H24/3230

MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
ON..... 7 June 2002 .
..... [redacted] .

Sec. Research Graduate School Committee

Analysis of Clostridial MLS Resistance Determinants

Kylie Ann Farrow B.Sc. (Hons.),
Bacterial Pathogenesis Research Group,
Department of Microbiology,
Monash University.

**Thesis submitted for the degree of Doctor of Philosophy,
Monash University,
Melbourne, Australia**

November 2001

Table of Contents

Johnson, S., Samore, M.H., Farrow, K.A., Killgore, G.E., Tenover, F.C., Lyras, D., Rood, J.I., DeGirolami, P.C., Baltch, A.L., Rafferty, M.E., Pear[illegible] M. and Gerding, D.N. (1999). Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. *N. Engl. J. Med.* 341 1645-1651.

Farrow, K.A., Lyras, D. and Rood, J.I. (2000). The macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium difficile* 630 contains two *erm*(B) genes. *Antimicrob. Agents Chemother.* **44,** 411-413.

Farrow, K.A., Lyras, D. and Rood, J.I. (2001). Genomic analysis of the erythromycin resistance element Tn*5398* from *Clostridium difficile*. *Microbiology* **147,** 2717-2728.

List of Figures

Chapter Four

Chapter Five

Chapter Six

List of Tables

Summary

Bacterial resistance to the macrolide-lincosamide-Streptogramin B (MLS) group of antibiotics is often mediated by *erm* genes, which encode rRNA methyltransferases that methylate the ribosomal RNA target of these antibiotics. In the pathogenic clostridial species, *Clostridium difficile* and *Clostridium perfringens*, MLS resistance is due to the presence of *erm* genes that belong to the Erm B class of Erm determinants.

In *C. difficile* erythromycin resistance has been shown to be transferable to *C. difficile* recipients and also to *Staphylococcus aureus* and *Bacillus subtilis*. Transfer of resistance occurs in the absence of detectable plasmid DNA and it has therefore been postulated that the Erm B determinant from *C. difficile* resides on a conjugative transposon, designated Tn*5398*. Tn*5398* was cloned from *C. difficile* strain 630. Sequence and genetic analysis of this element revealed that it was approximately 9.6 kb in size and did not appear to encode proteins that are typically involved in conjugative transposition, such as transposases, integrases, resolvases, or mobilization and transfer proteins. Analysis of the nucleotide sequence of Tn*5398* did, however, reveal the presence of two *oriT* sites that have similarity to the *oriT* sites on the conjugative transposons Tn*916* and Tn*5397*. Based on this analysis it is concluded that Tn*5398* is not a conjugative transposon but may be a mobilizable element that is transferred using proteins provided *in trans* by Tn*5397*, which is present in the same *C. difficile* strain, or by genes that are host encoded.

Sequence analysis of Tn*5398* also revealed that it carries a novel Erm B determinant. This determinant consists of two identical *erm*(B) genes, which are separated by a single direct repeat (DR) sequence and are flanked by variants of the

DR sequence. This is the first known Erm determinant that contains a duplicated *erm* structural gene.

The prevalence of the novel Erm B determinant and Tn*5398* in various *C. difficile* strains was investigated. A PCR and DNA hybridization based strategy was used to determine the genetic organization of the *erm*(B) gene region, and to detect the presence of Tn*5398*-like sequences, in 27 erythromycin resistant *C. difficile* isolates from different geographical origins. The results showed that there is considerable heterogeneity in the arrangement of the *erm*(B) gene region in *C. difficile* isolates and, furthermore, that not all *erm*(B) gene regions are associated with Tn*5398*-like elements. Tn*5398*-like elements were detected in three isolates; these elements appear to carry two different *erm*(B) gene regions.

In *C. perfringens*, the *erm*(B) gene is located on a large non-conjugative but mobilizable plasmid, pIP402. The *C. perfringens* Erm B determinant has been studied extensively in terms of its sequence and genetic organization. This thesis presents a mutational analysis of the *erm*(B) gene from *C. perfringens* with the aim of defining structurally and/or functionally significant amino acid residues of the Erm(B) protein. Random mutagenesis identified several residues that, when mutated, either significantly reduced or completely abolished erythromycin resistance. These residues were located in conserved methyltransferase motifs and structural modelling of the Erm(B) protein revealed that these amino acids are likely to be important in either the binding of the substrate *S*-adenosylmethionine (SAM), the transfer of the methyl group from SAM to the target rRNA residue, or in the binding of the rRNA target. These residues may represent good targets for the development of rRNA methyltransferase inhibitors.

List of Publications

The following publications have arisen from the research presented in this thesis:

Johnson, S., Samore, M.H., Farrow, K.A., Killgore, G.E., Tenover, F.C., Lyras, D., Rood, J.I., DeGirolami, P.C., Baltch, A.L., Rafferty, M.E., Pear, S.M. and Gerding, D.N. (1999). Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. *N. Engl. J. Med.* **341,** 1645-1651.

Farrow, K.A., Lyras, D. and Rood, J.I. (2000). The macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium difficile* 630 contains two *erm*(B) genes. *Antimicrob. Agents Chemother.* **44,** 411-413.

Farrow, K.A., Lyras, D. and Rood, J.I. (2001). Genomic analysis of the erythromycin resistance element Tn*5398* from *Clostridium difficile*. *Microbiology* **147,** 2717-2728.

The following conference abstracts have arisen from the research presented in this thesis:

Farrow, K.A., Lyras, D. and Rood, J.I. (1998). The Erm determinant from *Clostridium difficile* strain 630 contains two *ermBZ* genes. *1998 Annual Scientific Meeting & Exhibition of the Australian Society for Microbiology*, Hobart, Tasmania, Australia, p. A85.

Farrow, K.A., Polekhina, G., Lyras, D., Koutsis, K., Parker, M.W. and Rood, J.I. (1999). The ErmBP 23S RNA methyltransferase from *Clostridium perfringens* has conserved motifs containing functionally important residues. *5th Australian Conference on Molecular Analysis of Bacterial Pathogens*, Victor Harbor, South Australia, Australia, p. 39.

Rood, J.I., Lyras, D., Farrow, K.A. and Crellin, P.K. (1998). Antibiotic resistance and resistance gene transfer in *Clostridium*. *38th Interscience Conference on Antimicrobial Agents and Chemotherapy*, San Diego, California, U.S.A., p. 637.

Rood, J.I., Lyras, D., Johanesen, P.A. and Farrow, K.A. (1999). Molecular epidemiology of Clostridial antibiotic resistance determinants. *IXth International Congress of Bacteriology and Applied Microbiology*, Sydney, New South Wales, Australia, p. 6.

Declaration of Authenticity

I, Kylie Ann Farrow, declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to the best of my knowledge and belief contains no material which has been previously published or written by another person, except where due reference is made in the text of this thesis.

The hydroxylamine mutagenesis of pJIR418 (Chapter Three) was performed by Ms. Katerina Koutsis and I estimate this contribution as 10% of the work presented in this chapter. The structural modelling of the Erm(B) protein (Chapter Three) was performed by Dr. Galina Polekhina and Dr. Michael Parker, Biota Structural Biology Laboratory, St Vincent's Institute of Medical Research and The University of Melbourne. These collaborators also produced several of the figures (Figures 3.10, 3.11, 3.12 and 3.13) in this chapter and gave valuable theoretical input to the analysis of the mutant proteins. I estimate their contribution to be 20% of the work presented in Chapter Three.

[redacted]

Kylie A. Farrow

Acknowledgments

The successful completion of a Ph.D is both a professionally and personally challenging exercise that allows you to meet many people who either expand your mind or warm your heart. During my candidature I have been lucky to be associated with many such people.

I would firstly like to extend my sincere thanks and gratitude to my two supervisors, Professor Julian Rood and Dr. Dena Lyras. I thank you Julian for allowing me to undertake my Ph.D. in your laboratory, which provided an environment that was both scientifically challenging and also personally rewarding. I thank you also for your meticulous scrutiny of my research and for helpful scientific discussions, which have helped me mature as a scientist. Thank you for your patience and understanding in the latter stages of my candidature when circumstances began to overwhelm me. Special thanks to my co-supervisor Dena who is not only a reliable scientific sounding board but also an exceptional friend and confidant. I do not have the right words to say how much I appreciate your support, advice and friendship. Let's just say that my ship may have sunk if not for you.

So many people made my time in the laboratory enjoyable that I'm sure I'll forget at least a few, so I apologize in advance to anyone I may forget. To the current members of the extended Rood Laboratory, you are all my second family and I am greatly appreciative to you all for one reason or another. To Drs. Ruth Kennan and Trudi Bannam thank you both for your advice on how to survive a Ph.D. and come out again on the other side with at least the resemblance of a life, and also for your encouraging words when the chips were down. To my fellow Ph.D. students Sheena McGowan, Lisa Stubberfield, Fiona Glenister, Ishara Gunesekere, Angela Cox and Dr. Grant Jenkin, thank you all for sharing my frustrations and joys and for being around in the early mornings and late evenings when I was still here writing away. Special thanks go to my close friends Vicki Adams, Dr. Jackie Cheung and Milena Awad for being such wonderful people, and helping me in so many ways, both professionally and personally, that I cannot list them. Please know that your kindness and support have not gone unnoticed and I hope that one day I may be as good a friend to you all as you have been to me. Special thanks also go to Lina Laskos who has shared the writing up experience with me. You have often allowed me to see that my problems are not insurmountable, and that good humour can most

definitely conquer writer's block! To Dr. Isabelle Lucet I extend a huge thank you for allowing me to be so possessive of our office computer, and for being such a good friend over the last year. To Pauline Howarth I extend sincere thanks for her excellent technical assistance, without you Pauline, the lab would be in total chaos. To Soula Krejany thank you for understanding the stresses of the writing up period, and knowing that sometimes it's just better not to ask.

To the members of the extended Rood laboratory (Drs. Andrea Huggins, Joanne Johnston, Paul Crellin, Meredith Hunt, Chris Hamilton, Karena Waller and Tim Stinear) who have moved on to bigger and better things, thank you all for your support and advice during various stages of my Ph.D. candidature. Very special thanks go to my good friend Dr. Priscilla Johanesen for being a source of inspiration and courage when I had very little of either left. I would also like to thank the many honours students (Louise Hilton, Melissa Bateman, Chelsea Salvado, Dianne Abrahams, Jodie Nicholls, Tash Pincus, Kay Burdett, David Wong, Sacha Pidot, Dane Parker and Jenny O'Connor) that I have had the pleasure of being associated with during my candidature. I enjoyed interacting with you all in various capacities and hope that your careers develop the way you envisage.

During my Ph.D. I taught many undergraduate students, which I feel has improved my communication and teaching skills. I would like to extend many thanks to Mohamed Mohideen, Dr. Harley Dean and Lyn Howden for giving me the opportunity to be a part of the sessional teaching staff, and also for their support and encouragement. I would also like to thank the administrative staff, Margaret Dooley, Lesley Gladman and Beth Harrison, who tackled the tough task of making sure that I got paid, and chasing up my pay when I didn't. Special thanks to Tony Stott for his advice and assistance on all things computer related.

I would also like to extend special thanks to the staff of the oligonucleotide and sequencing facilities, Mark Cauchi, Heidi Dryburgh, Elise Webber, Vivien Vasic, Khim Hoe and Saw Eng Tan for their efficiency and diligence. Without their services sequencing 15 kb worth of DNA would have been very difficult!!

To my special friends in the Enrolments division of Staff and Student Services, John Paduano and Anna Carusello, I thank you both for being such wonderful support over the many years that I have had the pleasure to work with you. I'm sorry that

now my time as a student is over I will no longer be able to work with you during the summer period. Thank you for allowing me to have a casual job that was so much fun and enabled me to pay my student fees at the same time.

During the course of my Ph.D. I had the pleasure of having several profitable collaborations and scientific discussion. I would like to thank my collaborators, Dr. Galina Polekhina and Dr. Michael Parker, for their assistance with the protein structural modelling and for opening my mind to the world of protein biochemistry. I thank Dr. Stuart Johnson for giving me the opportunity to break into true medical research while working with some of the "big guns" in the field. I also thank Dr. Peter Mullany and Adam Roberts for providing Tn*5397* sequence data and also for many helpful scientific discussions.

A Ph.D. is not achievable without a wonderful support network at home as well as at work. I would like to thank my wonderful friends Rachel Alway and Donna Tymensen for their encouragement and support during my Ph.D. candidature. No matter how small the achievement you were both always so enthusiastic. I also thank you both for pulling me out of the Ph.D. world for brief periods, for lunch, coffee or dinner, so that I remained somewhat in touch with reality.

Last, but certainly not least, I would like to thank my family for their encouragement, unconditional love and support during this tumultuous time of my life. To my partner, Ian Wilson, thank you for your endless support both personally and professionally, and for not complaining on those many nights you were left to fend for yourself while I slaved at the computer. You mean the world to me.

To my Mum, thank you for your unwavering support and encouragement, and for getting me where I am today. The last six months have been difficult for us all but you still, as always, are more concerned for your children than for yourself. Thank you for everything. To dear Dad, although you will never read this, I thank you for the thirst for knowledge that you instilled in me from the time I was a small child. Your intelligence and wit were the qualities I admired in you most, and I think I have inherited them from you. Thank you for all that you have given me, and for the wonderful memories I have. I'm only sorry I didn't finish "the damn book" soon enough for you to read it. I hope, wherever you are, that you are proud of me.

To Melissa, Nathan and Angela, Jason and Megan, thank you all for your love and support over the many years it has taken me to finish this epic. You all always ask "How's it going?" well the answer can now be "It's finished!" instead of "Don't ask". I thank you all for understanding that I could not always be as close to you all as I would have liked to have been.

To my extended family, Grannie, Grandma and Grandpa, Craig, Belinda and kids, Glenda, Ken and kids, Brian, Gayelene and kids, Valerie, John and Marg thank you for your words of encouragement and support. I hope now that this chapter of my life is complete that I will be able to spend more time with you all.

Dedication

Dedicated to the loving memory of my father, Robert George Farrow (1948-2001), whose thirst for knowledge, intelligence and quick wit will be with me always.

List of Abbreviations

A	deoxyadenosine
AP	ammonium peroxodisulfate
Ap	ampicillin
~	approximately
ATP	adenosine triphosphate
BHI	brain heart infusion
BHIS	brain heart infusion with iron sulfate and L-cysteine
bp	base pair(s)
BSA	bovine serum albumin
C	deoxycytosine
°C	degrees Celsius
CAT	chloramphenicol acetyltransferase
Cbm	carbomycin
CDAD	*C. difficile*-associated disease
Cm	chloramphenicol
cm	centimetre(s)
CO_2	carbon dioxide
CTn	conjugative transposon
Δ	denotes a deletion in genotype descriptions
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphae
dGTP	deoxyguanidine triphosphate
dH_2O	distilled water
DIG	digoxigenin
DIG-dUTP	digoxigenin-11-dUTP
DTT	1,4-dithiothreitol
dTTP	deoxythymidine triphosphate
dUTP	deoxyuracil triphosphate
DNA	deoxyribonucleic acid
dNTP(s)	deoxynucleotide triphosphate(s)
DR	direct repeat
EDTA	ethylenediaminetetra-acetic acid

Em	erythromycin
FMN	flavin mononucleotide
FTG	fluid thioglycollate
G	deoxyguanidine
g	gram(s)
g	gravitational constant
H_2	hydrogen
h	hour(s)
HCl	hydrochloric acid
6 x His	hexahistidine
His	histidine
His-Erm(B)	N-terminal histidine-tagged Erm(B)
HRP	horse radish peroxidase
Ω	denotes an insertion event in genotype descriptions
IPTG	isopropyl-β-D-thiogalactoside
kb	kilobase pair(s)
kDa	kilodalton(s)
Kn	kanamycin
kV	kilovolt(s)
LB	Luria-Bertani medium
M	molar
μFD	microfarad(s)
MFS	Major Facilitator Superfamily
μg	micrograms(s)
mg	milligram(s)
MIC	Minimum Inhibitory Concentration
min	minute(s)
μl	microlitre(s)
ml	millilitre(s)
MLS	macrolide-lincosamide-Streptogramin B
μm	micrometre(s)
μM	micromolar
mm	millimetre(s)
mM	millimolar
MOPS	3-[N-Morpholino]propane sulfonic acid

mRNA	messenger ribonucleic acid
#	number
N_2	nitrogen
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanogram(s)
nm	nanometre(s)
NMR	nuclear magnetic resonance
ORF(s)	open reading frame(s)
ori	origin of replication
%	percent, percentage
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Brookhaven Protein Databank
PEG	polyethylene glycol
PFGE	pulsed-field gel electrophoresis
pH	$\log_{10}[H^+]$
PMC	pseudomembranous colitis
PP	protein purification
R	resistant
RBS	ribosome binding site
Rif	rifampicin
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal ribonucleic acid
RP	reverse primer
S	sensitive
SAM	*S*-adenosylmethionine
SAH	*S*-adenosylhomocysteine
SDS	sodium dodecyl sulphate
SEB	sucrose electroporation buffer
sec	second(s)
SOB	tryptone-yeast extract-NaCl
SOC	SOB + glucose

SSC	salt sodium citrate buffer
SQ	sequencing
*	stop codon
T	deoxythymidine
TAE	Tris acetic acid EDTA
TBS	Tris buffered saline
TE	Tris EDTA
Tel	telithromycin
TEMED	*N*, *N*, *N'*, *N'*-tetramethylethylenediamine
Tn	transposon
TNE	Tris NaCl EDTA
TPG	trypticase-peptone-glucose
Tris	Tris[hydroxymethyl]aminomethane
tRNA	transfer ribonucleic acid
$Tween_{20}$	polyoxyethylenesorbitan monolaurate
Tyl	tylosin
UP	universal primer
U.S.A	United States of America
UV	ultraviolet
V	volts
v/v	volume per volume
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YT	yeast extract tryptone

CHAPTER ONE

INTRODUCTION

Bacterial infections have caused substantial morbidity and mortality for thousands of years. Only in the last fifty years have we had available effective drugs, antimicrobial agents that have had a major impact in treating many of these infections. However, almost as soon as antimicrobials had been developed, bacteria began developing resistance mechanisms to combat these drugs. Bacterial resistance has developed to virtually all antimicrobial agents (Collignon, 1997) and many organisms have become resistant to more than one antimicrobial agent, presenting a real dilemma for disease treatment.

Over recent years this problem has become of great concern due to the discovery that bacteria are able to acquire exogenous genetic material, either from the environment or from other bacterial species, which enables them to rapidly exhibit resistance. Exchange of genetic information between bacteria by transformation, transduction and conjugation allows them to take up segments of genomic DNA, which enables them to express antimicrobial resistance genes from other environmental sources. This discovery has highlighted the need for close monitoring of resistant bacteria and for detailed study of resistance mechanisms. This study investigates a resistance mechanism utilized by two clostridial species, *Clostridium difficile* and *Clostridium perfringens*, to evade the antimicrobial action of the macrolide-lincosamide-Streptogramin B (MLS) group of antibiotics.

MLS resistance

Macrolide, lincosamide and Streptogramin B antibiotics are chemically distinct but have a similar mode of action. These antibiotics all have target sites on the 50S ribosomal subunit and inhibit protein synthesis by stimulating dissociation of the peptidyl-tRNA molecule from the ribosomes during elongation, resulting in chain termination and a reversible stoppage of protein synthesis (Cocito *et al.*, 1997; Weisblum, 1998). They have a narrow spectrum of activity that includes Gram-positive cocci (particularly staphylococci, streptococci and enterococci) and rods, and Gram-negative cocci (Leclercq and Courvalin, 1991). These drugs are also potent against anaerobic bacteria.

Macrolides (e.g. erythromycin) are composed of a minimum of two amino and/or neutral sugars attached to a lactone ring of variable size. Macrolides that are commercially available can be divided into 14-, 15-, and 16-membered lactone ring molecules, with each class differing in their pharmacokinetic properties and in their responses to bacterial resistance mechanisms (Leclercq and Courvalin, 1991). Lincosamides (e.g. lincomycin and clindamycin) are alkyl derivatives of proline and are devoid of a lactone ring. Streptogramin antibiotics (e.g. pristinamycin and virginiamycin) can be classified as A and B compounds according to their basic primary structure (Cocito, 1979). Compounds of the A and B groups bind different targets in the peptidyltransferase domain of the 50S ribosomal subunit and inhibit protein elongation at different steps. When used separately the A and B compounds are bacteriostatic, however, when used in combination they act synergistically and are bactericidal (Allignet and El Solh, 1997). Streptogramins are used in clinical practice only in a few countries, including Belgium and France.

Three mechanisms account for acquired resistance to MLS antibiotics, direct inactivation, active efflux, and modification of the target. The first two mechanisms confer resistance to structurally related antibiotics only, while for the third mechanism a single alteration in the 23S rRNA molecule confers broad cross-resistance to macrolides, lincosamides and Streptogramin B-type antibiotics, giving what is called the MLS cross-resistance phenotype.

Antibiotic inactivation

Several mechanisms, which usually confer resistance to only one of the three types of antibiotics (macrolides, lincosamides or streptogramins), have been described (Weisblum, 1998). The proteins involved in this type of mechanism inactivate the antibiotic rather than modifying the antibiotic target site.

a) Macrolide inactivating enzymes

Two main types of macrolide modifying enzymes have been identified, the erythromycin esterases and the macrolide phosphotransferases, each of which inactivate the lactone ring of 14- and 16-membered macrolides.

i) Erythromycin esterases

Two types of erythromycin esterases, type I encoded by the *ere*(A) gene and type II encoded by the *ere*(B) gene, were first described in *Escherichia coli* (Arthur *et al.*, 1986; Ounissi and Courvalin, 1985). Ere(A) and Ere(B) have esterase activity and are capable of hydrolyzing the lactone ring of 14- and 15- membered macrolides. Homologous genes have subsequently been detected in many members of the *Enterobacteriaceae* family (Arthur *et al.*, 1987) and also in *Staphylococcus aureus* (Schmitz *et al.*, 2000; Wondrack *et al.*, 1996).

ii) Macrolide 2'-phosphotransferases

Macrolide 2'-phosphotransferases inactivate macrolide antibiotics by phosphorylation of the 2'-OH group of the sugar moiety that is bound to the macrolide ring in the presence of ATP (Noguchi *et al.*, 1996). Two types of phosphotransferase, I and II, have been found in erythromycin resistant *E. coli*. Macrolide 2'-phosphotransferase I, encoded by the *mph*(A) gene, is more active against 14-membered macrolides and is inducibly expressed (Noguchi *et al.*, 1996), while macrolide 2'-phosphotransferase II, encoded by the *mph*(B) gene, is constitutively expressed and is active against both 14- and 16- membered macrolides (Noguchi *et al.*, 1996). The *mph*(B) gene has also been detected in *S. aureus* (Sutcliffe *et al.*, 1996). More recently, a third type of phosphotransferase, encoded by the *mph*(C) gene, has been detected in *S. aureus* (Matsuoka *et al.*, 1998) and *Stenotrophomonas maltophila* (Alonso *et al.*, 2000), however, the intracellular enzymatic activity of Mph(C) has not yet been reported.

b) Lincosamide inactivating enzymes

Specific resistance to lincosamide antibiotics is generally mediated by bacterial modification of these agents. Phosphorylation (Argoudelis and Coats, 1969) and nucleotidylation (Argoudelis *et al.*, 1977; Marshall *et al.*, 1989) of lincosamide molecules have been detected in several species of *Streptomyces*. Inactivation of lincosamides has also been observed in strains of staphylococci, streptococci, enterococci, and lactobacilli of animal origin (Devriese, 1980; Dutta and Devriese, 1981; Dutta and Devriese, 1982) and in staphylococci isolated from humans (Brisson-Noël *et al.*, 1988; Leclercq *et al.*, 1987; Leclercq *et al.*, 1985). Clinical isolates of *Staphylococcus haemolyticus* BM4610 and *S. aureus* BM4611 are highly resistant to lincomycin. In these strains lincosamide *O*-

nucleotidyltransferases encoded by two closely related *lnu*(A) genes (Roberts *et al.*, 1999) have been characterized (Brisson-Noël and Courvalin, 1986; Brisson-Noël *et al.*, 1988). The enzymes encoded by these genes inactivate lincomycin and clindamycin by converting them to lincomycin 3-(5'-adenylate) and clindamycin 4-(5'-adenylate) by using ATP, GTP, CTP, or UTP as a nucleotidyl donor and $MgCl_2$ as a co-factor (Brisson-Noël *et al.*, 1988). Recently, the *lnu*(B) gene from *Enterococcus faecium* has also been characterized (Bozdogan *et al.*, 1999). This gene also encodes an *O*-nucleotidyltransferase that inactivates lincomycin and clindamycin, however, the amino acid sequence of Lnu(B) is significantly different from that of Lnu(A).

c) Streptogramin inactivating enzymes

Enzymes that modify the streptogramin group of antibiotics generally only confer resistance to one component such as Streptogramin A, but not Streptogramin B, or *vice versa*. Enzymes that hydrolyze Streptogramin B or modify Streptogramin A by adding an acetyl group have been described. Many of the genes encoding these enzymes are plasmid borne and are often found in pairs that are capable of inactivating both types of streptogramins (Roberts *et al.*, 1999).

i) Streptogramin B hydrolases

Streptogramin B hydrolases, or lactonases, are capable of cleaving the macrocyclic lactone ring structure of type B streptogramins and have been identified in *Actinoplanes* (Hou *et al.*, 1970) and *Streptomyces* (Kim *et al.*, 1974). Two genes, *vgbA* (Allignet *et al.*, 1988) and *vgbB* (Allignet *et al.*, 1998), which encode Streptogramin B lactonases, have also been identified in staphylococci.

ii) Streptogramin A acetyltransferases

Streptogramin A acetyltransferases inactivate this class of antibiotic by adding an acetyl group. Genes encoding these enzymes (*vat* genes) have been detected in a wide variety of bacterial species. The *vat*(A) (Allignet *et al.*, 1993), *vat*(B) (Allignet and El Solh, 1995) and *vat*(C) (Allignet *et al.*, 1998) genes have all been isolated and characterized from staphylococcal strains, while the *vat*(D) (Rende-Fournier *et al.*, 1993) and *vat*(E) (Werner and Witte, 1999) genes have been isolated and characterized from enterococcal strains. The staphylococcal genes are all located on plasmids, however, genes with homology to the *vat* genes from staphylococci and enterococci have recently been detected in the chromosome of *Yersinia enterocolitica* and *Synechocystis* spp. and in the unfinished genome sequences of *Shewanella putrefaciens* and *Pasteurella multocida* (Seoane and García Lobo, 2000).

Active efflux of antibiotics

Several different antibiotic resistance genes code for transport or efflux proteins, which effectively remove MLS antibiotics from the bacterial cell. These proteins do not modify either the antibiotic or the antibiotic target, but instead pump the antibiotic out of the cell, keeping intracellular concentrations low and the ribosomes free from antibiotic. Most of these proteins have homology to the major facilitator superfamily (MFS) or are putative members of the ABC transporter superfamily (Roberts *et al.*, 1999).

Four different efflux systems that confer resistance to MLS antibiotics have been described for Gram-positive organisms. The first of these systems, Msr(A), is responsible for the active efflux of macrolide and Streptogramin B antibiotics. The

msr(A) gene, a member of the ATP-binding transport supergene family, was first identified in *Staphylococcus epidermidis* and confers the so-called MS phenotype, resistance to 14- and 15- membered ring macrolides and Streptogramin B after induction with erythromycin (Ross *et al.*, 1990). This gene has subsequently been detected in other species of both coagulase positive and negative staphylococci (Eady *et al.*, 1993; Lina *et al.*, 1999; Nawaz *et al.*, 2000). Several homologues of *msr*(A) have also been characterized, including *msrB* from *Staphylococcus xylosus* (Milton *et al.*, 1992), *msrSA* and *msrSA'* from *S. aureus* (Matsuoka *et al.*, 1998; Matsuoka *et al.*, 1999) and *msrC* from *E. faecium* (Portillo *et al.*, 2000). Due to the high level of homology between these genes they have all recently been renamed as *msr*(A) (Roberts *et al.*, 1999).

The second efflux system is that of Mef(A), which is responsible for the active efflux of macrolide antibiotics. The *mef* genes have been found in a variety of Gram positive genera, including corynebacteria, enterococci, micrococci and streptococcal species (Fraimow and Knob, 1997; Kataja *et al.*, 1998; Luna *et al.*, 1999; Shortridge *et al.*, 1996). Many of these genes are associated with conjugative elements located in the chromosome, which are readily transferred across species and genus barriers (Kataja *et al.*, 1998; Luna *et al.*, 1999). Two *mef* genes have been characterized, *mefA* from *Streptococcus pyogenes* (Clancy *et al.*, 1996) and *mefE* from *Streptococcus pneumoniae* (Tait-Kamradt *et al.*, 1997), both of which have been renamed *mef*(A) (Roberts *et al.*, 1999).

In addition to the Msr(A) efflux pumps, two efflux systems, Vga(A) (Allignet *et al.*, 1992) and Vga(B) (Allignet and El Solh, 1997), have been identified in staphylococci that confer resistance to Streptogramin A antibiotics. These proteins

are ATP-binding proteins that are thought to be involved in the active efflux of Streptogramin A compounds. They were originally characterized from *S. aureus*, but have now been detected in several different species of staphylococci (Haroche *et al.*, 2000; Lina *et al.*, 1999).

In antibiotic producers, there are also efflux pumps specific for various members of the MLS group of antibiotics. These proteins are generally members of the ABC transporter superfamily (Schoner *et al.*, 1992) and include the efflux pumps encoded by *car*(A) from *Streptomyces thermotolerans* (efflux of lincomycin) (Schoner *et al.*, 1992), *ole*(B) and *ole*(C) from *Streptomyces antibioticus* (efflux of oleandomycin) (Olano *et al.*, 1995; Rodriguez *et al.*, 1993), *srm*(B) from *Streptomyces ambofaciens* (efflux of spiramycin) (Richardson *et al.*, 1987), *lmr*(C) from *Streptomyces lincolnensis* (efflux of lincomycin) (Peschke *et al.*, 1995), and *tlr*(C) from *Streptomyces fradiae* (efflux of tylosin) (Rosteck Jnr. *et al.*, 1991).

