INSTITUT FÜR MIKROBIOLOGIE
FORSCHUNGSZENTRUM FÜR MILCH UND LEBENSMITTEL WEIHENSTEPHAN
TECHNISCHE UNIVERSITÄT MÜNCHEN

Complete Genome Sequence and Characterization of the Lysis System of the Temperate *Clostridium perfringens* Bacteriophage f3626

MARKUS ZIMMER

Vollständiger Abdruck der von der Fakultät
Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt
der Technischen Universität München zur Erlangung des akademischen Grades
eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Univ.-Prof. Dr. H. Daniel
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. S. Scherer
	2.	Univ.-Prof. Dr. J. Bauer

Die Dissertation wurde am 31.05.2002 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 29.07.2002 angenommen.

Für meine Frau Josi

"The enemy of my enemy is my friend"

- Arabian Proverb -

TABLE OF CONTENT

TABLE OF CONTENT I

LIST OF FIGURES V
LIST OF TABLES V

ABBREVIATIONS VI

SUMMARY 1
ZUSAMMENFASSUNG 3

1 INTRODUCTION 5

1.1 *Clostridium perfringens* 5
1.1.1 Genetic organization of virulence factors from *C. perfringens* 6
1.1.2 Histotoxic diseases in humans caused by *C. perfringens* 7
1.1.3 Food Poisoning 8
1.1.4 Diseases caused by *C. perfringens* in poultry 10

1.2 Bacteriophages 11
1.2.1 General Introduction 11
1.2.2 Life cycles 12
1.2.3 Bacteriophages of *C. perfringens* 13

1.3 Lysis by Bacteriophages 14
1.3.1 Holins 14
1.3.2 Endolysins 15

1.4 Aims of this work 16

2 MATERIALS AND METHODS 17

2.1 Organisms, bacteriophages and plasmids 17

2.2 Media and growth conditions 18

2.3 Isolation of bacteriophages 19

2.4 Propagation and purification of phages 20

2.5 Determination of the lytic range of the isolated bacteriophages 21

2.6 Electron microscopy 21

2.7 DNA manipulation techniques 22

2.7.1 DNA isolation from bacteriophages **22**

2.7.2 DNA isolation from *C. perfringens* **22**

2.7.3 DNA precipitation by Ethanol / Isopropanol **22**

2.7.4 Ligation **23**

2.7.5 Electrotransformation **23**

2.7.6 Separation of DNA fragments by agarose gel electrophoresis (AGE) **24**

2.7.7 Preparation of DNA fragments from agarose gels **26**

2.7.8 Polymerase Chain Reaction (PCR) **26**

2.7.9 Purification of PCR products **26**

2.8 Analysis and amino-terminal sequencing of f3626 structural proteins 26

2.9 Bioinformatics 29

2.10 Cloning, nucleotide sequencing, and identification of the *cos*-site 29

2.11 Identification of *attPP'* and *attBB'* 30

2.12 Cloning and functional analysis of the holin gene *hol3626* 31

2.13 Cloning of endolysin gene *ply3626* 31

2.14 Overexpression and purification of Ply3626 32

2.15 Lytic activity and substrate specificity 33

2.16 Nucleotide sequence accession numbers 34

3 RESULTS 35

3.1 Isolation of *C. perfringens* phages 35

3.2 Nucleotide sequencing and determination of the *cos*-sites 35

3.3 Identification and nucleotide sequence of the attachment sites *attPP'* and *attBB'* 38

3.4 Identification and organization of f3626 ORFs 40

3.5 Structural proteins of f3626 41

3.6 Bioinformatic analysis of f3626 gene products 42

3.6.1 ORF 1 and 2 **45**

3.6.2 ORF 3-5 **45**

3.6.3 ORF 7 and 9 **45**

3.6.4 ORF 12/12.1 **46**

3.6.5 ORF 13 **46**

3.6.6 ORF 15 **46**

3.6.7 ORF 19 and 20 **47**

3.6.8 ORF 22 **49**

3.6.9 ORF 23 **51**

3.6.10 ORF 24-25 **51**

3.6.11 ORF 26 **51**

3.6.12 ORF27 **51**

3.6.13 ORF 32 **52**

3.6.14 ORF 34 **52**

3.6.15 ORF 35 **52**

3.6.16 ORF 39 **52**

3.6.17 ORF 42 **52**

3.6.18 ORF 46 **53**

3.6.19 ORF 50 **53**

3.7 Other features of the f3626 genome 53

3.8 Functional characterization of the f3626 holin 54

3.9 Functional characterization of ply3626 55

3.9.1 Recombinant Ply3626 can only be synthesized in *E. coli* supplemented with rare tRNAs 55

3.9.2 The endolysin lyses *C. perfringens* cell walls 56

4 DISCUSSION 59

4.1 Genome organization of f3626 and relationships to other phages 59

4.2 Cohesive ends of the f3626 genome 60

4.3 Integration site 61

4.4 Gene products of f3626 62

4.5 A Link Between Sporulation and Lysogeny ? 63

4.6 Lysis system of f3626 65

REFERENCES 70

APPENDIX 88

ACKNOWLEDGMENTS VII

LIST OF FIGURES

Fig. 1 Electron micrograph and protein profile of ϕ3626 **36**
Fig. 2 Identification of single stranded cohesive ends in ϕ3626 DNA **37**
Fig. 3 Organization of phage and bacterial attachment sites **39**
Fig. 4 Schematic representation of the ϕ3626 genome **44**
Fig. 5 Putative translational frameshift **46**
Fig. 6 Bioinformatic analysis of Hol3626 **48**
Fig. 7 Alignment of Ply3626 with other amidases **50**
Fig. 8 Lysis profile of Hol3626 **54**
Fig. 9 Synthesis of recombinant endolysin and purification of the protein **56**
Fig. 10 Lysis assays with HPL3626 **58**
Fig. 11 The putative *cosB* sequence **60**

LIST OF TABLES

Tab. 1 Major lethal toxins and types of *C. perfringens* **5**
Tab. 2 Bacterial strains, bacteriophages, and plasmids used **17**
Tab. 3 Size of the fragments of the used molecular markers **25**
Tab. 4 Open reading frames, gene products, functional assignments, and amino acid sequence homologies of ϕ3626 **42**

ABBREVIATIONS

× *g*	relative centrifugal force
aa	amino acid(s)
Arg	arginine
bp	base pairs
C-terminal	carboxy-terminal
cfu	colony forming units
Cps	major capsid protein
CsCl	cesium chlorid
DNA	desoxyribonucleic acid
ds	double stranded
EDTA	Ethylenediaminetetraacetic acid
EN	enteritis necroticans
gp	gene product
HMM	Hidden Markov Modell
Ile	Isoleucine
kb	kilobase pairs
MOI	multiplicity of infection
MSD	membrane-spanning domain
N-terminal	amino-terminal
NE	necrotic enteritis
Ni-NTA	nickel-nitrilo-triacetic acid
nt	nucleotides
ORF	open reading frames
pfu	plaque-forming units
SDS	sodium dodecyl sulfate
TM	transmembrane
Tmp	tapemeasure protein
Tsh	major tail protein
Tris	tris(hydroxymethyl)aminomethane

SUMMARY

Two temperate viruses ϕ3626 and ϕ8533 have been isolated from lysogenic *Clostridium perfringens* strains. Phage ϕ3626 was chosen for a detailed analysis, and was inspected by electron microscopy, protein profiling, and host range determination. For the first time, the nucleotide sequence of a bacteriophage infecting a *Clostridium* species was determined. The virus belongs to the Siphoviridae family of the tailed phages, in the order Caudovirales. Its genome consists of a linear dsDNA molecules of 33,507 nucleotides, with invariable 3'-protruding cohesive ends of 9 residues. Fifty open reading frames were identified, which are organized in 3 major life cycle-specific gene clusters. The genes required for lytic development show an opposite orientation and arrangement as compared to the lysogeny control region. A function could be assigned to 19 gene products, based upon bioinformatic analyses, N-terminal amino acid sequencing, or direct experimental evidence. These include DNA packaging proteins, structural components, the dual lysis system, a putative lysogeny switch, and proteins that are involved in replication, recombination, and modification of phage DNA. The presence of a putative sigma factor related to sporulation-dependent sigma factors and a putative sporulation-dependent transcriptional regulator suggest a possible interaction of ϕ3626 with sporulation events in *C. perfringens*. It has been found that the ϕ3626 attachment site *attPP'* has been found to be located in a non-coding region immediately downstream of *int*. Integration occurs into the bacterial attachment site *attBB'*, located within the 3'-end of a *guaA* homologue. This essential housekeeping gene is functionally independent of the integration status, due to the reconstitution of its terminal codon by the phage sequence.

C. perfringens commonly occurs in food and feed, can produce an enterotoxin frequently implicated in foodborne disease, and has a substantial negative impact on the poultry industry. As a step towards new approaches for its control, the lysis system of *C. perfringens* bacteriophage ϕ3626 has been investigated, whose dual lysis gene cassette consists of a holin and an endolysin. Hol3626 features two membrane spanning domains (MSD) and represents a class

II holin. A positively charged beta-turn between the two MSDs suggests that both the amino- and the carboxy-terminus of Hol3626 might be located outside of the cell membrane, a very unusual holin topology. Holin function was experimentally demonstrated by its ability to complement deletion of the heterologous phage λ holin in λΔSthf. The endolysin gene *ply3626* was cloned in *E. coli*. However, protein synthesis only occurred when the bacteria were supplemented with rare $tRNA^{Arg}$ and $tRNA^{Ile}$ genes. Formation of inclusion bodies could be avoided by drastically lowering the expression level. Amino-terminal modification by a hexa-histidine tag did not affect enzyme activity, and enabled purification by metal-chelate affinity chromatography. Ply3626 features a N-terminal amidase domain and an unique C-terminal portion, which might be responsible for the specific lytic range of the enzyme. All forty-eight tested strains of *C. perfringens* were sensitive to the murein hydrolase, whereas other clostridia and bacteria of other genera were generally not affected. The highly specific activity towards *C. perfringens* might be useful for novel biocontrol measures in food, feed, and complex microbial communities.

ZUSAMMENFASSUNG

Aus lysogenen *C. perfringens* Stämmen konnten zwei temperente Phagen, ϕ3626 and ϕ8533, isoliert werden. Eine detaillierte Analyse einschließlich Elektronenmikroskopie, Bestimmung der elektrophoretischen Mobilität der strukturellen Proteine, sowie eine Bestimmung des Wirts-Spektrum wurde mit dem Phagen ϕ3626 durchgeführt. Erstmalig wurde die vollständige Nukleotid Sequenz eines Bakteriophagen bestimmt, welcher Organismen der Gattung *Clostridium* infizieren kann. Der Phage gehört in die Familie der Siphoviridae, innerhalb der Ordnung der Caudovirales (geschwänzte Phagen). Das Genom ist ein lineares dsDNS Molekül mit einer Größe von 33507 bp, welches einzelsträngige, 3'-überstehende, kohäsive Enden von 9 Nukleotiden besitzt. Fünfzig offene Leseraster (ORFs) konnten im ϕ3626 Genom identifiziert werden, deren Anordnung dieses in drei wesentliche Lebenszyklus-spezifische Genbereiche aufteilt. Der Genbereich verantwortlich für die Aufrechterhaltung des lysogenen Lebenszyklus ist im Vergleich zu den übrigen Genbereichen ("frühe" und "späte" Gene des lytischen Zyklus) gegenläufig organisiert. Aufgrund bioinformatischer Analysen, N-terminaler Aminosäure Sequenzierung und funktioneller Charakterisierungen konnten über 19 ORFs funktionelle Aussagen gemacht werden. So konnten folgende Proteine identifiziert werden: Proteine verantwortlich für die DNA Verpackung, strukturelle Proteine, die Proteine des Dualen-Lyse Systems, ein potentieller Lysogenie-Schalter und Proteine, die in der Replikation, Rekombination und Modifikation der Phagen DNA involviert sind. Die Gegenwart eines möglichen Sigma-Faktors, der Ähnlichkeiten aufweist zu Sporulations-abhängigen Sigma-Faktoren, und die eines weiteren Sporulations-abhängigen Transkriptionsfaktors weisen darauf hin, dass ϕ3626 möglicherweise einen Einfluss auf Sporulations-spezifische Prozesse in *C. perfringens* hat. Die Integrationsstelle *attPP'* ist in einem nicht-kodierenden Bereich vom ϕ3626 Genom lokalisiert, direkt in Leserichtung und unterhalb von *int*. Eine Integration findet an der bakteriellen *attBB'* Stelle statt, welche im 3' Ende eines *guaA*-ähnlichen Gens zu finden ist. Dieses essentielle „Haushaltungs"-Gen ist in seiner Funktion unabhängig vom Lysogeniestatus, da durch eine Integration das ursprüngliche Stopp-Codon an gleicher Stelle vom Phagen bereitgestellt wird.

C. perfringens kann häufig aus Futter- und Lebensmitteln isoliert werden und einige Stämme sind in der Lage ein Enterotoxin produzieren, welches ursächlich

ist für die vom Mikroorganismus hervorgerufenen Fälle von Lebensmittelvergiftungen. Das Bakterium wird aber auch mit Erkrankungen in Verbindung gebracht, welche zu beachtlichen wirtschaftlichen Einbussen in der Geflügel-Industrie führen. Daher wurde für die Kontrolle dieses Bakteriums das lytische System des temperenten Bakteriophagen ϕ3626 untersucht. Das zunächst durch Bioinformatik identifizierte Lyse Module codiert ein Holin und ein Endolysin. Hol3626 besitzt zwei transmembrane Bereiche (MSD) und wird folglich den Klasse II Holinen zugeordnet. Jedoch weist der positiv geladene Bereich zwischen den beiden MSD darauf hin, das sowohl das Amino- als auch das Carboxy-Ende des Hol3626 möglicherweise außerhalb der Zellmembran lokalisiert sind, was eine sehr außergewöhnlichen Topologie darstellen würde. Die Funktion des Holins konnte experimentell gezeigt werden durch seine Fähigkeit, das Fehlen des heterologen λ Holins in Testsystem $\lambda\Delta$Sthf zu komplementieren. Das Gen *ply3626,* welches das Endolysin kodiert, wurde zunächst in einen gängigen Stamm von *Escherichia coli* kloniert und exprimiert. Jedoch war eine Protein-Synthese nur möglich, nachdem das Bakterium mit den im Organismus selten vorkommenden $tRNA^{Arg}$ und $tRNA^{Ile}$ Genen sublementiert wurde. Die hierbei nach Induktion zunächst beobachtete Bildung von Einschlusskörpern konnte durch eine starke Absenkung der Expressionsrate vermieden werden. Eine Amino-terminale Veränderung des Enzymes durch das Hinzufügen eines Hexa-Histidine Markers (HPL3626) hatte keinen negativen Effekt auf die enzymatische Aktivität und erlaubte eine Aufreinigung mittels Nickel-Chelat Affinitätschromatographie. Das Endolysin Ply3626 besteht aus einer N-terminalen Amidase Domäne und einem C-terminalen Bereich, welcher möglicherweise für die spezifische Erkennung der *C. perfringens* Zellwand verantwortlich ist. Lebende Zellen von 48 getesteten *C. perfringens* Stämme waren sensitiv gegenüber der Murein Hydrolase, hingegen zeigten andere Clostridien Arten generell keine Sensitivität, und auch andere Bakterien waren nicht sensitiv. Diese hohe Spezifität von Ply3626 gegenüber *C. perfringens* ist von hohem Wert für nun mögliche Anwendungen, zum Beispiel die Entwicklung neuer biologischer Maßnahmen zur Kontrolle der Entwicklung von *C. perfringens* in Lebensmitteln, Futtermitteln oder auch komplexen mikrobiellen Lebensgemeinschaften.

1 INTRODUCTION

1.1 Clostridium perfringens

Clostridium perfringens is an anaerobic, gram-positive, spore-forming, non-motile rod which can be isolated from the environment, and is frequently found in the intestine of humans and domestic and feral animals (SONGER 1996; SONGER 1997; STEVENS 1997; MCCLANE et al. 2000; MCCLANE 2001). Spores can persist in soil, sediments, and areas subject to human or animal faecal pollution (10^4 -10^6 cfu/g are usually found in human feces) (MCCLANE 2001). *C. perfringens* type A can be isolated from virtually all examined soil samples (10^3-10^4 cfu/g) (LABBÉ 1989). Therefore, it is not surprising that *C. perfringens* can be found on many raw and processed foods. About 50 % of raw or frozen meat and poultry contain *C. perfringens* (LABBÉ 1989), making these meat products common vehicles of *C. perfringens* food poisoning.

Tab. 1 Major lethal toxins and types of *C. perfringens*

	Toxins produced			
Type	Alpha	Beta	Epsilon	Iota
A	+	-	-	-
B	+	+	+	-
C	+	+	-	-
D	+	-	+	-
E	+	-	-	+

+ : present; - : not produced

The species is divided into five types, A to E, based on the production of four major toxins (Tab. 1) (SMITH 1991). The **α-toxin** is a phospholipase C (PLC) and is produced by all types of *C. perfringens*. α-Toxin is hemolytic, destroys platelets and leukocytes, and increases capillary permeability, effects which are likely to be related to its ability to cleave sphingomyelin and the phosphoglycerides of cholin, ethonolamin and serine present in eukaryotic cell membranes (BRYANT and

STEVENS 1997). The b**-toxin** is produced by *C. perfringens* types B and C. It is recognized as the major virulence determinant in enterotoxaemia and necrotic enteritis in various animals species. The β-toxin is a very potent toxin. The LD_{50} of purified β-toxin in mice has been reported to be between 0.31 and 12.5 µg kg^{-1} by the intravenous route and between 4.5 and 93.5 µg kg^{-1} when administered intraperitoneally. Purified β-toxin has shown to cause hemorrhagic necrosis and severe destruction of intestinal villi and mucosae in guinea-pig intestinal loops (LEARY and TITBALL 1997). The e**-toxin** is the most potent clostridial toxin after botulinum und tetanus neurotoxins. The toxin has an LD_{50} in mice of 78 ng kg^{-1} when administered intravenously. The toxin is produced by *C. perfringens* types B and D. This toxin is the main virulence factor of *C. perfringens* type D, which is responsible for enterotoxemia in sheep, goat, and more rarely in cattle. In experimental animals, epsilon toxin elevates the blood pressure, increases the vascular permeability, and causes edema and congestion in various organs including lungs and kidneys. Necrosis of the kidneys is also observed in lambs that have died from enterotoxemia (PAYNE and OYSTON 1997). The i**-toxin** from *C. perfringens* type E is a binary toxin consisting of two independent proteins, an enzymatic Ia and binding Ib component. Ia catalyses ADP-ribosylation of actin monomers, thus disrupting the actin cytoskeleton (CARMAN et al. 1997).

Besides these major toxins, as many as 17 exotoxins have been described for *C. perfringens*, but their role in pathogeny of the organisms has been shown for only some of them (SONGER 1996).

1.1.1 Genetic organization of virulence factors from *C. perfringens*

Symptoms of a *C. perfringens* infection are caused by extracellular enzymes and toxins produced by the organism (PETIT et al. 1999). The genes for these enzymes can be chromosomal or plasmid-encoded, or, as the enterotoxin *cpe* gene, be located on a transposon (PETIT et al. 1999). It has been reasoned that the similarities of some toxins of *C. perfringens* with toxins found in other organisms is due to horizontal gene transfer based on conjugative plasmids, transposons or bacteriophages (PETIT et al. 1999). In other clostridial species,

toxins are known to be bacteriophage-encoded; prominent examples are the neurotoxins BoNT/C and BoNT/D of *C. botulinum* and the α-toxin from *C. novyi* (HENDERSON et al. 1997; JOHNSON 1997). In contrast to these findings, no toxin gene or other virulence factor has been reported to be encoded by a bacteriophage in *C. perfringens* (ROOD and COLE 1991).

1.1.2 Histotoxic diseases in humans caused by *C. perfringens*

1.1.2.1 Clostridial wound infection

Three types of clostridial wound infections are recognized (STEVENS 1997): **(i)** simple contamination of wounds with the organisms; **(ii)** clostridial (anaerobic) cellulitis, an infection localized in the skin and soft tissue. Some gas is produced by the growth of the clostridia in a wound, but invasion into healthy tissue and bacteremia does not occur; **(iii)** traumatic gas gangrene (clostridial myonecrosis, "Gasbrand"), the most common type of clostridal wound infection, occurs through direct inoculation of a contaminated, ischemic wound. Also, it can occasionally occur after a surgical procedure. Rapid onset of myonecrosis, gas production, and sepsis are the hallmarks of this disease. The toxins produced are responsible for decomposition of muscular tissue. An anaerobic environment is created, which is conducive to further growth of the bacteria. *C. perfringens* is found in 80 % of traumatic gas gangrene cases. The remaining cases are caused by other clostridial species, like *C. septicum, C. novyi, C. histolyticum, C. bifermentas, C. tertium* and *C. fallax* (STEVENS 1997).

Gas gangrene in humans is usually associated with *C. perfringens* type A and is mediated primarily by the α-toxin, and secondarily by the ?-toxin (perfringolysin O, PFO) (ROOD 1998; PETIT et al. 1999).

1.1.2.2 Enteritis Necroticans (EN)

A segmental necrotizing enteritis of the small intestine, also called "Darmbrand", was first described in 1949 in an epidemic in Northern Germany after World War II (KREFT et al. 2000). In the highlands of Papua New Guinea, EN was an endemic disease often called pig-bel, because of the increase of

incidences of EN primarily in children after ritual pork feasts (JOHNSON and GERDING 1997). The disease also occurs sporadically in Africa, South East Asia, and the United States (STEVENS 1997). Predisposing factors are helminthic co-infection (JOHNSON and GERDING 1997) and malnutrition, specifically in patients with diets low in protein and rich in trypsin inhibitor (such as sweet potatoes) (LAWRENCE 1997). This kind of disease is rare in the developed countries, where its occurrence is confined to adults with chronic illnesses like diabetes (PETRILLO et al. 2000; GUI et al. 2002).

The β-toxin produced by *C. perfringens* type C has been considered to be the major virulence factor for the occurrence of this disease in humans (MCCLANE 2001). A formation of holes in lipid bilayers has been shown to be caused by β-toxin and it has been reasoned that the lethal action of the β-toxin is based on the formation of pores in susceptible cells (SHATURSKY et al. 2000).

1.1.3 Food Poisoning

The *C. perfringens* type A-associated diarrhea is the most common disease caused by this organism. The surveillance reports of the Center of Disease Control ranked *C. perfringens* as one of top five common causes of food-borne disease in the US (BEAN et al. 1996; OLSEN et al. 2000).

Highly contaminated food (>10^6-10^7 *C. perfringens* per g food) is the basis of this infection. While sporulating, the ingested clostridia are able to produce an enterotoxin within the intestine (JOHNSON and GERDING 1997; MCCLANE 2001). Most frequently meat products, such as beef, turkey, chicken, or pork are the sources associated with *C. perfringens* food poisoning, but cases connected to seafood, dairy products, fruits and vegetables have also been described (JOHNSON and GERDING 1997). The diarrhea is usually self-limiting, and additional symptoms are abdominal pain, nausea, vomiting and fever.

The enterotoxin (Cpe) is the causative agent responsible for the symptoms of *C. perfringens* food poisoning. Unlike many other enterotoxins, Cpe is not an exotoxin that is produced by the organism and secreted. Its production has been found to be linked to the sporulation of the organism (DUNCAN et al. 1972; LABBÉ

and DUNCAN 1974). It has been shown that only sporulating cells and not vegetative cells are able to produce *cpe* mRNA (CZECZULIN et al. 1996). The enterotoxin is most probably released by the destruction of the mother cell at the final step of sporulation (MCCLANE et al. 2000). In the intestine, it binds to the epithelial cells where it induces morphological damage, which eventually is responsible for intestinal fluid loss, causing the diarrhea (MCCLANE 2001).

The regulation of the expression of the enterotoxin is based on a transcriptional level. Sporulation-dependent promoters upstream of the enterotoxin gene have been identified (ZHAO and MELVILLE 1998). It has been reasoned that, due to similarities of these promoters to the $\sigma^{E/K}$ dependent promoter of *B. subtilis*, the expression of *cpe* might depend on alternative sigma factors that are involved in the late stages of sporulation. This would explain the correlation of sporulation and enterotoxin formation found in the mother cell compartment of *C. perfringens* (ZHAO and MELVILLE 1998).

Sporulation of bacteria has been studied in detail in *Bacillus subtilis* (AGUILAR et al. 2001). The recently completed sequencing of *C. acetobutylicum* allowed the comparison of genes involved in sporulation. Many of the respective genes found in *B. subtilis* are absent in *C. acetobutylicum*, which suggests major differences in the sporulation process between these organisms. However, the repertoire of transcriptional regulators, in particular the sigma factors, are very similar (NÖLLING et al. 2001).

The enterotoxin gene *cpe* seems to have a variable localization; it can be found chromosomal, plasmid-borne, or is located on a transposon (CORNILLOT et al. 1995; PETIT et al. 1999). Interestingly, there seems to be a correlation of the localization of the genes and the diseases caused by CPE. Most food poisoning cases are caused by *C. perfringens* carrying *cpe* on the chromosome, whereas most isolates from non-food-poisoning disease (such as antibiotic-associated diarrhea), and veterinary disease harbor a plasmid-borne *cpe* gene (CORNILLOT et al. 1995; COLLIE et al. 1998; COLLIE and MCCLANE 1998).

1.1.4 Diseases caused by *C. perfringens* in poultry

C. perfringens is also responsible for several, mostly enteric, diseases in diverse animals (SONGER 1996; SONGER 1997). Enterotoxemic diseases associated to type A strains have been reported for lambs, goats, calves, horses, pigs, and dogs (SONGER 1996). Since necrotic enteritis (NE) was first reported in 1961 (PARISH 1961) it has to be regarded as an important, world wide disease of domestic chicken (NAIRN and BAMFORD 1967; BAINS 1968; KÖHLER ET AL. 1974A; KÖHLER ET AL. 1974B). Its occurrence has also been described also for other birds, such as turkeys or wild geese (WOBESER and RAINNIE 1987; GAZDZINSKI and JULIAN 1992). Predisposing factors for NE are intestinal mucosal damage caused by high-fiber litter (AL-SHEIKHLY and TRUSCOTT 1977), or coccidiosis (co-infection with protozoa like *Eimeria* spec.) (AL-SHEIKHLY and AL-SAIEG 1980).

The cause of NE is *C. perfringens* type A and C (FICKEN and WAGES 1997). Besides NE, *C. perfringens* is also able to cause a mild form of necrotic enteritis (KALDHUSDAL and HOFSHAGEN 1992). Another subclinical condition associated with *C. perfringens* infection has been described: hepatitis or cholangiohepatitis which is associated with liver lesions (LOVLAND and KALDHUSDAL 1999; LOVLAND and KALDHUSDAL 2001). Both conditions are thought be responsible for an impaired production performance (LOVLAND and KALDHUSDAL 2001), associated with an increased feed conversion ratio and retarded growth rate.

Antibiotics, to which in some cases development of resistance of *C. perfringens* strains has been described (DEVRIESE et al. 1993), and are commonly prophylactically used in poultry feed for the control of NE. It has been reasoned that use of antibiotics as growth promoters does increase development of cross-resistance of pathogenic bacteria against therapeutically used antibiotics (WITTE 1998; WEGENER et al. 1999). According to the World Health Organization, the use of any antimicrobial agent for growth promotion in animals should be terminated if it is: used in human therapeutics; or known to select for cross-resistance to antimicrobials used in human medicine (WHO 1997).

Therefore, there is a need for the development of alternatives to antibiotics used in feed, in order to control, for example, the development of NE in poultry. Specifically acting murein hydrolases (see 1.3.2) from bacteriophages infecting *C. perfringens* might be such an alternative.

1.2 Bacteriophages

1.2.1 General Introduction

Bacteriophages (or phages) are viruses that are able to infect bacteria and destroy them by lysis, or hydrolysis of the cell. They can be considered as obligate intracellular parasites, due to the fact that they solely rely on the metabolism of their host to proliferate (KLAUS et al. 1992). It has been assumed that per prokaryotic cell a 10-fold number of bacteriophages can be found, making the phages the most abundant "organism" on earth (BRÜSSOW and HENDRIX 2002).

Bacteriophages have been discovered independently by the microbiologists F. Twort (TWORT 1915) and Félix d'Hérelle (D'HÉRELLE 1917). Phages have frequently been used in the study of bacterial genetics and cellular control mechanisms, largely because their bacterial hosts can be easily grown and infected with phage in the laboratory. Phages were also used in an attempt to destroy bacteria causing diseases, but this approach was largely abandoned in the 1940s when antibacterial drugs became available. The possibility of "phage therapy" has recently attracted new interest among medical researchers, owing to the increasing threat posed by drug-resistant bacteria.

The primary criteria for a classification of bacteriophages into families are (i) the type of their genome (whether it is build up of RNA or DNA, and whether the nucleic acid molecule is double-stranded or single-stranded); and (ii), their morphology (ACKERMANN and DUBOW 1987). Of the bacterial viruses 96 % are tailed phages (Caudovirales) (BRÜSSOW and HENDRIX 2002), which consist of a head with an inner core of DNA - the phage genome - and a tail, whose morphology is a criterion for classification into Myoviridae (long contractile tail),

Siphoviridae (long non contractile flexible tail), and Podoviridiae (short non contractile tail).

The host range of bacteriophages is usually very limited, they frequently infect only one or a few related species of bacteria, sometimes they are even strain-specific (ACKERMANN and DUBOW 1987).

1.2.2 Life cycles

In general there are two major types of bacteriophages, virulent and temperate phages. Virulent phages are only able to propagate in a lytic life cycle, whereas temperate phages are additionally able to switch into a lysogenic life cycle.

1.2.2.1 Lytic life cycle of phages

The life cycle can be divided into four main steps: **(i) Absorption:** Phages adsorb on a specific receptor on the cell surface. **(ii) Infecton:** After a local disruption of the cell wall the phage injects its genome into the bacterium. **(iii) Replication and Maturation:** The phage uses the bacterium's chemical energy and biosynthetic machinery to produce viral proteins, as well as more phage nucleic acid. These produced molecules assemble within the bacterial host into new phage particles. **(iv) Release:** The progeny virions are released after maturation either by lysis, extrusion, or budding (ACKERMANN and DUBOW 1987). Lysis can be readily observed in bacterial cultures growing confluently on solid medium, where groups of lysed cells appear as clear areas, or plaques (ACKERMANN and DUBOW 1987).

1.2.2.2 Lysogenic life cycle of temperate phages

In temperate phages, the infected host is not used for the immediate proliferation of the phage, but the phage genome becomes usually integrated as a prophage into the host chromosome. In this state, known as lysogeny, most of the information contained in the viral nucleic acid is not expressed, only the proteins responsible for the maintenance of the lysogeny are produced. In some rather rare cases, the phage genome can become transiently located extrachromosomally as

an episome (pseudo-lysogeny), a status which seems to be dependent on environmental conditions (RIPP and MILLER 1998).

The establishment of the lysogenic life style and the latent ability of the prophage to change into the lytic life cycle resulting in the production of virus particles and lysis of the host is regulated in a complex way, best understood in phage λ. It depends on a concerted action of the repressors (lysogenic switch) expressed from the phage genome (PTASHNE 1992). In a growing culture the induction of the lytic life cycle is usually strong suppressed. Certain environmental conditions, which have a negative effect on the host, (such as stress, nutrient depletion, inappropriate growth temperatures or antimetabolites) might induce the lytic life cycle. This also occurs spontaneously at low rates (10^{-2} - 10^{-5} per bacterium per generation), but can also be induced with physical or chemical DNA damaging agents (e. g. UV light, mitomycin C) (ACKERMANN and DUBOW 1987).

1.2.3 Bacteriophages of *C. perfringens*

Bacteriophages of the genus *C. perfringens* have been described since the early 1940s, and both, temperate and virulent phages have been described in the following years (MAHONY 1979). However, besides the description of a preliminary typing scheme (YAN 1989) and a preliminary mapping of the integration sites of two temperate phages (CANARD and COLE 1990), no work has been performed on bacteriophages of *C. perfringens* for the last 20 years (H. W. ACKERMANN, personal communication). No sequences of bacteriophages infecting *C. perfringens* are available. Up to now not a single genome of phages infecting any species of the genus *Clostridium* was sequenced and analyzed.

In 1959, Smith screened 152 strains of *C. perfringens* for temperate phages and found that 32 were lysogenic. Additionally, the isolation of virulent phages was described by the author (SMITH 1959). The author also investigated the usefulness of these phages to establish a typing scheme, but found that only 61.2 % of the strains were susceptible to lysis (SMITH 1959). A very narrow host range, which seems to be common in clostridial phages (OGATA and HONGO 1979) was an obvious obstacle in the development of phage typing schemes.

The only example of phage conversion reported so far for *C. perfringens* were experiments performed by Stewart and Johnson. The researches tested a lysogenic strain and its cured and its re-lysogenized derivatives on their ability to sporulate. The results indicated that the lysogens produced spores with a higher efficiency and that the spores were more resistant to heat in comparison to the phage-free derivatives (STEWART and JOHNSON 1977).

1.3 Lysis by Bacteriophages

Generally, there are two different ways how bacteriophages can lyse their host cells to release their progeny at the end of their lytic life cycle. For both, the cell wall is the main target, but the disruption of the cell wall is achieved in different ways. The first way is a blockage of the enzymes involved in peptidoglycan synthesis, exploiting the bacteria's need to continuously synthesize these molecules (HATFULL 2001). This was recently demonstrated for phages Qβ and ϕX174 (BERNHARDT et al. 2001a; BERNHARDT et al. 2001b). The second, more common way, is the use of a cell wall hydrolyzing system, which usually consists of a holin and an endolysin (YOUNG 1992; YOUNG et al. 2000).

1.3.1 Holins

Holins are small hydrophobic proteins that are thought to forms stable and non-specific lesions in the cell membrane and thereby admit the endolysin to pass the membrane and access the cell wall target (YOUNG and BLÄSI 1995). The timing of lysis, which is critical for the viral reproduction, is somehow "programmed" into the structure of the holin and thus controls the length of the vegetative cycle (WANG et al. 2000; YOUNG et al. 2000). Holins are one of the most diverse functional groups, represented by more than 100 known or putative holin sequences, which form more than 30 ortholog groups (WANG et al. 2000). Many holins are organized in a dual gene cluster together with the endolysin (WANG et al. 2000). For the Siphoviridae, it appears as if this cluster is most often located at the 3'-end of the late gene region, next to the lysogeny module (LUCCHINI et al. 1999b).

All holins commonly possess at least one transmembrane domain (TM domains, or membrane spanning domain (MSD)) (YOUNG 1992). The number of TM domains is a criterion for the basic classification of the holins into class I and class II holins. Class I holins do have three TM domains and class II holins do have only two TM domains. Some other holins cannot be grouped into theses classes, since they have only one MSD, and likely function in different ways. (WANG et al. 2000).

A recently developed genetic system can be used to assess the functional properties of various holins of all classes in *E. coli* λΔSthf::*hol* lysogens, where a chromosomally deleted S holin is replaced with a desired holin. Thermal induction of a lysogen allows an observation of a synchronized lytic cycle, and permits reasonably precise assessment of the course and timing of lysis (VUKOV et al. 2000).

1.3.2 Endolysins

Endolysins are mureolytic enzymes. They access the cell wall after the formation of lesions by holins, and hydrolyze the peptidoglycan (YOUNG 1992). With respect to their substrate specificities, they can be grouped into several classes: amidases, glycosidases, muramidases and endopeptidases (YOUNG 1992). A novel activity, L-alanoyl-D-glutamate peptidases, has been described for *Listeria* phages (LOESSNER et al. 1995b).

