actin-recruitment to the inclusion (Clifton et al., 2004) and to the site of entry of the pathogen. Since actin-
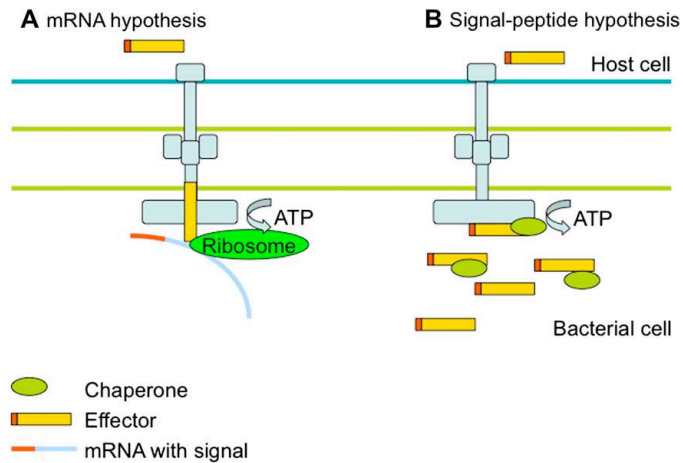
Figure 4 : Schematic illustration of the two hypotheses of the location of the amino terminal secretion signal: mRNA-based (A) and peptide-based (B). In (A), the effector mRNA, which carries the signal, is synthesized into the TTSS during transport. In (B), the effector is translated in the bacterial cytosol and recognized by a peptide born amino terminal signal. Chaperones play different roles, as enhancing signals or holding the protein in an unfolded, transportable state. Adapted from Arnold et al. (2010)

recruitment is necessary at early stages of internalization Clifton (2004) conceived that the needle-complex could be preloaded with this protein prior to infection of the host by the EB.

1.3 In silico Prediction of Effectors

In silico predictions of effectors can generate candidate lists that can then be subjected to in vivo tests. One approach is to screen for homologs of already known effectors between different species, based on the hypothesis that they either have a common ancestor or are distributed through LGTs (Tobe et al., 2006). A severe drawback is that some distinct effector families are unique to specific species or that so far undetected families will be skipped.

The TTSS as well as the effectors underlie a strictly regulated concerted activation. Next to the previously mentioned chaperones, which play an important role as control mechanism for temporal order, transcriptional control is also involved (Valls et al., 2006). This may include several regulators and sigma factors that bind to their substrate (gene or protein; both post-transcriptional and post-translational is possible). Once these regulators have been identified, identification of their binding sites can lead to the discovery of putative effectors. A complication of this approach is the often degenerated and unknown nature of these binding sequences.

Another approach makes use of putative chaperone-effector pairs clustered on the genome (Panina et al., 2005). However, it is likely that there exist also chaperone-independent substrate recognitions and that not all organisms possess clusters of effectors and chaperones.

Many experimental data support the hypothesis that the amino terminal sequence of effectors is recognized as a signal for TTS. Therefore, deciphering this amino terminal signal appears like the easiest way towards the identification of TTS substrates. As it has already been mentioned, the amino termini do unfortunately not uncover any obvious sequence similarities

that are common for all known effectors (Arnold et al., 2009). Consequently, a simple alignment of the sequences is not sufficient. However, machine learning approaches unravel relationships between amino termini and help to identify TTS features that are still puzzling. The limiting point is the lack of data they need to be fed with. EffectiveT3 was the first prediction software developed for type III secreted proteins (Arnold et al., 2009; Jehl et al., 2011). 100 amino termini of proteins that were known to be type III effectors were compared, and an enrichment of serine, threonine and proline was noted for animal effectors (only serine in plant effectors), while leucine residues were underrepresented in the amino termini of both groups. A binary classification algorithm was trained to recognize effectors on the basis of features as frequencies of amino acids, amino acid properties and short combinations of them (Arnold et al., 2009). This approach helps to conduct large-scale screens of prokaryotic proteomes and is therefore a valuable tool.

Cazalet et al. (Cazalet et al., 2004) showed that eukaryotic-like proteins are present in the genome of *Legionella pneumophila*, a facultative intracellular parasite. Ankyrin-rich regions typically found in eukaryotes are for example overrepresented in type IV secreted effectors (Lurie-Weinberger et al., 2010). Proteins with eukaryotic-like domains are likely to interfere with the eukaryotic host cell and may thus represent virulence factors. The software "Effective" also takes this into account and detects all protein domains that occur in eukaryotes, pathogens and symbionts but not or only rarely in non-pathogens. However, large-scale in vitro or in vivo validations of this elegant in silico approach have still not been conducted. There are only single examples of pathogens where eukaryotic-like domains were detected in secreted proteins. In this work we will perform type III secretion tests on candidates that were chosen by the last approach, the "eukaryotic-like domain approach" in order to evaluate its potential to identify TTS substrates.

A short overview of useful methods for in silico predictions of effectors has been given here. However, all these methods help to create lists of candidates, but none of them are generally applicable.

1.4 In vivo Test of Secretion

It has largely been explained that sequences of virulence effectors are very diverse. However, the TTS machinery seems to be quite conserved, even between very distinct species (Rosqvist et al., 1995; Frithz-Lindsten et al., 1998). Based on this observation, the ability of TTS effectors to be secreted by a heterologous TTSS of another species was tested. This is of particular interest for bacteria such as *Chlamydiae* that cannot be manipulated genetically.

Fields and Hackstadt (Fields and Hackstadt, 2000) showed that the chlamydial protein CopN can be secreted by *Yersinia enterocolitica*. Ho and Starnbach (Ho and Starnbach, 2005) confirmed the translocation by a type III secretion system of proteins of *C. trachomatis* in *Salmonella enterica*. Subtil et al. proved the validity of heterologous TTS of chlamydial effectors in *Shigella flexneri* with a large-scale screen where numerous positive and negative controls were included (Subtil et al., 2001 and 2005). In this secretion assay chimeras consisting out of the first 20 amino terminal amino acids of chlamydial candidate effectors and a reporter gene were transformed into two different *Shigella flexneri* strains. *ipaB* is a mutant secreting effectors in a deregulated manner (Ménard et al., 1994), whereas the TTSS of *mxiD* is totally impaired (Allaoui et al., 1993). Secretion of chimeras was tested on both mutants allowing for differentiation between proteins that are secreted by type III or by one of the other pathways.

1.5 Tripartite Symbiosis between Heterotrophic Organism, Cyanobiont and *Chlamydiae*

1.5.1 Hypothesis

It is widely accepted that the origin of photosynthetic organelles (plastids) in eukaryotes occurred via endosymbiosis of a cyanobiont. About 1 billion years ago a eukaryotic heterotrophic cell entered into a symbiontic relationship with a prokaryotic organism and instead of phagocytosing it, it took benefit out of the arising symbiosis (Cavalier-Smith, 2006; Yoon et al., 2004). Another case of endosymbiosis is the emergence of mitochondria, which are derived from α-proteobacteria (Gray, 1993). It has been suggested that the nucleus developed through endosymbiosis of bacterial or archaebacterial partners as well (Lake and Rivera, 1994). Even though there are some examples of this primary endosymbiosis, they stay limited. Selection pressure clearly favours organisms, which obtain a new source of energy as for example photosynthetic organelles. Still, the establishment of plastids seems to have been unique. Progenitors of Archaeplastida (comprising red algae, glaucophyta and green plants, all of them possess a functional or remnant plastid) and cyanobionts were quite abundant and thus had a lot of possibilities for physical contact. Why is it then such a rare event?

It has been suggested by Stephen Ball (University of Lille, France) that a key step of this metabolic symbiosis was the export of the metabolite ADP-glucose from the cyanobiont to the eukaryotic host cell, where it could be stored as polysaccharides, meaning that the host cell would benefit from the prokaryotes ability to photosynthesis. Still, ADP-glucose is a bacterial-specific metabolite, whereas the eukaryotic equivalent is UDP-glucose. Hence,

eukaryotes were not able to further use this metabolite. How can a symbiosis with a photosynthetic prokaryote then have been of benefit for the ancestor of Archaeplastida, when the latter was not able to use this new source of energy?

Different scenarios are possible. The archaeplastidal progenitor could have acquired appropriate genes from other prokaryotic organisms by lateral gene transfer (LGT) prior to or at the time of establishment of a heterotrophic/cyanobacterial symbiosis. Since these gene products gained a sudden function in the host cell, selection pressure would have prevented their loss.

Another possibility would be an adaptation of the eukaryotic genes to the new source. This includes several mutations and thus a low probability, but considering the idea that the uptake of a cyanobiont occurred frequently over a long time, selection pressure would finally strongly favour the ones that succeeded.

Previous phylogenetic studies showed that a surprisingly high amount of archaeplastidal genes have their closest homologs in *Chlamydiae* (21 according to Huang and Gogarten, 2007; 55 according to Moustafa et al., 2008), more precisely in *Candidatus Protochlamydia amoebophila* UWE25 and *Parachlamydia acanthamoeba* UV7. According to current phylogenetic analysis, Archaeplastida and *Chlamydiae* do not possess a direct common ancestor; the presence of homologs in their genome is thus not a sign for a close relationship. Different scenarios to explain this occurrence are conceivable:

1) The cyanobiont received chlamydial genes by LGT prior to the establishment of the symbiosis (or vice-versa).
2) LGT occurred between *Chlamydiae* and Archaeplastida at a later stage than the plastid endosymbiosis. Both directions of gene transfer are thinkable.
3) Infected insects played the role as vectors introducing chlamydial genes into plants.
4) A third party (eg. *Chlamydiae*) was involved at the time of establishment of the endosymbiosis (hypothesis of a tripartite symbiosis, supported by Prof. Steven Ball, University of Lille, France, unpublished).

Interestingly, chlamydial homologs in the archaeplastidal genome often contain a plastid-targeting signal (Huang and Gogarten, 2007; Moustafa et al., 2008), promoting the idea of an ancestral evolutionary relationship between cyanobacteria (plastids) and *Chlamydiae*. The chlamydial genes found in archaeplastidal genomes clearly have bacterial nature; hence, gene transfer has to have occurred from prokaryote (chlamydial or plastid genome) to eukaryote.

The hypothesis that a transfer was directed from the cyanobiont versus *Chlamydiae* can be declined by today's state of knowledge that cyanobacterial homologs are distinct from chlamydial homologs that are found in archeaplastidal genomes. Additionally, chlamydial proteins are much more related to proteins of other bacteria than to cyanobionts and some of the chlamydial homologs are not found in cyanobacteria. Consequently, this would not explain the occurrence of close homologs between archaeplastidal and chlamydial genes. These reasons suggest gene transfer from *Chlamydiae* to Archaeplastida or the archaeplastidal progenitor (Huang and Gogarten, 2007).

The hypothesis that LGT of chlamydial genes gave the archaeplastidal ancestor the ability to metabolise cyanobacterial products implies that infection with *Chlamydiae* and the uptake of the cyanobiont happened more or less at the same time within the same host cell. This cell would have been weakened by the parasitic *Chlamydiae*, but would also acquire a new source of energy, the cyanobiontic ADP-glucose, which could rescue the cell of the deletorious effects. The missing link would be enzymes or the respective genes that are contributed by *Chlamydiae* to render a usage of the prokaryotic product for the host cell possible. If this is to be true, *Chlamydiae* introduced these genes to the archaeplastidal ancestor by LGT. Phylogenetical analysis of plant-homologs could possibly clarify this point. Moreover, secretion of these enzymes by *Chlamydiae* into the host cell prior to LGT would strengthen this hypothesis by extending the time span for a stable integration into the genome. In this scenario, the host cell would incorporate a cyanobiont and additionally be infected by *Chlamydiae*, which secretes prokaryotic enzymes for the ADP-glucose metabolism in order to hijack the host metabolism. The host cell having the cyanobacterial energy source would benefit from this tripartite symbiosis, build up a stable endosymbiosis with the cyanobiont and integrate chlamydial genes into its genome. We will further examine the respective metabolisms of host cell and *Chlamydiae* in order to reveal which enzymes could have played a role in this scenario. Furthermore, the previously mentioned test of secretion could give information whether those enzymes are putative effectors and could have fulfilled the function suggested in our scenario.

1.5.2 Glycogen and Starch Metabolism

Today's green plants (Chloroplastida, green algae and land plants) use as storage polysaccharide starch, which consists out of 20-25% amylose and 75-80% amylopectine. The main storage equivalent in animals, fungi, bacteria, archaea and non-chloroplastidal protists is glycogen, a closely related but more branched polysaccharide. Interestingly, starch is

produced out of ADP-glucose, as is prokaryotic glycogen. In contrast, the base for eukaryotic glycogen is UDP-glucose.

1.5.2.1 Glycogen Metabolism in Bacteria and Eukaryotes

Figure 5 and 6 show the different biochemistry of glycogen metabolism in bacteria and eukaryotes (Preiss and Romeo, 1994; Ball et al., 2011). Both need an activated form of glucose, glucosyl-nucleotide, which is generated from glucose-1-P. This nucleotide-sugar (UDP-glucose for eukaryotes, ADP-glucose for bacteria) is synthesized through the action of UDP-glucose pyrophosphorylase and ADP-glucose pyrophosphorylase, respectively. A subsequent step of elongation by glycogen synthases leads to a transfer of the activated glucose to a pre-existant chain by splitting of the nucleotide. By cleaving some of the α-1,4 linked chains and substituting the bonding by a α-1,6 linkage, in both cases branching enzymes lead to a delinearization of the macromolecule. The product is the storage polysaccharide glycogen.

Glycogen degradation passes in both cases by an intermediate step of phosphorylase limit dextrin, which is produced by the action of glycogen phosphorylases, releasing one glucose-1-P molecule. Phosphorylase limit dextrin serves as a substrate for debranching enzymes, direct debranching enzymes in bacteria and indirect debranching enzymes in eukaryotes. In bacteria, this leads to a release of maltotetraose, which can (among others) be used for reactions catalyzed by α-1,4 glucanotransferases. One possible subsequent pathway results in the production of a glucose-molecule and by action of maltodextrin phosphorylase in the release of a newly synthesized glucose-1-P.

In eukaryotes, the next step differ between fungi/animals and amoebozoa. The latter seem to use an enzyme (transglucosidase, not depicted in figure 6) which belongs to the same family as the bacterial α-1,4 glucanotransferases. In any case, indirect debranching enzymes attack the outer chains generated on the glycogen particle and transfer them to neighboring glycogen chains, meanwhile releasing glucose residues. The newly synthesized long external chains can be further recessed by glycogen phosphorylases, producing again glucose-1-P.
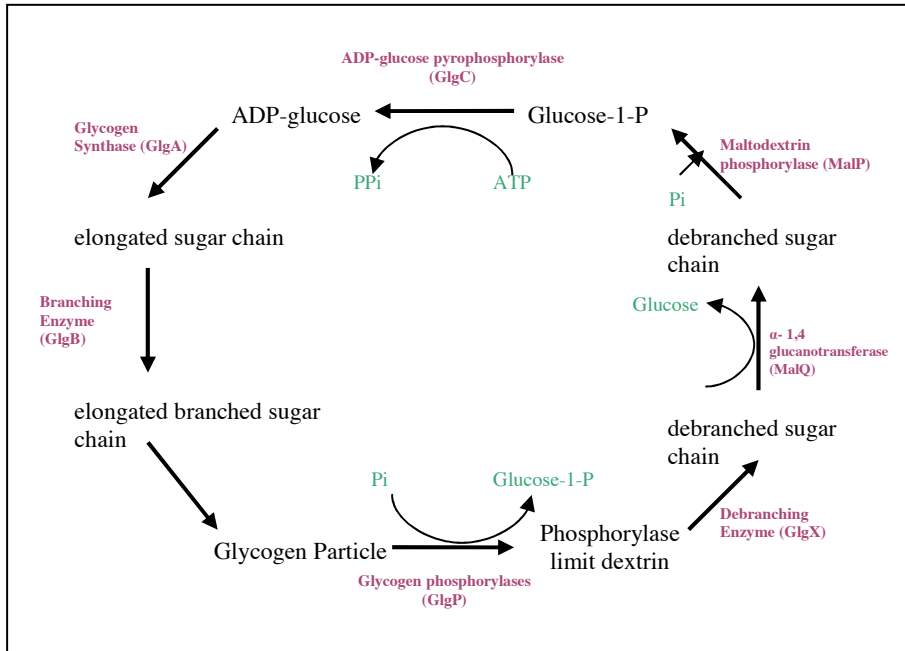
Figure 5: Glycogen metabolism in prokaryotes.
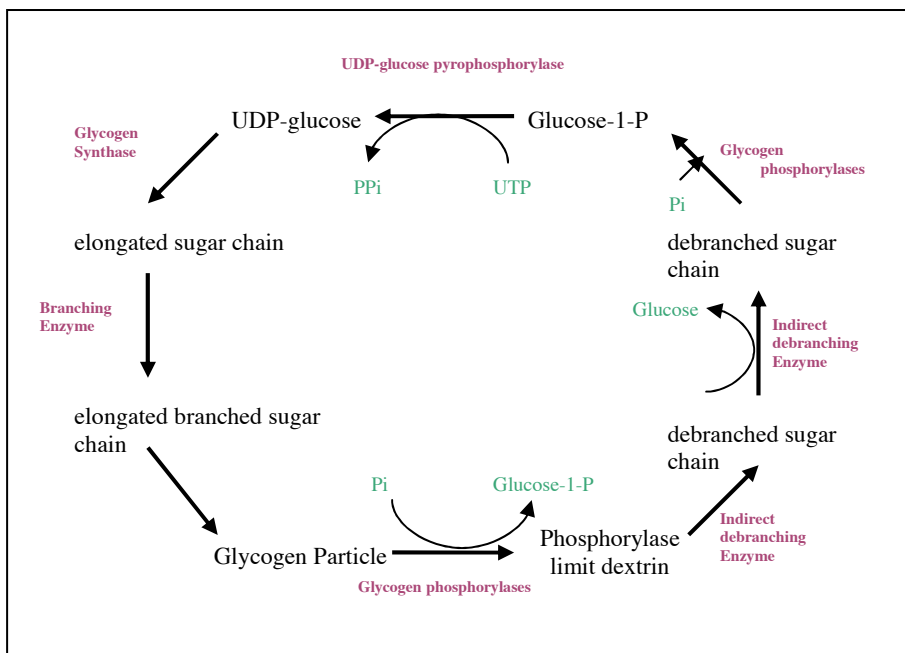
Figure 6: Glycogen metabolism in eukaryotes.

1.5.2.2 Starch Metabolism in Chloroplastida

As mentioned above, starch in Chloroplastida is derived from ADP-glucose (figure 7) (Ball and Morell, 2003; Preiss et al., 1987). Chloroplastida are the only eukaryotes that produce and metabolize this otherwise bacterial nucleotide-sugar. In general, starch and glycogen synthesis is strongly related. Different starch synthases using ADP-glucose as substrate are known:

Granule bound starch synthase, soluble starch synthase I/II (SSI/II) and soluble starch synthase III/IV (SSIII/IV). Starch synthases build up an amylose chain from activated ADP-glucose. A subsequent branching step between different amylose chains leads to amylopectine and finally the starch particle. This particle can be reshaped (transfer of glucose residues to different chains, modification of the glycosidic bonding) or broken down to glucose subunits by isoamylases and amylases (starch debranching enzymes).

Glucose-1-P adenylyltransferase

ADP-glucose ← Glucose-1-P

PPi   ATP

Starch Synthase

Growing amylose chain

Starch branching Enzyme

Amylose/Amylopektin

Starch Particle → Reshaping or break-down of starch

Starch debranching enzymes/Isoamylases
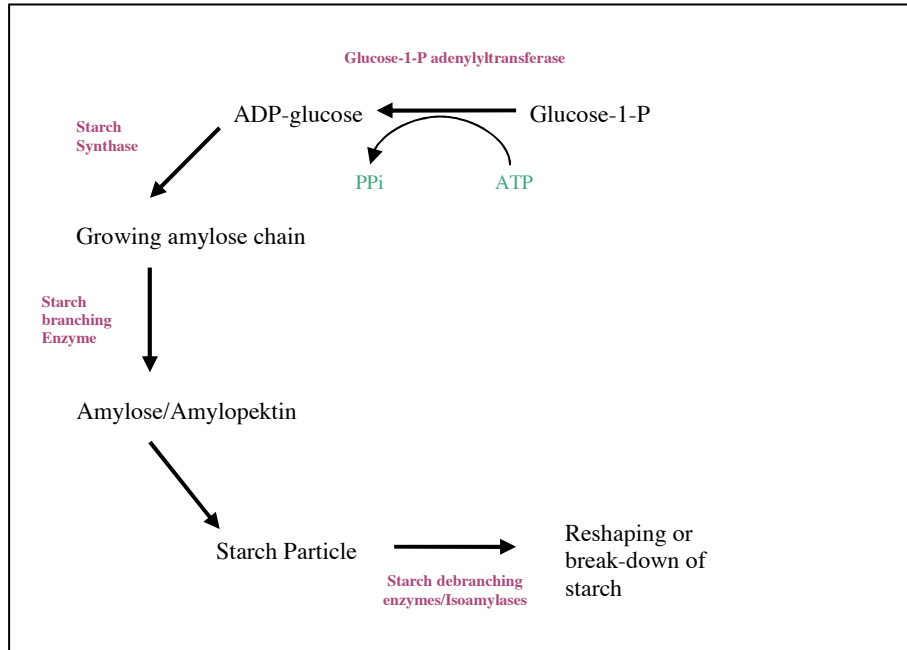
Figure 7: Starch metabolism in Chloroplastida.