Target modification

The most common mechanism of resistance to MLS antibiotics is modification of the antibiotic target site. Bacterial ribosomes are 70S particles comprising 50S (the target site of MLS antibiotics) and 30S units that join at the initiation step of protein synthesis and separate at the termination step. Each subunit is comprised of RNA (one 5S and one 23S rRNA molecule in 50S subunits and a 16S species in the 30S subunit) and ribosomal proteins (Yusupov *et al.*, 2001). On the basis of secondary structure, 23S rRNA has been subdivided into six domains (Noller *et al.*, 1981). In *E. coli* and other organisms, mutations in two of these domains, Domain II and Domain V, have been implicated in resistance to erythromycin (Douthwaite *et al.*, 1985). Modifications in Domain V can either be

due to the acquisition of a gene that encodes adenine-N^6 methyltransferase, an enzyme which post-transcriptionally methylates a residue in Domain V, or by direct mutation of residues in Domain V. Both mechanisms result in ribosomes which bind MLS antibiotics with reduced affinity (Weisblum, 1998).

a) Post-transcriptional methylation of Domain V

In *S. aureus* and other pathogens, modification involving methylation of A2058, which is located in Domain V, has been observed. The methylation of the 23S rRNA occurs post-transcriptionally at a site in the peptidyltransferase circle of 23S rRNA Domain V, which corresponds to A2058 based on the *E. coli* numbering scheme (Lai and Weisblum, 1971; Skinner *et al.*, 1983). The enzymes that catalyze the methylation belong to a family of enzymes that has been designated Erm, which stands for erythromycin resistance methylase. The methylation of the 23S rRNA presumably causes a conformational change in the ribosome and leads to co-resistance to MLS antibiotics because the binding sites of these drugs overlap.

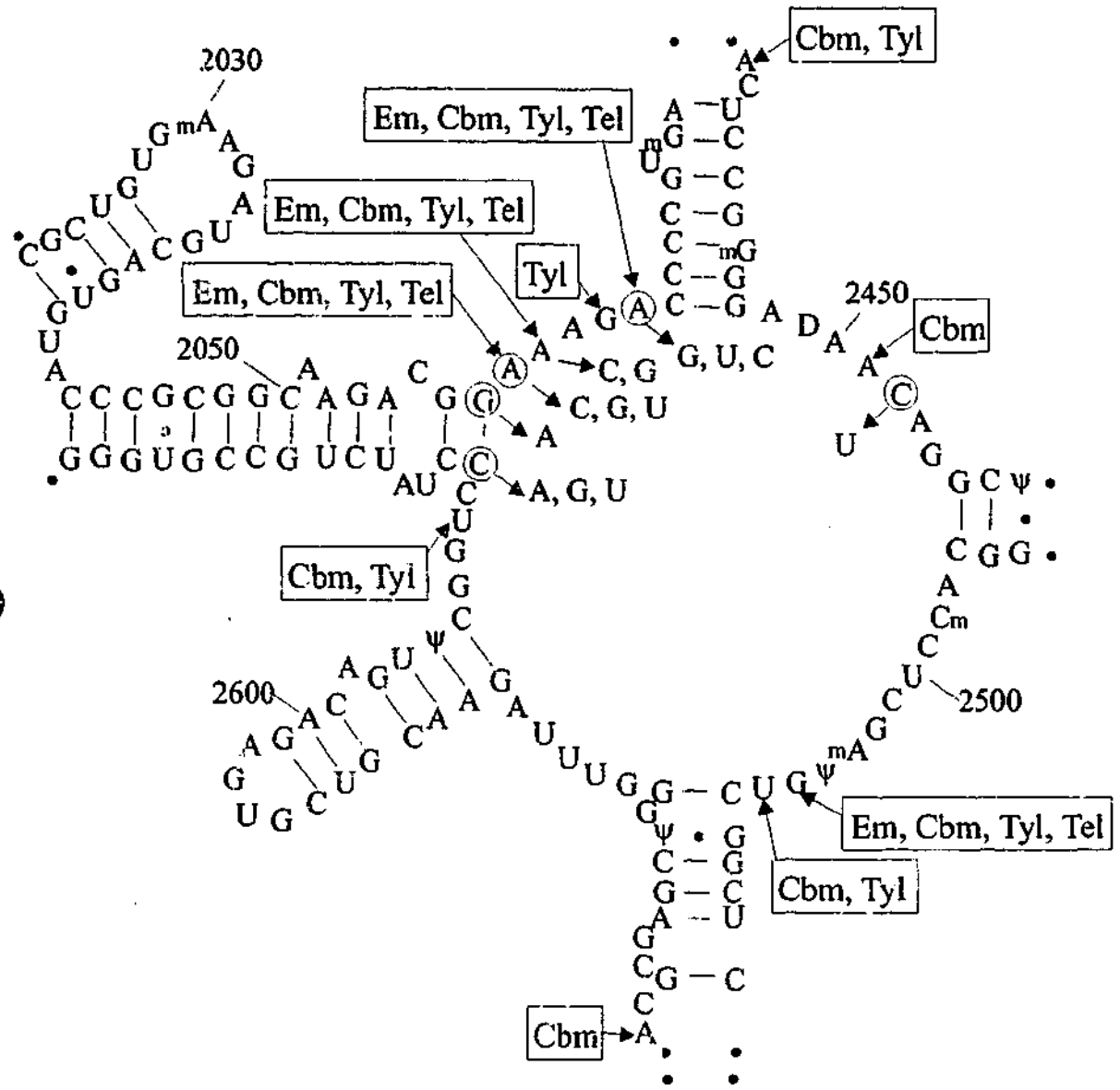
b) Mutations in Domain V

The second form of 23S rRNA alteration is intrinsic, results in the MLS cross resistance phenotype, and is based on a mutational alteration of 23S rRNA at A2058, which is the same site at which post-transcriptional methylation occurs. This form of resistance was first reported in the 23S rRNA of clinical isolates of *Mycobacterium intracellulare* (Meier *et al.*, 1994) and has been attributed to mutation to G, C, or U at A2058 (Figure 1.1). Similar mutations have also been observed in strains of the pathogenic organisms *Brachyspira hyodysenteriae* (Karlsson *et al.*, 1999), *Helicobacter pylori* (Wang and Taylor, 1998), *Mycobacterium abscessus* (Wallace Jnr. *et al.*, 1996), *Mycobacterium avium* (Nash and Inderlied, 1995), *Mycobacterium*

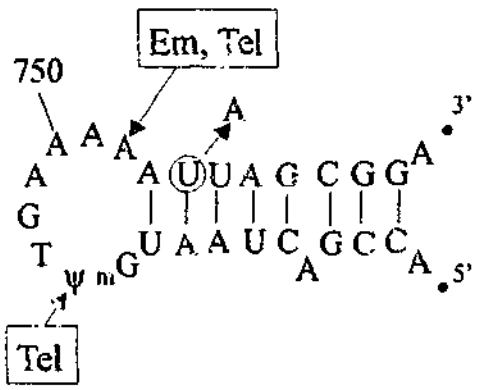
Figure 1.1 : Secondary structure models of the peptidyl transferase center of 23S rRNA.

Secondary-structure models of the peptidyl transferase center in Domain V (A) and hairpin 35 in Domain II (B) from *E. coli* 23S rRNA are shown. The red circled nucleotides indicate the positions of mutations that confer resistance to MLS antibiotics. Nucleotides at which the MLS antibiotics erythromycin (Em), carbomycin (Cbm), telithromycin (Tel) and tylosin (Tyl) interact (as defined by chemical footprinting experiments) are indicated by green arrows. Figure updated from the version presented by Vester and Douthwaite (2001) using information from Depardieu and Courvalin (2001) and Fumeri *et al.* (2001).

(A)


Cbm, Tyl
2030
Em, Cbm, Tyl, Tel
Em, Cbm, Tyl, Tel
Tyl
Em, Cbm, Tyl, Tel
2450
Cbm
2050
C, G
G, U, C
C, G, U
A, G, U
Cbm, Tyl
2600
2500
Em, Cbm, Tyl, Tel
Cbm, Tyl
Cbm

(B)


Em, Tel
750
3'
5'
Tel

chelonae (Wallace Jnr. *et al.*, 1996), *Mycobacterium kansasii* (Burman *et al.*, 1998), *Mycoplasma pneumoniae* (Lucier *et al.*, 1995), *Propionibacteria* sp. (Ross *et al.*, 1997) and *S. pneumoniae* (Tait-Kamradt *et al.*, 2000), as well as in other non-pathogenic organisms (Vester and Douthwaite, 2001).

While A2058 appears to be the most common site of mutation in the 23S rRNA, mutations in the neighbouring residues, G2057 and A2059, have also been reported for many of the pathogenic species listed above. Mutation of G2057 appears to result in ribosomes that are resistant to 14-membered macrolides, but sensitive to 16-membered macrolides, as well as lincosamide and Streptogramin B antibiotics. Mutation of A2059 results in ribosomes that are macrolide and lincosamide resistant but sensitive to Streptogramin B antibiotics (Figure 1.1). Mutation of A2062 in *S. pneumoniae* (Depardieu and Courvalin, 2001) and *Mycoplasma hominis* (Furneri *et al.*, 2001), C2452 in *Sulfolobus acidocaldarius* (Aagaard *et al.*, 1994), and C2611 in *Chlamydomonas moewusii* (Gauthier *et al.*, 1988), *Chlamydomonas reinhardtii* (Harris *et al.*, 1989), *E. coli* (Vannuffel *et al.*, 1992), *S. pneumoniae* (Tait-Kamradt *et al.*, 2000) and *Saccharomyces cerevisiae* (Sor and Fukuhara, 1984) have been reported and all lead to resistance to one or more of the MLS antibiotics (Figure 1.1).

c) Mutations in Domain II

Mutations in Domain II of the 23S rRNA that cause resistance to erythromycin do so in a manner fundamentally different from mutations in Domain V. Domain II mutations are generally located in a hairpin structure between nucleotides 1198 and 1247 (Dam *et al.*, 1996). This hairpin is close to a short open reading frame in the 23S rRNA that encodes a pentapeptide (E-peptide) whose

expression *in vivo* renders cells resistant to erythromycin (Tenson *et al.*, 1996). Mutations in nucleotides within the E-peptide gene, and in the ribosome binding site, appear to increase the accessibility of the ribosome binding site and E-peptide gene. Consequently, mutations in these regions appear to mediate erythromycin resistance by increasing expression of the rRNA-encoded E-peptide (Dam *et al.*, 1996).

Mutation of U754 in Domain II has also been reported in *E. coli* (Xiong *et al.*, 1999), resulting in ribosomes that are resistant to erythromycin. U754 is in hairpin 35 in Domain II (Figure 1.1). The results suggest that this region of Domain II is in close proximity to Domain V and may constitute part of the ribosomal peptidyltransferase centre.

Erythromycin resistance methylase (erm) *genes*

The most common mechanism of resistance to MLS antibiotics is the acquisition of an *erm* gene, the product of which catalyzes the post-transcriptional methylation of a specific adenine in the 23S rRNA, leading to cross resistance to all MLS antibiotics. Over the last 30 years, a large number of different *erm* genes have been isolated from many different bacterial species that range from *E. coli* to *Haemophilus influenzae* in Gram-negative species and from *S. pneumoniae* to *Corynebacterium* spp. in Gram-positive bacteria (Table 1.1). In addition, several Gram-positive and Gram-negative anaerobic organisms and even spirochaetes, such as *Borrelia burgdorferi* and *Treponema denticola*, have been shown to carry *erm* genes (Roberts *et al.*, 1999).

Table 1.1 : Ribosomal RNA methylase genes involved in MLS resistance.

Class	Protein	Gene Name	Gene(s) Included*	% Homology		Plasmid or Transposon	Reference	Species
				DNA	Amino Acid			
A	Erm(A)	*erm*(A)	*ermA*	83	81	Tn*554*	(Murphy, 1985)	*S. aureus*
							(Thakker-Varia *et al.*, 1987)	Coagulase negative staphylococci
							(Roe *et al.*, 1996)	*Actinobacillus actinomycetemcomitans*
							(Wasteson *et al.*, 1996)	*Actinobacillus pleuropneumoniae*
			ermTR				(Seppälä *et al.*, 1998)	*S. pyo*
							(Reig *et al.*, 2001)	*Peptos.* spp.
B	Erm(B)	*erm*(B)	*ermAM*	98-100	98-100	Tn*1545*	(Trieu-Cuot *et al.*, 1990)	*S. pneumoniae*
						pAMβ1	(Brehm *et al.*, 1987)	*E. faecalis*
						pAM77	(Horinouchi *et al.*, 1983)	*Streptococcus sanguis*
			ermB			Tn*917*	(Shaw and Clewell, 1985)	*E. faecalis*
						pIP501	(Brantl *et al.*, 1994)	*Streptococcus agalactiae*
			ermAMR				(Oh *et al.*, 1998)	*E. faecalis*
			ermBC			pIP1527	(Brisson-Noël and Courvalin, 1988)	*E. coli*
			ermP, ermBP *			pIP402	(Berryman and Rood, 1995)	*C. perfringens*
			ermIP			pIP501	(Pujol *et al.*, 1994)	*S. agalactiae*
			ermZ, ermBZ1, ermBZ2 *			Tn*5398*	(Farrow *et al.*, 2000)	*C. difficile*
			erm			pLEM3	(Fons *et al.*, 1997)	*Lactobacillus fermentum*
			erm2			pBT233	(Ceglowski *et al.*, 1993)	*S. pyogenes*
						pMD101	(Ceglowski and Alonso, 1994)	*S. pyogenes*

C	Erm(C)	*erm*(C)	*ermC*	99-100	98-100	pE194	(Horinouchi and Weisblum, 1982)	*S. aureus*
						pT48	(Catchpole *et al.*, 1988)	*S. aureus*
						pE5	(Projan *et al.*, 1987)	*S. aureus*
						pJR5	(Oliveira *et al.*, 1993)	*S. aurues*
						pA22	(Catchpole and Dyke, 1990)	*S. aureus*
						pSES6	(Lodder *et al.*, 1996)	*Staphylococcus equorum*
						pSES5	(Lodder *et al.*, 1997)	*Staphylococcus hominis*
						pSES4a	(Lodder *et al.*, 1997)	*S. haemolyticus*
						pSES21	(Lodder *et al.*, 1997)	*Staphylococcus hyicus*
						pOX7	(Needham *et al.*, 1995)	*S. aureus*
			ermIM			pIM13	(Projan *et al.*, 1987)	*Bacillus subtilis*
			ermM			pNE131	(Lampson and Parisi, 1986)	*S. epidermidis*
						pPV141	(Somkuti *et al.*, 1997)	*Staphylococcus chromogenes*
						pPV142	(Somkuti *et al.*, 1998)	*Staphylococcus simulans*
D	Erm(D)	*erm*(D)	*ermD*	97-99	97-99	pBD90	(Docherty *et al.*, 1981)	*Bacillus licheniformis*
			ermJ			pBA423	(Kim *et al.*, 1993)	*Bacillus anthracis*
			ermK				(Kwak *et al.*, 1991)	*B. licheniformis*
E	Erm(E)	*erm*(E)	*ermE*	99	96	pUC31, pIJ43	(Dhillon and Leadlay, 1990)	*Saccharopolyspora erythraea*
			ermE2				(Bibb *et al.*, 1985)	*Streptomyces erythraeus*
F	Erm(F)	*erm*(F)	*ermF*	98-100	97-100	pBF4	(Rasmussen *et al.*, 1986)	*Bacteroides fragilis*
						Tn*4351*	(Rasmussen *et al.*, 1987)	*B. fragilis*
			ermFS			Tn*4551*	(Smith, 1987)	*B. fragilis*
			ermFU			Chromosomal	(Halula *et al.*, 1991)	*B. fragilis*
G	Erm(G)	*erm*(G)	*ermG*	99	99	pBD370	(Monod *et al.*, 1987)	*Bacillus sphaericus*
						Tn*7853*	(Cooper *et al.*, 1997)	*Bacteroides thetaiotaomicron*
H	Erm(H)	*erm*(H)	*carB*			pOJ159	(Epp *et al.*, 1987)	*S. thermotolerans*

I	Erm(I)	*erm*(I)	*mdmA*				(Hara and Hutchinson, 1990)	*Streptomyces* spp.
N	Erm(N)	*erm*(N)	*tlrD*				(Zalacain and Cundliffe, 1991)	*S. fradiae*
O	Erm(O)	*erm*(O)	*lrm* *srmA*	84	84	pLST391	(Jenkins and Cundliffe, 1991) (Pernodet *et al.*, 1999)	*Streptomyces lividans* *S. ambofaciens*
Q	Erm(Q)	*erm*(Q)	*ermQ*			Chromosomal	(Berryman *et al.*, 1994)	*C. perfringens*
R	Erm(R)	*erm*(R)	*ermA*				(Roberts *et al.*, 1985)	*Arthrobacter* sp.
S	Erm(S)	*erm*(S)	*ermSF* *tlrA*	100	100	pET23	(Kamimiya and Weisblum, 1988) (Kovalic *et al.*, 1994)	*S. fradiae* *S. fradiae*
T	Erm(T)	*erm*(T)	*ermGT*			pGT633	(Tannock *et al.*, 1994)	*Lactobacillus reuteri*
U	Erm(U)	*erm*(U)	*lmrB*			pPZ303	(Peschke *et al.*, 1995)	*S. lincolnensis*
V	Erm(V)	*erm*(V)	*ermSV*				(Fujisawa and Weisblum, 1981)	*Streptomyces viridochromogenes*
W	Erm(W)	*erm*(W)	*myrB*				(Inouye *et al.*, 1994)	*Micromonospora griseorubida.*
X	Erm(X)	*erm*(X)	*ermCD* *ermA* *ermCX*	99-100	99-100	pNG2 pNG2 Tn*5432*	(Serwold-Davis and Groman, 1988) (Hodgson *et al.*, 1990) (Tauch *et al.*, 1995)	*Corynebacterium diphtheriae* *C. diphtheriae* *Corynebacterium xerosis*
Y	Erm(Y)	*erm*(Y)	*ermGM*			pMS97	(Matsuoka *et al.*, 1998)	*S. aureus*
Z	Erm(2)	*erm*(2)	*srmD*				(Pernodet *et al.*, 1993)	*S. ambofaciens*

* When two or three gene names are listed under "Gene(s) Included" it means that the same gene was designated by two or three different names in the literature; for *erm*(B), *erm*(C), and *erm*(M), multiple related genes have been sequenced and a selection of these are listed. Table modified from Roberts *et al.* (1999).

Classification, nomenclature and distribution of *erm* genes

Due to the large number of *erm* genes that had been described and characterized in the literature, the classification and nomenclature of *erm* genes has been very confusing. Originally the *erm* genes were divided into eight hybridization classes, ErmA-ErmG and ErmQ, on the basis of DNA-DNA hybridization. These classes were assigned according to the prototype gene in each class (Arthur *et al.*, 1990; Hächler and Kayser, 1993; Leclercq and Courvalin, 1991; Mabilat and Courvalin, 1988). It was common practice for investigators to give their *erm* gene a new name regardless of the nucleotide and predicted amino acid sequence similarity to previously characterized *erm* genes and Erm proteins and without regard to whether the gene resided in a different isolate, species or genus. The result was that unrelated genes were often given the same name or very similar genes were given different names (Roberts *et al.*, 1999), which has caused significant confusion about the relationship between different *erm* genes.

In a recent review (Roberts *et al.*, 1999) a more rational classification and nomenclature system for the *erm* genes was proposed. This system divided the *erm* genes into 21 classes (A-I, N,O, Q-Z) on the basis of DNA and amino acid sequence homology (Table 1.1). Genes that have greater than 80% amino acid sequence similarity have been placed in the same class and are now given the same gene and protein name. For example, all genes in class A are named *erm*(A) and the proteins encoded by these genes are all named Erm(A). This system is based on the current classification and nomenclature system for identifying and naming tetracycline resistance genes (Levy *et al.*, 1999; Levy *et al.*, 1989) and serves to clarify the relationships between the *erm* genes.

In general each class of *erm* genes is loosely associated with a particular bacterial genus with the exception of classes B, C and F (Table 1.2). The broad distribution of the *erm* genes in these classes indicates that they are readily transferable between different genéra. Many of the *erm* genes in classes B, C and F have been shown to be associated with either conjugative or non-conjugative transposons and also with broad host range plasmids, which would provide a means for transfer of the *erm* genes from one bacterial species to another.

Regulation of erythromycin resistance

The expression of several *erm* genes is induced by exposure to subinhibitory concentrations of MLS antibiotics (Dubnau and Monod, 1986). Erythromycin resistance is generally negatively regulated by attenuators. In the absence of erythromycin, the requisite mRNA is either synthesized in an inactive conformation in which translation cannot be efficiently initiated, or the synthesis of the requisite mRNA is not completed owing to rho-factor-independent termination before synthesis of the message is complete. In both cases, the state of association of inverted complementary repeat sequences located between the transcriptional and translational start signal of the resistance gene determines whether its transcription is completed (transcriptional attenuation), and, if completed, whether the message is translated (translational attenuation) (Weisblum, 1998).

a) Translational attenuation

The best studied example of regulation of erythromycin resistance is that of the translational attenuation mechanism exhibited by *erm*(C). Analysis of the nucleotide sequence upstream of the *erm*(C) structural gene reveals an open reading frame (ORF) that encodes a peptide referred to as the Erm leader peptide. The Erm

Table 1.2 : Genus distribution of ribosomal RNA methyltransferases.

Gene	Genus or Genera
erm(A)	*Actinobacillus*, *Staphylococcus*, *Streptococcus*
erm(B)	*Actinobacillus*, *Clostridium*, *Escherichia*, *Enterococcus*, *Klebsiella*, *Neisseria*, *Pediococcus*, *Staphylococcus*, *Streptococcus*, *Wolinella*
erm(C)	*Actinobacillus*, *Bacillus*, *Eubacterium*, *Lactobacillus*, *Neisseria*, *Staphylococcus*, *Streptococcus*, *Wolinella*
erm(D)	*Bacillus*
erm(E)	*Streptomyces*
erm(F)	*Actinobacillus*, *Actinomyces*, *Bacteroides*, *Clostridium*, *Eubacterium*, *Fusobacterium*, *Gardnerella*, *Haemophilus*, *Neisseria*, *Porphyromonas*, *Prevotella*, *Peptostreptococcus*, *Selenomonas*, *Streptococcus*, *Treponema*, *Veillonella*, *Wolinella*
erm(G)	*Bacillus*, *Bacteroides*
erm(H)	*Streptomyces*
erm(I)	*Streptomyces*
erm(N)	*Streptomyces*
erm(O)	*Streptomyces*
erm(Q)	*Actinobacillus*, *Clostridium*, *Streptococcus*, *Wolinella*
erm(R)	*Arthrobacter*
erm(S)	*Streptomyces*
erm(T)	*Lactobacillus*
erm(U)	*Streptomyces*
erm(V)	*Streptomyces*
erm(W)	*Micromonospora*
erm(X)	*Corynebacterium*
erm(Y)	*Staphylococcus*
erm(2)	*Streptomyces*

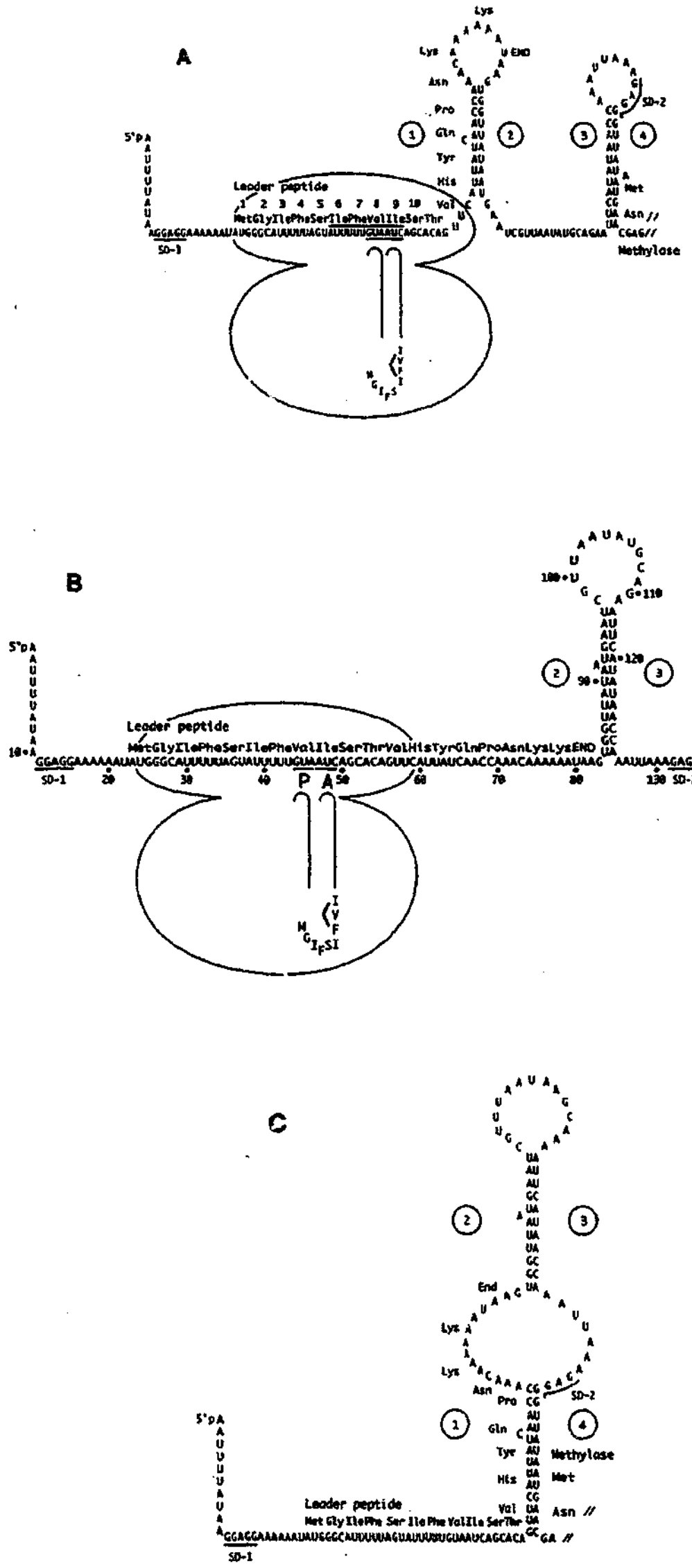
N.B. Table taken from Roberts *et al.* (1999).

leader peptide ORF is constitutively co-transcribed with the *erm*(C) gene and contains a number of inverted repeats that enable the mRNA molecule to assume different stable conformations (Figure 1.2). In the uninduced, or nascent state, the *erm*(C) mRNA leader is expected to assume the conformation shown in Figure 1.2A. In this state, translation of the Erm(C) protein is initiated at low efficiency because the first two *erm*(C) codons, AUG and AAC, as well as the *erm*(C) ribosome binding site, are sequestered by the secondary structure (Weisblum, 1998). Induction provides conditions that favour a translationally active conformation of the *erm*(C) mRNA leader, as shown in Figure 1.2B. Induction of *erm*(C) begins with the binding of erythromycin to ribosomes that synthesize the 19 amino acid leader peptide. This binding causes the ribosome to stall while transcribing the leader peptide, thus preventing the antibiotic-bound ribosomes from completely translating the leader peptide. The ribosome stops at a point in the leader peptide sequence and causes a conformational change in the mRNA, making available the *erm*(C) ribosome binding site and the first codons of *erm*(C) (Mayford and Weisblum, 1990). The result is increased efficiency of *erm*(C) translation. After induction, when the concentration of erythromycin has been reduced and can no longer support induction or when all available 23S rRNA molecules have been methylated (maximal methylation), the leader region can refold to assume an inactive conformation shown in Figure 1.2C.

Based on the similarity of the upstream region of other *erm* genes to the leader peptide sequence upstream of *erm*(C), other *erm* genes have also been proposed to be regulated by translational attenuation, including *erm*(A) (Murphy, 1985), *erm*(B) from *Streptococcus sanguis* (Horinouchi *et al.*, 1983), *erm*(G) (Monod *et al.*, 1987), *erm*(S) (Kamimiya and Weisblum, 1988), and *erm*(V) (Kamimiya and

Figure 1.2 : Proposed conformational transitions of the *erm*(C) leader sequence during induction.

Early during induction by erythromycin, the *erm*(C) mRNA leader sequence is proposed to take up the conformation shown in (A). The four major segments of the *erm*(C) attenuator are paired as segment 1:2 and segments 3:4, reflecting the temporal order of their synthesis. The ribosome is shown stalled during the addition of isoleucine 9 to the growing peptide chain. In the fully induced state the *erm*(C) leader is proposed to have the conformation shown in (B). As a consequence of stable complex formation between the erythromycin-ribosome complex and the *erm*(C) message, the association between segments 1 and 2 is prevented. Instead association between segments 2 and 3 occurs, which uncovers the ribosome binding site (SD-2) and first two codons of the *erm*(C) message encoded by segment 4. The removal of erythromycin or maximal methylation of 23S rRNA is proposed to result in the *erm*(C) mRNA leader sequence taking up the inactive conformation shown in (C). Figure taken from Weisblum (1995).


A
Leader peptide
MetGlyIlePheSerIlePheValIleSerThr
SD-1
Methylase
B
Leader peptide
MetGlyIlePheSerIlePheValIleSerThrValHisTyrGlnProAsnLysLysEND
MetAsnGln //
SD-1
SD-2
P
A
C
Leader peptide
Met Gly Ile Phe Ser Ile Phe Val Ile Ser Thr
SD-1
Methylase
Met
Asn //

Weisblum, 1997). Examination of MLS resistance determinants that express erythromycin resistance constitutively has revealed that the leader peptide sequence is either absent from the region upstream of the *erm* structural gene (Brehm *et al.*, 1987; Martin *et al.*, 1987) or has mutations that abolish its potential to take up the conformation that sequesters the *erm* ribosome binding site (Brisson-Noël and Courvalin, 1988; Kamimiya and Weisblum, 1988; Mayford and Weisblum, 1990; Murphy, 1985; Rasmussen *et al.*, 1986).

b) Transcriptional attenuation

Transcriptional attenuation has been shown to be the regulatory mechanism that mediates expression of *erm*(D) from *Bacillus licheniformis*. This mechanism contrasts with translational attenuation in that the ribosomal pause is linked functionally to termination of transcription of the *erm*(D) message. In the absence of erythromycin, the transcribed *erm*(D) mRNA consists of two truncated classes, corresponding to rho-independent transcription termination sites located in the leader sequence of *erm*(D). In response to erythromycin, ribosomes stall during the translation of the leader peptide, resulting in a conformational change in the *erm*(D) mRNA and transcription of full length *erm*(D) message (Kwak *et al.*, 1991).