For endolysins, an astonishing substrate specific lysis actitvity has been described, which is thought to be based upon a modular structure found in many phages lysins from different bacteria (LOESSNER et al. 1995b; LOESSNER et al. 1997; LOEFFLER et al. 2001; NELSON et al. 2001). The enzymatically active polypeptide domain is linked to a corresponding cell wall binding domain, which targets the protein to its site of action (Garcia et al. 1990; Loessner et al. 2002). Consequently, the recognition specificity of the hydrolases is largely conferred by these cell wall binding domains (Baba and Schneewind 1996; Loessner et al. 2002).

Besides the described lysis from within, endolysins are also able to quickly lyse gram-positive phage host cells when applied exogenously (LOESSNER et al. 1995a). Recently, streptococcal phage lysins have shown their potential use in control of pathogenic bacteria *in situ* (LOEFFLER et al. 2001; NELSON et al. 2001). Here, mice were treated nasally and pharyngeally with endolysin obtained from streptococcal phages, which led to a reduction, or even eradication, of previously inoculated streptococci (LOEFFLER et al. 2001; NELSON et al. 2001). In another approach *Lactococcus lactis* has been used to produce and secret a *Listeria* phage amidase (Ply511), which remained fully active and killed *L. monocytogenes* cells in the medium (GAENG et al. 2000).

1.4 Aims of this work

One major aim was to collect essential information on phages infecting *C. perfringens*, with respect to basic morphological characteristics, nucleotide sequence, and potential effect on lysogenized host cells. For this purpose strains of *C. perfringens* were screened for lysogeny and bacteriophages could be obtained from these strains. Further steps were the complete nucleotide sequencing of a bacteriophage genome and its detailed analysis by bioinformatics and experimental studies.

In the prospect of a potential application the lysis system of the bacteriophage was studied in detail. The respective genes were identified and cloned, and the protein tested for activity and its substrate specificity. Here, the goal was to evaluate the potential use of a recombinant endolysin as a novel antimicrobial substance to control *C. perfringens*.

2 MATERIALS AND METHODS

2.1 Organisms, bacteriophages and plasmids

All bacterial strains, phages, and plasmids used are listed in Table 2. Fifty-one *Clostridium perfringens* strains have been obtained from various sources (see Tab. 2). *Escherichia coli* strain DH5αMCR (Invitrogen) in combination with plasmid pBluescript II SK- (Stratagene) was used for cloning. The endolysin gene *ply3626* was cloned into pQE-30 and pSP72, and overexpressed in *E. coli* strains JM109 or JM109 (DE3), respectively. *E. coli* LE392 was used for plating of λ phage and for the generation of lysogens

Tab. 2 Bacterial strains, bacteriophages, and plasmids used

Strain/phage/plasmid	Genotype or relevant properties	Source/reference
Bacteria		
E. coli DH5αMCR	F⁻ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR recA*1 *endA*1 *supE*44 λ⁻ *thi*-1 *gyrA*96 *relA*1	Invitrogen
E. coli JM109(DE3)	JM109 with λDE3	Promega
E. coli JM109	*rec*A1 *end*A1 *gyr*A96 *thi hsd*R17 (r_K- m_K+) *sup*E44 *rel*A1 Δ(*lac-pro*AB) [F' *tra*D36 *pro*AB *lac*I^{q}ZΔM15]	Laboratory stock
E. coli LE392	*hsd*R514(r_K-, m_K+) *sup*E44 *sup*F5*8* *lac*Y1 *gal*K2 *gal*T22 *met*B1 *trp*R55 (phage host; permissive for λgt11)	Promega
C. perfringens WS 2981 to 3002		Twenty-two isolates from various food sources
C. perfringens WS 3006 to 3008		Three isolates from the caecum/ileum of poultry with necrotic enteritis
C. perfringens WS 2955 to 2963		Nine clinical isolates from humans
C. perfringens ATCC 3626, 3628	Toxin type B and C type strains	ATCC, USA
C. perfringens NCTC strains 528, 3110, 6719, 6785, 8081, 8346, 8533, 10240, 10612, 10614, 10719, 11144	Strains with various toxin types: A, B, C, E; NCTC3110: propagation strain for ϕ3626, used as standard in the lysis assay	NCTC, London
C. perfringens DSM strains 756^T, 798, 2943		DSMZ, Heidelberg, Germany

Tab. 2 continued; Bacterial strains, bacteriophages, and plasmids used

Strain/phage/plasmid	Genotype or relevant properties	Source/reference
Phages		
ϕ3626	From *C. perfringens* ATCC3626	This work
ϕ8533	From *C. perfringens* NCTC8533	This work
λΔSthf	λS, single *Eco*RI site, *c*Its857, Amp^r	(Vukov et al. 2000)
λΔSthf::S	native *S* inserted into the *Eco*RI site of λΔSthf	(Vukov et al. 2000)
λΔSthf::*hol3626*	*hol3626* cloned into the *Eco*RI site of λΔSthf	This work
Plasmids		
pBluescript II SK (-)	3.0-kbp cloning vector; Amp^r	Stratagene
pACYC-IRL10	encodes additional copies of ileX, argU and leuW, Cmr	(ZDANOVSKY AND ZDANOVSKAIA 2000)
pQE-30	3.4-kb coning and expression vector; T5 promoter/lac operator element; 5'- 6x His-tag coding sequence; Ampr	Qiagen
pSP72	2.4-kb cloning and expression vector; T7 promoter Ampr	Promega
pHPL3626	Ply3626 cloned into the BamHI and SalI site of pQE-30	This study
pHPL3626ΔC	5' fragment (540 bp) of Ply3626 probably encoding an amidase domain cloned into the BamHI and SalI site of pQE-30	This study
pSPply3626	Ply3626 cloned into the EcoRI site of pSP72	This study

2.2 Media and growth conditions

C. perfringens strains were grown at 37°C in tryptone-yeast extract medium (FACH and GUILLOU 1993) in anaerobic jars (Oxoid) using the Anaerocult A system (Merck). Most of the handling was performed in a flexible vinyl glove-chamber (Coy Laboratories, Michigan), which contained a 95 % N_2 / 5 % H_2 atmosphere. Luria-Bertani (LB) medium was used for the incubation of *E. coli* cells at 37°C. For production of recombinant endolysin, temperature was reduced to 21°C, and to 30°C for propagating lysogens. For the growth of LE392, 0.2 % maltose and 10 mM $MgSO_4$ were added. Plasmid-bearing cells were selected using 100 µg ml^{-1} ampicillin and/or 10 µg ml^{-1} chloramphenicol, when appropriate. Lysogens were selected on media containing 30 µg ml^{-1} ampicillin. Phage stocks were stored at 4°C in SM Buffer.

All media were prepared using demineralized water, all buffers or solutions were prepared using MilliQ water.

TY (per l)

Trypton	30	g
Yeast extract	20	g
HCl cysteine	0.5	g

pH 7.2 ± 0.2

For plates add agarose 1.5 % (w/v) or for top agar 0.4 %.

LB broth (per l)

Trypton	10	g
Yeast extract	5	g
NaCl	8	g

pH 7.4

For plates add agarose 1.5 % (w/v) or for top agar 0.4 %.

SM buffer (per l)

Tris HCl stock solution	50.0	ml
NaCl	5.8	g
$MgSO_4 \times H_2O$	2.0	g

Tris HCl stock solution: 1 M (per l)

Tris	121.1	g

pH 7,5 titrate with HCl

2.3 Isolation of bacteriophages

A total of 51 clostridial strains clostridial strains were screened for lysogeny by UV-irradiation as previously described (LOESSNER and BUSSE 1990). Exponentially growing cells (10 ml) were exposed to UV-light (254 nm, 0.02 J cm^{-2}) for 5 min. After 3 h incubation (37°C) in the dark, cultures were centrifuged (10 min, 8,000 × *g*), and supernatants cleared by filtration (0.2 µm).

Phage activity was tested by the spot-on-the-lawn method, against all available *C. perfringens* strains. 100 µl of exponentially growing *C. perfringens* cultures were spread on a agar plate and allowed to dry for thirty minutes. On the confluently growing bacteria 5-25 µl of phage containing solution were spotted. Lytic activity was observed by the formation of macroplaques after an overnight incubation. To verify whether the macroplaques were formed due to the multiplication of the phages the according solutions were consequently used for the soft agar layer technique (see 2.4).

2.4 Propagation and purification of phages

The soft agar layer technique (ADAMS 1959) was used for the purification and propagation of the phages. Dilutions of the supernatants displaying lytic activity were added to 3.5 ml of molten soft agar inoculated with 0.1 ml log-phase culture of the propagation strain. The mixture was poured on TY plates and incubated overnight. Single plaques were picked and placed into 0.45 ml of TY medium. The phages were allowed to diffuse from the agar into the medium by incubation of 4 h at 4°C. After centrifugation at 8000 $\times g$ for 15 min the supernatant was cleared by sterile filtration (0.22 µm) and used for a several rounds of purification.

Bacteriophage ϕ8533 was propagated by this technique to high titers (10^9 pfu ml^{-1}). This was achieved by using a soft agar layers of confluently lysed plates for the propagation. The complete soft agar layers were transferred into disposable 50 ml centrifuge tubes containing SM buffer. After incubation (4°C, 4 h) the solution was centrifuged (8000 $\times g$; 15 min) and cleared by sterile filtration.

For high titer stocks of ϕ3626 (>10^9 pfu ml^{-1}), liquid cultures were used. Cultures were infected at an OD_{600} of 0.1 using an MOI of 1. Afterwards, growth was monitored photometrically, and following lysis, phages were harvested by centrifugation (10,000 × *g*, 10 min) and sterile-filtration of the culture supernatant.

Purification of viruses from high titer stocks has been described earlier (ZINK and LOESSNER 1992). The phages of the high titer stock solutions were precipitated using the standard protocol for the isolation of lambda bacteriophages

(SAMBROOK et al. 1989) by adding 1 M NaCl and 10 % (w/v) polyethylene glycol 8000. After an incubation for overnight at 4°C, with a slight agitation for the first 10 hours of incubation, the phages were collected by centrifugation (10000 × *g*, 15 min, 4°C) and the pellet was resuspended carefully in 5 ml of SM buffer. These preparations could be used for the electron microscopy (see 2.6).

For further molecular work ϕ3626 was purified using a pre-formed stepped CsCl gradient (densities: 1.7/1.5/1.45/1.4/1.3/1.2 g cm^{-3}) at 85000 × g_{max} for 16 h at 10°C. A blue-greyish band at a buoyant density of approximately 1.45 g cm^{-3} was carefully withdrawn by a sterile syringe (~ 5 ml). The CsCl was removed from the purified bacteriophages by dialysis against SM buffer (2 l) for 2 h and a second dialysis for overnight (4°C) in autoclaved cellulose ester tubings.

2.5 Determination of the lytic range of the isolated bacteriophages

The ability of the two phages to lyse *C. perfringens* strains was tested by the drop-on-the-lawn-technique. Ten µl of the purified phage stocks (10^7 pfu ml^{-1}, see 2.4) were placed on the plates inoculated with *C. perfringens* strains. The lytic activity was observed after overnight incubation.

2.6 Electron microscopy

Phage particles were examined by electron microscopy (EM) as reported before (ZINK and LOESSNER 1992). Briefly, a small drop of the CsCl solution (5 µl) was placed on top of a carbon film fixed on a copper grid (400 mesh) for 1.5 min, to allow phage to attach to the carbon film. Excess solution was removed. The surface of the grid was repeatedly washed with water, and finally negatively stained using 2 % uranyl acetate. Pictures of the virus particles were taken with a transmission electron microscope (Zeiss EM-10A) at an acceleration voltage of 60 kV, with a magnification of 100,000 fold.

2.7 DNA manipulation techniques

2.7.1 DNA isolation from bacteriophages

The DNA of ϕ3626 was obtained using the standard method for the extraction of bacteriophage λ DNA (SAMBROOK et al. 1989) with the exception of doing the extraction with one time phenol, twice with phenol/chloroform/isoamylalcohol (50:48:2, v/v/v) and once with chloroform. No dialysis was performed and the DNA was precipitated with EtOH. Due to a contamination with RNA an additional treatment with RNAse A was necessary and the DNA was finally isopropanol precipitated.

2.7.2 DNA isolation from *C. perfringens*

For the identification of the attachment site it was necessary to isolate and purify the genomic DNA of its lysogenic host *C. perfringens* ATCC3626. This was done according to an alternative protocol for the isolation of DNA from gram positives (FLISS et al. 1991). A 25 ml culture grown overnight at 37°C was harvested by centrifugation at 10000 g_{max} for 10 min. The pellet was washed once with 0.1 × SSC (for 20 × SCC: 3 M NaCl; 0.3 M Na-citrate; pH 7.0) recentrifuged and resuspended in 500 µl TES (30 mM Tris/HCl [pH8]; 50 mM NaCl; 5 mM EDTA [pH8]; 20 % saccharose [w/v]) containing 8.4 µg mutanolysin and 0.6 mg lysozyme to ensure the destruction of the cell wall. After an incubation of 30 min at 37°C 5 µl of RNAse A (~ 80 U) was added. After another incubation for 30 min 4.5 ml of lysis buffer (10 mM Tris/HCl [pH8]; 1 mM EDTA [pH8]; 0.8 % SDS [w/v]) and 300 µg proteinase K was added. The cells lysed within the following incubation at 50°C for 2 h (gently agitated). The DNA was extracted using a phenol, (2 ×) phenol/chloroform/isoamylalcohol (50/48/2, [v/v/v]), chloroform extraction. Finally the DNA was ethanol precipitated.

2.7.3 DNA precipitation by Ethanol / Isopropanol

DNA in a sample was precipitated by the addition of 0.3 M sodium acetate (pH 5.2, 0.1 volume of 3 M) along with 2 to 2.5 volumes of ethanol (-20°C) and incubation for 30 min at 21°C. The sample was then centrifuged at 4°C (maximum

speed) in a microcentrifuge (Eppendorf) and washed once with 70 % ethanol. After a second centrifugation the pellet was left to dry and finally resuspended in MilliQ water or ½ TE. Instead of using 2/2.5 volumes of ethanol 1 volume of isopropanol can also be used for precipitation of DNA.

TE buffer

10 mM	TRIS/HCl
1 mM	EDTA
pH 8.0	

2.7.4 Ligation

Ligations were performed using 1 U T4 DNA ligase (Promega, Madison, USA) and 10x reaction buffer (Promega) for 4 h at 21°C or for 18 h at 4°C.

2.7.5 Electrotransformation

Transformation of *E. coli* was carried out by electroporation according to the protocols by (DOWER ET AL. 1988) using a BioRad Gene Pulser and sterile, individually wrapped 2-mm electroporation cuvettes (EquiBio Ltd., Kent, UK). Selection for resistance to antibiotics in *E. coli* was performed with ampicillin (100 µg ml^{-1}). Colonies harboring recombinant plasmids were selected by blue-white screening according to standard methods (SAMBROOK et al. 1989). Recombinant plasmids were isolated using the Qiaprep Miniprep Kit (Qiagen, Hilden, Germany).

2.7.5.1 Preparation of competent cells

500 ml of LB (*E. coli*) broth were inoculated with 1 ml of an overnight liquid culture of the *E. coli* and incubated at 37°C shaking at 150 rpm. After reaching an optical density of OD_{600} = 0.5-0.6, cells were harvested by centrifugation (8000 × g, 4°C, 10 min) and stored on ice. Cell Pellet was resuspended and washed with: 500 ml 5 % glycerol (4°C), 250 ml 5 % glycerol (4°C), 10 ml 10 % glycerol (4°C). Finally the pellet was resuspended in 1 ml 10% glycerol (4°C) and stored in 40 µl aliquots at -70°C

2.7.5.2 Electroporation

20 µl of the ligation mixture or plasmid was dialyzed for 30 min using dialysis membrane filter (0.025 µm pore size, Millipore) and electoporated into 40 µl of *E. coli* competent cells (see 2.7.5.1) on ice using a Gene-Pulser (BioRad) under the following conditions:

Resistance:	200 Ω
Capacitance:	25 µFD
Volt:	2.5 kV (12.5 kV/cm)
Time constant:	~4.7

Following electroporation, 500 µl LB-medium was added directly to the electroporation cuvettes. For regeneration of the host cells, the mixture was incubated at 37°C for 1 h. 100 µl aliquots (and ten-fold dilutions) were plated on LB plates supplemented with the corresponding antibiotic. Plates were incubated for 18 h at 37°C.

2.7.6 Separation of DNA fragments by agarose gel electrophoresis (AGE)

To prepare a 1.2 % gel, 0.5 g agarose (SeaKem LE, FMC BioProducts) were dissolved in 40 ml 1 x TAE buffer using a microwave. The hot agarose solution was poured into a gel tray with slot-formers. 5 µl of DNA containing samples were mixed with 1 µl 6x gel loading buffer (GLB) Blue/orange 6x loading dye (Promega, Madison, USA) and pipetted into the gel slots. Additionally an appropriate molecular weight marker was loaded (see Tab. 3) After electrophoresis, the gel was stained with ethidiumbromide EtBr (0.5 µg/ml) for 30-45 min and visualized by UV-translumination using the ImageMaster® VDS (Pharmacia, Biotech). Electrophoresis was performed using the following equipment: GNA-100 submarine electrophoresis unit (Pharmacia Biotech); Easy-Cast™ Electrophoresis system, Model #B1A (Owl Scientific, Inc.); Electrophoresis Power Supply EPS 600 (Pharmacia Biotech).

TAE (Tris/acetate/EDTA) electrophoresis buffer 50 x stock solution (1 l):

2 M	Tris-base
1 M	NaAc
50 mM	EDTA
(pH 8.0)	

Gel loading buffer: (GLB 6x)

0.10 %	Xylen-Cyanol FF (XCFF)
0.25 %	Bromophenolblue
0.20 %	SDS
5 mM	EDTA
50% (v/v)	Glycerol
(pH 8.0)	

Tab. 3 Size of the fragments of the used molecular markers

Marker		Fragment size (bp)
DNA Ladder (MBI Fermentas, Vilnius)	100 bp	100, 200, 300, 400, 500, 600, 700, 800, 900, 1031, 1200, 1500, 2000, 3000 bp
DNA Ladder (MBI Fermentas, Vilnius)	1 kb	250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000, 10.000 bp

2.7.7 Preparation of DNA fragments from agarose gels

Gel extraction was performed by using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) according to the instructions of the manufacturer:

2.7.8 Polymerase Chain Reaction (PCR)

Oligonucleotides used for PCR and sequencing reactions were obtained from MWG Biotech (Ebersberg, Germany). PCR amplification was performed with a Techne Progene automated thermocycler with 0.2-ml thin-walled PCR tubes (Advanced Biotechnologies). Reactions were carried out in 50-µl volumes containing 5 µl 10x PCR reaction buffer (containing 2.0 mM $MgCl_2$), 100 pmol of each oligonuleotide primer, 0.2 mM of each deoxynucleotide triphosphate (dATP, dCTP, dGTP, dTTP, Eurogentech) and 1 U of *Taq* DNA polymerase (Qiagen). The standard PCR protocol: After an initial denaturation step (94°C, 5 min) amplification was carried out with 25-35 cycles (denaturation at 94°C for 450s, annealing at 50°C for 60s and extension at 72°C for 30s) followed by a final extension step at 72°C for 10 min. The annealing temperature was adjusted to the melting temperature of the primers and the extention time to the size of the expected fragment.

2.7.9 Purification of PCR products

PCR products were purified using spin columns with the QIAquick-spin PCR Purification Kit (QIAGEN, Hilden, Germany) according to the instructions of the manufacturer.

2.8 Analysis and amino-terminal sequencing of f3626 structural proteins

The isolation and purification of the structural proteins was performed essentially as described earlier (ZINK and LOESSNER 1992; LOESSNER et al. 1994). Virion proteins were separated by Sodium dodecyl sulphate polyacrylamid gel electrophoresis (SDS-PAGE) according to (LAEMMLI 1970) on a discontinous 8-18% gradient gel (Excel Gel, Pharmacia LKB). Equal volumes of the peptide

containing samples and gel loading buffer (2x GLB) were mixed and boiled at 100°C for 10 min. Electrophoresis was performed at 12°C on a horizontal system (2217 Ultrophor Electrofocusing Unit, LKB Broma) at constant voltage (200 V for the first 20 min, 600 V for the remaining 60 min, 50 mA, 35 W) using the 2303 Multidrive XL 3.5 kV power supply (LKB, Broma). As molecular-weight standard SDS-*7L* (Sigma) was used.

Electrophoresis buffer: (10x Tris-Glycine-buffer) (per l)

144.2 g	Glycine
30.3 g	Tris
10.0 g	SDS

SDS-GLB (Gel-loading-buffer) (50 ml)

2,5 ml	Tris-HCl (1.25 M, pH 6,8)
1,0 g	SDS
2,5 ml	2-Mercaptoethanol
5,8 ml	glycerole (87%)
5 mg	bromphenolblue

Gels that were not used for semidry-blotting were dyed for 25-30 min using prewarmed (60°C) coomassie R-350 dye solution. The background dye was removed by washing (2 × 30 min) the gel with prewarmed wash solution and finally the gel was placed for 30 min in a preserve solution. The percentage of the major structural bands were determined by densitometrical scanning.

Coomassie R350 dye solution

coomassie R-350 tablets	4 ×
methanol	25 %
acetic acid	10 %

Wash solution

methanol	25 %
acetic acid	10 %

Preserve solution

methanol	25 %
acetic acid	10 %
glycerol	10 %

For microsequencing proteins were electroblotted on a polyvinylidene difluoride (PVDF) membrane (GelmanSciences) using the 2117-250 Novablot electrophoretic transfer kit (LKB) in a Multiphor II device (Pharmacia) according to (KYHSE-ANDERSEN 1984) at 20°C for 1 h at 0.8 mA cm^{-2}. The blot was build up of the immobilizing membrane and the gel placed between filters soaked with transferbuffers that are in direct contact to the graphite-electrodes, according to the manufacturers manual (Pharmacia).

Anodic buffer I

0.3 M	Tris
20%	methanol

Anodic buffer II

25 mM	Tris
20 %	methanol

Cathodic buffer

40 mM	6-Aminocapronic acid
0.01 %	SDS
20 %	methanol

The membrane was rinsed with methanol once and dyed using 1/2 Coomassie R-350 dye solution Three major bands were excised from the

membrane and the dye was removed by washing the membrane in 45 % methanol and 10 % acetic acid. The first 10 amino acids (aa) each of the individual proteins were determined using an automated sequencer (Applied Biosystems Procise 492-01) by the service department of the Arbeitsgruppe Proteinanalytik (Universität Münster)

2.9 Bioinformatics

The program DNASIS and the Husar Analysis Package (version 4.0; http://genome.dkfz-heidelberg.de) were used for analysis of the nucleotide and amino acid sequences. The BLAST algorithms (ALTSCHUL et al. 1997) were used for similarity searches in the databases available through the NCBI (http://www.ncbi.nlm.nih.gov), or the Husar Analysis Package. COILSCAN (LUPAS et al. 1991) was used to predict probabilities to form a coiled-coil structure. The HMMscan program (Pfam database; release 5; http://pfam.wustl.edu) identifies protein families using the Hidden Markov Model (DURBIN et al. 1998), and TmHMM (v. 2.0) protein analysis uses this model for prediction of transmembrane domains (SONNHAMMER et al. 1998).

2.10 Cloning, nucleotide sequencing, and identification of the *cos*-site

The DNA of ϕ3626 was extracted and purified using standard techniques (see 2.7), and the construction of genomic libraries of ϕ3626 was performed essentially as described previously (LOESSNER et al. 2000). Here, limited digests with *Tsp*509I (New England Biolabs), and complete digests using *Hin*dIII (MBI Fermentas) or *Taq*I (Roche) were performed, fragments of 1-2 kb length ligated into pBluescript, and transformed into *E. coli*. Blue-white screening on Xgal-containing agar plates was used for the identification of insert-bearing clones. Plasmids from small-scale cultures were digested with *Pau*I (MBI), and fifty-eight clones carrying inserts varying in size were identified by agarose gelelectrophoresis. These plasmids were used for sequencing using IRD-800-

labelled primers complementary to sequences flanking the multiple cloning site. Sequencing was performed using a heat-stable polymerase (SequiTherm EXCEL II; Epicentre Technologies) on an automated DNA sequencer (4200 IR^2; LI-OR).

The obtained sequences were edited and aligned using the software DNASIS (version 2.1; Hitachi). Gaps were closed by direct sequencing of ϕ3626 chromosomal DNA, with the aid of specific primers derived from the contigs. Distinct chain termination signals were generated at the ends of the molecule, i.e., the putative single stranded ends (*cos*-sites). The genome sequence was finalized by the determination of the sequence of the *cos*-site overlaps, by PCR amplification of DNA from lysogenic host bacteria (see 2.11), using primers complementary to sequences upstream and downstream of the *cos*-site.

2.11 Identification of *attPP'* and *attBB'*

Purified genomic DNA of ϕ3626 was used as template for the identification of the attachment site using inverse PCR (OCHMAN et al. 1988). *AttPP'* was expected to be located in a non-coding region immediately downstream of *int*. A *Sau*3AI restriction site is present within the *int* gene, and this enzyme (Roche) was used for complete digestion of the bacterial DNA. Fragments were treated with T4 DNA ligase to obtain self-ligated circular molecules. Divergent primers were designed, complementary to an area located within *int*: Att3up (5'-CTCAAATGATAGCAA-CAACAGG-3') and Att3dw (5'-CTTTTACTTTTAGGAGTTTGGG-3'), and used for PCR amplification of ligated fragments. The product were purified and sequenced using the same primers. The obtained sequence contained the *attBP'* site, and the non-prophage part of the sequence displayed 100 % identity over 663 nucleotides to a sequence of the unfinished *C. perfringens* genome available from the Institute for Genome Research (TIGR, http://www.tigr.org). Additional sequence was obtained, in order to design a primer attB1 (5'-GACAATCATATTAAAATGACT-GCC-3') that, in combination with primers complementary to the prophage DNA, att5dw (5'-CTCAAATGATAGCAACAACAGG-3'), produced a fragment containing the *attPB'* site, which was also purified and sequenced.

2.12 Cloning and functional analysis of the holin gene *hol3626*

To study the function of the holin, a specialized vector system (λΔSthf) was used (VUKOV et al. 2000). *Hol3626* was amplified by PCR using primers hol3626-up (5'-ATCAGAATTC**TTA**ATTTTCTTTTTTATTAAATCCTTCTTT-3'), and hol3626-down (5'-ATCAGAATTC**ATG**TTTAAATTTATACCAGAAGTAATAAGT-3') (*Eco*RI sites are underlined, start and stop codons are in boldface [the native TAG stop codon was changed to TAA]). The product was cloned into the *Eco*RI site of λΔSthf, and packaged using a λ DNA packaging extract (Stratagene). A preliminary test for holin function was the ability to form plaques on *E. coli* LE392. Single plaques were isolated, and identity of *hol3626* verified by sequencing using primers as described earlier (VUKOV et al. 2000). Purified λΔSthf*::hol3626* was then used to lysogenize LE392. Growth and lysis kinetics of the lysogenic cells in liquid cultures was monitored by measuring optical density (OD_{600}). Lytic development was induced by a temperature shift from 30°C to 42°C for 20 min, followed by incubation at 37°C for 120 min. Lysogens carrying the wildtype *S* holin gene from λ (λΔSthf::*S*) served as positive control, and lysogens carrying λΔSthf without holin as negative control.

2.13 Cloning of endolysin gene *ply3626*

For cloning of *ply3626* into pSP72, primers PlyEcoStart (5'- ATCAGAATT-C*GAGGAG*AAATTACT**ATG**AAGATAGCAGAAAGAGGCGGTCATAAT-3'), and PlyEcoStop (5'-ATCAGAATTC**TTA**TATTCTTTCTAAATATTTAGCTGTAAA-3') (*Eco*RI sites are underlined, start/stop codons are in boldface, ribosome binding site is in italics) were used for amplification of the endolysin gene, from purified phage DNA. The PCR product was digested with *Eco*RI (Roche), and ligated into the *Eco*RI site of pSP72, which has been prepared by treatment with shrimp alkaline phosphate (U.S. Biochemicals). Ligation reactions were transformed into *E. coli* JM109(DE3). For cloning into pQE-30, primers PlyBamStart (5'-TCTAGGATCC**ATG**AAGATAGCAGAAAGAGGCGGTCATAAT-3'), PlySalStop (5'-

ATAAGTGTCGAC**TTA**TATTCTTTCTAAATATTTAGCTGTAAA-3') were used. A truncated Ply 5'-fragment encoding only the first 180 aa of the endolysin was amplified using PlyBamStart and PlySalMid (5'-AAGTGTCGAC**TTA**TTGTCCCCC-AGCAACTTTACAAACTCC-3'), the latter of which is complementary to bp 513-540 of *ply3626* and equipped with a stop codon. The PCR products were purified, digested with *Bam*HI and *Sal*I (MBI Fermentas), and ligated into pQE-30. Ligation reactions were transformed into *E. coli* JM109(pACYC-IRL10). Correct insertion and sequence of the genes was checked by nucleotide sequencing of the plasmid inserts.

2.14 Overexpression and purification of Ply3626

JM109(DE3)(pPly3626) failed to produce enzymatically active protein upon induction of protein synthesis by addition of IPTG (1 mM). With respect to JM109(pACYC-IRL10)(pHPL3626), IPTG induction at 37°C incubation temperature resulted in the rapid formation of inclusion bodies. Incubation temperature was then reduced to room temperature (21°C), and growth was monitored photometrically (OD_{600}) over a period of 20 hours. At regular intervals, samples of the growing cultures were taken, and protein synthesis documented by SDS-PAGE. At a final OD of approximately 1.0, cells were harvested by centrifugation (8,000 × *g*, 15 min at 4°C), and pellets frozen at -20°C overnight. After thawing, cells were resuspended in 1/50 volume in PBS (pH 8.0) or in Ni-NTA-buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, pH 8.0). Crude extracts were prepared by single passage through a French-Press cell (SLM Aminco) at 100 MPa. Debris was removed by centrifugation (30,000 × *g*, 10°C, 25 min), and supernatants containing the active enzyme were cleared by filtration (0.22 µm pore size), and stored at -20°C.

Purification of the hexa-His-tagged endolysin was carried out using immobilized metal-chelate affinity chromatography (IMAC), as previously described (LOESSNER et al. 1996). Individual fractions were checked by SDS-PAGE and assayed for lytic activity. Protein concentration was determined using

with a modified Bradford Assay (Nanoquant; Roth), and buffer exchange and concentration of the proteins were performed using centrifugal concentrators (Fugisep Maxi, 10 kDa cut-off, Intersep Systems).

2.15 Lytic activity and substrate specificity

A total of 48 *C. perfringens* strains, 30 strains of other *Clostridium* species. (*C. absonum, C. barati, C. beijerinckii, C. bifermentans, C. botulinum, C. butyricum, C. difficile, C. fallax, C. novyi, C. pasteurianum, C. sporogenes, C. tertium, C. tetani, C. tetanomorphum* and *C. tyrobutyricum*), and 34 bacterial strains of other genera (*Bacillus cereus, B. megaterium, B. polymyxa, B. subtilis, B. thuringiensis, B. weihenstephanensis, Bifidobacterium gallicum, Bf. gallinarum, Bf. ruminantium, Brochothrix thermosphacta, Campylobacter jejuni, Enterobacter cloacae, E. gallinarum, Enterococcus faecalis, Ec. faecium, Escherichia coli, Lactobacillus aviarius, Lb. plantarum, Lb. ruminis, Lactococcus lactis, Leuconostoc carnosum, Leu. citreum, Leu. mesenteroides, Listeria innocua, L. ivanovii, L. monocytogenes, Pediococcus pentosaceus, Salmonella enterica, Staphylococcus aureus, S. epidermidis, S. gallinarum, Streptococcus pyogenes* and *Strep. thermophilus*) were selected and tested for sensitivity against the recombinant phage lysin. Cells were resuspended in PBS to an optical density of approximately 1, and directly used as substrate, that were carried out in microplates at 37°C. Cell suspension (180 µl) were mixed with 20 µl (1:10) of crude extract of JM109(pACYC-IRL10)(pHPL3626), or 5 µl purified HPL3626. The decrease in optical density was monitored using a Victor^2 1420 platereader (Perkin-Elmer). As a negative control, cell extract of JM109(pACYC-IRL10) was used. The activity of the C-terminally truncated endolysin was determined with cell extracts from JM109(pACYC-IRL)(pHPL3626ΔC).

2.16 Nucleotide sequence accession numbers

The DNA sequences reported here can be accessed in the EMBL, Genbank, and DDBJ databases under accession numbers AY082069 and AY082070 (Appendix).

3 RESULTS

3.1 Isolation of *C. perfringens* phages

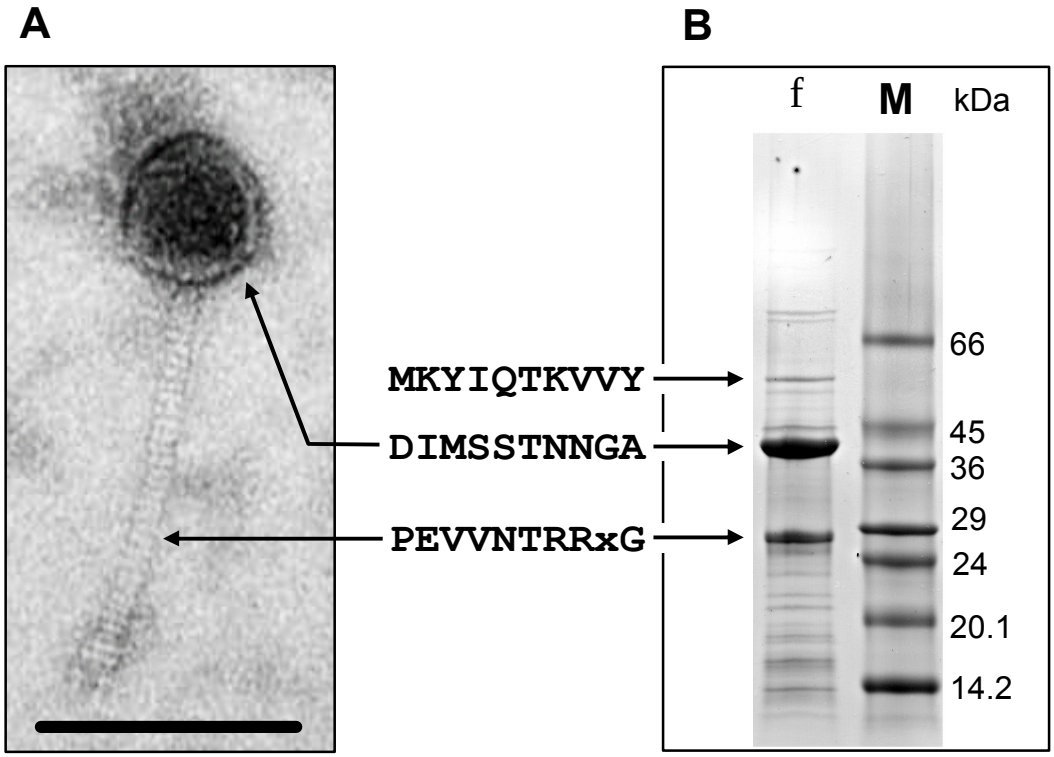
Lytic activity was observed in UV-induced culture supernatant of two of the tested strains (ATCC 3626 and NCTC 8533), which was shown to be due to the presence of bacteriophages. Their lytic ranges on 51 *C. perfringens* strains were different: Phage ϕ3626 was lytic against 11 strains (21.6 %) while phage ϕ8533 was able to lyse only 4 strains (7.8 %), which were also all sensitive to ϕ3626. The optimal propagation hosts were determined to be *C. perfringens* NCTC 3110 for ϕ3626, and ATCC 3628 for ϕ8533.

Since phage ϕ3626 displayed a broader host range and, at least in the framework of this project, was easier to propagate, it was chosen for further studies. Electron microscopy (Fig. 1) revealed that the ϕ3626 virion has an isometric capsid (55 nm diameter ± 2 nm; a total 7 virion particles were measured) and a long, flexible, non-contractile tail (170 nm ± 5 nm length). Therefore, ϕ3626 belongs to the Siphoviridae family of doublestranded DNA bacterial viruses in the order Caudovirales (ACKERMANN 1998).