1.6 Objective

In this work we use the data of the software "Effective" that has kindly been provided by the group of Prof. Dr. Thomas Rattei (University of Vienna, Austria). The whole genomes of *C. trachomatis* D/UW-3, *Chlamydophila caviae* GPIC and *Chlamydophila pneumoniae* CWL029 will be screened for eukaryotic-like domains. The hits will give rise to a pre-selection of candidates we will reasonably revise in order to obtain a final list of candidates. These final candidates will be tested for secretion by a TTSS in a heterologous TTSS of *Shigella flexneri*. We will further compare our obtained secretion results with computational predictions of TTS achieved with the software EffectiveT3. A focus on the eukaryotic-like domains of positive candidates will probably account for the uncovering of the candidate proteins' function in a eukaryotic host cell.

Additionally, we will test an assortment of chlamydial proteins engaged in glycogen metabolism for secretion into their eukaryotic host. If these enzymes possess TTS signals, it could provide strong evidence for an indispensable role of *Chlamydiae* in today's plants' history of development. This part of the work is done in collaboration with Prof. Steven Ball (University of Lille, France).

We will further examine if these proteins have different secretion characteristics in *C. trachomatis* and *C. pneumoniae*, knowing that only *C. trachomatis* accumulates glycogen.

2. Material and Methods

2.1 Material

2.1.1 Equipment

Autoclave	Matachana
Balance	Ohaus
Blotting Gadget	BioRad
Centrifuge Sorvall	ThermoScientific
Centrifuge Heraeus Megafuge	ThermoElectron Corporation
Centrifuge Tabletop 5415D	Eppendorf
Gel Chambers Agarose Gel SubCell	Bio-Rad
Gel Chambers Polyacrylamide Gel	Bio-Rad
GenePulser XCell™	Bio-Rad
Hypercassette™	Amersham Biosciences
Incubator	Memmert
Microwave	Easytronic
NanoDrop 2000	Labtech
Pipettes	Gilson
Powersupply Agarose Gel	Pharmacia
Powersupply Polyacrylamide Gel	Prolabo
SmartSpec™ 3000	Bio-Rad
Storm™	Molecular Dynamics
ThermalCycler 2720	Applied Biosciences
UV-Light	Bio-Rad
Waterbath	Polystat

2.1.2 Expendable Materials

Cuvettes 0,2 cm	Bio-Rad
Expendable Pipettes	Costar
Falcon™ Tubes	Sarstedt

Gloves	Kimtech
Hyperfilms™ MP	Amersham GE Healthcare
Inoculation Loop	Sarstedt
Microcentrifuge Tubes	Eppendorf
Millipore Filter	Sartorius Stedim
Paper 3MM	Whatman
Parafilm	Pechiney
Petri Dish 90mm	Graner
Polystyrene Round-Bottom Tubes	Becton Dickinson
Polyvinylidene Fluoride (PVDF) Membrane	Amersham GE Healthcare
Sterile Tips	StarLab
Syringe	Terumo
Thermo-Strip 0,2 mL PCR Tubes	ThermoScientific

2.1.3 Chemicals

Acrylamide/Bisacrylamide solution (40%)	Euromedex
Agar	Difco
Agarose	Sigma
Ammonium persulfate (APS)	Sigma
β-Mercaptoethanol	Merck-Schuchardt
Bromophenol Blue	Sigma-Aldrich
Chloroform	Sigma-Aldrich
Congo Red	Serva
ECF™ substrate	Amersham GE Healthcare
ECL™ substrate	Amersham GE Healthcare
Ethanol	TechniSorv
Ethylenediaminetetraacetic acid (EDTA)	Invitrogen
Ethidium bromide	Eurobio
Hydrochloric acid	Panreac
Glycerol	Sigma-Aldrich

Glycine	Fluka
LB broth base	Invitrogen
Phenol	Gibco
Potassium chloride	VWR
Potassium dihydrogen phosphate	Rectapur
Sodium chloride	Fluka
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich
Sodium hydrogen phosphate dihydrate	VWR
Sodium hydroxide	Prolabo
Tetramethylethylenediamine (TEMED)	Euromedex
Trichloroacetic acid	Sigma-Aldrich
Tris(hydroxymethyl)aminomethane (Tris)	Sigma
Tryptic soy broth	Merck
Tris Acetate EDTA Buffer (TAE)	Euromedex
Tween 20	Sigma-Aldrich

2.1.4 Antibiotics

Ampicillin	Sigma-Aldrich
Kanamycin	Sigma-Aldrich

2.1.5 Media, Buffer and Solutions

2.1.5.1 Media

<u>LB plates</u>

2,5% LB broth base and 1,5% agar are mixed in water and the pH is adjusted to 7,2 with NaOH. The mixture is autoclaved at 121°C for 15 min. Before adding antibiotics it has to cool down to 56°C (Ampicillin 100 µg/ml, Kanamycin 30 μg/ml).

LB medium

2,5% LB broth base are mixed in water and the pH is adjusted to 7,2 with NaOH. The mixture is autoclaved at 121°C for 15 min.

Congo red plates

0,01% Congo Red, 1,5% Agar, 3% Tryptic soy broth are mixed in water and the pH is adjusted to 7,3 with NaOH. Instead of autoclaving it, the mixture is boiled three times with a time interval of 5 min. Antibiotics are added as described in « LB plates ».

2.1.5.2 Buffer and Solutions

All buffers and solutions are diluted in pyrolysed water and autoclaved at 120°C for 20 min if not designated elsewise.

Phosphate-buffered saline (PBS) 1X

0,13 M NaCl
2,68 mM KCL
6,5 mM Na_2HPO_4 $2H_2O$
1,46mM KH_2PO_4

PBS Tween 0,1%

99,9% PBS 1X
0,1% Tween 20

Blocking solution

0,1 g/mL milk powder in PBS Tween 0,1%

Tris 3 M pH 8,8

3 M Tris adjusted to pH 8,8 with HCl

Tris 2 M pH 6,7

2 M Tris adjusted to pH 6,7 with HCl

Tris HCl pH 6,8

2 M Tris adjusted to pH 6,8 with HCl

Running buffer 1X (SDS-PAGE)

0,2 M Glycine
25 mM Tris base
0,1% SDS 20%

<u>Transfer buffer (Western Blot)</u>

25 mM Tris base
190 mM Glycine
20% Ethanol

<u>Loading Buffer 5X (SDS-PAGE)</u>

0,31 M Tris HCl pH 6,8
10% SDS
50% Glycerol
0,05% Bromophenol blue 1%

The ingredients are mixed and heated to 60°C before addition of Bromophenol blue.

<u>Loading Buffer 10X (Agarose gel)</u>

43,75% Glycerol 87%
31,25% Bromophenol blue 0,4%
25% EDTA 0,5 M

2.1.6 Enzymes and Antibodies

2.1.6.1 Restrictionenzymes

HindIII	`5'-A*AGCTT-3'`	Roche
	`3'-TTCGA*A-5'`	
XbaI	`5'-T*CTAGA-3'`	Roche
	`3'-AGATC*T-5'`	
BsaI	`5'-GGTCTC(N)`$_1$`*-3'`	Roche
	`3'-CCAGAG(N)`$_5$`*-5'`	

2.1.6.2 other Enzymes

Prime-Star DNA Polymerase	Takara
GoTaq DNA Polymerase	Promega
T4 DNA Ligase	New England Biolabs
Calf Intestinal Alkaline Phosphatase	New England Biolabs

2.1.6.3 Antibodies

Anti-cyclase antibody	By courtesy of N. Guiso, Institut Pasteur
Anti-CRP antibody	By courtesy of A. Ullmann, Institut Pasteur
Anti-IpaD antibody	By courtesy of C. Parsot, Institut Pasteur
Horseradish peroxidase-linked antibody	Amersham Pharmacia Biotech
Alkaline phosphatase-linked antibody	Pierce

2.1.7 Marker

| 1 kb DNA ladder | Invitrogen |
| Precision Plus Proteins Standard | Bio-Rad |

2.1.8 Kits

| QIAquick® PCR Purification Kit | Qiagen |
| PureLink™ Quick Plasmid Miniprep Kit | Invitrogen |

2.1.9 Organisms and Plasmids

2.1.9.1 Bacteria

Escherichia coli TG1	New England Biolabs
Escherichia coli DH5α	New England Biolabs
Shigella flexneri ipaB	By courtesy of C. Parsot, Institut Pasteur
Shigella flexneri mxiD	By courtesy of C. Parsot, Institut Pasteur

2.1.9.2 Plasmids, Sequences and *Chlamydiales* Genomes

2.1.9.2.1 Plasmids
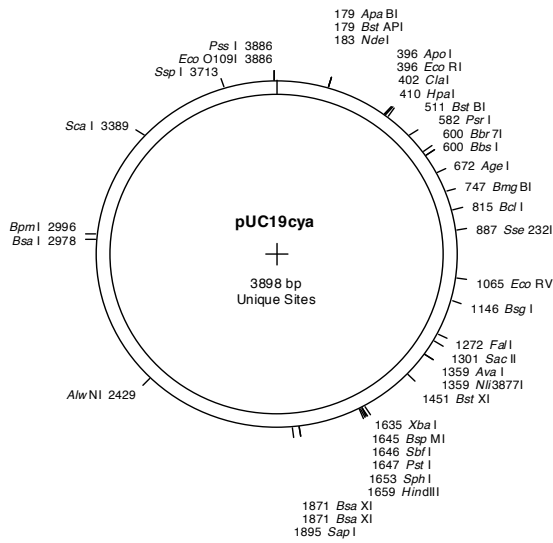
Figure 8: Vectormap of pUC19cya (not depicted Ampicillin Resistance, Origin of Replication)
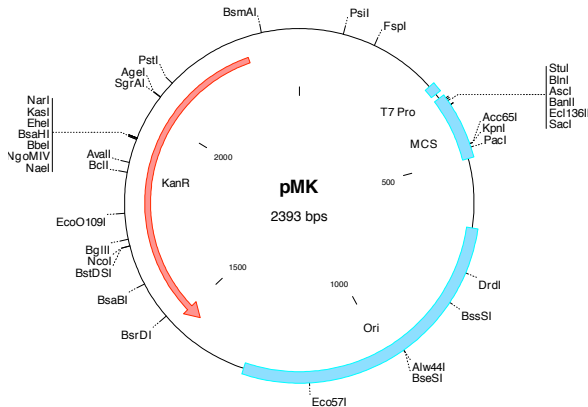
Figure 9: Vectormap of GeneArt® pMK (KanR = Kanamycin Resistance, Ori = Origin of Replication, MCS = Multiple cloning site, T7 Pro = T7 Promotor)

2.1.9.2.2 Sequences

Truncated sequence of calmodulin-dependent adenylate cyclase of *Bordetella pertussis* used for the pUC19cya vector:

```
GCCGTGGCGAAGGAAAAAAACGCCACATTGATGTTCCGCCTGGTCAACCCCCATTCCACCAGCCTGATTGCCGAAGGGGTGGC
CACCAAAGGATTGGGCGTGCACGCCAAGTCGTCCGATTGGGGGTTGCAGGCGGGCTACATTCCCGTCAACCCGAATCTTTCCA
AACTGTTCGGCCGTGCGCCCGAGGTGATCGCGCGGGCCGACAACGACGTCAACAGCAGCCTGGCGCATGGCCATACCGCGGTC
GACCTGACGCTGTCGAAAGAGCGGCTTGACTATCTGCGGCAAGCGGGCCTGGTCACCGGCATGGCCGATGGCGTGGTCGCGAG
CAACCACGCAGGCTACGAGCAGTTCGAGTTTCGCGTGAAGGAAACCTCGGACGGGCGCTATGCCGTGCAGTATCGCCGCAAGG
GCGGCGACGATTTCGAGGCGGTCAAGGTGATCGGCAATGCCGCCGGTATTCCACTGACGGCGGATATCGACATGTTCGCCATT
ATGCCGCATCTGTCCAACTTCCGCGACTCGGCGCGCAGTTCGGTGACCAGCGGCGATTCGGTGACCGATTACCTGGCGCGCAC
GCGGCGGGCCGCCAGCGAGGCCACGGGCGGCCTGGATCGCGAACGCATCGACTTGTTGTGGAAAATCGCTCGCGCCGGCGCCC
GTTCCGCAGTGGGCACCGAGGCGCGTCGCCAGTTCCGCTACGACGGCGACATGAATATCGGCGTGATCACCGATTTCGAGCTG
GAAGTGCGCAATGCGCTGAACAGGCGGGCGCACGCCGTCGGCGCGCAGGACGTGGTCCAGCATGGCACTGAGCAGAACAATCC
```

23

TTTCCCGGAGGCAGATGAGAAGATTTTCGTCGTATCGGCCACCGGTGAAAGCCAGATGCTCACGCGCGGGCAACTGAAGGAAT
ACATTGGCCAGCAGCGCGGCGAGGGCTATGTCTTCTACGAGAACCGTGCATACGGCGTGGCGGGGAAAAGCCTGTTCGACGAT
GGGCTGGGAGCCGCGCCCGGCGTGCCGAGCGGACGTTCGAAGTTCTCGCCGGATGTACTGGAAACGGTGCCGGCGTCACCCGG
ATTGCGGCGGCCGTCGCTGGGCGCAGTGGAACGCCAGGATTCCGGCTATGACAG

2.1.9.2.3 *Chlamydiales* Genomes

Chlamydophila pneumoniae CWL029: NC_000922

Chlamydophila caviae GPIC: NC_003361.3

Chlamydia trachomatis D/UW-3: NC_000117.1

Parachlamydia acanthamoebae UV7: NC_015702.1

Candidatus Protochlamydia amoebophila UWE25: NC_005861.1

2.1.9.3 Oligonucleotides

All oligonucleotides contain restriction sites for HindIII, XbaI or BsaI (underlined):

	Forward	**Reverse**
CT576	AGTCAAGCTTTTAGGAATTATCGCGATGAGCA	AGTCTCTAGACAAGCGGCTACGTGATTTTT
CT153	AGTCAAGCTTGTTCCTTCCATTATAGGGTGTCA	AGTCTCTAGAAGAGAAACGTCCTAATCGTGGA
CT305	AGTCAAGCTTTTAGGAGTAGTCTGCATGCGCGTAG	AGTCTCTAGAAAATTCGACAGCCCCAATCT
CT035	AGTCAAGCTTATAGAGGCAAAGAGATGAAGCGTAT	AGTCTCTAGACTGGAATGGAGCAGCTACCT
CT460	AGTCAAGCTTCTTTTCTTATCATCTTCTTTAACTAGGAGTC	AGTCTCTAGAACCAACGATGGCAGCTAAAT
CT862-61	AGTCAAGCTTATAGGCCATCTCCAAGAAGTGTCT	AGTCTCTAGACATATTATCGGGAAGCGGAC
CT862-87	AGTCAAGCTTATAGAGGACACCTTTATCTACAACATGC	AGTCTCTAGACATATTATCGGGAAGCGGAC
CPn0811	AGTCAAGCTTTTAGGAATTAGATCGATGAGCAAG	AGTCTCTAGATAGACGGCTTCGCGTTTTT
CPn0091	AGTCAAGCTTAAGATAGAACTCCGGAAAGCAA	AGTCTCTAGAACCAAGCTCTCTACTTGCAGAA
CPn0128	AGTCAAGCTTATAGCGCTTTTCTGTTCGAGAGG	AGTCTCTAGAACGTATCGTATGCCGTAAATAATA
CPn0489	AGTCAAGCTTGTTACAATGGAGGATTGGCTAAG	AGTCTCTAGAACACGCTGTGACCTCATCC
CPn0577	AGTCAAGCTTGTAGTCAGGTCTCTTCTTACCACCTT	AGTCTCTAGACATAGGTCCCTTGCCAACTA
CPn0769	AGTCAAGCTTATAGAGGTCCTCACGCAATTAGATTA	AGTCTCTAGAGGCAAAAACAAATTCACTCC
CPn0856	AGTCAAGCTTGGATCAAGAGATAGGAACGTAAGG	AGTCTCTAGATATGGCTTTGAGCTTATCTGC

CPn1018	AGTC<u>AAGCTT</u>CTGGTTTACAAGTGGAAAATGA	AGTC<u>TCTAGA</u>CTTGGCAAGTAAGCCTAAAGAC
CPn1021	AGTC<u>AAGCTT</u>ATAGAGGACACCTTCATTTACAACCTACG	AGTC<u>TCTAGA</u>ATCATCTGGGAAAGGGAAGTC
CPn0176	AGTC<u>GGTCTC</u>TAGCTAGGCGAAGTTAAGGAAGCTCTG	AGTC<u>TCTAGA</u>AAACTCATAGGAAAAACTAGCTGAA
CPn0562	AGTC<u>AAGCTT</u>ATAGACTACATTCTGCATGAAATTATAAGG	AGTC<u>TCTAGA</u>CATGGCGATCGAAGGTTTA
CPn0927	AGTC<u>AAGCTT</u>CTTTTGTTCTCTTTATACTCTGTGTACC	AGTC<u>GGTCTC</u>ACTAGACGGCTTTGGATCCGTCT
CPn0928	AGTC<u>AAGCTT</u>AGAGACGTAAGGATTTCGCATT	AGTC<u>TCTAGA</u>TGACGAAAAATAGGAAACTGG
CPn0929	AGTC<u>AAGCTT</u>ATAGTTAAATTTTATAGGAAAAGTTCATGG	AGTC<u>TCTAGA</u>AAAATAAGTCGCTTGTGGAGA
CCA00743	AGTC<u>AAGCTT</u>ATAGACCGATTTACAAGTTGAAGAAGA	AGTC<u>TCTAGA</u>ATGCATTCCTACTCGCAATA
CCA00911	AGTC<u>AAGCTT</u>ATCTAACATCAAGAGATAGGAACGC	AGTC<u>TCTAGA</u>ATGCTCTTGATTAATAGACTTGAGTT
CTMalQ	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGCCGTCATTATCCCAAT	AGTC<u>TCTAGA</u>GCCATGCTTAGGAGAAGTATCTATC
CTGlgA	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGAAAATTATTCACACAGCTATCG	AGTC<u>TCTAGA</u>TGCTAGTCCGTATAGCGCG
CTGlgC	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGCTGGTAGACGGACG	AGTC<u>TCTAGA</u>TCCGCATAAGACAATAACTCCTAC
CTGlgP	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGTATTTCGATCGGACAAAG	AGTC<u>TCTAGA</u>TTGAGGAGTCTGGACTACCC
CTGlgX	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGAATCTTTGTCTGTTCGTT	AGTC<u>TCTAGA</u>AGAAAATAGAGAAAAACGGTAGCG
CTGlgB	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGATCCTTTTTTCTTAAATACTCAAC	AGTC<u>TCTAGA</u>AGAAACAATTCCCAAAAGATCTTG
CPnGlgX	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGAAAAAGTTTCTTCTTATCCC	AGTC<u>TCTAGA</u>AGCATATAAAGCAAATCGATAGC
CPnGlgA	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGAGAATCGTACAAGTCGCTGT	AGTC<u>TCTAGA</u>TAACTCCTTAGATAGACTAGCTACAGCATC
CPnGlgB	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTAGTGTTGATAAACTGATCCATCC	AGTC<u>TCTAGA</u>AGCAAGGATCCCTAAGAGTTTATG
CPnGlgC	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGATAGAAAACGATTTTCCGG	AGTC<u>TCTAGA</u>CCCTCCACACAAGATAATTACTCC
CPnMalQ	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGAATGTTTTAAAATACACAAAACACTC	AGTC<u>TCTAGA</u>TGGGAGATAAATCCCGTGTTTA
CPnGlgP	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGAAGATTTTTCGAGTTTTGATAA	AGTC<u>TCTAGA</u>TACAACACTTAAATACAGACGATCTAAAA
ParaGlgAΔ10	AGTC<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGCACCACTTGCCAAAGTC	GGTACC<u>TCTAGA</u>TAGTTCACGAGAGAG
ParaMalQBIS	AGTC<u>AAGCTT</u>GTAATTACTGAGGAATGGTAACGATGACAAACTTTTT	AGTC<u>TCTAGA</u>GAAAATGGGAAGGCAAA

2.1.9.4 Genes synthesized by GeneArt® Gene Synthesis

GeneArt® Gene Synthesis synthesized the amino terminal sequences of the genes derived

from *Parachlamydia acanthamoebae* and *Candidatus Protochlamydia amoebophila* genome

(all inserts are listed in the appendix). Inserts were integrated into the pMK vector bearing a Kanamycin resistance. By performing a restriction digest we cleaved out the insert and ligated it into pUC19cya (see 2.2.1.5 and 2.2.1.6).