The leader peptide sequence found upstream of a second *erm*(D) gene cloned from *B. licheniformis* on plasmid pDB90 (Docherty *et al.*, 1981) differs from the *erm*(D) leader peptide sequence described by Kwak *et al.* (1991) at only three nucleotide positions and would not be predicted to significantly alter the regulatory mechanism. However expression of the pDB90-derived *erm*(D) gene appears to be regulated by translational attenuation rather than transcriptional attenuation (Hue and Bechhofer, 1992). Transcriptional termination is observed by these researchers,

however, they find no difference in the amount of full length *erm*(D) message produced in induced and non-induced cells, indicating that translational attenuation is the more likely regulatory mechanism. The role of transcriptional termination in the pDB90-derived *erm*(D) leader region is not well understood.

Production of Erm(S) in *S. fradiae* is also regulated by a transcriptional attenuation mechanism. In the uninduced state, transcription terminates 27 nucleotides into the *erm*(S) coding sequence. Induction of *erm*(S) is proposed to involve a ribosome-mediated conformational change within the mRNA leader that allows transcription to continue beyond the attenuation site, resulting in transcription of the full length *erm*(S) message (Kelemen *et al.*, 1994).

Dissemination of antibiotic resistance genes

Bacterial antibiotic resistance is often caused by the acquisition of new genes rather than by mutation. The way in which antibiotic resistance genes are acquired and spread among different bacterial species is of interest as it furthers our understanding of the emergence of resistant bacterial isolates (Recchia and Hall, 1997).

MLS resistance determinants have a widespread distribution amongst the different bacteria isolated from humans, animals and the environment (Tables 1.1 and 1.2). In particular, the Erm B, Erm C and Erm F determinants appear to be found in a diverse range of bacterial species and genera, indicating extensive horizontal transfer, which has occurred as a result of the association of the MLS resistance determinants with either conjugative or mobilizable genetic elements.

Conjugative transposons

Conjugative transposons are discrete DNA elements that are normally integrated into either the bacterial chromosome or plasmids. They are characterized by their ability to encode their own movement from one bacterial cell to another by a process requiring cell to cell contact. They are found in both Gram-positive and Gram-negative bacteria and range in size from 18 kb to greater than 150 kb. The majority of these transposons carry antibiotic resistance determinants and therefore have been shown to contribute significantly to the spread of antibiotic resistance genes between bacterial genera (Clewell and Flannagan, 1993; Clewell *et al.*, 1995; Rice, 1998; Salyers and Shoemaker, 1995; Scott and Churchward, 1995; Waters, 1999).

a) The Tn916/Tn1545 family of conjugative transposons

The best studied conjugative transposons are Tn*916* (18 kb) from *Enterococcus faecalis* (Franke and Clewell, 1981) and the closely related element, Tn*1545* (25.2 kb), from *S. pneumoniae* (Courvalin and Carlier, 1986; Courvalin and Carlier, 1987). Both elements carry the *tet*(M) tetracycline resistance gene, however, Tn*1545* also carries kanamycin (*aphA-3*) and MLS (*erm*(B)) resistance genes (Courvalin and Carlier, 1987). Related conjugative elements have been detected in several bacterial genera and form a group of transposons that are collectively known as the Tn*916*/Tn*1545* family (Clewell *et al.*, 1995; Rice, 1998). This family of transposons is widely distributed and its members have been found naturally, or have been introduced into, over 50 different species and 24 bacterial genera (Clewell *et al.*, 1995).

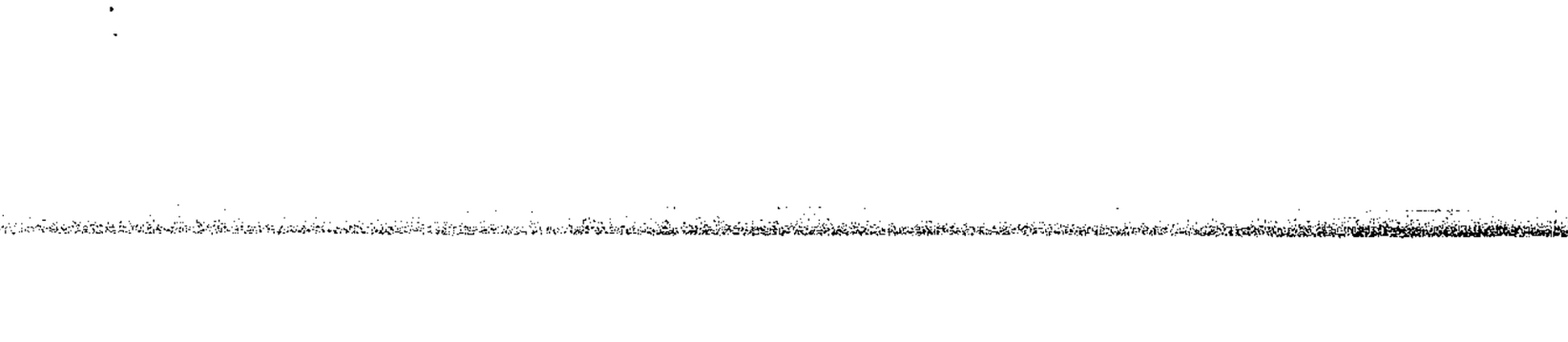
The nucleotide sequence of Tn*916* has been determined. It is 18,032 bp in length and contains 24 ORFs (Figure 1.3). Genes encoding proteins involved in conjugative transfer make up most of the transposon (Figure 1.3) (Clewell *et al.*, 1995; Rice, 1998; Senghas *et al.*, 1988; Yamamoto *et al.*, 1987) and are located to the left of the *tet*(M) gene. To the right of *tet*(M) are genes encoding proteins that are involved in regulation and transposition , including *int* and *xis* (Celli and Trieu-Cuot, 1998; Jaworski *et al.*, 1996).

The conjugative transposition of Tn*916* has been shown to be stimulated in the presence of tetracycline (Manganelli *et al.*, 1995; Showsh and Andrews Jnr., 1992) and involves three distinct stages (Figure 1.4). The first stage involves the excision of the transposon from the donor DNA. Staggered cleavage at both ends of the element by the Int protein (Taylor and Churchward, 1997) results in 5' 6 bp overhangs (Manganelli *et al.*, 1996; Rudy and Scott, 1994), which are referred to as coupling sequences. Following excision, the coupling sequences covalently link to circularize the transposon, resulting in a heteroduplex at the junction point (Caparon and Scott, 1989). Transfer of one strand of the circularized transposon to a new host, after it is nicked at the *oriT* site, then occurs (Jaworski and Clewell, 1995). In the new host the complementary strand is synthesized (Scott *et al.*, 1994) and the transposon then inserts into a target site, which is usually A-T rich (Lu and Churchward, 1995).

In *E. coli* and *Bacillus subtilis*, Int has been shown to be required for the integration and excision of Tn*916* (Poyart-Salmeron *et al.*, 1989; Rudy *et al.*, 1997; Storrs *et al.*, 1991; Su and Clewell, 1993). In *E. coli*, Int alone can carry out all the functions of transposition but at a much lower frequency compared to when Xis is

Figure 1.3 : Genetic organization of the conjugative transposon Tn*916*.

The 24 ORFs and their direction of transcription are indicated as block arrows. ORF3 and ORF4 are located within ORF2 and are in the same frame. Areas that have been identified as having specific functions are indicated by the boxes below the map. The tetracycline resistance gene, *tet*(M), is shown in red and the excisionase and integrase genes, *xis* and *int*, are shown in blue. The origin of transfer is indicated by the black triangle. Figure modified from Jaworski and Clewell (1995).



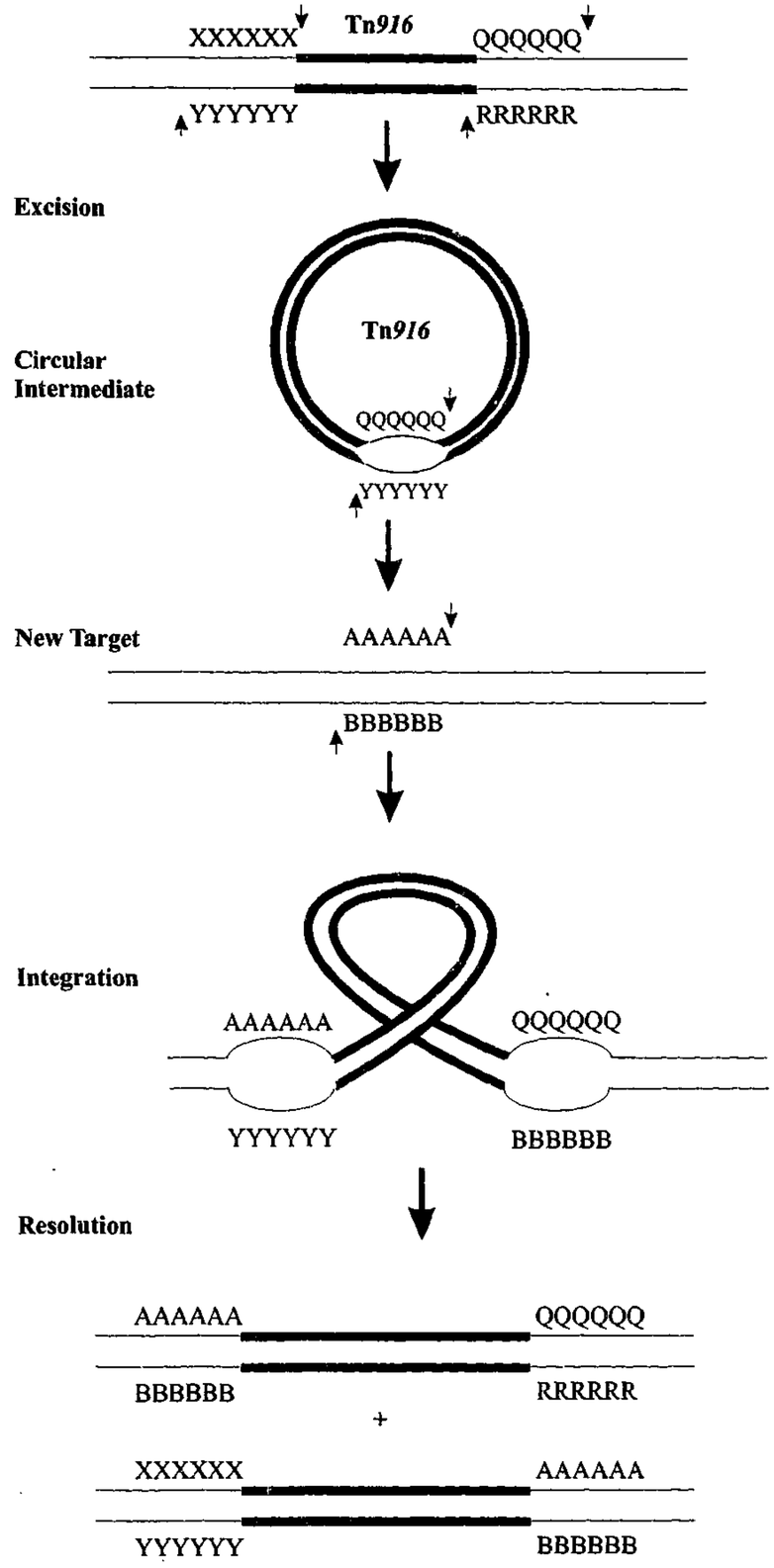


Figure 1.4 : Model for excision and integration of Tn*916*.

The thick green lines represent Tn*916* and the thin black lines represent the DNA adjacent to the transposon. Coupling sequences are indicated by the hypothetical complementary nucleotide pairs X-Y, Q-R and A-B. Excision involves staggered cleavages (small arrows) at the 5' side of the coupling sequence on both strands. The ends are joined to create a circular intermediate, which contains a heteroduplex consisting of the base pairs present in the coupling sequences. Insertion of Tn*916* into a new target site involves staggered cleavages in the circular intermediate and target site followed by ligation. This process results in a new insertion of Tn*916* with a heteroduplex at each end. Replication resolves the heteroduplexes and generates a pair of molecules in which each member is flanked by the target sequence at one end and a coupling sequence at the other end. Figure based on Scott (1992).

present (Poyart-Salmeron *et al.*, 1989; Su and Clewell, 1993). In Gram-positive hosts, however, it has been shown that Int cannot act alone but requires Xis for the excision of Tn*916* (Hinerfeld and Churchward, 2001; Jaworski *et al.*, 1996; Marra *et al.*, 1999). Xis is not only required for excision of Tn*916* but also, when in high concentrations, negatively regulates the excision process by binding to the right end of the transposon (Hinerfeld and Churchward, 2001).

In addition to being capable of catalyzing its own conjugative transposition Tn*916* has also been shown to be capable of enhancing the transfer of other homologous conjugative transposons that are co-resident in the cell (Flannagan and Clewell, 1991), and of mobilizing non-conjugative plasmids (Jaworski and Clewell, 1995; Naglich and Andrews Jnr., 1988; Showsh and Andrews Jnr., 1999). Mobilization of plasmids does not appear to be dependent on the plasmid possessing a functional *mob* gene, but does require the presence of a sequence similar to the *oriT* sequence present on Tn*916*. It is postulated that the same protein or proteins involved in nicking of the Tn*916* circular intermediates at *oriT* during transfer nicks similar sequences present on co-resident plasmids. Once nicked at this site, the plasmid then assumes a relaxed form that is capable of being transferred during conjugation (Showsh and Andrews Jnr., 1999). ORF23 of Tn*916* has similarity to the MbeA mobilization protein of ColE1 (Flannagan *et al.*, 1994) and could be acting as the necessary Mob protein in this process (Showsh and Andrews Jnr., 1999).

b) Bacteroides *conjugative transposons*

Conjugative transposons are not confined to Gram-positive bacteria. A distinctive group of conjugative transposons, which are completely unrelated to the Tn*916*/Tn*1545* family transposons, have been found in Gram-negative anaerobes

belonging to *Bacteroides* sp. The *Bacteroides* conjugative transposons range in size from 65 kb to over 150 kb and have considerable regions of sequence similarity, but they are not identical (Salyers *et al.*, 1995a; Smith *et al.*, 1998) (Figure 1.5). Most of them carry a tetracycline resistance gene, *tet*(Q), which is distantly related to *tet*(M). Some elements also carry Erm F or Erm G MLS resistance determinants.

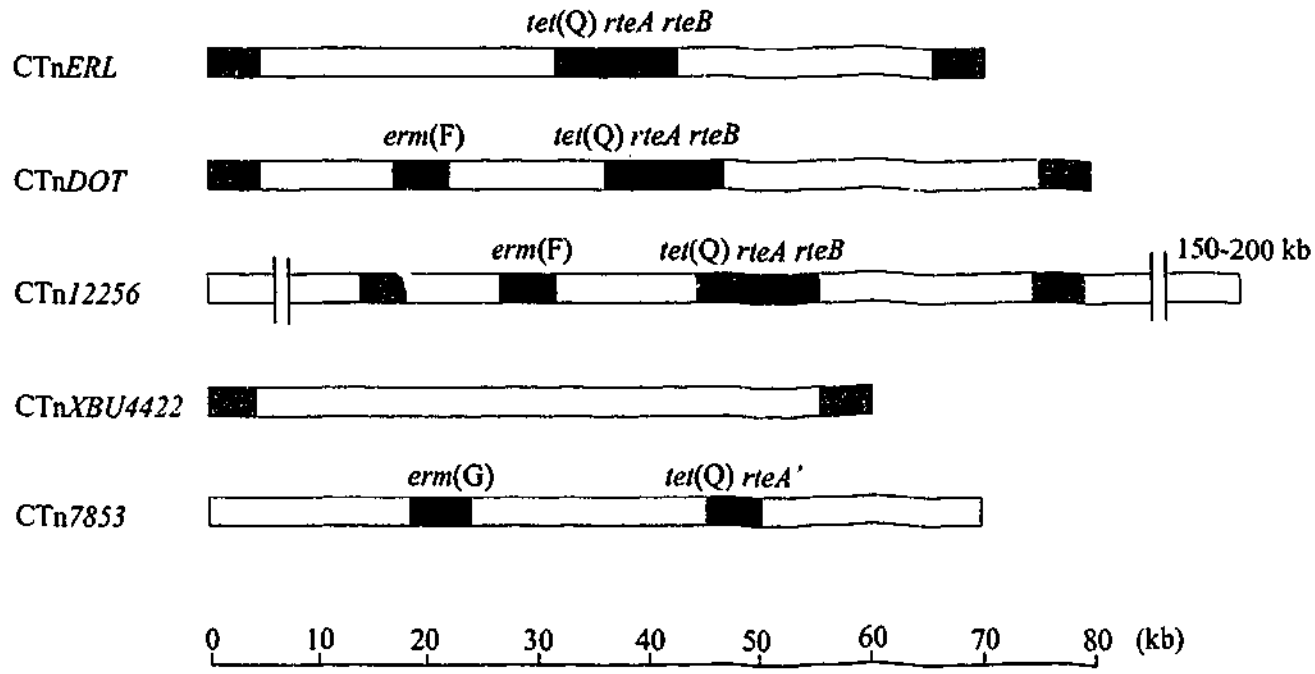
There are at least two distinct families of conjugative transposons in the bacteroides. One family is exemplified by the elements CTn*ERL* and CTn*DOT* (Figure 1.5) (Salyers *et al.*, 1995a). These elements are very similar to each other and are virtually identical in the region that contains *tet*(Q), *rteA* and *rteB*. CTn*DOT* contains a 13 kb insertion that carries *erm*(F) and appears to have resulted from the integration of mobilizable and non-mobilizable *Bacteroides* elements (Whittle *et al.*, 2001). This family has at least one cryptic member, CTn*XBU4422* (Salyers *et al.*, 1995a). The second family is typified by CTn*7853* (Figure 1.5) (Nikolich *et al.*, 1994), which appears to be completely unrelated to the CTn*ERL*/CTn*DOT* family except in the region immediately around *tet*(Q).

The large conjugative transposon, CTn*12256* (also called Tn*5030*) (Figure 1.5) (Bedzyk *et al.*, 1992; Macrina, 1993), is a compound element that consists of a CTn*DOT* type element inserted into another larger element. The latter element appears different from CTn*DOT* and CTn*7853* and could represent a third family of *Bacteroides* elements.

The mechanisms of integration and excision are less well established for *Bacteroides* conjugative transposons than they are for Tn*916*. Recent work has suggested that the *Bacteroides* elements integrate and excise by a similar mechanism

Figure 1.5 : The *Bacteroides* conjugative transposons.

Schematic representations of the known *Bacteroides* conjugative transposons are shown. Areas of similarity between the different transposons are indicated by similar coloring. CTn*DOT* has a 13 kb insertion that contains *erm*(F). CTn*12256* is a hybrid element with a CTn*DOT*-type element embedded in another similarly sized element. CTn*XBU4422* is a cryptic element that cross-hybridizes with CTn*ERL* but is more distantly related to it than is CTn*DOT*. CTn*7853* is unrelated to the other elements except that it contains *tet*(Q). Based on Figure 3 of Salyers *et al.* (1999).


tet(Q) rteA rteB
CTnERL
erm(F)
tet(Q) rteA rteB
CTnDOT
erm(F)
tet(Q) rteA rteB
150-200 kb
CTn12256
CTnXBU4422
erm(G)
tet(Q) rteA'
CTn7853
0
10
20
30
40
50
60
70
80
(kb)

to Tn*916* except that the coupling sequences are only 4 or 5 bp long (Cheng *et al.*, 2000). However, the biochemical processes of CTn*DOT* and Tn*916* excision are different. CTn*DOT* requires the products of the *int* and *exc* (predicted to encode a topoisomerase) genes and also the *rteA*, *rteB* and *rteC* genes. The transfer of CTn*DOT* is regulated by tetracycline, with pre-exposure to this antibiotic resulting in a 1,000-10,000 fold increase in the transfer frequency of the element. Tetracycline regulation of CTn*DOT* transfer is mediated by the regulatory genes, *rteA*, *rteB* and *rteC*. The *rteA* and *rteB* genes are part of an operon that also contains the *tet*(Q) gene and encode a two-component system in which RteA is the sensor and RteB is the response regulator (Salyers *et al.*, 1995c). Exposure of cells to tetracycline acts by increasing transcription of the entire operon by an unknown mechanism that does not require the participation of either RteA or RteB (Salyers *et al.*, 1995a), subsequently RteB activates the expression of the *rteC* gene (Stevens *et al.*, 1993).

Bacteroides conjugative transposons also appear to be more site specific than Tn*916*. There are usually about five to eight preferred integration sites in the *Bacteroides* genome (Salyers and Shoemaker, 1995). Site selectivity appears to be mediated by a 10 bp segment that is about 5 bp from one end of the conjugative transposon and which has a high level of sequence similarity to a 10 bp segment adjacent to the site where the conjugative transposon integrates (Cheng *et al.*, 2000). *Bacteroides* conjugative transposons have also been shown to mediate the transfer of co-resident plasmids, mobilizable transposons and the non-replicating *Bacteroides* units (NBUs) (Salyers and Shoemaker, 1995; Salyers *et al.*, 1995b; Smith and Parker, 1993).

Mobilizable transposons and elements

Conjugative transfer of plasmids and transposons occurs as a multistep process requiring specific DNA sequences and multiple gene products. These include a *cis*-acting origin of transfer, *oriT*, and *trans*-acting mobilization or Mob proteins, which are involved in the initiation of DNA transfer and replication in the recipient. In addition, other *trans*-acting proteins that form the conjugation pore or mating apparatus are also required. Conjugative plasmids and transposons encode all of these required proteins and are said to be self-transmissible, since their proteins can perform all initiation and termination functions and also can assemble the conjugation apparatus. Unlike conjugative elements, mobilizable transposons and elements only harbor an *oriT* site and genes required for the initiation and termination of transfer (Vedantam *et al.*, 1999). They generally carry antibiotic resistance genes and are capable of cell-to-cell movement only if there is a co-resident conjugative plasmid or transposon present in the cell (Shoemaker *et al.*, 2000; Tribble *et al.*, 1999; Vedantam *et al.*, 1999) to provide the mating pore proteins, and, where a conjugative transposon integrates into a plasmid (*cis* mobilization), the *oriT* site and proteins that initiate plasmid transfer (Salyers *et al.*, 1995a). There are several examples of mobilizable transposons, which range in size from Tn*5220* at 4.7 kb (Vedantam *et al.*, 1999), to Tn*4555* at 12.1 kb (Tribble *et al.*, 1999).

a) Bacteroides *mobilizable elements*

Many mobilizable elements are found in *Bacteroides* sp. (Smith *et al.*, 1998) and these include the NBUs, NBU1, NBU2, and NBU3, Tn*4399*, Tn*4555*, and Tn*5220* (Salyers *et al.*, 1995b; Smith *et al.*, 1998; Vedantam *et al.*, 1999). Tn*5220*, the NBUs and Tn*4555*, which are 5 to 12 kb in size, are integrated elements that can

be excised and mobilized in *trans* by the *Bacteroides* tetracycline-inducible conjugative transposons (Salyers *et al.*, 1995b). The mobilization region is located near the middle of these elements and consists of an *oriT* site and a single Mob protein. This region is all that is required for the circular form of these elements to be mobilized by the conjugative transposon.

In these elements the Mob proteins are multifunctional and appear to perform most, if not all, of the reactions required for the initiation of DNA transfer (recognition, binding, and specific cutting at the nick site) (Vedantam *et al.*, 1999). The multiple functions of these Mob proteins is somewhat unusual because most mobilizable plasmids and other mobilizable elements encode at least two mobilization genes, one that binds and nicks at the *oriT* site and one that encodes a helicase, which aids in strand separation (Salyers *et al.*, 1999). For example, Tn*4399* encodes two such proteins involved in mobilization, MocA and MocB (Vedantam *et al.*, 1999).

While the mobilization regions of these elements are similar, the integration mechanisms appear to be different. Tn*4555* integrates and excises similarly to Tn*916* and integrates into many different sites (Tribble *et al.*, 1997). By contrast, NBU1 integrates in *Bacteroides* in a site-specific manner into the 3'end of a leucine t-RNA gene, which is typical of the insertion of lamboid phages (Salyers *et al.*, 1999). The integration mechanism used by NBU2 has not been determined.

b) Clostridial mobilizable transposons

A second family of mobilizable transposons exists in the clostridia, the Tn*4451*/Tn*4453* family. This family includes the elements Tn*4451* and Tn*4452* from

C. perfringens and Tn*4453a* and Tn*4453b* from *C. difficile*, all of which carry the chloramphenicol resistance gene, *catP*. Like the *Bacteroides* mobilizable elements, these mobilizable transposons also appear to encode only one mobilization protein, TnpZ.

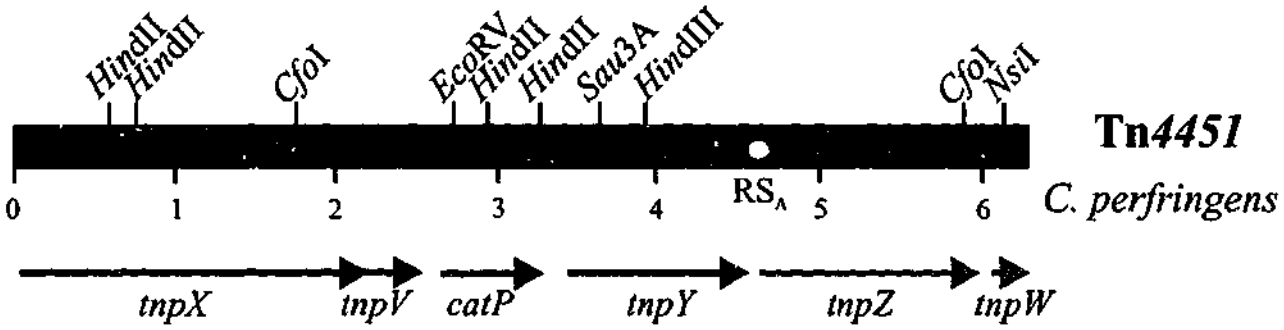
The nucleotide sequence of both Tn*4451* and Tn*4453a* has been determined (Bannam *et al.*, 1995; Lyras and Rood, 2000). These elements are closely related (89% identity at the nucleotide level) and have a similar genetic organization (Figure 1.6). However, based on functional analyses using a transposition assay system in *E. coli*, it has been shown that Tn*4453a* transposes at a higher frequency than does Tn*4451*. The molecular basis for this difference is unknown (Lyras and Rood, 2000).

In addition to the *catP* gene, both transposons carry a gene encoding a site-specific recombinase, *tnpX*, a mobilization gene, *tnpZ*, and three genes of unknown function, *tnpV*, *tnpY* and *tnpW* (Figure 1.6) (Bannam *et al.*, 1995; Lyras and Rood, 2000). Upstream of the *tnpZ* gene is an RS_A site (Bannam *et al.*, 1995; Crellin and Rood, 1998), which is essential for plasmid mobilization because it is the site of nicking by the TnpZ mobilization protein. The TnpZ proteins and the RS_A sites have been shown to be functional in *E. coli*, based on the mobilization of recombinant plasmids carrying these elements in the presence of a chromosomally integrated copy of the broad host range plasmid RP4 (Crellin and Rood, 1998; Lyras *et al.*, 1998).

TnpX is a member of the large resolvase family of site-specific recombinases and is responsible for the precise excision of the Tn*4451*/Tn*4453* family of transposons (Bannam *et al.*, 1995; Crellin and Rood, 1997; Lyras and Rood, 2000).

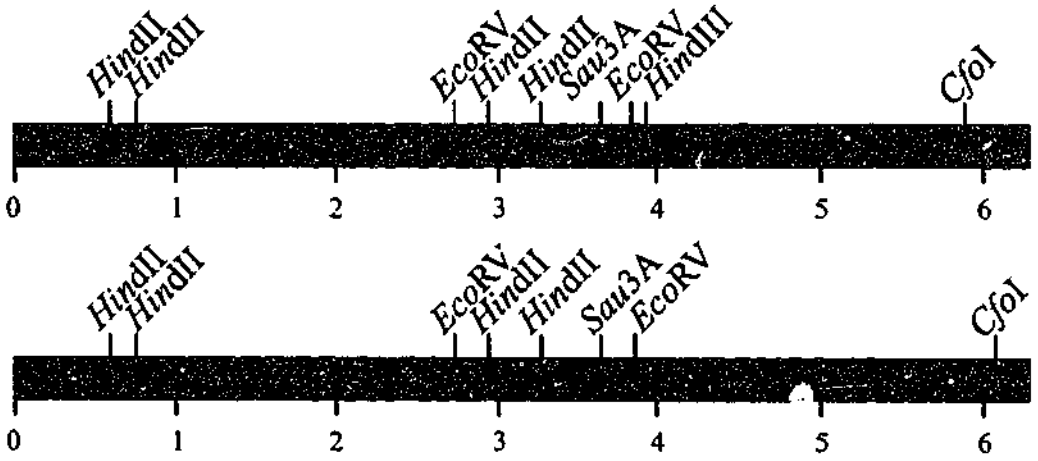
Figure 1.6 : The Tn*4451*/Tn*4453* family of mobilizable transposons.

Tn*4451*, Tn*4453a* and Tn*4453b* are shown as linear maps. The approximate extent of each of the ORFs is indicated by the arrows below the Tn*4451* map. The genes encoding chloramphenicol acetyltransferase, *catP*, the large resolvase, *tnpX*, and the mobilization protein, *tnpZ*, are indicated by green, blue and orange arrows respectively. The RS_A site at which the TnpZ protein acts is indicated by a yellow circle in the Tn*4451* map. Restriction sites for *Cfo*I, *Eco*RV, *Hin*dIII, *Nsi*I and *Sau*3A are indicated. Modified from Figure 1 of Lyras *et al.* (1998).



Tn*4451*

C. perfringens



Tn*4453a*

C. difficile

Tn*4453b*

TnpX catalyzes the formation of 2 bp staggered nicks on either side of conserved GA dinucleotides, which are located on either end of both mobilizable elements (Bannam *et al.*, 1995; Crellin and Rood, 1997; Lyras *et al.*, 1998). Excision of the elements results in the formation of circular transposition intermediates (Lyras and Rood, 2000) that carry a GA dinucleotide at their junctions (Crellin and Rood, 1997).