3.2 Nucleotide sequencing and determination of the *cos*-sites

The genome of ϕ3626 was sequenced by a shotgun approach. Sequencing the various plasmid inserts yielded contigs which allowed the design of primers to close the gaps in-between the contigs. The sequence was finalized by determination of the *cos*-site core sequence from amplified bacterial chromosomal DNA (Fig. 2). Predicted restriction maps of the ϕ3626 DNA were in perfect agreement with the experimental achieved pattern, indicating that the sequences were assembled correctly (results not shown). The complete genome has a size of 33,507 nucleotides with 3'-protruding, singlestranded cohesive ends of 9

nucleotides (Fig. 2). Its average molar G+C content of 28.4 mol% is slightly higher than the 24-27 mol% reported for its host (HIPPE et al. 1991), or for the clostridial plasmid pIP404 (25 mol%).



C

In silico Translated Sequence	ORF	Functional Assignment
1-MKYIQTKVVY	16	Minor Structural Protein
115-DIMSSTNNGA	6	Major Head Protein [Cps]
2-PEVVNTRRCG	11	Major Tail Protein [Tsh]

Fig. 1 Electron micrograph and protein profile of f3626

(A) Electron micrograph of ϕ3626, showing its isometric capsid and flexible tail. Scale marker is 100 nm. **(B)** SDS-PAGE showing the protein profile of ϕ3626 (left lane) and molecular mass marker proteins (right lane). N-terminal amino acid sequences from selected structural proteins are shown between the images. The arrows point to the respective SDS-PAGE bands and individual, corresponding viral components, as deduced by bioinformatic analysis. **(C)** Genetic identity and functional assignments for the three proteins are given. The mature major head protein begins with amino acid D115 of the deduced sequence, indicating proteolytic processing.

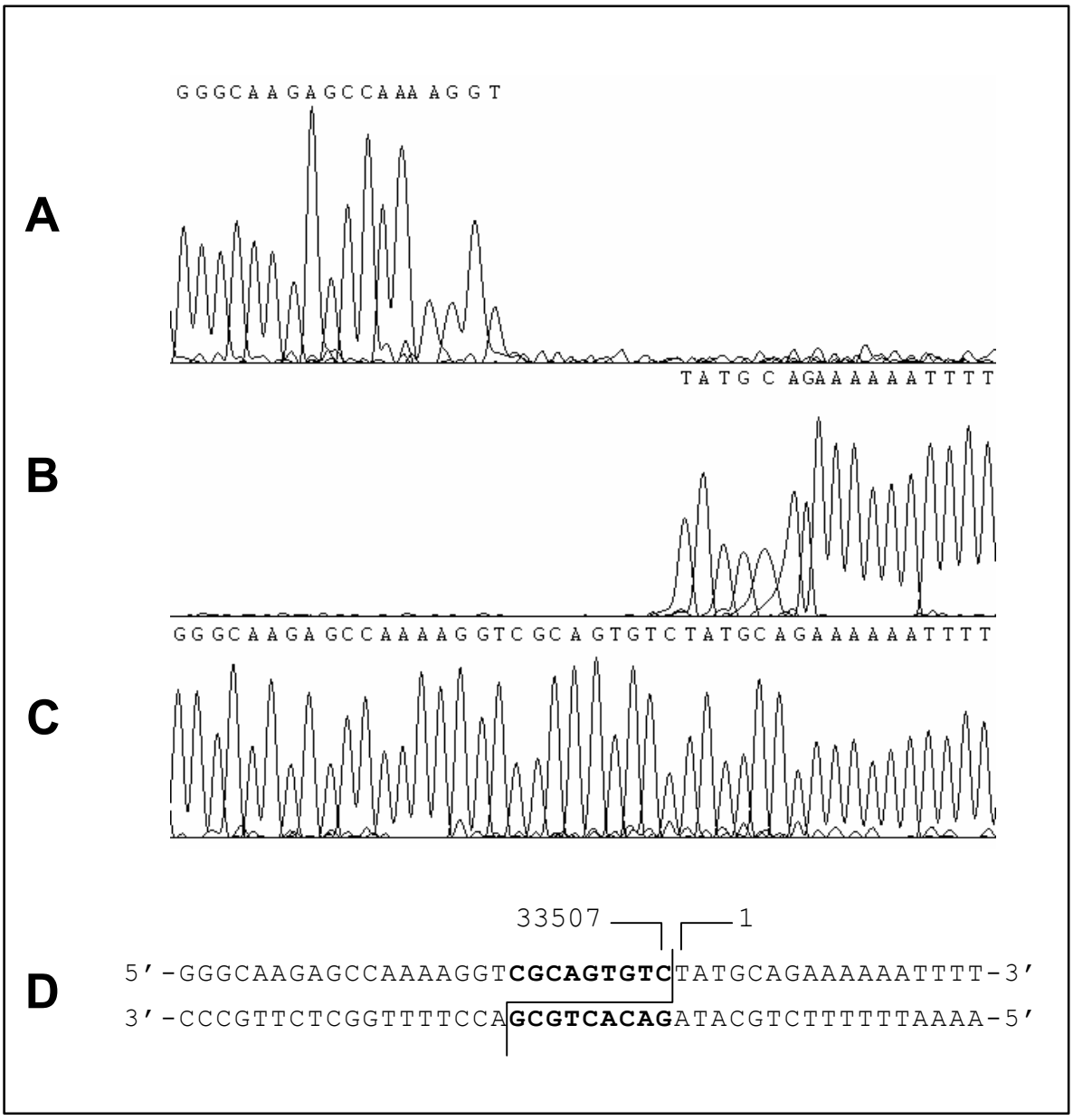


Fig. 2 Identification of single stranded cohesive ends in ϕ3626 DNA.

(A) Chromatogram showing the sequence from the right end of the genome towards the cos-site (coordinates 33481-33498). **(B)** Chromatogram of the sequence from the left end of the non-circular genome towards the cos-site (shown inversed; coordinates 1-17). **(C)** Sequence of a PCR product from ϕ3626 prophage, spanning the entire cos-site. **(D)** Corresponding sequence of the ligated cohesive ends, joining the left and right arm of the DNA molecule. The single stranded 3'-protruding ends are printed in bold letters.

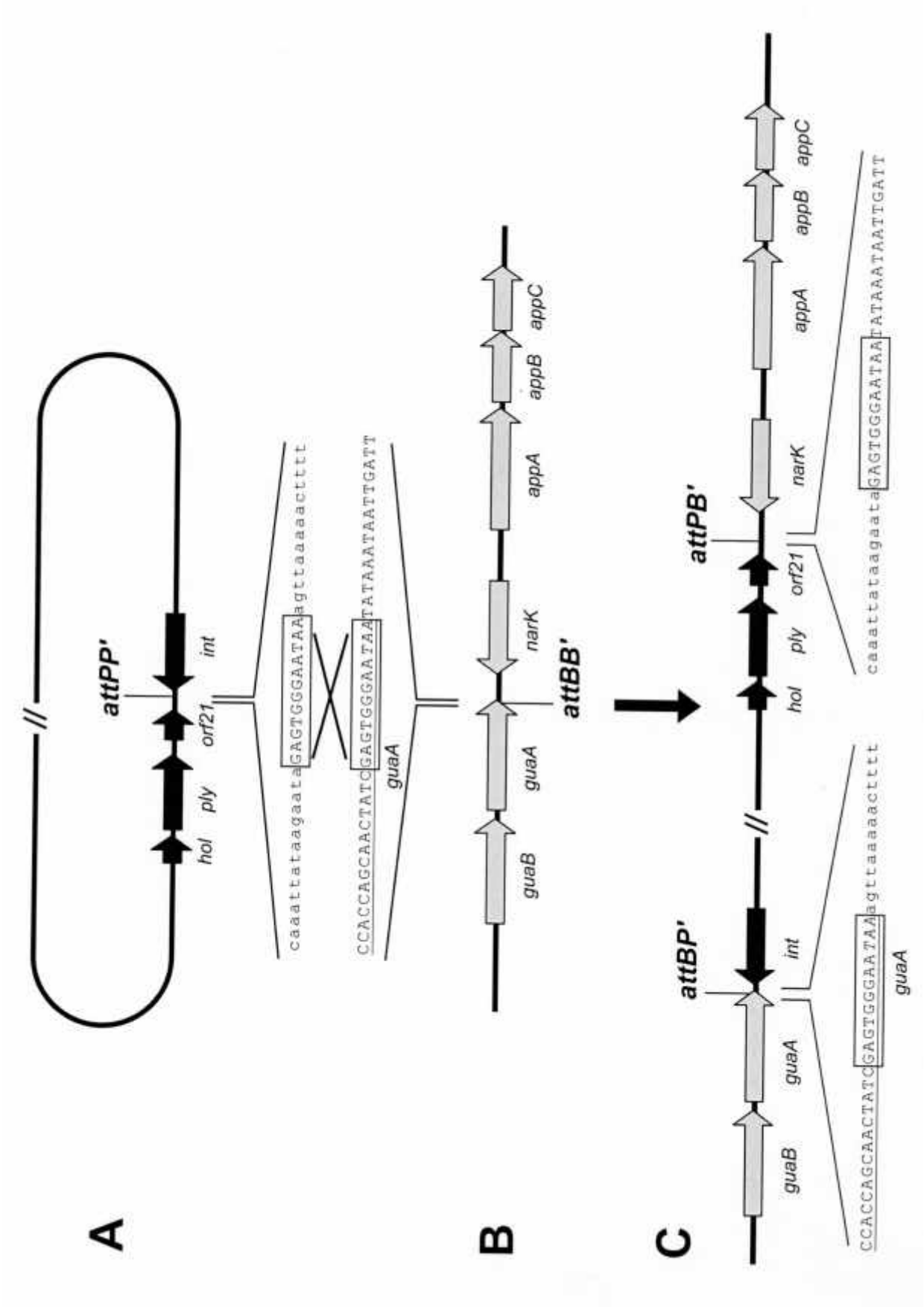
3.3 Identification and nucleotide sequence of the attachment sites *attPP'* and *attBB'*

The integration site of the bacteriophage ϕ3626 was identified using an inverse PCR approach. Sequencing of the first PCR product yielded 663 nt bacterial sequence. This ("left-arm") prophage-host junction was designated *attBP'* (Fig. 3C). The bacterial sequence was found to be 100 % identical to a portion of a contig (10.804 bp) of the unfinished genome of *C. perfringens* (TIGR, http://www.tigr.org). A second PCR product yielded additional sequence information of the host (183 nt), which was also identical over its full length to the *C. perfringens* sequence, and encompassed the *attPB'* site, at the junction of the "right arm" of ϕ3626 and the bacterial chromosome.

Alignment of the two *att* site flanking sequences revealed a core sequence of 12 nt (Fig. 3). On the circularized phage genome, *attPP'* is located in a non-coding region of 273 bp between ORF 21 and *int*. Sequence analysis showed that the *attBP'* is encompassed within the 3'-end of a *C. acetobutylicum guaA* homologue, encoding a guanosine monophosphate (GMP) synthetase (89 % similarity over 220 aa). The core sequence of 12 bp represents the terminal 4 tripletts of *guaA*, including a TAA stop codon (Fig. 3). The region immediately downstream of *attPB'* (189 nt) does not contain coding capacity.

Fig. 3 Organization of phage and bacterial attachment sites (next page)

(A) Schematic representation of the circularized ϕ3626 genome with its *attPP'* site (see Table 2 for identity of ORFs). **(B)** Partial sequence of the *C. perfringens* genome, encompassing attBB' and surrounding genes (see text). **(C)** Integrated prophage status within *attBP'* and *attPB'*. The ORFs flanking *attPP'* and *attBB'* are indicated by arrows (black arrows, phage/prophage ORFs; grey-shaded arrows, *C. perfringens* ORFs). Partial sequences of junction fragments from the phage/prophage are denoted in small letters, host sequence in capital letters, the homologous att site core sequence (12 bp) is boxed. Underlined sequence corresponds to the 3'-end of *guaA*, the stop codon is shown in italic letters. Figure is not to scale.


A
attPP'
hol
ply
orf21
int
caaattataagaaataGAGTGGGAATAAagttaaaaactttt
CCACCAGCAACTATCGAGTGGGAATAATATAAATAATTGATT
guaA
B
guaB
guaA
narK
attBB'
appA
appB
appC
C
attBP'
guaB
guaA
int
attPB'
hol
ply
orf21
narK
appA
appB
appC
caaattataagaaataGAGTGGGAATAATATAAATAATTGATT
CCACCAGCAACTATCGAGTGGGAATAAagttaaaaactttt
guaA

In order to obtain a more detailed picture of the chromosomal localization of *attBB'* and *guaA*, a 10.8 kb sequence contig of the unfinished genome of *C. perfringens* (The Institute for Genomic Research, TIGR) was annotated. Upstream of *guaA*, *guaB* could be identified, encoding an inosin-5'-monophosphate (IMP) dehydrogenase (82 % similarity over 478 aa, with *C. actetobutylicum* GuaB (NÖLLING et al. 2001)). Downstream of *attBB'*, four coding regions could be identified, the products of which are similar to NarK of *Pseudomonas aeruginosa* (Acc.-No. Y15252; 40 % over 315 aa), and AppA, AppB, and AppC, which are part of an oligopeptide transporter operon present for example in *C. actetobutylicum* (NÖLLING et al. 2001) (similarities of 53 % over 580 aa, 63 % over 288 aa, and 71 % over 323 aa, respectively).

3.4 Identification and organization of ϕ3626 ORFs

Bioinformatic analysis revealed the existence of 50 putative protein coding regions (Tab. 4) on the ϕ3626 genome, covering 94.1 % of the sequence. The criteria for the characterization of a potential ORF were the existence of an ATG, GTG (5 present), or TTG (3 present) start codon, and a minimum coding capacity of 40 amino acids. Except for ORF41, all ORFs where proceeded by a recognizable potential ribosome binding site complementary to the 3'-end of *C. perfringens* 16S rRNA (GARNIER et al. 1991).

On the whole, the transcriptional units of ϕ3626 (as deduced from bioinformatic analysis) can be organized into three major functional clusters, which is also reflected by the transcriptional direction of the ORFs (Fig. 4). The first cluster from the *cos*-site at coordinate 1 up to position 19804 is transcribed rightward in the genomic map (Fig. 4), and represents genes encoding structural proteins and the lysis system. These genes can be summarized as 'late genes'. The second cluster (nt 19805 to 23645) likely encodes products responsible for the control of lysogeny including the *att*-site, an integrase, the repressor and a putative *cro*-like protein. The third cluster (nt 23680 to 33507) includes rightward-

facing ORFs, whose putative products represent the early genes, involved in replication, recombination and modification of phage DNA.

3.5 Structural proteins of f3626

The proteins building up the virion particle were separated by SDS polyacrylamide electrophoresis (Fig. 1). Microsequencing yielded N-terminal sequences of three structural proteins, which enabled identification of the corresponding genes.

The major capsid component Cps (NH_2-DIMSSTNNGA, encoded by ORF6) resembles 43.3 % of total phage protein (Fig. 1). The N-terminal sequence indicated that the gene product must be post-translationally processed between R_{114} and D_{115}, which results in a decrease in size from 47.7 kDa to 34.3 kDa. Cps displays a high probability to form a coiled-coil structure in the N-terminal portion removed by processing, a finding similar to what has been reported for phages Sfi21 and HK97 (CONWAY et al. 1995; DESIERE et al. 1999).

The major tail protein Tsh (NH_2-PEVVNTRRXG, encoded by ORF11) corresponds to 12.7 % of the total protein, with an apparent size of 27 kDa (Fig. 1). The predicted size was 22.5 kDa. This observation was also made for other phages, such as *Listeria* phage A118 (LOESSNER et al. 2000). The amino acid sequence differs from the predicted sequence only by the absence of the initiator methionine, as is often observed in prokaryotes with proline as the penultimate amino acid (HIREL et al. 1989)

A minor structural protein (Gp16, 2.1 % total protein content) was also isolated, and its N-terminal sequence (NH_2-MKYIQTKVVY) was in perfect agreement with the putative product of ORF16. The predicted size of 55.1 kDa is identical to the experimentally determined size.

3.6 Bioinformatic analysis of f3626 gene products

Deduced amino acid sequences of the fifty ORFs were compared with known sequences from the databases to uncover similarities to genes with known function. Functional assignments and significant homologies to other proteins are listed in Tab. 4. The findings are described below in more detail.

Tab. 4 Open reading frames, gene products, functional assignments, and amino acid sequence homologies of ϕ3626

ORF	Start	Stop	Gp (kDa)	Gp (pI)[a]	Functional assignments	Homologies (gp/orfs are given in brackets)[b]
1	54	557	19.2	7.2	Terminase, small subunit	ϕ105 (gp21) Sfi19 (gp161) bIL309 (gp38)
2	550	2295	67.7	5.9	Terminase, large subunit	ϕ105 (gp22 & 23) ϕSLT (gp563) D3 (gp2)
3	2296	3564	48.7	8.3	Portal protein	ϕ105 (gp25) ϕSLT (gp412) D3 (gp4)
4	3554	3715	6.4	9.7		
5	3720	4325	23.1	4.7	Putative prohead protease	ϕPV83 (gp41) ϕC31 (gp35) ϕPVL (gp5a)
6[c]	4366	5631	47.7	4.9	Major capsid protein [Cps]	ϕ105 (gp27) bIL285 (gp44) ϕadh (gp395)
			(34.3)	4.7	Processed form, cleaved between R_{114} and D_{115}	
7	5641	5919	10.8	4.9		ϕ105 (gp30)
8	5912	6235	13.2	8.8		
9	6236	6685	17.0	8.7		ϕ105 (gp32)
10	6687	7037	14.0	4.4		
11[c]	7054	7650	22.5	4.9	Major tail protein [Tsh]	ϕ105 (gp34)
12	7665	7979	11.9	9.7		
12.1[d]	7665	8485	32.2	4.3		
13	8515	11400	103.8	10.0	Tail tape-measure protein [Tmp]	ϕSLT (gp1374) TP901-1 (TMP) ϕPVL (gp15)
14	11401	12102	27.4	4.8		PSA (gp13) bIL285 (gp53)
15	12102	15053	110.0	6.2		PSA (gp14) bIL285 (gp54) ϕg1e (gp1608)
16[c]	15106	16584	55.1	5.5	Minor structural protein	ϕ105 gp42
17	16597	16905	12.1	4.8		
18	16914	17669	28.2	4.8		bIL170 (gpI12, accession number AF009630)
19	17684	18061	14.3	9.6	Holin	ϕ105 (gp45)
20	18149	19192	38.8	5.9	Endolysin	pIP404 (ORF10) *C. perfringens* (ORF 10C) PSA (Ply) 12826 (Ply12)
21	19382	19804	15.9	4.8		
22	21119	20070	41.3	9.9	Integrase	T12 (Int) PSA (gp24) T270 (Int)
23	21587	21135	17.6	5.2		PSA (gp26) ϕSLT (gp153) A118 (gp35)

continued

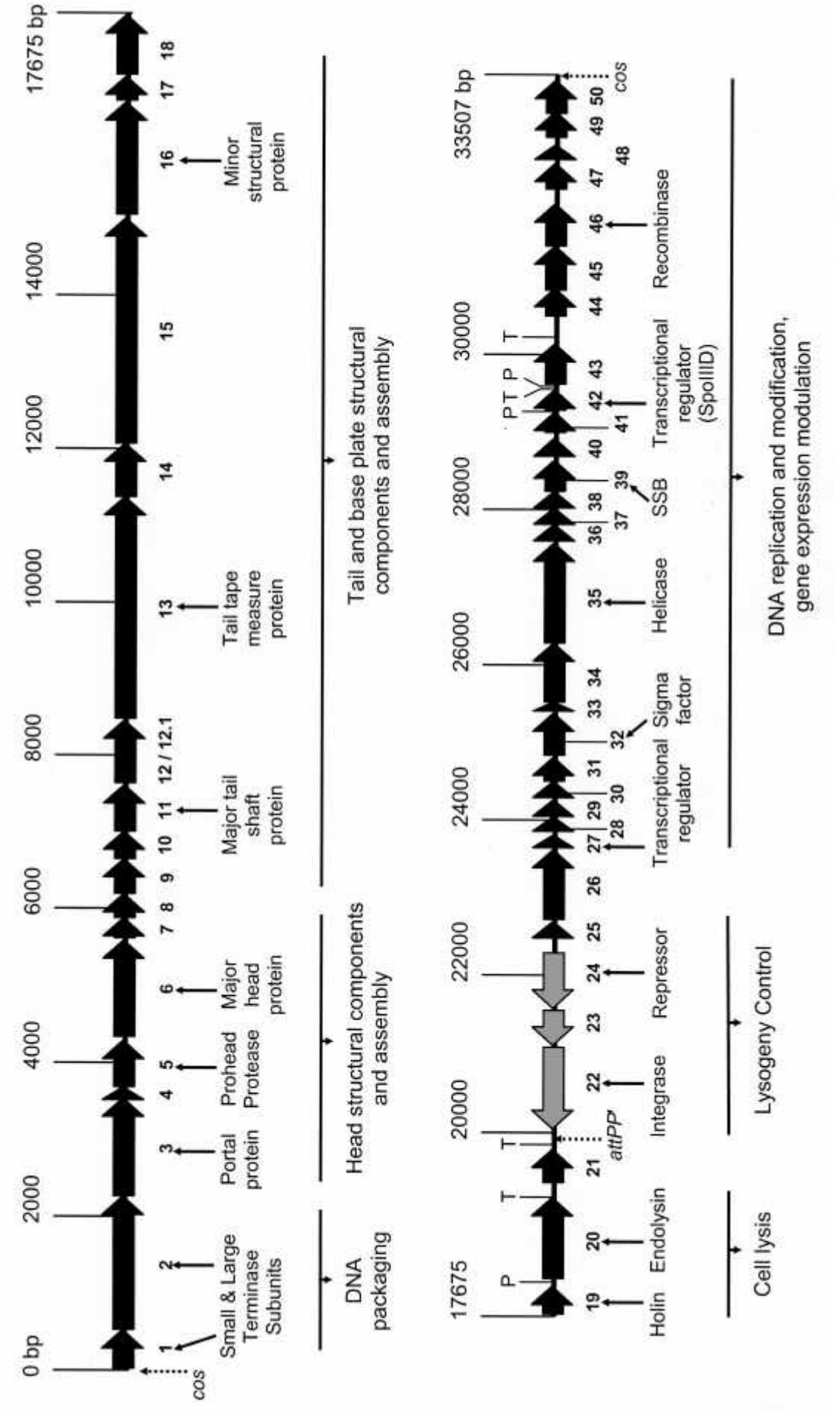
Tab. 4 continued

ORF	Start	Stop	Gp (kDa)	Gp (pI)[a]	Functional assignments	Homologies (gp/orfs are given in brackets)[b]
24	22330	21608	27.4	5.1	Repressor	ϕSLT (gp101a) ϕg1e (gp132), PBSX (Xre)
25	22504	22722	8.3	9.2	Cro	A2 (Cro)
26	22749	23645	34.5	8.7		ϕSLT (gp81a) ϕadh (Tec)
27	23680	23856	6.8	9.6	Transcriptional regulator	ϕETA (gp13,) ϕPV83 (gp12)
28	23903	24082	6.7	9.8		
29	24093	24299	7.9	9.2		
30	24319	24537	8.5	10.3		
31	24527	24853	13.1	4.5		
32	24867	25412	21.3	9.4	Sigma factor	*C. acetobutylicum*, *B. subtilis* and of *C. perfringens*
33	25428	25547	4.6	8.9		
34	25563	26315	29.3	9.4		ϕETA (gp22) ϕSLT (gp256) r1t (gp11)
35	26325	27608	48.3	8.3	Helicase	P1 (Ban) SPP1 (gp40) D3 (gp74)
36	27625	27825	7.9	4.4		
37	27825	28061	9.3	4.9		
38	28058	28270	7.9	6.4		
39	28263	28679	15.4	4.9	SSB	Ul36 (gp141) bIL286 (SSB) A118 (SSB)
40	28706	28942	9.1	4.1		
41	29020	29295	10.9	10.0		
42	29304	29576	10.5	10.9	SpoIIID	*C acetobutylicum*, *B. subtilis* and other Bacilli
43	29638	30168	21.1	8.7		
44	30515	30865	13.5	9.5		
45	30858	31424	21.4	4.7		
46	31428	31973	21.4	10.2	Recombinase	L2 (gp5) Wϕ (Int) 186 (Int,)
47	32153	32500	13.4	9.5		
48	32533	32718	7.4	4.6		
49	32735	33058	12.6	9.9		
50	33039	33467	16.6	9.8		ϕSLT (gp104b) bIL286 (gp39) ϕadh (gp170)

Tab. 4 continued [a] Predicted by computer analysis; [b] For reference or accession numbers, see text; [c] Experimentally shown; [d] Translational frameshift might result in two C-terminally different products (see 3.5.4).

Fig. 4 Schematic representation of the f3626 genome. (next page)

Assumed ORFs, some functional assignments and overall genetic organization. The ORFs are numbered consecutively (see Table 3) and are indicated by arrows or arrowheads that point in the direction of transcription. Black arrows indicate rightward transcription, grey-shaded ORFs are oriented leftwards. Their relative position on the genome (33,507 nt) is indicated by coordinates. The *attPP'* and the *cos*-sites are shown by a dashed arrows. P indicates some specific putative sporulation dependent promoters, and T stands for putative rho-independent transcription terminators (see text for explanation)


0 bp
2000
4000
6000
8000
10000
12000
14000
17675 bp
cos
1
2
3
4
5
6
7
8
9
10
11
12 / 12.1
13
14
15
16
17
18
Small & Large Terminase Subunits
Portal protein
Prohead Protease
Major head protein
Major tail shaft protein
Tail tape measure protein
Minor structural protein
DNA packaging
Head structural components and assembly
Tail and base plate structural components and assembly
17675
20000
22000
24000
26000
28000
30000
33507 bp
P
T
T
19
20
21
attPP'
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
39
38
40
41
42
43
P
T
T
44
45
46
48
47
49
50
cos
Holin
Endolysin
Integrase
Repressor
Transcriptional regulator
Sigma factor
Helicase
SSB
Transcriptional regulator (SpoIIID)
Recombinase
Cell lysis
Lysogeny Control
DNA replication and modification, gene expression modulation

3.6.1 ORF 1 and 2

Fifty-four nucleotides downstream of the *cos*-site, two ORFs have been allocated that likely represent the small and large subunit of the terminase, which introduces specific cuts into the concatemeric DNA at the *cos*-site to initiate genome packaging. Gp1 shares similarities with gp21 of *Bacillus subtilis* phage ϕ105, gp161 of phage Sfi19 of *Streptococcus thermophilus* (DESIERE et al. 1998) and gp38 of prophage bIL309 of *Lactococcus lactis* (CHOPIN et al. 2001), with similarities between 43-47 % over stretches of 132-169 aa. The large subunit (gp2) displayed significant similarities to the gp22 and gp23 ϕ105 (55 % over 416 aa and 48 % over 88 aa, respectively), and to terminases from various other phages.

3.6.2 ORF 3-5

Gp3 encodes a putative portal protein, based on similarities to gp25 of ϕ105, and the portal proteins of ϕSLT (NARITA et al. 2001) and D3 (KROPINSKI 2000) (39-45 % identity over 313-397 aa). The localization of ORF4 disrupts the pattern found in ϕC31, D3 and HK97, which consists of a consecutive order of portal protein, prohead protease and a N-terminally processed major capsid protein (DUDA et al. 1995; SMITH et al. 1999; KROPINSKI 2000). No similarities to any potential protein could be found for Gp4. Gp5 is similar to putative prohead proteases from various phages, including *S. aureus* phages ϕPV83 gp41 (KANEKO et al. 1997) and ϕPVL gp5a (KANEKO et al. 1998), *Streptomyces* phage ϕC31 gp35 and bIL309 gp36. Homologies range from 58-62 % similarity over stretches from 147-172 aa.

3.6.3 ORF 7 and 9

For these ORFs, no functional assignments could be made. Only a congruent gene order is obvious by the similarities to putative gene products 30 and 32 of ϕ105 with similarities of 50 % over 87 aa and 45 % over 121 aa, respectively. Also, the deduced sequence of the major tail protein (encoded by ORF11) displays shares similarities to gp34 of ϕ105 (42 % over 146 aa).

3.6.4 ORF 12/12.1

Tailed bacteriophages frequently have a pair of overlapping ORFs between the major tail protein gene and the tail length tape measure gene that are expressed by a translational frameshift (JUHALA et al. 2000). This seems to be the case for ORF12/12.1, resembling the situation in λ phage (LEVIN et al. 1993). ORF12 starts at nt 7665, and the obvious stop codon is located at 7979. However, just 27 nucleotides upstream of this stop codon, a putative heptanucleotide 'slippery sequence' (GGGUUUU) can be allocated, encoding the dipeptide Gly-Phe in both, the zero and -1 reading frames (Fig. 5). At this site ribosomes might shift frames and continue in the -1 frame until termination at nt 8485, resulting in a larger gene product as described for the gpG-T of λ phage (LEVIN et al. 1993).

```
    7921  ATGATAAAGCTTTTAAAGCGAGTGC***GGGTTTT***ATGAAAGGCAAGAAATTCAAGAGA**TAG**G

 0 FRAME  N  D  K  A  F  K  A  S  A  G  F  M  K  G  K  K  F  K  R  *
-1 FRAME                           G  F  Y  E  R  Q  E  I  Q  E  I  G
```

Fig. 5 Putative translational frameshift.
A fragment of the ORF12/12.1 is shown and the deduced amino acid sequence. The "slippery sequence" is in bold and italic letters. The stop codon is in bold letters and the amino acids encoded by the "slippery sequence" are in italic and underlined

3.6.5 ORF 13

The gene product of ORF13 encodes a large protein (103.8 kDa) sharing similarities with the tape measure proteins (Tmp) of lactococcal bacteriophage TP901-1 (PEDERSEN et al. 2000), and gp1374 of ϕSLT, gp15 of ϕPVL, and many other phages.

3.6.6 ORF 15

The deduced gp15 seems to consist of a mosaic sequence. A HMMscan indicated two areas with similarities to different protein families. The area between aa 892-975 displayed a high probability being a peptidase M37, and the area from aa 681-801 displayed a similarity to a phage lysozyme domain. The latter domain showed similarities to the *Salmonella* phage P22 endolysin (VANDER BYL and

KROPINSKI 2000), and to the lactococcal phage C2 endolysin (WARD et al. 1993), (41-42% over 157-169 aa).

3.6.7 ORF 19 and 20

These two ORFs located at the distal end of the "late gene" region encode a dual lysis system consisting of a holin (Hol) and an endolysin (Ply). This genetic organization is common to many viruses of the Siphoviridae family (LUCCHINI et al. 1999b). Together, they are responsible for the release of phage progeny from infected host cells.

The deduced amino acid sequence from ORF19 displayed similarity to a putative holin of *B. subtilis* phage ϕ105 (KOBAYASHI et al. 1998) (50 % over 105 aa). A dual start motif (YOUNG and BLÄSI 1995) is not present in hol3626. Bioinformatics revealed that hol3626 encodes a protein with two highly hydrophobic amino acid sequence regions (Fig 6 B) which, by TmHMM analysis, were strongly suggested to function as membrane spanning domains (MSD). Applying the "positive-inside" rule (VON HEIJNE 1992), the positively charged amino acids located in the putative beta-turn region between the two MSDs (coordinates 33-46; Fig. 6A) suggest a location of this domain on the "inside" of the cytoplasmic membrane, when the polypeptide is inserted into the membrane (Fig. 6 C). This suggests that both the N- and the C-terminus are most likely located on the "outside" of the membrane (Fig. 6 C)

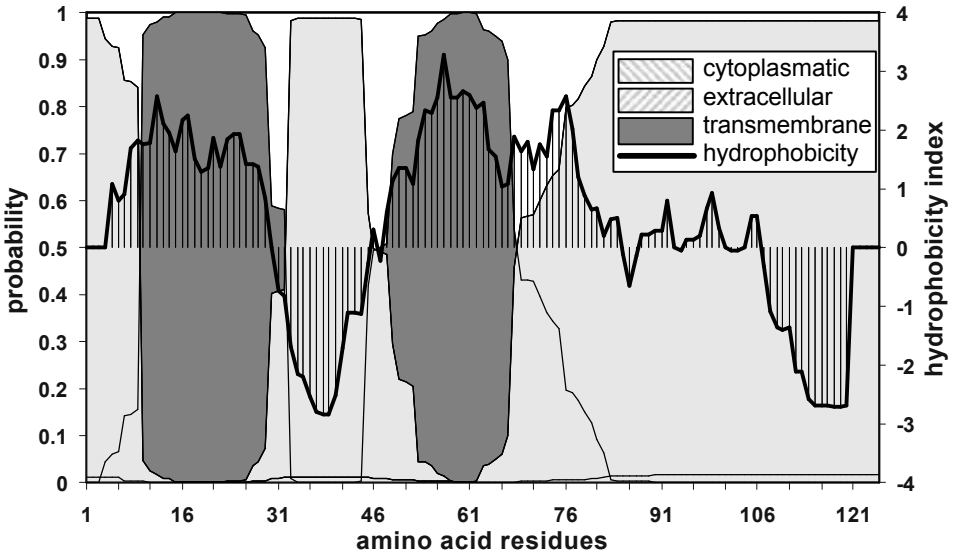
Fig. 6 Bioinformatic analysis of Hol3626. (next page)
(A) Amino acid sequence of Hol3626. Charged amino acid residues are indicated by + and - signs, respectively, and sequences with a high probability to form MSD (by TmHMM, see text) are boxed. **(B)** The probability to form MSD (left y-axis, dark grey areas) was predicted by TmHMM, and probable cytoplasmic and extracellular regions of the polypeptide are indicated by grey diagonal stripes. An hydrophobicity plot (right y-axis) (KYTE and DOOLITTLE 1982) was overlaid (bold black line). Hydrophobic region are in the positive area, whereas a negative hydrophobicity indicates a hydrophilic region. **(C)** The putative model of Hol3626 topology when inserted in the cell membrane (see text). Grey bars indicated the membrane spanning domains. Charged amino acid residues in the hydrophilic area between the two putative MSD are indicated.

A

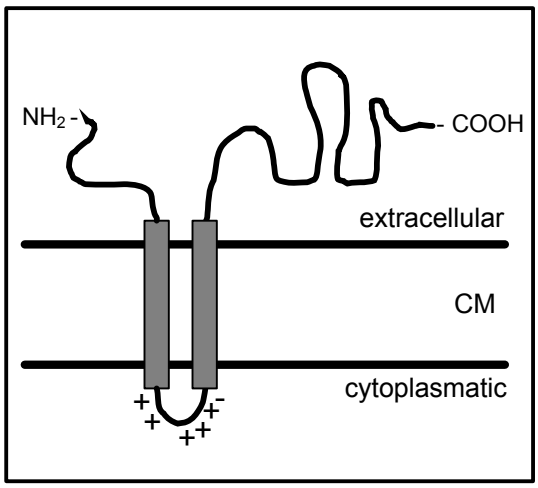
```
          10        20        30        40        50        60
MFKFIPEVISWLLVLYIGFKIIDMILGVLKTIKNKNYRSRKMRDGIIRWVAELMAIAFVL
  +   -           +  -      +  + +  + ++ +-   +   -

          70        80        90       100       110       120
ILDMFLGLKFTVIGVTLALFAYKEAGSIVENLGECGVELPEIVSEKLEVLNKNNKNKEGFNKKEN
  -     +            +-     -   -   -  -   -+ -   +  + +-   ++-
```

B



C



ORF20 probably encodes an N-acetylmuramoyl-L-alanine amidase. A HMMscan analysis revealed a N-acetylmuramoyl-L-alanine amidase domain in a 116 amino acid N-terminal portion of Ply3626 (Fig. 7). This assignment was further supported by sequence alignments, which also indicated relatedness of the Ply3626 N-terminal region to amidases, in particular the PlyPSA enyzme from *L. monocytogenes* phage PSA (SATTELBERGER et al. 2002), the *B. subtilis* autolysin CwlB, and the *B. cereus* phage 12862 endolysin (KURODA and SEKIGUCHI 1991; LOESSNER et al. 1997; SATTELBERGER et al. 2002). Interestingly, extended similarities (72-75% over 265-346 amino acids) exist to hypothetical proteins of unknown function of *C. perfringens* or the *C. perfringens* plasmid pIP404 (GARNIER and COLE 1988; LYRISTIS et al. 1994; SHIMIZU et al. 1994).