2.1.10 Computer Programs, Softwares and Databases

2.1.10.1 Computer Programs and Softwares

- DNA-Strider 1.4f3 : work on DNA and protein sequences
- Primer3Plus : primer design
- Quantity One : Visualization of agarose gel bands
- National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/ : nucleotide BLAST

2.1.10.2 Databases

- National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/ : genomes
- Effective http://effectors.org : effector predictions

2.2 Methods

2.2.1 Cloning Methods

The following cloning methods are listed in a chronological order to obtain chimeras of the candidate proteins which were then tested for secretion. Chimeras consisted out of the approximately first 20 amino terminal amino acids of the candidate protein (20 N-ter) and a truncated form of the calmodulin-dependent adenylate-cyclase (Cya) of *Bordetella pertussis* (figure 10; see 2.1.9.2.2 for its sequence). All constructs were verified by sequencing (all inserts are listed in the appendix).
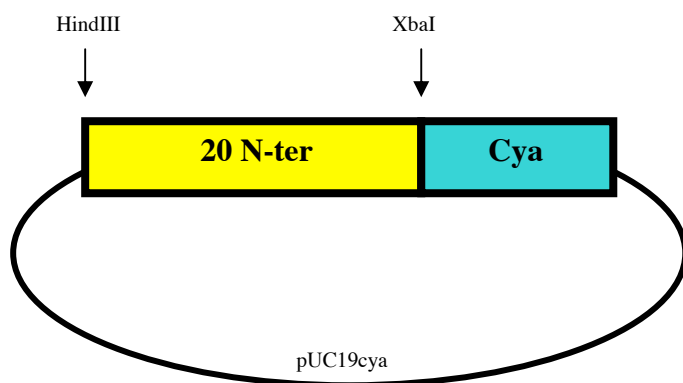
Figure 10: Scheme of pUC19cya-chimera. The nucleotide sequence of the 20 amino terminal amino acids are cloned into the vector pUC19cya by using HindIII and XbaI restriction sites.

2.2.1.1 Constructs and Primer Design

2.2.1.1.1 Constructs

The FASTA nucleotide sequence in NCBI of each candidate was taken and saved as a DNA-Strider file. Complement sequences were transformed into the antiparallel strand. After having transcribed them into amino acids the largest open reading frame (ORF) was assumed to depict the protein. Approximately the first 20 amino terminal amino acids from the start-codon on were taken into account and screened for restriction sites with HindIII and XbaI (BsaI if the sequence was cut by one of the first two).

2.2.1.1.2 Primer Design

In order to design primers for the subsequent Polymerase Chain Reaction (PCR) the software Primer3Plus was used. Our inserts were to be ligated into vector pUC19cya, that possesses an α-Galactosidase (α-Gal) site upstream of the ligation site which we did not want to be expressed. For this reason we were looking for stop-codons in frame with α-Gal located at least 10 and maximum 21 bases upstream of the start-codon of our candidate sequence. In case of absence a stop codon had to be integrated into the forward primer. We avoided the sequences to be in frame with the α-Gal site. The forward primers contained also a HindIII restriction site for further cloning procedures. The reverse primers featured a XbaI restriction site to put the insert in frame with the on the vector downstream located cyclase-tag (cya, calmodulin-dependent adenylate cyclase of *Bordetella pertussis*). In two cases a BsaI restriction site had to be integrated instead of HindIII and XbaI, respectively. The melting temperatures (Tm) for the primers were chosen in a range from 55° to 63° C.

2.2.1.2 Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction is a technique in molecular biology to amplify pieces of DNA. This can be used to detect the presence or absence of a certain piece or for cloning methods. In this work parts of the *C. caviae*, *C. pneumoniae* and *C. trachomatis* genome, respectively, were amplified producing the sequences of interest flanked by two restriction sites. A DNA Polymerase with high fidelity was chosen.

PCR machine : 2720 Thermal Cycle, Applied Biosystems

Reaction mix (per tube) :

Buffer Prime-Star 5x	5 μL
dNTP (2,5 mM)	2,5 μL
Prime-Star DNA Polymerase (2,5 U/μL)	0,25 μL
Forward Primer (10 μM)	1 μL
Reverse Primer (10 μM)	1 μL
Template DNA (10-100ng)	0,25 μL
ddH2O	16 μL
Total	25 μL

PCR conditions :

First Denaturation	98°C	5 min
Denaturation step	98°C	10 s
Annealing step	55°C	10 s
Elongation step	72°C	1 min

} 30 cycles

For runs with a less accurate DNA Polymerase (to validate the presence of a certain insert for example) the DNA Polymerase GoTaq was used.

Reaction mix (per tube):

Buffer GoTaq 5x	5 μL
MgCl2	2 μL
dNTP (2,5 mM)	0,5 μL
GoTaq DNA Polymerase (5 U/μL)	0,13 μL
Forward Primer (10 μM)	2 μL
Reverse Primer (10 μM)	2 μL
Template (boiled bacteria)	5 μL
ddH2O	8,37 μL
Total	25 μL

PCR conditions :

First Denaturation	95°C	5 min	
Denaturation step	95°C	30 s	
Annealing step	55°C	60 s	} 30 cycles
Elongation step	72°C	60 s	
Final Elongation	72°C	10 min	

2.2.1.3 Electrophoretical Separation of DNA Molecules by means of an Agarose Gel

This technique is used to separate DNA or RNA fragments according to their size. Agarose in a concentration of 1-3% is put in Tris Acetate EDTA Buffer (TAE) and carefully heated in the microwave until everything is dissolved. The solution is poured into a gel chamber. After cooling down the gel is solid and samples can be run at a constant voltage from 80-120 V with TAE as running buffer. The gel is then incubated in water or TAE with ethidium bromide (1-10 μg/mL) for 15 min and washed twice with water or TAE. Afterwards the bands are revealed under UV-light.

2.2.1.4 Purification of PCR Products

In order to remove the reagents of a previous PCR reaction and to have the DNA sufficiantly purified for a subsequent restriciton digestion QIAquick® PCR Purification Kit is used. By a simple bind-wash-procedure DNA fragments in a range of 100 bp to 10 kb can be purified with a yield of up to 95%.

Five volumes of binding buffer (PB) are added directly to the PCR reaction and the mixture is applied to the silica based spin column. The high-salt conditions provided by the buffer lead to an adsorption of the DNA to the silica-gel membrane and impurities can be washed away with the buffer PE. The bound DNA is eluted with 30 μL of water.

2.2.1.5 Restriction Endonucleases

Restriction endonucleases are enzymes that specifically recognize DNA sequences (restriction sites) and hydrolyse the DNA backbone producing either sticky or blunt ends. Different providers recommend different reaction conditions and temperatures depending on the enzyme, therefore the respective conditions should be consulted prior to an experiment.

2.2.1.5.1 Digestion of PCR and GeneArt® Products

Purified PCR products and GeneArt® plasmids with insert were digested with HindIII and XbaI (or BsaI if the construct possessed a restriction site for one of the others) in order to ligate them into the vector.

Reaction mix (per tube) :

HindIII/XbaI		BsaI/XbaI		HindIII/BsaI	
Buffer 2 10x	2 μL	Buffer 2 10x	2 μL	Buffer 2 10x	2 μL
BSA 100x	0,2 μL	BSA 100x	0,2 μL	BSA 100x	0,2 μL
HindIII (10 U/μL)	0,1 μL			HindIII (10 U/μL)	0,1 μL
XbaI (10 U/μL)	0,1 μL	XbaI (10 U/μL)	0,1 μL		
H2O	1,6 μL	H2O	1,6 μL	H2O	1,6 μL
DNA	16 μL	DNA	16 μL	DNA	16 μL
Digestion : 90 min 37°C					
		Buffer 3 10x	1 μL		
		BsaI (10 U/μL)	0,2 μL		
		Digestion : 60 min 50°C			
Inactivation : 30 min 65°C					

Buffer 2, 3 and BSA 100x : New England Biolabs

DNA: either 16 μL of PCR-product or 0,5 μg of GeneArt® plasmid.

2.2.1.5.2 Digestion and Phosphorylation of Vector

The vector pUC19cya was digested with HindIII and XbaI and two phenol/chloroform extractions were performed. The vector was precipitated with ethanol and treated with calf intestinal alkaline phosphatase to prevent self-ligation. Again two phenol/chloroform extractions and a precipitation step with ethanol were conducted and the DNA is diluted in 40 μL of Buffer TE.

2.2.1.6 Ligation of Insert and Vector

In cloning procedures ligases are applied to create a phosphodiester bond between two double-stranded DNA fragments. This offers the possibility to integrate an insert of interest into a specific vector. As a rule of thumb the ratio « insert : vector » is in a range from 1 : 1 to 3 : 1.

For 10 μL of ligation mix with T4 DNA Ligase (New England Biolabs) the reaction buffer (1 μL), the T4 DNA Ligase (0,05 μL, 400 U/μL), nuclease free water, vector pUC19cya (~0,025

pmol) and insert (~0,075 pmol) are carefully mixed and incubated at 16°C for at least 30 min (up to overnight).

2.2.1.7 Transformation in *Escherichia coli*

A transformation in general is the uptake of free DNA by competent bacteria. This procedure can be applied for different purposes, among it the amplification of a plasmid. Different methods can be used, which differ in the choice of bacteria and reaction conditions. The used methods here are the so-called Heat-Shock Transformation and the Electroporation.

2.2.1.7.1 Heat-Shock Transformation in *TG1*

Ligated plasmids are amplified in heat-shock competent *E. coli TG1*. For this purpose 7 μL of the ligation products are mixed with 40 μL of *TG1*, kept on ice for 30 min and transferred to 42°C for exactly 45 s. The sample is immediately put back on ice and as fast as possible 300 μL of LB medium are added. In order to allow bacteria to divide and to produce the antibiotic resistance they are incubated on 37°C for 40 min. 60 to 300 μL of the mixture are plated out on LB plates with the respective antibody for selection and put in a 37°C incubator overnight.

2.2.1.7.2 Electroporation in *DH5α*

Electroporation with electroporation competent *E. coli DH5α* is used to amplify plasmids. An externally applied electric field leads to a permeabilization of the cell membrane, rendering DNA uptake possible. 40 μL of *DH5α* are mixed with 10-100 ng DNA, transferred to a 2 mm cuvette and pulsed with BioRad GenePulser (2500 V, 25 μF, 200 Ω). Cells are immediately put back on ice and 300 μL of LB medium are added. In order to allow bacteria to divide and to produce the antibiotic resistance they are incubated on 37°C for 40 min. 60 to 300 μL of the mixture are plated out on LB plates with the respective antibody for selection and put in a 37°C incubator overnight.

2.2.1.8 Purification of Plasmid via Miniprep

The used PureLink™ Quick Plasmid Miniprep Kit renders it possible to isolate plasmids out of a bacterial culture. For this purpose 1,5 mL of an overnight culture are pelleted (5 min, 13200 rpm) and resuspended in resuspension buffer. Cells are lysed via an alkaline/SDS procedure with lysis buffer. The precipitation step occurs with the respective precipitation buffer and after centrifugation the supernatant is applied to a silica membrane column that selectively binds plasmid DNA. Two washing steps with washing buffers remove the

contaminants. The bound DNA is eluted with 30 μL of water and stored at either -20°C or 4°C (to avoid frequent freezing/thawing).

2.2.2 Working with *Shigella flexneri*

Shigella flexneri belong to the part of family of *enterobacteriaceae* which are able to secret the Shiga-toxin. Therefore they should be handled with care.

2.2.2.1 Preparation of electrocompetent *ipaB* and *mxiD*

The procedure for both strains is identical and has to be carried out on ice or 4°C.

5 mL of an overnight culture (30°C) are seeded out in 500 mL of LB broth/antibiotic. After about 3 h at 37°C the culture should have an optical density (OD at 600 nm) between 0,6 and 1. Cells are centrifuged at 6000 rpm for 10 min at 4°C and the supernatant is removed. The pellet is resuspended in 250 mL of icecold water and centrifuged again. This washing step has to be carried out twice. Afterwards one washing step with water/glycerol 10% for 15 min 4000 rpm is conducted and the pellet is resolved in 2 mL of water/glycerol 10%. Aliquots of the now electrocompetent bacteria are made and stored at -80°C.

2.2.2.2 Transformation in *Shigella flexneri* via Electroporation

Same procedure as for *E. coli DH5α*, see 2.2.1.7.2.

2.2.2.1 Growing of *Shigella flexneri ipaB* and *mxiD*

Shigella flexneri ipaB and *mxiD* liquid cultures were grown at 30°C overnight before shifting them to 37°C. By experience this decreases the risk of losing the plasmid encoding the TTSS. *Shigella flexneri ipaB* were grown on Congo Red plates. Colonies that possess the TTSS plasmid turn red upon secretion, non-secreting colonies are bigger in size and white.

2.2.3 Test of Cloning Efficiency by PCR

To validate the presence of the candidate sequences in *Shigella flexneri* a PCR with the respective primers is run. One colony of a plate is picked, put in 35 μL of water and boiled for 5 min. 5 μL are taken to serve as template. For reaction mix and PCR conditions see 2.2.1.2 GoTaq. 10 μL of the PCR products are run on a 3% agarose gel.

2.2.4 Secretion Test

2.2.4.1 Solid Test

Shigella flexneri ipaB colonies transformed with different constructs were picked in the morning on an LB plate with the respective antibiotic and incubated for 8 h at 37°C. In the evening the plate was covered with a polyvinylidene fluoride (PVDF) membrane and the colonies were allowed to grow overnight at 37°C. The following day the membrane was soaked for 5 min into 100% ethanol and washed three times with 0,1% Tween 20 in phosphate-buffered saline (PBS). After saturation in blocking solution (see 2.1.5.2) for at least 30 min at room temperature the membrane was probed with anti-Cya antibody and either alkaline phosphatase-linked or horseradish peroxidase-linked secondary antibodies. For revelation procedure and the exact usage of antibodies see 2.2.5.3 and 2.2.5.2, respectively.

2.2.4.2 Liquid Test

1 mL of a 30°C overnight culture of *Shigella flexneri ipaB* or *mxiD* transformed with different constructs was inoculated in 30 mL of LB broth/antibiotic and incubated at 37°C for 4 h. 1 mL was removed and pelleted (5 min, 13200 rpm). The pellet was then resolved in 500 μL of loading buffer (see 2.1.5.2) and boiled for 20 min before being stored at -20°C. 25 mL of overnight culture were centrifuged for 20 min at 4000 rpm and 4°C. The supernatant was filtered with a Millipore filter (0,2 μm) attached to a syringe. To precipitate the proteins 3 mL of trichloroacetic acid were added and allowed to incubate on ice for 30 min. The sample was centrifuged for 15 min at 10000 rpm and 4°C, the supernatant removed, the pellet dried and resolved in 500 μL of loading buffer with 3 μL of NaOH 10 N in order to adjust the pH. Before storing them at -20°C they were boiled for 5 min.

2.2.5 Protein-Biochemical Methods

2.2.5.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE is a means to separate proteins according to their size in an electric field. Since proteins possess, unlike nucleic acids, varying charges and shapes according to their secondary and tertiary structures, they may not migrate into the polyacrylamide gel at similar rates. Therefore the proteins are usually denatured and coated with a negative charge, so that their separation is now dependent on the size. Sodium dodecyl sulfate (SDS) is a detergent that fulfills both functions, and is added to the sample as an ingredient of the loading dye.

The gel itself consists out of an acrylamide/bisacrylamide solution (40%), which upon polymerization builds up a dense mesh with concentration-dependent pore size, tris (hydroxy methyl) aminomethane (Tris) as buffer, SDS, water, TEMED (N, N, N', N'-tetramethylethylenediamine) that accelerates the polymerization and APS (Ammonium persulphate) to initiate the reaction. APS and TEMED are added at the very end.

The gel is composed of two different parts : the upper part called stacking gel leads to a concentration of the proteins on a thin starting zone, the lower part (running gel) leads to a separation of the proteins according to their size. This difference is obtained by the usage of a low percentage gel (big pores) and a lower pH for the stacking gel.

The samples are mixed with the loading dye and 5% β-mercaptoethanol which has the capacity to cleave disulfide bonds and boiled for 5 min. They are loaded on the gel, the gadget filled with migration buffer and run at 100-130 V.

Valid for one gel :

	8% Running gel	12% Running gel	Stacking gel
Tris 3M pH 8,8	950 μL	950 μL	-
Tris 2M pH 6,7	-	-	156 μL
SDS (10%)	75 μL	75 μL	25 μL
acrylamide/bisacrylamide (40%)	1,5 mL	2,25 mL	235 μL
Water	4,95 mL	4,15 mL	2,06 mL
APS (10%)	75 μL	75 μL	25 μL
TEMED	7,5 μL	7,5 μL	2,5 μL

2.2.5.2 Western Blot

The western blot is a widely used analytical technique to detect specific proteins in a given sample. Proteins coated with a negative charge are transferred from a polyacrylamide gel to a PVDF membrane via an electric current (electroblotting) where they are probed using antibodies specific to the target protein.

The used blotting apparature is the so called tank-blotting system (in contrast to a semi-dry blot). Therefor the PVDF membrane is activated by soaking it with 100% ethanol for 5 min. After washing it twice with water the « sandwich » can be assembled, consisting out of Whatman filter paper (3 layers), PVDF membrane, gel and again 3 layers of Whatman filter papers, all soaked in transfer buffer. The sandwich is put in the appropriate gadget and the whole tank-blotting system is filled up with transfer buffer. The transfer takes place at either 30 V overnight or 120 V for 1-1,5 h (depending on the percentage of the gel).

Since the membrane has the ability to bind every protein a subsequent step has to be undertaken to prevent the membrane to bind to other proteins than the ones present in the gel. For this purpose the membrane is blocked after the transfer for at least 30 min in a saturating blocking solution before being probed with antibodies.

Primary antibodies conduct a protein-specific binding. The used secondary antibodies bind to the primary antibody and are linked to a reporter enzyme which when exposed to an appropriate substrate allows the detection of the protein. Both antibodies are diluted in 0,1% Tween 20 in phosphate-buffered saline (PBS). Between and after the incubation of the membrane with the different antibodies (1 h at room temperatur each) 3 washing steps with 0,1% Tween 20 in phosphate-buffered saline (PBS) for 10 min were conducted. The solutions of primary antibodies were reused and stored at -20°C. For secondary antibodies either alkaline phosphatase-linked or horseradish peroxidase-linked antibodies were applied.

2.2.5.3 Revelation of PVDF membranes

2.2.5.3.1 by Storm™

The adequate substrate in case of labelling with alkaline phosphatase-linked secondary antibodies is ECF™ substrate which is applied dropwise (~1 mL/membrane) on a plastic membrane. After treatment with antibodies the PVDF membrane is put with the protein side down on the substrate drops and incubated for 1 min. The proteins on the PVDF membrane can now be visualized by Storm™ .

2.2.5.3.2 with a film

Horseradish peroxidase is used to cleave a chemiluminescent agent (ECL™ substrate) which can be visualized by placing a sensitive sheet of photographic film against the membrane. Exposure to the light leads to an image of the antibodies bound to the blot.

The procedure for adding the substrate can be seen under 2.2.5.3.1.

3. Results

It has largely been shown that *Chlamydiae* use a TTS machinery to translocate effector proteins into the host cell, provoking different effects which are up to date still subject to many investigations. Due to a lack of genetic tools to manipulate *Chlamydiae* a wide screen of putative effectors has only once been attempted so far, even though the revelation of potentially secreted proteins could open the field for further functional studies (Subtil et al., 2005).