Integrons

Integrons are specialized gene capture and expression elements that are generally found in Gram-negative bacteria (Hall and Collis, 1995). They consist of an integrase gene, *intI*, a recombination site known as *attI*, a promoter, a 3' conserved region and one or more gene cassettes, which are mobile elements in their own right (Brown *et al.*, 1996; Hall and Collis, 1995; Recchia and Hall, 1995). Mobile gene cassettes normally encode only two functional components, a resistance gene and a recombination site (known as a 59-base element, or 59-be) that is located downstream of the gene (Recchia and Hall, 1997). Cassettes can exist in two forms, either as a free circular molecule that is unable to replicate, or integrated at the *attI* site in an integron. Most cassettes include very little non-coding sequence and do not contain a promoter. Therefore, the expression of cassette-associated genes is dependent on integration of the cassette, *via* site-specific recombination between the *attI* site and the 59-be, in the correct orientation into an integron that supplies an upstream promoter (Recchia and Hall, 1997). The majority of known cassette-associated genes encode resistance to an antibiotic and include inner membrane efflux pumps, acetyltransferases that modify chloramphenicol or aminoglycosides, adenylyltransferases that modify aminoglycosides, trimethoprim-resistance dihydrofolate reductases and β-lactamases (Recchia and Hall, 1997).

There are four types of integrons, designated as groups I-IV (Recchia and Hall, 1995). The basic integron, In0, includes no gene cassettes and consists of a 5' conserved sequence, including the *intI* gene, *attI* site and the promoter P_{ant}, followed by a 3' conserved sequence, which, in the case of Group I integrons, consists of a sulphonamide resistance gene, *sulI*, and several complete or partial ORFs of unknown function (Bennett, 1999). Group I integrons encode IntI1 and normally one or more gene cassettes. Group I integrons are well studied and have been shown to excise and integrate their gene cassettes (Recchia and Hall, 1995). Group II integrons consist of elements encoding a defective integrase gene, *intI2*, are found at the left end of Tn*7*-like elements, and also lack the 3' conserved region (Bennett, 1999). There is only one Group III integron identified (Bennett, 1999), which encodes the *intI3* gene (similar to *intI1*). However, the element has not been fully sequenced and so further study is necessary (Recchia and Hall, 1995). The super-integrons, or Group IV integrons, are so-called because of their extremely large size, which is due to the presence of hundreds of integrated gene cassettes (Rowe-Magnus *et al.*, 1999). For example, *Vibrio cholerae* contains a super-integron that is 126 kb in length and consists of at least 179 gene cassettes, corresponding to approximately 10% of the organism's genome (Rowe-Magnus *et al.*, 1999). An integrase gene, *intI4*, is located upstream of the first cassette along with an *attI* site and a promoter (Rowe-Magnus *et al.*, 1999), similar to the structure of the other three integron groups.

The Clostridia

The genus *Clostridium* consists of a diverse group of organisms that are classically defined as Gram-positive, endospore-forming, anaerobic, rod-shaped

bacteria (Stackebrandt and Rainey, 1997). Although most species within the genus are obligate anaerobes, some are aerotolerant and a few, such as *Clostridium carnis*, *Clostridium histolyticum* and *Clostridium tertium* are able to grow under aerobic conditions (Wells and Wilkins, 1991).

Of the 120 species that make up the genus, 35 are considered pathogenic for either humans or animals (Stackebrandt and Rainey, 1997). The major pathogens include the neurotoxigenic clostridia, *Clostridium botulinum* and *Clostridium tetani*, the enterotoxigenic *C. difficile* and the enterotoxic and histotoxic *C. perfringens*.

Clostridium perfringens

C. perfringens is commonly found in the gastrointestinal tract of humans and animals as well as in the soil and sewage. *C. perfringens* is different from many other clostridia in that it is non-motile and sporulation *in vitro* occurs only in specialized culture media (Rood and Cole, 1991). It is the causative agent of a number of human diseases, including gas gangrene (clostridial myonecrosis), food poisoning, necrotizing enterocolitis of infants and enteritis necroticans (pigbel) (Rood and Cole, 1991). It is also responsible for a number of animal diseases such as lamb dysentery, ovine enterotoxaemia (struck) and ovine pulpy kidney disease (Songer, 1996).

The ability of this organism to cause disease is due to the production of numerous extracellular toxins and enzymes including α-toxin, β-toxin, ε-toxin, θ-toxin, κ-toxin, λ-toxin, ι-toxin, μ-toxin and sialidase (Rood, 1998). Isolates of *C. perfringens* can be divided into five types (A to E) based on the particular

extracellular toxins which they produce, and each of these toxin types is responsible for specific disease syndromes (Rood and Cole, 1991). Type A strains are generally responsible for disease in humans and type B, C, D and E are associated with animal syndromes (Rood and Cole, 1991).

C. perfringens is considered the paradigm species for genetic studies on the pathogenic clostridia, primarily because of its oxygen tolerance, relatively fast growth rate and ability to be genetically manipulated (Rood, 1998). The development of methods for the introduction of plasmid DNA, such as electroporation-mediated transformation (Allen and Blaschek, 1988; Scott and Rood, 1989) and *E. coli-C. perfringens* conjugation (Lyras and Rood, 1998), the construction of several well characterized *E. coli-C. perfringens* shuttle vectors (Bannam and Rood, 1993; Lyras and Rood, 1998; Sloan *et al.*, 1992), and the development of methods for transposon mutagenesis and homologous recombination (Awad *et al.*, 2000; Awad and Rood, 1997; Lyristis *et al.*, 1994) has enabled the analysis of genes involved in *C. perfringens* toxin production as well as many other *C. perfringens* genes (Rood, 1997).

Clostridium difficile

C. difficile is commonly found in the intestine and faeces of healthy infants, in the hospital environment and also in the soil (Brazier, 1998). It is not a common gastrointestinal tract inhabitant of healthy adults. *C. difficile* has been identified as the causative agent of a spectrum of chronic gastrointestinal syndromes in humans, ranging from mild diarrhoea, through moderately severe disease with watery diarrhoea, abdominal pain and systemic upset, to life-threatening and sometimes fatal pseudomembranous colitis (PMC) (Borriello, 1998). *C. difficile*-associated disease

(CDAD) is also of veterinary significance and has been detected in hamsters (Bartlett *et al.*, 1977), dogs (Berry and Levett, 1986), horses (Madewell *et al.*, 1995) and captive ostriches (Frazier *et al.*, 1993).

C. difficile is unique among enteric pathogens in that disease development is almost always associated with prior antimicrobial therapy. One of the main defences against *C. difficile* colonization is the maintenance of a normal intestinal ecosystem. Even when it is present in the colon, *C. difficile* is normally suppressed by other components of the intestinal flora, so-called 'colonization resistance', and usually produces no symptoms. The use of antibiotics disrupts this normal ecosystem and allows *C. difficile* to become established and colonize the intestinal tract. Although most antibiotics have been associated with predisposition to *C. difficile* infection, the most commonly implicated have been clindamycin, cephalosporins and ampicillin (Spencer, 1998b). Rarely implicated antibiotics include quinolones, rifampicin, trimethoprim, aminoglycosides and the antibiotics that are normally used for treatment of CDAD, vancomycin and metronidazole (Spencer, 1998b).

In hospitals, patients receive antibiotics in an environment where *C. difficile* is highly prevalent, and as a result it is the most commonly diagnosed cause of infectious diarrhoea in hospitalized patients (Kelly *et al.*, 1994; McFarland *et al.*, 1989). *C. difficile* is recognized as the major cause of nosocomial diarrhoea in the U.S.A. (Gorbach, 1999) and is a significant nosocomial pathogen in both British (Wilcox, 1998) and Australian hospitals (Riley *et al.*, 1995). It has a significant economic impact, estimated at approximately $1 million per year for the average regional hospital (Riley *et al.*, 1995; Spencer, 1998a).

C. difficile produces a number of factors that contribute to its virulence. Some factors contribute directly to the pathology associated with infection, while others enable *C. difficile* to colonize and produce the toxins that directly contribute to disease. Adhesins mediating adherence to mucosa (Eveillard *et al.*, 1993; Karjalainen *et al.*, 1994; Waligora *et al.*, 1999; Waligora *et al.*, 2001), fimbriae (Borriello *et al.*, 1988), flagella (Tasteyre *et al.*, 2000a; Tasteyre *et al.*, 2000b; Tasteyre *et al.*, 2001), capsules (Davies and Borriello, 1990), S-layer proteins (Kawata *et al.*, 1984; McCoubrey and Poxton, 2001; Sharp and Poxton, 1988) and proteolytic enzymes (Poilane *et al.*, 1998; Seddon and Borriello, 1992) have all been identified in various *C. difficile* strains but the function of these factors in pathogenesis has not been elucidated. Presumably one or more of these factors enable *C. difficile* to colonize and penetrate the intestinal mucosa.

In addition to these accessory virulence factors, *C. difficile* produces at least five toxins (Borriello, 1998), although only toxins A and B have been studied in any detail. The other three toxins are an unstable, enterotoxic protein (Banno *et al.*, 1984; Giuliano *et al.*, 1988; Mitchell *et al.*, 1987), a high molecular weight protein that causes changes in electrical potential in isolated segments of rabbit intestine (Justus *et al.*, 1982), and an actin-specific ADP-ribosyltransferase (Popoff *et al.*, 1988; Braun *et al.*, 2000; Gülke *et al.*, 2001; Perelle *et al.*, 1997; Stubbs *et al.*, 2000).

Toxins A and B are the largest known single polypeptide bacterial toxins. Toxin A is an extremely potent enterotoxin and causes extensive damage to the intestine (Lima *et al.*, 1988; Lyerly *et al.*, 1982). The villus tips of the epithelium are initially disrupted, followed by damage to the brush border membrane. The mucosa eventually becomes eroded. Damage to the intestinal mucosa is accompanied by

extensive infiltration with inflammatory neutrophils, which probably plays an important role in the extensive damage that is caused. The fluid response presumably results from the tissue damage, however, it has been observed that the cytotoxic activity of toxin A results in disruption of tight junctions and this alone may play a role in the fluid loss (Hecht *et al.*, 1988). Toxin B is a potent cytotoxin, however, it does not cause a fluid response in animal models (Lima *et al.*, 1988; Lyerly *et al.*, 1982). Toxin A, which is able to bind to specific carbohydrate receptors on the surface of intestinal cells, initiates damage to the intestine and toxin B then gains access to the underlying tissues and contributes to the extensive damage during the course of disease (Moncrief *et al.*, 1997). Both toxin A and toxin B are UDP-glucose dependent monoglucosyltransferases that glucosylate the RhoA protein, a small GTP-binding protein that is required for actin polymerization (Aktories *et al.*, 1997; Just *et al.*, 1995). The net result is F-actin depolymerization, destruction of gastrointestinal tight junctions, and fluid loss to the intestinal lumen and diarrhoea.

Unlike *C. perfringens*, the genetics of *C. difficile* is poorly developed, which has significantly hampered the analysis of the genes encoding the toxins and other virulence factors. There are no transformation methods available and *C. difficile* remains one of the few bacterial species that at this time is still not amenable to electroporation. Two methods for the introduction of recombinant DNA molecules into *C. difficile* have been reported. The first of these methods (Mullany *et al.*, 1994) involves cloning of the required fragment into an *E. coli* plasmid that contains a portion of the conjugative enterococcal transposon Tn*916*, introduction of the recombinant plasmid onto the *B. subtilis* chromosome by homologous recombination into a resident copy of Tn*916*ΔE, and the conjugative transfer of that derivative into

C. difficile followed by its transposition onto the *C. difficile* chromosome. This method is obviously very cumbersome and does not enable introduction of DNA at a high frequency.

More recently a *C. perfringens-E. coli* shuttle vector has been constructed that can be transferred by conjugation from *E. coli* to *C. perfringens* and is stably maintained as a plasmid in both species (Lyras and Rood, 1998). The RP4-mediated mobilization system can be used to transfer genes from *E. coli* to *C. botulinum* (Bradshaw *et al.*, 1998), *Clostridium septicum* and, most importantly, to *C. difficile* (D. Lyras and J. Rood, unpublished results). Transfer to *C. difficile* occurs at a very low frequency, with only a few transconjugants obtained from each mating, which is not high enough to enable chromosomal mutants to be constructed by insertional inactivation and homologous recombination. However, the conjugation method is simple and reproducible and the introduced plasmid is stably maintained in at least one strain of *C. difficile* (D. Lyras and J. Rood, unpublished results).

The first report of a targeted gene disruption in the *C. difficile* chromosome was made earlier this year (Liyanage *et al.*, 2001). These researchers used homologous recombination to inactivate the glycerol dehydrogenase gene, *gldA*, in the chromosome of strain CD37. A copy of *gldA* with an internal deletion was introduced into *C. difficile* by conjugation on a plasmid that was unable to replicate in Gram-positive organisms. The plasmid was able to integrate into the *gldA* gene on the *C. difficile* chromosome, thereby inactivating the chromosomal copy of this gene. The development of this technique represents a significant advancement in *C. difficile* genetics and may allow the analysis of many other genes. Unfortunately,

CD37 is a non-toxigenic strain of *C. difficile*, however, this technique may prove to be of value in more virulent isolates.

Antibiotic resistance determinants of C. perfringens *and* C. difficile

The work undertaken in this thesis focused on erythromycin (MLS) resistance determinants from *C. difficile* and *C. perfringens*. Antibiotic resistance determinants that have been characterized from these organisms not only include those that mediate resistance to erythromycin, but also determinants that mediate resistance to chloramphenicol and tetracycline (Lyras and Rood, 1997).

Chloramphenicol resistance in *C. perfringens* and *C. difficile*

Chloramphenicol inhibits bacterial growth at the level of protein synthesis by binding to the 50S ribosomal subunit and blocking elongation of the growing peptide chain by inhibiting peptidyl transferase (Gale *et al.*, 1981). Resistance to chloramphenicol is commonly mediated by the enzymatic modification and inactivation of the antibiotic, as a result of the action of chloramphenicol acetyltransferase (CAT). This enzyme catalyzes the transfer of an acetyl group from acetyl coenzyme A to the primary hydroxyl group of chloramphenicol (Shaw and Leslie, 1991). The acetylated chloramphenicol derivatives do not bind to bacterial ribosomes and therefore do not exhibit antimicrobial activity.

Chloramphenicol resistance in *C. perfringens* is not as common as erythromycin or tetracycline resistance (Rood *et al.*, 1985; Rood *et al.*, 1978a) and has been shown to be mediated by the production of CAT enzymes (Rood *et al.*,

1978b; Steffen and Matzura, 1989; Zaidenzaig *et al.*, 1979). Two resistance genes from distinct hybridization classes have been identified, *catP* (Rood *et al.*, 1989) and *catQ* (Bannam and Rood, 1991). The *catP* gene is located on the mobilizable transposons Tn*4451* (Figure 1.6) and Tn*4452*, which are located on the large conjugative plasmids pIP401 and pJIR27, respectively (Abraham and Rood, 1987; Bannam *et al.*, 1995). The *catQ* gene belongs to a different hybridization class, is chromosomally located, and does not appear to be associated with a transposon (Rood *et al.*, 1989).

The *catP* gene is also found in *C. difficile* although it was originally called *catD* in that organism. The *C. difficile catP* gene is identical to *C. perfringens catP* and has been shown to be associated with two mobilizable transposons, Tn*4453a* and Tn*4453b*, which are structurally and functionally similar to Tn*4451* (Figure 1.6) (Lyras and Rood, 2000; Lyras *et al.*, 1998). Unlike the *C. perfringens catP* or *catQ* genes, *catP* is present in multiple copies on the *C. difficile* chromosome (Lyras *et al.*, 1998; Wren *et al.*, 1988).

Tetracycline resistance in *C. perfringens* and *C. difficile*

The tetracyclines inhibit protein synthesis by binding to the 30S ribosomal subunit thereby blocking the binding of aminoacyl-tRNA molecules (Levy, 1984). Bacterial resistance to the tetracyclines is mediated by two major mechanisms involving either the active efflux of the drug out of the cell (Levy, 1992) or ribosomal modification, which protects the ribosomes from the action of the antibiotic (Chopra and Roberts, 2001).

a) The C. perfringens *Tet P determinant*

Tetracycline resistance is the most common antibiotic resistance phenotype observed in *C. perfringens* (Rood, 1983; Rood and Cole, 1991). In most strains this resistance is non-transferable. However, conjugative transfer of tetracycline resistance has been shown to occur (Brefort *et al.*, 1977; Rood, 1983; Rood *et al.*, 1985; Rood *et al.*, 1978b). In all the conjugative strains examined, transfer of resistance is mediated by large conjugative plasmids, which are either identical or closely related to the prototype conjugative *C. perfringens* R-plasmid, pCW3 (Abraham and Rood, 1985b). The best characterized tetracycline resistance determinant from *C. perfringens*, Tet P, was isolated from pCW3 and has been found on all known tetracycline resistance plasmids from this organism (Abraham *et al.*, 1988; Abraham and Rood, 1985a; Abraham and Rood, 1985b; Saksena and Truffaut, 1992). The Tet P determinant encodes two functional tetracycline resistance genes, designated *tetA*(P) and *tetB*(P) (Sloan *et al.*, 1994). These genes overlap by 17 bp and comprise an operon, which is transcribed from a single promoter located upstream of the *tetA*(P) start codon (Johanesen *et al.*, 2001). The presence of two functional tetracycline resistance genes in an operon represents a novel genetic arrangement for tetracycline resistance determinants (Sloan *et al.*, 1994).

The *tetA*(P) gene encodes a 420 amino acid, 46 kDa protein, TetA(P), which is responsible for the active efflux of tetracycline from the cell (Sloan *et al.*, 1994). TetA(P) is predicted to have 12 membrane spanning helices, which is typical of tetracycline efflux proteins, however, the prototype tetracycline efflux proteins all have two related six transmembrane domains that are separated by a large central hydrophilic loop. Instead, TetA(P) is predicted to have two major hydrophilic domains that are not centrally located (Bannam and Rood, 1999; Kennan *et al.*, 1997;

Sloan *et al.*, 1994). Site-directed (Kennan *et al.*, 1997) and random mutagenesis (Bannam and Rood, 1999) has identified many residues, which, when mutated, abolish tetracycline resistance, suggesting they are important for tetracycline efflux.

The *tetB*(P) gene encodes a putative 72.6 kDa hydrophilic protein that has significant amino acid sequence identity to Tet M-like cytoplasmic tetracycline resistance proteins that mediate tetracycline resistance by ribosomal modification/protection. Due to its sequence identity, TetB(P) is proposed to function in a similar manner (Sloan *et al.*, 1994), however, the function of TetB(P) has not been experimentally proven. Note that *tetB*(P) does not hybridize with the *tet*(M) gene from Tn*916*, indicating that *tetB*(P) is significantly different from the *tet*(M)-like genes and supporting its designation in a separate hybridization class (Lyras and Rood, 1996).

In a study on the distribution of the Tet P determinant in *C. perfringens*, all 81 tetracycline resistant isolates examined were shown to carry the *tetA*(P) gene. Of these strains, 93% were also found to carry a second tetracycline resistance gene, with 53% carrying the *tetB*(P) gene and 40% carrying a *tet*(M)-like gene (Lyras and Rood, 1996). Hybridization analysis with other tetracycline resistant clostridial isolates has shown the presence of *tetA*(P) in *Clostridium paraputrificum* (Lyras and Rood, 1996), *Clostridium septicum* (Sasaki *et al.*, 2001) and *Clostridium sordellii* (Sasaki *et al.*, 2001), and the presence of the *tetB*(P) gene in *C. septicum* and *C. sordellii* (Sasaki *et al.*, 2001) indicating that these genes are not confined to *C. perfringens*. No hybridization of the *tetA*(P) and *tetB*(P) probes was observed in nine *C. difficile* isolates, all of which hybridized to the *tet*(M) probe (Lyras and Rood, 1996).

b) The C. perfringens *Tet M determinant*

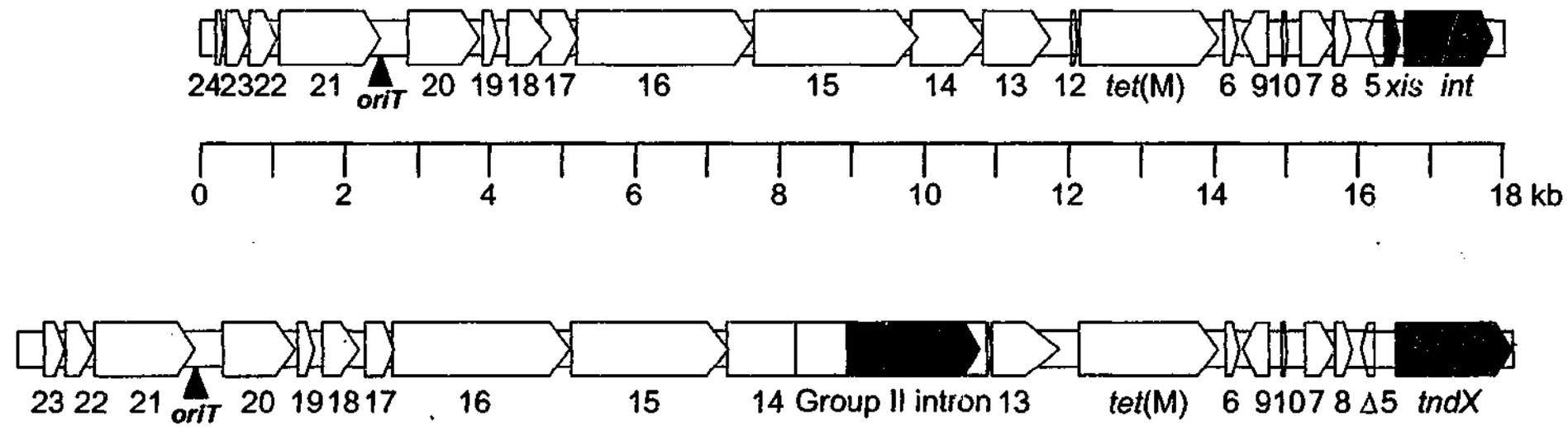
The detection of a *tet*(M)-like gene in *C. perfringens* (Lyras and Rood, 1996) was unexpected as this resistance determinant had not been previously reported in this organism. Due to the common association of the *tet*(M) gene with conjugative transposons, hybridization analyses were performed using probes specific for the left and right ends of Tn*916*. These probes hybridized to *tet*(M) carrying strains, indicating that the *C. perfringens tet*(M)-like gene may be associated with a conjugative transposon (D. Lyras and J.I. Rood, unpublished data). Recently, the tetracycline resistance determinant from one of these strains, CW459, has been cloned and analyzed. The *tet*(M)-like gene is associated with an element with similarity to Tn*916* and also to Tn*5397* from *C. difficile* (Figure 1.7) (Roberts *et al.*, 2001). This study revealed that the three elements have similar conjugation regions but different insertion and excision modules. In Tn*916* the products of the *xis* and *int* genes are essential for excision and insertion. In the clostridial elements, these genes have been replaced with other genes that encode different site-specific recombinases, *tndX* on Tn*5397*, and *int459* on the *C. perfringens tet*(M) element. However, conjugative transfer of the *C. perfringens tet*(M) element has not been demonstrated so it appears that the *int459* protein either is not produced, is not functional, or is not sufficient on its own for transposition (Roberts *et al.*, 2001).

c) The C. difficile *Tet M determinant*

Studies on tetracycline resistance in *C. difficile* are not as well advanced as those in *C. perfringens*. Three resistance determinants of classes L, K, and M have been reported (Hächler *et al.*, 1987b; Roberts *et al.*, 1994), however, only the presence of the *tet*(M) gene has been confirmed by cloning and sequence analysis. Early studies on the tetracycline resistant strain, 630, revealed that resistance was

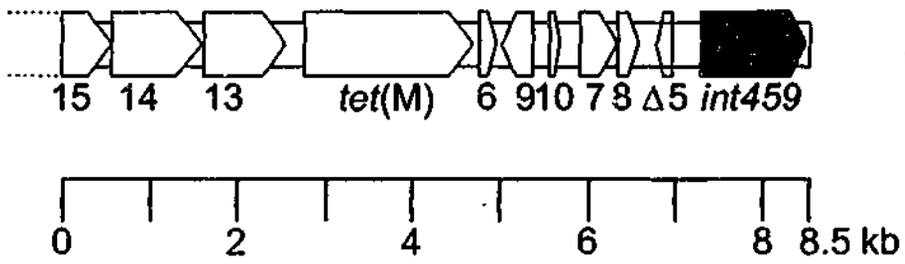
Figure 1.7 : Genetic organization of Tn*916*, Tn*5397* and the *C. perfringens tet*(M) element.

The genetic organization of the three tetracycline resistance elements is shown schematically to scale. The size of each of the elements is indicated by the scale bar below each element. Each of the ORFs is indicated by a block arrow, with the proposed direction of transcription shown by the arrow. The ORFs are labelled below the block arrows. The ORFs comprising the insertion and excision module for each element are colored red. The group II intron in Tn*5397* is colored green. The functional *oriT* sites on Tn*916* and Tn*5397* are indicated by black vertical arrows. Based on Figure 2 from Roberts *et al.* (2001).



Tn*916*

Tn*5397*



***C. perfringens*-*tet*(M) element**

transferable (Smith *et al.*, 1981; Wüst and Hardegger, 1983), and subsequent experiments showed that the resistance determinant was encoded by a *tet*(M) gene located on a Tn*916*-like conjugative transposon, Tn*5397* (Hächler *et al.*, 1987b; Mullany *et al.*, 1996; Mullany *et al.*, 1990).

Tn*5397* is a modular transposon whose central region is very similar to that of Tn*916* but whose ends are different (Mullany *et al.*, 1990) (Figure 1.7). As previously mentioned, Tn*5397* does not contain the *xis* and *int* genes that are required for the excision and integration of Tn*916*. Instead, it contains the *tndX* gene, the product of which is a member of the large resolvase family of site-specific recombinases. TndX is most closely related to TnpX from Tn*4451* and, like TnpX, is responsible for the excision and circularization of a transposon, Tn*5397* (Wang and Mullany, 2000). Tn*5397* also differs from Tn*916* in that it contains a group II intron inserted into a gene that is almost identical to *orf14* from Tn*916* (Mullany *et al.*, 1996).

Hybridization analysis of other *tet*(M) carrying *C. difficile* isolates has revealed that other Tn*916*-like conjugative elements, which differ from Tn*5397*, may also be present in *C. difficile*, as some isolates, while hybridizing to probes specific for the right and left ends of Tn*916*, do not hybridize to a Tn*5397*-specific probe (D. Lyras, P. Mullany and J.I. Rood, unpublished results). Recently, a second element, designated Tn*916*CD, has been characterized from an environmental isolate of *C. difficile* and has been shown by PCR and sequence analysis to be identical to Tn*916* (Wang *et al.*, 2000b).

MLS resistance in *C. perfringens* and *C. difficile*

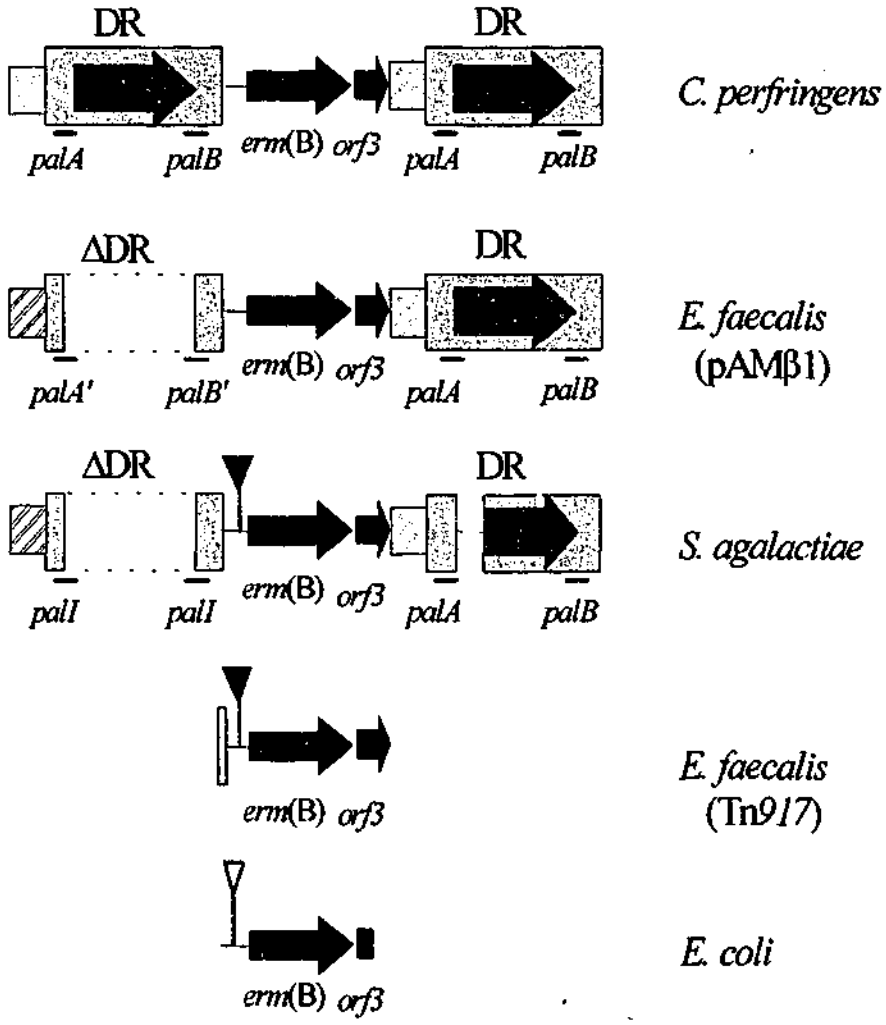
a) C. perfringens MLS resistance genes

Resistance to erythromycin was first reported in *C. perfringens* (Sebald *et al.*, 1975) in strain CP590, which carried not only the conjugative tetracycline and chloramphenicol resistance plasmid pIP401 but was also resistant to MLS antibiotics. Further studies on this strain indicated that the MLS resistance determinant was carried on a large non-conjugative plasmid, pIP402, (Brefort *et al.*, 1977), which is not widespread in *C. perfringens* isolates. Cloning and molecular analysis of this determinant indicated that it belonged to the Erm B class of MLS resistance determinants and it was designated *ermBP* (*ermBP* is now called *erm*(B) in accordance with the revised nomenclature (Roberts *et al.*, 1999)).

The *C. perfringens erm*(B) gene is identical to the *erm*(B) gene from the promiscuous *E. faecalis* plasmid pAMβ1 and has at least 98% nucleotide sequence identity to other members of the Erm B class of MLS resistance genes (Berryman and Rood, 1995). It is not preceded by a leader peptide sequence, which suggests that like the *erm*(B) gene from pAMβ1 (Brehm *et al.*, 1987; Martin *et al.*, 1987) it is constitutively expressed. The *C. perfringens* Erm B determinant consists of the *erm*(B) gene flanked by two, almost identical, directly repeated sequences designated DR1 and DR2, respectively (Figure 1.8) (Berryman and Rood, 1995). Downstream of the *erm*(B) gene is a small ORF, designated *orf3*. Each DR contains an open reading frame, *orf298*, flanked by the highly palindromic sequences *palA* and *palB* (Figure 1.8)(Berryman and Rood, 1995). The putative protein encoded by *orf298* has low levels of identity to ParA (Easter *et al.*, 1998) and Soj (Sharpe and Errington, 1996) proteins, which are involved in plasmid and chromosomal partitioning.