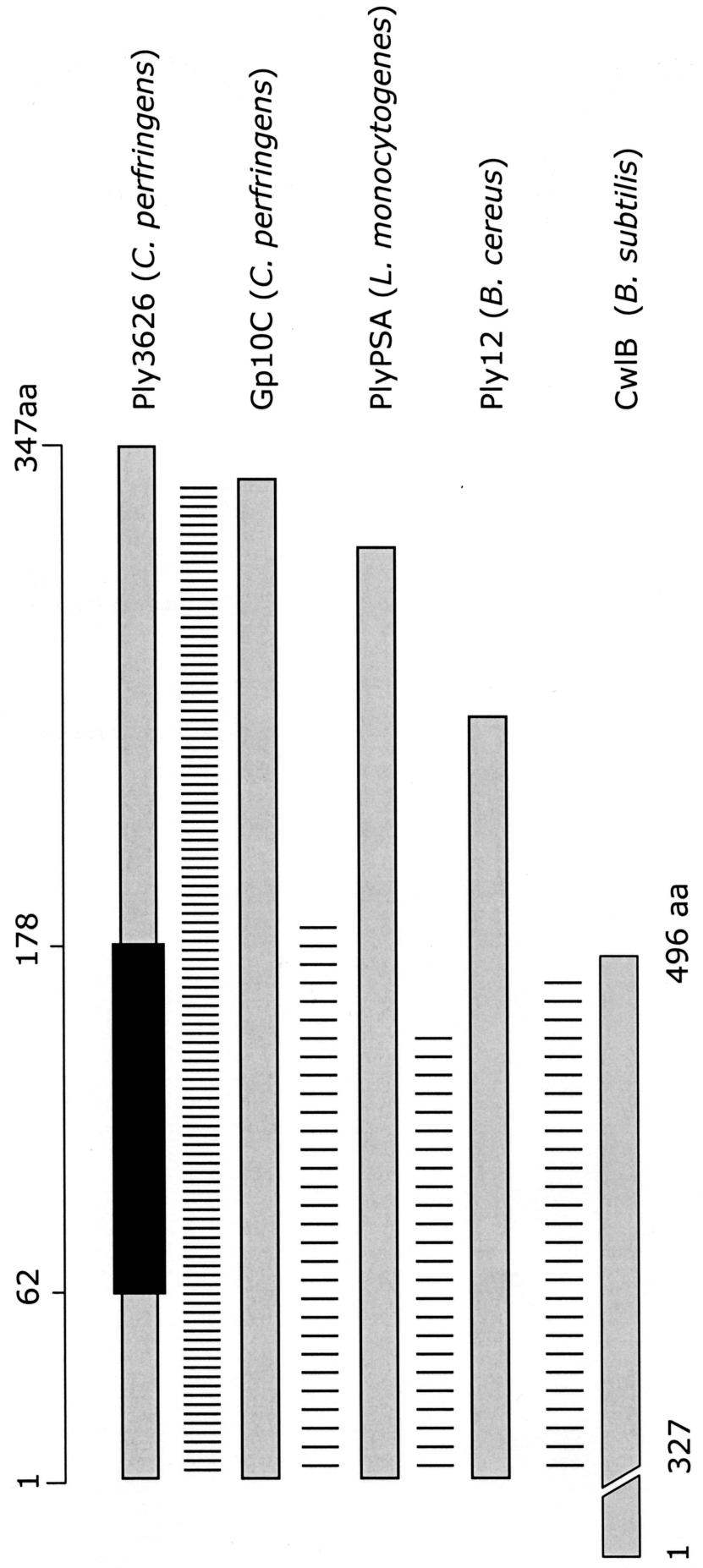
The two lysis genes have been cloned and their function was demonstrated experimentally (see 3.8 and 3.9)

3.6.8 ORF 22

This ORF is located immediately upstream of the attachment site (see 3.3). A HMMscan indicated that it encodes a phage integrase belonging to the λ integrase family, responsible for the site-specific recombination of ϕ3626 into the *C. perfringens* chromosome. Furthermore a relationship to Int459 of the transposon like element CW459*tet*(M) from *C. perfringens* was found (ROBERTS et al. 2001) (similarity of 47 % over 363 aa). Many other integrases of phage origins also displayed homology

Fig. 7 Alignment of Ply3626 with other amidases (next page)

Alignment of the amino acid sequence comparisons of Ply3626, with the gene product of ORF 10C from *C. perfringens*, and three amidases of other origin. Size of the proteins (in aa) are indicated by the ruler. The black boxed area indicates the N-acetylmuramoyl-L-alanine amidase domain identified by HMMscan. The vertical bars indicate the extend and degree of similarities among the different polypeptides.


Ply3626 (C. perfringens)
Gp10C (C. perfringens)
PlyPSA (L. monocytogenes)
Ply12 (B. cereus)
CwlB (B. subtilis)
1
62
178
347aa
1
327
496 aa

3.6.9 ORF 23

No functional assignment was possible with gp23, but similar putative gene products are found in the same loci between the integrase and the repressor in other bacteriophages, such as PSA (gp26) (SATTELBERGER et al. 2002), ϕSLT (gp153) and A118 (gp35) with similarities of 45-49% over 90-123 aa

3.6.10 ORF 24-25

Gp24 displays in the N-terminus homologies to several repressors of phages. Strongest hits are the repressor of ϕSLT, the repressor Xre of the *Bacillus* prophage PBSX (TAKAMI et al. 2000) and the repressor of ϕg1e (KODAIRA et al. 1997)) (similarities from 53-64 % over 62-70 aa). In the opposite direction, ORF25 encodes a product with similarity to Cro of *Lactobacillus casei* phage A2 (LADERO et al. 1999) (47 % similarity over 70 aa). Both proteins contain putative H-T-H motifs similar to CI and Cro of λ (results not shown), indicating their potential to bind to DNA.

3.6.11 ORF 26

The product of this gene displays similarity to a putative anti-repressor of phage ϕSLT (56 % similarity over 71 aa). The amino terminus also displayed similarity (67 % over 60 aa) to the Tec protein (topological equivalent of *cro*) found in ϕadh. This implicates that gp26 might also be involved in lysogeny control.

3.6.12 ORF27

Gp27 displays similarities (64 % over 54 aa) to a transcriptional regulator of *B. subtilis* (KUNST et al. 1997), but also to phage proteins with unknown function, such as the gp13 form *S. aureus* phage ϕETA (YAMAGUCHI et al. 2000), and gp12 of ϕPV83, both with 66% similarity over a length of 45 aa.

3.6.13 ORF 32

Both HMMscan and BLAST searches suggest that gp32 is a sigma factor, similar to sporulation-specific sigma factors of *C. acetobutylicum* (σ^F) (NÖLLING et al. 2001), *B. subtilis* (σ^E) (KUNST et al. 1997) and *C. perfringens* (σ^K) (STIREWALT and MELVILLE 2000) (45 % to 53 % over 174-219 aa).

3.6.14 ORF 34

ORFs with similar gene products can be found in ϕETA (gp22), in ϕSLT (gp256) and in the phage r1t from *Lactococcus lactis* (VAN SINDEREN et al. 1996), all with no known function.

3.6.15 ORF 35

Gp35 seems to be a helicase, responsible for the unwinding of DNA before replication. The strongest similarities are with DnaC of *C. acetobutylicum* (NÖLLING et al. 2001) and *B. subtilis* (KUNST et al. 1997) (51-55 % over 424 aa), but also to phage helicases from SPP1 (ALONSO et al. 1997) and others.

3.6.16 ORF 39

HMMscan and BLAST analysis suggest that this gene product most probably is a single stranded binding protein (SSB). Many phage-encoded SSB show a similarity to gp39; some strongest hits are the *L. lactis* phage ul36 (BOUCHARD and MOINEAU 2000), bIL286 (CHOPIN et al. 2001) and A118 (LOESSNER et al. 2000), with similarities of 62-65% over 125-144 amino acids.

3.6.17 ORF 42

Gp42 shows significant similarities (66-71 % over 74 aa) to σ^E/σ^K-dependent transcriptional regulators, also known as a stage III sporulation protein D (SpoIIID), found in *C. acetobutylicum* (NÖLLING et al. 2001), *B. subtilis* (KUNST et al. 1997), and other Bacilli (YOSHISUE et al. 1995; LINDBACK and KOLSTO 1999; TAKAMI et al. 2000).

3.6.18 ORF 46

As gp22 the gp46 probably belongs to the family of λ integrases. Similarities are found to several bacterial recombinases; C. actetobutylicum, B. haolurans or S. aureus (TAKAMI et al. 2000; KURODA et al. 2001; NÖLLING et al. 2001) (53-61 % similarity over 174-179 aa). The mycoplasma virus L2 integrase (MANILOFF et al. 1994), the coliphage Wϕ integrase (LIU and HAGGARD-LJUNGQUIST 1999), and the integrase of the coliphage 168 (KALIONIS et al. 1986) share sequence similarities ranging from 42-47 % over 147-152 aa.

3.6.19 ORF 50

Although no function could be assigned to gp50, similar proteins are encoded by many phages, such as ϕSLT (gp104b), bIL286 (gp39) and ϕadh (gp170), with similarities from 45-59 % over 91-133 aa.

3.7 Other features of the f3626 genome

Several potential stem-loop forming sequences were identified at the ends of possible transcriptional units, likely representing rho-independent transcription terminators: (i) between the lysis cassette and ORF 21, at position 19195-19230 (ΔG = -32.3 kcal mol^{-1}); (ii) between ORF 21 and the lysogeny control region (19856 - 19888; ΔG = -18.0 kcal mol^{-1}); (iii) downstream of ORF 42 (SpoIIID homologue) (29587-29620; ΔG = -17.1 kcal mol^{-1}); and (iv) in a non-coding region downstream of ORF43 (30462-30488; ΔG = -15.4 kcal mol^{-1}). Sequences similar to the P1 promoter upstream of the *C. perfringens* enterotoxin gene *cpe* and to the *B. subtilis* σ^K dependent promoter (ZHAO and MELVILLE 1998) have been allocated upstream of ORF42, possibly encoding a σ^K/σ^E regulated SpoIIID homologue (coordinates 29260-29287), but could also be found upstream of *ply* (18007-18033) and upstream of ORF43 (29688-29713). Interestingly, putative stem-loop structures are present downstream of all these coding regions, suggesting a tightly regulated, spatial expression of specific genes.

3.8 Functional characterization of the f3626 holin

For functional analysis of the holin in an established background, *hol3626* was cloned into λΔSthf background (VUKOV et al. 2000). Recombinant λΔSthf::hol3626 was able to form normal sized plaques on *E. coli* LE392, apparently due to the ability of *hol3626* to support R-mediated lysis in *E. coli*. Thermal induction of the lytic cycle of λΔSthf::*hol3626* in lysogenized LE392 resulted in lysis kinetics relatively similar to λΔSthf::S with a slight delay in lysis onset of approximately 15 min (Fig. 8). Taken together, these results clearly demonstrated that the *hol3626* gene product function as a holin.

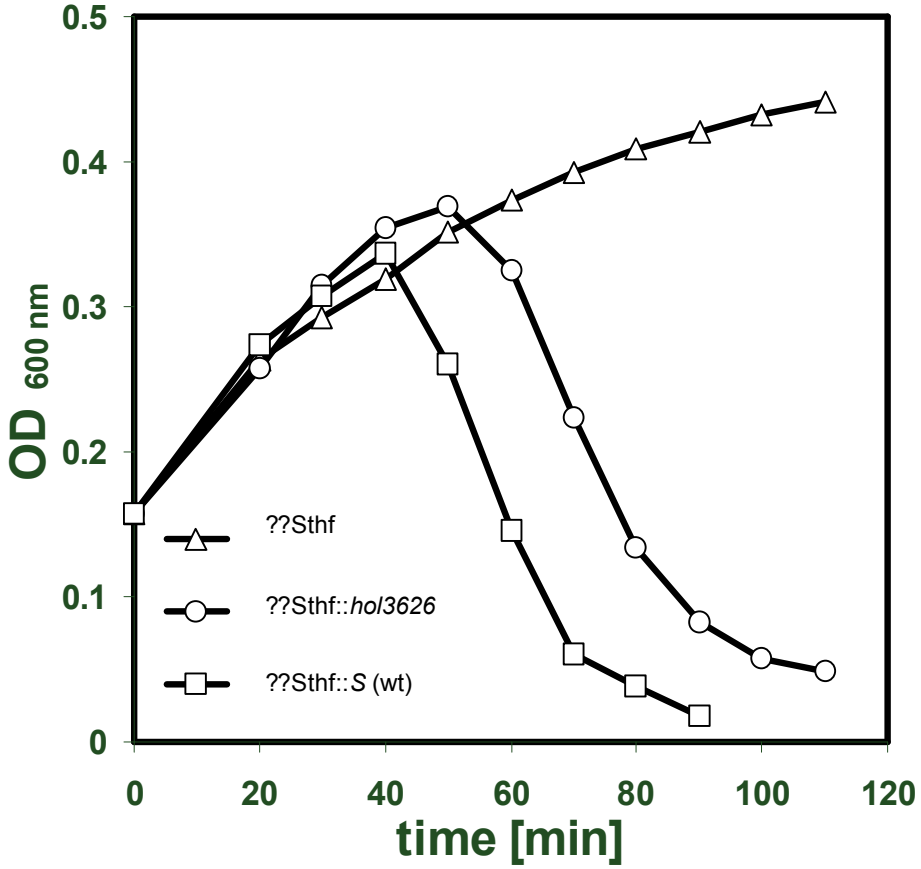


Fig. 8 Lysis profile of Hol3626

Lysis profiles observed following thermal induction of *E. coli* LE392 lysogenized with λΔSthf::hol3626 (squares), λΔSthf::S (circles) and λΔSthf (control without holin) (triangles).

3.9 Functional characterization of ply3626

3.9.1 Recombinant Ply3626 can only be synthesized in *E. coli* supplemented with rare tRNAs

First attempts to produce functional, recombinant Ply3626 were based upon a combination of JM109(DE3) and pSP72, which had repeatedly shown to be useful for this purpose (LOESSNER et al. 1995b; LOESSNER et al. 1997; LOESSNER et al. 1998). However, cloning of *ply3626* did not yield the desired enzymatically active product. In fact, SDS-Page revealed that no appreciable amount of recombinant protein was synthesized (results not shown). Next, a pQE-30 backbone was used, and the N-terminus of Ply3626 was modified by adding a hexa-histidine tag (yielding HPL3626), in order to allow concentration of enzyme from dilute cell extracts. However, still no lytic activity could be found. One reasonable explanation was that the failure to express a low G+C gene of *C. perfringens* origin in an *E. coli* background might be due to different codon usage. In fact, ply3626 contains 12 copies of the isoleucine codon ATA (4.8 % in *C. perfringens*), and 16 copies of the arginine codon AGA (2.7 % in *C. perfringens*). In contrast, these codons are rarely used in *E. coli* (0.7 % and 0.4 %, respectively) (see Codon Usage Database at http://www.kazusa.or.jp/codon). Thus, the third approach was designed such that proper tRNA genes were substituted in order to efficiently express ply3626 in E. coli. These were provided on the compatible plasmid pACYC-IRL10, encoding tRNAs for isoleucine, arginine, and leucine (ZDANOVSKY and ZDANOVSKAIA 2000). For this purpose, pACYC-IRL10 and pHPL3626 were sequentially introduced into *E. coli* JM109. However, induction with IPTG resulted in rapid formation of insoluble inclusion bodies. These contained inactive enzyme, and attempts to resolubilize the polypeptides into an active form were unsuccessful (results not shown). This phenomenon could be avoided by culturing the bacteria at room temperature (21°C) without addition of IPTG. Synthesis of the desired protein was monitored over time (Fig. 9), and revealed that approximately 20 h incubation are required for sufficient protein production. Surprisingly, under these conditions

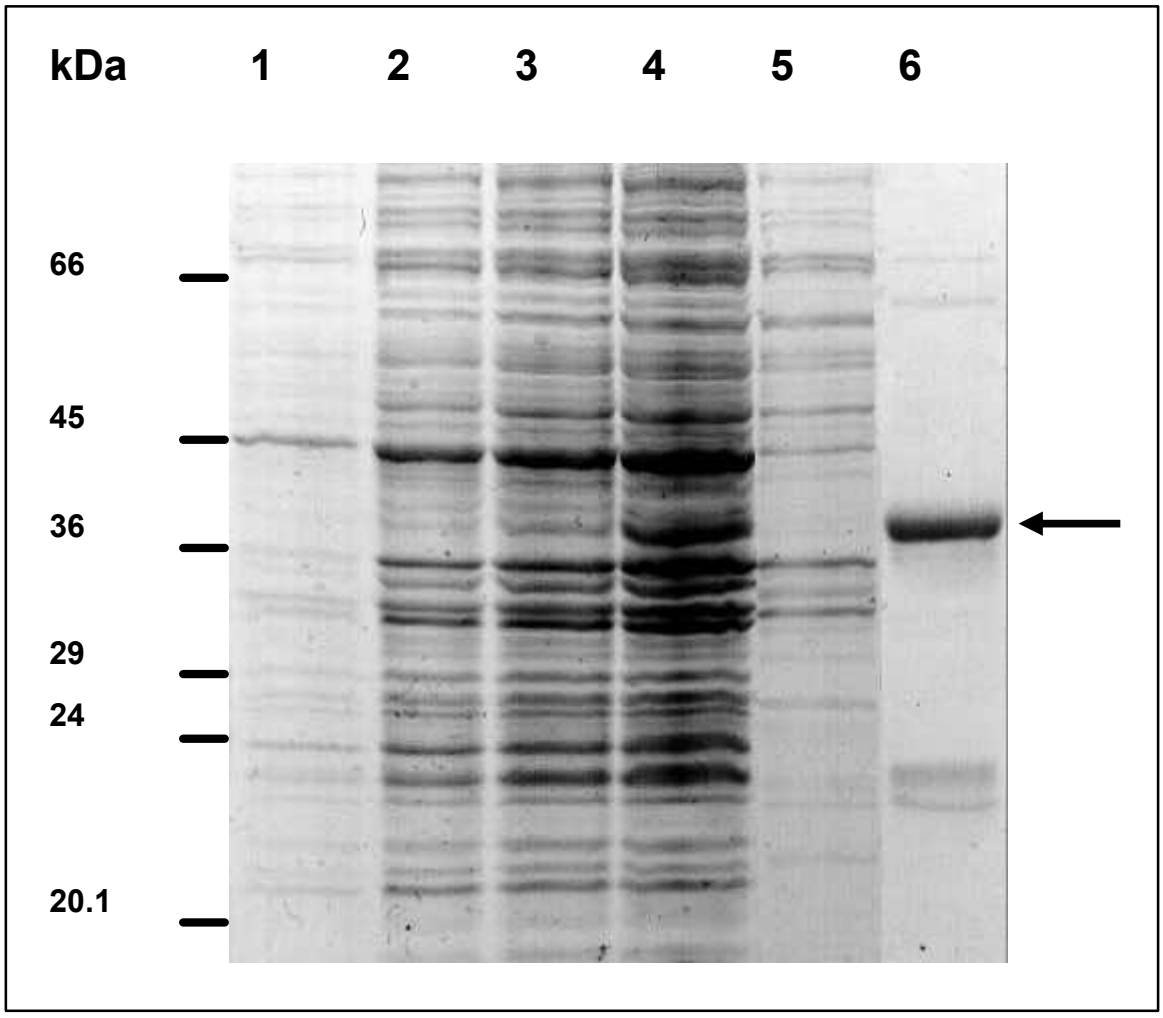


Fig. 9 Synthesis of recombinant endolysin in *E. coli* (pACYC-IRL10) (pHPL-3626), and purification of the protein (indicated by black arrow). Lanes 1-4, samples of the culture at different time point after induction (0, 10, 15, 20 hours). Lane 5, wash fraction from Ni-NTA affinity chromatography. Lane 6, HPL3626 after elution and buffer exchange

background expression from the strong T5 promoter on pQE-30 was sufficient for appreciable gene expression. Crude cell extracts were then prepared, and HPL3626 purified by Ni-NTA affinity chromatography (Fig. 9). Using this method, a total yield of approximately 1.3 mg purified enzyme per liter of culture could be obtained.

3.9.2 The endolysin lyses *C. perfringens* cell walls

For the sensitivity testing towards enzymatic lysis, bacterial cell suspensions were exposed to extracts of *E. coli* (pACYC-IRL10)(pHPL3626). The rapid

decrease in optical density of a cell suspension of *C. perfringens* NCTC 3110 (standard propagation strain for ϕ3626) is shown in Fig. 10. Further tests showed that HPL3626 was able to lyse all 48 tested strains of *C. perfringens* although some variation in the sensitivity of the different strains was observed (see Fig. 10). None of the other bacteria were affected by the enzyme. A single exception was a strain of *Clostridium fallax,* which also displayed sensitivity. Nevertheless, the lytic activity of Ply3626 seems to be generally restricted to *C. perfringens.* Identical results were obtained using purified HPL3626, again demonstrating the nature of the lytic activity. In contrast, the C-terminally truncated form HPL3626ΔC was found to be inactive.

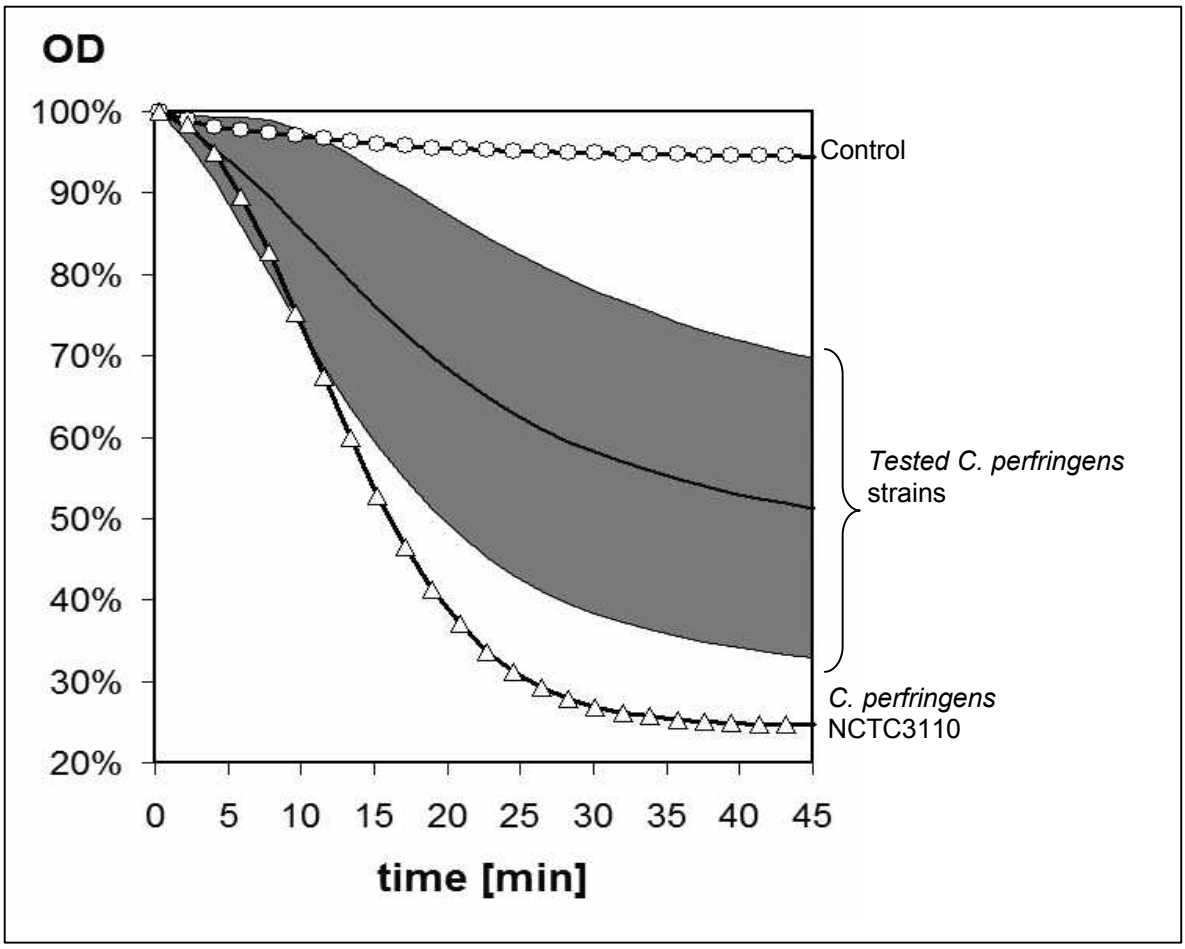


Fig. 10 Lysis Assays with HPL3626

Summary of lysis assays using the enzyme produced by *E. coli* (JM109 pACYC-IRL10) (pHPL3626). The decrease in percent optical density (OD, y-axis) over time is shown here, following addition of enzyme to standardized cell suspensions of C. perfringens. Open triangles indicate the lysis obtained with *C. perfringens* NCTC 3110 cells, the propagation host of ϕ3626. The control assay (extract from JM109 (pACYC-IRL10) (pQE-30)) is indicated by open circles. The average cell lysis response of 48 additional, different *C. perfringens* strains is represented by a black line, and the standard deviation as a grey shaded area

4 DISCUSSION

4.1 Genome organization of f3626 and relationships to other phages

In the framework of this thesis, two novel bacteriophages were isolated from lysogenic strains of the pathogen *C. perfringens*. ϕ3626, a member of Siphoviridae, is the first *C. perfringens* phage characterized on a molecular level, and it is the first bacteriophage of the genus *Clostridium* whose complete genome sequence is available. While 19 of the fifty potential polypeptides allowed functional assignments, twenty-two gene products found no match in the current databases and represent new entries. Most similarities were found with proteins of other phages infecting low G+C gram-positive bacteria, from the genera *Bacillus, Streptococcus, Staphylococcus, Lactococcus, Listeria* and *Lactobacillus*, but also, to a more limited extend, to phages from *E. coli* or *Pseudomonas*. Intergeneric relationships of ϕ3626 genomic modules are particularly pronounced in the "left-end" module, encompassing the DNA packaging and capsid building machinery. The consecutive order of the genes encoding small and large terminase, portal, prohead protease and major head protein is common to phages from low G+C gram-positives (DESIERE et al. 2001), and to lambdoid viruses in general (JUHALA et al. 2000). Here, it is disrupted by ORF 4 with unknown function.

From an evolutionary point of view, it is interesting to note that most amino acid sequence similarities were found to proteins from other Sfi21-like *cos*-site phages (BRÜSSOW and DESIERE 2001) from low G+C bacterial hosts, in particular *B. subtilis* phage ϕ105 and *S. aureus* phages ϕSLT, ϕPVL and ϕPV83. The convincing relatedness of DNA packaging and head proteins points to a vertical passage and evolution of at least this module (BRÜSSOW and DESIERE 2001). However, clostridia occupy different ecological niches and, due to their anaerobic nature, require different growth conditions than the other hosts. Nevertheless, the conserved gene order among their viruses implies a rather tight genetic

conservation. However, it should be noted that the similarities of ϕ3626 are solely based on amino acid sequences, whereas alignment of nucleotide sequences did generally not yield significant homologies. This finding suggests a lack of recent horizontal genetic exchange and therefore implicates that at least the line of viruses represented by ϕ3626 may have diverged at an earlier point. The foreseeable availability of more clostridial phage sequences in the nearer future may help to clarify this situation. Generally, however, these findings are in agreement with the hypothesis (HENDRIX et al. 1999) that most tailed phage genomes are genetic mosaics which have been build from a large common pool by genetic exchange, but were access was, and is, not uniform among the different host-dependent taxonomical groups.

4.2 Cohesive ends of the f3626 genome

The ϕ3626 chromosome features nine nucleotides 3'-protruding, single-stranded, cohesive end. Thirty-five bp upstream of this *cosN* site a nucleotide sequence G_4AG_4 can be allocated, and 37 bp downstream a run of 5 consecutive G nucleotides occur (Fig. 11). A situation reminiscent to bacteriophage BK5-T, where a C_3TC_5 sequence and a run of 7 Gs was found (MAHANIVONG et al. 2001). Also in the bacteriophages Sfi19 and Sfi21 similar findings have been described, where the sequence C_5GC_5 is located in close proximity to the *cosN*-site (LUCCHINI et al. 1999a). MAHANIVONG et al. speculated that in the background of low-G+C content organisms theses findings could be interpreted that these G+C rich

```
 -87 ATACTCAAGA TGAATTGTTT AAGCTAATTC CTAAT***GGGGA*** ***GGGG***TAGGGC AAAAAGTTTT
                                    +1
 -27 GGGCAAGAGC CAAAAGGT**CG CAGTGTC**TAT GCAGAAAAAA TTTTTTTCGG TTTATAAAGT

  34 GAAA***GGGGG***T ATATTATAGTTTGGCTAGAAA GATGATTTCA GTTAATCAAA TTATTGCAAA
```

Fig. 11 The putative *cosB* sequence. The *cosN* sequence is shown in bold. Underlined residues indicate runs of at least four identical bases. The G_4AG_4 and G_5 sequences are indicated in bold italics.

sequences might constitute the recognition site for terminase binding (*cosB*). Additionally, in BK5-T and also in λ the *cosN*-site is directly surrounded by a number of runs of four to seven identical bases of adenines and thymines that are important to DNA bending which occurs upon integration host factor binding (YEO et al. 1990; MAHANIVONG et al. 2001).

4.3 Integration site

Phage ϕ3626 was isolated from a temperate *C. perfringens* strain, where, in the lysogenic stage, the prophage is integrated into the host chromosome within the 3' terminus of *guaA.* The *attPP'* site complements the disrupted coding sequence, which enables termination of *guaA* translation at the expected site (see Fig. 3), independent of the integration status. *GuaA* encodes a GMP synthetase, a housekeeping protein responsible for *de-novo* biosynthesis of the purine nucleotide GMP. *GuaA* null mutants become guanine auxotrophs (NYGAARD 1993), and it is obvious that integration of phage must not destroy its function. Phage ϕ3626 exhibits a particularly elegant way to retain genetic function by duplication of terminal codons including a translational stop signal. This strategy is reminiscent of phages such as PSA of *L. monocytogenes* which integrate into the 3'-ends of a tRNA gene without negative effects (LAUER et al. 2002).

The putative integrase present in ϕ3626 displays similarities to *int459* of the transposon-like element CW459*tet*(M) of *C. perfringens* CW459 (ROBERTS et al. 2001). This element has been partially sequenced, and downstream of the integrase gene, part of *guaA* (*gmp*) can be allocated. These two genes are transcribed in opposite directions, just like the situation found in the ϕ3626 prophage. The coding region of *guaA* (from CW459*tet*(M)) displays high nucleotide homology (98% identical over 151 nucleotides) to the sequence upstream of *attBP'*. Interestingly, for CW459*tet*(M), the nucleotide sequence in the area between *int* and *guaA* differs from the one found in the lysogenic host (rendering the stop codon of *guaA* functional). This might be the reason why CW459*tet*(M) seems to be unable of transposition of its *tet*(M) gene (ROBERTS et al. 2001).

In the phage genome, *attPP'* is located immediately downstream of *int*. This is a common organization and forms an ideal basis for building a site-specific integration vector. No such system is available for *C. perfringens*, and its construction could be of interest for molecular and genetic research on this pathogen.

4.4 Gene products of f3626

The ORF upstream of *int* might encode the repressor and a Cro homologue, being important for the control of lysogeny. Both share helix-turn-helix (HTH) motifs known to be responsible for the necessary protein-DNA interaction. Additional evidence that supports this functional assignments is derived from using non-sequence alignment based homology parameters, such as size and charge of the proteins (CHANDRY et al. 1997). Gp24 and gp25 have biochemical properties that are quite similar to the λ proteins, i.e., λ CI is acidic (predicted pI 4.9) similar to gp24 (pI 5.1), and λ Cro is very basic (pI 10.2) as is gp25 (pI 9.2). Relative sizes of these two proteins are also analogous. These data taken together, it is tempting to speculate that gp24 and gp25 govern the genetic switch of A118.

Post-translational processing of the major head protein is frequently found in bacteriophages. During capsid maturation, ϕ3626 Cps is processed by removal of the first 114 residues. A similar processing can be observed in many other phages which lack a scaffold protein. As described for Sfi21, ϕPVL (DESIERE et al. 1999) and HK97 (CONWAY et al. 1995), ϕ3626 Cps revealed a possible coiled-coil structure in the amino terminal part of the protein which is removed. It has been assumed by Duda and co-workers (DUDA et al. 1995) that this domain might be a functional equivalent of a scaffold protein, fused to the capsid protein.

The length of the phage tail seems to be determined by a ruler-mechanism, dependent on the size of the so-called tape measure protein (KATSURA 1987). Findings in the analysis of the phage support this theory: the designated Tmp (gp13, 962 aa) is about 13 % larger than the size of the λ protein (gpH, 853 aa)

and the ϕ3626 tail is approximately 13 % longer than the λ tail (170 nm and 150 nm, respectively).

The finding of a putative translational frameshift in the overlapping pair of ORF 12/12.1 which are located between the *tsh* and the *tmp* allows the speculation that this resembles the situation found in λ.(LEVIN et al. 1993) and other bacteria (JUHALA et al. 2000). It was postulated that a programmed frameshift allows a regulation of protein expression on a post-transcriptional level and leads to the production of two proteins which share a identical sequence similarity in the N-terminus. In the λ system, the frameshift occurs with a low frequency (4 %). The recently sequenced *Listeria* phage PSA uses this mechanism for the production of its major structural proteins of the head and the tail (SATTELBERGER et al. 2002). In ϕ3626, ORF12 is followed by an overlapping ORF12.1. Just nine codons upstream of the stop codon of ORF12, a 'slippery' sequence could be identified which encodes the dipeptide -G-F- in both the zero and the -1 reading frame (Fig. 5).

Bioinformatics suggest that the product of ORF15 is a "mosaic" sequence which may have at least two functional domains. A phage lysozyme domain to which several phage endolysins showed similarities was found and additionally an endopeptidase domain was located. Both domains were located in the C-terminus of the gene product. The localization of ORF15 between *tsh* and the lysis genes suggest that gp15 might have a function comparable to the T4 tail lysozyme (NAKAGAWA et al. 1985).

4.5 A Link Between Sporulation and Lysogeny ?

The finding of an ORF possibly encoding a putative sporulation-associated sigma factor within the early genes (ORF32) suggest a possible function in 'programming' the RNA polymerase, as shown for sigma gp28 of *B. subtilis* phage SPO1 (COSTANZO and PERO 1983). ORF42 encodes a protein with convincing similarity to a *B. subtilis* sporulation-dependent transcriptional regulator (SpoIIID), which is part of the sigma factor cascade resulting in sporulation. Moreover, sequence motifs possibly reflecting sporulation- or σ^K-dependent promoters (ZHAO

and MELVILLE 1998) are present in the ϕ3626 genome. Interestingly, SpoIIID and some of the bacterial sigma factors showing homology are mother-cell specific, and are made only in the late stages of sporulation, after which the mother cell lyses. Albeit it is not clear what the precise function of homologous factors in ϕ3626 could be, an earlier study on the effect of lysogeny on sporulation of *C. perfringens* offers a possible explanation: Curing of a lysogenic strain resulted in less efficient sporulation and decreased heat resistance of spores. Re-introducing the prophage reversed the effect (STEWART and JOHNSON 1977). Unfortunately, no further details on the nature of the phage or its interaction with the host are known. However, similar findings were reported for some Bacilli, were spore-converting bacteriophages were isolated, enhancing the sporulation of infected cells (SILVER-MYSLIWIEC and BRAMUCCI 1990).