In this work we apply a TTSS-screen on 46 different candidates of 5 different species of the order *Chlamydiales*. This screen is based on the fact that effectors can be secreted by a heterologous TTSS, meaning that the candidates can also be secreted by another bacterium possessing a TTSS (Subtil et al., 2001). We chose *Shigella flexneri* for our experiments and designed 51 constructs for our 46 candidates. Some candidates were tested with two constructs because the translational start was ambigous.

The selection of the candidates was made from two different aspects. For one series they were selected based on the presence of eukaryotic-like, for the other series we chose the candidates according to their function in the glycogen metabolism. Both rationales are discussed in more detail below.

3.1 Selection of the Candidates

3.1.1 via Computational Analysis

In order to identify proteins with eukaryotic-like domains, computational analysis of genomes of *C. caviae*, *C. trachomatis* and *C. pneumoniae* was performed with the database "Effective". This database classifies all organisms of which completely sequenced genomes are listed in the RefSeq md the GenBank database into eukaryotes, pathogens, symbiotic and non-pathogenic bacteria (Jehl et al., 2011). The whole proteomes of *C. caviae*, *C. trachomatis* and *C. pneumoniae* were screened and eukaryotic-like domains extracted using signatures detected by Pfam. The calculation was restricted to protein domains which are detected in pathogenic/symbiontic genomes as well as in at least 3 eukaryotic genomes, in order to eliminate the influence of bacterial contaminations in eukaryotic genomes.

Each detected domain annotated with its Pfam name yields a specific score. This so called domain enrichment score S is calculated as the number of standard deviations σ of the background frequency in non-pathogen genomes in which the domain frequency in that

particular pathogenic/symbiontic genome *n* differs from the background frequency η in non-pathogen genomes:

$$S = (n - \eta) / \sigma$$

This score allows distinguishing between domains that are uniformly distributed over eukaryotic, non-pathogenic and pathogenic genomes and domains that are only or mainly present in eukaryotes and pathogens/symbionts. Domains with a score higher than 3 were considered to be enriched and taken into account, which resulted in 38 hits for the genomes of *C. caviae* (12 hits), *C. trachomatis* (10 hits) and *C. pneumoniae* (16 hits). Domains exclusively found in pathogens/symbionts and eukaryotes get a score of 10000.

For all these hits a computational prediction of TTS signals was performed. Only few type III effectors are known so far, mainly due to the fact that the secretion signal is still puzzling. According to the best established model that the first 20 amino terminal amino acids of a protein determine its ability to be secreted by this system, we used the EffectiveT3 software, which is accessible through the database "Effective". This software detects putative TTS signals using an algorithm that is trained to divide secreted and non-secreted proteins by rating a combination of discriminative sequence properties of the amino termini. The score gives a value between 0 and 1 for each candidate, 1 being the most confident for secretion. The cut-off value was set at 0.99.

In table 3 all 38 hits for the extraction of eukaryotic-like domains are depicted with their locus tag, EffectiveT3 score *T3*, the evaluation of this score considering its cut-off, Pfam of the eukaryotic-like domain and the domain enrichment score *S*. For clearer view the candidates predicted to be secreted are highlighted in yellow, the non-secreted in blue.

Locus tag	*T3*	T3 Pred.	Pfam	*S*
		C. caviae		
CCA00164	0,996046511	Y	PF02201	5
CCA00180	0,999893146	Y	PF05677	10000
CCA00261	1	Y	PF10275	10000
CCA00718	0,999987251	Y	PF02902	17
CCA00911	0,999999154	Y	PF01704	4
CCA00254	0,437064582	N	PF03690	5
CCA00648	2,60E-07	N	PF09825	5
CCA00681	1,02E-07	N	PF01496	4
CCA00740	7,92E-06	N	PF07720	4
CCA00743	6,22E-06	N	PF08123	7
CCA00952	0,095512048	N	PF07720	4

CCA00988	1,75E-13	N	PF02201	5
C. pneumoniae				
CPn0577	0,99824817	Y	PF02201	5
CPn0811	0,999875252	Y	PF07720	7
CPn0856	0,99306126	Y	PF01704	4
CPn0928	0,999029796	Y	PF05677	10000
CPn0929	0,999997291	Y	PF05677	10000
CPn0091	9,67E-07	N	PF01496	4
CPn0128	1,05E-04	N	PF09825	5
CPn0176	3,57E-10	N	PF01823	10
CPn0483	1,84E-06	N	PF10275	10000
CPn0489	0,297280059	N	PF03690	5
CPn0562	8,84E-09	N	PF05677	10000
CPn0769	1,68E-08	N	PF02201	5
CPn0887	5,00E-15	N	PF07720	7
CPn0927	0,905041597	N	PF05677	10000
CPn1018	5,27E-10	N	PF08123	7
CPn1021	3,65E-06	N	PF07720	7
C. trachomatis				
CT576	0,99828678	Y	PF07720	4
CT867	0,9988277	Y	PF02902	34
CT868	0,999995295	Y	PF02902	34
CT035	5,77E-08	N	PF09825	5
CT153	4,18E-08	N	PF01823	10
CT305	2,32E-11	N	PF01496	4
CT386	0,293770526	N	PF03690	5
CT460	0,618098361	N	PF02201	5
CT643	9,00E-15	N	PF02201	5
CT862	0,856131679	N	PF07720	4

Table 3: Extraction of eukaryotic-like domains in *C. caviae*, *C. trachomatis* and *C. pneumoniae*. Extraction of eukaryotic-like domains and prediction of TTS signals were performed with "Effective", cut-off values amounted to 4 and 0,99, respectively. Proteins are given with their locus tag, domains with their Pfam name. *S* = domain enrichment score, *T3* = EffectiveT3 score, Proteins predicted to have a TTS signal are yellow, the rest blue.

We aligned all homolog proteins being part of these hits between the three different species on NCBI Blast and checked for their consistency regarding the first 20 amino terminal amino acids. Proteins being very conserved between two or all three species were supposed to have the same secretion properties and only one construct for the subsequent screen was designed (Subtil et al., 2005). Table 4 depicts all homologs amongst the 38 hits highlighting in green the ones that were very conserved, and in this case giving an alignment of the first 20 amino terminal amino acids.

Chlamydia trachomatis	Chlamydophila caviae	Chlamydophila pneumoniae	
CT576	CCA00952	CPn0811	
CT867	CCA00718		
CT868			
CT035	CCA00648	CPn0128	
CT153		CPn0176	
CT305	CCA00681	CPn0091	
CT386	CCA00254	CPn0489	`MQIPRSVGTHDGSFHADEVT   CT386` `MQIPRSIGTHDGSFHADEVT   CPn0489`
CT460	CCA00164	CPn0577	
CT862	CCA00740	CPn1021	
CT643	CCA00988	CPn0769	`MKKSLIIVESPAKIKTLRKL   CT643` `MKKSLIIVESPAKIKTLQKL   CPn0769`
		CPn0562	
	CCA00180	CPn0927	
		CPn0928	
		CPn0929	
	CCA00911	CPn0856	
	CCA00743	CPn1018	
	CCA00261	CPn0483	
		CPn0887	

Table 4: Homologies between the *C. caviae*, *C. trachomatis* and *C. pneumoniae* hits chosen by "Effective". Highlighted in green are homologs with high conservation. Highlighted in yellow are deviating amino acids.

Hence, two pairs of homologs were sufficiently conserved (CT386 – CPn0489 and CT643 – CPn0769), and only one construct of each pair was selected: CPn0489 and CPn0769. Amongst the remaining hits we focused mainly on *C. trachomatis* and *C. pneumoniae*, selecting only two of the *C. caviae* hits for further examination: CCA00911 and CCA00743. The translational start for CT862 was ambigous, so we chose to design two constructs (CT862-61 and CT862-87). This gave rise to a set of 26 candidates we chose for the in vitro assay of TTS, highlighted the ones which were additionally predicted to have a TTS signal (table 5).

Chlamydia trachomatis	Chlamydophila caviae	Chlamydophila pneumoniae
CT576		CPn0811
CT867		
CT868		
CT035		CPn0128
CT153		CPn0176
CT305		CPn0091
		CPn0489
CT460		CPn0577
CT862		CPn1021
		CPn0769
		CPn0562
		CPn0927
		CPn0928
		CPn0929
	CCA00911	CPn0856
	CCA00743	CPn1018
		CPn0483
		CPn0887

Table 5: *C. caviae*, *C. trachomatis* and *C. pneumoniae* proteins chosen for secretion tests due to eukaryotic-like domains in their sequence. Highlighted in yellow the ones with an additional putative TTS signal.

3.1.2 via their Role in the Glycogen Metabolism

Considering the hypothesis that a tripartite symbiosis might have been essential for the arising of Archaeplastida, we asked the question at which point the selection pressure was so favorable compared to the simple symbiosis between heterotrophic organism and cyanobiont. The advantage for the host to interrupt phagocytosis of the cyanobiont in order to profit from its capability of photosynthesis is obvious – a new source of energy. But this source is only within reach if the host can metabolise the cyanobiont's product, ADP-glucose. It is most unlikely that the eukaryotic cell adapted quickly to this new situation by modifying its own glycogen enzymes, which use UDP-glucose as a substrate. More likely is the hypothesis that these enzymes or their genes were contributed by a prokaryotic partner, either directly by LGT, or with a preceding step of secretion of the effectors into the host. Interestingly, no other eukaryotes obtained the capacity of using ADP-glucose as substrate.

According to unpublished data of Steven Ball, the groups of soluble starch synthases (SS) III/IV as well as starch debranching enzymes (both archaeplastidal) are derived from prokaryotes. Phylogenetic analysis further suggests a chlamydial origin, more precisely a

source within the environmental *Chlamydiales* such as *Candidatus Protochlamydia amoebophila* and *Parachlamydia acanthamoeba*. These data support additionally a monophyly of all Archaeplastida through a common chlamydial LGT.

The group SSIII/IV is a group of ADP-glucose-utilizing starch synthases that seem to be most related to the chlamydial GlgA (glycogen synthase). Furthermore, analysis suggests that archaeplastidal starch debranching enzymes are derived from chlamydial GlgX (direct debranching enzyme).

Again, in the scenario of a tripartite symbiosis time and probabilities play an important role. If *Chlamydiae* secrete enzymes (like GlgA and GlgX) into the host, that help this cell to metabolise the newly acquired cyanobiont energy source, time pressure would be less. The system would then have more time to stably integrate these genes into the host genome via LGT. A test of secretion clarifies the point whether these enzymes are putative effectors.

We chose to test MalQ, GlgA, GlgB, GlgC, GlgP and GlgX of *Candidatus Protochlamydia amoebophila* and *Parachlamydia acanthamoeba* (except *C. P. amoebophila* GlgA, for which the sequence was highly similar to *P. acanthamoeba*, *C. P. amoebophila* GlgC for which the amino terminal sequence was to uncertain, and *C. P. amoebophila* GlgP).

To gain more insight into the possibly different ways of glycogen metabolism of *C. trachomatis* and *C. pneumoniae* we also chose to test the previous mentioned 6 different glycogen-linked enzymes in these two species.

Summing it up we had 21 candidates, amongst which we had one with an ambigous translation start (*P. acanthamoeba* MalQ and MalQBis), one for which it was not clear whether the gene locus was correctly annotated (*P. acanthamoeba* GlgP1 and GlgP2) and one for which we deleted (additionally to the normal construct *P. acanthamoeba* GlgA) the first 10 amino terminal amino acids. This served us as a negative control.

Table 6 shows all candidates we chose to test, with their protein name and their locus tag.

	Chlamydia trachomatis	*Chlamydophila pneumoniae*	*Parachlamydia acanthamoeba*	*Candidatus Protochlamydia amoebophila*
MalQ	CT087	CPn0326	PUV_07340	pc0745
GlgA	CT798	CPn0948	PUV_18990	-
GlgB	CT866	CPn0475	PUV_16710	pc1761
GlgC	CT489	CPn0607	PUV_08690	-
GlgP	CT248	CPn0307	PUV_22600	-
GlgX	CT042	CPn0388	PUV_16520	pc1106

Table 6: List of *Chlamydia trachomatis, Chlamydophila pneumoniae, Parachlamydia acanthamoeba and Candidatus Protochlamydia amoebophila* proteins involved in glycogen metabolism chosen for secretion test. Proteins are given with their protein name and their locus tag on the genome.

3.2 Results of Cloning Step

We constructed chimeras for each of our *C. caviae*, *C. trachomatis* and *C. pneumoniae* candidates consisting approximately out of the first 20 amino terminal amino acids and a cyclase reporter molecule.

Performing a PCR with the respective primers we obtained the 5 prime end of the chimeric gene bearing the sequence of chlamydial origin flanked by restrictions sites. The efficiency of the PCR reaction was tested on agarose gels. A band at about 120 bp was detected for all of our candidates, which corresponds to the expected size of the fragment. In a digestion step with restriction enzymes we prepared the PCR products as well as the synthesized genes for *Candidatus Protochlamydia amoebophila* and *Parachlamydia acanthamoeba* for a subsequent ligation into the vector pUC19cya. The ligation product bearing an antibiotic resistance as selection marker was transformed into *E. coli*. Transformation efficiency was good for all candidates except for CPn0577 where only very few colonies grew. Expression of the chimera was tested by Western Blot, where we could detect the proper band at about 50 kD for all candidates except CPn0577 and to a much less extent CPn0769. A PCR was rerun for CPn0577 and digestion, ligation and transformation steps repeated. Cloning efficiency was better but never reached the level of the other samples. Upon transformation into *Shigella flexneri* mutant *ipaB* and plating on Congo Red plates we could distinguish between colonies that lost their plasmid encoding the TTS apparatus (big white colonies) and colonies that incorporated our construct and were still in possession of a functional TTSS (little and red). About 10-30% of the colonies lost their TTS plasmid. Transformed *mxiD* were plated out on normal LB plates. As control we run a PCR with the bacterial colonies to see whether there was no mixing up of the different candidates. All constructs except the ones that were designed with GeneArt® were verified by sequencing. The first PCR was run with GoTaq, where the sequencing revealed 5 mutated constructs out of 27. The PCR for these candidates was rerun with Prime-Star. At a second PCR run CPn1018 still incorporated a mutation and was discarded of the list. The chimeras of CT868, CT867, CPn0887 and CPn0483 were already available. Out of a total of 51 constructs to be designed and cloned, only one failed.

3.3 Secretion Test of "Effective" candidates

The main goal of this work was to test candidate proteins for secretion via a TTSS. The secretion tests assume the universality of the secretion signal recognized by TTS machineries, giving the possibility to use a heterologous TTSS. The chimeras consisting out of the first 20

amino terminal amino acids and the cyclase reporter molecule were tested in *Shigella flexneri*, revealing the properties of their secretion by a TTS machinery.

3.3.1 Solid Test

The solid test of secretion is a possibility to screen rapidly for secretion for a big amount of candidate proteins. The chimeras were transformed into the *Shigella flexneri* strain *ipaB* and colonies grew overnight on a LB plate covered with a PVDF membrane, which was probed with the antibody against the cyclase reporter molecule the day after. Secreted chimeras appeared as a halo surrounding the dot where the colony grew, giving sometimes a very clear signal, sometimes a signal harder to define. Chimeras that were not secreted at all just appeared as the dot where the colony grew.

Importantly, previous experiments revealed that the occurrence of false positives is below 5% (Subtil et al., 2005). The solid test was exclusively performed for the *C. pneumoniae*, *C. caviae* and *C. trachomatis* candidates (for which we obtained clones) possessing a eukaryotic-like domain. Five constructs for which no results are available had a mutation at that time and were only subjected to the next test. Some of the listed candidates had already been tested, serving as positive and negative controls in our assay. Additionally, we tested the plasmid pUC19cya without insert for secretion.

We obtained 10 positives (amongst them the four positive controls), 2 unclear, 11 negatives (including the 3 negative controls) out of 23 tested constructs. Depicted in figure 11 are the positive, the unclear candidates and pUC19cya. Table 7 gives an overview of the results.
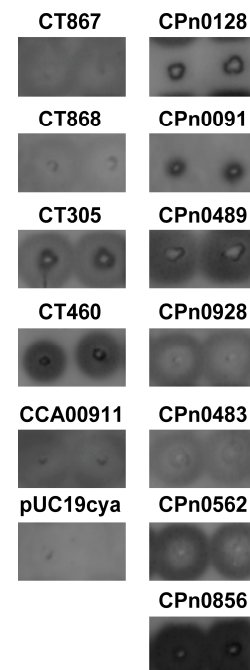
Figure 11: Solid test of secretion on colonies. *ipaB* strain was transformed with chimeras, colonies grew overnight covered with a PVDF membrane. The membrane was used the next day to reveal the localization of the chimeras by using anti-Cya antibodies. Dot-shaped patterns show that constucts were not secreted, halo-shaped patterns demonstrate secretion of the construct into the medium. CPn0128 and CPn0091 are unclear; pUC19cya is a negative control; the rest is positive.

CT862-61	-	CPn0577	-
CT862-87	-	CPn0091	unclear
CT035	NT	CPn0928	+
CT153	NT	CPn0769	NT
CT460	+	CPn0176	-
CT305	+	CPn0927	NT
CT576	-	CPn0811 (- ctrl)	-
CCA00911	+	CPn0887 (- ctrl)	-
CCA00743	-	pUC19cya (- ctrl)	-
CPn0929	-	CPn0489 (+ ctrl)	+
CPn1021	-	CPn0483 (+ ctrl)	+
CPn0562	+	CT868 (+ ctrl)	+
CPn0128	unclear	CT867 (+ ctrl)	+
CPn0856	+		

Table 7: Secretion results for solid test. Tested were the candidates that were chosen due to eukaryotic-like domains in their sequence. Positive for secretion (+), negative for secretion (-), not tested (NT), signal too feeble to be interpreted (unclear).

3.3.2 Liquid Test

The liquid test is more sensitive than the solid test but also more time consuming. To investigate whether our hybrid proteins are produced and secreted we fractionized exponentially growing liquid cultures of *Shigella flexneri ipaB* expressing the construct of interest into pellet and supernatant. The supernatant fraction was concentrated 25-fold compared to the pellet fraction. Hybrid proteins were subsequently subjected to a Western Blot probing them with antibodies against cyclase providing information about secretion. In order to exclude a signal in the supernatant fraction due to leakage or lysis of the cell, we also probed the membrane with antibodies against a cytosolic bacterial protein, cAMP receptor protein (CRP). To check whether the introduced construct hampered secretion we also checked the presence of IpaD, one of the *Shigella flexneri* proteins that is secreted by a TTS machinery in the supernatant. Candidates having a band for the chimera and IpaD in the supernatant (possibly also to a certain amount in the pellet) and a signal for CRP restricted to the pellet fraction were considered as positive for secretion.

In order to test whether this secretion occured via a TTSS, the chimera was also introduced into *Shigella flexneri mxiD* possessing a totally impaired TTS machinery. If the chimera signal was only in the pellet and not in the supernatant, the candidate was considered as positive for TTS.

Regarding the hits of proteins with a eukaryotic-like domain we chose to test all *C. pneumoniae* candidates, except CPn0483 and CPn0887 (which had already been tested previously). Additionally, we included CCA00911, CT878, CT876 and the candidates that had not been subjected to the solid test (CT035, CT153). This resulted in a list of 18 candidates.