Figure 1.8 : Comparative genetic organization of the Erm B determinants.

The approximate extent and organization of the Erm B determinants from *C. perfringens* (Berryman and Rood, 1995), pAMβ1 (Martin *et al.*, 1987), *S. agalactiae* plasmid pIP501 (Pujol *et al.*, 1994), *E. faecalis* transposon Tn*917* (Shaw and Clewell, 1985) and the *E. coli* plasmid pIP1527 (Brisson-Noël and Courvalin, 1988) are shown, not necessarily to scale. Regions of similarity are indicated by similar colors. The solid arrows represent the individual ORFs and their respective directions of transcription. The approximate locations of the palindromic sequences (*palA* and *palB*) are indicated by the black lines below the grey filled rectangles. The *palA'/palB'* and *palI* sequences represent the portions of the *C. perfringens* derived *palA* and *palB* homologues that are present at the ends of the deletions in the DRs from the pAMβ1 and pIP501 variants. Functional and non-functional leader peptide sequences are represented by the solid and open blue triangles, respectively. Based on Figure 2 from Berryman and Rood (1995).


DR
DR
palA
palB
ermA(B)
orf3
palA
palB
C. perfringens
ΔDR
DR
palA'
palB'
ermA(B)
orf3
palA
palB
E. faecalis
(pAMβ1)
ΔDR
DR
palI
palI
ermA(B)
orf3
palA
palB
S. agalactiae
ermA(B)
orf3
E. faecalis
(Tn917)
ermA(B)
orf3
E. coli

Comparative analysis has shown that both the pAMβ1 (*E. faecalis*) and pIP501 (*Streptococcus agalactiae*) Erm B determinants have DR2 but have an internal deletion in DR1 that removes *orf298* (Figure 1.8). The deletion endpoints are both located within the *palA* and *palB* sequences, but they appear to have arisen from separate deletion events because the deletion points are different. It has been suggested that the *C. perfringens* Erm B determinant represents the progenitor Erm B determinant and that other Erm B determinants have arisen through homologous recombination events between the *palA* and *palB* sites of the progenitor (Berryman and Rood, 1995).

Examination of erythromycin resistant *C. perfringens* strains that did not hybridize with a *C. perfringens erm*(B) probe, or with probes from any of the other Erm hybridization classes, led to the cloning and sequencing of the *erm*(Q) gene (Berryman *et al.*, 1994). The Erm Q determinant has between 31% and 57% sequence identity at the nucleotide level, and 20% to 43% identity at the amino acid level, with determinants from other Erm classes, confirming that Erm Q represents a different class of MLS resistance determinant (Berryman *et al.*, 1994). The Erm Q determinant represents the most common erythromycin resistance determinant in *C. perfringens*, which may reflect differences in the mechanisms by which *erm*(Q) and *erm*(B) are disseminated (Berryman *et al.*, 1994).

b) MLS resistance in C. difficile

Resistance to MLS antibiotics in *C. difficile* has been of great interest to many researchers because of the association between clindamycin, and to a lesser degree erythromycin, and the development of *C. difficile* antibiotic-associated diarrhoea or colitis. Initial studies of erythromycin resistance in *C. difficile*

demonstrated that MLS resistance could be transferred by conjugation to *C. difficile* (Wüst and Hardegger, 1983), *S. aureus* (Hächler *et al.*, 1987a) and *B. subtilis* (Mullany *et al.*, 1995) recipients. Transfer of MLS resistance was observed to occur in the absence of detectable plasmid DNA (Hächler *et al.*, 1987a; Mullany *et al.*, 1995; Wüst and Hardegger, 1983) and appeared to involve a chromosomal determinant, which was subsequently designated Tn*5398* (Mullany *et al.*, 1995). This element has not been analyzed by cloning or sequence analysis and its size is unknown. Tn*5398* was identified in strain 630, which also carries the tetracycline resistance transposon Tn*5397*.

The MLS resistance gene carried on Tn*5398* has been shown to belong to the Erm B class of MLS resistance determinants using the *erm*(B) genes from Tn*551* (Hächler *et al.*, 1987a) and pIP402 (Berryman and Rood, 1989) as probes. The *C. difficile* MLS resistance gene has been designated as *erm*(B) (Roberts *et al.*, 1999), but was previously known as *ermZ* (Hächler *et al.*, 1987a) and *ermBZ* (Berryman and Rood, 1989). The *erm*(B) gene has been cloned and sequenced from *C. difficile* strain 630 (Farrow, 1995). Hybridization analysis has shown that the *erm*(B) gene is widespread amongst erythromycin resistant *C. difficile* isolates (Berryman and Rood, 1995; Hächler *et al.*, 1987a; Roberts *et al.*, 1994), which is not unexpected due to its putative association with a conjugative transposon.

Hybridization analysis of a large range of *C. difficile* isolates has also suggested the presence of other MLS resistance determinants. These include both the *erm*(Q) gene and *erm*(F) from *Bacteroides fragilis* (Roberts *et al.*, 1994). However, the presence of these genes has not been confirmed by either cloning or sequence analysis (Lyras and Rood, 1997).

Aims and objectives

Cross resistance to MLS antibiotics is becoming an increasingly common phenotype in many bacteria due to the acquisition of MLS resistance determinants. Many of these determinants are located on transposable and mobilizable elements that can be transferred both inter- and intra-generically, which may explain the large number of bacterial species now exhibiting this phenotype (Roberts *et al.*, 1999).

Resistance to MLS antibiotics has been detected in both *C. difficile* and *C. perfringens* and has been shown to be mediated, in some resistant strains, by the presence of *erm* genes from the Erm B class of MLS resistance determinants. The *erm*(B) gene from *C. perfringens* is located on the large mobilizable plasmid pIP402 and has been cloned and sequenced (Berryman and Rood, 1995). It shares significant identity (>98%) to most other members of the Erm B class of determinants. The work presented in Chapter Three of this thesis aims to determine functionally and structurally important residues of the Erm(B) protein. Random mutagenesis performed on an *E. coli-C. perfringens* shuttle vector carrying the *C. perfringens erm*(B) gene identified several residues of the Erm(B) protein that, when mutated, either significantly reduced or completely abolished erythromycin resistance. The location of these residues in the structure of the Erm(B) protein suggests that these amino acids are likely to be important in either the binding of the substrate *S*-adenosylmethionine (SAM), the transfer of the methyl group from SAM to the target rRNA residue, or in the binding of the rRNA target.

MLS resistance in *C. difficile* has been shown by hybridization analysis to be mediated by an *erm*(B) gene (Berryman and Rood, 1989; Hächler *et al.*, 1987a; Mullany *et al.*, 1995) that can be transferred both inter- and intra-generically in the

absence of plasmid DNA (Hächler *et al.*, 1987a; Mullany *et al.*, 1995; Wüst and Hardegger, 1983). The *erm*(B) gene has been postulated to reside on a conjugative transposon, Tn*5398* (Mullany *et al.*, 1995), however, neither the Erm B determinant nor Tn*5398* have been characterized. The work presented in Chapter Four of this thesis aimed to clone Tn*5398* from *C. difficile* strain 630, with the objectives of delineating this conjugative element and determining which genes were potentially involved in its transfer, and analyzing the Erm B determinant that it carries. The results showed that Tn*5398* is approximately 9.6 kb in size and carries a novel Erm B determinant that contains two identical *erm*(B) genes. Tn*5398* does not appear to encode proteins that are typically involved in conjugative transposition, such as transposases, integrases, resolvases, or mobilization and transfer proteins. It appears from this analysis that Tn*5398* is not a conjugative transposon but may be a mobilizable element.

It is not known if the *erm*(B) gene represents the most common MLS resistance determinant in *C. difficile*. The work presented in Chapter Five aims to give a broader view of the type of Erm determinants carried by *C. difficile* isolates from different geographical locations. The results of PCR and DNA hybridization studies conducted showed that there was considerable heterogeneity in the arrangement of the *erm*(B) gene region in *C. difficile* isolates and, furthermore, that not all *erm*(B) gene regions were associated with Tn*5398*-like elements.

CHAPTER TWO

MATERIALS AND METHODS

Bacterial strains and culture conditions

The *E. coli*, *C. difficile* and *C. perfringens* strains used in this study are listed in Table 2.1. *E. coli* strains were cultured at 37°C in 2 x YT agar or broth (Miller, 1972), or in SOC broth (Sambrook *et al.*, 1989), supplemented with ampicillin (100 μg/ml), erythromycin (150 μg/ml) or chloramphenicol (30 μg/ml). *C. difficile* strains were grown at 37°C in BHIS agar or broth (Smith *et al.*, 1981), supplemented with erythromycin (50 μg/ml), tetracycline (10 μg/ml) or rifampicin (20 μg/ml). *C. perfringens* strains were grown at 37°C in trypticase-peptone-glucose broth (TPG) (Rood *et al.*, 1978a), Brain Heart Infusion broth (BHI) (Oxoid, Hampshire, England), fluid thioglycollate medium (FTG) (Difco, Michigan, USA) or nutrient agar (Rood, 1983) supplemented with erythromycin (50 μg/ml), chloramphenicol (5 μg/ml), naladixic acid (10 μg/ml) or rifampicin (10 μg/ml). All agar cultures of *C. difficile* or *C. perfringens* were incubated in an atmosphere of 10% (v/v) H_2, 10% (v/v) CO_2 in N_2 in an anaerobic chamber (Coy Laboratory Products Inc., Michigan, U.S.A.) or in anaerobic jars (Oxoid, Hampshire, England). All media were sterilized by autoclaving at 121°C for 20 min.

E. coli and *C. perfringens* strains were stored in glycerol storage broth (3.7% (w/v) BHI broth, 50% (v/v) glycerol) at -20°C, or as freeze-dried cultures at room temperature. The latter were prepared by resuspending the cells from one agar

Table 2.1 : Characteristics and origin of bacterial strains used in this study.

Strain	Relevant Characteristics	Reference/Origin
E. coli		
DH5α	F⁻ ϕ80 d*lacZ*ΔM15Δ(*lacZYA –argF*)U169 *endA1 recA1 hsdR17* ($r_k^- m_k^+$)*deoR thi-1 supE44⁻ gyrA96 relA1*	Bethesda Research Laboratories
S17-1	RP4-2 (Tc::Mu, Km::Tn7) Tp^R, Sm^R	(Simon *et al.*, 1983)
BL21(DE3)	F⁻ *ompT hsdS*$_E$($r_B^- M_B^-$)*gal dcm* (DE3)	Novagen
DH12S	*mcr*Δ(*mrr-hsdRMS-mcrBC*) ϕ80 d*lacZ* Δ*mis* Δ*lacx74 dcoR recA1 araD139* Δ(*ara, leu*) *7697 galU galK* λ⁻ *nupG*/F' *proAB*$^+$ *lacI*^{q}*Z* Δ*mis*	Bethesda Research Laboratories
XL1-Red	*endA1 gyrA96 thi-1 hsdr17 supE44 relA1 lac mutD5 mutS mutT* Tn*10* (Tet^R)	Stratagene
LT101	F⁻ *hsdS20* ($r_B^- m_B^-$) *leu supE44 ara14 galK2 lacYI proA2 rpsL20 xyl-5 mtl-1 recA13 mcrB* Sm^R, Rif^R	(Palombo *et al.*, 1989)
JIR5268	BL21(DE3)(pRSETA), Ap^R	This Study
JIR5753	BL21(DE3)(pJIR1626), Ap^R Em^R	This Study
C. difficile		
630	Clinical Isolate, Em^R Tet^R Rif^6 Cm^S	(Wüst and Hardegger, 1983) (Zurich, Switzerland)
CD37	Clinical Isolate, Em^S Tet^S Rif^R Cm^S	(Smith *et al.*, 1981) (U.S.A.)
JIR1162	630 X CD37 Transconjugant, Em^R Tet^S Rif^R	This Study
JIR1164	630 X CD37 Transconjugant, Em^R Tet^R Rif^R	This Study
JIR1182	630 X CD37 Transconjugant, Em^R Tet^R Rif^R	This Study
JIR1184	630 X CD37 Transconjugant, Em^R Tet^R Rif^R	This Study
L289	Clinical Isolate, Em^R Tet^S	(Hayter and Dale, 1984) (Surrey, U.K.)
662	Clinical Isolate, Em^R Tet^R	(Wüst and Hardegger, 1983) (Zurich, Switzerland)
AM140	Clinical Isolate, Em^R Tet^S	Wilkinson, K. unpublished (U.S.A.)
AM480	Clinical Isolate, Em^R Tet^S	Institute of Medical and Veterinary Science (Adelaide, Australia)
AM1180	Clinical Isolate, Em^R Tet^R	LaTrobe Valley Hospital (Sale, Australia)

AM1182	Clinical Isolate, Em^R Tet^R	Royal Melbourne Hospital (Melbourne, Australia)
AM1185	Clinical Isolate, Em^R Tet^R	Royal Melbourne Hospital (Melbourne, Australia)
SGC0545	Clinical Isolate, Em^R Cm^R	(Wren *et al.*, 1988) (Brussells, Belgium)
B1	Clinical Isolate, Em^R	(Borriello, 1998) (U.K.)
KZ1604	Clinical Isolate, Em^R Tet^S Cm^S	(Nakamura *et al.*, 1987) (Japan)
KZ1610	Clinical Isolate, Em^R Tet^R Cm^S	(Nakamura *et al.*, 1987) (Japan)
KZ1614	Clinical Isolate, Em^R Tet^S Cm^R	(Nakamura *et al.*, 1987) (Japan)
KZ1623	Clinical Isolate, Em^R Tet^R Cm^S	(Nakamura *et al.*, 1987) (Japan)
KZ1655	Clinical Isolate, Em^R Tet^S Cm^S	(Nakamura *et al.*, 1987) (Japan)
660/2	Clinical Isolate, Em^R Tet^R	Pasteur Institute (France)
685	Clinical Isolate, Em^R Tet^R	Pasteur Institute (France)
24/5-507	Clinical Isolate, Em^R Tet^R Cm^S	Monash Medical Centre (Melbourne, Australia)
R5948	Clinical Isolate, Em^R Tet^S Cm^S	Public Health Laboratory Service. (Cardiff, U.K.)
J9/5602	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (New York, U.S.A.)
J9/5610	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (New York, U.S.A.)
J9/5627	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Massachusetts, U.S.A.)
J9/4478	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Massachusetts, U.S.A.)
J9p2/5644	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Florida, U.S.A.)
J9p2/5650	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Florida, U.S.A.)
J7/4224	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Arizona, U.S.A.)
J7/4290	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Arizona, U.S.A.)
B1/832	Clinical Isolate, Em^R Tet^S	(Johnson *et al.*, 1999) (Minneapolis, U.S.A.)
Y4/1323	Clinical Isolate, Em^S Tet^S	(Johnson *et al.*, 1999) (Minneapolis, U.S.A.)

K12p/5672	Clinical Isolate, Em^S Tet^S	(Johnson *et al.*, 1999) (Chicago, U.S.A.)
C. perfringens		
JIR325	Strain 13 Nal^R Rif^R	(Lyristis *et al.*, 1994)
CP592	CP590 (pIP402, pIP403), Em^R Tet^S Cm^S	(Brefort *et al.*, 1977)

Em^R, Tet^R, Rif^R, Cm^R, Nal^R, Tp^R, Sm^R: resistant to erythromycin, tetracycline, rifampicin, chloramphenicol, naladixic acid, trimethoprim, and streptomycin respectively.
Tet^S, Rif^S, Cm^S: sensitive to tetracycline, rifampicin and chloramphenicol respectively.

plate in 1 ml of Mist Dessicans solution [7.5% (w/v) D-glucose (Ajax Chemicals, New South Wales, Australia) in horse serum (CSL Ltd., Victoria, Australia)]. The cells were freeze dried in Samco Freeze Drying Ampules (Crown Scientific [Pharmaglass], Sydney, Australia) using a Speedy Centrifugal Freeze Dryer, model 5PS (Edwards High Vacuum Ltd., Crawley, England). *C. perfringens* strains were also stored in cooked meat media (Becton Dickinson, New Jersey, U.S.A.). C. *difficile* strains were stored as freeze dried cultures at room temperature, or in cooked meat media (Department of Microbiology, University of Melbourne, Victoria, Australia).

Recombinant plasmids and cloning vectors

Routine cloning experiments were carried out using the *E. coli* vector pWSK29 (Wang and Kushner, 1991). The T7 expression vector, pRSETA (Invitrogen, Groningen, Netherlands), was used to facilitate the overexpression and production of recombinant protein in *E. coli* cells. The *C. perfringens-E. coli* shuttle vectors, pJIR418 (Sloan *et al.*, 1992) and pJIR750 (Bannam and Rood, 1993), were used for the introduction of recombinant plasmids into *C. perfringens*. All plasmids used in this study are shown in Table 2.2.

Molecular methods

Isolation of chromosomal DNA from *C. difficile* and *C. perfringens*.

Chromosomal DNA was extracted from *C. difficile* and *C. perfringens* strains using a sarkosyl lysis method (Abraham and Rood, 1985b) with a few modifications. *C. difficile* strains were inoculated on to three BHIS plates (supplemented with the

Table 2.2 : Characteristics and origin of recombinant plasmids.

Plasmid	Relevant Characteristics*#	Reference/Origin
pWSK29	Cloning vector, 5.4 kb, Ap^R	(Wang and Kushner, 1991)
pRSETA	pUC derived expression vector, N-terminal 6xHis tag, 2.9 kb, Ap^R.	Invitrogen
pJIR418	*C. perfringens-E. coli* shuttle vector, 7.3 kb, Em^R Cm^R	(Sloan *et al.*, 1992)
pJIR750	*C. perfringens-E. coli* shuttle vector, 6.5 kb, Cm^R	(Bannam and Rood, 1993)
pJIR883	pJIR418(*erm*(B)-*E58K*), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR932	pJIR418 (*erm*(B)-*87**), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR934	pJIR418(*erm*(B)-*P164S*), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR971	pJIR418(*erm*(B)-*138**), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR973	pJIR418(*erm*(B)-*138**), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR977	pJIR418(*erm*(B)-*P164S*), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR1377	pUC18Ω(*Eco*RI: Tn*4453* from *C. difficile* strain W1, ~6.0 kb), 8.7 kb, Ap^R Cm^R	(Lyras *et al.*, 1998)
pJIR1537	pSU39Ω(*Eco*RI/*Xba*I: *tndX* from Tn*5397*, 2.3kb), Kn^R	(Wang *et al.*, 2000a)
pJIR1551	pJIR418(*erm*(B)-*138**), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR1570	pJIR418(*erm*(B)-*87**), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR1571	pJIR418(*erm*(B)-*P164S*), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR1576	pJIR418(*erm*(B)-*G37E*), Em^S Cm^R	Hydroxylamine Random Mutant
pJIR1594	pWSK29Ω(*Hin*dIII: *C. difficile* strain 630, 9.7 kb), Em^R	Recombinant
pJIR1595	pJIR418(*erm*(B)-*114+1**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1596	pJIR418(*erm*(B)-*16+4**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1597	pJIR418(*erm*(B)-*114+1**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1598	pJIR418(*erm*(B)-*16+4**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1599	pJIR418(*erm*(B)-*9+2**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1600	pJIR418(*erm*(B)-*114+1**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1602	pJIR418(*erm*(B)-*147**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1603	pJIR418(*erm*(B)-*L63P*), Em^S Cm^R	XL1-Red Random Mutant
pJIR1604	pJIR418(*erm*(B)-*114+9**), Em^S Cm^R	XL1-Red Random Mutant

pJIR1605	pJIR418(*erm*(B)-*16+4**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1606	pJIR418(*erm*(B)-*Q148P*), Em^S Cm^R	XL1-Red Random Mutant
pJIR1607	pJIR418(*erm*(B)-*112+3**), Em^S Cm^R	XL1-RedRandom Mutant
pJIR1608	pJIR418(*erm*(B)-*203**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1609	pJIR418(*erm*(B)-*203**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1610	pJIR418(*erm*(B)-*229**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1611	pJIR418(*erm*(B)-*87**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1613	pJIR418(*erm*(B)-*S170I*), Em^S Cm^R	XL1-Red Random Mutant
pJIR1614	pJIR418(*erm*(B)-*229**), Em^S Cm^R	XL1-Red Random Mutant
pJIR1615	pJIR418(*erm*(B)-*H42Y*), Em^S Cm^R	XL1-Red Random Mutant
pJIR1626	pRSETAΩ(*Bam*HI/*Asp*718: pJIR418 #6356/#6357 PCR product, 0.76 kb), (*erm*(B) wild-type)	Recombinant
pJIR1790	pWSK29Ω(*Asp*718: *C. difficile* strain 630, ~23 kb), Em^R	Recombinant
pJIR1847	pJIR750Ω(*Asp*718: pJIR418 #10515/#6357 PCR product, 0.79 kb), (*erm*(B) wild-type)	Recombinant
pJIR1848	pJIR750Ω(*Asp*718: pJIR883 #10515/#6357 PCR product, 0.79 kb), (*erm*(B)-*E58K*), Em^S Cm^R	Recombinant
pJIR1850	pJIR750Ω(*Asp*718: pJIR1576 #10515/#6357 PCR product, 0.79 kb), (*erm*(B)-*G37E*), Em^S Cm^R	Recombinant
pJIR1851	pJIR750Ω(*Asp*718: pJIR1606 #10515/#6357 PCR product, 0.79 kb), (*erm*(B)-*Q148P*), Em^S Cm^R	Recombinant
pJIR1852	pJIR750Ω(*Asp*718: pJIR1613 #10515/#6357 PCR product, 0.79 kb), (*erm*(B)-*S170I*), Em^S Cm^R	Recombinant
pJIR1853	pJIR750Ω(*Asp*718: pJIR1615 #10515/#6357 PCR product, 0.79 kb), (*erm*(B)-*H42Y*), Em^S Cm^R	Recombinant
pJIR1874	pJIR750Ω(*Asp*718: pJIR977 #10515/#6357 PCR product, 0.79 kb), (*erm*(B)-*P164S*), Em^S Cm^R	Recombinant

*Random mutants of pJIR418 have either introduced stop codons (*), e.g. *erm*(B)-*87**, or a frameshift which leads to a certain number of normal Erm(B) residues plus residues following the frameshift up to the next stop codon, e.g. *erm*(B)-*9+2**, or missense mutations, e.g. (*erm*(B)-*G37E*)

#Ap^R, Em^R, Cm^R, Kn^R: resistance to ampicillin, erythromycin, chloramphenicol and kanamycin respectively

appropriate antibiotics), and grown anaerobically overnight at 37°C. The growth from each plate was resuspended in BHIS broth and was then used to inoculate 100 ml of BHIS broth. *C. perfringens* strains were used to inoculate 6 ml of FTG medium and were grown overnight at 37°C. The entire FTG culture was then used to inoculate 100 ml of TPG broth. These cultures were grown at 37°C until late log phase. The cells were harvested by centrifugation at 4100 *g* for 15 min at room temperature and the supernatant discarded. The cell pellets were resuspended in 10 ml of TES buffer (500 mM Tris (pH 8.0), 5 mM Na_2EDTA, 50 mM NaCl) and washed by centrifugation at 3000 *g* for 15 min at room temperature. The supernatants were discarded and the cell pellets were stored at -20°C overnight, or until required. The cells were thawed and resuspended in 2 ml of 25% (w/v) sucrose in TES to which 0.4 ml of freshly prepared lysozyme solution (10 mg/ml in TES) was added. The suspension was incubated at 37°C for 30 min. EDTA (0.8 ml, 0.25 M) was then added and the suspension was incubated for a further 30 min. Finally, 3.6 ml of 2% (w/v) sarkosyl in TES was added and the suspension was incubated for 10 min at 37°C to lyse the cells. Cellular debris was removed by centrifugation at 27,000 *g* for 30 min at room temperature.

The supernatant was transferred to a graduated test tube to which 7.79 g of caesium chloride (Cabot, Pennsylvania, U.S.A.) had previously been added. The contents were mixed by inversion until all of the caesium chloride had dissolved and the volume was then adjusted to 11 ml with TES. The solution was placed in a Quick-Seal™ Centrifuge Tube (Beckman, California, U.S.A.) and overlayed with paraffin oil. Ethidium bromide (0.25 ml of a 10 mg/ml solution) was then added and the tube was heat sealed. Centrifugation was carried out at 260,000 *g* for 20 h at 20°C in a Beckman L8-70M ultracentrifuge, using a 70.1 Ti rotor.

After centrifugation the DNA was visualized under ultraviolet light and the chromosomal DNA band was extracted using an 18 gauge needle. Ethidium bromide was removed by repeated extraction with sodium chloride-saturated isopropanol until the preparation was colourless. The preparation was dialyzed against weak TE buffer (10 μM EDTA, 100 μM Tris, pH 8.0) to remove the caesium chloride and was concentrated to approximately 200 to 500 μl by pervaporation at room temperature. The DNA samples were stored at -20°C until required.

Plasmid DNA isolation and manipulation

a) Small scale isolation of* E. coli *plasmid DNA

i) Alkaline lysis method

E. coli plasmid DNA was routinely extracted using an alkaline lysis method (Morelle, 1989). The appropriate *E. coli* strain was grown overnight at 37°C in 10 ml of 2 x YT broth supplemented with the relevant antibiotics. The cells were harvested by centrifugation at 1,500 *g* for 10 min at room temperature and the supernatant discarded. The cell pellet was resuspended in 200 μl of lysis buffer (50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA), and 400 μl of freshly prepared alkaline solution (0.2 M NaOH, 1% (w/v) SDS) was then added. The solution was mixed by inversion and then incubated on ice for 5 min. Ammonium acetate (300 μl of a 7.5 M, pH 7.8 solution) was then added and the solution was again mixed by inversion. The solution was then incubated on ice for 10 min prior to centrifugation at 12,000 *g* for 5 min. The supernatant was transferred to a fresh microfuge tube and RNase A (Sigma Chemical Co., Missouri, U.S.A.) was added to a final concentration of 20 μg/ml. The sample was incubated at room temperature for 10 min prior to the addition of 0.6 volumes of isopropanol to precipitate the DNA. The solution was

incubated for a further 10 min at room temperature and then centrifuged at room temperature for 10 min at 12,000 *g* to pellet the DNA. The DNA pellet was washed with cold 70% (v/v) ethanol, dried under vacuum using a Savant Speedvac SVC100 for approximately 5 min, and then resuspended in 100 µl of distilled water (dH_2O). The DNA preparation was incubated at 37°C for 10 min, then centrifuged at 12,000 *g* for 3 min at room temperature. The clear supernatant containing the purified DNA was transferred to a fresh tube and was stored at -20°C until use.

ii) Applied Biosystems method

When high quality plasmid DNA was required for cloning or automated sequencing purposes, DNA was extracted using the modified alkaline lysis/PEG precipitation procedure outlined in the PRISM™ Ready Reaction Dye Deoxy™ Terminator Cycle Sequencing Kit manual (Applied Biosystems, California, U.S.A.). The appropriate *E. coli* strain was grown overnight at 37°C in 10 ml of 2 x YT broth supplemented with the relevant antibiotics. The cells were harvested by centrifugation at 1,500 *g* for 10 min at room temperature and the supernatant discarded. The cell pellet was resuspended in 200 µl of lysis buffer, and 300 µl of freshly prepared alkaline lysis solution was added as before. The sample was mixed by inversion and then incubated on ice for 5 min before adding 300 µl of 3.0 M potassium acetate, pH 4.8. The contents were mixed by inverting the tube, then incubated on ice for 5 min. Cellular debris was removed by centrifugation at 12,000 *g* for 10 min at room temperature and the clear supernatant was transferred to a fresh tube. RNase A was added to a final concentration of 20 µg/ml and the solution was incubated at 37°C for 20 min. The supernatant was extracted twice with 400 µl of chloroform, where the layers were mixed for 30 sec by inversion following each extraction. The tube was centrifuged at 12,000 *g* for 1 min at room temperature

to separate the phases and the aqueous phase was transferred to a new tube. An equal volume of isopropanol was added to precipitate the DNA and the sample was immediately centrifuged at 12,000 g for 10 min at room temperature. The DNA pellet was washed with 500 µl of cold 70% ethanol, dried under vacuum, and then resuspended in 32 µl of deionized water. The plasmid DNA was then selectively precipitated by the addition of 8.0 µl of 4 M NaCl and 40 µl of 13% (w/v) PEG_{8000}. The sample was incubated on ice for 20 min and then centrifuged at 12,000 *g* for 15 min at 4°C. The supernatant was removed and the plasmid DNA pellet washed with 500 µl of cold 70% ethanol. The pellet was dried under vacuum and then resuspended in 50 µl of dH_2O. The DNA preparation was stored at -20°C until required.

***b) Small scale isolation of plasmid DNA from* C. perfringens**

Plasmid DNA from *C. perfringens* strains was prepared as previously described (Lyristis *et al.*, 1994). A single colony of the appropriate *C. perfringens* strain was used to inoculate 6 ml of pre-boiled FTG medium and was grown overnight at 37°C. Twenty ml of pre-boiled BHI broth supplemented with 1.5% glucose was inoculated with 3 ml of the overnight culture and the cells were grown at 37°C until mid-log phase before being harvested by centrifugation at 1,500 *g* for 10 min at room temperature. The supernatant was discarded and the cell pellet was resuspended in 200 µl of TES buffer. A sample (100 µl) of the cell suspension was transferred to a microcentrifuge tube and 200 µl of 25% (w/v) sucrose in TES containing 10 mg/ml of lysozyme was added. The suspension was incubated at 37°C for 30 min after which 200 µl of 2% (w/v) sarkosyl in TES was added. The suspension was incubated for a further 30 min at 37°C , followed by the addition of 200 µl of 2 M NaOH. The sample was mixed by inversion and 200 µl of

neutralization solution from the Magic™ Mini Prep Kit (Promega Corporation, Wisconsin, U.S.A.) was added to the suspension. Plasmid DNA was then isolated according to the procedure provided by the manufacturer.