There are additional findings that seem to support this theory. Lewis et al. considered an evolutionary link between prophage induction and sporulation (LEWIS et al. 1998), based on the findings that in *Bacillus subtilis* the repressor proteins of phage 434 and SinR responsible for the sporulation inhibition, displays a high structural similarity within their DNA binding domains and share functional similarities. Loessner et al. described that the endolysins of several *Bacillus* phages displayed similarities to autolysins of *Bacillus* species (LOESSNER et al. 1997). Some autolysin are known to be involved in the process of differentiation leading to sporulation and finally lysis of the mother cell. It is important to mention that Ply3626 also displayed similarities to autolysins, such as CwlB (LytC) from *Bacillus*. Another interesting finding is that the *skin* element in *Bacillus subtilis* is a cryptic prophage. This prophage is integrated into the gene encoding the mother cell specific RNA polymerase sigma factor σ^K and must be excised by a site-specific recombinase, encoded in the *skin* element itself. This maturation results in a functional σ^K, which is needed for the transcription of late sporulation genes (STRAGIER et al. 1989; KROGH et al. 1996).

Altogether, these findings point towards a potential effect of lysogeny on sporulation. In fact, correlation of prophage carrier state and sporulation ability is

presently being investigated, and preliminary results (not included in this thesis) support the above hypothesis.

No direct evidence for an influence of ϕ3626 on the pathogenicity or virulence of *C. perfringens* is yet available. However, it is noteworthy that the expression of the Cpe enterotoxin, which is responsible for the food poisoning effect of *C. perfringens*, is also dependent on sporulation events, and transcription of *cpe* relies on the activition of sporulation-associated promoters (ZHAO and MELVILLE 1998). Further studies are needed to identify the precise role of these sequences in ϕ3626. It is intriguing to note that they are homologous to the recognition sites of sporulation-dependent sigma factors from *B. subtilis*, which are also similar to the proposed ϕ3626 sigma factor Gp32.

4.6 Lysis system of ϕ3626

New approaches are needed for the control of disease-causing organisms such as *Clostridium perfringens*, which produces an enterotoxin frequently implicated in foodborne disease and has a strong impact on the poultry industry (SONGER 1996; STEVENS 1997; MCCLANE et al. 2000; MCCLANE 2001). As a first step towards this approach, the cell wall lytic system of *C. perfringens* phage ϕ3626 has been investigated. This is the first description of lysis functions from a clostridial bacteriophage. The genes are localized at the 3'-end of the "late" genes region next to the lysogeny module, in a dual gene cluster. These findings are consistent with the situation in several other bacteriophages of the Caudovirales family (LUCCHINI et al. 1999b; WANG et al. 2000).

The function of the proposed holin and endolysin gene products could been demonstrated experimentally. Holins are proteins which are thought to first insert themselves into the cytoplasmic membrane, and subsequently oligomerize to form unspecific lesions to permit the cognate endolysin to access the murein (YOUNG and BLÄSI 1995). The ϕ3626 *hol* gene does not contain a dual start motif, which plays an important role in S-mediated lysis timing and was also found in many other holin genes (YOUNG and BLÄSI 1995; BLÄSI and YOUNG 1996). However,

expression in a λ background showed that Hol3636 can substitute for S in R-mediated cell lysis, indicating not only that the clostridial phage holin is fully functional in an *E. coli,* but also suggesting that it contains its own intrinsic timing function. Moreover, one can speculate that its mode of action might be substantially different from S. The primary structure of Hol3626 features two MSDs, which classifies it as a novel entry among the group II holins (YOUNG and BLÄSI 1995). As reasoned in 3.6.7, the "positive-inside" rule (VON HEIJNE 1992) suggests that the charged domain in between the MSDs might be located in the cytoplasm, which would mean an "outside" location of the termini (see Fig. 6 C). According to WANG et al. the most conserved feature of holins, regardless of their class, is that they possess a highly charged C-terminal domain, which with its multiple basic residues, seems likely to be disposed in the cytoplasm (WANG et al. 2000). As shown by protease accessibility in spheroblasts and inverted membrane vesicles this is at least true for the S holin from bacteriophage λ, the class I prototype holin (BLÄSI et al. 1999). Additionally, for λS, it has been shown that the C-terminal sequences define a cytoplasmic regulatory domain which affects the timing of lysis. Also Phage 21 S, the prototype class II holin seems to have the C-terminus located in the cytoplasm (RAMANCULOV and YOUNG 2001). Interestingly, the T4 holin, which seems to have only one MSD, is thought to be orientated in a way that the C-terminus is also not located in the cytoplasm (RAMANCULOV and YOUNG 2001). All these findings indicate that Hol3626, with a "C-outside, N-outside" orientation, represents an unusual and novel topology among the extremely diverse family of holin proteins which certainly warrants further investigations.

Located immediately downstream of the holin gene, *ply* encodes the ϕ3626 endolysin, which probably functions as an amidase. In contrast to the N-terminus, the Ply3626 C-terminal portion is unrelated to other lytic enzymes. It is known that bacterial peptidoglycan hydrolases, in particular those encoded by phages, display a modular domain structure. They usually harbor at least one enzymatically active domain, linked to a corresponding cell wall binding domain, which targets the enzyme to its site of action (GARCIA et al. 1990; LOESSNER et al. 2002). It has been

demonstrated that such binding domains confer recognition specificity on the hydrolases (BABA and SCHNEEWIND 1996; LOESSNER et al. 2002). Since Ply3626 exhibits a very highly evolved specificity towards *C. perfringens* cell walls, it is reasonable to assume that its unique C-terminus harbors a cell wall binding domain, which targets the enzyme to a specific receptor found exclusively on these cells. Loss of lytic activity in the truncated HPL3626ΔC enzyme (containing the isolated amidase module) also indicated the crucial role of the C-terminus in lysis. This hypothesis is also supported by the similar structure and function of two *Listeria* phage endolysins (Loessner et al. 2002). However, it is in contrast to the findings for the multidomain murein hydrolases of *Staphylococcus aureus* phages, where up to 75% of the enzymes could be deleted, sometimes with even increased lytic activity (BABA and SCHNEEWIND 1996; LOESSNER et al. 1998; LOESSNER et al. 1999). This difference of the staphylococcal endolysins to the ϕ3626 endolysin might be explained by the multiple domain architecture of these relative large enzymes, which allows the removal of some portions of the native proteins without losing the actual lytic action (NAVARRE et al. 1999). It has been shown that this is not the case for the listerial endolysins (LOESSNER et al. 2002) and it also seems to be the case for *ply3626.*

The gene encoding for Ply3626 was cloned and expressed in *E. coli.* Efficient production of native and modified (6×His-tag) endolysin was only possible under conditions compensating for the unusual codon usage by *ply3626*, due to a very low G+C content. An *E. coli* vector containing additional copies of the rarely used t-RNA genes *ileX* and *argU* under control of a constitutive promoter became recently available (ZDANOVSKY and ZDANOVSKAIA 2000). Although these observations were similar to the difficulties encountered by other researchers attempting expression of clostridial proteins in *E. coli* (GARNIER and COLE 1986; GARNIER and COLE 1988; ZDANOVSKY and ZDANOVSKAIA 2000), the next hurdle in production of these enzymes, i.e., formation of insoluble inclusion bodies after induction, was unexpected. This seems to be rare with respect to highly soluble and charged proteins such as cytoplasmic endolysin. Lowering the expression level by omitting IPTG and reducing growth temperature apparently promoted

correct protein folding and solubility, so that inclusion bodies did not form. Still, more research is needed to better understand the molecular and biochemical properties of the Ply3626 enzyme, in particular its stability and aggregation behavior.

HPL3626 exhibited stringent substrate specificity towards *C. perfringens* cells, although some variation in lysis sensitivity was observed among the different strains. Besides *C. perfringens*, only *Clostridium fallax* was found to be sensitive to the murein hydrolase. The latter organism is also regarded as a pathogen associated with cases of gas gangrene (STEVENS 1997), albeit with a lower frequency. Most of the clostridial species tested in this study belong to a phylogenetically closely related 16S rDNA cluster I (STACKEBRANDT and RAINEY 1997), encompassing more than half of the pathogenic species (STACKEBRANDT et al. 1999). Here it is interesting to note that *C. perfringens* and some strains of *C. fallax* share the same type of peptidoglycan (A3?), which features LL-diaminopimelic acid-glycine bridge to the terminal alanine of the opposite peptide chain, instead of meso-diaminopimelic acid as found in most other clostridia (SCHLEIFER and KANDLER 1972). These peptide subunits are in close proximity to the bond that is cleaved by the amidase, and, beside the putative cell wall binding domain, might also have an effect on substrate specificity.

The enzyme is very active on clostridial cells when added exogenously, because of the direct accessibility of cell wall peptidoglycan from "without". The high sensitivity to lytic enzymes under otherwise non-disruptive conditions is a particular feature of most gram-positive bacteria, and is particularly "useful" when applying native or recombinant phage endolysins for cell lysis (LOESSNER et al. 1995b; LOESSNER et al. 1997; LOESSNER et al. 1998; LOESSNER et al. 1999; LOEFFLER et al. 2001; NELSON et al. 2001). The highly specific action of Ply3626 on *C. perfringens* cells forms the basis for several potential applications of this enzyme, preferably as a recombinant product. One of the foremost goals for future developments will be the evaluation of its usefulness as a novel antimicrobial additive, with respect to the control or growth prevention of *C. perfringens* in poultry intestine. In such applications, it is of utmost importance not to affect

organisms other than the target bacteria, in order to leave the natural microbial communities undisturbed. Based on the present data, Ply3626 may fulfill this requirement. The application as a novel biopreservative in food items such as raw chicken or turkey may also be considered. The enzyme could be directly added to food or feed (LOEFFLER et al. 2001; NELSON et al. 2001), or, alternatively, be produced by recombinant bacteria. *Lactococcus lactis* has been used to produce and secret a *Listeria* phage amidase endolysin, which remained fully active and was able to kill *L. monocytogenes* cells in the medium (GAENG et al. 2000).

The use of the bacteriophage itself, in a "therapeutic" approach to control *C. perfringens,* may be also be considered as a possible way to harness the specificity of bacterial viruses (SUMMERS 2001). However, due to its narrow lytic host range, ϕ3626 would reach only approximately 22 % of the tested *C. perfringens* strains, and the lytic range of the other *C. perfringens* phage ϕ8533 is also only 8 %. These relatively narrow host ranges would require pre-therapeutic testing, or pooling of a larger number of different phages. Moreover, the likely appearance of resistance and lysogenization would limit the killing potential of native, unmodified phage. On the other hand, development of a resistance against the activity of bacteriophage endolysins has not been reported. This finding has been linked to the necessity of the phage not to be trapped inside the host cell and thereby targeting unique and highly conserved bonds in the cell wall (LOEFFLER et al. 2001).

The results of the study of the lytic system of ϕ3626 are certainly very promising. However, the potential use of Ply3626 for the prevention or biological control of the pathogen *C. perfringens* requires further studies, in particular the function of the enzyme in a food environment, and its potential effects in the intestine of infected poultry.

REFERENCES

Ackermann, H. W. (1998). "Tailed bacteriophages: the order caudovirales." Adv. Virus Res. 51: 135-201.

Ackermann, H. W. and M. S. DuBow (1987). Virus of Prokaryotes. Boca Raton, CRC Press Inc.

Adams, M. H. (1959). Methods of study of bacterial viruses. Bacteriophages. New York, Intersciences Publishers, Inc.**:** 443-457.

Aguilar, P. S., S.-I. Aizawa, P. Babitzke, A. Bolhuis and e. al. (2001). *Bacillus subtilis* and Its Closest Relatives. Washington D.C., ASM Press.

Alonso, J. C., G. Luder, A. C. Stiege, S. Chai, F. Weise and T. A. Trautner (1997). "The complete nucleotide sequence and functional organization of *Bacillus subtilis* bacteriophage SPP1." Gene 204(1-2): 201-12.

Al-Sheikhly, F. and A. Al-Saieg (1980). "Role of Coccidia in the occurrence of necrotic enteritis of chickens." Avian Dis. 24(2): 324-33.

Al-Sheikhly, F. and R. B. Truscott (1977). "The interaction of *Clostridium perfringens* and its toxins in the production of necrotic enteritis of chickens." Avian Dis. 21(2): 256-63.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25(17): 3389-402.

Baba, T. and O. Schneewind (1996). "Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*." Embo J. 15(18): 4789-97.

Bains, B. S. (1968). "Necrotic enteritis of chickens." Aust. Vet. J. 44: 40.

Bean, N. H., J. S. Goulding, C. Lao and F. J. Angulo (1996). "Surveillance for foodborne-disease outbreaks--United States, 1988-1992." Mor. Mortal. Wkly Rep. CDC Surveill. Summ. 45(5): 1-66.

Bernhardt, T. G., D. K. Struck and R. Young (2001a). "The lysis protein E of phi X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis." J. Biol. Chem. 276(9): 6093-7.

Bernhardt, T. G., I. N. Wang, D. K. Struck and R. Young (2001b). "A protein antibiotic in the phage Qbeta virion: diversity in lysis targets." Science 292(5525): 2326-9.

Bläsi, U., P. Fraisl, C. Y. Chang, N. Zhang and R. Young (1999). "The C-terminal sequence of the lambda holin constitutes a cytoplasmic regulatory domain." J. Bacteriol. 181(9): 2922-9.

Bläsi, U. and R. Young (1996). "Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis." Mol. Microbiol. 21(4): 675-82.

Bouchard, J. D. and S. Moineau (2000). "Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain." Virology 270(1): 65-75.

Brüssow, H. and F. Desiere (2001). "Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages." Mol. Microbiol. 39(2): 213-22.

Brüssow, H. and R. W. Hendrix (2002). "Phage genomics: small is beautiful." Cell 108(1): 13-6.

Bryant, A. E. and D. L. Stevens (1997). The Pathogenesis of Gas Gangrene. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 185-196.

Canard, B. and S. T. Cole (1990). "Lysogenic phages of *Clostridium perfringens*: mapping of the chromosomal attachment sites." FEMS Microbiol. Lett. 54(1-3): 323-6.

Carman, R. J., S. Perelle and M. R. Popoff (1997). Binary Toxins from *Clostridium spiroforme* and *Clostridium perfringens*. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 359-367.

Chandry, P. S., S. C. Moore, J. D. Boyce, B. E. Davidson and A. J. Hillier (1997). "Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1." Mol. Microbiol. 26(1): 49-64.

Chopin, A., A. Bolotin, A. Sorokin, S. D. Ehrlich and M. Chopin (2001). "Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations." Nucleic Acids Res. 29(3): 644-51.

Collie, R. E., J. F. Kokai-Kun and B. A. McClane (1998). "Evidence that the enterotoxin gene can be episomal in *Clostridium perfringens* isolates from non-fooborne human gastrointestinal diseases." Anaerobe 4(2): 69-79.

Collie, R. E. and B. A. McClane (1998). "Evidence that the enterotoxin gene can be episomal in *Clostridium perfringens* isolates associated with non-food-borne human gastrointestinal diseases." J. Clin. Microbiol. 36(1): 30-6.

Conway, J. F., R. L. Duda, N. Cheng, R. W. Hendrix and A. C. Steven (1995). "Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system." J. Mol. Biol. 253(1): 86-99.

Cornillot, E., B. Saint-Joanis, G. Daube, S. Katayama, P. E. Granum, B. Canard and S. T. Cole (1995). "The enterotoxin gene (cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne." Mol. Microbiol. 15(4): 639-47.

Costanzo, M. and J. Pero (1983). "Structure of a *Bacillus subtilis* bacteriophage SPO1 gene encoding RNA polymerase sigma factor." Proc. Natl. Acad. Sci. USA 80(5): 1236-40.

Czeczulin, J. R., R. E. Collie and B. A. McClane (1996). "Regulated expression of *Clostridium perfringens* enterotoxin in naturally cpe-negative type A, B, and C isolates of *C. perfringens*." Infect. Immun. 64(8): 3301-9.

Desiere, F., S. Lucchini and H. Brüssow (1998). "Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions." Virology 241(2): 345-56.

Desiere, F., S. Lucchini and H. Brüssow (1999). "Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate cos-site *Streptococcus thermophilus* bacteriophage Sfi21." Virology 260(2): 244-53.

Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson and H. Brüssow (2001). "Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T." Virology 283(2): 240-52.

Devriese, L. A., G. Daube, J. Hommez and F. Haesebrouck (1993). "In vitro susceptibility of *Clostridium perfringens* isolated from farm animals to growth-enhancing antibiotics." J. Appl. Bacteriol. 75(1): 55-7.

d'Hérelle, F. (1917). "Sur un microbe invisble antagoniste des bacilles dysentérques." C R. Acad. Sci. Ser. D. 165: 373-373.

Dower, W. J., J. F. Miller and C. W. Ragsdale (1988). "High efficiency transformation of E. coli by high voltage electroporation." Nucleic Acids Res 16(13): 6127-45.

Duda, R. L., K. Martincic and R. W. Hendrix (1995). "Genetic basis of bacteriophage HK97 prohead assembly." J. Mol. Biol. 247(4): 636-47.

Duncan, C. L., D. H. Strong and M. Sebald (1972). "Sporulation and enterotoxin production by mutants of *Clostridium perfringens*." J. Bacteriol. 110(1): 378-91.

Durbin, R., S. Eddy, A. Krogh and G. Mitchison (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge, Cambridge University Press.

Fach, P. and J. P. Guillou (1993). "Detection by in vitro amplification of the alpha-toxin (phospholipase C) gene from *Clostridium perfringens*." J. Appl. Bacteriol. 74(1): 61-6.

Ficken, M. D. and D. P. Wages (1997). Necrotic enteritis. Diseases of poultry. B. W. Calnek. Ames, IA, Iowa State University Press**:** 261-264.

Fliss, I., E. Emond, R. E. Simard and S. Pandian (1991). "A rapid and efficient method of lysis of *Listeria* and other gram-positive bacteria using mutanolysin." Biotechniques 11(4): 453, 456-7.

Gaeng, S., S. Scherer, H. Neve and M. J. Loessner (2000). "Gene cloning and expression and secretion of *Listeria monocytogenes* bacteriophage-lytic enzymes in *Lactococcus lactis*." Appl. Environ. Microbiol. 66(7): 2951-8.

Garcia, P., J. L. Garcia, E. Garcia, J. M. Sanchez-Puelles and R. Lopez (1990). "Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages." Gene 86(1): 81-8.

Garnier, T., B. Canard and S. T. Cole (1991). "Cloning, mapping, and molecular characterization of the rRNA operons of *Clostridium perfringens*." J. Bacteriol. 173(17): 5431-8.

Garnier, T. and S. T. Cole (1986). "Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene." J. Bacteriol. 168(3): 1189-96.

Garnier, T. and S. T. Cole (1988). "Complete nucleotide sequence and genetic organization of the bacteriocinogenic plasmid, pIP404, from *Clostridium perfringens*." Plasmid 19(2): 134-50.

Gazdzinski, P. and R. J. Julian (1992). "Necrotic enteritis in turkeys." Avian Dis. 36(3): 792-8.

Gui, L., C. Subramony, J. Fratkin and M. D. Hughson (2002). "Fatal enteritis necroticans (pigbel) in a diabetic adult." Mod. Pathol. 15(1): 66-70.

Hatfull, G. F. (2001). "Microbiology. The great escape." Science 292(5525): 2263-4.

Henderson, I., T. Davis, M. Elmore and N. P. Minton (1997). The Genetic Basis of Toxin Production in *Clostridium botulinum* and *Clostridium tetani*. The Clostridia, Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 261-294.

Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford and G. F. Hatfull (1999). "Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage." Proc. Natl. Acad. Sci. USA 96(5): 2192-7.

Hippe, H., J. R. Andreesen and G. Gottschalk (1991). The Genus *Clostridium* - Nonmedical. The Prokaryotes. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K.-H. Schleifer. New York, Springer-Verlag. 2**:** 1800-1866.

Hirel, P. H., M. J. Schmitter, P. Dessen, G. Fayat and S. Blanquet (1989). "Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid." Proc. Natl. Acad. Sci. USA 86(21): 8247-51.

Johnson, E. A. (1997). Extrachromosomal Virulence Determinants in the Clostridia. The Clostridia, Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 35-48.

Johnson, S. and D. N. Gerding (1997). Enterotoxemic Infections. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 117-140.

Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull and R. W. Hendrix (2000). "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages." J. Mol. Biol. 299(1): 27-51.

Kaldhusdal, M. and M. Hofshagen (1992). "Barley inclusion and avoparcin supplementation in broiler diets. 2. Clinical, pathological, and bacteriological findings in a mild form of necrotic enteritis." Poult. Sci. 71(7): 1145-53.

Kalionis, B., I. B. Dodd and J. B. Egan (1986). "Control of gene expression in the P2-related template coliphages. III. DNA sequence of the major control region of phage 186." J. Mol. Biol. 191(2): 199-209.

Kaneko, J., T. Kimura, Y. Kawakami, T. Tomita and Y. Kamio (1997). "Panton-valentine leukocidin genes in a phage-like particle isolated from mitomycin C-treated *Staphylococcus aureus* V8 (ATCC 49775)." Biosci. Biotechnol. Biochem. 61(11): 1960-2.

Kaneko, J., T. Kimura, S. Narita, T. Tomita and Y. Kamio (1998). "Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton-Valentine leukocidin genes." Gene 215(1): 57-67.

Katsura, I. (1987). "Determination of bacteriophage lambda tail length by a protein ruler." Nature 327(6117): 73-5.

Klaus, S., D. H. Krüger and J. Meyer (1992). Bakterienviren. Stuttgart, Gustav Fischer Verlag Jena.

Kobayashi, K., K. Okamura, T. Inoue, T. Sato and Y. Kobayashi (1998). Complete nucleotide sequence of *Bacillus subtilis* phage phi-105, unpublished, Accession Number AB016282.

Kodaira, K. I., M. Oki, M. Kakikawa, N. Watanabe, M. Hirakawa, K. Yamada and A. Taketo (1997). "Genome structure of the *Lactobacillus* temperate phage phi g1e: the whole genome sequence and the putative promoter/repressor system." Gene 187(1): 45-53.

Köhler, B., S. Kolbach and J. Meine (1974a). "Untersuchungen zur nekrotischen Enteritis der Hühner. 2. Mikrobiologische Aspekte." Monatsh. Veterinärmed. 29: 385-391.

Köhler, B., G. Marx, S. Kolbach and E. Böttcher (1974b). "Untersuchungen zur nekrotischen Enteritis der Hühner. 1. Diagnostik und Bekämpfung." Monatsh. Veterinärmed. 29: 380-384.

Kreft, B., K. Dalhoff and K. Sack (2000). "Darmbrand (Enteritis necroticans): Eine historische und aktuelle Übersicht." Med. Klin. 95: 435-41.

Krogh, S., M. O'Reilly, N. Nolan and K. M. Devine (1996). "The phage-like element PBSX and part of the skin element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous." Microbiology 142 (Pt 8): 2031-40.

Kropinski, A. M. (2000). "Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3." J. Bacteriol. 182(21): 6066-74.

Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, A. Danchin and et al. (1997). "The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*." Nature 390(6657): 249-56.

Kuroda, A. and J. Sekiguchi (1991). "Molecular cloning and sequencing of a major *Bacillus subtilis* autolysin gene." J. Bacteriol. 173(22): 7304-12.

Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi and K. Hiramatsu (2001). "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*." Lancet 357(9264): 1225-40.

Kyhse-Andersen, J. (1984). "Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose." J. Biochem. Biophys. Methods 10(3-4): 203-9.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J. Mol. Biol. 157(1): 105-32.

Labbé, R. G. (1989). *Clostridium perfringens*. Foodborne Bacterial Pathogens. M. P. Doyle. New York, Marcel Dekker Inc.**:** 191-234.

Labbé, R. G. and C. L. Duncan (1974). "Sporulation and enterotoxin production by *Clostridium perfringens* type A under conditions of controlled pH and temperature." Can. J. Microbiol. 20(11): 1493-501.

Ladero, V., P. Garcia, J. C. Alonso and J. E. Suarez (1999). "A2 cro, the lysogenic cycle repressor, specifically binds to the genetic switch region of *Lactobacillus casei* bacteriophage A2." Virology 262(1): 220-9.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature 227(259): 680-5.

Lauer, P., M. Y. N. Chow, M. J. Loessner, D. A. Portnoy and R. Calendar (2002). "Construction, Characterization, and Use of Two *Listeria monocytogenes* site-specific phage integration vectors." J. Bacteriol.: submitted for publication.

Lawrence, G. W. (1997). The Pathogenesis of Enteritis Necroticans. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 197-207.

Leary, S. E. C. and R. W. Titball (1997). The *Clostridium perfringens* β-toxin. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 241-250.

Levin, M. E., R. W. Hendrix and S. R. Casjens (1993). "A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein." J. Mol. Biol. 234(1): 124-39.

Lewis, R. J., J. A. Brannigan, W. A. Offen, I. Smith and A. J. Wilkinson (1998). "An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex." J. Mol. Biol. 283(5): 907-12.

Lindback, T. and A. B. Kolsto (1999). "Characterization of the mbl determinant and cloning of the *spoIIID* gene from *Bacillus cereus* ATCC 10987." Curr. Microbiol. 38(4): 224-7.

Liu, T. and E. Haggard-Ljungquist (1999). "The transcriptional switch of bacteriophage WPhi, a P2-related but heteroimmune coliphage." J. Virol. 73(12): 9816-26.

Loeffler, J. M., D. Nelson and V. A. Fischetti (2001). "Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase." Science 294(5549): 2170-2.

Loessner, M. J. and M. Busse (1990). "Bacteriophage typing of *Listeria* species." Appl. Environ. Microbiol. 56(6): 1912-8.

Loessner, M. J., S. Gaeng and S. Scherer (1999). "Evidence for a holin-like protein gene fully embedded out of frame in the endolysin gene of *Staphylococcus aureus* bacteriophage 187." J. Bacteriol. 181(15): 4452-60.

Loessner, M. J., S. Gaeng, G. Wendlinger, S. K. Maier and S. Scherer (1998). "The two-component lysis system of *Staphylococcus aureus* bacteriophage Twort: a large TTG-start holin and an associated amidase endolysin." FEMS Microbiol. Lett. 162(2): 265-74.

Loessner, M. J., R. B. Inman, P. Lauer and R. Calendar (2000). "Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution." Mol. Microbiol. 35(2): 324-40.

Loessner, M. J., K. Kramer, F. Ebel and S. Scherer (2002). "C-terminal domains of *Listeria monocytogenes* bacteriophage murein hydrolases determine specific recognition and high affinity binding to bacterial cell wall carbohydrates." Mol. Microbiol. 44: 335-349.

Loessner, M. J., I. B. Krause, T. Henle and S. Scherer (1994). "Structural proteins and DNA characteristics of 14 *Listeria* typing bacteriophages." J. Gen. Virol. 75: 701-10.

Loessner, M. J., S. K. Maier, H. Daubek-Puza, G. Wendlinger and S. Scherer (1997). "Three *Bacillus cereus* bacteriophage endolysins are unrelated but reveal high homology to cell wall hydrolases from different bacilli." J. Bacteriol. 179(9): 2845-51.

Loessner, M. J., A. Schneider and S. Scherer (1995a). "A new procedure for efficient recovery of DNA, RNA, and proteins from *Listeria* cells by rapid lysis with a recombinant bacteriophage endolysin." Appl. Environ. Microbiol. 61(3): 1150-2.

Loessner, M. J., A. Schneider and S. Scherer (1996). "Modified *Listeria* bacteriophage lysin genes (ply) allow efficient overexpression and one-step purification of biochemically active fusion proteins." Appl. Environ. Microbiol. 62(8): 3057-60.

Loessner, M. J., G. Wendlinger and S. Scherer (1995b). "Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes." Mol. Microbiol. 16(6): 1231-41.

Lovland, A. and M. Kaldhusdal (1999). "Liver lesions seen at slaughter as an indicator of necrotic enteritis in broiler flocks." FEMS Immunol. Med. Microbiol. 24(3): 345-51.

Lovland, A. and M. Kaldhusdal (2001). "Severely impaired production performance in broiler flocks with high incidence of *Clostridium perfringens*-associated hepatitis." Avian Pathology 30: 73-81.

Lucchini, S., F. Desiere and H. Brüssow (1999a). "The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison of cos-site phages Sfi19 and Sfi21." Virology 260(2): 232-43.

Lucchini, S., F. Desiere and H. Brüssow (1999b). "Similarly organized lysogeny modules in temperate Siphoviridae from low GC content gram-positive bacteria." Virology 263(2): 427-35.

Lupas, A., M. Van Dyke and J. Stock (1991). "Predicting coiled coils from protein sequences." Science 252(5010): 1162-4.

Lyristis, M., A. E. Bryant, J. Sloan, M. M. Awad, I. T. Nisbet, D. L. Stevens and J. I. Rood (1994). "Identification and molecular analysis of a locus that regulates extracellular toxin production in *Clostridium perfringens*." Mol. Microbiol. 12(5): 761-77.

Mahanivong, C., J. D. Boyce, B. E. Davidson and A. J. Hillier (2001). "Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T." Appl. Environ. Microbiol. 67(8): 3564-76.

Mahony, D. E. (1979). Bacteriocin, Bacteriophage and other Epidemiological Typing Methods for the Genus *Clostridium*. Methods in Microbiology. T. Bergan and J. R. Norris. London, Academic Press. 13**:** 1-30.

Maniloff, J., G. J. Kampo and C. C. Dascher (1994). "Sequence analysis of a unique temperature phage: mycoplasma virus L2." Gene 141(1): 1-8.

McClane, B. A. (2001). *Clostridium perfringens*. Food Microbiology: Fundamentals and Frontiers. M. P. Doyle, L. R. Beuchat and T. J. Montville. Washington, D.C., ASM Press**:** 351-372.

McClane, B. A., D. M. Lyerly, J. S. Moncrief and T. D. Wilkins (2000). Enterotoxic Clostridia: *Clostridium perfringens* Type A and *Clostridium difficile*. Gram-Positive Pathogens. V. A. Fischetti, R. P. Novick, F. J. J., A. D. Portnoy and J. I. Rood. Washington, American Society for Microbiology**:** 551-562.

Nairn, M. E. and V. W. Bamford (1967). "Necrotic enteritis of broiler chickens in western Australia." Aust. Vet. J. 43(2): 49-54.

Nakagawa, H., F. Arisaka and S. Ishii (1985). "Isolation and characterization of the bacteriophage T4 tail-associated lysozyme." J. Virol. 54(2): 460-6.

Narita, S., J. Kaneko, J. Chiba, Y. Piemont, S. Jarraud, J. Etienne and Y. Kamio (2001). "Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, phiSLT." Gene 268(1-2): 195-206.

Navarre, W. W., H. Ton-That, K. F. Faull and O. Schneewind (1999). "Multiple enzymatic activities of the murein hydrolase from staphylococcal phage phi11. Identification of a D-alanyl-glycine endopeptidase activity." J. Biol. Chem. 274(22): 15847-56.

Nelson, D., L. Loomis and V. A. Fischetti (2001). "Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme." Proc. Natl. Acad. Sci. USA 98(7): 4107-12.

Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin and D. R. Smith (2001). "Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*." J. Bacteriol. 183(16): 4823-38.

Nygaard, P. (1993). Purine and Pyrimidine Salvage Pathways. *Bacillus subtilis* and Other Gram-Positive Bacteria. A. I. Sonenshein, J. A. Hoch and R. Losick. Washington, D.C., American Society for Microbiology.

Ochman, H., A. S. Gerber and D. L. Hartl (1988). "Genetic applications of an inverse polymerase chain reaction." Genetics 120(3): 621-3.

Ogata, S. and M. Hongo (1979). "Bacteriophages of the Genus *Clostridium*." Adv. Appl. Microbiol. 25: 241-273.

Olsen, S. J., L. C. MacKinnon, J. S. Goulding, N. H. Bean and L. Slutsker (2000). "Surveillance for foodborne-disease outbreaks--United States, 1993-1997." Mor. Mortal. Wkly Rep. CDC Surveill. Summ. 49(1): 1-62.

Parish, W. E. (1961). "Necrotic enteritis in the fowl (*Gallus gallus domesticus*) I. Histopathologiy of the diseas and isolation of a strain of *Clostridium welchii*." J. Comp. Path. 71: 394-404.

Payne, D. and P. Oyston (1997). The *Clostridium perfringens* ε-Toxin. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 439-447.

Pedersen, M., S. Ostergaard, J. Bresciani and F. K. Vogensen (2000). "Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly." Virology 276(2): 315-28.

Petit, L., M. Gibert and M. R. Popoff (1999). "*Clostridium perfringens*: toxinotype and genotype." Trends Microbiol. 7(3): 104-10.

Petrillo, T. M., C. M. Beck-Sague, J. G. Songer, C. Abramowsky, J. D. Fortenberry, L. Meacham, A. G. Dean, H. Lee, D. M. Bueschel and S. R. Nesheim (2000). "Enteritis necroticans (pigbel) in a diabetic child." N Engl J Med 342(17): 1250-3.

Ptashne, M. (1992). A Genetic Switch Phage Lambda and Higher Organisms. Cambridge, Cell Press & Blackwell Sci. Publ.

Ramanculov, E. and R. Young (2001). "Genetic analysis of the T4 holin: timing and topology." Gene 265(1-2): 25-36.

Ripp, S. and R. V. Miller (1998). "Dynamics of the pseudolysogenic response in slowly growing cells of *Pseudomonas aeruginosa*." Microbiology 144 (Pt 8): 2225-32.

Roberts, A. P., P. A. Johanesen, D. Lyras, P. Mullany and J. I. Rood (2001). "Comparison of Tn5397 from *Clostridium difficile*, Tn916 from *Enterococcus faecalis* and the CW459tet(M) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules." Microbiology 147(Pt 5): 1243-51.

Rood, J. I. (1998). "Virulence genes of *Clostridium perfringens*." Annu. Rev. Microbiol. 52: 333-60.

Rood, J. I. and S. T. Cole (1991). "Molecular genetics and pathogenesis of *Clostridium perfringens*." Microbiol. Rev. 55(4): 621-48.

Sambrook, J., T. Maniatis and E. F. Fritsch (1989). Molecular Cloning: a Laboratory Manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Sattelberger, E., M. Zimmer, R. Calendar, R. B. Inman, S. Scherer and M. J. Loessner (2002). Genome and proteome of *Listeria monocytogenes* bacteriophage PSA: A case for an unusual form of programmed translational frameshifting during structural protein synthesis. submitted for publication.

Schleifer, K. H. and O. Kandler (1972). "Peptidoglycan types of bacterial cell walls and their taxonomic implications." Bacteriol. Rev. 36(4): 407-77.

Shatursky, O., R. Bayles, M. Rogers, B. H. Jost, J. G. Songer and R. K. Tweten (2000). "*Clostridium perfringens* beta-toxin forms potential-dependent, cation-selective channels in lipid bilayers." Infect. Immun. 68(10): 5546-51.

Shimizu, T., W. Ba-Thein, M. Tamaki and H. Hayashi (1994). "The *virR* gene, a member of a class of two-component response regulators, regulates the production of perfringolysin O, collagenase, and hemagglutinin in *Clostridium perfringens*." J. Bacteriol. 176(6): 1616-23.

Silver-Mysliwiec, T. H. and M. G. Bramucci (1990). "Bacteriophage-enhanced sporulation: comparison of spore-converting bacteriophages PMB12 and SP10." J. Bacteriol. 172(4): 1948-53.