Out of our hits we chose with the database „Effective" we obtained 13 positive and 3 negative results regarding secretion (figure 12, table 8). Importantly, for all of them secretion in *mxiD* was negative, proving that it occured via a TTSS. Controls by IpaD and CRP validated the results. We could not determine the properties of secretion for two constructs (CPn0176 and CPn0577, not depicted in figure 12) since they were neither detected in the pellet nor in the supernatant of *ipaB* and *mxiD*. In conclusion, the results of these two from the solid test should be discarded.
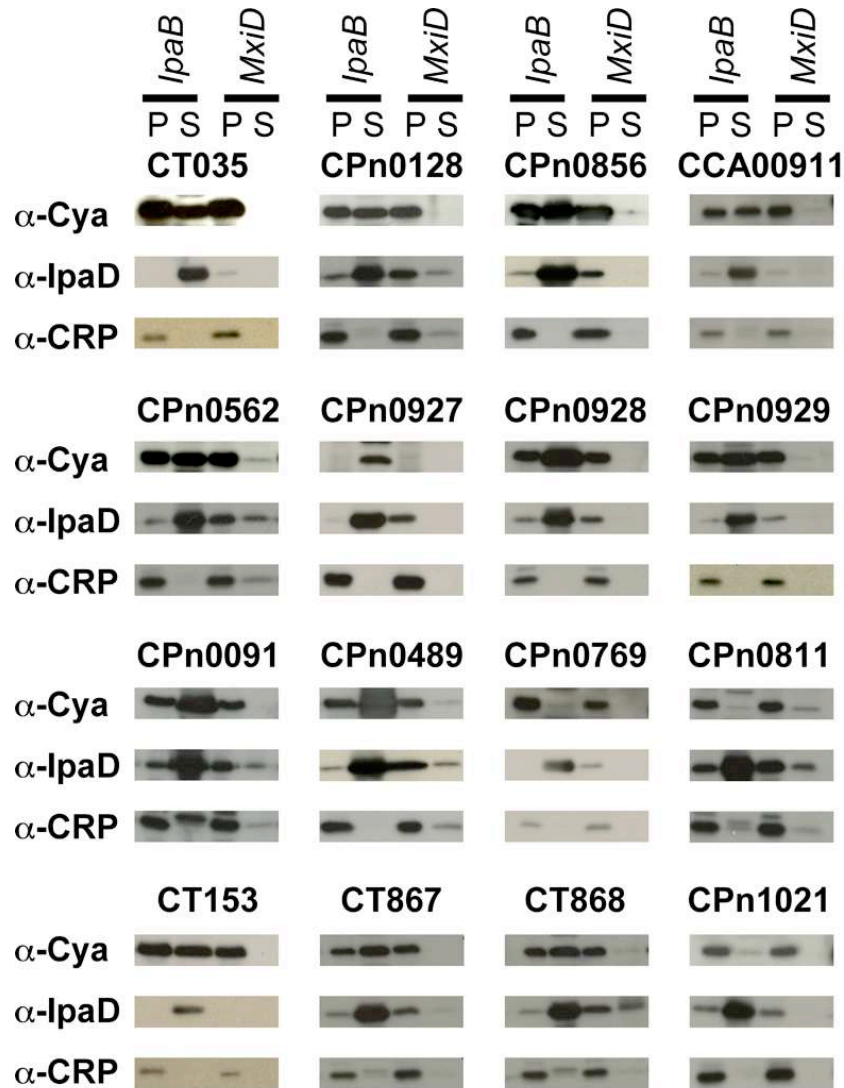
Figure 12: Liquid test of secretion on candidates chosen due to eukaryotic-like domains in their sequence. Liquid cultures of *ipaB* and *mxiD* expressing the indicated construct were franctionized into pellet (P) and supernatant (S) and run on a SDS-PAGE as described in Material and Methods. Supernatant fraction was concentrated 25-fold compared to pellet fraction. The membrane was probed with antibodies against Cya, IpaD and CRP. CT = *Chlamydia trachomatis*, CPn = *Chlamydophila pneumoniae*

CT867	+	CPn0489	+
CT868	+	CPn0769	-
CT035	+	CPn1021	-
CT153	+	CPn0562	+
CCA00911	+	CPn0927	+
CPn0811	-	CPn0928	+
CPn0128	+	CPn0929	+
CPn0091	+	CPn0856	+

Table 8: Secretion results for liquid test. Tested were the candidates that were chosen due to eukaryotic-like domains in their sequence. Positive for secretion (+), negative for secretion (-).

By comparison of the valid results for the solid and the liquid test (table 9) we found a very good correlation. All candidates being positive in the solid test also revealed secretion in the liquid test. Only one candidate was negative in the solid and positive in the liquid test (CPn0929) confirming the higher sensitivity of the latter.

It is to mention that we can only exclude secretion by another secretory pathway for candidates we tested in the liquid test. The solid test was only performed with *ipaB*, not in the *mxiD* control. Still, the constructs are designed in the context of a TTS signal, thus the probability of secretion of the chimeras by another pathway remains very low. This is confirmed by the absence of signal in the *mxiD* supernatant we obtain in all our 13 positively tested constructs.

	Solid Test	Liquid Test		Solid Test	Liquid Test
CT862-61	-	NT	CPn0856	+	+
CT862-87	-	NT	CPn0091	unclear	+
CT035	NT	+	CPn0928	+	+
CT153	NT	+	CPn0489	+	+
CT460	+	NT	CPn0769	NT	-
CT305	+	NT	CPn0811	-	-
CT576	-	NT	CPn0927	NT	+
CT868	+	+	CPn0483	+	NT
CT867	+	+	CPn0887	-	NT
CPn0929	-	+	CCA00911	+	+
CPn1021	-	-	CCA00743	-	-
CPn0562	+	+	pUC19cya	-	NT
CPn0128	unclear	+			

Table 9: Comparison of results for solid and liquid test of secretion. Candidates were chosen due to eukaryotic-like domains in their sequences. Positive for secretion (+), negative for secretion (-), not tested (NT), signal too feeble to be interpreted (unclear).

3.4 Results for Glycogen Metabolism Candidates

3.4.1 Liquid Test

In a second set of the liquid test we tested all candidates chosen by their role in the glycogen metabolism (24 hits). 16/24 candidates were positive for secretion (figure 13, table 10), amongst which one failed the *mxiD* control (PcMalQ). It is therefore possible to confirm its secretion, but not to specify if this secretion occurs via a TTSS. Interestingly, we could detect that PUVGlgP2 was positive, whereas PUVGlgP1 does not even seem to be well expressed. Hence we assume that the nucleotide sequence for PUVGlgP2 is the right one for this protein. PUVGlgAΔ10 lacking the first 10 amino acids was not secreted neither (in contrast to PUVGlgA) proving again clearly the reliability of this assay.

	Chlamydia trachomatis	*Chlamydophila pneumoniae*	*Parachlamydia acanthamoeba*	*Candidatus Protochlamydia amoebophila*
MalQ	+	+	- MalQ - MalQBis	+ (not clear if TTS)
GlgA	+	+	+ GlgA - GlgAΔ10	NT
GlgB	+	-	+	+
GlgC	-	-	+	NT
GlgP	+	+	+ GlgP2 - GlgP1	NT
GlgX	+	+	unclear	+

Table 10: Secretion results for liquid test. Tested were the candidates of *Chlamydia trachomatis, Chlamydophila pneumoniae, Parachlamydia acanthamoeba* and *Candidatus Protochlamydia amoebophila* proteins involved in glycogen metabolism. Positive for secretion (+), negative for secretion (-), not tested (NT), signal too feeble to be interpreted (unclear).
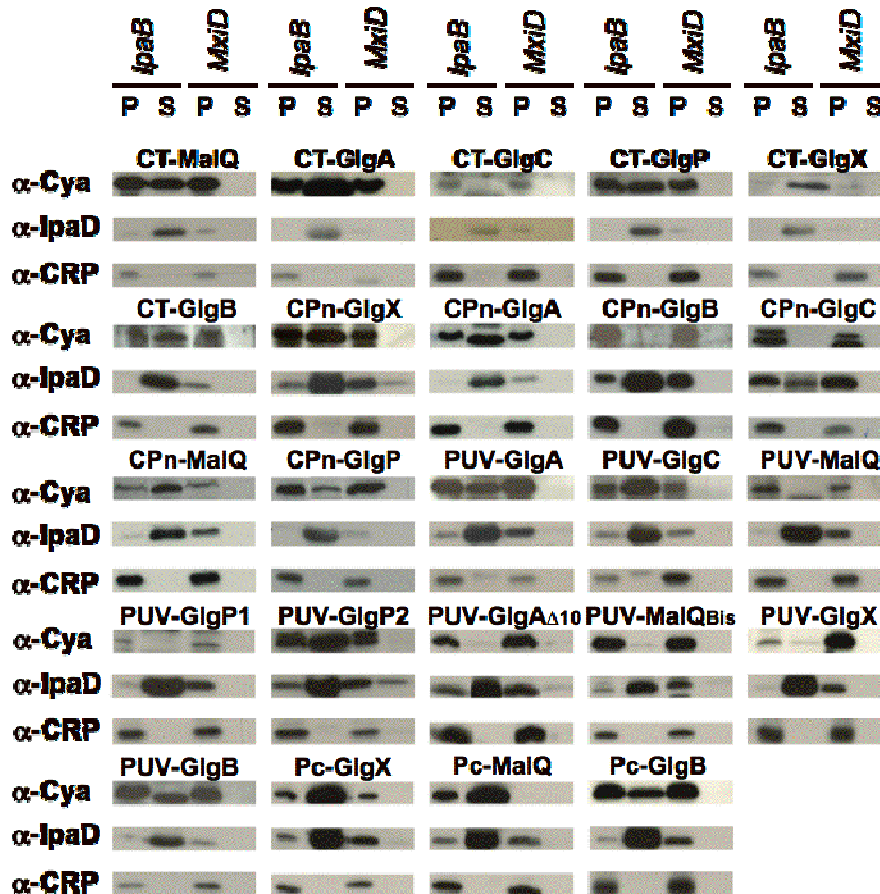
Figure 13: Liquid test of secretion on candidates involved in glycogen metabolism. Liquid cultures of *ipaB* and *mxiD* expressing the indicated construct were franctionized into pellet (P) and supernatant (S) and run on a SDS-PAGE as described in Material and Methods. Supernatant fraction was concentrated 25-fold compared to pellet fraction. The membrane was probed with antibodies against Cya, IpaD and CRP. CT = *Chlamydia trachomatis*, CPn = *Chlamydophila pneumoniae*, PUV = *Parachlamydia acanthamoeba*, Pc = *Candidatus Protochlamydia amoebophila*

All 21 candidates involved in the glycogen metabolism were also screened for TTS signals with EffectiveT3. Only three of the candidates had a predicted TTS signal (cut-off value at 0.99): CTMalQ (T3 = 1), CPnGlgX (T3 = 0.99835) and PcGlgX (T3 = 0.9999).

3.4.2 Phylogenetic Analysis of Glycogen Metabolism Candidates

We showed in the previous paragraph that chlamydial proteins engaged in the glycogen metabolism could have been secreted into the eukaryotic host. Here we will examine whether our phylogenetic analysis support a stable integration of the chlamydial genes encoding these proteins into the host genome.

In collaboration with Prof. Thomas Rattei (University of Vienna, Austria) phylogenetic trees of *Candidatus Protochlamydia amoebophila* GlgA, GlgB, GlgC, GlgP, GlgX and MalQ were established (see appendix). Due to their large size only details are depicted. The tree for GlgA reveals a close grouping of Archaeplastida and *Chlamydiae* GlgA, especially to the environmental *Chlamydiae Candidatus Protochlamydia amoebophila* and *Parachlamydia acanthamoeba*. Similar results are obtained for GlgB. There is no specific relationship existing between archaeplastidal and chlamydial GlgC and GlgP according to our trees. "Environmental chlamydial" GlgX seems to be closely grouped with archaeplastidal GlgX. Interestingly, all *Chlamydiales* (except *Simkaniaea*) group together with a part of Archaeplastida in the tree for MalQ, but there is also another part of Archaeplastida which seems to have a MalQ more closely related to other bacteria.

4. Discussion

4.1 Secretion Results of Computational Predictions

4.1.1 Evaluation of Extraction of Eukaryotic-like Domains and Prediction of TTS Signal

By focusing on eukaryotic-like domains that are enriched in the proteomes of pathogens and symbionts compared to non-pathogens we obtained a list of 38 proteins of *C. caviae*, *C. trachomatis* and *C. pneumoniae*. We chose to test 26 (two of them had a homolog being sufficiently conserved to assume the same secretion results) for secretion by a TTSS. This screen was performed in a heterologous TTSS of *Shigella flexneri* that has been shown to be functional for chlamydial effectors. Three of our candidates failed to be tested, for reasons of cloning difficulties or insufficient expression of the chimera (CPn0577, CPn0176, CPn1018). When transformed into the *ipaB* strain, chimeras positive for secretion were detected around the colony (solid test) or in the supernatant (liquid test). In the liquid test none of our positive chimeras was detected in the supernatant of TTS deficient *mxiD*, proving a secretion by the TTS machinery. We report the following observations of the identification of secreted chlamydial effectors (included two that were not tested, but whose amino termini are very conserved so that we made the assumption that they behave like the homologous protein that was tested):

(i) 17/25 candidates are secreted by a TTSS, 8/25 were not secreted,

(ii) 8/25 tested proteins had additionally a predicted TTS signal, under which 6/8 chimeras were indeed positive for secretion,

(iii) all 7 candidates that reached a domain enrichment score of 10000 were secreted, 4/6 proteins with a domain enrichment score of 4 (the threshold chosen for candidate selection) were secreted,

(iv) when homologs of different chlamydial species were tested, they always showed consistent results.

We can therefore assume that about 70% (17/25) of the chlamydial proteins that had an enrichment of a eukaryotic-like domain were secreted by a TTSS. We cannot exclude the secretion of the remaining 30% of candidates by another secretion pathway.

LGT from a eukaryotic cell can occur to pathogenic/symbiontic and eventually also to non-pathogenic organisms, but only be retained by pathogens/symbionts, giving rise to a domain enrichment score of 10000 (a domain enrichment score of 10000 excludes the presence of this domain in non-pathogens). However, a slight enrichment just above the cut-off of 3 also

seems sufficiently significant. It is indeed likely that eukaryotic-like domains being enriched in pathogens/symbionts strengthen their virulence or play at least a role within the process of communication between host cell and intruder. Since this communication often occurs via secretion of effectors into the host, the extraction of eukaryotic-like domains to pre-select putative effectors is a powerful tool. However, this tool only takes into account a subset of effectors and can thus only be additionally used.

A method to obtain a more exhaustive list of candidates that are secreted by a TTSS would be based on a reliable prediction of a TTS signal. EffectiveT3 is a machine learning approach based on amino terminal features as frequencies of amino acids, amino acid properties and short combinations of them. This learning approach requires a huge data set of TTS effectors. Still, no extensive list of TTS effectors is available so far. By including the predictions of a putative TTS signal for our tested candidates we can state that the proposed secretion signal is rather too restrictive. 11/17 positively tested proteins were predicted to be negative. However, 6/8 proteins that appeared as type III secreted in EffectiveT3 were indeed positive, indicating that this approach is very selective. The appearance of false negatives (or positives) can be due to wrong annotations of the translational start sites. It is therefore highly recommended to manually verify it before applying EffectiveT3. We took this into account and could only find one gene (CT862), where the translational start was ambiguous. Hence, we designed two different constructs, both revealing the same result (negative). We manually verified the remaining genes and found out that their translational starts were consistent with the ones of the source that was used for EffectiveT3. We can thus minimalize the bioinformatical bias of wrongly annotated translational starts or open reading frames.

It has been suggested that a putative secretion signal partially (or fully) depends on the mRNA sequence. There are indications of the correctness of both theories, the mRNA signal and the peptide signal hypothesis. Consequently, this is an issue that still remains to be investigated in detail. However, we only looked at a putative TTS signal under the aspect of a strictly proteinaceous secretion signal. To date the idea that the mRNA sequence is partially also necessary for a proper secretion cannot be rejected totally. It might even be possible that the predominance of mRNA- or peptide-based sequences is dependent on the protein. Thus, it cannot be completely ruled out that the low number of predicted type III secreted proteins is due to an exclusive focus on a proteinaceous signal.

4.1.2 Secreted Proteins, their Eukaryotic-like Domains and putative Function

Table 11 shows all tested candidates with their locus tag, description of the protein's function indicated on GenBank, the eukaryotic-like domain with its Pfam name and its function as given on the Sanger Institute Pfam database. Highlighted in blue are candidates that appeared to be negative and highlighted in yellow the positive ones. In the following we will give a short overview of each group of homologs (data about Pfam ID and domain function: CDD conserved protein domain database of NCBI).

Chlamydophila pneumoniae	*Chlamydia trachomatis*	*Chlamydophila caviae*	**Pfam domain**
CPn0811 low calcium response protein H	CT576 low calcium response protein H	-	PF07720 Tetratricopeptide repeat
-	CT867 hypothetical protein	-	PF02902 Ulp1 protease family
-	CT868 hypothetical protein	-	PF02902 Ulp1 protease family
CPn0128 biotin protein ligase	CT035 biotin protein ligase	-	PF09825 Biotin-protein ligase
-	CT153 MAC/perforin family protein	-	PF01823 MACPF protein superfamily
CPn0091 V-type ATP synthase subunit I	CT305 V-type ATP synthase subunit I	-	PF01496 V-type ATPase
CPn0489 hypothetical protein	CT386 metal dependent hydrolase	-	PF03690 Uncharacterised protein family (UPF0160)
-	CT460 SWIB (YM74) complex protein	-	PF02201 SWIB/MDM2 domain
CPn0769 DNA topoisomerase I/SWI domain fusion protein	CT643 DNA topoisomerase I/SWI domain fusion protein	-	PF02201 SWIB/MDM2 domain
CPn1021 low calcium response protein H	CT862 type III secretion chaperone	-	PF07720 Tetratricopeptide repeat
CPn0562 CHLPS 43 kDa protein homolog_1	-	-	PF05677 Domain of unknown function (DUF)
CPn0927 CHLPS 43 kDa protein homolog_2	-	-	PF05677 Domain of unknown function (DUF)
CPn0928 CHLPS 43 kDa protein homolog_3	-	-	PF05677 Domain of unknown function (DUF)
CPn0929 CHLPS 43 kDa protein homolog_4	-	-	PF05677 Domain of unknown function (DUF)
CPn0856 UTP-glucose-1-phosphate uridylyltransferase	-	CCA00911 UTP-glucose-1-phosphate uridylyltransferase	PF01704 UTP-glucose-1-phosphate uridylyltransferase
CPn0483 hypothetical protein	-	-	PF10275 Peptidase C65 Otubain
-	-	CCA00743 hypothetical protein	PF08123 Histone methylation protein DOT1
CPn0887 CHLTR phosphoprotein	-	-	PF07720 Tetratricopeptide repeat

Table 11: Tested proteins and eukaryotic-like domains. Homologs are in the same line with their eukaryotic-like domain. Highlighted in yellow: positive for secretion. Highlighted in blue: negative for secretion.

Proteins that are secreted:

CT867 – CT868

CT867 and CT868 are two adjacent homologous ORFs. They possess deubiquinating and deneddylating activity (Misaghi et al., 2006). Ubiquitination and neddylation only occur in eukaryotes, therefore our results that these deubiquitinating and deneddylating proteins are secreted support previous findings of their activity. Interestingly, homologs of these proteins have been found only in *C. trachomatis* and *C. muridarum*. Their eukaryotic-like domain belongs to the Ulp1 protease family containing the catalytic triad Cys-His-Asn.

CPn0128 – CT035

The attachment of biotin to requiring proteins is triggered by biotin ligase proteins, in eukaryotes as well as in prokaryotes. It has been shown that biotin is a cofactor in the metabolism of fatty acids, leucine and interestingly, also in gluconeogenesis (Pacheco-Alvarez et al., 2002). According to Belland et al. (Belland et al., 2003) CT035 belongs to the immediate early expressed genes in the infection cycle of *C. trachomatis*. Translocation of this biotin ligase protein into the host cell might trigger the production of glucose or fatty acids in the cytoplasm, which could then be available for the chlamydial pathogen. This would be an intriguing subject for further investigation. Interestingly, the enriched eukaryotic-like domain is annotated as the domain that is found N-terminal of the catalytic site of the biotin protein ligase.

CT153

The eukaryotic-like domain belongs to the MAC/perforin superfamily. Their members are proteins that oligomerize from monomers to oligomeric membrane-spanning pores, exhibiting cytolytic activity in vitro. Taylor et al. (Taylor et al., 2010) showed that CT153 is already present in EBs and is proteolytically processed immediately following infection, suggesting a role in the very early pathogen/host cell interactions. They propose that it is an important factor for the acquisition or modification of host cell derived lipids. The mechanism by which CT153 access the host cell cytosol was not discussed, our results strongly argue for a TTS mechanism.

CPn0091 – CT305

Both proteins were positive in our test of secretion. They are annotated as V-type ATP synthase subunit I which is part of the membrane proton channel due to their enrichment of the eukaryotic-like domain "V-type ATPase". ATP synthases in general have the function to produce ATP from ADP in the presence of a proton gradient across a membrane. It would be interesting to investigate whether this subunit is the only one to be secreted or if the remaining parts which are clustered with subunit I in the genome are also effectors. *Chlamydia* have the capacity to take up host cell derived ATP (Tjaden et al., 1999; Hatch et al., 1982). Consequently, it would be interesting to investigate whether *Chlamydia* enhance the host cell's ATP production in order to transport this source of energy into the bacterial cytosol, and which proton gradient is used to do so.

CPn0489 - CT386

CPn0489 is annotated as hypothetical protein, whereas the predicted function for CT386 is that of a metal dependent hydrolase. Its composite domain can be found in several bacterial and fungal enzymes, e.g. in the virulence factor urease of *Helicobacter pylori* (Davies et al., 2002) . More functional studies have to be done on CPn0489 and CT386 before suggesting a role of the secreted proteins in the host cell. Their eukaryotic-like domain is of an uncharacterized protein family (UPF0160) containing a large number of metal binding residues. The patterns suggest a phosphoesterase function.