Ethanol and isopropanol precipitation of DNA

Chromosomal DNA was ethanol precipitated by the addition of 0.1 volumes of 7.5 M ammonium acetate (pH 7.8) and 2.5 volumes of cold 100% ethanol. The mixture was incubated at -70°C for 60 min. DNA was collected by centrifugation at 12,000 *g* for 30 min at 4°C. The DNA pellet was washed with cold 70% ethanol, dried under vacuum and resuspended in the appropriate volume of dH_2O or TE (1 mM EDTA, 10 mM Tris, pH 8.0).

Plasmid DNA was ethanol or isopropanol precipitated by the addition of 0.1 volumes of 3 M sodium acetate (pH 5.8) and 2.5 volumes of cold 100% ethanol or 1.0 volumes of isopropanol, respectively. The mixture was then incubated at -70°C for 60 min for ethanol precipitation or 30 min for isopropanol precipitation. DNA was collected by centrifugation at 12,000 *g* for 30 min for ethanol precipitations or 15 min for isopropanol precipitations, at 4°C. The DNA pellet was washed and dried as before, and resuspended in an appropriate volume of dH_2O or TE.

Phenol-chloroform extraction of DNA

When required, DNA was further purified by phenol-chloroform extraction. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the DNA preparation, vortexed vigorously and the phases separated by centrifugation at

12,000 *g* for 1 min at room temperature. The aqueous phase was transferred to a new tube, to which an equal volume of chloroform:isoamyl alcohol (24:1) was added and the DNA extracted as before. The aqueous layer was transferred to a new tube and the DNA was isolated by ethanol or isopropanol precipitation.

Recombinant DNA techniques

Restriction endonuclease digestion

DNA was digested with various restriction endonucleases under the conditions outlined by the manufacturers (Roche Diagnostics Australia, New South Wales, Australia, or New England Biolabs Inc. (NEB), Massachusetts, U.S.A.). Reactions were terminated either by the addition of 0.2 volumes of stop mix (0.1 M EDTA, 0.05% (w/v) bromophenol blue (Progen Industries Ltd., Queensland, Australia), 50% (w/v) sucrose pH 7.0) or by heat inactivation for 15 min at 65°C, followed by phenol-chloroform extraction and precipitation with ethanol.

Agarose gel electrophoresis

DNA samples were analyzed by electrophoresis using 0.8%-1.0% (w/v) agarose (FMC BioProducts, Maine, U.S.A.) gels in TAE buffer (1.0 mM EDTA, 38.2 mM Tris-HCl, 16.6 mM sodium acetate, pH 7.8). Mini sub-gel electrophoresis was carried out at a constant voltage of 100 V for 30-60 min, while large gels were subjected to 130 V for 2.5-3.0 h. Following electrophoresis, gels were stained with 10 μg/ml ethidium bromide (Progen), destained in dH_2O and the DNA visualized under UV light using a Spectroline Ultraviolet Transilluminator (Medos Company Pty. Ltd., Victoria, Australia). DNA fragment sizes were estimated by comparison

with λ*cI857* DNA (Promega Corporation) digested with *Hin*dIII or PCR markers (Promega Corporation) as molecular size standards using the SEQAID II program (D. Rhoads and D. Roufa, Kansas State University, U.S.A.).

Isolation of DNA fragments from agarose gels

When necessary, DNA fragments were isolated from agarose gels following electrophoresis. The appropriate DNA fragments were excised from the gel using a scalpel blade. DNA was isolated using the QIAquick Gel Extraction Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions.

Dephosphorylation of plasmid DNA with alkaline phosphatase

Removal of 5' phosphate groups to prevent recircularization of digested vector DNA was achieved by treatment of the DNA with alkaline phosphatase. Following digestion of vector DNA with the appropriate restriction endonuclease, 6 µl of 10X dephosphorylation buffer (Roche Diagnostics Australia) (0.5 M Tris-HCl, 1 mM EDTA, pH 8.5) and 20 units of calf intestine alkaline phosphatase (Roche Diagnostics Australia) were added. The volume was then adjusted to 60 µl with dH_2O and the reaction incubated at 37°C for 1 h. The alkaline phosphatase was inactivated by the addition of one-ninth volume of 10X TNE (100 mM Tris-HCl, 1 M NaCl, 10 mM EDTA, pH 8.0) followed by heat inactivation at 65°C for 15 min. To remove the inactivated alkaline phosphatase and restriction endonucleases, the reaction was extracted with phenol:chloroform:isoamyl alcohol and chloroform:isoamyl alcohol and the DNA was precipitated with ethanol as previously described.

DNA ligation

DNA ligations were carried out in reactions containing 3 units of T4 DNA ligase (3 units/µl, Promega Corporation), 0.1 volumes of the 10X ligase buffer (300 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP) supplied by the manufacturer, and vector and insert DNA at a ratio of 1:5. Ligation of the DNA was performed at 16°C overnight. The ligation reaction was terminated by heat inactivation at 65°C for 10 min, and the ligated DNA was precipitated using isopropanol, as previously described, prior to transformation into the appropriate cells.

Synthesis of oligonucleotide primers

The oligonucleotide primers used in this study (Table 2.3) were synthesized using a 392 DNA/RNA Synthesizer (Applied Biosystems). Following synthesis, the oligonucleotide primers were deprotected by incubation at 55°C for 2 h and dried in a Heto Maxi-Dry Plus vacuum concentrator. When required, the primer was resuspended in 100 µl of dH_2O and the concentration determined by measurement of the absorbance at 260 nm. An absorbance reading of 1.0 at 260 nm was equated to a single stranded DNA concentration of 33 µg/ml (Sambrook *et al.*, 1989). Dried and resuspended oligonucleotide primers were stored at -20°C until required.

Nucleotide sequencing

a) Automated sequencing

Plasmid DNA to be sequenced was obtained using the Applied Biosystems method previously described. PCR products to be sequenced were either purified directly from the PCR reaction using the Magic™ PCR Preps Purification System

Table 2.3 : Oligonucleotide primers.

Primer	5'-3' Sequence	Characteristics/Reference/Use
UP	GTTGTAAAACGACGGCCAGT	Universal Primer, SQ
RP	AGCGGATAACAATTTCACACAGGA	Reverse Primer, SQ
#2980	AATAAGTAAACAGGTAACGTCT	Internal *erm*(B), SQ & PCR
#2981	GCTCCTTGGAAGCTGTCAGTAG	Internal *orf3*, SQ & PCR
#3046	GCTGCCAGCTGAATGCTTTCAT	Internal *erm*(B), SQ
#3049	GAGACTTGAGTGTGCAAGAGCA	Internal *erm*(B), SQ
#3105	CTTGGTGAATTAAAGTGACACG	Internal *erm*(B), SQ
#3106	CGGGAGGAAATAATTCTATGAG	3' end *erm*(B)/5' end *orf3*, SQ
#3139	ACTTACCCGCCATACCACAGAT	Internal *erm*(B), SQ & PCR
#3140	ATTTTATACCTCTGTTTGTTAG	Internal *erm*(B), SQ & PCR
#3226	AAGAATTACTGGAGGGAAAAGA	Intergenic space between *orf3* and ΔDR, SQ
#3248	TTACAACGGCATTGTAGGGCTT	Internal DR sequence, SQ
#3715	TACCAAACCATACACCATCCTC	Intergenic space between ΔDR and *orf13*, SQ
#3716	AAGTGATTTGTGATTGTTGATG	Internal ΔDR, SQ
#4191	CGTTGTAAAAATTGGGGAAAAG	Internal DR sequence, SQ
#4192	CAAGTCGGCACGAACACGAACC	Internal DR sequence, SQ & PCR
#4210	TCAATAGACGTTACCTGTTTAC	Internal *erm*(B), SQ & PCR
#4348	GGTTCGTGTTCGTGCCGACTTG	Internal DR sequence, SQ
#4349	CATGAGCGAGTTAATTTTGGCA	Internal DR sequence, SQ & PCR
#4350	TGCCAAAATTAACTCGCTCATG	Internal DR sequence, SQ & PCR
#4451	CTGCTTGTAAAGGGATCATAAC	Internal DR sequence, SQ
#4537	GTCAAGTAAGCAAACATAGTCG	Internal DR sequence, SQ & PCR
#4538	CGACTATGTTTGCTTACTTGAC	Internal DR sequence, SQ & PCR
#6018	AATGGCTGGTTCTACAAATACA	Internal *ilvD*, SQ & PCR
#6019	ACTCTGCCTGACAAAACATCTG	Internal *effR*, SQ
#6114	ACTGACACACTGACCTTGAGAT	Internal *orf13*, SQ
#6115	TGAAATAGAAAATGAAGAAGTA	Internal *ilvD*, SQ
#6182	GCACTTCTTACTGATGGTCGTT	Internal *ilvD*, SQ
#6183	GTCCCTAAATCTACGGTCACTT	Intergenic space between ΔDR and *orf13*, SQ
#6260	GTATGAAAAACACAGCAAAATC	Intergenic space between *ilvD* and *hydR*, SQ & PCR
#6278	GATTTTGCTGTGTTTTTCATAC	Intergenic space between *ilvD* and *hydR*, SQ & PCR
#6306	CATTTTCACTATTTTCGTCTAA	Internal *hydR*, SQ & PCR
#6339	ATGCTCGTTTTTAGTATTGAT	Internal *hydR*, SQ & PCR
#6356	GCGGGGGATCCATGAACAAAAATATAAAAT	5' end *erm*(B), PP
#6357	CGCTTGGTACCTTATTTCCTCCCGTTA	3' end *erm*(B), PP
#6371	GAAAGCAGAAGTAATGGGTGTT	Internal *hydD*, SQ

#6427	AGGGATTGGGACACGCTACATA	Internal *hydD*, SQ & PCR
#6495	CTACTAATGAATGAGCCTTGAT	Intergenic space between *hydD* and Erm leader peptide, SQ
#6552	GCACTATCAACACACTCTT	Erm leader peptide promoter region, SQ
#6554	GAAACTGCCTATTGCGTGAACA	Intergenic space between *hydD* and Erm leader peptide, SQ
#6604	TAAGAGTGTGTTGATAGTGC	Erm leader peptide promoter region, SQ & PCR
#6784	AAGTTTGTATGAGAAGCAGTAT	Intergenic space between *hydD* and Erm leader peptide, SQ
#6785	TTAGGGACACTTACTGATGAAT	Intergenic space between ΔDR and *orf13*, SQ & PCR
#6940	TAGCGTGTCCCAATCCCTCATA	Internal *hydD*, SQ & PCR
#7390	AGTCACAGATAAAAACGGTCAG	Intergenic space between ΔDR and *orf13*, SQ
#7391	ATCAAGGCTCATTCATTAGTAG	Intergenic space between *hydD* and Erm leader peptide, SQ & PCR
#7449	CGTATTTATTTATCTGCGTA	Internal Erm leader peptide, SQ
#7716	GGCTAGCATGACTGGTGG	Upstream of multiple cloning site in pRSETA, SQ
#7774	ATAATCTCAAGGTCAGTGTGTC	Internal *orf13*, SQ
#8752	TATTGTTGTATTGGTAAAGCACT	Intergenic space between *hydD* and Erm leader peptide, SQ
#8753	TCAGCAACCGAAAACAGACTATC	Internal *effR*, SQ
#8885	TGGTTCATTTTGTTCGTCTCC	5'end *effD*,SQ & PCR
#9069	TACTGGCTTTTAGACGCACCTG	Internal *effD*, SQ & PCR
#9153	CTTCTCGGTGCTGATAGTAATA	Internal *effD*, SQ
#9251	TATCGCTGTTGCCTATGGTGC	Internal *effD*, SQ
#9341	ACAAAGTTAGTGATGGTTAT	Intergenic space between *effD* and *orf9*, SQ
#9371	GATAGAAATACTCGTCAACAGA	Internal *orf9*, SQ & PCR
#9387	ATTTTTTATTTTTAGGAGTCAT	Intergenic space between *orf7* and *ispD*, SQ & PCR
#9409	TACTATTTTCACAGGTTTGCTC	Intergenic space between *orf9* and *orf7*, SQ & PCR
#9493	AACCATCAGACTTCCAAAA	Internal *orf7*, SQ & PCR
#9586	CACACCCCTTTCGCTATG	5' end *orf7*, SQ
#9587	GGATGATTACGAAAGTGAC	Internal *orf7*, SQ
#9782	CAAGGGCTGATGATAAACTA	Intergenic space between *orf9* and *orf7*, SQ & PCR
#9896	GCCGTCAGATAGATTCGTCACT	Internal *orf9*, SQ
#10166	TTTCTTCATCAATAGTGGCTTC	Intergenic space between *effD* and *orf9*, SQ
#10327	CATAACGGACATAACAACAGCC	Internal *effD*, SQ & PCR
#10515	GGGGTACCAGATGCTAAAAATTTGTA	5' end *erm*(B), PCR
#11546	ATGACTCCTAAAAATAAAAAAT	Intergenic space between *orf7* and *ispD*, SQ
#11547	CGGCAAGCACATAATCTCCATA	Internal *effD*, SQ
#11617	CCAAACAGGAAAGATAGCCATA	Internal *effD*, SQ

#11662	TGTGGGATGAAGGTTAT	Internal *ispD*, SQ & PCR
#11795	AGGTAATAATGGAGATGGTGAT	Internal *ispD*, SQ
#11864	AGTATCCATTTCCTTGTTC	Internal *ispD*, SQ & PCR
#11865	GAACAAGGAAATGGATACT	Internal *ispD*, SQ & PCR
#12142	GGAGTGGAACAGGAAATAC	Internal *flxD*, SQ
#12143	GTATTTCCTGTTCCACTCC	Internal *flxD*, SQ & PCR
#12262	TTCCCCCAAGAGACATAG	Internal *ispD*, SQ
#16525	TACTTCTTCATTTTCTATTTCA	Internal *ilvD*, SQ
#16526	AACGACCATCAGTAAGAAGTGC	Internal *ilvD*, SQ
#16527	TTAGACGAAAATAGTGAAAATG	Internal *hydR*, SQ
#16528	AACACCCATTACTTCTGCTTTC	Internal *hydD*, SQ
#16533	ATCAATACTAAAAACGAGCAT	Internal *hydR*, SQ
#16534	TTCAAAAACAACATTAAATTAT	Intergenic space following *flxD*, SQ

*SQ: Sequencing; PCR: Polymerase Chain Reaction; PP: Protein purification.

(Promega Corporation) according to the manufacturer's instructions, or were isolated from agarose gels using a QIAquick Gel Extraction Kit (Qiagen). Cycle sequencing reactions were carried out on a GeneAmp PCR System 2400 (Perkin Elmer Corporation, California, U.S.A), using a PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase, FS (Applied Biosystems) according to the manufacturer's instructions. Excess dye terminators were removed by ethanol/sodium acetate precipitation using a protocol supplied by the manufacturer, with the exception that extension products were precipitated at room temperature for 15 min prior to centrifugation. Sequencing samples were resolved and analyzed on a 373 DNA STRETCH Sequencer (Applied Biosystems).

b) Sequence analysis

Nucleotide sequences were compiled using Sequencher™3.1 software (GeneCodes Corporation, Michigan, U.S.A.). Nucleotide and amino acid sequence database searches were performed using the BLAST program (Altschul *et al.*, 1997) at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Amino acid sequences were analyzed using Network Protein Sequence @nalysis (NPS@) (http://npsa-pbil.ibcp.fr/) and the Expert Protein Analysis System (ExPASy) proteomics server of the Swiss Institute of Bioinformatics (http://www.expasy.ch/). Multiple nucleotide and amino acid sequence alignments were achieved using the CLUSTAL W program (Thompson *et al.*, 1994) at NPS@. Prediction of transmembrane regions was conducted using the SOSUI System (Takatsugu *et al.*, 1998) (http://sosui.proteome.bio.tuat.ac.jp/).

Amplification of DNA by Polymerase Chain Reaction (PCR)

a) PCR

PCR amplification was carried out using *Taq* DNA polymerase (5 units/µl, Roche Diagnostics Australia) in a total volume of 100 µl. Each reaction consisted of 10 µl of the 10X reaction buffer supplied by the manufacturer, 0.2 mM dNTP's, 0.4 µM of each oligonucleotide primer, between 0.05 and 1.0 µg of template DNA and sterile dH_2O. Amplification was performed in a GeneAmp PCR System 2400 (Perkin Elmer Corporation). The DNA template was first denatured at 94°C for 3 min, then the temperature held at 70°C , at which time 5 units of *Taq* DNA polymerase was added to the reaction mixture. Amplification was carried out over 30 cycles consisting of a 94°C denaturation step (1 min), 50°C annealing step (2 min) and a 72°C extension step (3 min). The final cycle consisted of 2 min annealing at 50°C and 5 min of extension at 72°C. When required, the specificity of the PCR reaction was either increased or decreased by changing the annealing temperature to 55°C or 47°C, respectively. PCR products were detected by running a sample (5-10 µl) on an agarose gel.

b) Capillary PCR

A capillary PCR method was occasionally used for initial screening of recombinant clones. Crude cell extracts were obtained from selected colonies by resuspending the cells in 50 µl of sterile dH_2O, vortexing the suspension vigorously for several seconds and then lysing the cells by boiling for 10 min. The cellular debris was then removed by centrifugation at 12,000 *g* for 10 min at room temperature. The clear supernatant was removed and stored on ice or at -20°C until required.

A PCR master mix was prepared and aliquoted into 18 µl reaction mixtures that consisted of 2 µl of 10X PCR reaction buffer, 0.2 mM dNTPs, 1 unit of *Taq* DNA polymerase, 0.4 µm of each oligonucleotide primer and the appropriate amount of sterile dH_2O. To each reaction, 2 µl of the cellular extract was added as the DNA template and the reaction mixture was then drawn up in a capillary PCR tube by use of a CP-1 Cycle Prep Auto Gun (Corbett Research, New South Wales, Australia). The tubes were then placed in a FTS-1 Thermal Sequencer (Corbett Research) and the products were amplified. The PCR program consisted of an initial cycle of denaturation at 94°C for 5 min, annealing at 47°C for 2 min and extension at 72°C for 3 min. The program then continued for a further 30 cycles of denaturation at 94°C for 1 min, annealing at 47°C for 1 min and extension at 72°C for 2 min. A final cycle consisted of denaturation at 94°C for 1 min, annealing at 47°C for 1 min and extension at 72°C for 5 min. The amplification of PCR products was detected by examining the entire 20 µl reaction by agarose gel electrophoresis.

c) Purification of PCR products

PCR products were purified for cloning or sequencing either directly from the PCR using Magic™ PCR Preps Purification System (Promega Corporation), or were extracted from agarose gels using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions.

d) Digoxigenin labelling of probes using PCR

PCR fragments to be used as probes were labelled with digoxigenin-11-dUTP (DIG-dUTP) during PCR amplification. PCR amplification was carried out using 5 units of *Taq* DNA polymerase (5 units/µl, Roche Diagnostics Australia) in a total volume of 100 µl. Each reaction consisted of 10 µl of the 10X reaction buffer

supplied by the manufacturer, 0.2 mM dNTPs, 0.4 μM of each oligonucleotide primer, between 0.05 and 1.0 μg of template DNA and sterile dH_2O. In addition, 2 μl of DIG DNA Labelling Mix (1mM each dATP, dCTP, dGTP, 0.65 mM dTTP, 0.35 mM DIG-dUTP)(Roche Diagnostics Australia) was added to each reaction. The DNA was amplified in a GeneAmp PCR System 2400 (Perkin Elmer Corporation) as previously described. Amplification of DNA was confirmed by electrophoresis in agarose gels and the PCR product was then purified directly from the reaction using the Magic™ PCR Preps Purification System (Promega Corporation). The efficiency (ng labelled DNA/μl) of the labelling reaction was determined using the quantitation protocol supplied by the manufacturer (Roche Diagnostics Australia).

DNA hybridization analysis

a) Southern hybridization analysis

Southern hybridization anlaysis was performed by the method of Southern (1975) with several modifications. Restriction endonuclease digested DNA was subjected to electrophoresis alongside DIG-labelled λ *Hin*dIII DNA molecular size markers (Roche Diagnostics Australia), stained with ethidium bromide (10 μg/ml), and the agarose gel photographed. To prepare the DNA for transfer to a nylon membrane, the DNA fragments of high molecular weight were first depurinated by incubating the agarose gel (at a slight angle such that the high molecular weight DNA end of the gel was fully submerged, but the low molecular weight DNA end of the gel was out of the solution) in 0.25 M HCl for 10-15 min at room temperature on a shaker. The gel was rinsed briefly in dH_2O and then incubated in denaturation solution (0.2 M NaOH, 0.6 M NaCl) for 30 min at room temperature on a shaker. The gel was again rinsed briefly in dH_2O and was then

neutralized by incubating twice in neutralization solution (165 mM tri-sodium citrate, 1.5 M NaCl, 0.25 M Tris-HCl (pH 7.5)) for 20 min. Each incubation was carried out at room temperature on a shaker.

Once the gel was prepared, the DNA was transferred to a Hybond-N+ membrane (Amersham Pharmacia Biotech, Buckinghamshire, U.K.) overnight on a support covered with a wick of Whatman 3MM chromatography paper placed within a container that held 500-600 ml of 10 x SSC (165 mM tri-sodium citrate, 1.5 M NaCl). The gel was placed upside down on the wick and the nylon membrane, which was presoaked in 2 x SSC, was placed over the gel. Two pieces of Whatman 3MM paper presoaked in 2 x SSC, a stack of paper towels and a weight (approximately 300 g) were then placed over the membrane. The DNA was allowed to transfer from the gel to the nylon membrane by capillary action overnight. Following transfer, the nylon membrane was removed and air-dried. The DNA was cross-linked to the membrane by exposure to UV light (312 nm) for 3-5 min.

For hybridization, the membrane was pre-hybridized for a minimum of 3 h at 65°C in pre-hybridization solution (83 mM tri-sodium citrate, 0.75 M NaCl, 0.05% (w/v) Blocking Reagent (Roche Diagnostics Australia), 1% (v/v) N-lauroylsarcosine, 0.02% (v/v) SDS). Between 10-50 ng of probe DNA was added to 5 ml of pre-hybridization solution (to create the hybridization solution) and denatured by boiling for 10 min followed by immediate incubation on ice for 5 min. Following pre-hybridization, the pre-hybridization solution was removed from the membrane and the hybridization solution was added. Hybridization was carried out overnight under high stringency conditions (65°C). Following hybridization the

membrane was washed twice for 5 min at room temperature in 2 x SSC, 0.1% SDS, and twice for 15 min at 65°C in 0.2 x SSC, 0.1% SDS.

DNA-DNA hybrids were identified by chemiluminescent detection using CDP-*Star*™ (Roche Diagnostics Australia) in accordance with the manufacturer's specifications. When necessary, following detection, the membranes were stripped and reprobed as follows. The membrane was rinsed for 1 min in sterile H_2O, followed by two 10 min washes at 37°C in a freshly prepared solution of 0.2 M NaOH, 0.1% SDS. The membrane was then rinsed thoroughly for 5 min in 2 x SSC, placed back into pre-hybridization solution and then hybridized with a different probe.

b) DNA dot blots

DNA samples were applied to a Hybond-N+ nylon membrane (Amersham Pharmacia Biotech) using a dot blot apparatus (SRC 96D Minifold I, Schleicher and Schuell, Dassel, Germany) as follows. Two pieces of Whatman 3MM paper pre-soaked in 2 x SSC were placed over the base of the manifold, a pre-soaked nylon membrane (Amersham Pharmacia Biotech) cut to the appropriate size was then added and the apparatus assembled. The wells of the apparatus were washed twice with 2 x SSC, and the DNA samples, which had been adjusted to a volume of 100 µl, were then applied to the membrane. After application of the DNA samples, the apparatus was disassembled, the membrane removed and air dried at room temperature. The DNA was crosslinked to the membrane by exposure to UV light (312 nm) for 5 min. Pre-hybridization and hybridization were carried out as described for Southern hybridization analysis. When necessary, blots were stripped and reprobed using the procedure outlined before.

Genetic methods

Transformation procedures

a) Preparation and transformation of rubidium chloride-competent* E. coli *cells

Rubidium chloride (RbCl)-competent *E. coli* cells were prepared using the method described by Hanahan (1985). The appropriate *E. coli* strain was subcultured on SOB (10 mM NaCl, 2.5 mM KCl, 2% (w/v) Tryptone, 0.5% (w/v) Yeast Extract) (Sambrook *et al.*, 1989) agar and grown at 37°C for 16-20 h. Ten colonies were resuspended in 1 ml of SOB broth, mixed thoroughly by vortexing and subsequently used to inoculate a 2 L flask containing 100 ml of SOB broth, to which 1 ml of a Mg^{2+} solution (1 M $MgCl_2$, 1 M $MgSO_4$) had been added. The cells were grown at 37°C with moderate agitation until the turbidity of the culture at 600 nm was 0.3. The culture was then transferred to sterile 50 ml tubes and chilled on ice for 10-15 min. The tubes were then centrifuged at 12,000 *g* for 15 min at 4°C , the supernatant was discarded and the cell pellet thoroughly drained by inverting the tube. The pellet was resuspended in 0.33 volumes of filter sterilized RF1 (10 mM RbCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2.2H_2O$, 15% glycerol, pH 5.8 with 0.2 M acetic acid), mixed by moderate vortexing and incubated on ice for 1 h. The cells were centrifuged as before and the resulting cell pellet was resuspended in filter sterilized RF2 (8 mM RbCl, 75 mM $CaCl_2.2H_2O$, 10 mM 3-(N-Morpholino) propane-sulfonic acid (MOPS), 15% glycerol, pH 6.8) to 1/12.5 of the original volume. The suspension was then incubated on ice for 15 min, dispensed as 100 µl aliquots into pre-chilled microfuge tubes and snap frozen in a dry ice/ethanol bath. The cells were stored at -70°C until use.

Transformation experiments were conducted essentially as previously described (Hanahan, 1985). RbCl-competent cells were thawed on ice prior to the addition of plasmid DNA. The plasmid/competent cell mixture was incubated on ice for 60 min. The cells were heat shocked at 37°C for 2 min and then immediately chilled on ice for 5 min. To allow the expression of antibiotic resistance genes, 900 µl of 2 x YT broth was added and the cell mixture was incubated, with shaking, at 37°C for 1 h. Several dilutions of the transformation mixture were prepared and 100 µl aliquots were spread onto 2 x YT agar supplemented with the appropriate antibiotic. To facilitate 'blue-white' selection of recombinant clones, 50 µl of 2% (w/v) 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (Progen Industries Ltd.) and, when required, 10 µl of 100 mM isopropyl-β-D-thiogalactoside (IPTG) (Progen Industries Ltd.), were spread onto the surface of agar plates prior to inoculation with the transformation mixture. The cultures were then incubated for 24-48 h at 37°C.

b) Preparation and transformation of electrocompetent* E. coli *cells

Electrocompetent *E. coli* cells were prepared using the procedure outlined by Smith *et al.*, (1990) with several modifications. A single *E. coli* colony was used to inoculate 10 ml of SOB broth supplemented with appropriate antibiotics. The cells were grown overnight at 37°C and 1 ml of the resulting culture was used to inoculate 500 ml of SOB broth. The culture was grown at 37°C with moderate agitation until the turbidity at 550 nm was approximately 0.8. The cells were harvested by centrifugation at 4,000 *g* for 10 min at 4°C. The supernatant was discarded and the pellet resuspended in 250 ml of ice-cold 10% (v/v) glycerol. The cells were harvested as before, the supernatant was discarded and the cell pellet was again resuspended in 250 ml of ice-cold 10% (v/v) glycerol. The cells were collected by

centrifugation as before and the cell pellet was resuspended in a final volume of 2 ml of ice-cold 10% (v/v) glycerol. The cells were snap frozen as 100 µl aliquots in a dry/ice ethanol bath and were stored at -70°C until use.

Prior to use the cells were thawed on ice. Ethanol or isopropanol precipitated DNA was then added to a 20 µl aliquot of electrocompetent cells and mixed gently. The mixture was transferred to a pre-chilled 0.1 cm electroporation cuvette (Bio-Rad Laboratories, California, U.S.A.), which was subsequently placed in a Bio-Rad Gene Pulser™ and subjected to electroporation at 1.8 kV/cm, a resistance of 200 ohms and a capacitance of 25 µFD. The cells were then subcultured into 2 ml of SOC broth (SOB broth supplemented with 20 mM glucose) and incubated at 37°C with shaking for 1 h. The transformation mixture was then plated, in 100 µl aliquots, onto 2 x YT agar plates supplemented with the appropriate antibiotic. When required, 50 µl of 2% X-Gal and/or 10 µl of 100 mM IPTG was added to the agar prior to inoculation with the transformation mixture, to facilitate 'blue-white' selection of recombinant clones as before. The cultures were incubated for 24-48 h at 37°C.

c) Preparation and transformation of electrocompetent C. perfringens *cells*

Electrocompetent *C. perfringens* cells were prepared and transformed as before (Scott and Rood, 1989). A single colony of the chosen *C. perfringens* strain was used to inoculate 6 ml of pre-boiled FTG broth and was grown overnight at 37°C. The overnight culture was then used to inoculate a pre-boiled 100 ml TPG broth supplemented with 1.5% (v/v) glucose. The culture was grown until the turbidity at 600 nm was between 0.15 and 0.2. The cells were then harvested by centrifugation at 16,300 *g* for 15 min at room temperature and the supernatant discarded. The cells were then washed with 10 ml of SEB buffer (272 mM sucrose,

7 mM Na_2HPO_4 (pH 7.4), 1 mM $MgCl_2$) and centrifuged as before. The pellet was resuspended in 10 ml of SEB buffer containing 10 μg/ml lysostaphin (Sigma Chemical Co.) and incubated at 37°C for 1 h. The cells were then washed twice in 10 ml of SEB buffer and pelleted by centrifugation at 12,000 *g* for 10 min at room temperature after each wash. Following the final wash the cell pellet was resuspended in 10 ml of SEB buffer.

The resuspended cells were then used immediately in transformation experiments. An appropriate amount of plasmid DNA (5-10 μg) was mixed with 400 μl of the electrocompetent *C. perfringens* cells. The mixture was transferred to pre-chilled 0.2 cm electroporation cuvettes and incubated on ice for 10 min. Electroporation of the cells was carried out at 2.5 kV/cm, a resistance of 200 ohms, and a capacitance of 25 μFD in a Bio-Rad Gene Pulser™. Following electroporation, the cells were incubated on ice for 10 min, before being used to inoculate 20 ml of pre-boiled BHI broth supplemented with 1.5% glucose. The cultures were incubated overnight at 37°C to allow the cells to recover and to express antibiotic resistance genes. The following day, the cells were harvested by centrifugation at 1,500 *g* for 10 min at room temperature and resuspended in 3 ml of BHI diluent (one-tenth strength BHI broth). Dilutions of the transformation mixture were made and 100 μl aliquots of the cell suspensions were spread onto nutrient agar supplemented with the appropriate antibiotics. The cultures were incubated overnight at 37°C under anaerobic conditions.