Smith, H. W. (1959). "The Bacteriophages of *Clostridium perfringens*." J. Gen. Microbiol. 21: 622-630.

Smith, L. D. S. (1991). The Genus *Clostridium* - Medical. The Prokaryotes: a handbook on the biology of bacteria: ecophysiology, isolation, identification, applications. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K.-H. Schleifer. New York, Springer-Verlag.

Smith, M. C., R. N. Burns, S. E. Wilson and M. A. Gregory (1999). "The complete genome sequence of the *Streptomyces* temperate phage phi-C31: evolutionary relationships to other viruses." Nucleic Acids Res. 27(10): 2145-55.

Songer, J. G. (1996). "Clostridial enteric diseases of domestic animals." Clin. Microbiol. Rev. 9(2): 216-34.

Songer, J. G. (1997). Clostridial Diseases of Animals. The Clostridia: Molecular Biology and Pahtogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, Academic Press**:** 153-182.

Sonnhammer, E. L., G. von Heijne and A. Krogh (1998). "A hidden Markov model for predicting transmembrane helices in protein sequences." Proc. Int. Conf. Intell. Syst. Mol. Biol. 6: 175-82.

Stackebrandt, E., I. Kramer, J. Swiderski and H. Hippe (1999). "Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*." FEMS Immunol. Med. Microbiol. 24(3): 253-8.

Stackebrandt, E. and F. A. Rainey (1997). Phylogenetic Relationships. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, CA, USA, Academic Press, Inc.**:** 3-19.

Stevens, D. L. (1997). Necrotizing Clostridial Soft Tissue Infections. The Clostridia: Molecular Biology and Pathogenesis. J. I. Rood, B. A. McClane, J. G. Songer and R. W. Titball. San Diego, CA, USA, Academic Press**:** 141-151.

Stewart, A. W. and M. G. Johnson (1977). "Increased numbers of heat-resistant spores produced by two strains of *Clostridium perfringens* bearing temperate phage s9." J. Gen. Microbiol. 103(1): 45-50.

Stirewalt, V. L. and S. B. Melville (2000). Cloning and sequence analysis of the *sigK* gene of *Clostridium perfringens*.

Stragier, P., B. Kunkel, L. Kroos and R. Losick (1989). "Chromosomal rearrangement generating a composite gene for a developmental transcription factor." Science 243(4890): 507-12.

Summers, W. C. (2001). "Bacteriophage therapy." Annu. Rev. Microbiol. 55: 437-51.

Takami, H., K. Nakasone, Y. Takaki, G. Maeno, R. Sasaki, N. Masui, F. Fuji, C. Hirama, Y. Nakamura, N. Ogasawara, S. Kuhara and K. Horikoshi (2000). "Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*." Nucleic Acids Res. 28(21): 4317-31.

Twort, F. W. (1915). "An investigation on the nature of ultra-microscopic viruses." Lancet 2: 1241-1243.

van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Ruiters, G. Venema and A. Nauta (1996). "Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t." Mol. Microbiol. 19(6): 1343-55.

Vander Byl, C. and A. M. Kropinski (2000). "Sequence of the genome of *Salmonella* bacteriophage P22." J. Bacteriol. 182(22): 6472-81.

von Heijne, G. (1992). "Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule." J. Mol. Biol. 225(2): 487-94.

Vukov, N., S. Scherer, E. Hibbert and M. J. Loessner (2000). "Functional analysis of heterologous holin proteins in a λ∆S genetic background." FEMS Microbiol. Lett. 184(2): 179-86.

Wang, I. N., D. L. Smith and R. Young (2000). "Holins: the protein clocks of bacteriophage infections." Annu. Rev. Microbiol. 54: 799-825.

Ward, L. J., T. P. Beresford, M. W. Lubbers, B. D. Jarvis and A. W. Jarvis (1993). "Sequence analysis of the lysin gene region of the prolate lactococcal bacteriophage c2." Can. J. Microbiol. 39(8): 767-74.

Wegener, H. C., F. M. Aarestrup, L. B. Jensen, A. M. Hammerum and F. Bager (1999). "Use of antimicrobial growth promoters in food animals and *Enterococcus faecium* resistance to therapeutic antimicrobial drugs in Europe." Emerg. Infect. Dis. 5(3): 329-35.

WHO (1997). WHO Meeting: The Medical Impat of Antimicrobial Use in Food Animals, Berlin, Germany.

Witte, W. (1998). "Medical consequences of antibiotic use in agriculture." Science 279(5353): 996-7.

Wobeser, G. and D. J. Rainnie (1987). "Epizootic necrotic enteritis in wild geese." J. Wildl. Dis. 23(3): 376-85.

Yamaguchi, T., T. Hayashi, H. Takami, K. Nakasone, M. Ohnishi, K. Nakayama, S. Yamada, H. Komatsuzawa and M. Sugai (2000). "Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*." Mol. Microbiol. 38(4): 694-705.

Yan, W. K. (1989). "Use of host modified bacteriophages in development of a phage typing scheme for *Clostridium perfringens*." Med. Lab. Sci. 46(3): 186-93.

Yeo, A., L. D. Kosturka and M. Feiss (1990). "Structure of the bacteriophage lambda cohesive end site: bent DNA on both sides of the site, cosN, at which terminase introduces nicks during chromosome maturation." Virology 174(1): 329-34.

Yoshisue, H., K. Ihara, T. Nishimoto, H. Sakai and T. Komano (1995). "Cloning and characterization of a *Bacillus thuringiensis* homolog of the *spoIIID* gene from *Bacillus subtilis*." Gene 154(1): 23-9.

Young, I., I. Wang and W. D. Roof (2000). "Phages will out: strategies of host cell lysis." Trends Microbiol 8(3): 120-8.

Young, R. (1992). "Bacteriophage lysis: mechanism and regulation." Microbiol. Rev. 56(3): 430-81.

Young, R. and U. Bläsi (1995). "Holins: form and function in bacteriophage lysis." FEMS Microbiol. Rev. 17(1-2): 191-205.

Zdanovsky, A. G. and M. V. Zdanovskaia (2000). "Simple and efficient method for heterologous expression of clostridial proteins." Appl. Environ. Microbiol. 66(8): 3166-73.

Zhao, Y. and S. B. Melville (1998). "Identification and characterization of sporulation-dependent promoters upstream of the enterotoxin gene (*cpe*) of *Clostridium perfringens*." J. Bacteriol. 180(1): 136-42.

Zink, R. and M. J. Loessner (1992). "Classification of virulent and temperate bacteriophages of *Listeria* spp. on the basis of morphology and protein analysis." Appl. Environ. Microbiol. 58(1): 296-302.

APPENDIX

Following Genbank entries can also be retrieved from the National Center for Biotechnology Information (NCBI) under:

AY082069

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=nucleotide&list_uids=19908292&dopt=GenBank

AY082070

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=nucleotide&list_uids=19908294&dopt=GenBank

Genbank entry AY082069

```
LOCUS       AY082069                834 bp    DNA     linear   BCT 02-APR-2002
DEFINITION  Clostridium perfringens GMP synthetase (guaA) gene, partial cds.
ACCESSION   AY082069
VERSION     AY082069.1  GI:19908292
KEYWORDS    .
SOURCE      Clostridium perfringens.
  ORGANISM  Clostridium perfringens
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
            Clostridiales; Clostridiaceae; Clostridium.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   Zimmer,M., Scherer,S. and Loessner,M.J.
  TITLE     Complete genome sequence, molecular analysis and integration
            properties of Clostridium perfringens bacteriophage phi3626
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 834)
  AUTHORS   Zimmer,M. and Loessner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Institute of Microbiology, FML
            Weihenstephan, Technical University of Munich, Weihenstephaner Berg
            3, Freising 85350, Germany
FEATURES             Location/Qualifiers
     source          1..834
                     /organism="Clostridium perfringens"
                     /strain="ATCC 3626"
                     /db_xref="ATCC:3626"
                     /db_xref="taxon:1502"
     gene            <1..663
                     /gene="guaA"
     CDS             <1..663
                     /gene="guaA"
                     /note="GuaA"
                     /codon_start=1
                     /transl_table=11
                     /product="GMP synthetase"
                     /protein_id="AAL96770.1"
                     /db_xref="GI:19908293"
                     /translation="DPETKRKIIGEEFIRVFEEEAGKLGDIKFLVQGTIYPDVVESGT
                     DTSAVIKSHHNVGGLPEDMEFSLIEPLRELFKDEVRAVGEELGIPHHLVWRQPFPGPG
                     LAIRVLGEVTEDKLEVVREADAIFREEIALAGLESEIWQYFAVLPNIQSVGVMGDERT
```

```
                     YCHTVGLRAVTSSDGMTSNWAHIPYEVIDKVSRRIVNEVKGVNRIVYDVTSKPPATIE
                     WE"
     misc feature    652..663
                     /gene="guaA"
                     /note="bacteriophage phi3626 attachment site"
BASE COUNT      326 a    107 c    166 g    235 t
ORIGIN
        1 gatccagaaa ctaagagaaa gataatcgga gaagaattca taagagtatt tgaagaagaa
       61 gctggaaaat taggagatat taaattctta gttcaaggaa caatttatcc agacgttgtt
      121 gaaagtggaa cagatacttc agcagtaatt aagagtcacc ataacgttgg aggacttcca
      181 gaagatatgg aattctcatt aatagagcca ttaagagaat tattcaaaga tgaagtaaga
      241 gctgtaggag aagaattagg aattcctcac cacttagtat ggagacagcc attcccagga
      301 ccaggattag ctataagagt tctaggagaa gtaactgaag ataaattaga agttgttaga
      361 gaagctgatg ctatatttag agaagaaata gctttagcag gtcttgaaag tgaaatatgg
      421 caatacttcg cagtattacc taacattcaa tcagttggcg ttatgggaga cgaaagaact
      481 tattgccaca ctgtaggatt aagagcagtt acatcatcag acggaatgac ttcaaactgg
      541 gctcatatac catatgaagt tatagataaa gtatcaagaa gaattgtaaa tgaagttaag
      601 ggagttaaca gaattgttta cgatgtaact tcaaaaccac cagcaactat cgagtgggaa
      661 taatataaat aattgattca gataggtatt aatcctatct gaattttaat ataatttatc
      721 tggttaaaaa ttcttaatag gtctaaatca gataaaattt aagtatccaa actaatcaaa
      781 agatatataa atttttatta cggataaata ggcagtcatt ttaatatgat tgtc
//
```