CT460

The eukaryotic-like SWIB (YM74) domain is a conserved region of the mammal protein BAF60b, which plays a role in chromatin remodelling. It has been suggested by Bennett-Lovsey et al. (Lovsey et al., 2002) that the SWIB domain in *Chlamydia* is derived from eukaryotes and acts on the condensation and decondensation of the chlamydial genome during its developmental cycle. They also gave evidence for the homology of the SWIB protein domain to MDM2, a eukaryotic inhibitor of the tumour suppressor p53. p53 leads in cooperation with NF-κB to apoptosis of the cell (Ryan et al., 2000). Further studies have to be conducted in order to specify the function that a secreted CT460 could fulfill within the host cell.

CPn0562 – CPn0927 – CPn0928 – CPn0929

These four proteins are homologs with unknown function. They all possess the same domain and belong to the DUF818 (domain of unknown function) superfamily. No homologs are present in the *Chlamydia trachomatis* genome. Strikingly, all four homologs are positive and possess different amino termini, indicating that they are bona fide effectors.

CPn0856 – CCA00911

These homologs are annotated as UTP-glucose-1-phosphate uridylyltransferase (PF01704), which is another name for UDP-glucose pyrophosphorylase. Its function is described above (1.5.2.1). Remarkably, UDP-glucose is a eukaryotic metabolite, hence secretion into the eukaryotic host cell is very likely to occur. This is supported by our secretion results.

CPn0483

Unpublished data of Subtil et al. give evidence for a secretion of this protein at the entry step of *C. pneumoniae*. CPn0483 interacts with NDP52. Mammalian cells ubiquitinate bacterial intruders for distruction by autophagy. Ivanov and Roy (Ivanov and Roy, 2009) show that NDP52 binds to these ubiquinated bacteria and facilitates their degradation. It is thus conceivable that CPn0483 hampers this process and plays a role in the bacterial evasion of the host cell's response. Its eukaryotic-like domain is the peptidase C65 Otubin, a highly specific ubiquitin isopeptidase that removes ubiquitin from proteins.

Proteins that are not secreted:

CPn0811 – CT576

The homologous pair CPn0811 and CT576 are both annotated as "low calcium response protein H (LcrH)" possessing a eukaryotic-like domain that is characterized by tetratricopeptide repeats. These repeats are generally found in chaperones. It has been shown that CPn0811 functions as a chaperone for the TTSS, which is expressed from the middle to late stages of the chlamydial developmental cycle (Faludi et al., 2009). Both were negative for secretion. As mentioned above, chaperones are not secreted and are released from the substrate-chaperone complex by ATPase activity. The description about the function of both proteins fits well with our findings.

CPn0769 – CT643

Both proteins have the predicted function of a DNA topoisomerase I/SWI domain fusion protein. This family of proteins catalyzes the ATP-dependent breakage of single-stranded DNA followed by passage and rejoining and is therefore part of the DNA damage response (Bugreev and Nevinsky, 2009). This process is likely to take place within the bacterium, which supports our negative secretion results. Interestingly, they possess the previously mentioned SWIB/MDM2 domain, which is also present in a positive candidate (CT460).

CPn1021 – CT862

Slepenkin et al. (2005) showed that CPn1021 interacts with chlamydial CopN, a TTSS effector, suggesting that CPn1021 functions as a chaperone. Additionally, Fields and Hackstadt (Fields and Hackstadt, 2000) proposed the same function and interaction for CT862 in *Chlamydia trachomatis*. Both homologs possess the domain including tetratricopeptide repeats (mentioned above) found in chaperones. This would be consistent with our findings.

CPn0887

CPn0887 is annotated as a CHLTR phosphoprotein (NCBI). Phosphoproteins are proteins that are modified post-translationally by phosphorylation. CPn0887 has homologs in *C. caviae* and *C. trachomatis*, both of which possess no enrichment of a eukaryotic-like domain. CPn0887 reveals an enrichment of tetratricopeptide repeats usually found in chaperones.

CCA00743

CCA00743 is a hypothetical protein, no reports about functional studies are available. A BLAST-research did not reveal any homology to bacterial proteins except for *Chlamydophila*. Its eukaryotic-like domain is annotated as Histone methylation protein DOT1 regulating gene expression by methylating histone H3.

In the following we will concentrate on the eukaryotic-like domains that were present in our candidates. Table 12 gives the numbers of positive and negative results we obtained for each domain.

Domain	Positive	Negative	Domain	Positive	Negative
PF01704	2	0	PF08123	0	1
PF07720	0	5	PF05677	4	0
PF01496	1	0	PF09825	2	0
PF10275	1	0	PF03690	2	0
PF02201	1	2	PF02902	2	0
PF01823	1	0	PF01496	1	0

Table 12: Eukaryotic-like domains present in tested candidates. Given are the numbers of positive and negative results obtained in the secretion test.

Interestingly there was one domain (PF02201), which was present in one secreted and in two not secreted candidates. The remaining domains were consistent in their presence in either secreted or non-secreted proteins.

4.2 Tripartite Symbiosis and Enzymes Engaged in Glycogen Metabolism

4.2.1 Secretion Results of Enzymes Engaged in Glycogen Metabolism

The heterologous secretion test revealed that GlgA, GlgB, GlgC and GlgP but not MalQ were secreted by *P. acanthamoeba* (table 13). We cannot make any statement about GlgX, since the secretion results were not clear. Not all of these candidates were tested for *Candidatus Protochlamydia amoebophila*. However, the tested ones (GlgB, GlgX and MalQ) were all positive. What could be the function of these secreted proteins during infection by *Chlamydiales* in the host cell?

4.2.2 Putative Role of Enzymes in *Chlamydiae* Infected Cells

In the beginning of an infection GlgC can promote the synthesis of ADP-glucose in presence of a high cytosolic ATP level (Ballicora et al., 2003). ADP-glucose is neither recognized nor used by the host cell. GlgA and GlgB would increase the glycogen production in the host cell as long as the ATP level remains high. Upon decrease of the host cell's energy stock (decrease of the ratio ATP to Pi) degradation of glycogen to subunits occurs by the function of GlgX, which releases glucose-1-P and maltotetraose (Ball et al., 2011). Interestingly, maltotetraose is not a eukaryotic metabolite and could therefore represent a substrate for import into the chlamydial invader. However, we cannot exclude that host enzymes are able to metabolize maltotetraose into smaller molecules which could also be substrates for import.

Substrate	Metabolized by	Secreted (Pa/CP)	Phylogenetics
Glucose-1-P	GlgC	+/?	-
ADP-glucose	GlgA	+/?	++
Linear glucans	GlgB	+/+	++
Glycogen	GlgP	+/?	-
Phosphorylase limit dextrin	GlgX	?/+	++
Debranched glucans	MalQ	-/+	+

Table 13: Secretion results and phylogenetic relationship between chlamydial protein and archaeplastidal homolog. Given is the substrate of the respective enzyme ("metabolized by"), the secretion results for *P. acanthamoeba* (Pa) and *C. P. amoebophila* (CP), (+) positive for secretion, (-) negative for secretion, (?) unclear or not tested; Phylogenetic analysis: (++) very close grouping between chlamydial and archaeplastidal homolog, (+) grouping between chlamydial and archaeplastidal homolog, (-) no grouping between chlamydial and archaeplastidal homolog.

Further recession and break-down of the sugar chain occurs by secretion of MalQ (not for *P. acanthamoeba*). Manipulation of the host cell's glycogen pool would be an effective way to stock all the ATP which is available at the beginning of the infection in metabolites that cannot be used by the host cell. Even when the energy level of the host cytosol becomes critically low, the parasite would still be provided with carbon. Little is known about suitable transporters in the inclusion and chlamydial membranes to bring these metabolites to the parasite. However, their presence would be a prerequisite for this scenario.

4.2.3 Probability of LGT from *Chlamydiae* to Host Cell

Our phylogenetic analysis suggests that GlgA, GlgB, GlgX and eventually MalQ could have been transferred from *Chlamydiales* (especially the environmental *Chlamydiae*) to the archaeplastidal ancestor by LGT. This is not the case for GlgC and GlgP. We proposed the theory that ADP-glucose was contributed to the host by the cyanobiont. Hence, the side of ADP-glucose production would not be in the eukaryotic host cell. This would explain that the transfer of GlgC was not necessary. The host would nevertheless need enzymes to metabolize this product. LGT of GlgA and GlgB would ensure the synthesis of glycogen. Catabolic enzymes such as GlgX and MalQ would deliver smaller products of degradation. Still, eukaryotic cells possess their own glycogen metabolism. These second "newly-obtained" enzymes engaged in the glycogen metabolism are somehow auxiliary, being consequently prone to genetic changes since they only function as an additional set for glycogen production. It is not the today's glycogen metabolism engaged enzymes in plants that are related to chlamydial genes, but the enzymes in plants engaged in the starch metabolism (see Ball and Morell, 2003 for a review of starch metabolism).

We showed here that it is possible that *Chlamydiae* secreted enzymes engaged in the glycogen metabolism into the host cell. These enzymes might have rescued the host as a secondary effect, since they gave the host cell the opportunity to acquire energy out of another source: the cyanobiont. Subsequently, specific genes of these enzymes could have been transferred by LGT to the host cell. We can support this hypothesis by close relationships of chlamydial GlgA, GlgB, GlgX and MalQ to archaeplastidal homologs. At the beginning they could have fulfilled an auxiliary function in the eukaryotic glycogen synthesis, evolving finally to enzymes necessary for starch production. Stephen Ball's analysis, which we repeated, suggested that chlamydial GlgA is closely related to archaeplastidal soluble starch synthase III/IV according to phylogenetic trees. Archaeplastidal isoamylases (starch debranching enzymes) are also likely to be derived from chlamydial GlgX (Stephen Ball, unpublished data, and our own analysis). The archaeplastidal homolog to GlgB is annotated as either glycogen branching enzyme or starch branching enzyme. The closest archaeplastidal homologs of MalQ are amylomaltose, also called 4-α-glucanotransferase.

So far, the hypothesis of a tripartite symbiosis between a heterotrophic organism, the cyanobiont and a chlamydial parasite is a conceivable scenario, which we strengthened with our demonstration that chlamydial enzymes possess functional TTS signals. However, some links are still missing. We introduced cyanobacterial ADP-glucose as the key substrate that would enforce a tripartite symbiosis. However, it is not known whether ADP-glucose could have been transported into the host cytosol in such an early stage of "preliminary endosymbiosis". Not a lot is known about chlamydial enzymes engaged in glycogen synthesis and their function in the eukaryotic cytosol. A very important aspect would be to track the way of the glycogen pathway in *Chlamydiae* and the host cell upon an infection. For example, are there transporters that can bring the catabolic products of the glycogen metabolism from the cytosol into the bacteria? Is it hence indisputable that secretion of these enzymes takes place, to the profit of the parasite? The unquestionably most important question is however: can we be sure that since over one billion years (the time when this scenario was proposed to take place) the genome of our tested *Chlamydiae* stayed sufficiently conserved to make predictions about their characteristics at that time? The proposed hypothesis is an intriguing subject, which deserves attention and more profound research.

4.3 Secretion Results of *C. trachomatis* and *C. pneumoniae* "Glycogen-Enzymes"

It has been reported that glycogen particels can be detected upon infection with *C. trachomatis* but not with *C. pneumoniae*. These glycogen particles are visible in the inclusion,

in EBs and RBs (Chiappino et al., 1995). They appear at 20-30 hours post-infection and peak at 30-60 hours post-infection (Iliffe-Lee and McClarty, 2000). It is typically assumed that the particles in the inclusion lumen are due to ruptured bacteria, and that these particles are originally synthesized within the bacteria. We will here briefly discuss the secretion results we obtained for the *C. trachomatis* and *C. pneumoniae* enzymes engaged in the gycogen metabolism. Both, *C. trachomatis* and *C. pneumoniae* possess a full set of homologs for GlgA, GlgB, GlgC, GlgP, GlgX and MalQ, even though glycogen has only been detected in *C. trachomatis*. GlgC seems not to be secreted in neither *C. trachomatis* nor *C. pneumoniae*. Interestingly, GlgA is secreted in both, GlgB only in *C. trachomatis*. We here ask the unsolved question why GlgA, which uses ADP-glucose as substrate, is secreted. ADP-glucose is not available in the hosts of *C. trachomatis* and *C. pneumoniae*. In a previous test we demonstrated that CPn0856, a UDP-glucose-pyrophosphorylase producing UDP-glucose as a substrate is secreted. Is it possible that GlgA of *C. trachomatis* and *C. pneumoniae* are able to use mammal UDP-glucose as a substrate? This has never been tested. It is to mention that these *Chlamydiales* are much younger in their evolution than environmental *Chlamydiae*. Their GlgA homologs could possibly have evolved as an adaptation to use UDP-glucose as substrate, which is indeed produced by their host cell, eventually also due to secretion of the chlamydial UDP-glucose pyrophosphorylase. GlgB is only secreted by *C. trachomatis*. Glycogen particles are only found in *C. trachomatis*. Is it possible that the detected glycogen particles in the inclusion are no artefact of lysed bacteria, but that glycogen is actually stored in the inclusion of *C. trachomatis*? Does in contrast *C. pneumoniae* synthesize unbranched glycogen in the inclusion, which have never been detected? Or does another pathway exist for further processing of these elongated unbranched sugar chains in *C. pneumoniae*?

Both pathogens secrete the catabolic enzymes necessary for glycogen break-down (GlgP, GlgX, MalQ). Glycogen is a product of their eukaryotic host, allowing the intruders to parasite this energy source. Again, this theory is only coherent when transporters for the break-down products into the bacteria are present.

We asked here a lot of unsolved questions, which we will partially try to further examine in the future.

4.4 Efficiency of Heterologous Secretion Test

Due to the lack of genetic means to manipulate *Chlamydiae* the heterologous screen is a very valuable tool to test whether proteins are or are not secreted by a type III machinery. Positive and negative controls that were included in this and in previous tests proved its reliability.

However, this screen is performed on chimeras containing the amino terminus of the protein of interest. As it has already been mentioned, there have been hints for the existance of a second signal, the CBD (chaperone-binding domain). For several full-length effectors it has been shown that the association with a specific chaperone is important for an efficient secretion. In our clones the CBD is excluded. Additionally, it is not likely that all homologs of chlamydial chaperones are encoded in the *Shigella* genome. Even though we can suppose that our constructs are too short to be obligatory unfolded for functional translocation, thereby bypassing the need for CBD, we do not know whether the amino terminal secretion signal is sufficient in all proteins.

Additionally, we cannot exclude the occurence of false positives. However, Subtil et al. (Subtil et al., 2005) showed that this occurence is below 5%.

5. Summary

Chlamydiae are obligate intracellular parasites infecting a broad spectrum of organisms such as animals, insects and amoeba. These gram-negative bacteria are worldwide a major cause of preventable blindness and infertility in humans. Upon infection of a eukaryotic host cell *Chlamydiae* multiply within a parasitophorous compartment termed inclusion, where they also undergo conversion from an infectious, metabolically rather inert form to a non-infectious, metabolically active form. Both forms possess a type III secretion system in order to translocate potentially toxic effector proteins to targets within the host cell.

No genetic tools to manipulate *Chlamydiae* are available so far due to their obligate intracellular lifestyle, hampering the examination of putative secreted effectors. The identification of type III secreted proteins is additionally complicated by the fact that the molecular recognition of effectors by a type III machinery is still elusive. However, various experiments suggest a signal in the 20 amino terminal amino acids of the effectors.

We made use of the already published approach of testing putative chlamydial effectors in a heterologous type III secretion system of *Shigella flexneri* in order to identify novel type III secreted proteins. For this purpose we designed chimeras consisting out of the 20 amino terminal amino acids of a candidate protein fused to a reporter, the calmodulin-dependent adenylate cyclase (Cya) of *Bordetella pertussis*. These chimeras were expressed in different strains of *Shigella flexneri*. By applying an antibody against Cya we could localize the chimeras and determine their characteristics of secretion.

In order to create a list of candidates that we subsequently subjected to this secretion test we made use of the finding that many effectors have been shown to contain protein domain signatures that are typically found in eukaryotes. The software "Effective" gave us precalculated lists for proteins enriched in these "eukaryotic-like domains" in the proteomes of *Chlamydia trachomatis*, *Chlamydophila pneumoniae* and *Chlamydophila caviae*. With the previously mentioned heterologous secretion test we demonstrated that 17/25 chosen candidates were positive for type III secretion. The software "Effective" is thus a very useful approach, but it can only be used additionally, since other proteins we tested as positive were not listed by this software.

Additionally, we tested an assortment of proteins of environmental *Chlamydiae* engaged in glycogen metabolism for secretion. This assortment was chosen regarding the function these enzymes could have fulfilled to stabilize the formation of endosymbiosis between a

cyanobiont and a heterotrophic organism. We showed that almost all of these enzymes are indeed secreted and that some of them seem to be closest related to plant homologs, strengthening the hypothesis of an indispensable role of *Chlamydiae* in the early history of development of today's plants.

We further examined if these enzymes have different secretion characteristics in *C. trachomatis* and *C. pneumoniae*, knowing that only *C. trachomatis* accumulates glycogen. We could state differences here and gave suggestions for further thoughts.

6. References

Aizawa, S.I. (2001) Bacterial flagella and type III secretion systems. FEMS Microbiol Lett. 202(2):157-64

Akeda, Y. and Galàn, J.E. (2005) Chaperone release and unfolding of substrates in type III secretion. Nature. 437(7060):911-5

Alloui, A., Sansonetti, P., Parsot, C. (1993) MxiD, an outer membrane protein necessary for the secretion of the Shigella flexneri Ipa invasins. Mol Microbiol. 7(1):59-68

Anderson, D.M. and Schneewind, O. (1997) A mRNA signal for the type III secretion of Yop proteins by Yersinia enterocolitica. Science. 278(5340):1140-3

Arnold, R., Brandmaier, S., Kleine, F., Tischler, P., Heinz, E., Behrens, S., Ninikoski, A., Mewes, H.W., Horn, M., Rattei, T. (2009) Sequence-based prediction of type III secereted proteins. PloS Pathog. 5(4):e1000376

Arnold, R., Jehl, A., Rattei, T. (2010) Targeting effectors: the molecular recognition of Type III secreted proteins. Microbes Infect. 12(5):346-58

Ball, S., Colleoni, C., Cenci, U., Raj, N.M., Tirtiaux, C. (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J Exp Bot. 62(6):1775-801

Ball, S. and Morell, M.K. (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Plant Cell. 54:207-33

Ballicora, M.A., Iglesias, A.A., Preiss, J (2003) ADP-Glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogensynthesis.
Belland, R.J., Zhong, G., Crane, D.D., Hogan, D., Sturdevant, D., Sharma, J., Beatty, W.L., Caldwell, H.D. (2003) Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis. Proc Natl Acad Sci U S A. 100(14):8478-83

Bennett-Lovsey, R., Hart, S.E., Shirai, H., Mizuguchi, K. (2002) The SWIB and the MDM2 domains are homologous and share a common fold. Bioinformatics. 18(4):626-30

Betts-Hampikian, H.J. and Fields, K.A. (2010) The Chlamydial Type III Secretion Mechanism: Reveiling Cracks in a Tough Nut. Front Microbiol. 1:114

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P., Parsot, C. (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. Mol Microbiol. 38(4):760-71

Bugreev, D.V. and Nevinsky, G.A. (2009) Structure and mechanism of action of type IA DNA topoisomerases. Biochemistry (Mosc). 74(13):1467-81
Bush, R.M. and Everett, K.D. (2001) Molecular evolution of the Chlamydiaceae. Int J Syst Evol Microbiol. 51(Pt 1):203-20

References

Byrne, G.I. and Ojcius, D.M. (2004) Chlamydia and apoptosis: life and death decisions of an intracellular pathogen. Nat Rev Microbiol. 2(10):802-8

Carabeo, R.A., Grieshaber, S.S., Fischer, E., Hackstadt, T. (2002) Chlamydia trachomatis induces remodelin of the actin cytoskeleton during attachement and entry into HeLa cells. Infect Immun. 70(7):3793-803

Cavalier-Smith, T. (2006) Cell evolution and Earth history: stasis and revolution. Philos Trans R Soc Lond B Biol Sci. 361(1470):969-1006

Cazalet, C., Rusniok, C., Brüggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F., Kunst, F., Etienne, J., Glaser, P., Buchrieser, C. (2004) Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. Nat Genet. 36(11):1165-73