Conjugative transfer experiments

a)* C. difficile *filter matings

Transfer of DNA from *C. difficile* strains to *C. difficile* recipients was performed using a filter mating protocol as described previously (Mullany *et al.*, 1990) with a few modifications. Where possible, all manipulations were conducted in an anaerobic chamber. The donor and recipient strains were grown overnight at 37°C on BHIS agar supplemented with appropriate antibiotics. Five colonies from the overnight culture were used to inoculate a pre-boiled 20 ml BHIS broth and were grown at 37°C until the turbidity at 650 nm was approximately 0.45 (mid-exponential phase). The cells were then harvested by centrifugation at 1,500 *g* for 10 min at room temperature. The supernatant was discarded and the resulting cell pellet was resuspended in 1 ml of BHIS broth. Aliquots (100 µl) of the donor and recipient suspensions were mixed together on sterile 0.45 µm nitrocellulose filters (Whatman International Ltd., Kent, U.K.), which had been placed on the surface of thick (approximately 40 ml of agar per plate) BHIS agar plates. As controls, 100 µl of the donor and recipient suspensions were added separately to nitrocellulose filters on thick BHIS agar plates. The plates containing the filters were then incubated at 37°C for 24 h. Following incubation, each filter containing bacterial growth was removed to separate sterile bottles containing 1 ml of BHIS broth. The growth from the filter was resuspended by moderate vortexing. Transconjugants were selected by plating 100 µl aliquots of the resuspended cultures onto BHIS agar supplemented with appropriate antibiotics and then incubating the plates for 48 h. As negative controls 100 µl aliquots derived from filters containing the donor or recipient cells only were also plated on to the selective media.

b) **E. coli *filter matings***

Filter matings were carried to out mobilize shuttle vectors by conjugation from *E. coli* strain S17-1 to an appropriate recipient strain. These matings were performed essentially as previously described (Palombo *et al.*, 1989). The donor and recipient strains were cultured separately on suitable selective media. After overnight incubation, individual colonies were used to inoculate 10 ml of 2 x YT broth supplemented with appropriate antibiotics and were grown overnight at 37°C. A 500 µl aliquot from each of the overnight cultures was used to separately inoculate 10 ml 2 x YT broths supplemented with appropriate antibiotics. The cultures were incubated at 37°C with shaking until mid-exponential phase (approximately 3-4 h). Following incubation, 500 µl of the donor and recipient cultures were filtered through the same 0.45 µm nitrocellulose filter (Whatman International Ltd.), which was then incubated, bacteria side up, on 2 x YT agar at 37°C for 2 h. The growth was then resuspended in 2 ml of 2 x YT broth by moderate vortexing and appropriate dilutions were plated on to 2 x YT agar supplemented with antibiotics.

Induction experiments

Induction experiments were performed based on the method of Rood (1983). Each *C. difficile* strain was grown on BHIS agar containing erythromycin (50 µg/ml) and a single colony then used to inoculate a 20 ml pre-boiled BHIS broth. The culture was grown overnight at 37°C in an anaerobic chamber. To provide starter cultures, 5 ml aliquots of the overnight cultures were used to inoculate two 90 ml BHIS broths, one containing erythromycin (50 µg/ml) and the other without antibiotic selection. The cultures were grown in the anaerobic chamber at 37°C until they had grown to the mid-exponential phase (approximately 3-4 h). The turbidity of

each culture at 600 nm was determined and the starter cultures diluted so that the same mass of cells was used to inoculate subsequent cultures (inoculum volume = 2.5 / turbidity$_{600nm}$). The diluted starter cultures were inoculated separately into 90 ml of BHIS broth and 90 ml of BHIS broth containing erythromycin (50 µg/ml). Growth was monitored by measuring turbidity at 600 nm for 4 to 10 h as required.

Determination of Minimum Inhibitory Concentrations (MICs)

MIC assays were performed in both *E. coli* and *C. perfringens* based on the method of Kennan *et al.*, (1997). *E. coli* strains were inoculated into 10 ml 2 x YT broths supplemented with chloramphenicol (30 µg/ml) and grown overnight at 37°C. The cultures were then diluted 1 in 25 into fresh 2 x YT broth containing chloramphenicol (30 µg/ml) and grown at 37°C until the turbidity at 550 nm was 0.7-0.8. Cultures were then diluted 1 in 100 in fresh 2 x YT broth. Duplicate aliquots of 10 µl were placed on 2 x YT agar containing chloramphenicol (30 µg/ml) and erythromycin at concentrations ranging from 0 to 1280 µg/ml. The cultures were incubated for 18-20 h at 37°C, after which the MIC was determined as the lowest concentration of erythromycin that completely inhibited growth. The assay was repeated three times for each strain. For *C. perfringens* strains, an identical procedure was used with the exception that BHI media was substituted for 2 x YT media, the cultures were grown anaerobically at 37°C, and the turbidity was determined at 600 nm.

Random mutagenesis

a) Random mutagenesis using the mutator strain XL1-Red

Mutagenesis of pJIR418 was achieved by passage of the plasmid through the *E. coli* mutator strain XL-1 Red (Stratagene, California, U.S.A.) according to the manufacturer's instructions, with a few modifications. The Epicurian Coli® XL1-Red competent cells were thawed on ice, 3.4 µl of the β-mercaptoethanol provided with the kit was added to 200 µl of the thawed cells and mixed gently by inversion. The cells were aliquoted in 40 µl volumes in pre-chilled 15 ml Falcon polypropylene tubes and the contents of the tubes were swirled gently every two minutes while incubating on ice for 10 min. pJIR418 DNA (26 ng), which had been isolated using the Applied Biosystems method, was added to the cells, the contents were swirled gently, and the tube incubated on ice for 30 min. The cells were heat shocked at 42°C for 50 sec and immediately chilled on ice for 2 min. Pre-heated (42°C) SOC broth (0.9 ml) was added and the transformation mixture was incubated at 37°C for 1 h with moderate agitation. The transformation mixture was concentrated by centrifugation at 12,000 *g* for 10 min at room temperature, the supernatant discarded and the resulting cell pellet resuspended in 100 µl of fresh SOC broth. The entire transformation mixture was plated onto a single LB (1% (w/v) NaCl, 1% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 2% (w/v) agar, pH 7.0) agar plate, which was supplemented with chloramphenicol (30 µg/ml), and incubated for 24-30 h at 37°C.

Following incubation, approximately 200 of the resulting colonies were used to inoculate 10 ml of LB broth supplemented with chloramphenicol (30 µg/ml) and were grown overnight at 37°C. Plasmid DNA was extracted from the overnight culture using the Applied Biosystems method and 2 µl of the resulting plasmid DNA was used to transform *E. coli* DH5α cells to chloramphenicol resistance.

Appropriate dilutions of the transformation mixture were made and 100 µl aliquots were plated on 2 x YT agar supplemented with chloramphenicol (30 µg/ml), and incubated at 37°C overnight.

The resultant colonies were patched on to 2 x YT agar supplemented with either chloramphenicol (30 µg/ml) or erythromycin (150 µg/ml) to screen for erythromycin-sensitive mutants, and the cultures incubated overnight at 37°C. Plasmid DNA was extracted and analyzed from those *E. coli* isolates that displayed a chloramphenicol-resistant, erythromycin-sensitive phenotype.

b) Random mutagenesis using hydroxylamine

Hydroxylamine mutagenesis of pJIR418 was carried out essentially as previously described (Humphreys *et al.*, 1976). Plasmid DNA was extracted from 500 ml *E. coli* cultures using a modified large-scale alkaline lysis method (Sambrook *et al.*, 1989) and was further purified by equilibrium centrifugation on caesium chloride-ethidium bromide density gradients. Samples of plasmid DNA (3 µg) were added to 5 volumes of 0.1 M sodium phosphate buffer (pH 6.0) containing 1 mM EDTA and 4 volumes of hydroxylamine hydrochloride (1 M adjusted to pH 6.0 with NaOH). The mixtures were incubated at 70°C for periods of 120 or 180 mins, after which the reaction was terminated by ethanol precipitation. The mutated plasmid DNA was further purified by phenol/chloroform extraction. The DNA was used to transform *E. coli* DH5α cells to chloramphenicol resistance (30 µg/ml). Chloramphenicol resistant colonies were then cross-patched on to 2 xYT agar containing either chloramphenicol (30 µg/ml) or erythromycin (150 µg/ml) to screen for erythromycin-sensitive mutants as before.

Analysis and detection of proteins

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

To visualize purified proteins and whole cell lysates, protein samples were separated by electrophoresis in 0.75 mm thick, 12% denaturing polyacrylamide gels (Laemmli, 1970). Mini gels were prepared in a mini Protean II casting stand (Bio-Rad Laboratories), which was assembled as described by the manufacturer. The 12% resolving gel solution consisted of 4 ml of resolving gel buffer (1.5 M Tris-HCl, 0.4% (w/v) SDS, pH 8.8), 4.8 ml of 40% acrylamide solution (Amresco, Ohio, U.S.A.) (acrylamide: N,N'-methylene bis-acrylamide, 37:5:1) and 7.2 ml of dH_2O. To catalyze polymerization, 60 µl of 10% ammonium peroxodisulfate (AP) (MERCK KgaA, Darmstadt, Germany) and 17.5 µl of *N, N, N', N'*-tetramethylethylenediamine (TEMED) (Amresco) were added to the acrylamide solution, which was subsequently poured between the glass plates and overlaid with H_2O.

Once the resolving gel had set, a 4% stacking gel was prepared from 1.25 ml of stacking gel buffer (0.5 M Tris-HCl, 0.4% (w/v) SDS, pH 6.8), 500 µl of 40% acrylamide solution and 3.25 ml of dH_2O. To initiate polymerization, 30 µl of 10% AP and 5 µl of TEMED were added to the stacking gel solution. The mixture was then poured over the resolving gel and a 10 well Teflon comb was inserted between the glass plates into the stacking gel solution.

While the stacking gel was polymerizing, the protein samples were prepared by the addition of an equal volume of 2X sample buffer (0.013% bromophenol blue, 10% β-mercaptoethanol, 20% glycerol, 2.5% SDS, 4 ml stacking gel buffer). Cell

lysates were denatured by boiling for 5 min, followed by centrifugation at 12,000 g for 2 min to pellet the cellular debris. Samples from protein purification fractions were incubated at 37°C for 10 min immediately before loading the gel, as the acid labile bonds in samples containing imidazole are partially hydrolyzed when boiled (QIAexpressionist Protein Purification Manual, Qiagen).

Protein samples were applied to the gel, which was then subjected to electrophoresis in Tris-Glycine protein gel electrophoresis buffer (12.5 mM Tris-HCl, 100 mM Glycine, 0.05% SDS, pH 8.3) in a mini Protean II Electrophoresis Cell (Bio-Rad Laboratories) at 200V until the bromophenol blue dye front had reached the bottom of the gel. To estimate protein sizes, 5 µl of low molecular size standards from the LMW Electrophoresis Calibration Kit (Amersham Pharmacia Biotech), or 15 µl of SeeBlue™ pre-stained standards (Invitrogen) were run alongside the protein samples for coomassie stained gels, and for gels used in Western blot analysis, respectively.

To visualize the protein bands, the gels were stained with coomassie brilliant blue (0.2% coomassie brilliant blue (BDH Chemicals Australia Pty. Ltd., Victoria, Australia), 25% (v/v) isopropanol, 7% (v/v) glacial acetic acid) for 30 min at room temperature, followed by destaining in a 33% (v/v) methanol, 7% (v/v) glacial acetic acid solution. Alternatively, the proteins were transferred to a nitrocellulose membrane (Schleicher and Schuell) for Western blot analysis.

Western blot analysis

Proteins that had been separated by SDS-PAGE were transferred to nitrocellulose membranes using a mini Trans-Blot® Electrophoresis Transfer Cell (Bio-Rad Laboratories). Prior to transfer, the stacking gel was removed from the protein gel and discarded. Transfer was performed in transfer buffer (12.5 mM Tris-HCl, 100 mM glycine, 15% methanol, pH 8.3) at 100V for 1 h. Following transfer, the membrane was incubated overnight in blocking solution (5% instant skim milk (Bonlac [illegible]ria, Australia) in TBS-Tween [5 mM Tris-HCl, 15 mM NaCl, pH 7 [illegible] at 4°C.

[illegible]ve all traces of blocking solution, TBS-Tween was used to wash the membrane once for 15 min and then twice for 5 min at room temperature. The washed blot was then incubated for 1 h at room temperature with the primary antibody, which was either mouse monoclonal T7 tag antibody (Novagen, Massachusetts, U.S.A.) that had been diluted 1/3000 with TBS-Tween, or polyclonal Erm(B) antiserum that had been diluted 1/5000 with TBS-Tween. After washing, the membrane was incubated in a 1/2000 dilution of affinity isolated, horse radish peroxidase (HRP) conjugated, anti-mouse or anti-rabbit immunoglobulin (Silenus, Victoria, Australia) in TBS-Tween for 1 h at room temperature. To remove excess secondary antibody, the membrane was washed with TBS-Tween once for 15 min and four times for 5 min at room temperature.

The bands were developed using the Renaissance® Western Blot Chemiluminescence Reagent (NEN™ Life Science Products, Massachusetts, U.S.A.). Equal volumes of the detection solutions, sufficient to cover the blot, were mixed and incubated with the membrane for 1 min at room temperature. To

visualize the immunoreactive band, the blot was then wrapped in plastic film and the membrane exposed to Fuji X-Ray film for an appropriate amount of time (approximately 30 sec to 5 min).

Protein purification

Solubility analysis of the Erm(B) protein

The following protocol was performed to check the solubility of the recombinant Erm(B) protein. A single colony of strain JIR5735 was used to inoculate 10 ml of 2 x YT broth supplemented with ampicillin (100 µg/ml) and erythromycin (150 µg/ml), and incubated overnight at 37°C. This culture was diluted 1:10 in fresh medium and grown at 37°C for 1 h. To induce production of the Erm(B) protein, IPTG was added to a final concentration of 2 mM. After incubation at 37°C for 3 h, the cells were harvested by centrifugation at 12,000 *g* for 10 min at 4°C. The cells were resuspended and washed with 10 ml of phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4) and centrifuged as before. The cells were finally resuspended in 1 ml of PBS.

The cells were lysed by passing the cell suspension through a French Press. The cell lysate was then centrifuged at 12,000 *g* for 15 min at 4°C to separate soluble and insoluble proteins. The supernatant was transferred to a new tube and the volume adjusted to 1 ml with PBS, before being stored on ice. The pellet was washed with 1 ml of PBS, centrifuged as before, and resuspended in 1 ml of PBS.

Equal volumes of each protein sample were combined with 2X sample buffer, boiled for 5 min and centrifuged at 12,000 *g* for 2 min at room temperature. Aliquots

of each sample were separated by SDS-PAGE and the immunoreactive proteins were anlayzed by Western blotting as previously described.

Induction of Erm(B) production and preparation of whole cell lysates

E. coli BL21(DE3) cells (Novagen) harbouring pJIR1626 were cultured overnight at 37°C in 100 ml of 2 x YT broth supplemented with ampicillin (100 μg/ml) and erythromycin (150 μg/ml). Following incubation, the culture was used to inoculate 900 ml of the same medium and grown at 37°C for 1 h. The production of His-tagged protein was induced by the addition of 2 mM IPTG (final concentration) and the culture was grown at 37°C for a further 3 h. The cells were harvested by centrifugation at 16,300 *g* for 10 min at 4°C and the supernatant discarded. The cell pellet was resuspended and washed in 10 ml of PBS and centrifuged as before. The supernatant was discarded and the cell pellet stored overnight at -70°C. The cell pellet was thawed and resuspended in 10 ml of lysis buffer (20 mM Tris-HCl, 0.3 M NaCl, 10% glycerol, 5 mM imidazole, pH 7.9) and lysed by passage twice through a French Press. Cellular debris was removed by centrifugation at 4°C at 12,000 *g* for 10 min.

Purification of His-tagged Erm(B)

The cell lysate was added to 1 ml of Talon™ (ClonTech, California, U.S.A.) affinity resin, which had been previously equilibrated with lysis buffer. The cell lysate and resin were mixed and sealed in a Poly-prep chromatography column (Bio-Rad Laboratories). Proteins were allowed to bind under rotation for 1 h at 4°C, the column allowed to settle, and the unbound proteins eluted from the column. The column was then washed three times with 5 ml of lysis buffer to remove all of the

unbound protein. His-tagged proteins were sequentially eluted with 5 ml of elution buffer (20 mM Tris-HCl, 0.3 M NaCl, 10% glycerol, pH 7.9) supplemented with 20 mM, 60 mM, 100 mM, or 200 mM imidazole and 1 ml fractions were collected.

Samples of each fraction were mixed with equal volumes of 2X sample buffer and subjected to SDS-PAGE. The 12% SDS-PAGE gels were stained with coomassie brilliant blue as described previously. Fractions containing highly purified His-tagged protein were then pooled and dialyzed overnight in dialysis buffer (100 mM NaCl, 20 mM Tris pH 7.5, 1 mM EDTA, 10% glycerol) at 4°C to remove the imidazole. Dialyzed protein samples were concentrated by use of a Biomax-10 Ultrafree®-15 centrifugal filter device (Millipore Corporation, Massachusetts, U.S.A.) as per the manufacturer's instructions and stored at –20°C in 1 ml aliquots until use.

Determination of protein concentration

Protein concentrations were determined using the microwell plate protocol from the BCA Protein Assay Kit (Pierce, Illinois, U.S.A.). The bovine serum albumin (BSA) standards (Pierce) were prepared for the enhanced protocol as per the manufacturer's instructions. A 25 µl aliquot from each standard and also from the appropriately diluted protein sample, was transferred to the wells of a microtitre plate. The working reagent was prepared by mixing BCA reagents A and B (Pierce) at a 50:1 ratio. The working reagent (200 µl) was then dispensed into each of the wells containing the protein samples, and mixed by shaking for 30 s. The covered microtitre plate was incubated at 37°C for 30 min, cooled to room temperature, and the absorbance of each well at 570 nm determined in a microtitre plate reader. The

concentrations of the unknown protein samples were determined by interpolation from a standard curve prepared from the analysis of the BSA standards.

Production of Erm(B) antiserum

Antiserum to the purified His-Erm(B) protein was raised in two female New Zealand White rabbits. Pre-bleed samples were taken from each rabbit *via* the central ear artery (approximately 10 ml). Rabbits were initially injected intramuscularly with approximately 50 μg of His-Erm(B) protein emulsified in Freund's complete adjuvant. Four weeks later the rabbits were boosted by intramuscular injection of approximately 50 μg of His-Erm(B) protein emulsified in Freund's incomplete adjuvant. Following a two week interval, test bleed samples were taken from each rabbit *via* the central ear artery. The sera from these samples and from the pre-bleed samples were diluted 1 in 500 and to determine if antibodies were being produced were used in Western blot analysis, as previously described; using purified His-Erm(B) protein. A subsequent boost of approximately 50 μg of His-Erm(B) protein emulsified in Freund's incomplete adjuvant was administered and, following a two week interval, approximately 100 ml of blood was collected from each rabbit *via* cardiac puncture.

Adsorption of the Erm(B) antiserum

To remove non-specific antibodies the Erm(B) antiserum was adsorbed against a whole cell lysate prepared from the base strain, BL21(DE3) carrying pRSETA (designated JIR5268). A single colony of JIR5268 was cultured overnight at 37°C in 10 ml of 2 x YT broth supplemented with ampicillin (100 μg/ml). Following incubation, 5 ml of the culture was used to inoculate 45 ml of the same

medium and was grown at 37°C for 1 h. After the addition of IPTG (2 mM final concentration) the culture was grown at 37°C for a further 3 h. The cells were harvested by centrifugation at 16,300 *g* for 10 min at room temperature and the supernatant discarded. The cell pellet was resuspended and washed in 10 ml of PBS and centrifuged at 12,000 *g* for 10 min at room temperature. The supernatant was discarded and the cell pellet was resuspended in 10 ml of PBS. The cells were lysed by passage twice through a French Press. Cellular debris was removed by centrifugation at 4°C at 12,000 *g* for 10 min. The supernatant was transferred to a Blue Max™ 50 ml Polypropylene Conical Tube (Becton Dickinson), and two pieces of nitrocellulose (Schleicher and Schuell), cut to 10 cm X 10 cm, were added. Proteins were allowed to bind to the nitrocellulose under rotation for 20 min at 4°C. The antiserum was diluted 1 in 5 in a sterile petri dish using TBS-Tween. One piece of nitrocellulose was added to the diluted antiserum and incubated at 4°C with shaking to allow antibodies to bind. After 1 h, the nitrocellulose was replaced with the second filter and the solution incubated for a further 1 h. The adsorbed antiserum was removed to sterile 5 ml polypropylene containers (Disposable Products Pty. Ltd., South Australia, Australia) and stored at -20°C until use.

CHAPTER THREE

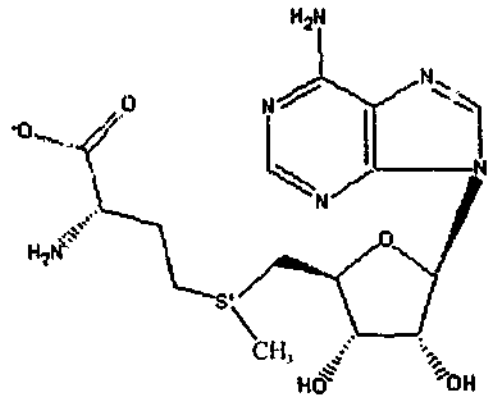
IDENTIFICATION OF FUNCTIONALLY AND STRUCTURALLY SIGNIFICANT RESIDUES OF THE Erm(B) METHYLTRANSFERASE FROM *C. perfringens*

Introduction

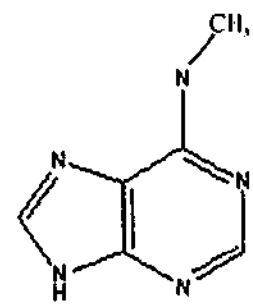
Methyltransferases are enzymes that can methylate DNA or RNA (mRNA, rRNA and tRNA) using *S*-adenosyl-L-methionine (SAM) as the universal methyl donor and releasing *S*-adenosyl-L-homocysteine (SAH) as a reaction product (Bussiere *et al.*, 1998) (Figure 3.1). A structure guided analysis of over 40 DNA methyltransferases proteins (Malone *et al.*, 1995) revealed the presence of nine conserved sequence motifs that are important in target sequence specificity, catalysis and SAM binding of these enzymes (Figure 3.2). Motif I is highly conserved and forms a secondary structure known as the G-loop, which binds the methionine moiety of SAM. Motifs II and III are less conserved, with Motif II containing a negatively charged amino acid that interacts with the ribose hydroxyls of SAM, and a bulky hydrophobic side chain that makes van der Waals contacts with the SAM adenine, and Motif III containing a conserved residue that interacts directly with the exocyclic amino group of the SAM adenine (Bussiere *et al.*, 1998). Motif IV contains a string of highly conserved amino acid residues and creates a structure known as the P-loop, which along with Motifs VI and VIII forms the active or catalytic site of the enzyme. Motif V contains a conserved consensus sequence that is involved in van der Waals contacts to the SAM adenine and also interacts with

Figure 3.1 : Methylation of target adenine residues.

Methyltransferase proteins catalyze the transfer of a methyl group (shown in red) from the substrate *S*-adenosylmethionine (SAM) to an adenine residue in either RNA or DNA targets, yielding a methylated adenine residue and *S*-adenosylhomocysteine (SAH) as the reaction products.



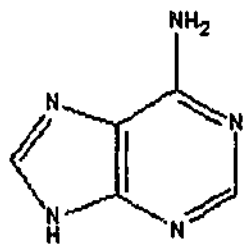
S-adenosylmethionine
(SAM)



Methylated Adenine
Residue

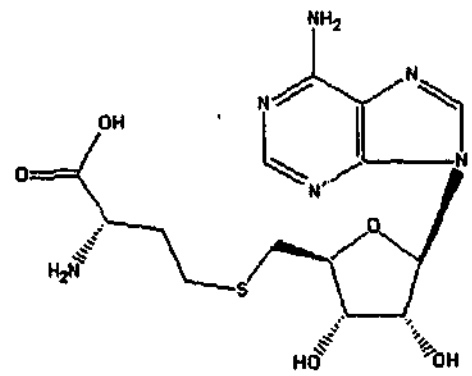
Methyltransferase

\+



Target Adenine
Residue

\+



S-adenosylhomocysteine
(SAH)

Figure 3.2 : Sequence alignment of the DNA methyltransferases M-TaqI and M-HhaI, and the rRNA methyltransferases Erm(B) and Erm(C). The amino acid sequences of M-TaqI(JN0257) and M-HhaI (AAA24989) are aligned with the Erm(B) methyltransferases from *C. perfringens* (S16033) and *S. pneumoniae* (Yu *et al.*, 1997) and the Erm(C) methyltransferase from *B. subtilis* (P13956). The position and extent of each of the conserved motifs, I-VIII and X, are indicated by dark blue arrows below the sequence alignment. The superscript numbers to the left of the start of each M-TaqI or M-HhaI sequence indicate the starting residue number. Regions of identity between the DNA and rRNA methyltransferases are indicated in red. Amino acids in the *C. perfringens* Erm(B) sequence that differ from the *S. pneumoniae* Erm(B) amino acid sequence are indicated in light blue. Amino acid mutations that were obtained in this study are indicated by a green asterisk above the mutated residue and the single letter code of the residue to which the amino acid was mutated. Modified from Bussiere *et al.* (1998).

```
                                                                                  E    Y
                                                                                  *    *
C. perfringens Erm(B)   MN KNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLA   48
S. pneumoniae Erm(B)    MN KNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLA   48
Erm(C)                  MNEKNIKHSQNFITSKHNIDKIMTNIRLNEHDNIFEIGSGKGHFTLELV   49
M-TaqI                   19 GRVETPPEVVDFMVSLAE     39 GGRVLEPACAHGPFLRAF
M-HhaI                  303 GNSVVINVLQYIAYNIG      9 LTGLRFIDLFAGLGGFRLALE
                            <---------------->              <------------->
                                                                   I
                                 K   X P
                                 *     *
C. perfringens Erm(B)   KISKQVTSIELDSHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKI   97
S. pneumoniae Erm(B)    KISKQVTSIELDSHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKI   97
Erm(C)                  QRCNFVTAIEIDHKLCKTTENKLVDHDNFQVLNKDILQFKFPKNQSYKI   98
M-TaqI                   65 YRFVGVEIDPKALD         82 WAEGILADFLLQEPGE 100DLI
M-HhaI                  32 GAECVYSNEWDKYAQEVYEMNF     57 PEGDITQVNE   73 DIL
                                 <--->                  <------>        <-
                                  II                      III

C. perfringens Erm(B)   VGSIPYHLSTQIIKKVVFESHASDIYLIVEEGFYKRTLDIHRTLGLLLH  146
S. pneumoniae Erm(B)    VGNIPYHLSTQIIKKVVFESRASDIYLIVEEGFYKRTLDIHRTLGLLLH  146
Erm(C)                  FGNIPYNISTDIIRKIVFDSIADEIYLIVEYGFAKRLLNTKRSLALFLM  147
M-TaqI                  LGNPPY143YGAFLEKAVR   158GVLVFVVPATW
M-HhaI                  CAGFPCQ 100LFFDIARIVREKKPKVVFMEN         175ALLREFLA
                        ------> <-------->   <--->
                        P  IV       V     S    I VI
                        *                 *    *
C. perfringens Erm(B)   TQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTTDVPDKYWKLYTYFVS  195
S. pneumoniae Erm(B)    TQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTTDVPDKYWKLYTYFVS  195
Erm(C)                  AEVDISILSMVPREYFHPKPKVNSSLIRLNRKKSRISHKDKQKYNYFVM  196
M-TaqI                  REGKTSV        196FPQK KVSAVVIRFQKS
M-HhaI                  144DYSFHA             165RTYMICFR
                          <--->          <------>
                           VII             VIII

C. perfringens Erm(B)   KWVNREYRQLFTKNQFHQAMKHAKVNNLSTVTYEQVLSIFNSYLLFNGRK 245
S. pneumoniae Erm(B)    KWVNREYRQLFTKNQFHQAMKHAKVNNLSTITYEQVLSIFNSYLLFNGRK 245
Erm(C)                  KWVNKEYKKIFTKNQFNNSLKHAGIDDLNNISFEQFLSLFNSYKLFNK   244
```

Motif VII to form the SAM binding site. Motif VI consists of a cluster of three hydrophobic residues, which have been suggested to be involved in placing the target adenine ring on the side opposite Motif IV. Motif VII is not strongly conserved but is believed to be involved in the folding of the catalytic region (Cheng, 1995). Motif VIII is highly conserved and contains a phenylalanine residue, which is proposed to interact with the target adenine residue. This region forms a loop that hangs over the active site and is referred to as the "adenine-binding loop" (Bussiere *et al.*, 1998). Motif X is not well defined, however, in all methyltransferases the loop formed by Motif X, along with the G-loop of Motif I and the P-loop of Motif IV, forms the sides of the binding pocket for the methionine moiety of SAM.

Motifs I, II, III and X are primarily responsible for binding SAM (Cheng *et al.*, 1993; Klimasauskas *et al.*, 1994; Labahn *et al.*, 1994; Schluckebier *et al.*, 1995) and have been collectively termed the SAM-binding region (Malone *et al.*, 1995). Motifs IV, V, VI, VII and VIII are primarily responsible for catalysis (Schluckebier *et al.*, 1995) as they form the active site and have been collectively termed the catalytic region (Schluckebier *et al.*, 1995). Previous work has also shown that the methyltransferases belong to three groups that are distinguished by differences in the order of the conserved motifs in the relative linear order of the SAM-binding region, the catalytic region and the target recognition region (Malone *et al.*, 1995). Group α is arranged in the order (amino to carboxy terminal) SAM-binding region, target recognition region, and then catalytic region. In Group β the catalytic region is followed by the target recognition region and then the SAM-binding region. Group γ is arranged in the order SAM-binding region, catalytic region and then target recognition region (Malone *et al.*, 1995).

In *C. perfringens* MLS resistance is mediated by the Erm(B) rRNA methyltransferase (Berryman and Rood, 1989), which, based on its amino acid sequence and motif order, belongs to the γ group of methyltransferases. This chapter presents the results of mutational analysis of the *C. perfringens erm*(B) gene and identifies several amino acid residues located in conserved motifs that are important for either the structure or function of the Erm(B) protein.