Genbank Enty AY082070:

```
LOCUS       AY082070               33507 bp    DNA     linear   PHG 02-APR-2002
DEFINITION  Bacteriophage phi3626, complete genome.
ACCESSION   AY082070
VERSION     AY082070.1  GI:19908294
KEYWORDS    .
SOURCE      bacteriophage phi3626.
  ORGANISM  bacteriophage phi3626
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
REFERENCE   1  (bases 1 to 33507)
  AUTHORS   Zimmer,M., Scherer,S. and Loessner,M.J.
  TITLE     Bacteriophage phi3626 complete genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 33507)
  AUTHORS   Zimmer,M., Scherer,S. and Loessner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Institute of Microbiology, FML
            Weihenstephan, Technical University of Munich, Weihenstephaner Berg
            3, Freising 85350, Germany
FEATURES             Location/Qualifiers
     source          1..33507
                     /organism="bacteriophage phi3626"
                     /virion
                     /specific_host="Clostridium perfringens"
                     /db_xref="taxon:190478"
     gene            54..557
                     /gene="terS"
     CDS             54..557
                     /gene="terS"
                     /note="ORF1; TerS"
                     /codon_start=1
                     /transl_table=11
                     /product="putative terminase small subunit"
                     /protein_id="AAL96771.1"
                     /db_xref="GI:19908295"
                     /translation="MARKMISVNQIIANGNKSHLPNEEIEKRKEQEEKLKKLPRDKIR
                     PPTWLSKEGKSIFKKIVKELEVVDILANIDNYNLAVLANAIEQYIECTRKLNCDELTI
                     THINKRGFETTQKNPLISIQIQYADVIKKLGSEFGLSPAARLKIIQEASDISDDEKEL
                     NKDFEDV"
     gene            550..2295
                     /gene="terL"
     CDS             550..2295
                     /gene="terL"
                     /function="involved in DNA packaging"
```

```
                     /function="cleaves viral DNA at cos-site"
                     /note="ORF2; TerL"
                     /codon_start=1
                     /transl_table=11
                     /product="putative terminase large subunit"
                     /protein_id="AAL96772.1"
                     /db_xref="GI:19908296"
                     /translation="MFNTVLEELIDYSNSILDGKINACKRHKQACKRFLNDLERMESD
                     DFEYYWDEEEAQRIVKWYGHCRHSKGVLEGKPIILNSWAKFVVCNIEAWKHKDTNYRR
                     FRFAFIQVARKNQKSQLEAGMAGYEIGAKGYNAAEVYTLGVERDQAKIVFEEWELMAS
                     KPLKKRFKFTQKVIKHKKSNSFIKHLSKKAGKTGDGKNPQMAVIDEYHAHPNSKMYDV
                     MKSGMMARTEPLLVIITTAGEDYEETACYYEYLDCCSILDGTFENEKYFVMICELEDE
                     DDPFDEKAWIKANPVLCTYEEGIESMRQNANLAKNTSNEEKRIEFLTKNCNIYVAAGE
                     KKYVDVEFWKACKENITLEKFRGHDCYIGMDLSKSGDLTSIAFEFPYLDEGIRKYALF
                     GKSFIPAEVVKEKMKTDNVPYNIWSSKEKGWLIKTEANEGQIVDLWAILNTIESIVKE
                     YELNVIEVSYDPHGAAMLVSELERRDYNCVECGQSCAKLNEATVNFRDLMKIKQIVHD
                     ENNLMTWCVQNAEKDTNSFGEIKISKKSRFKRIDPLASCIFAHNRAMTYWNRENLDVS
                     EFADEDFLKKLWRKK"
CDS                  2296..3564
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="putative portal protein"
                     /protein_id="AAL96773.1"
                     /db_xref="GI:19908297"
                     /translation="MGFLRGLFNKKNNNDEKRSNYDEDIGIDISDSNFWEKFGIKLNF
                     SVRGKRALKENTVYVCTKIRAESIGKLSLKIYKDKEEYKEHELYYLLRYKPNPLMSSI
                     NFWKCLETQRTLKGNAYAYIERDRKGKIIGLYPINSDNVTKIIDDDNFLSSLSKVWYV
                     VTDKNGKEHKLLPDEMLHFIGDITLDGLIGIKPLDYLRCTIENGRATQEFINKFFKNG
                     LSIKGIVQYVGDLDEKAKKIFKKEFESMSNGLENAHSISLLPFGYQFQPISLSMADAQ
                     FLENSKLTKRELAATFGMKSYHLNDLERATFNNLTEQQKDFYVTTLQSSLTVYEQEIQ
                     DKLFSQYETLQDVKAEFNVDTILRSDIKTRYEAYRIGIQGGFIEANEARRRENLPPVE
                     GGDRLLVNGNMIPIEMAGEQYKKGGEKGGK"
CDS                  3554..3715
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp4 protein"
                     /protein_id="AAL96774.1"
                     /db_xref="GI:19908298"
                     /translation="MENKYEDYKNEIKEGKRTLNQTRKELGLPPINGGNTEFINREIA
                     FCRKIQERR"
CDS                  3720..4325
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="putative prohead protease"
                     /protein_id="AAL96775.1"
                     /db_xref="GI:19908299"
                     /translation="MEHKEVKREVRHLISNIEVRATGDENTKTITGYASKYNVESQIL
                     RDWWGDKFVEVVAEGAFDNSLRSNTIKALYNHNTDNVLGSTKSGTLRLESDSIGLRFE
                     IDLPNTTVANDLYESVKRGDVDGTSFGFKVLDDKWSKVEIDGEEIMKRTLLEIDLYEI
                     SPTPFPAYEDTEVDCRSLEKIKFNENKKEEKRSSLLELMYC"
gene                 4366..5631
                     /gene="cps"
CDS                  4366..5631
                     /gene="cps"
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="major capsid protein"
                     /protein_id="AAL96776.1"
                     /db_xref="GI:19908300"
                     /translation="MNLFERLKELRAKKKELEEKRCGIVEEIRSLAKEKKEEEARSKA
                     LEREKIEARMEIIEEEIESVMTAIDEERKNTNFTGGRVIINGDSKEEKRSLQLSAMSK
                     TIRGIQLSEEERDIMSSTNNGAVIPQEFVNEFEKLKEGYPSLKEHCHVIPVNRNAGKM
                     PVRAGASVDKLANLAKDTELVKAMLKTQPMAYDIDDYGLLAPIDNSLLEDSEINFLEF
```

```
                VNEEFAEFAVNTENAEIVKQAKAVLAEETINDYAGLVKTINSLVPNARKRAIIVTNSD
                GRAYLDGLMDKQGRPLLKELSDGGDLVFKGRPVIELEESIFDVGDETKFIVSDFKTLI
                KFMDRKQYLIDQSKEAGYTKNETIARIIERFDVNSPLDKSSDAEKIRKFGVIVKLQEV
                LKSSPRSGKNKNESKEEIKEEGEATQQNE"
mat_peptide     4708..5628
                /gene="cps"
                /evidence=experimental
                /product="major capsid protein"
CDS             5641..5919
                /note="ORF7"
                /codon_start=1
                /transl_table=11
                /product="Gp7 protein"
                /protein_id="AAL96777.1"
                /db_xref="GI:19908301"
                /translation="MDLRELKEYLRIDFEEDDILLRSLLLAAEEYLYNAGIKRDYKKS
                LYSLAIKILVKHWYDNRDCVVAGNVNNKLEYSLNAILTQLRYCGDDNG"
CDS             5912..6235
                /note="ORF8"
                /codon_start=1
                /transl_table=11
                /product="Gp8 protein"
                /protein_id="AAL96778.1"
                /db_xref="GI:19908302"
                /translation="MARYERISIKKLEEKNIKGRRQEECLIPFYDCWAEILDLYGQEL
                YGALQMKLENTIIFKIRYCKKVEELRNKENFIVEWQGRKYEIYYPDFLGYNKQFVKLK
                CKEVL"
CDS             6236..6685
                /note="ORF9"
                /codon_start=1
                /transl_table=11
                /product="Gp9 protein"
                /protein_id="AAL96779.1"
                /db_xref="GI:19908303"
                /translation="MSDGWEIKFEGLDDLIKTFEQLGTEKENEDVEKSILKECGDLAK
                KTVAPLIHISDDNSKSGRKGSRPPGHAANNIPEPKIRKKKGNLQCVVGWEKSDNTPFY
                YMKMEEWGTSERPPHHAFGKTNKILKRVYDNIAQKKYDNFVKEKLGD"
CDS             6687..7037
                /note="ORF10"
                /codon_start=1
                /transl_table=11
                /product="Gp10 protein"
                /protein_id="AAL96780.1"
                /db_xref="GI:19908304"
                /translation="MEDFDIIALVYECLECLNVPVIEGWYDEELNKTHITVHEYLEQD
                ESFEDDEAREEEHNIQIDVWSKDSLEAFKLKKAIKKLLKKNNFYFDSSEDFYETKTRI
                YHKGLRFSYISEIS"
gene            7054..7650
                /gene="tsh"
CDS             7054..7650
                /gene="tsh"
                /note="orf11; intitiator methionine is removed by
                post-translational processing"
                /codon_start=1
                /transl_table=11
                /product="major tail protein"
                /protein_id="AAL96781.1"
                /db_xref="GI:19908305"
                /translation="MPEVVNTRRCGCKDIYIAIVNKNTATEYATATPIKLGRALNAKV
                TTKKTVESTESDDTIEETFESFDSIEIEFDVNKLSPEQKAILRGATYEDGFLINNVDD
                QAKEVAIGWRARQTNGKYEFVWYYCGKFNSGWDDQYETNGKKPKTQTDKMKGTFYGRQ
                KDGNTSIEVDETFLLESHNDAKSSITNWFSKVQEKEAA"
CDS             7665..7979
                /note="ORF12"
                /codon_start=1
                /transl_table=11
                /product="Gp12 protein"
```

```
                     /protein_id="AAL96782.1"
                     /db_xref="GI:19908306"
                     /translation="MKITVKGKQYDSGKMLFDKYDVYTKARNKLAEKELFNSDDINLM
                     IDTLVFIFDNKFTAKDIKSEMDISDIIFNFSSVDIEIAGKSNDKAFKASAGFMKGKKF
                     KR"
     CDS             8515..11400
                     /function="determines length of phage tail during particle
                     morphogenesis"
                     /note="ORF13"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tape-measure protein"
                     /protein_id="AAL96783.1"
                     /db_xref="GI:19908307"
                     /translation="MRVIGANVKIGANTSSFQKQMKEMAQELKRVGSSYNLANTQAKL
                     FGSRTDLLKSQQSELTAKMKIQNNMIKTQGDYLTKLNSDLDKQKSKQSELAKKISETS
                     DKYKKSVQETGKNSEESKKLKEELKQLEAEYTKNNKAIDTNVIKISNAEIKLNNSKKA
                     LEENKKALENINKELDKSKIDKFSEGIGKAGEKAGNLSNKMKPASVAMIGLGSAATFA
                     NIKFEDSLANINTLLDDQSHLESYKNKIRELSDETGMSLETVSEGMYTAISSIGDGGK
                     ETEKIFETMAKSAKAGGAEVQDAVSLISAGMKGYNQVNDETAKKISDLAFQTAKLGVT
                     TFPEMAKSMQQLFPLANNLELSLEELFTDMATLTGVTGNTAEVSTQLKAVFSNLIKPT
                     TDMQKLMQKYGFQNGQAMLKSEGLIGTLKILQKETGGQSDKMGKLFSSTEGLTALTAL
                     TGSQFDTLVDKSKKMKESIGTTDSALTKINSTLGNDLRTSLNLVRNSFVDFGEILAPF
                     ISIGANTLSTIITMINGLSEGQKKLVVGFGASFIGINLLLGGFSKLSKGIQSNIKFTK
                     DMIKVTKEGVSKLKDFGKAAREGTNLIGKFGKGIVTTAKSVASFTLELIKSAGKGLAS
                     LGKGVLNCIKSMGSFTLELIKNTAIGIKNGAIWVGNMAKMLAYKAAQVTVTVATKAMT
                     VAQKALNLAMAMNPITLIIGLLVALGGTFILLYNKCEWFRNGVNNVWNSITSIFTKFD
                     NFLTGIFTRDWTQSFGLLGVPLNAFLGTVGSIWEGVKGVFNGVIRFFKGVFTGNWREV
                     FQGLSDIVSSIFGAIGGIIKAPINAAISGINWVIRQINKISFDVPSWIPGIGGSHFGI
                     NLPTIPALAEGGIVTQATMALVGEGKEHEAVIPLSKLDKLVTSSVKKVLENYSNKKDE
                     GDFKSEPKVFQIIVPIDGRVLAEIIIDEHGNVLTKKVKDRRVVRGKI"
     CDS             11401..12102
                     /note="ORF14"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp14 protein"
                     /protein_id="AAL96784.1"
                     /db_xref="GI:19908308"
                     /translation="MIDLIFNNKSLRHDFKVSIVKRPPTMFSIKNKRKVEVLGGDGDL
                     YLELGGQQDITIPVECNFICSPNEIRDRFRLLKSWLNNVEDDKLIFTDDHNWFYKVVD
                     INISSMDVIKRRKGKFTINFICRGYHFKLDGDEFIDIFNETILFNEYDLAKPLILLRG
                     NGTIEVDINSNKFEVEVNEFVYIDSELEVAYKYEEDSYNLDIGDYPVLIHGENLISYK
                     GNLDSFEIKCRWRCV"
     CDS             12102..15053
                     /note="ORF15"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp15 protein"
                     /protein_id="AAL96785.1"
                     /db_xref="GI:19908309"
                     /translation="MQYYKSNNTNFEKNGDITLQLLSGVLRLELNSGLNEIEFEIPYD
                     KQGRWKKVEEWGVVKQKVFYTKNKQLYRIYEKNKGMFGLRLKARHIFFDLVKHTILDK
                     RAVNLNGQRALERILFETKFKAHSDIDINNTAYFVNLNGIEAINGDNENSFRNRWGGE
                     ILNDNFDVYINKKIGRDNGVKIRFGNNLEDINFKIDRDPIITRAYPKAYDGIMLPEKF
                     IDSEHINKYPIICEGFVDVSDLKLKDSENPEDEGFETKEELYEAMRRRIVDLYNNGLD
                     RPKVSGNIKIVNLENTNKYKNFKNLKHIGIGDTITINLTEIGIDAKTRFISTEWDLVT
                     EEHNNMEFGEVEVNYFDKQEIVNGRLDNILNQNGSVKSSELEGIINAMKAPLVAQKDI
                     AKNLDCVAWKCEVNDPNDPNFGYLGAGTKGLMLANSKTPDKREWDYRTVITPDGIIAD
                     WLIGKLKTVLIENADGSFAIDLNKPGGAIFRNNGVDAIKIENNMVQLGNWGKNGELIG
                     GLMSLLNNSDPSKPIVALGNALNSATMITYPKEGTITHPSYITFDKFNIMGDKNGVAI
                     RIYEDIDFKNNDLYNAVLSSENNENFIYTNNDKTVIKHGSVFIKLENGGATISGDLTL
                     DGVIKNTSGQVILDPTNPSIGGGDGDWQNGVISSNGFRFMKGYEGLGRYLYYDSGGIA
                     TIGYGVTMSEPTVFNKLKANQPVPEEMAAKESYNLKIRDYGKPIIQRCKELGITRQQQ
                     FDALCDLAFNAGTGRILANNSLTNAIMRNPNDEAYIRPIWEKFIIKDAAGNILNGLKA
                     RRKAECDIYFSGIYEKRPILTINSNGSYGSPVTENNGDGWLPSVTENYKTVNNDYGTG
                     WVIPTTGTITATFPTYPSGGKHDGIDFGTPIGTPVRVSKDGVVIKRRELTTSYGKYLF
                     VDHGGGLITIYAHNSELLVNEGDRVKRGQIIAKTGNTGNSTGPHLHWELRVNGIPQNV
```

```
                APSLRVGQVIR"
CDS             15106..16584
                /note="ORF16"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="minor structural protein"
                /protein_id="AAL96786.1"
                /db_xref="GI:19908310"
                /translation="MKYIQTKVVYIDRDEQIEIKAIEHDIKTRFINFKFIGENKILDI
                SHCIVRIYALNSKYNEVFNNLKIIDGAKGIAQLELTDALLIEGVTKYQLKIYDDNGGI
                LSSKIFKLIVDKNLMVDNAIEGSNEYRALDEALKTVGELNSMKVSIGENTNAIKEHDI
                FMRESRKEINDNKAIFDKFVEDSREFDETMKETSCFNYVSNPIFSTGSLKGWELWGNG
                ATCTVLSDTSLSHKYSLKIKCDSPSQGVTQTISGLVYGKHYTFKAKVWVESGTPGIMV
                RNDDSWGGAIFESSQGYSKWVELKMKFWAKEGTTPIYIGTVNNESKNASFWVSEIMLH
                EGYGEIPFIDNMKELYSEKFQVDLKGMVFGDGDGTYSNSNNGDLEYTTGGITNKYPAL
                TYISAFSIPAGNPGRVNIKLPKEFVKRKSSLKWGVVPKGYYYNTSGNFFPFHVAVNAV
                GEAYEEGGFIICPVEGYCRIQNGANAGDVQNTSINAVLIALA"
CDS             16597..16905
                /note="ORF17"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAL96787.1"
                /db_xref="GI:19908311"
                /translation="MSKKILKEEKVNLENKIEEDSTQDFNSKMTIFYSPKTGKIAHVP
                DGKQDMDYFAEDKDDFNYEFLVVEKDLYIKNNPEKFIVQDGEVRLKLNPELDKYNVAR
                "
CDS             16914..17669
                /note="ORF18"
                /codon_start=1
                /transl_table=11
                /product="Gp18 protein"
                /protein_id="AAL96788.1"
                /db_xref="GI:19908312"
                /translation="MVNEIFINIDELNNERIVTTQNNNNCETYKIYIIQNNKRIDLTG
                KTVELAYLKNDNKNGNIIDLNITNARDGEITLEIDKLLTKQDGVYSCQLAIYKENYLE
                HSATFEMIIKESLFAKISGEIEENNFEILNNLISNVERNTTDIEELKKVVKKEAKLKL
                TGGFEITVSRVNNMVILDIYNAATNGQMSGEGSLPVWALPPRTLGCSISASGPNFNGV
                LNVIKSGKIYYYVATATSEVGVYCGQLVYFTDK"
gene            17684..18061
                /gene="hol"
CDS             17684..18061
                /gene="hol"
                /function="forms lesions into cytoplasmic membrane"
                /note="ORF19"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="holin"
                /protein_id="AAL96789.1"
                /db_xref="GI:19908313"
                /translation="MFKFIPEVISWLLVLYIGFKIIDMILGVLKTIKNKNYRSRKMRD
                GIIRWVAELMAIAFVLILDMFLGLKFTVIGVTLALFAYKEAGSIVENLGECGVELPEI
                VSEKLEVLNKNNKNKEGFNKKEN"
gene            18149..19192
                /gene="ply"
CDS             18149..19192
                /gene="ply"
                /note="ORF20"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="endolysin"
                /protein_id="AAL96790.1"
                /db_xref="GI:19908314"
                /translation="MKIAERGGHNFQATGAVGLINETVEDRKVLAAAYKYTKAAGYDV
```

```
                     LDVTPGNCDSNTDLILGVNKAERFGAELFLSYHFDKCYDEYNGALGVACWICATGGKA
                     EEYAKSIVDTIAAGTGLKNRGVKVNPKLYELRKTSMPAVIVEVCFCEATEDVRIYKEK
                     GADLIGKLIAEGVCKVAGGQVPGTVIENVEYEVQESKPVPVYDRNKFKTNARALVNLD
                     PRDRASGIYEDLGEIYKDERFYVLPEVCDKGDYLPVLYWKDGANRASNKVWVSSKQKY
                     MMIDTYHRVVNVVTELDARYEPSPNSNRMGYVCNAERVYVHKIEGNYALCTYFAGEGY
                     KTAWFTAKYLERI"
CDS                  19382..19804
                     /note="ORF21"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp21 protein"
                     /protein_id="AAL96791.1"
                     /db_xref="GI:19908315"
                     /translation="MRKIIAILLSVLVFGSLLWGCGDKEETTNLNINQNDNSMISQSE
                     KDSYKEEVKKVLNGLTLNDIEILTQNKTKKPIVMVQVIFDGTKGNINKSEVEEFIKTI
                     KEKIETISNLYDITILDKNTQMIATTGYENEGKIYFFD"
misc feature         19984..19995
                     /note="attP site; attachment site"
                     /evidence=experimental
CDS                  complement(20070..21119)
                     /note="ORF22"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integrase"
                     /protein_id="AAL96792.1"
                     /db_xref="GI:19908316"
                     /translation="MPTYKNKERGTYYCSFYYKDWTGKRKRKKKEGFKTKKEADKFER
                     EFLVKVQENSDIDFKTLTEVYLEDAKIRVRETTLENKKYLINLKILPYFKDKLISNIT
                     PNDIRQWQNTLKSMNYSLTYLKTINNQLTAIFNFGIKFYGLNTNPCHRAGSMGKKNAD
                     DMKFWTTEQFKTFIEYNDKPISKLAFEVLFWTGMRSGELMALTFNDIDLKNKTININK
                     TYTRLNGKDIINPPKTPKSKRKVSIPNALCSYIETYITKLYDYNKSDRIFSFTKYFLV
                     HEMKRVCKKSNIPIIRIHDLRHSHASLLIELGFTPLLISERLGHEKVQTTLDTYSHLY
                     PNKDTMVSNKLDELI"
CDS                  complement(21135..21587)
                     /note="ORF23"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp23 protein"
                     /protein_id="AAL96793.1"
                     /db_xref="GI:19908317"
                     /translation="MTKYEELIEEIASYGVKVIEMRLGQDCGYCCNDVIFINETSPEK
                     TKYCILAEEIGHYFTNHGNITNLSKIENIHQENKARAWAYERLVSPEALIEALLKGVN
                     STEEITEYFNITKDFFLEAIKYYRKRYGIYYVSKECVLNLEPLKLIKF"
CDS                  complement(21608..22330)
                     /note="ORF24"
                     /codon_start=1
                     /transl_table=11
                     /product="putative CI-like transcriptional repressor"
                     /protein_id="AAL96794.1"
                     /db_xref="GI:19908318"
                     /translation="MNIGQRIKEKRKSMGLTVDEVAKKLNKNRATVYRYESNDIENLP
                     ISILEPLAEILQTTPSYLMGWEDNELNSYENLNKSKKLKELILSRYNSIREFSKIVGI
                     PNSTLTSALDKSIGGMAVDRIIKICDVLNIDVKTFEPLENVSSNTNNYSEEEQNHIED
                     LRKLNDIGKNKVITYTKDLIDNPKFINEPIQKQSENEISATSIDDFEPYLLAAHSDGL
                     DEETNKANIEKIKEMYLKIKGK"
CDS                  22504..22722
                     /note="ORF25"
                     /codon_start=1
                     /transl_table=11
                     /product="putative Cro-like protein"
                     /protein_id="AAL96795.1"
                     /db_xref="GI:19908319"
                     /translation="MVNLNKLKLKIFENGMNVSQLAEKIGMDRSTLYRKFQKNGDNIS
                     IKEANLMVETLHLTVEEANAIFFSNYVA"
CDS                  22749..23645
                     /note="ORF26"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="Gp26 protein"
                     /protein_id="AAL96796.1"
                     /db_xref="GI:19908320"
                     /translation="MNNLVLKEKQEFIGIQIPVIEGGFGENQKVILGKTIAEIHELRL
                     DKINSLINNNLEEFEEGIDILDLKNSNLTEVSLLELGFTKQTIANSKNIYLLSEQGYM
                     ALVGLMRTDKAKEIRRKLRREYFAMKAEIKENSFDISELSPELQMFNKMLQAVAKSEL
                     EQKRIKQQLNEVNYHALEAKEKADKSIEEIQAMREVVATNTTNWREECRKIIVKIAHK
                     LGGNSYIHDVQTEVYSLMRTRLRCKLDVKLTNMRRRMAEEGICKSKRDKKNYLDVIED
                     DARLKEGYIAIVKDLAFKYGVA"
CDS                  23680..23856
                     /note="ORF27"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp27 protein"
                     /protein_id="AAL96797.1"
                     /db_xref="GI:19908321"
                     /translation="MAVKRTYYTVDDIMSFLGVGQSKAYSIIRELNNELKSKGFITIA
                     GRVPKKYFDEKYYC"
CDS                  23903..24082
                     /note="ORF28"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp28 protein"
                     /protein_id="AAL96798.1"
                     /db_xref="GI:19908322"
                     /translation="MLTPIGEVVLGTISIATTLFLTVFFLEKYLEERNSKKRTKYLIL
                     SIANILSLLFVSNVI"
CDS                  24093..24299
                     /note="ORF29"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp29 protein"
                     /protein_id="AAL96799.1"
                     /db_xref="GI:19908323"
                     /translation="MFRVDFIDIGASNKNFSKYIKEINHENLYHAVKPYLQSRNIWFT
                     CRDGEGVVFAGFQTVGKFKVIIEI"
CDS                  24319..24537
                     /note="ORF30"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp30 protein"
                     /protein_id="AAL96800.1"
                     /db_xref="GI:19908324"
                     /translation="MLKVAVVGFEDSRKTFLYKTKFLDLKPGDYVWVKDSKNGFKRGI
                     FQNYTNAISRIRQAENWIIDRKDIKNGY"
CDS                  24527..24853
                     /note="ORF31"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp31 protein"
                     /protein_id="AAL96801.1"
                     /db_xref="GI:19908325"
                     /translation="MDINYISKKQSEEFINNWLSGNTLPLEKYISCYGTNQYVAIDNS
                     TNECWTEEFKTKEGCERYLLYFEDVEEVRAWEENRLRKIEISIYGVYYLLIFSMILVL
                     FYLLRI"
CDS                  24867..25412
                     /note="ORF32; similar to sporulation-specific sigma
                     factors"
                     /codon_start=1
                     /transl_table=11
                     /product="putative sigma factor"
                     /protein_id="AAL96802.1"
                     /db_xref="GI:19908326"
                     /translation="MINVEDYIGAAYKVASKFYSKNWDMSIDEINSIAFLGLLKAKKR
                     FEYKGVEFVTYAMKTIEYEIKNSLYRDRSKFRRKNIAGKDVYERIYTDSLDKEINSAE
                     GVKLIDCLVEESNTDEYVENIDLKIAINKLNTYQRKILKMLYFDEKTQRETAKILSVH
```

```
                EKTISREKKKILNLLRETLTA"
     CDS        25428..25547
                /note="ORF33"
                /codon_start=1
                /transl_table=11
                /product="Gp33 protein"
                /protein_id="AAL96803.1"
                /db_xref="GI:19908327"
                /translation="MDKLLNILIEKKRECIGTDQSNQCTQIMNQSYFNILNKK"
     CDS        25563..26315
                /note="ORF34"
                /codon_start=1
                /transl_table=11
                /product="Gp34 protein"
                /protein_id="AAL96804.1"
                /db_xref="GI:19908328"
                /translation="MKYTIHGFQQNKLIELGLDNNDALILSVIRDMYASQSIEFQIID
                GERYIWVNQSNLILDQIPIIGHRTSLLRRLSNLEEKGIIIRKLSHEKLNKKDKQFKKG
                NFCFIKLTSKLDYLMEYDLVAKCNNPCSKIQQPLVAKCDNKDSPIKDSNNIYSQVIDY
                LNSKAGKSFKYTTKKTQDLIKARLADGFKKEEFFKVIDNKVKEWQGTEYEKYLRPETL
                FGNKFEGYLNQGVNKKSYEERKENSSKDSFDY"
     CDS        26325..27608
                /note="ORF35"
                /codon_start=1
                /transl_table=11
                /product="putative helicase"
                /protein_id="AAL96805.1"
                /db_xref="GI:19908329"
                /translation="MEKNYSIESEKIILGTIIQDNELMLKAINILGENDFYNSKHKII
                FRAIKELYKNNISFDLTILSEKLSNDIKKQVLTLSDLTEISYHTSRGAFDSHLSLVKE
                KSKERKLNETCKSILTTPGTIESKIDFLQNTLLELNSSEDKEKVYTMSEVMEKTFEKI
                EKAFSSKSGITGISTGIRKIDNATNGLEKKEFIVLGARPSMGKTALSLAIMENIQGNV
                LYIQLDMSKEGMGQRLLASNSYIENGKIARGKFNDNEMKSIINVFDRLSRKENIFIYE
                PASITVNQIRLIAKQIKIKHGLDVIIVDHIGKIKPTTKGTKYEQSSYISNTLKSIAKE
                LNVAMVGLCQLSRAAEQRADHRPMLSDLRDTGSIEEDADVIGLLYRDGYYRAREDKED
                IVGDVLEINFAKCRNGRTGVVELQYHLPTQRLSEF"
     CDS        27625..27825
                /note="ORF36"
                /codon_start=1
                /transl_table=11
                /product="Gp36 protein"
                /protein_id="AAL96806.1"
                /db_xref="GI:19908330"
                /translation="MCELYENMLNEILSKQYDICKDKWNCLDCIFENECDKVFVSDRP
                DSICDKETFIKKIKEEMLRGCE"
     CDS        27825..28061
                /note="ORF37"
                /codon_start=1
                /transl_table=11
                /product="Gp37 protein"
                /protein_id="AAL96807.1"
                /db_xref="GI:19908331"
                /translation="MTRYEKTINSLMKIRDKFSIGSVEWDEIDNGLQLVTAIMAENKA
                LQKAYNEKKEMLEIHEWAFEGLRQAYENREEKEL"
     CDS        28058..28270
                /note="ORF38"
                /codon_start=1
                /transl_table=11
                /product="Gp38 protein"
                /protein_id="AAL96808.1"
                /db_xref="GI:19908332"
                /translation="MSCRACYGGMERKCRVTGDRCVFMIPDEKLCYELLGEGPLEYEK
                NIGIDLAKGKDKTGIVRKVRGDLENE"
     CDS        28263..28679
                /note="ORF39"
                /codon_start=1
                /transl_table=11
```

```
                     /product="putative single stranded DNA binding protein"
                     /protein_id="AAL96809.1"
                     /db_xref="GI:19908333"
                     /translation="MNKVVLIGRLTKDPELRFTATKGTAVARFTLAVNRKYKGENGEY
                     EADFIQCIAYGKPGEIIAQYMTKGRQLAIVGNIRTGSYEDKEGTKRYTTGVVVEGFDF
                     IDYAKRENDFGGAFSTPDNSFNNDDMVPIDDGDVPF"
     CDS             28706..28942
                     /note="ORF40"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp40 protein"
                     /protein_id="AAL96810.1"
                     /db_xref="GI:19908334"
                     /translation="MAKFKITINEIVNFNHEMTVEAKSESELNKVLDKIEREANYRDD
                     VDYILEEHGIKILDFNEDGSGEVNIEVPDLEEVE"
     CDS             29020..29295
                     /note="ORF41"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp41 protein"
                     /protein_id="AAL96811.1"
                     /db_xref="GI:19908335"
                     /translation="MDKKYFEHIANIEELGKYHYKVLLLLITDEYTQSDVRKLLNLRK
                     QNVNRVFRELEKLGLIEVKEIKGANKYYRLVDVKRLKLNIVGQMKII"
     CDS             29304..29576
                     /note="ORF42; similar to sporulation specific
                     transcriptional regulator SpoIIID"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator"
                     /protein_id="AAL96812.1"
                     /db_xref="GI:19908336"
                     /translation="MYKRKYKSRQVEIRVLNVVDYIINNKVTIREAAKVFGVSKSTVY
                     LDSTSRILEINPQKAMEVEKIILQNKSKRAMRGVKARKIKSLDRTS"
     CDS             29638..30168
                     /note="ORF43"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp43 protein"
                     /protein_id="AAL96813.1"
                     /db_xref="GI:19908337"
                     /translation="MVIDKKMFRLAEERIRNYYRKNAKIKSLKEKISFLENSIKIIEA
                     KMENNYFNIPSNLKSPDFVEKIQSSCVGSPIEQSMIKVLEGYENKIAAYIKEIGYLEN
                     LIFEIEEDYKIIECNMLNSLEEEERNFLEEAYKKRIPNWKLANKYHMSEVSCTRKKKR
                     LIREIIIWEKDRLSEI"
     CDS             30515..30865
                     /note="ORF44"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp44 protein"
                     /protein_id="AAL96820.1"
                     /db_xref="GI:19908344"
                     /translation="MKKLLINILGGVTVGAIAIVSLFGIGKISNIYDRTINKESVSIQ
                     REKFKESKSYVESMINTLAEYKREYERAKTKEEKEQIRNLIDSKFSNFNIDYIENKDL
                     YNFLQQMRGGNFNE"
     CDS             30858..31424
                     /note="ORF45"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp45 protein"
                     /protein_id="AAL96814.1"
                     /db_xref="GI:19908338"
                     /translation="MNKKRVTAILLVLSLGIVGVGCGEVKDKKDVGSQETINTQQEQL
                     LEESNRQLGMPEITNFFEKDMAKKIWELRDDPQLTTYAYFKSMDGKFVYLGRCIGFGL
                     PYTTQYTNPEKMDTVKGGEYARNPYVLPQADPNGLFSSETTNATWLMVVNEETNEPEI
                     MYSEPEIIVTQSKLNRRLVEEWSLPANY"
     CDS             31428..31973
```

```
                     /note="ORF46"
                     /codon_start=1
                     /transl_table=11
                     /product="putative recombinase"
                     /protein_id="AAL96815.1"
                     /db_xref="GI:19908339"
                     /translation="MNYVEPIRNSSDVNALANKLRKDSERNYIMFIIGIYSGLRISDI
                     LRLRVRDVKGKEQIAIREKKTGKQKMFPINPILKKELKKYCSDKELDEFLIKSRNGYN
                     RAIGRERAYTILREAGEELGIYNLATHTLRKTFGYHYYMQYNDIAELQKIFNHSDPAV
                     TLRYIGVEKSRIDRRIKSFKI"
     CDS             32153..32500
                     /note="ORF47"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp47 protein"
                     /protein_id="AAL96816.1"
                     /db_xref="GI:19908340"
                     /translation="MNLEEVKESRLNELEKKYKPIIEAYNDTVEMVNSIEIKDNDSID
                     IIICKLYILTNNVNVTLKLLSKSEYRINNRKVNSSDISEILNSISKKETNQIKAFAKK
                     QFLNNKKKSIKLC"
     CDS             32533..32718
                     /note="ORF48"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp48 protein"
                     /protein_id="AAL96817.1"
                     /db_xref="GI:19908341"
                     /translation="MEKNKEFLRVRDIFRECADIMDKVIDLEKREEKGEDVTPETERL
                     MGRYMMLLMELNSLTNN"
     CDS             32735..33058
                     /note="ORF49"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp49 protein"
                     /protein_id="AAL96818.1"
                     /db_xref="GI:19908342"
                     /translation="MKKTKYRCCNCGLEYEFEGKAKEGLSCEICTEPLRPLVQERELK
                     DKRRKLIKSVTKDFLSQEMKLGIREFMSLINEIESLEKELEKNKKIIAKLKKALRKRN
                     AKESM"
     CDS             33039..33467
                     /note="ORF50"
                     /codon_start=1
                     /transl_table=11
                     /product="Gp50 protein"
                     /protein_id="AAL96819.1"
                     /db_xref="GI:19908343"
                     /translation="MPKKVCNYSGCTNLISVKETYCSKHRVESNKERHKCYKVRRKDK
                     EEQAFYNSKEWRIVRGFVINRDNGLCKLCLSKKKIRVADVVHHVIELKERRELGLETG
                     NLLSLCNSCHQEIHSKYNKGMVTKKNTQDELFKLIPNGEG"
     misc feature    33499..33507
                     /note="3' overlapping ssDNA (cohesive end) at (right) end
                     of viral DNA molecule; hybridizes to complementary
                     sequence at the other (left) end of the molecule upon
                     entry of viral DNA into host to form circular molecule"
                     /evidence=experimental
BASE COUNT    13735 a   3112 c   6398 g  10262 t
ORIGIN
        1 tatgcagaaa aaatttttt cggtttataa agtgaaaggg ggtatattat agtttggcta
       61 gaaagatgat ttcagttaat caaattattg caaatggaaa taagagccat ttacctaatg
      121 aagaaattga aaaaagaaag gaacaggaag aaaaattaaa aaaattacct agagataaaa
      181 taagaccacc aacatggcta tcaaaagagg gcaaaagcat atttaagaaa attgtcaagg
      241 agttagaagt tgtagatata ctcgcaaata tagataatta taatttagct gttctagcta
      301 atgctataga gcagtatata gaatgtacaa gaaaattaaa ttgtgatgaa cttacaatta
      361 ctcatattaa taaaagaggt tttgaaacta ctcaaaagaa tcctttaata tctattcaaa
      421 ttcaatatgc agatgttatt aaaaaattag gaagtgagtt tggactaagt ccagcagcta
      481 gattaaaaat tattcaagaa gcaagtgata tttctgatga tgaaaaagaa ttaaataagg
      541 attttgaaga tgtttaatac agttcttgag gaacttatag attattcaaa ttcaatatta
```

```
 601 gatggaaaga ttaatgcttg taaaaggcat aaacaagctt gcaaaaggtt cttgaatgat
 661 ttagaaagaa tggaaagtga tgattttgag tattattggg atgaagaaga agctcaaaga
 721 atagttaaat ggtatggaca ttgtagacat tcaaagggtg tattagaagg aaaaccaata
 781 atattaaact catgggctaa atttgttgtt tgtaatatag aagcttggaa gcataaagat
 841 acaaattaca gaagatttag gtttgctttt atacaagtag caagaaaaaa tcaaaaatca
 901 cagttagaag ctggtatggc tggttatgaa ataggagcaa aaggatataa tgcagcagaa
 961 gtttatactt taggcgttga aagagatcag gctaaaatag tttttgagga atgggagtta
1021 atggcttcta aaccattgaa aaaaagattt aagtttactc aaaaagtaat aaaacataaa
1081 aaaagtaata gttttattaa acacttaagt aaaaaggcag gtaaaactgg tgatggtaag
1141 aatcctcaaa tggctgttat agatgagtat catgcacatc ctaatagtaa aatgtatgat
1201 gttatgaaat caggtatgat ggcaagaaca gaacctttat tagtaataat aactacggct
1261 ggagaagact atgaagaaac agcttgttac tatgagtatt tagattgttg ttctatctta
1321 gatggaactt ttgaaaatga aaaatacttt gtaatgattt gtgagctaga agatgaagat
1381 gatccttttg acgagaaagc gtggataaaa gctaatccag ttttatgcac atatgaagaa
1441 ggcattgaaa gtatgagaca aaatgctaat ctagctaaaa atacaagtaa tgaggaaaaa
1501 agaatagaat ttttaactaa gaactgtaat atatatgttg cggctggaga gaagaaatat
1561 gttgatgttg aattttggaa ggcttgtaaa gagaatataa ccttagagaa atttagaggt
1621 catgattgtt atattggtat ggacttatca aaatcgggag atttaacatc aatagctttt
1681 gagtttcctt acttagatga aggaattaga aaatatgctt tatttggtaa gtcatttata
1741 ccagcagaag ttgttaagga gaaaatgaaa acagataatg taccttacaa catttggagt
1801 agtaaagaga aaggatggct tataaaaaca gaggcaaatg agggacaaat agtagattta
1861 tgggcaattt taaatactat agagagtatt gttaaagaat atgaactaaa tgttattgaa
1921 gttagttatg accctcatgg agctgcaatg ttagttagtg agttagaaag aagagattat
1981 aattgcgtag agtgtggaca aagttgtgcg aaacttaatg aagcaactgt caactttaga
2041 gacttaatga agataaaaca aattgttcat gatgaaaata atcttatgac atggtgtgtt
2101 caaaatgcag aaaaagatac aaactcattt ggagaaataa aaataagtaa aaagagtaga
2161 tttaaaagaa ttgacccatt agcttcttgt atattcgcac acaacagagc tatgacttat
2221 tggaatagag aaaatttaga tgtgagtgaa tttgcagatg aagatttctt aaaaaaatta
2281 tggaggaaaa aatagatggg tttcttaaga ggattattta acaagaaaaa taataatgat
2341 gaaaaaagaa gtaattatga tgaagatatt ggtattgata tatcggattc taacttttgg
2401 gagaaattcg gtattaaatt gaatttttca gtaagaggaa aaagagcttt aaaagaaaat
2461 acagtttatg tatgtaccaa gattagggct gaaagcatag ggaaattatc tttaaagatt
2521 tataaagata aagaagagta taaagagcat gaattatatt atctattgag gtataaacct
2581 aatccattga tgagttcaat aaacttttgg aagtgtttag aaactcaaag aacattgaaa
2641 ggaaatgctt atgcttatat agagagagat agaaaaggaa aaattatagg attatatcca
2701 attaattcag ataatgttac taaaataatt gatgatgata attttctaag tagtttaagt
2761 aaggtctggt atgtagtaac agataaaaat gggaaggaac ataaattact tccagatgaa
2821 atgttacact ttataggtga tattacttta gatggattaa taggaataaa gcccttggat
2881 tatttaagat gtactatcga aaatgggaga gcaactcaag agtttataaa taaattcttt
2941 aaaaatggat taagtattaa aggaattgtt caatatgtag gtgatttaga tgaaaaagct
3001 aaaaaaatat tcaaaaagga atttgaatca atgagtaatg gcttagaaaa tgctcattct
3061 atttcattac ttccttttgg gtatcaattt cagccaatat cattaagtat ggcagatgct
3121 caattcttgg aaaactcaaa gttaacaaaa agagaattgg cagcaacatt tggtatgaaa
3181 tcataccatc ttaatgattt agaaagagct acatttaata atcttacaga acaacaaaaa
3241 gatttttatg taacaacact tcaatcttct cttactgtat atgaacagga aatacaagat
3301 aagttattca gtcaatatga aaccttacag gatgtaaaag ctgaatttaa tgttgatact
3361 attctaagaa gtgatattaa aactagatat gaagcttata gaataggaat acaaggtgga
3421 tttattgagg caaatgaagc tagaaggaga gaaaatttac caccagtaga aggtggagat
3481 aggttacttg ttaatgggaa tatgattcct atagaaatgg ctggagaaca atataagaag
3541 ggaggggaaa agggtggaaa ataagtatga ggattataaa aatgaaatca aagaaggtaa
3601 aagaactctt aatcaaacaa gaaaagagtt aggattacca cctattaatg gaggaaatac
3661 agaatttatt aatagagaaa ttgctttctg tagaaaaata caagaaagga gatgagtata
3721 tggagcataa agaagttaaa agagaagtta gacatcttat ttctaatatt gaggtaagag
3781 caactggaga tgaaaatact aaaacaataa ctggttacgc ttcaaagtat aacgttgaat
3841 ctcagatatt aagagattgg tggggcgata agtttgttga agtagttgca gaaggagctt
3901 ttgataatag cttaagaagt aacacaataa aagctttata taatcacaat actgataatg
3961 tattaggttc aacaaagagt ggtactttaa gattagaaag tgattcaata ggacttagat
4021 ttgaaataga tttacctaat acaacagtag caaatgattt atatgagagt gtaaagcgtg
4081 gtgatgtaga tggtacttca tttggattta aagttcttga tgataagtgg tcaaaagttg
4141 aaatagatgg tgaggagata atgaaaagaa cattattaga gattgattta tatgaaatat
4201 caccaacacc atttccagct tatgaagata cagaagttga ttgtagaagt ttagaaaaaa
4261 ttaaatttaa tgaaaataaa aaagaagaaa aaagaagtag tttattagag cttatgtatt
4321 gttaaggctc ttttttttatt acttaaaatt aggaggattt tataaatgaa tttatttgaa
4381 agattaaaag aattaagagc aaagaagaaa gagttagaag aaaaaagatg tggaattgta
4441 gaagaaatta gatcattagc taaagagaaa aaggaagaag aagctagaag taaagcatta
4501 gaaagagaaa agatagaagc tagaatggaa ataattgagg aagaaattga gtccgttatg
4561 acagctattg atgaagaaag aaaaaacact aattttacag gtggaagagt aataattaat
```

```
4621 ggtgattcta aggaagaaaa aagaagttta caattaagtg caatgagtaa aactattaga
4681 ggtatacaat taagtgaaga agaaagggat attatgtcat caactaataa tggagctgta
4741 atacctcaag aatttgtaaa tgaatttgaa aagttaaaag aaggatatcc ttcacttaaa
4801 gaacattgtc atgttatacc agtaaatagg aatgcgggaa aaatgccagt aagagcagga
4861 gcatcagttg ataaacttgc taacttggca aaagatacag agttagtgaa agctatgctt
4921 aaaactcaac ctatggctta tgatatagat gactatgggc tacttgcacc gatagataat
4981 tcattattag aagatagtga aataaacttc ttagagtttg taaatgaaga atttgcagag
5041 tttgctgtta atacggaaaa tgcagagatt gttaaacaag ctaaggctgt attagcagaa
5101 gaaactataa atgattatgc tggacttgtt aaaacaataa actcattagt ccctaatgca
5161 agaaaaagag caattattgt aactaattca gatggaagag cttatttaga tggacttatg
5221 gataagcaag gaagaccttt attaaaagaa ttatctgatg gcggagattt agtatttaag
5281 ggtagaccag ttatcgaatt agaagaatca atatttgatg ttggtgatga aactaagttt
5341 atagtttcag attttaaaac actaattaaa ttcatggata gaaaacagta tctaatagat
5401 caatctaaag aagctggata cactaaaaat gaaactattg ctagaataat tgaaagattt
5461 gatgttaact caccacttga taaatcttct gatgcagaga aaataagaaa gtttggagta
5521 attgtaaaat tacaagaagt attaaaaagt agtccaaggt caggtaagaa caagaacgaa
5581 tcaaaagaag aaataaaaga agaaggagaa gctactcaac aaaatgagta gttcttgctt
5641 atggatttaa gagagttaaa agagtattta agaattgatt tcgaagaaga tgatatactt
5701 cttcgaagtc ttttattagc agcagaagaa tacttgtata atgctggaat aaaaagagac
5761 tataaaaagt cgttatatag tcttgctata aagatacttg ttaagcattg gtatgataat
5821 agagattgtg ttgttgctgg aaatgtgaat aataaattgg aatattcatt aaatgcaata
5881 ttaacacaac ttagatattg tggtgatgat aatggctaga tatgaaagaa tttcaataaa
5941 aaagctagaa gaaaaaaata ttaaggggag aaggcaagag gaatgcctaa ttccctttta
6001 tgattgttgg gcagaaatat tagatttgta tggacaagaa ttatatggtg ctcttcaaat
6061 gaaactagaa aatactatta tttttaagat tagatactgt aaaaaggtag aagaacttag
6121 aaataaagag aattttatag ttgaatggca agggagaaaa tatgagatat attatcctga
6181 ttttttaggt tacaataaac aatttgtaaa actaaagtgt aaagaggttt tataaatgag
6241 tgatggatgg gaaattaagt ttgaaggttt ggatgattta ataaaaactt ttgaacagtt
6301 aggaacagaa aaggaaaatg aagatgttga aaaaagtatt cttaaagagt gtggagattt
6361 agcgaaaaaa acagtagctc cattaattca tataagtgat gacaatagta aaagtggaag
6421 aaaaggaagt agaccaccag gacacgcagc taataatatt cctgaaccta aaataagaaa
6481 aaagaaaggt aatttacagt gtgttgttgg gtgggaaaaa tcagataata caccgtttta
6541 ctatatgaaa atggaagaat ggggaacatc agaaagacca cctcatcatg cttttggaaa
6601 aactaataaa atattaaaac gagtttatga taatatcgcc caaaagaaat atgataattt
6661 tgttaaagag aagttaggtg attaatttgg aggattttga tattatcgcc ttagtctatg
6721 agtgtttaga atgtttaaat gttccagtta tagaggggtg gtatgatgaa gagcttaata
6781 aaactcatat aacagtacat gaatatttag aacaggatga atcatttgaa gatgatgaag
6841 ctagagaaga agaacataat attcaaattg atgtatggag taaagatagt ttagaagctt
6901 ttaagttaaa gaaggctatt aaaaaacttt tgaaaaagaa taatttttat tttgattcaa
6961 gtgaggattt ttacgaaaca aaaacaagaa tataccataa aggtttaaga tttagctata
7021 taagtgaaat tagctaaata gaaaggagaa aatatgccag aagtagttaa tacaagaaga
7081 tgtggatgta aagatatata tattgcaatt gttaataaga atacagcaac agagtacgct
7141 acggctacac caattaaact aggaagagca ttaaatgcta aagtaacaac taaaaagact
7201 gtagaatcaa cagaaagtga tgatactatt gaggaaactt ttgaaagttt tgatagtatt
7261 gaaatagaat ttgatgtaaa taagcttagt ccagaacaaa aagctatatt aagaggggca
7321 acttatgaag atggattttt aattaataat gttgatgatc aagctaaaga agtagctatt
7381 ggttggagag caagacaaac aaatggtaag tacgagtttg tttggtacta ttgtggtaag
7441 tttaatagtg gttgggatga ccaatacgaa acaaatggta agaaacctaa aacacaaact
7501 gataaaatga aaggtacatt ctatggtaga caaaaagatg gtaatactag catagaagta
7561 gatgaaactt tcttattaga atcacataat gatgcaaaaa gttctataac aaattggttt
7621 tcaaaggtac aagaaaaaga ggctgcataa aaggaggata taaaatgaaa ataacagtaa
7681 agggtaaaca atatgatagt ggaaaaatgc tatttgataa gtatgatgtt tatactaaag
7741 ctagaaataa attagcagaa aaagagctat ttaatagtga tgatataaat ttaatgatag
7801 atactttggt atttatattt gacaacaaat ttactgcaaa agatattaag agtgaaatgg
7861 atatatcaga tataatattt aatttctcta gtgttgatat tgagattgct ggtaaatcaa
7921 atgataaagc ttttaaagcg agtgcgggtt ttatgaaagg caagaaattc aagagatagg
7981 gatttcttgt cttaatttat atactagcaa atttacaata aggaaatata aaaagtatat
8041 agagatagaa aagcgaatta gagaaaatgg aaatgacata tataatgtat gttcatttct
8101 aacagttttc tttgataata aagtttcatt atcagagcta aaaaatatag atgtttttga
8161 aatattagaa tgttatcttt ggattaaaat gttcattgat aaaattattg tagatagaat
8221 agctgatata ttctgtgaga aagtagaata cgagaaaagt gcctttgatg aatatgatga
8281 aatagaagga gttagtgatg aagagctaga taattacggt gattatgaag aaaaccaaat
8341 agagatttta aatagtattg taaattattc tataaaaaaa tatagagata gtttaaatac
8401 tgtattagat atggatttat tagaattact tgattatata gaaagtgata ttaataaaga
8461 tgatgaagga gaagagtact attaactctt cttcttttat tttaaatgag ggaggtgaga
8521 gttataggag ctaatgtaaa aataggagca aacacctcga gctttcagaa gcaaatgaaa
8581 gaaatggctc aagaacttaa aagagttgga agtagttata atttagccaa tacacaagca
```

```
 8641 aagttattcg gtagtagaac tgatttatta aaaagtcagc aatcagaatt aacagctaaa
 8701 atgaaaatac agaacaatat gattaaaact caaggagatt acttaactaa attaaatagt
 8761 gatttagata agcagaagtc aaaacagagt gaattagcta aaaaaataag tgaaacaagt
 8821 gataaatata aaaagagtgt tcaagaaact ggtaagaaca gtgaagaatc aaaaaaatta
 8881 aaagaagaat taaagcaact agaagctgaa tatacaaaga ataataaagc gatagatact
 8941 aatgtaataa aaataagtaa tgcagagata aaattaaata attcaaaaaa agctttagaa
 9001 gaaaataaaa aggctttgga aaatataaat aaagaattag ataagtctaa aatagataaa
 9061 ttctctgaag gaattgggaa agctggagaa aaggctggga atttatctaa taaaatgaag
 9121 ccagctagtg tagctatgat tggtttaggt agtgctgcta catttgcgaa tattaaattt
 9181 gaggatagtc ttgcaaatat aaatacactt ttagacgacc aaagtcattt agaatcttat
 9241 aaaaataaaa taagagagct ttctgatgaa acaggaatga gcctagaaac agtttcagaa
 9301 gggatgtaca ccgcaatatc ttcaatagga gatggaggaa aagaaacaga aaagattttt
 9361 gaaacaatgg ctaaaagtgc taaagctggc ggagcagaag tacaagatgc agtttcatta
 9421 ataagtgctg gaatgaaggg atataatcaa gttaatgatg aaaccgctaa aaaaataagt
 9481 gatttagcat ttcaaactgc aaaactagga gttacaactt tccctgaaat ggctaaatct
 9541 atgcaacagt tatttccttt agcgaataat ttagagcttt cactagaaga gttatttaca
 9601 gatatggcta cattaactgg ggttacaggt aatacggcag aagtaagtac acaattaaag
 9661 gctgttttta gtaatttaat aaaaccaact acagatatgc aaaagttaat gcaaaaatat
 9721 ggttttcaaa atggtcaagc aatgttgaaa agcgaaggat taattggaac tttaaaaatt
 9781 ttacaaaaag aaactggtgg acaatcagat aaaatgggta agctttttag tagtacagaa
 9841 ggattaactg cacttacagc attaacaggt tcacagtttg acacattggt agataagtct
 9901 aagaaaatga aagaatctat aggaacaact gattctgctc ttacaaaaat taatagtaca
 9961 ctaggaaatg atttaagaac atcactaaac ttagttagaa attcatttgt tgattttggt
10021 gaaatactag ctccttttat atcaatagga gcaaatactt taagtactat aataacaatg
10081 ataaatgggc ttagtgaagg tcagaaaaag ttagttgttg gattcggagc atcttttata
10141 ggaataaatc ttttgttagg aggattctct aagttatcaa aagggataca gagtaatatt
10201 aagtttacaa aagatatgat taaggttact aaagagggag tttctaagtt aaaagatttt
10261 ggaaaggctg caagagaggg tactaattta attggaaagt ttggaaaagg tatagtaact
10321 acagcgaagt cagttgctag ttttactttg gaattaataa aaagtgctgg taaagggcta
10381 gcttctttag gaaagggcgt acttaattgt attaagtcaa tgggtagctt tactttagaa
10441 cttattaaaa atacggctat tggtattaaa aatggtgcta tatgggtagg taatatggct
10501 aaaatgttag cctacaaagc ggcacaagta acggttacag tagcaacaaa agctatgaca
10561 gtagcacaaa aggctttaaa tttagctatg gctatgaatc ctattacttt aattattgga
10621 ttattagttg ctttaggagg aacatttata cttctataca ataaatgtga gtggttcaga
10681 aatggagtta ataatgtttg gaattctata acatcaattt tcactaaatt tgacaacttt
10741 cttacaggaa tttttactag agattggacg caaagttttg gtttattagg tgtaccttta
10801 aacgcttttc ttggaacagt aggttcaata tgggagggcg tgaaaggcgt attcaatgga
10861 gtaataagat tcttcaaggg ggtatttaca ggaaattgga gagaagtttt tcaaggtcta
10921 agtgatattg taagtagtat tttcggagct ataggtggaa taattaaagc tccaattaat
10981 gcagctataa gcggaatcaa ctgggttata aggcaaataa acaaaataag ttttgatgtt
11041 ccttcatgga tacctggtat tggtggaagt cattttggaa ttaatttacc tactatacca
11101 gcacttgcag aaggggggaat tgtaacacaa gcaactatgg ctttagttgg tgagggtaaa
11161 gagcatgagg ctgttattcc tttaagtaaa ctagataagt tagttacttc tagtgtaaag
11221 aaagtattag agaattacag taataaaaaa gatgaagggg attttaaatc agagcctaaa
11281 gtttttcaaa taatagtacc tatagatggt agagtacttg cagaaattat tattgatgaa
11341 catggaaatg tgttaactaa aaaagttaag gatagaagag tagttagggg gaaaatataa
11401 ttgattgatt taatatttaa taataaaagt ttgagacatg attttaaagt tagtatagta
11461 aaaagacctc caactatgtt ttctattaaa aataaaagaa aagttgaagt tttaggggga
11521 gatggagatt tatatttaga gttaggaggt cagcaggata taactatacc agttgaatgt
11581 aactttatat gctcaccaaa tgaaataagg gataggttta gattattaaa aagctggtta
11641 aacaatgtag aagatgataa attaatttct actgatgatc ataattggtt ttataaggtt
11701 gtagatataa atataagttc tatggatgta ataaaaagaa gaaaaggaaa gtttactata
11761 aattttattt gcagaggata tcattttaaa ttagatggtg acgagtttat tgatattttt
11821 aatgaaacaa tattatttaa tgagtatgat ttagcaaaac ctttaatatt gttaagaggc
11881 aatggaacta tagaggttga tataaatagc aataagtttg aggtggaagt taatgagttt
11941 gtatatattg atagtgaatt agaagttgct tataagtatg aagaagattc ttataattta
12001 gatattggag attatccagt attgatacat ggagaaaatt taatttcata taaaggcaat
12061 ttagatagtt ttgaaattaa atgtagatgg agatgtgtgt aatgcaatac tataagagta
12121 ataatacaaa ttttgaaaag aatggggaca taacacttca attacttagt ggagtattaa
12181 ggctagaatt aaatagtgga cttaatgaaa tagagtttga aatcccttat gataaacagg
12241 gaaggtggaa gaaggttgag gaatggggag ttgttaaaca aaaagttttt tacactaaga
12301 ataaacaatt atatagaatt tatgaaaaaa ataaaggtat gtttggcttg agactaaaag
12361 ctaggcatat ctttttgat ttagtaaaac atactatttt agataaaaga gcagtaaatt
12421 taaatggaca aagagcatta gagagaattc tttttgaaac taagtttaaa gctcatagtg
12481 acattgatat aaataataca gcttattttg ttaatttaaa tggtatagaa gcaataaatg
12541 gagataatga aaatagcttt agaaatcgtt ggggtggtga aatattaaac gataattttg
12601 atgtatatat aaataaaaaa attggtagag ataatggcgt taaaattaga tttggtaata
```

```
12661 atttagaaga tattaatttt aaaattgata gagacccaat aataacaaga gcatatccta
12721 aagcttacga tggaataatg ttaccagaga aatttattga tagtgaacat attaataaat
12781 atcctattat atgtgaggga tttgtagatg taagtgactt aaaattaaaa gactctgaaa
12841 atccagaaga tgaaggtttt gaaacaaaag aagaactcta tgaagcaatg aggagacgta
12901 tagttgattt atataataat ggattagata gacctaaagt aagtggaaat ataaaaatag
12961 taaatttaga aaataccaat aaatataaaa actttaaaaa tcttaaacat attggaattg
13021 gtgatacaat tacaataaat ttaaccgaaa ttggtataga tgctaaaaca agatttataa
13081 gtactgagtg ggatttagta acagaagaac ataataatat ggagtttgga gaagtagaag
13141 taaattattt tgataagcaa gaaattgtta atggcaggtt ggataatatt ttaaaccaaa
13201 atggcagcgt taaaagttct gagttagaag gaatcataaa tgctatgaaa gctccattag
13261 tagcacaaaa agatatagct aaaaatttag attgtgttgc gtggaagtgt gaagttaatg
13321 atcctaatga cccaaacttt ggctatttag gtgctgggac taaaggactc atgttagcaa
13381 atagtaaaac tccagataag agagaatggg attatagaac agttataacc ccagatggaa
13441 ttatagctga ttggcttatt ggtaaattga aaactgttct tatagaaaat gctgatggaa
13501 gttttgcaat agatttaaat aagccaggag gagcgatttt tagaaataat ggtgtagatg
13561 ctataaaaat tgaaaacaat atggttcaat taggaaattg gggtaagaat ggagagctaa
13621 tagggggatt aatgtcttta ctaaataata gtgatccaag taaaccaatt gtagctttag
13681 gtaatgcttt aaatagtgca actatgatta catatcctaa agaaggaact attacacacc
13741 cttcttatat tacttttgat aagtttaata ttatgggaga taaaaatgga gtagccatta
13801 ggatatatga agatatagat tttaaaaata atgacttata taatgctgtt ttaagttcag
13861 agaacaatga gaactttatt tatacaaata atgataaaac agtaataaag catggaagtg
13921 tatttataaa acttgaaaat ggaggagcca ccataagtgg agatttaact ttagatggag
13981 ttataaagaa tacaagtggt caagttattt tagatcctac taacccatca attggaggtg
14041 gtgatggtga ttggcaaaat ggtgttatat caagtaatgg attcagattc atgaagggat
14101 atgagggatt aggaagatat ttatattatg atagtggtgg aattgctact attggttatg
14161 gagtaacaat gagtgaacca acagtattta ataagttaaa agctaatcaa ccagttccag
14221 aagagatggc agctaaagaa agttataact taaaaattcg agattatggt aagccaatta
14281 tacagagatg caaagaatta ggaataacta gacaacaaca atttgatgct ttatgtgatt
14341 tagcttttaa tgctgggaca ggaagaatat tagcaaataa ttctttaaca aatgctataa
14401 tgagaaatcc taatgatgaa gcatatataa gacctatttg ggagaaattt ataataaaag
14461 atgcggctgg gaatatacta aatggactta aggcaagaag aaaagcagag tgtgatattt
14521 attttagtgg aatatatgaa aaaagaccta tactgactat taattcaaat ggaagctatg
14581 gttctccagt aacagaaaat aatggggatg gttggttacc ttctgtaaca gaaaattata
14641 agactgttaa taatgattat ggcactggat gggtcatacc aaccactgga acaattacag
14701 caacattccc aacatatcca agtggaggta aacatgatgg aatagatttt ggaactccaa
14761 ttggaactcc agttagagta agtaaagatg gtgttgtaat aaaaagaagg gagttaacga
14821 cttcttatgg taaatattta tttgtagatc atggaggagg attgattaca atatatgcac
14881 ataatagtga gttgcttgta aatgagggtg atagggttaa aagaggacaa attatcgcta
14941 aaacaggaaa tacagggaat tcaactggtc ctcacttaca ttgggagctt agagtaaatg
15001 gaattccaca aaatgtagca ccttcattaa gggtaggtca ggtaattaga tagttagagt
15061 agcttaaaag gtgctctttt tttattgtaa gaaaggtgga gtaatatgaa atatattcaa
15121 actaaagtag tctatataga ccgagatgag caaatagaga ttaaggcaat agaacatgat
15181 ataaaaacga gatttatcaa ctttaaattt ataggagaaa acaagatatt agatataagt
15241 cattgtatag ttagaattta tgctcttaat agtaagtata atgaggtatt taataattta
15301 aaaattattg atggagctaa ggggatagca caactagaat taacagatgc tttattgatt
15361 gaaggagtta caaagtatca gttgaaaatt tatgatgata atgggggaat tcttagttca
15421 aagatattta aactcattgt tgataaaaat ttaatggttg acaatgctat tgaaggaagt
15481 aatgaatata gagctttaga cgaagctctt aaaaccgttg gtgaacttaa ttctatgaaa
15541 gttagtatag gagaaaatac taatgcaata aaagagcatg atatttttat gagggaaagt
15601 agaaaagaaa taaatgataa taaagctatt ttcgataagt ttgtagaaga tagtagagag
15661 tttgatgaaa ctatgaaaga aacatcttgc tttaattatg taagtaatcc tatatttagc
15721 acaggaagct tgaaaggttg ggaattatgg gggaatggtg ctacttgcac ggtgctaagt
15781 gatacttcct taagccataa atattctttg aagattaaat gtgattcacc atcacaaggt
15841 gttacacaaa ctatttctgg tttagtatat ggaaaacatt atacttttaa agctaaagtt
15901 tgggttgaaa gcgggacacc tggaattatg gttaggaatg atgattcatg gggaggtgct
15961 atttttgaaa gttctcaagg atatagtaaa tgggtagagc ttaaaatgaa gttttgggca
16021 aaggaaggaa ctacacccat ttatatagga acggttaata atgagtctaa aaatgctagt
16081 ttttgggttt ctgaaattat gttacatgaa ggttatggag agattccatt tatagataat
16141 atgaaagagc tatatagtga aaaatttcaa gttgatttaa aaggaatggt atttggagat
16201 ggagatggga cttattcaaa tagtaataat ggagatttgg aatatacaac aggaggaata
16261 actaataaat atccagcatt aacatatatt tcggcttttt caataccagc tggtaatcca
16321 ggaagagtta atattaaact acctaaagaa tttgtaaaaa gaaaaagcag tttgaaatgg
16381 ggagtagttc cgaagggata ttattataat acaagtggaa actttttcc ttttcatgta
16441 gctgttaatg cagtaggaga agcatatgaa gaaggtggat ttattatttg tccagttgaa
16501 ggttattgta gaatacaaaa tggagctaat gcgggagatg tacagaacac ttctataaat
16561 gctgtattaa ttgcattggc ttaaggagga ataaaagtgt caaagaagat tttaaaggaa
16621 gaaaaagtta atttggaaaa taaaattgaa gaagattcaa ctcaagattt taatagtaaa
```

```
16681 atgactatat tctatagtcc taaaacaggt aaaattgcac atgtaccaga tggtaaacaa
16741 gatatggatt actttgcaga ggataaggat gattttaatt atgaattttt ggttgtagaa
16801 aaagatttat atataaaaaa taatcctgaa aaatttattg tgcaagatgg agaagttagg
16861 ctaaaactta atccagaatt agataaatat aatgttgcaa gataggagga aatatggtta
16921 atgaaatttt tataaatata gatgagttaa acaatgaacg tatagtaact acacaaaata
16981 ataataattg tgaaacatat aaaatttata ttattcaaaa taataaacgt attgatttga
17041 ctggaaaaac tgtagagtta gcttatctaa aaaatgataa taaaaatgga aatataattg
17101 atttgaatat aacaaatgca agagatggag aaataacttt agaaatagat aaacttctaa
17161 caaagcaaga tggagtatat tcttgtcaat tagctattta taaagaaaat tatttagaac
17221 actcagctac ttttgagatg attataaaag aaagtttatt tgcaaaaata agtggagaaa
17281 tagaagaaaa taattttgag attttaaata atctcataag taatgttgaa agaaatacta
17341 cagatataga agaactaaaa aaagttgtga agaaagaagc taaattaaaa ttaactggtg
17401 gatttgagat aactgttagt agagtaaaca atatggtaat attagatatt tataatgcag
17461 caactaatgg acaaatgagt ggagaaggaa gtttaccagt ttgggcttta ccacctagaa
17521 ctttaggatg ctctataagt gcaagtggac ctaattttaa tggagtttta aatgttatta
17581 aaagtgggaa aatttattat tatgttgcta cagcgacgtc agaagtagga gtatattgtg
17641 gacaattagt atattttaca gataagtagg gaggaaaaag aatatgttta aatttatacc
17701 agaagtaata agttggttat tggtacttta tataggattt aaaataatag atatgatatt
17761 aggagtatta aaaactataa agaataaaaa ttacagaagt agaaaaatga gagatggaat
17821 aataaggtgg gtagcagagt taatggctat agcctttgta ttaatattgg atatgttttt
17881 aggacttaaa tttacagtta taggtgttac attagcacta tttgcatata aagaagccgg
17941 aagtatagta gaaaatttgg gggaatgtgg agtggaatta ccagagattg tttcggaaaa
18001 attggaagta ctcaataaga ataacaaaaa taaagaagga tttaataaaa aagaaaatta
18061 gttaaggtaa gacaatgaat taggaactgg aaagtagttc tttttttatt gtctttaatt
18121 atataaaaat atgaaaggaa tgattaatat gaagatagca gaaagaggcg gtcataattt
18181 tcaagctaca ggagcagtag gattaataaa tgaaacagta gaggatagaa aagtattagc
18241 agctgcatac aaatatacta aagcggcagg atacgatgta ttagatgtaa caccaggcaa
18301 ttgtgattct aatacagatt taattttagg agttaacaag gctgaaagat ttggagctga
18361 attattctta agttatcatt ttgataaatg ctatgatgaa tacaatggag ctttaggagt
18421 tgcttgttgg atttgtgcta ctggtgggaa agcagaagaa tacgctaaaa gcatagttga
18481 tactatagca gcaggaacag gattaaaaaa tagaggggtt aaagtaaatc ctaagcttta
18541 tgaattaagg aaaacatcta tgccagcagt tatagttgag gtatgcttct gtgaagcgac
18601 tgaggatgtt agaatttaca aagaaaaagg tgcagattta ataggtaaat taatagcaga
18661 aggagtttgt aaagttgctg ggggacaagt tccaggaaca gtaatagaaa atgtagaata
18721 tgaagtgcaa gaatctaaac cagttccagt ttatgataga aataaattta aaactaatgc
18781 aagagcttta gttaatttag atccaagaga tagagcaagt ggaatatatg aagatttagg
18841 cgaaatttat aaggatgaaa gattttatgt acttccagag gtttgtgata aaggtgatta
18901 tctgccagtt ctttattgga aagatggagc aaatagagca tctaataaag tatgggtaag
18961 tagtaaacaa aaatatatga tgatagacac ttatcataga gtagttaatg ttgttacaga
19021 gttagatgct agatatgagc cttctcctaa ctcaaataga atgggctacg tatgcaatgc
19081 tgaaagagta tatgttcata agatagaagg aaactatgca ttatgtacat attttgcagg
19141 agaaggctat aaaacagcat ggtttacagc taaatattta gaaagaatat aagcaaaagg
19201 gctagatagg aaatcctatc tagccctttt ttgcatttta ggagtgatat ttaaataaaa
19261 gtagtctcta acttatatta aaattataga cataatatat aattttaaac atagtaatca
19321 aaacttttt tacttacaat aatgttatat aatgttaata tattttgttt aggaggattt
19381 tatgagaaag ataatagcaa ttttattaag tgtattagta tttggaagtt tactttgggg
19441 gtgtggagat aaagaagaaa ctactaatct aaatataaat caaaatgata attctatgat
19501 aagccaaagt gaaaaagatt cttataaaga agaggtaaag aaggttttaa atggattgac
19561 tttaaatgat atagagatat taactcaaaa taaaactaaa aaaccaatag taatggttca
19621 agttatttt gatggaacga aaggcaatat aaacaaaagt gaagtagaag aatttataaa
19681 aaccataaaa gaaaaaatag aaacaatttc aaatttatat gatattacaa tattagataa
19741 aaatactcaa atgatagcaa caacaggata tgaaaatgaa ggaaaaatat atttttttga
19801 ctagaataaa gactaataat taaagttaat acatatttaa acgaggtatt ataaaattag
19861 aagagtacta aaaacgtact cttctaataa gaaaactatt gtaatttgtg taaaataaga
19921 aataataata taataaatca tatgttaaca cttaaaataa atagaaatca aattataaga
19981 atagagtggg aataaagtta aaaactttta taaactagat gaatactgag ttttaaaaaa
20041 tttaaaaaac aaaacgtact aaaaacgtac taaattaatt catctagttt atttgatacc
20101 atagtatctt tattaggata taaatgtgag tacgtatcta aagtagtttg tacttttca
20161 tgtcctaatc tttcagaaat taaaagaggg gtaaatccta gttctattaa tagagaagcg
20221 tgagagtgtc ttaaatcatg tattctaata ataggtatat tacttttctt acaaactctt
20281 ttcatttcat ggactagaaa gtactttgta aaactaaaaa ttctatcaga tttattataa
20341 tcatatagtt tagttatata agtttctata taggagcata aagcatttgg tatagaaact
20401 tttcttttac ttttaggagt tttgggagga ttaataatat ctttaccatt taatcttgta
20461 taagtcttat ttatatttat agttttattt tttagatcaa tatcattaaa tgttaaagcc
20521 attaattcac cacttctcat accagtccaa aataaaactt caaaagcaag tttagatatt
20581 ggtttatcat tatattctat aaatgtttta aattgttcag tagtccaaaa cttcatgtca
20641 tcagcatttt tctttcccat acttccagct ctatgacaag gattagtatt taatccataa
```

```
20701 aatttaattc caaaattaaa tatagctgtt aattgattgt ttatagtttt taaataagta
20761 agagagtaat tcatagattt taatgtattt tgccattggc gtatatcatt tggagttata
20821 ttagatataa gtttatcttt gaagtatgga agaatcttta aatttataag atatttttg
20881 ttttctaaag tagtttctct aactcttatt tttgcatctt ctaaataaac ctcagttaat
20941 gttttaaaat ctatatcaga attttcttgt accttaacta aaaattctct ctcaaattta
21001 tcagcttctt ttttagtttt aaatccttct tttttttac gttttctttt accagtccaa
21061 tctttataat aaaatgaaca ataataagtt cctctttctt tatttttata agtaggcata
21121 aatttaattt ctccttaaaa ttttattaac tttagtggtt ctaaatttaa tacgcattct
21181 ttactaacat aataaatccc atatcgctta cgataatatt taatagcttc aaggaaaaaa
21241 tcttttgtta tattaaaata ttcagttatt tcttctgtag aatttactcc ttttagtaat
21301 gcttctatca aagcttcggg agagactaat ctttcatatg cccatgcacg agctttattt
21361 tcttgatgaa tattttctat ttttgaaaga tttgttatat ttccatgatt agtaaaataa
21421 tgtcctattt cttcagctaa tatacagtat tttgtttttt cgggtgaagt ttcatttata
21481 aaaataacat cattacagca atagccacag tcttgaccta atctcatttc aattacttt
21541 acaccataag aggcaatttc ttctataagt tcctcatatt ttgtcatctt ttaacctcta
21601 atatagttta tttgccttta atctttaagt acatctcttt aatttttct atattagctt
21661 tattagtttc ttcatcaagt ccatcactgt gagcagcaag taaatatggt tcaaaatcat
21721 ctatagaagt tgctgaaatt tcattttcac tttgcttctg tataggttca tttataaatt
21781 taggattatc aattaaatct ttagtgtaag taataacttt atttttacct atatcattta
21841 attttcttaa atcttctata tgattttgtt cttcctctga ataattatta gtattactag
21901 atacattttc taatggttca aatgttttta catcaatatt taaaacatca caaatttaa
21961 ttattctatc aactgccatt cctccaatac ttttgtctaa tgcacttgtt aaagttgagt
22021 ttggaattcc tactatttta gaaaattctc ttatactatt atatctactt aaaattaatt
22081 cttttagctt ttttgactta tttaaatttt cgtaagaatt taattcatta tcttcccatc
22141 ccataaggta tgatggagtg gtttgaagaa tttcggctaa aggctcaaga attgatattg
22201 gtaaattttc gatatcatta ctttcatatc tataaacagt agctctattt ttatttaatt
22261 ttttagcgac ttcatctaca gtaagtccca ttgattttct tttttctttt attctttgac
22321 ctatattcat taaaaaacct ccttgtaaac atatattacc tataaattcg cttatatgca
22381 acataaatat taaaaaaaag attttttttg cgaaaaatgc gaaaaatata ttgacaagcc
22441 caaaatagga tgttaccata aaaacaataa atcgcatttt atgcgacaaa aagaggtgag
22501 tatatggtaa atttaaataa gttaaaattg aaaatttttg aaaatggaat gaatgtaagt
22561 caacttgctg aaaagattgg aatggataga tcaacattgt atagaaagtt tcaaaaaaat
22621 ggggataata tttcgataaa agaagctaat ctaatggttg aaactttaca tttaacagtt
22681 gaagaagcga atgctatttt ttttagtaat tatgtcgcat aaaatgcgaa aaagtgaggg
22741 tgagttagat gaataattta gttttaaaag aaaaacaaga atttattgga attcaaattc
22801 cagtaataga aggtggcttt ggagaaaatc aaaaagttat cttaggaaaa acaatagctg
22861 aaatacatga gttaagatta gacaaaatta attcattaat caacaacaac ctagaagagt
22921 ttgaagaagg aatagatatt ttagatttaa aaaactcaaa ccttaccgag gtttcacttt
22981 tagagttagg atttaccaaa caaacaatag caaatagtaa aaatatatat ttactttcag
23041 aacaagggta tatggcttta gttggattaa tgagaacaga taaagctaag gaaataagaa
23101 gaaagttaag aagagaatat tttgctatga aagcagaaat aaaagaaaat agctttgata
23161 tttctgaatt aagtccagaa cttcaaatgt ttaataaaat gcttcaagct gtagcaaaga
23221 gtgagctaga gcaaaagaga ataaagcaac agcttaatga agtaaattat catgctttag
23281 aagctaaaga gaaagctgat aagagtatag aagaaataca agctatgaga gaggttgtgg
23341 ctactaatac aactaattgg agagaagagt gtagaaagat aatagttaaa atagctcata
23401 aattaggagg aaattcctat atacatgatg tgcagacaga agtgtattca cttatgagaa
23461 ctaggttaag atgtaaactt gatgttaagc ttacgaatat gcgtagaaga atggcggaag
23521 aagggatttg taaatcgaag agagataaga agaattattt agatgtaatt gaggatgatg
23581 ctagattgaa agaaggatat attgcaatag taaaagattt agcatttaaa tatggagttg
23641 cttaaaatta ttaaattgat ttaggggggat ggagagaaaa tggcagtaaa aagaacatat
23701 tacacagtag atgacattat gagttttcta ggagttgggc aatctaaggc atatagcata
23761 ataagagaac ttaataatga gttgaaaagt aaagggttta taactattgc aggaagagtt
23821 ccaaaaaagt attttgatga aaaatattat tgttaaaatt agaattttta attgaataat
23881 tgaggttaat ttaggaggta gaatgttaac accgatagga gaagtagttt taggaactat
23941 atctatagca acaacattgt ttttaactgt attcttctta gagaaatacc tggaagagag
24001 aaacagtaag aaaagaacta aatacttaat tttaagtatt gcaaatatat taagtttact
24061 ttttgtatca aatgtaattt aggaggaaag aaatgtttag agtagatttt atagatattg
24121 gagcaagtaa taaaaatttt tctaaatata taaaagaaat aaatcacgaa aatctttatc
24181 atgcagttaa accatattta caaagtagga atatttggtt cacttgtaga gatggagaag
24241 gtgttgtatt tgcaggattt caaactgtag gtaaattcaa agtgattatt gaaatatagg
24301 attcatagga ggattaaagt gttaaaagta gcagtagtag gatttgaaga tagcagaaaa
24361 acatttttgt ataaaactaa attcttagat ttgaaaccag gagattatgt ttgggtaaaa
24421 gattcgaaga atggttttaa gagaggaatt tttcagaatt atacaaacgc tataagtaga
24481 attaggcaag cagaaaattg gattattgat agaaaggata taaaaaatgg atattaatta
24541 tataagcaag aaacaatcag aggaatttat taataattgg ttatctggca atactttacc
24601 attagaaaaa tatatatcat gttatggaac taatcagtac gttgcaatag ataattctac
24661 gaatgagtgt tggacagaag agttcaaaac aaaagaaggt tgtgagaggt atcttttata
```

```
24721 ttttgaagat gtagaagaag taagagcatg ggaagaaaac agattaagaa aaatagagat
24781 aagtatttat ggagtatatt atcttttaat atttagcatg atactagttt tattttactt
24841 attaagaatt taggaggtag ggaaacatga taaatgttga agattatatt ggagcagctt
24901 ataaggtagc atctaaattt tatagtaaga attgggatat gagtattgat gaaattaatt
24961 caattgcatt tttaggatta cttaaagcaa aaaaaagatt tgagtataaa ggtgttgagt
25021 ttgtaacata tgcaatgaaa actattgagt atgagattaa aaatagtctt tatagagata
25081 gaagtaaatt tagaagaaaa aatatagctg gtaaagatgt atatgaaagg atatatacag
25141 attccttaga taaagaaata aattcagcag aaggggttaa gcttattgat tgccttgttg
25201 aagaatctaa tactgatgaa tatgtagaaa atatagattt aaaaatagca ataaataaat
25261 taaatacgta tcaaagaaaa atattaaaga tgctttattt tgatgagaaa actcaaagag
25321 aaacagcaaa aatattaagt gttcatgaaa aaacaatatc aagagaaaag aaaaaaatac
25381 ttaatttatt aagagagaca cttacagcat aagaaggtga agaaatagtg gacaaacttt
25441 taaatatttt aatagagaaa aaaagagagt gcattggtac cgaccaaagt aaccaatgta
25501 ctcaaataat gaatcaatct tattttaaca ttttaaataa gaaatagaaa gagggtgatt
25561 gtatgaaata tacaatacat ggatttcaac agaataaatt aattgaactt ggattagata
25621 ataatgatgc tcttatccta agtgttatta gggatatgta tgctagtcaa tctatagaat
25681 ttcaaattat agatggagaa agatatattt gggtaaatca atcaaattta atattagatc
25741 agataccaat aataggtcat agaacttcat tgttaagaag attaagtaat cttgaagaaa
25801 aagggattat aattagaaaa ctttcacatg aaaaattaaa taaaaaagat aagcaattta
25861 agaaaggaaa cttttgcttt ataaaattaa catcaaaatt agattactta atggagtatg
25921 accttgtagc aaaatgcaac aacccttgta gcaaaatata acaacccctt gtagcaaaat
25981 gcgacaacaa agattctcct ataaaagatt ctaataatat atatagtcaa gttatagatt
26041 accttaattc taaagcagga aaatcattca aatatacaac taaaaaaaca caagatttaa
26101 ttaaggctag attagcagat ggatttaaaa aggaagaatt ctttaaggtt atagataaca
26161 aagttaaaga atggcaaggt actgaatatg aaaaatattt aagaccagaa actctatttg
26221 gaaataaatt tgagggatat ttaaatcaag gagttaataa aaaaagttac gaagaaagaa
26281 aagaaaatag ttcaaaagat agttttgatt actagggggt aaacatggag aaaaattatt
26341 cgattgaaag tgaaaaaata attttgggaa cgataatata agataatgag cttatgttaa
26401 aagctataaa tatactaggt gaaaatgact tctataattc aaaacataaa attatttta
26461 gagctataaa agagctatac aagaataaca taagctttga tttaacaata ttatctgaaa
26521 agttatcaaa tgatattaaa aaacaagtgc ttacattatc agatttaaca gaaatatcat
26581 atcacacatc taggggagca tttgatagtc atttaagttt agttaaagaa aaaagcaaag
26641 agagaaaact aaatgaaact tgcaaaagta tattaacaac accaggaact atagaaagta
26701 aaatagactt tttacaaaat actttattag agctaaattc atcagaagat aaagaaaaag
26761 tttatacaat gagcgaagtt atggaaaaga cttttgaaaa gatagaaaaa gcctttagta
26821 gtaaaagtgg tattactgga ataagtacag gtattagaaa aatagataat gcaactaatg
26881 gtttagagaa aaaagaattt atagtattag gagcaagacc aagtatgggg aagacagccc
26941 tttctctagc aataatggaa aatatacaag gtaatgtttt atatattcaa ttagatatga
27001 gtaaagaagg aatgggacaa aggttattag cttctaactc atatattgaa aatgggaaaa
27061 ttgctagggg aaaattcaat gataatgaaa tgaagagtat tataaatgtc tttgatagat
27121 taagccgaaa agaaaatata tttatttatg aaccagcatc tataacagta aatcaaataa
27181 ggttaattgc aaaacagata aaaataaagc atggattaga tgttattatc gtagatcata
27241 taggaaaaat caaaccaaca actaaaggta ctaaatatga acaatcaagt tatatatcca
27301 atacattaaa atctattgca aaagaattaa atgttgcaat ggttggatta tgtcagcttt
27361 caagggcagc agagcagaga gcagaccata gaccgatgtt atcagattta agagatactg
27421 gaagtataga agaagatgcg gatgtaattg gattactcta tagagatggt tattataggg
27481 caagagaaga taaagaagat atagttggtg atgttttaga aattaatttt gctaaatgta
27541 gaaacggaag aactggggta gtggaattac aatatcattt accaacacaa aggttatcag
27601 aattttagtt gtaaggagat atgaatgtgt gaattatatg aaaatatgtt aaatgaaatt
27661 ttaagtaaac aatatgatat ttgtaaagat aaatggaatt gtttagattg tatttttgaa
27721 aatgaatgtg acaaagtttt tgttagtgat agaccagatt ctatttgtga taaagaaact
27781 tttataaaaa agataaaaga ggaaatgctt agggggtgtg agtgatgact agatatgaaa
27841 aaacaataaa ttcattaatg aaaattagag ataagttttc tataggaagt gtagagtggg
27901 atgaaattga caatggacta caattagtaa ctgctataat ggctgaaaat aaagcattgc
27961 aaaaagctta taacgaaaaa aaagaaatgt tagaaataca tgagtgggct tttgaagggc
28021 ttagacaagc ttacgaaaac agggaggaga aagaactatg agttgtagag catgttatgg
28081 aggtatggaa cgtaaatgta gagtaactgg tgatagatgt gtatttatga ttccagatga
28141 aaaattatgt tatgagcttt tgggagaagg accattggaa tatgaaaaaa atataggcat
28201 tgacttagcc aaaggaaaag ataagacagg tattgttaga aaagttagag gagatttaga
28261 gaatgaataa agtggtttta ataggtaggt taacaaagga tccagaatta agatttacag
28321 ctactaaggg aacagcagtt gcaagattta ctttagctgt aaataggaaa tataaaggtg
28381 aaaatggtga gtatgaagct gactttatac aatgtattgc atatgggaaa ccaggagaaa
28441 taatagcaca atatatgact aaagggagac agttagccat tgtaggaaac ataagaactg
28501 gaagctatga agataaagaa ggaactaaaa gatatacaac aggagtagtt gtagaagggt
28561 ttgattttat tgattatgca aaaagagaaa atgactttgg aggagctttt agtacaccag
28621 acaattcatt taacaatgat gatatggtac caatagatga tggagatgta ccattttaaa
28681 aataactttt caggaggaat taagaatggc aaaatttaaa ataactataa atgaaattgt
```

```
28741 taacttcaat catgaaatga cagttgaggc aaaaagtgaa agtgaattaa ataaggtttt
28801 agacaaaata gaaagagaag ccaattatag agatgatgta gattatattt tagaagagca
28861 tggaattaag atattagatt ttaatgaaga tggaagtgga gaagtgaata tagaagttcc
28921 agacttagag gaagttgaat aggaaaattt gataatggag ttaagagtaa tcaaatatta
28981 ttactagagt aatcatattt gattactcaa taaaaatata tggataagaa atattttgaa
29041 catatagcta acattgaaga attaggaaaa tatcattata aagtattatt gctattaatt
29101 acagatgagt atactcaatc tgatgttaga aaattactta atttaagaaa acagaatgta
29161 aatagagtat ttagagagtt agaaaaatta ggtttgatag aagttaaaga aataaaagga
29221 gctaataaat attataggct agtagatgtt aaaagattaa aactaaatat agttggtcaa
29281 atgaaaatta tttaggagga gttatgtata aaagaaaata taagagtagg caagttgaga
29341 taagagtttt aaatgttgtt gattatatta taaataataa agttactatt agagaagctg
29401 ctaaggtatt tggagtaagt aaaagtacag tttatttaga ttcaacaagt agaattttag
29461 aaataaatcc tcaaaaagct atggaagttg aaaagattat tttacaaaat aaaagtaaaa
29521 gagctatgag aggagtaaaa gcaagaaaaa taaagagttt agataggaca agttaaagat
29581 gatgatataa tacaagatat tattaaatat cttgtattat ttttttgagg tgattttatg
29641 gttatagata aaaagatgtt tagattagca gaagaaagaa tacgtaacta ctatagaaaa
29701 aatgcaaaga taaagagttt aaaagaaaag ataagttttt tagaaaatag tataaaaata
29761 attgaagcta aaatggaaaa taattatttt aatataccat caaatttaaa aagtccagat
29821 tttgtggaga aaatacaaag ttcatgtgtt ggaagtccta tagagcaatc aatgataaag
29881 gtattggaag gctatgaaaa taaaattgca gcatatataa aagaaattgg ttatttagaa
29941 aacttaatat ttgaaataga agaagattat aagattatag aatgtaatat gttaaattct
30001 ttggaagaag aagagagaaa ctttttagaa gaagcatata agaaacgtat acctaattgg
30061 aaacttgcaa ataaatatca tatgagtgag gtttcatgta caagaaagaa aaagagatta
30121 ataagagaaa taattatatg ggagaaagat agattaagtg aaatttagtg aaatattttg
30181 aaataaaatg aaatgaattg aaatgttttg aaataaattg aaaaaagatg ataaatcaca
30241 ttaaaaaaga tgtgagataa tattattgta aaaaaatatt tttttacgct tccttgcaaa
30301 ttatataatg tcagaatgta aagcaaggca caaaagcctt gctaacatgg aaggtctcca
30361 gttgggtgtg caaaaggtgg ttcgaatcca atgccttcct tcaatcgcct tttttcttat
30421 ttaaaatgta aatactctta ataatcctaa gttaagagat tagagaatct agttttacta
30481 gattctcttt ttatttaagg aggtattttt gataatgaaa aaattgttaa ttaatatttt
30541 aggaggagtt acagtaggag caatagctat tgtaagtctt tttggtattg gaaaaatttc
30601 taatatttat gaccgaacga taaataaaga aagtgttagt attcaaagag aaaaatttaa
30661 agagagcaag tcttatgtag aaagtatgat aaatacttta gcagagtata aaagagagta
30721 tgaaagagct aaaactaaag aagagaaaga acaaataaga aacttaatag atagtaagtt
30781 tagtaatttt aatattgatt atatagaaaa taaagattta tataattttt tacaacaaat
30841 gagaggagga aattttaatg aataaaaaaa gagttacagc tattttatta gttctatcgc
30901 taggtatagt aggagttggg tgtggtgaag ttaaagataa aaaagatgta ggttcacaag
30961 aaactataaa tactcagcaa gagcaattac tagaagaatc taatagacaa ttaggtatgc
31021 cagagataac taatttcttt gaaaaggata tggctaaaaa aatatgggag ttaagagatg
31081 accctcaatt aacaacatat gcttatttta agagtatgga tggaaaattt gtatacttag
31141 gtagatgtat aggatttgga cttccatata ctacacaata tactaatcca gaaaagatgg
31201 atactgtaaa aggtggagaa tatgctagga atccatatgt tttaccacaa gcagacccta
31261 atggattatt tagttcagaa actacgaatg ctacttggtt aatggtagtt aatgaagaaa
31321 caaatgagcc agagataatg tatagtgagc ctgaaataat agttacccaa agtaaactta
31381 atagaagatt agttgaggaa tggagcttac cagctaatta ttaagatatg aattatgttg
31441 aaccaataag aaattctagt gatgttaatg cgttagcaaa taagttaaga aaagatagtg
31501 aacgaaacta tattatgttt attattggga tatattctgg tttgagaata agcgatatat
31561 tgagattaag agttagagat gttaagggaa aagaacaaat tgccattaga gagaagaaaa
31621 cagggaaaca aaaaatgttc cctattaatc caatattgaa aaaagaattg aaaaaatatt
31681 gttcggataa ggagttagat gaattcctta taaaatcaag aaatggatat aatagagcaa
31741 taggcagaga aagagcgtat acaattctta gagaagctgg tgaagaatta ggaatatata
31801 atttggcaac tcacacttta agaaaaacat ttggatatca ttattatatg caatataacg
31861 atatagctga actacaaaaa atatttaacc attctgatcc agctgtaact ttacgttata
31921 taggtgttga gaaaagtaga atagatagaa gaatcaagag ttttaaaata tagcgaacac
31981 aacatattta gcgtacgttg tgttcaaaat taataatgat taagaagtaa gatatatcaa
32041 tggttataaa agattggtga acacaacaga agttaaatta tgttgtgttc aatgcaagaa
32101 gaaaattata aaattgcaaa aaaaaaatat atatctttta aaggagttaa aaatgaattt
32161 agaagaagta aaagaaagca gattaaatga acttgagaaa aaatataaac ctataattga
32221 agcttataat gatactgttg aaatggttaa tagtatagaa ataaaagata atgatagtat
32281 tgatattata atttgtaaat tatacatatt aactaataat gttaatgtta cattaaaact
32341 tctatcaaaa tctgaatatc gtattaacaa tagaaaagtt aattcttctg atatatccga
32401 aatattaaat agtatttcaa aaaaagaaac taatcaaatt aaagcttttg caaagaaaca
32461 gtttttaaat aataagaaaa agtcaataaa gctttgttga taactaagtt taaatttaaa
32521 ggagttgttt ttatggaaaa aaataaagag tttttaagag ttagagatat atttagagaa
32581 tgtgcagata ttatggataa ggtaattgat ttagaaaaaa gagaggaaaa aggggaagat
32641 gttactccag aaactgagag gttaatgggt agatatatga tgctattaat ggaattaaat
32701 agtctaacta ataattaata ggaggaagga cattatgaag aaaactaagt atagatgttg
```

```
    32761 taattgtgga ttggaatatg agtttgaagg aaaagctaaa gaggggttat catgtgaaat
    32821 atgtacagaa ccattacgcc ctttagttca agagagagaa ttaaaggata agagaagaaa
    32881 gttgattaaa tctgttacga aagatttttt atcacaagaa atgaagttag gaataagaga
    32941 gtttatgagc cttattaatg aaatagagag tttagagaaa gaattagaga agaataagaa
    33001 aataatagca aaactaaaga aagcactgag gaaaagaaat gccaaagaaa gtatgtaact
    33061 atagtggatg tactaatctt ataagtgtaa aagaaactta ttgtagcaag catagagtag
    33121 agagcaataa ggaaagacat aagtgctata aggttaggag gaaggacaaa gaggaacagg
    33181 ctttttataa ttctaaagag tggcgtatag ttagaggttt tgttattaat agagataatg
    33241 ggttatgtaa gttgtgttta agtaaaaaga aaataagagt tgctgatgtt gttcatcatg
    33301 taatagaact aaaagaaaga agggagcttg ggctagagac aggtaacctt ctttctttat
    33361 gtaattcatg tcatcaagaa atacactcta agtataacaa gggtatggtt actaaaaaga
    33421 atactcaaga tgaattgttt aagctaattc ctaatgggga ggggtagggc aaaaagtttt
    33481 gggcaagagc caaaaggtcg cagtgtc
//
```