Cheng, L.W., Anderson, D.M., Schneewind, O. (1997) Two independent type III secretion mechanisms for YopE in Yersinia enterocolitica. Mol Microbiol. 24(4):757-65

Chiappino, M.L., Dawson, C., Schachter, J., Nichols, B.A. (1995) Cytochemical localization of glycogen in Chlamydia trachomatis inclusions. J Bacteriol. 177(18):5358-63

Clifton, D.R., Fields, K.A., Grieshaber, S.S., Dooley, C.A., Fischer, E.R., Mead, D.J., Carabeo, R.A., Hackstadt, T. (2004) A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. Proc Natu Acad Sci U S A. 101(27):10166-71

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R.C., Read, T.D., Bavoil, P.M., Sachse, K., Kahane, S., Friedman, M.G., Rattei, T., Myers, G.S., Horn, M. (2011) Unity in Variety – the Pan-Genome of the Chlamydiae. Mol Biol Evol. (Epub)

Collingro, A., Toenshoff, E.R., Taylor, M.W., Fritsche, T.R., Wagner, M., Horn, M. (2005) 'Candidatus Protochlamydia amoebophila', an endosymbiont of Acanthamoeba spp. Int J Syst Microbiol. 55(Pt 5):1863-6

Davies, B.J., de Vries, N., Rijpkema, S.G., van Vliet, A.H., Penn, C.W. (2002) Transcriptional and mutational analysis of the Helicobacter pylori urease promoter. FEMS Microbiol Lett. 213(1):27-32

Dunn, J.D. and Valdivia, R.H. (2010) Uncivil engineers: Chlamydia, Salmonella and Shigella alter cytoskeleton architecture to invade epithelial cells. Future Microbiol. 5(8):1219-32

Faludi, I., Csanádi, A., Szabó, A.M., Burián, K., Endrész, V., Miczák, A. (2009) Production and purification of low calcium response protein H of Chlamydophila pneumoniae. Acta Microbiol Immunol Hung. 56(4):389-97

Fields, K.A. and Hackstadt, T. (2000) Evidence for the secretion of Chlamydia trachomatis CopN by a type III secretion mechanism. Mol Microbiol. 38(5):1048-60

Frithz-Lindsten, E., Holmström, A., Jacobsson, L., Soltani, M., Olsson, J., Rosqvist, R., Forsberg, A. (1998) Functional conservation of the effector protein translocators PopB/YopB

References

and PopD/YopD of Pseudomonas aeruginosa and Yersinia pseudotuberculosis. Mol Microbiol. 29(5):1155-65

Galàn, J.E., Collmer, A. (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. Science. 284(5418):1322-8

Gray, M.W. (1993) Origin and evolution of organelle genomes. Curr Opin Genet Dev. 3(6):884-90

Gregory, D.W. and Schaffner, W. (1997) Psittacosis. Semin Respir Infect. 12(1):7-11

Haider, S., Wagner, M., Schmid, M.C., Sixt, B.S., Christian, J.G., Häcker, G., Pichler, P., Mechtler, K., Müller, A., Baranyi, C., Toenshoff, E.R., Montanaro, J., Horn, M. (2010) Raman microspectroscopy reveals long-term extracellular activity of chlamydiae. Mol Microbiol. 77(3):687-700

Hatch, T.P., Al-Hossainy, E., Silverman, J.A. (1982) Adenine nucleotide and lysine transport in Chlamydia psittaci. J Bacteriol. 150(2):662-70

Hayashi, Y., Nakamura, S., Matsuo, J., Fukumoto, T., Yoshida, M., Takahashi, K., Mizutani, Y., Yao, T., Yamaguchi, H. (2010) Host range of obligate intracellular bacterium Parachlamydia acanthamoebae. Microbiol Immunol. 54(11):707-13

Ho, T.D. and Starnbach, M.N. (2005) The Salmonella enterica serovar typhimurium-encoded type III secretion systems can translocate Chlamydia trachomatis proteins into the cytosol of host cells. Infect Immun. 73(2):905-11

Horn, M (2008) Chlamydiae as Symbionts in Eukaryotes. Annu Rev Microbiol. 2008;62:113-31.

Huang, J. and Gogarten, J.P. (2007) Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol. 8(6):R99

Iliffe-Lee, E.R. and McClarty, G. (2000) Regulation of carbon metabolism in Chlamydia trachomatis. Mol Microbiol. 38(1):20-30

Ivanov, S. and Roy, C.R. (2009) NDP52: the missing link between ubiquitinated bacteria and autophagy. Nat Immunol. 10(11):1137-9

Jehl, A., Arnold, R., Rattei, T., (2011) Effective—a database of predicted secreted bacterial proteins. Nucleic Acids Res. 39(Database issue):D591-5

Kuo, C.C., Jackson, L.A., Campbell, L.A., Grayston, J.T. (1995) Chlamydia pneumoniae (TWAR). Clin Microbiol Rev. 8(4):451-61

Lake, J.A. and Rivera, M.C. (1994) Was the nucleus the first endosymbiont? Proc Natl Acad Sci U S A. 12;91(8):2880-1

Lee, S.H. and Galàn, J.E. (2004) Salmonella type III secretion-associated chaperones confer secretion pathway specifity. Mol Microbiol. 51(2):483-95

<center>References</center>

Leitsch, D., Köhsler, M., Marchetti-Deschmann, M., Deutsch, A., Allmaier, G., König, L., Sixt, B.S., Duchêne, M., Walochnik, J. (2010) Proteomic aspects of Parachlamydia acanthamoebae infection in Acanthamoeba spp. ISME J. 4(11):1366-74

Lloyd, S.A., Norman, M., Rosqvist, R., Wolf-Watz, H. (2001) Yersinia YopE is targeted for type III secretion by N-terminal, not mRNA, signals. Mol Microbiol. 39(2):520-31

Lurie-Weinberger, M.N., Gomez-Valero, L., Merault, N., Glöckner, G., Buchrieser, C., Gophna, U. (2010) The origins of eukaryotic-like proteins in Legionella pneumophila. Int J Med Microbiol. 300(7):470-81

Ménard, R., Sansonetti, P., Parsot, C. (1994) The secretion of Shigella flexneri Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. EMBO J. 13(22):5293-302

Misaghi, S., Balsara, Z.R., Catic, A., Spooner, E., Ploegh, H.L., Starnbach, M.N. (2006) Chlamydia trachomatis-derived deubiquitinating enzymes in mammalian cells during infection. Mol Microbiol. 61(1):142-50

Moustafa, A., Reyes-Prieto, A., Bhattacharya, D. (2008) Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS One. 3(5):e2205

Nguyen, L., Paulsen, I.T., Tchieu, J., Hueck, C.J., Saier, M.H. Jr. (2000) Phylogenetic analyses of the constituents of Type III protein secretion systems. J Mol Microbiol Biotechnol. 2(2):125-44

Pacheco-Alvarez, D., Solórzano-Vargas, R.S., Del Río, A.L. (2002) Biotin in metabolism and its relationship to human disease. Arch Med Res. 33(5):439-47

Pandey, D., Shetty, J., Pai, M.V., Pratapkumar (2009) Chlamydia in female reproductive tract . The Internet Journal of Infectious Diseases.Volume 7 Number 1

Panina, E.M., Mattoo, S., Griffith, N., Kozak, N.A., Yuk, M.H., Miller, J.F. (2005) A genome –wide screen identifies a Bordetella type III secretion effector and candidate effectors in other species. Mol Microbiol. 58(1):267-79

Peters, J., Wilson D.P., Myers, G., Timms, P., Bavoil, P.M. (2007) Type III secretion à la Chlamydia. Trends Microbiol. 15(6):241-51

Ramamurthi, K.S., Schneewind, O. (2002) Yersinia enterocolitica type III secretion: mutational analysis of the yopQ secretion signal. J Bacteriol. 184(12):3321-8

Ramamurthi, K.S., Schneewind, O. (2003) Substrate recognition by the Yersinia type III protein secretion machinery. Mol Microbiol. 50(4):1095-102

Rogers, D.G., Andersen, A.A., Hunsaker, B.D. (1996) Lung and nasal lesions caused by a swine chlamydial isolate in gnotobiotic pigs. J Vet Diagn Invest. 8(1):45-55

Rosqvist, R., Hakansson, S., Forsberg, A., Wolf-Watz, H. (1995) Functional conservation of the secretion and translocation machinery for virulence proteins of yersiniae, salmonellae and shigellea. EMBO J. 14(17):4187-4195

References

Rüssmann, H., Kubori, T., Sauer, J., Galàn, J.E. (2002) Molecular and functional analysis of the type III secretion signal of the Salmonella enterica InvJ protein. Mol Microbiol. 46(3):769-79

Ryan, K.M., Ernst, M.K., Rice, N.R., Vousden, K.H. (2000) Role of NF-kappaB in p53-mediated programmed cell death. Nature. 404(6780):892-7

Schoberg, R.V. (2011) Chlamydia persistence – a tool to dissect chlamydia – host interactions. Microbes Infect. 13(7):649-62

Shea, J.E., Hensel, M., Gleeson, C., Holden, D.W. (1996) Identification of a virulence locus encoding a second type III secretion system in Salmonella typhimurium. Proc Natl Acad Sci U S A. 93(6):2593-7

Slepenkin, A., de la Maza, L.M., Peterson, E.M. (2005) Interaction between components of the type III secretion system of Chlamydiaceae. J Bacteriol. 187(2):473-9

Stamm, W.E. (1999) Chlamydia trachomatis infection: progress and problems. J Infect Dis. 179 Suppl 2:S380-3

Stebbins, C.E. and Galàn, J.E. (2001) Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. Nature. 414(6859):77-81

Stephens, R.S., Kalman S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V., Davis, R.W. (1998) Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Science. 282(5389):754-9

Preiss, J. and Romeo, T. (1994) Molecular biology and regulatory aspects of glycogen biosynthesis in bacteria. Prog Nucleic Acid Res Mol Biol. 47:299-329

Preiss, J., Bloom, M., Morell, M., Knowles, V.L., Plaxton, W.C., Okita, T.W., Larsen, R., Harmon, A.C., Putnam-Evans, C. (1987) Regulation of starch synthesis:enzymological and genetic studies. Basic Life Sci. 41:133-52

Stephens, R.S., Myers, G., Eppinger, M., Bavoil, P.M. (2009) Divergence without difference: phylogenetics and taxonomy of Chlamydia resolved. FEMS Immunol Med Microbiol. 55(2):115-9

Subtil, A. and Dautry-Varsat, A. (2004) Chlamydia: five years A.G. (after genome). Curr Opin Microbiol. 7(1):85-92

Subtil, A., Delevoye, C., Balaná, M.E., Tastevin, L., Perrinet, S., Dautry-Varsat, A. (2005) A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates. Mol Microbiol. 56(6):1636-47

Subtil, A., Parsot, C., Dautry-Varsat, A. (2001) Secretion of predicted Inc proteins of Chlamydia pneumoniae by a heterologous type III machinery. Mol Microbiol. 39(3):792-800

References

Taylor, L.D., Nelson, D.E., Dorward, D.W., Whitmire, W.M., Caldwell, H.D. (2010) Biological characterization of Chlamydia trachomatis plasticity zone MACPF domain family protein CT153. Infect Immun. 78(6):2691-9

Tjaden, J., Winkler, H.H., Schwöppe, C., Van Der Laan, M., Möhlmann, T., Neuhaus, H.E. (1999) Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy. J Bacteriol. 181(4):1196-202.

Tobe, T., Beatson, S.A., Taniguchi, H., Abe, H., Bailey, C.M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., Pallen, M.J. (2006) An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. Proc Natl Sci U S A. 103(40):14941-6

Tseng, T., Tyler, B.M., Setubal, J.C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol. 19;9 Suppl 1:S2

Valls, M., Genin, S., Boucher, C. (2006) Integrated regulation of the type III secretion system and other virulence determinants in Ralstonia solanacearum. PloS Pathog. 2(8):e82

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., Bhattacharya, D. (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol. 21(5):809-18

7. Appendix

7.1 Sequences of the inserts cloned in the pUC19cya vector

Upstream of the initiation codon, the sequences provide a HindIII restriction site (if not BsaI, marked with an asterisk), a stop codon terminating translation from the α-Galactosidase sequence in pUC19, and about 10 nucleotides as spacer between this stop codon and the initiation codon of the gene of interest. The stop codon was spared when present in this spacer sequence. The constructs include approximately the first 20-30 codons of the gene of interest, followed by a XbaI site (if not BsaI, marked with an asterisk), for cloning into the pUC19cya vector. If site of initiation could not be unambigously identified, two constructs were designed (marked with a plus). Restriction sites are underlined.

Parachlamydia acanthamoebae:

GlgA
AAGCTTGTAATAGTTTTGTTTTTATGCATATTATTCATATCGCAGCAGAACTTGCA
CCACTTGCCAAAGTCGGCGGCCTTGCAGATGTTGTTCTCGGACTCTCTCGTGAACT
ATCTAGA

GlgAΔ10
AAGCTTGTAATAGTTTTGTTTTTATGGCACCACTTGCCAAAGTCGGCGGCCTTGCA
GATGTTGTTCTCGGACTCTCTCGTGAACTATCTAGA

GlgC
AAGCTTGTAATAGTTTTGTTTTTATGTCTTTGTTAACCACACCCCATGTAAAAACG
ACTCCACTAACACAAACAATCAATTTGCACACACACCGCACAGATCGGGTTGCAT
CCTCTAGA

MalQ⁺
AAGCTTGTAATAGTTTTGTTTTTATGACAAACTTTTTTCTGCAGCAATTGCTTGAA
AATCCTGATAGCCTTCCACACCATGGAATTTGCCTTCCCATTTTCTCTTTACACTC
CTCTAGA

MalQbis⁺
AGTCAAGCTTGTAATTACTGAGGAATGGTAACGATGACAAACTTTTTTCTGCAGC
AATTGCTTGAAAATCCTGATAGCCTTCCACACCATGGAATTTGCCTTCCCATTTTC
TCTAGAGACT

GlgP₁
AAGCTTGTAATAGTTTTGTTTTTATGAGTACATTAGCGGATTTGCCCTCTCAGGAG
AAGGTTAGTCAGGATAAATCGATTCGTACTGGATTAAGTGTTGAATCTTTAAAAA
AATCTAGA

GlgP$_2$

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGACGCAACCAGCACCCGATCTTGATTATCAA
GCAGAAATGTTAGCTGCTAAACAAAGCATTATTTGATTACAACCATGGGGCGTA
TC<u>TCTAGA</u>

GlgX

<u>AAGCTT</u>CTAATTCCTTTAATTTAGGCAAATACGTGTCAGAGCTTCAGTCTTTTCAA
GTAGAAAAGGATCCCCTCTTAACTTAGGTATCTTTGGTGATTGCGGCGGAATCA
ATTTTTCCCTT<u>TCTAGA</u>

GlgB

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGAAACGCACACATCTCCCCTTACCCACACT
CAATTTGATTCATTACTTGCTGGAGAAGCCTTTGACCCCCACCAATTTTTAGGACT
A<u>TCTAGA</u>

Candidatus Protochlamydia amoebophila

GlgX

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGAATATACAAATAACACCTGGATCCCCCTTT
CCTTTTGGTGCAAATATACAAGAAGGAAAGGTTAATTTTGCTCTTTATGCTAAAA
ATT<u>CTAGA</u>

MalQ

<u>AAGCTT</u>GTAATAGAAGCGAAAATTTGGTCCGATGCTCTCAATCAAGGAAATGTTC
CGTATATTACAGAAATTACCAATACTCCTCCTGCACCTCCATTTGGTTATCAAAGA
TGATAGATCCTACTTTCTTACTTCATTCTCTTGCGGCTAAACAGTGGGAGCGAATC
GGTATCAAGCACCATCATGGAATCAATGTTCCT<u>TCTAGA</u>

GlgB

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGATGACTATGCAACAGACTGAATTTGACTCT
CAATTTAATGAACACATTTATCGTATCGTTCATGTTGTTCATCACCAACCTCATGC
TT<u>CTAGA</u>

Chlamydia trachomatis

CT862-61[+]

<u>AAGCTT</u>ATAGGCCATCTCCAAGAAGTGTCTCAATGCCACCAAGCAAGATCCAATG
TCTTGAAACTTTTAAAAGAACTTATGGACACCTTTATCTACAACATGCGTCCCTAA
TGCGTCATTTAGCCTATCTACTCGATAAAATTGCTCGCTCTTACCATCATATGTGT
CCGCTTCCCGATAATATG<u>TCTAGA</u>

CT862-87[+]

<u>AAGCTT</u>ATAGAGGACACCTTTATCTACAACATGCGTCCCTAATGCGTCATTTAGC
CTATCTACTCGATAAAATTGCTCGCTCTTACCATCATATGTGTCCGCTTCCCGATA
ATATG<u>TCTAGA</u>

CT035

<u>AAGCTT</u>ATAGAGGCAAAGAGATGAAGCGTATCTTAGTGTATTCGGATAGAGGAG
TTTCTCCTTACTATTTGCGCCATACTGTTCGCTGGTTGAAGCAGGTAGCTGCTCCA
TTCCAGT<u>TCTAGA</u>

CT153

AAGCTTGTTCCTTCCATTATAGGGTGTCAATCACTGTCCAGCCCGGAGAGTTTTTA
ATGACTAAGCCTTCTTTCTTATACGTTATTCAACCTTTTTCCGTATTTAATCCACGA
TTAGGACGTTTCTCTTCTAGA

CT460

AAGCTTCTTTTCTTATCATCTTCTTTAACTAGGAGTCATCCATGAGTCAAAATAAG
AACTCTGCTTTCATGCAGCCTGTGAACGTATCCGCTGATTTAGCTGCCATCGTTGG
TTCTAGA

CT305

AAGCTTTTAGGAGTAGTCTGCATGCGCGTAGATGTGGATAAATATCTATTTATTG
GACGTGAGAAGTCTGAATTTTTCTCTGCATGTCGAGAGATTGGGGCTGTCGAATT
TTCTAGA

CT576

AAGCTTTTAGGAATTATCGCGATGAGCACTCCATCTTCTAATAATTCTAAAAAAC
CTTCGGCCTCTTTTAATAAAAAATCACGTAGCCGCTTGTCTAGA

MalQ

AAGCTTGTAATAGTTTTGTTTTTATGCCGTCATTATCCCAATCCCGACGTATCATC
CAGCAATCTTCCATTCGAAAGATTTGGAATCAGATAGATACTTCTCCTAAGCATG
GCTCTAGA

GlgA

AAGCTTGTAATAGTTTTGTTTTTATGAAAATTATTCACACAGCTATCGAATTTGCT
CCGGTAATCAAAGCCGGAGGCCTGGGAGACGCGCTATACGGACTAGCATCTAGA

GlgB

AAGCTTGTAATAGTTTTGTTTTTATGGATCCTTTTTTCTTAAATACTCAACACGTG
GAACTTCTCGTTTCTGGTAAACAGAGCAGTCCACAAGATCTTTTGGGAATTGTTTC
TTCTAGA

GlgC

AAGCTTGTAATAGTTTTGTTTTTATGGCTGGTAGACGGACGAAAGAAGAGCAGAT
CAATCGAAAACGATCGCATTTCTATCGAGATAACGTAGGAGTTATTGTCTTATGC
GGATCTAGA

GlgP

AAGCTTGTAATAGTTTTGTTTTTATGTATTTCGATCGGACAAAGATCAATGTTGAA
TCTATGAAGCAAGCTATCCTCGAAAGGGTATATTGTGGGGTAGTCCAGACTCCTC
AATCTAGA

GlgX

AAGCTTGTAATAGTTTTGTTTTTATGGAATCTTTGTCTGTTCGTTCCACTATCCCTT
TACCTCTAGGAGCCAAAAAGCTCTCCGCTGATCGCTACCGTTTTTCTCTATTTTCT
TCTAGA

Chlamydophila caviae

CCA00911

<u>AAGCTT</u>ATCTAACATCAAGAGATAGGAACGCAAGGCTAACTGTAATGACTGACTC
TGTAACCTTTCCTTCTGCTGTGGAAATGTCTTCACTAACAGAAAAACTCAAGTCTA
TTAATCAAGAGCATT<u>CTAGA</u>

CCA00743

<u>AAGCTT</u>ATAGACCGATTTACAAGTTGAAGAAGATATATTTGCTATAGAAGATGTC
GTATTTCAACCGCCTAAAGAGCTCCGTAATTAAAAGTCTTCATTCTTACATATAT
TGCGAGTAGGAATGCAT<u>TCTAGA</u>