ACKNOWLEDGMENTS

The research work resulting in this PhD thesis was performed at the Institute of Microbiology, Research Centre for Milk and Food Weihenstephan, Technische Universität München under the supervision of Prof. Dr. Siegfried Scherer.

Whenever you start to work scientifically you have to rely on the professional as well as moral support of colleagues, professors, family and friends. Therefore I would like to thank heartily all my colleagues at the institute for always supporting and encouraging me during the time of my work at the FML.

I am very grateful to my supervisor Prof. Dr. Scherer for offering me the interesting PhD project and giving me the opportunity to work at his institute. I want to thank him for his continuing encouragement, his interest and support.

My sincere thanks goes to my "personal" supervisor PD Dr. Martin J. Loessner. His affection to bacteriophages and their potentials is highly contagious! I want to thank him for his encouragement, his interest and his support, both as a supervisor and as a colleague.

A big "Thanks" goes to Nataša Vukov, for her expert assistance in the holin part and the constructive discussions on this topic. I want to thank Waltraut Knapp for her professional assistance with all aspects of electron microscopy, and Simon Maier for the introduction into anaerobic techniques and Ingo Krause for the densitometrical scanning. I also want to thank Peter Pfaller for supplying the *C. perfringens* strains from human isolates. I am most grateful to Andrew Morgan for continuous encouragement and support of this work, and to Roger Hendrix for his help and many useful tips in bioinformatic analysis.

A very heartfelt "Thank you" goes to my wife Josi, for loving me and understanding that the hours spent in the evening in the lab or the office were, at least most of the time ;-), necessary. A little "Thanks" goes to my yet unborn child for motivating me getting the last bits and pieces together to finish this PhD thesis.

CURRICULUM VITAE

MARKUS ZIMMER

PERSÖNLICHE ANGABEN

- Familienstand: verheiratet
- Staatsangehörigkeit: deutsch
- Geburtsdatum: 14.01.1971
- Geburtsort: Gerolstein

AUSBILDUNG

09/1981 - 05/1990	Staatl. neusprachliches Sankt-Matthias-Gymnasium, Gerolstein *Abschluss mit der Allgemeinen Hochschulreife*
10/1991 - 9/1997	Rheinische Friedrich-Wilhelms-Universität, Bonn *Studium der Biologie*
1/1995 - 12/1995	University of New South Wales, Sydney (Australien) *wissenschaftliche Aus- und Fortbildung im Ausland*
09/1996 - 09/1997	Institut für Mikrobiologie & Biotechnologie, Bonn *Diplomarbeit bei Prof. EA. Galinski,* *Prof. Dr. Dr. H. G. Trüper, mit Abschluss Diplom-Biologe (Dipl. rer. nat.)*
05/1998 - 02/2002	Institut für Mikrobiologie, Forschungszentrum für Milch und Lebensmittel Weihenstephan, TU München *Anfertigung der Dissertation, Prof. S.Scherer, PD Dr. M. Loessner*