Chlamydophila pneumoniae

CPn0929

<u>AAGCTT</u>ATAGTTAAATTTTATAGGAAAAGTTCATGGCTCCAATTCACGGAAGTAA
TGCGTTTGTTGAGGATATTTTACATTCCCACCCTTCTCCACAAGCGACTTATTTT<u>TC
TAGA</u>

CPn1021

<u>AAGCTT</u>ATAGAGGACACCTTCATTTACAACCTACGCCCCTAATGTCACATTTAAA
TTATTTACTAGAAAAAATCGCTGCATCTTCCAAGGAAGACTTCCCTTTCCCAGATG
AT<u>TCTAGA</u>

CPn0562

<u>AAGCTT</u>ATAGACTACATTCTGCATGAAATTATAAGGTAATAATGTCAATAGCTAT
TGCAAGGGAACAATACGCAGCTATATTGGATATGCATCCTAAACCTTCGATCGCC
ATG<u>TCTAGA</u>

CPn0128

<u>AAGCTT</u>ATAGCGCTTTTCTGTTCGAGAGGAAAATAAGAAGATAGGGGAAGAATG
TTAAGGAATCAGGTACTTGTTTACTGTAGTGAGGGTGTTTCTCCTTATTATTTACG
GCATACGATACGT<u>TCTAGA</u>

CPn0856

<u>AAGCTT</u>GGATCAAGAGATAGGAACGTAAGGCTAACTGTAATGACTGAATCGGTA
TATTCGCCCTCTGCTATGCATGTAAACTCTCTAGCAGATAAGCTCAAAGCCATA<u>TC
TAGA</u>

CPn0091

<u>AAGCTT</u>AAGATAGAACTCCGGAAAGCAAGGGGGGATGAGTGCGTTTAAATATAC
ATAAGTATCTCTTTATAGGACGCAATAAGGCGGATTTTTTTTCTGCAAGTAGAGA
GCTTGGT<u>TCTAGA</u>

CPn0928

<u>AAGCTT</u>AGAGACGTAAGGATTTCGCATTCACTCTTTTGAATCTTTCAAACAGGTCT
GATATTTTGTCAGGAATATTTTCAAATCCTCATCCAGTTTCCTATTTTTCGTCA<u>TCT
AGA</u>

CPn0489

AAGCTTGTTACAATGGAGGATTGGCTAAGGAGGATAGTAGGTATGCAGATTCCA
AGAAGCATTGGTACTCACGATGGTTCTTTCCATGCGGATGAGGTCACAGCGTGTT
CTAGA

CPn0811
AAGCTTTTAGGAATTAGATCGATGAGCAAGCCCTCTCCTCGTAATGCCAATCAAC
CTCAAAAACCTTCAGCCTCTTTCAATAAAAAAACGCGAAGCCGTCTATCTAGA

CPn0927*
AAGCTTCTTTTGTTCTCTTTATACTCTGTGTACCTAAAATTTAGGACTCTTGTATGA
TCCCATCCCCTACCCCAATAAACTTTCGTGATGATACGATTCTAGAGACGGATCC
AAAGCCGTCTAGTGAGACC

CPn0577
AAGCTTGTAGTCAGGTCTCTTCTTACCACCTTTACTAGGAGTCACCAATGAGTCAA
AAAAATAAAAACTCTGCTTTTATGCATCCCGTGAATATTTCCACAGATTTAGCAG
TTATAGTTGGCAAGGGACCTATGTCTAGA

CPn0176*
GGTCTCTAGCTAGGCGAAGTTAAGGAAGCTCTGAAACGATGATGATCCTATGGAT
GAATCCGATGGAGAAGAAGCTTCAAAAGATTCTGCATTTTCAGCTAGTTTTTCCT
ATGAGTTTTCTAGA

CPn1018
AAGCTTCTGGTTTACAAGTGGAAAATGATATATTTGCTGTAGAGGATGTCGTATT
TCAATTACCAAAAGAACTCCGTTGTTCTTAGGTCTTTAGGCTTACTTGCCAAGTCT
AGA

CPn0769
AAGCTTATAGAGGTCCTCACGCAATTAGATTAATGAAAAAGTCCTTAATTATAGT
AGAATCACCTGCAAAAATTAAAACGCTACAAAAATTATTAGGGAGTGAATTTGTT
TTTGCCTCTAGA

MalQ
AAGCTTGTAATAGTTTTGTTTTTATGAATGTTTTAAAATACACAAAACACTCACCC
TCAGCACATGCTTGGAAACTTATAGGAACCTCTCCTAAACACGGGATTTATCTCC
CATCTAGA

GlgA
AAGCTTGTAATAGTTTTGTTTTTATGAGAATCGTACAAGTCGCTGTAGAATTCACT
CCAATCGTTAAAGTAGGCGGTCTAGGCGATGCTGTAGCTAGTCTATCTAAGGAGT
TATCTAGA

GlgB
AAGCTTGTAATAGTTTTGTTTTTAGTGTTGATAAACTGATCCATCCTTGGGATCTT
GATCTGCTCGTCTCAGGACGACAGAAGATCCCCATAAACTCTTAGGGATCCTTG
CTTCTAGA

GlgC

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGATAGAAAACGATTTTCCGGAGGCCTCAAAT
TTTGAGAGCTCTCATTTTTATCGAGATAAGGTTGGAGTAATTATCTTGTGTGGAGG
G<u>TCTAGA</u>

GlgP

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGAAGATTTTTCGAGTTTTGATAAGAACAAA
GTCAGTGTTGACTCTATGAAACGGGCGATTTTAGATCGTCTGTATTTAAGTGTTGT
A<u>TCTAGA</u>

GlgX

<u>AAGCTT</u>GTAATAGTTTTGTTTTTATGGAAAAAGTTTCTTCTTATCCCTCAGTTCCTT
TACCTCTTGGGGCTTCTAAAATTTCCCCAAACCGCTATCGATTTGCTTTATATGCT
<u>TCTAGA</u>

7.2 Details of phylogenetic trees

All trees are cropped for lack of space. Bootstrap cut-off: 50%.

GlgA – Glycogen Synthase

Selaginella moellendorffii
Picea glauca
Solanum lycopersicum
Populus trichocarpa
Vitis vinifera
Chlamydomonas reinhardtii
Volvox carteri f nagariensis
Chlorella variabilis
Ostreococcus lucimarinus CCE9901
Micromonas pusilla CCMP1545
Micromonas sp RCC299
Physcomitrella patens subsp patens
Physcomitrella patens subsp patens
Physcomitrella patens subsp patens
Physcomitrella patens subsp patens
Physcomitrella patens subsp patens
Physcomitrella patens subsp patens
Physcomitrella patens subsp patens
Hordeum vulgare subsp vulgare
Oryza sativa Japonica Group
Sorghum bicolor
Oryza sativa Japonica Group
Oryza sativa Japonica Group
Solanum lycopersicum
Arabidopsis thaliana
Vitis vinifera
Vigna unguiculata
Populus trichocarpa
Ricinus communis
Ricinus communis
Chlamydia trachomatis D-EC
Chlamydia muridarum
Chlamydophila pneumoniae
Chlamydophila psittaci 6BC
Chlamydophila caviae
Chlamydophila felis FeC-56
Candidatus Protochlamydia amoebophila UWE25
Parachlamydia acanthamoebae str Halls coccus
Allochromatium vinosum DSM 180
Magnetococcus sp MC-1
Allochromatium vinosum DSM 180
Methylococcus capsulatus
Nitrosococcus halophilus Nc4
Nitrosococcus oceani ATCC 19707
Cyanothece sp PCC 7425
Synechococcus sp JA-3-3Ab
Lyngbya sp PCC 8106
Arthrospira platensis NIES-39
Microcoleus chthonoplastes PCC 7420
Cylindrospermopsis raciborskii CS-505
Anabaena variabilis ATCC 29413
Microcystis aeruginosa PCC 7806
Cyanothece sp PCC 8802
Cyanothece sp CCY0110
Crocosphaera watsonii WH 8501
Cyanothece sp PCC 7424
Acaryochloris marina MBIC11017
Synechococcus sp PCC 7002
Synechococcus sp PCC 7335
Synechocystis sp PCC 6803

GlgB – Branching Enzyme

Waddlia chondrophila WSU 86-1044
Candidatus Protochlamydia amoebophila UWE25
Trichoplax adhaerens
Chlamydophila pneumoniae
Chlamydia trachomatis D-LC
Chlamydophila caviae
Chlamydophila psittaci 6BC
Chlamydophila felis FeC-56
Ostreococcus lucimarinus CCE9901
Micromonas pusilla CCMP1545
Micromonas sp RCC299
Selaginella moellendorffii
Physcomitrella patens subsp patens
Solanum tuberosum
Zea mays
Hordeum vulgare
Arabidopsis lyrata subsp lyrata
Ricinus communis
Phaseolus vulgaris
Pisum sativum

GlgC – ADP-Glucose pyrophosphorylase

Acetabularia acetabulum
Salinibacter ruber M8
Rhodothermus marinus DSM 4252
Desulfobacterium autotrophicum HRM2
Sorangium cellulosum So ce 56
Anaerolinea thermophila UNI-1
Dehalogenimonas lykanthroporepellens BL-DC-9
Spirochaeta smaragdinae DSM 11293
Spirochaeta sp Buddy
Spirochaeta thermophila DSM 6578
Treponema succinifaciens DSM 2489
Treponema denticola F0402
Treponema phagedenis F0421
Treponema vincentii ATCC 35580
Brachyspira pilosicoli 951000
Methylacidiphilum infernorum V4
Chthoniobacter flavus Ellin428
Pedosphaera parvula Ellin514
Pedosphaera parvula Ellin514
Verrucomicrobiae bacterium DG1235
Opitutus terrae PB90-1
Opitutaceae bacterium TAV2
Coraliomargarita akajimensis DSM 45221
uncultured Verrucomicrobiales bacterium HF0130 14P10
uncultured verrucomicrobium HF0130 25O04
Isosphaera pallida ATCC 43644
Planctomyces brasiliensis DSM 5305
Planctomyces limnophilus DSM 3776
Planctomyces maris DSM 8797
Pirellula staleyi DSM 6068
Rhodopirellula baltica
Blastopirellula marina DSM 3645
Candidatus Kuenenia stuttgartiensis
Leptospira biflexa serovar Patoc strain Patoc 1 (Paris)
Parachlamydia acanthamoebae str Halls coccus
Candidatus Protochlamydia amoebophila UWE25
Chlamydia muridarum
Chlamydophila abortus
Chlamydophila pneumoniae
Chryseobacterium gleum ATCC 35910
Flavobacteriaceae bacterium 3519-10
Pedobacter saltans DSM 12145
Sphingobacterium spiritivorum ATCC 33861
Chitinophaga pinensis DSM 2588
Mucilaginibacter paludis DSM 18603
Pedobacter heparinus DSM 2366
Pedobacter sp BAL39
Flavobacterium johnsoniae UW101
Dokdonia donghaensis MED134
Flavobacteria bacterium BBFL7
Psychroflexus torquis ATCC 700755
Kordia algicida OT-1
Polaribacter irgensii 23-P
Polaribacter sp MED152

GlgX – Debranching Enzyme

Candidatus Protochlamydia amoebophila UWE25
Waddlia chondrophila WSU 86-1044
Parachlamydia acanthamoebae str Halls coccus
Chlorella variabilis
Volvox carteri f nagariensis
Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Micromonas sp RCC299
Micromonas pusilla CCMP1545
Ostreococcus lucimarinus CCE9901
Ostreococcus tauri
Triticum aestivum
Zea mays
Oryza sativa Japonica Group
Oryza sativa Japonica Group
Ricinus communis
Populus trichocarpa
Populus trichocarpa
Vitis vinifera
Ipomoea batatas
Solanum tuberosum

Appendix

GlgP – Glycogen Phosphorylase

MalQ – α -1,4 Glucanotransferase

Leptospira biflexa serovar Patoc strain Patoc 1 (Paris)
Candidatus Protochlamydia amoebophila UWE25
Waddlia chondrophila WSU 86-1044
Parachlamydia acanthamoebae str Halls coccus
Waddlia chondrophila WSU 86-1044
Chlamydophila pneumoniae
Chlamydia trachomatis BJali20OT
Chlamydia muridarum
Chlamydophila felis FeC-56
Chlamydophila abortus
Chlamydophila caviae
Treponema succinifaciens DSM 2489
Treponema vincentii ATCC 35580
Treponema phagedenis F0421
Fibrobacter succinogenes subsp succinogenes S85
Trichomonas vaginalis
Trichomonas vaginalis
Trichomonas vaginalis
Trichomonas vaginalis
Trichomonas vaginalis
Trichomonas vaginalis
Polysphondylium pallidum
Dictyostelium purpureum
Dictyostelium discoideum
Entamoeba histolytica
Entamoeba dispar SAW760
Chlorella variabilis
Volvox carteri f nagariensis
Chlamydomonas reinhardtii
Micromonas pusilla CCMP1545
Micromonas sp RCC299
Ostreococcus tauri
Ostreococcus lucimarinus CCE9901
Selaginella moellendorffii
Selaginella moellendorffii
Physcomitrella patens subsp patens
Oryza sativa Japonica Group
Sorghum bicolor
Hordeum vulgare subsp vulgare
Annona cherimola
Arabidopsis thaliana
Solanum tuberosum
Vitis vinifera
Ricinus communis
Populus trichocarpa
Populus trichocarpa
Paludibacter propionicigenes WB4
Flavobacteriaceae bacterium 3519-10
Chryseobacterium gleum ATCC 35910
Bacteroidetes oral taxon 274 str F0058
Lactococcus lactis subsp lactis CV56
Lactococcus lactis subsp cremoris SK11
Paludibacter propionicigenes WB4
Alistipes sp HGB5
Alistipes shahii WAL 8301
Bacteroides sp 20 3
Parabacteroides johnsonii DSM 18315
Candidatus Azobacteroides pseudotrichonymphae geno
Porphyromonas gingivalis ATCC 33277
Prevotella tannerae ATCC 51259
Prevotella ruminicola 23
Bacteroides sp 3 1 33FAA
Succinimonas hippei YIT 12066
Bacteroides plebeius DSM 17135
Bacteroides salanitronis DSM 18170
Bacteroides salanitronis DSM 18170
Bacteroides coprocola DSM 17136
Bacteroides sp 1 1 14
Bacteroides caccae ATCC 43185
Bacteroides ovatus SD CMC 3f
Bacteroides finegoldii DSM 17565
Bacteroides fragilis 638R
Bacteroides intestinalis DSM 17393
Bacteroides stercoris ATCC 43183
Bacteroides helcogenes P 36-108
Bacteroides uniformis ATCC 8492
Prevotella ruminicola 23
Prevotella oralis ATCC 33269
Prevotella salivae DSM 15606
Prevotella oris C735
Prevotella buccae ATCC 33574
Prevotella copri DSM 18205
Prevotella bryantii B14
Prevotella bergensis DSM 17361
Prevotella sp oral taxon 472 str F0295
Prevotella sp oral taxon 299 str F0039
Prevotella buccalis ATCC 35310
Prevotella marshii DSM 16973
Prevotella veroralis F0319
Prevotella melaninogenica D18
Prevotella denticola F0289
Prevotella disiens FB035-09AN
Prevotella bivia JCVIHMP010
Prevotella amnii CRIS 21A-A

Zusammenfassung

Chlamydiae sind obligat intrazelluläre Parasiten, welche ein breites Spektrum von Organismen wie Tiere, Insekten oder Amöben infizieren. Diese Gram-negativen Bakterien stellen eine weltweit vermeidbare Hauptursache für Blindheit und Unfruchtbarkeit in Menschen dar. Infizieren *Chlamydiae* eine eukaryotische Wirtszelle, so bilden sie ein spezielles Kompartiment aus, die Inklusion, in der sie sich vermehren und von der infektiösen, metabolisch inaktiven Form in die nicht-infektiöse, aber metabolisch aktive Form konvertieren. Beide Formen besitzen ein Typ III Sekretionssystem, mit dem sie potentiell toxische Effektorproteine in die Wirtszelle translozieren.

Aufgrund des intrazellulären Lebensstils von *Chlamydiae* konnten bis heute noch keine Strategien entwickelt werden, sie genetisch zu manipulieren. Dies wirkt sich erschwerend auf die Identifikation neuer Effektorproteine aus. Eine zusätzliche Komplikation wird dadurch hervorgerufen, dass die molekulare Erkennung von Effektoren durch die Typ III Sekretionsmaschinerie noch ungeklärt ist. Dennoch wird die Hypothese, dass sich dieses Erkennungssignal in den ersten 20 Aminosäuren des N-Terminus befinde, von einigen Versuchen gestützt.

Um neue Typ III sezernierte Proteine zu identifizieren, bedienten wir uns des schon veröffentlichten Versuchsansatzes, bei dem der Sekretionstest von *Chlamydiae*-Effektoren in einem heterologen Typ III Sekretionssystem von *Shigella flexneri* durchgeführt wurde. Hierfür erstellten wir Chimären, welche aus den ersten 20 Aminosäuren des N-Terminus des potentiellen Effektors und einem Reporter-Protein, der Calmodulin-abhängigen Adenylatcyclase (Cya) von *Bordetella pertussis*, bestanden. Diese Chimären exprimierten wir in unterschiedlichen *Shigella flexneri*-Stämmen. Durch den Gebrauch eines Antikörpers gegen Cya konnten wir die Chimären lokalisieren und ihre Sekretionseigenschaften bestimmen.

Um eine Liste von zu testenden Kandidaten zu erstellen machten wir uns das Wissen zunutze, dass viele Effektoren Signaturen von Proteindomänen, die typischerweise in Eukaryoten gefunden werden, besitzen. Durch die Software „Effective" erhielten wir vorausberechnete Listen von Proteinen aus den Proteomen von *Chlamydia trachomatis*, *Chlamydophila pneumoniae* und *Chlamydophila caviae*, welche eine Anreicherung dieser eukaryotisch-ähnlichen Domänen besitzen. Anhand des oben erwähnten heterologen Sekretionstests konnten wir zeigen, dass 17/25 getesteten Kandidaten tatsächlich durch ein Typ III

Zusammenfassung

Sekretionssystem sezerniert wurden. Die Software „Effective" stellt daher eine sehr nützliche Methode dar. Jedoch kann sie nur als Zusatz verwendet werden, da wir auch andere Proteine positiv getestet haben, die nicht von dem Effective-Programm aufgelistet wurden.

Zusätzlich dazu unterzogen wir eine Auswahl an Proteinen der Umweltchlamydien, welche eine Rolle im Glykogenmetabolismus spielen, ebenfalls dem Sekretionstest. Diese Auswahl wurde hinsichtlich der Funktionen dieser Enzyme getroffen, welche diese bei einer Stabilisierung der Endosymbiosebildung zwischen einem Cyanobiont und einem heterotrophen Organismus erfüllt haben könnten. Wir zeigten, dass fast alle diese Enzyme sezerniert wurden, und dass einige von ihnen am engsten mit Pflanzenhomologen verwandt zu sein scheinen. Dies unterstützt die Hypothese, dass *Chlamydiae* eine unabdingbare Rolle in der frühen Entstehungsgeschichte der heutigen Pflanze einnahm.

Weiters untersuchten wir Homologe dieser Enzyme in *C. trachomatis* und *C. pneumoniae* auf unterschiedliche Sekretionseigenschaften, da es bekannt ist, dass nur *C. trachomatis* Glykogen akkumuliert. Wir zeigten Unterschiede auf und gaben Anregungen für weitere Überlegungen.

CURRICULUM VITAE

Persönliche Daten

Lena Gehre
28 Rue Durantin
75018 Paris
Frankreich

Telefon: 0033658813597
Email: lena-gehre@gmx.at
Geburtstag, -ort: 13.06.85, Landshut/Deutschland

Studium

Dezember 2011:	Beginn der Diplomarbeit am Institut Pasteur, Paris
Wintersemester 2009/10:	Rückkehr an die Universität Wien
Sommersemester 2008:	Beginn eines einjährigen Auslandsaufenthaltes (ERASMUS) in Lyon/Frankreich zum Studium der Neurowissenschaften
Sommersemester 2007:	Wechsel an die Universität Wien (Diplomstudium Genetik und Mikrobiologie)
Wintersemester 2006/07:	Vordiplom (mit Auszeichnung bestanden)
Sommersemester 2005:	Beginn des Studiums Diplombiologie an der Universität Hamburg

Schulausbildung

2004:   Abitur (2,0)

1995-04: Hans-Leinberger-Gymnasium, Landshut

1991-95: Grundschule Berg, Landshut