Results

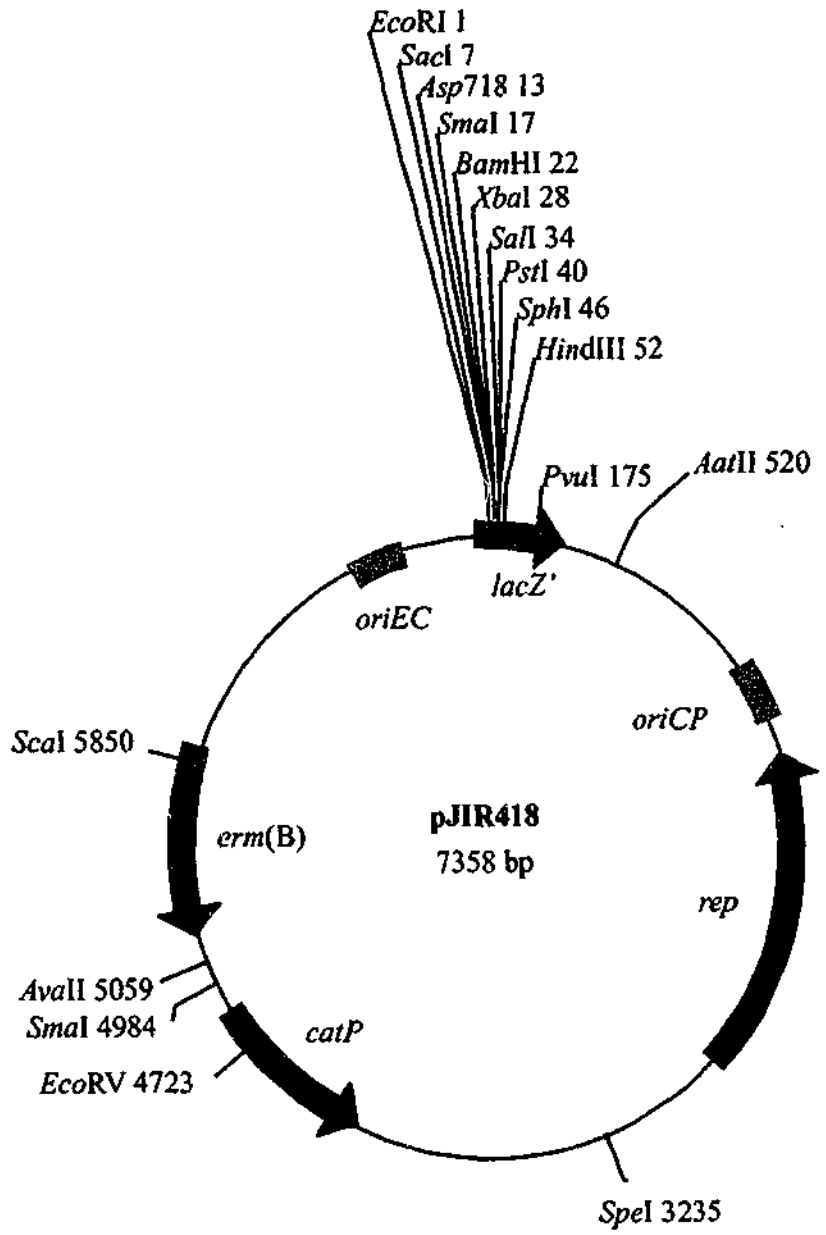
Mutagenesis of pJIR418

Mutagenesis was performed on the *E. coli-C. perfringens* shuttle plasmid pJIR418 (Figure 3.3), using *in vitro* hydroxylamine mutagenesis (Koutsis, 1993) and random mutagenesis *via* passage through the mutator strain XL1-Red. After mutagenesis, purified plasmid DNA was used to transform competent *E. coli* DH5α cells to chloramphenicol resistance, which is encoded by the *catP* gene located on pJIR418 (Figure 3.3). The resultant chloramphenicol resistant colonies were patched onto medium containing either chloramphenicol or erythromycin. Colonies that grew on chloramphenicol but not on erythromycin were selected as potential *erm*(B) mutants. Over 9000 chloramphenicol resistant *E. coli* colonies that were derived from three independent hydroxylamine and three independent random mutagenesis reactions were screened in this manner, resulting in the isolation of 38 chloramphenicol-resistant, erythromycin-sensitive colonies.

To reduce the possibility that the erythromycin-sensitive phenotype had resulted from gross changes to pJIR418, plasmid DNA was extracted from each of the 38 strains and restriction endonuclease analysis was used to check the overall

Figure 3.3 : Physical map of pJIR418.

Plasmid pJIR418 (Sloan *et al.*, 1992) is a shuttle vector which contains origins of replication for *E. coli* (*oriEC*) and *C. perfringens* (*oriCP*) (indicated by orange boxes) and the replication gene from the *C. perfringens* plasmid pIP404 (*rep*) (indicated by the green arrow). It encodes resistance to both erythromycin (*erm*(B)) and chloramphenicol (*catP*) (shown as blue arrows). It also contains a pUC18-derived multiple cloning site. The positions of restriction sites are indicated, relative to the single *Eco*RI site.


EcoRI 1
SacI 7
Asp718 13
SmaI 17
BamHI 22
XbaI 28
SalI 34
PstI 40
SphI 46
HindIII 52
PvuI 175
AatII 520
lacZ'
oriEC
oriCP
ScaI 5850
pJIR418
7358 bp
erm(B)
rep
AvaII 5059
SmaI 4984
catP
EcoRV 4723
SpeI 3235

profile of the plasmids. Digestion of pJIR418 with *Sma*I should result in two DNA fragments of 4967 bp and 2391 bp. The results showed that the plasmids from eight of the strains no longer had the correct profile (data not shown). These strains were eliminated from the study. The remaining plasmids all had the expected profile.

Sequence analysis of the mutated pJIR418 derivatives

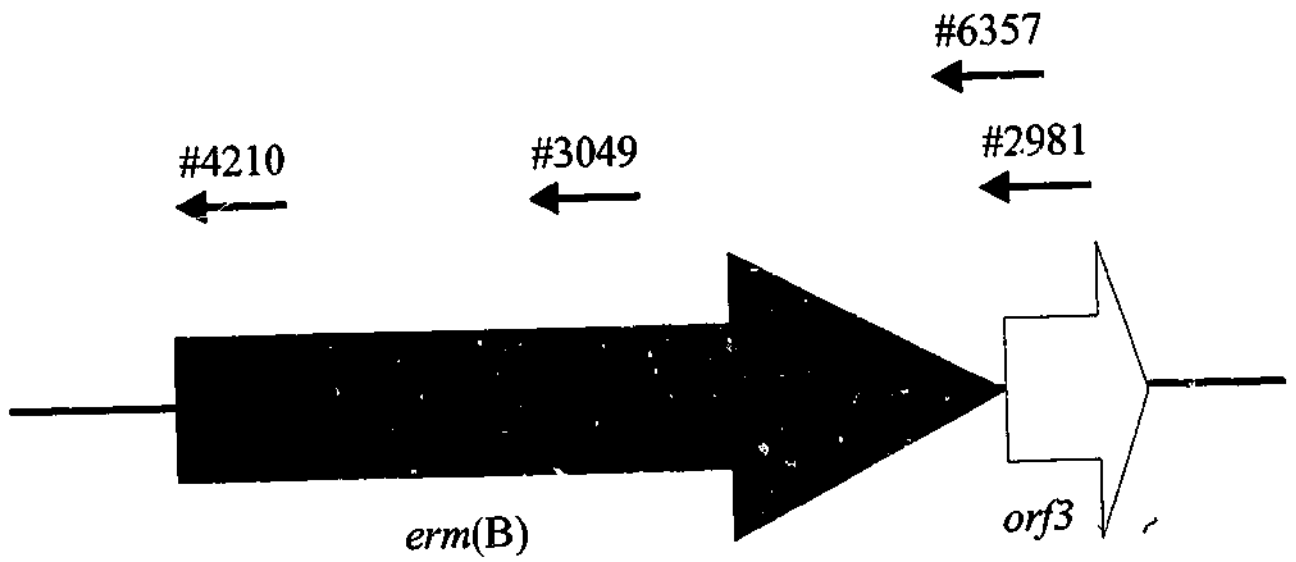
The remaining derivatives were sequenced using the oligonucleotide primers #2981, #3049 and #4210 (Figure 3.4) to identify the *erm*(B) mutation. When mutations were found, a second sequencing reaction was conducted using an appropriately positioned oligonucleotide primer on the opposite strand to confirm the mutation (Figure 3.4).

Two categories of erythromycin-sensitive mutants were identified. The first contained nine pJIR418 derivatives, each of which contained a single mutation in the *erm*(B) structural gene (Table 3.1). These mutations were scattered throughout the *erm*(B) gene. Three of these derivatives, pJIR977, pJIR934, and pJIR1571, were found to contain the same mutation.

The second category contained 19 of the mutated pJIR418 derivatives. These plasmids were found to have nucleotide changes that resulted in the introduction of a stop codon, or led to a frameshift resulting from the insertion or deletion of a single base (Table 3.2). These mutations would result in the production of truncated Erm(B) proteins (Table 3.2, Figure 3.5). Many derivatives in this category contained the same mutations. Two additional pJIR418 mutants were not studied further because they either contained multiple mutations (pJIR1573) or had no detectable mutations (pJIR972) (data not shown).

Figure 3.4 : Sequencing and amplification of the *erm*(B) gene.

Sequencing and amplification of the *erm*(B) gene (blue arrow) was undertaken using oligonucleotide primers that bound either within the *erm*(B) structural gene or within the downstream ORF, *orf3* (white arrow). The approximate binding sites, and polarity of these oligonucleotide primers, are indicated by their number and an arrow.


#6357
#4210
#3049
#2981
erm(B)
orf3

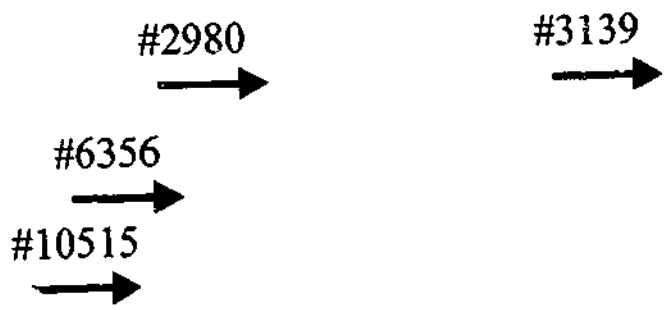

#2980
#3139
#6356
#10515

Table 3.1 : Characteristics of *E. coli* and *C. perfringens* strains carrying mutated pJIR418 derivatives that contained single missense mutations in the *erm*(B) gene.

Plasmid	Nucleotide Mutation*	Erm(B) derivative†	Erythromycin MIC (μg/ml)	
			E. coli	*C. perfringens*
pJIR418	N/A	Wild-type	>1280	>1280
pJIR750	N/A	Negative Control	160	<5
pJIR1576	G1620A	G37E	160	<5
pJIR1615	C1634T	H42Y	160	<5
pJIR883	G1682A	E58K	80	<5
pJIR1603	T1698C	L63P	640	640
pJIR1606	A1953C	Q148P	320	<5
pJIR977, pJIR934, pJIR1571	C2000T	P164S	640	<5
pJIR1613	G2019T	S170I	640	<5

*Nucleotide positions refer to the previously published sequence of the *erm*(B) gene region from *C. perfringens* (Berryman and Rood, 1995) (GenBank accession no. U18931). Nucleotide position 1511 is the first base of the *erm*(B) structural gene.
†Original Erm(B) residue, number of the residue, mutant residue.

Table 3.2 : Characteristics of *E. coli* and *C. perfringens* strains carrying mutated pJIR418 derivatives that contained nonsense mutations in the *erm*(B) gene.

Plasmid	Nucleotide Mutation*	Erm(B) derivative†	Erythromycin MIC (μg/ml)	
			E. coli	*C. perfringens*
pJIR418	N/A	Wild-type	>1280	>1280
pJIR750	N/A	Negative Control	160	<5
pJIR1599	1536ΔA	Erm(B)9+2*	160	<5
pJIR1596,			160	<5
pJIR1598,	1554ΔA	Erm(B)16+4*	160	<5
pJIR1605			160	<5
pJIR932,			80	<5
pJIR1570,	C1772T	Erm(B)87*	80	<5
pJIR1611			80	<5
pJIR1607	1841+A	Erm(B)112+3*	640	<5
pJIR1595,			160	<5
pJIR1597,	1851+T	Erm(B)114+1*	160	<5
pJIR1600			160	<5
pJIR1604	1851ΔT	Erm(B)114+9*	160	<5
pJIR971,			320	<5
pJIR973,	C1925T	Erm(B)138*	320	10
pJIR1551			160	<5
pJIR1	C1951T	Erm(B)147*	160	<5
pJIR1608,	C2120T	Erm(B)203*	160	<5
pJIR1609			320	<5
pJIR1610,	C2198T	Erm(B)229*	320	<5
pJIR1614			320	<5

*Nucleotide positions refer to the previously published sequence of the *erm*(B) gene region from *C. perfringens* (Berryman and Rood, 1995) (GenBank accession no. U18931). Nucleotide position 1511 is the first base of the *erm*(B) structural gene.

†Mutations have introduced either stop codons, e.g. Erm(B)88*, or a frameshift which leads to a certain number of normal Erm(B)-encoded residues plus residues following the frameshift up to the next stop codon, e.g. Erm(B)8+3*, where * indicates a stop codon.

Figure 3.5 : Truncated Erm(B) derivatives produced in this study.

A diagrammatic representation of the truncated Erm(B) derivatives is shown. The blue bar represents the wild-type Erm(B) protein (245 amino acids). The remaining bars represent truncated derivatives of the Erm(B) protein. Where the amino acid sequence of the derivative is the same as the wild-type protein the bar is colored grey. Where the amino acid sequence of the derivative is different from the wild-type protein the bar is colored red. For example, the truncated derivative Erm(B)9+2* is an 11 amino acid Erm(B) derivative of which the first eight residues are the same as the wild-type protein and the remaining three residues are different.

	Plasmid	Erm(B) Derivative
[redacted]	pJIR418	Erm(B)
[redacted]	pJIR1599	Erm(B)9+2*
[redacted]	pJIR1596, pJIR1598, pJIR1605	Erm(B)16+4*
[redacted]	pJIR932, pJIR1570, pJIR1611	Erm(B)87*
[redacted]	pJIR1607	Erm(B)112+3*
[redacted]	pJIR1595, pJIR1597, pJIR1600	Erm(B)114+1*
[redacted]	pJIR1604	Erm(B)114+9*
[redacted]	pJIR971, pJIR973, pJIR1551	Erm(B)138*
[redacted]	pJIR1602	Erm(B)147*
[redacted]	pJIR1608, pJIR1609	Erm(B)203*
[redacted]	pJIR1610, pJIR1614	Erm(B)229*

Effect of the *erm*(B) mutations on erythromycin resistance

To examine the quantitative effect of each of the *erm*(B) mutations on erythromycin resistance, MICs were determined in both *E. coli* and *C. perfringens*. MIC experiments were first performed on those *E. coli* and *C. perfringens* strains that carried pJIR418 derivatives with single point mutations. Each of the *E. coli* and *C. perfringens* strains carrying these mutated pJIR418 derivatives had an MIC of between 80 and 640 μg/ml in the *E. coli* background and <5 μg/ml in the *C. perfringens* background, with the exception of strains carrying pJIR1603, which had MICs of 640 μg/ml in both backgrounds (Table 3.1). The negative control plasmid used in these experiments, pJIR750 (Bannam and Rood, 1993), is a derivative of pJIR418 from which the *Ava*II-*Sca*I fragment containing the *erm*(B) gene has been deleted.

To ensure that the erythromycin-sensitive phenotype exhibited by these derivatives resulted from the single base mutations in the *erm*(B) genes and not from mutations elsewhere on pJIR418, the *erm*(B) gene from pJIR418 and from all of the missense mutants, except pJIR1603, were PCR amplified using the oligonucleotide primers #10515, which binds immediately upstream of the *erm*(B) RBS, and #6357, which binds immediately downstream of the *erm*(B) stop codon (Figure 3.4). These oligonucleotide primers introduced *Asp*718 sites to the ends of the *erm*(B) PCR products, enabling them to be cloned into the unique *Asp*718 site of pJIR750, adjacent to the *lac* promoter. Each of the pJIR750 derivatives were then sequenced across the *erm*(B) gene to ensure that no additional mutations had been introduced during PCR. These plasmids were used to transform *E. coli* DH5α cells to chloramphenicol resistance and the erythromycin MIC of each of the resultant strains was determined (Table 3.3). All pJIR750 derivatives carrying the mutated *erm*(B)

Table 3.3 : Characteristics of *E. coli* strains carrying pJIR750 derivatives into which mutated *erm*(B) genes were cloned.

Plasmid	Erm(B) derivative	Erythromycin MIC (μg/ml) in *E. coli*
pJIR1847	Wild-type	1280
pJIR750	Negative Control	80
pJIR1850	G37E	80
pJIR1853	H42Y	80
pJIR1848	E58K	80
pJIR1851	Q148P	320
pJIR1874	P164S	160
pJIR1852	S170I	80

genes conferred resistance to erythromycin at 320 μg/ml or less (Table 3.3), indicating that the mutations in the *erm*(B) gene were producing the erythromycin-sensitive phenotype in the mutated pJIR418 derivatives.

MIC experiments were also performed on the nonsense mutants (Table 3.2). These mutants conferred little or no resistance to erythromycin, depending on the position at which the *erm*(B) gene was truncated. Generally, those derivatives truncated within the first two-thirds of the *erm*(B) gene did not confer resistance, while those truncated in the last third of the *erm*(B) gene conferred a low level of resistance in the *E. coli* background (Table 3.2). In the *C. perfringens* background none of the truncated derivatives conferred resistance (Table 3.2).

Detection of mutant Erm(B) proteins by immunoblotting

The erythromycin-sensitive phenotype exhibited by the missense mutants could result from the production of either a stable but non-functional Erm(B) protein, or an unstable Erm(B) protein. To determine which of these possibilities was responsible for the phenotype, it was necessary to examine the proteins produced by these *erm*(B) derivatives. To detect the protein that was being produced, an Erm(B)-specific antiserum was required.

a) Production of the Erm(B) protein

To raise an Erm(B) specific antiserum it was first necessary to purify the Erm(B) protein. The *erm*(B) gene was amplified from pJIR418 using the oligonucleotide primers #6356 and #6357. Primer #6356 binds at the 5' end of the *erm*(B) structural gene and introduces a *Bam*HI site immediately preceding the ATG start codon, while primer #6357 binds at the 3' end of the *erm*(B) structural gene and

introduces an *Asp*718 site immediately following the TAA stop codon (Figure 3.4). Following amplification, the PCR product was digested with *Bam*HI and *Asp*718 and cloned into the *Bam*HI/*Asp*718 sites of the expression vector pRSETA. Sequencing of the resulting plasmid construct, pJIR1626, using oligonucleotide primer #7716, which binds upstream of the multiple cloning site in pRSETA, confirmed that the *erm*(B) gene had been cloned in the correct location for expression from the T7 promoter. Sequencing of the remainder of the gene confirmed that no mutations had been introduced during PCR amplification.

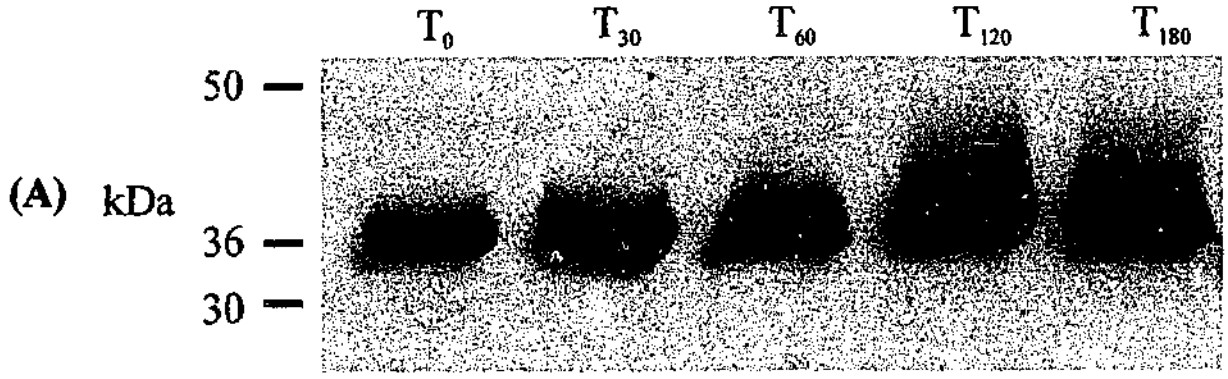
The plasmid pJIR1626 was introduced into the IPTG-inducible *E. coli* expression strain BL21(DE3). Erm(B) production from the resultant strain, JIR5753, was monitored over a 3 h time course following induction. The production of the His-Erm(B) protein was detected by immunoblot analysis of whole cell extracts using antibodies to the T7 tag, which is fused to the N-terminus of proteins produced from pRSETA. The results showed that a protein of 36-38 kDa was produced at all time points sampled and appeared to be produced in increased amounts as time progressed following induction (Figure 3.6A). The predicted molecular weight of the His-tagged Erm(B) protein is 37.92 kDa, which is in agreement with that observed in the T7 tag immunoblot.

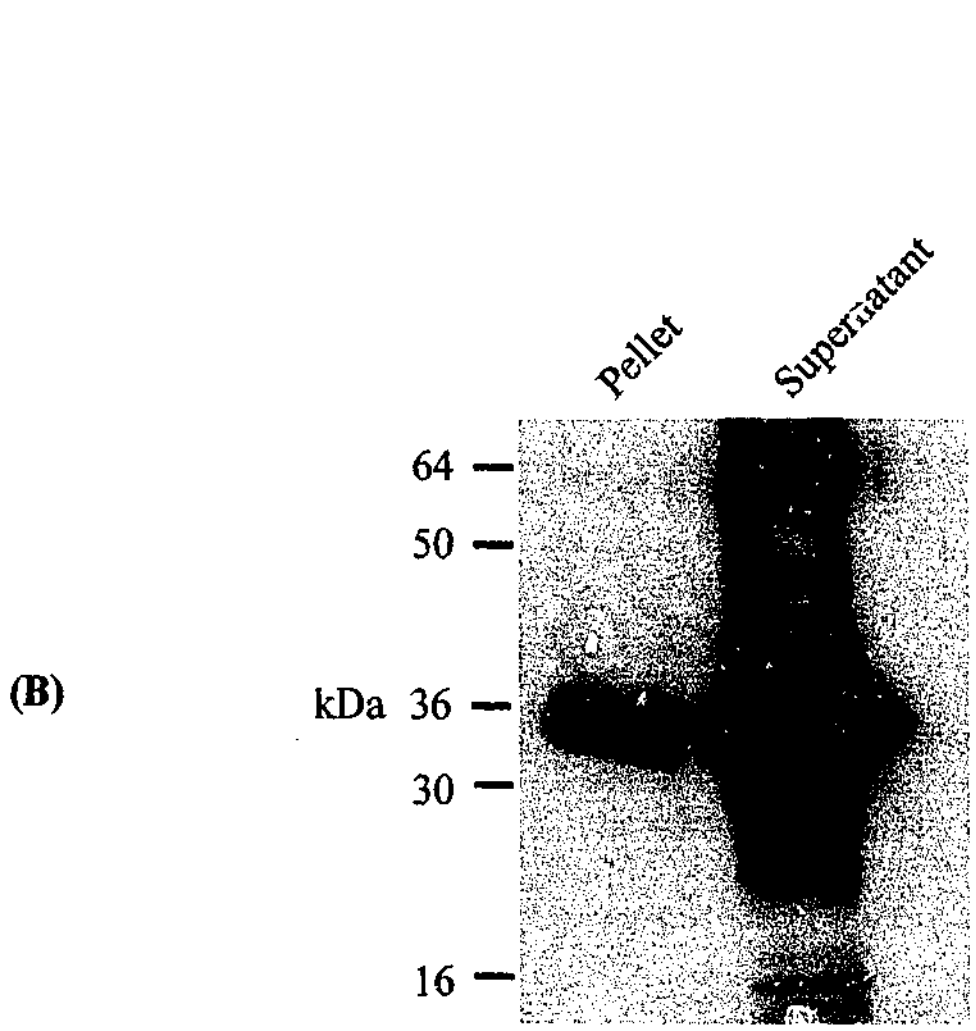
To determine the solubility of the His-Erm(B) protein, a whole cell lysate was prepared from a culture of JIR5753 that had been induced and grown for 3 h. The lysate was separated into pellet and supernatant fractions by centrifugation. Immunoblot analysis of these fractions using the T7 tag antibodies (Figure 3.6B) revealed that the majority of the His-Erm(B) protein was in the supernatant, indicating that it was soluble and therefore suitable for use in purification studies.

Figure 3.6 : Production of the Erm(B) protein.

(A) An overnight culture of *E. coli* strain JIR5753 was diluted 1 in 10 in 2YT broth containing ampicillin. The culture was grown for 1 h at 37°C, a 0 min (T_0) sample was removed, and protein production was then induced by the addition of IPTG (2 mM final concentration). Further samples were removed at 30 min (T_{30}), 60 min (T_{60}), 120 min (T_{120}) and 180 min (T_{180}) after induction, to determine when the His-Erm(B) protein was maximally produced. Cells were harvested by centrifugation (see Chapter 2), and the pellet resuspended in 2X sample buffer and boiled, to prepare whole cell extracts. The proteins present in the extracts were separated on 12% SDS-PAGE gels alongside appropriate size markers, and were then transferred to nitrocellulose. The production of the His-Erm(B) protein was detected by immunoblotting using the T7-tag antibody.

(B) Cells from a culture of JIR5735 that had been induced for 3 h were harvested by centrifugation at 4°C, and the pellet resuspended in PBS. The cells were then lysed using a French Pressure cell and the resultant lysate centrifuged at 4°C. The pellet was resuspended in PBS, and a portion of the supernatant retained. Proteins present in samples of the pellet and supernatant were separated on 12% SDS-PAGE gels alongside appropriate size markers, and were then transferred to nitrocellulose. The His-Erm(B) protein was detected by immunoblotting using the T7-tag antibody.


(A)
kDa
50
36
30
T_0
T_{30}
T_{60}
T_{120}
T_{180}


(B)
kDa
64
50
36
30
16
Pellet
Supernatant

b) Purification of the Erm(B) protein

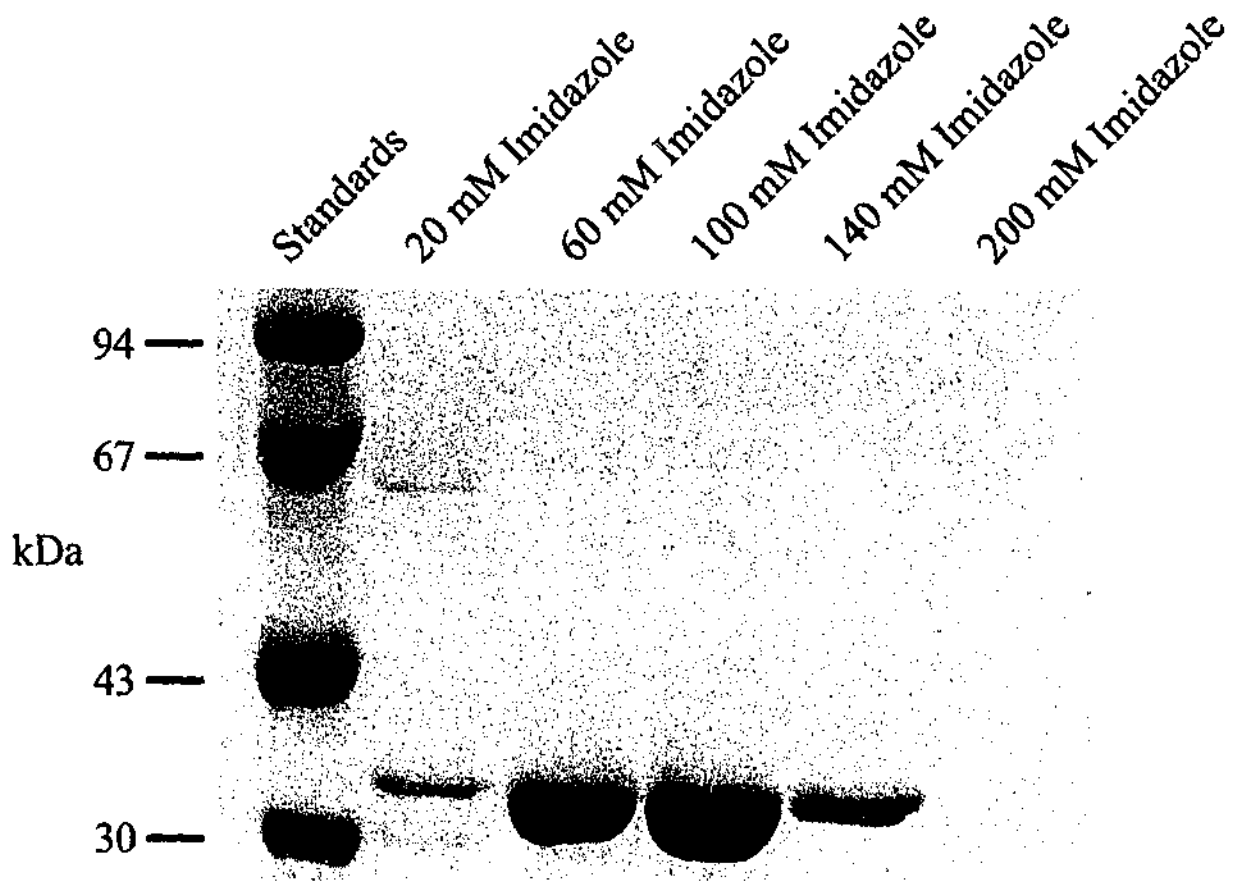
The His-Erm(B) protein was subsequently purified under non-denaturing conditions using Talon™ Metal Affinity Resin (Chapter 2) from a 1 L culture of JIR5753, which had been induced with 2 mM IPTG and grown at 37°C for 3 h. The His-Erm(B) protein was eluted from the Talon™ column in five 1 ml fractions using 20 mM, 60 mM, 100 mM, 140 mM and 200 mM imidazole. Proteins present in samples of each fraction were separated on a 12% SDS-PAGE gel and the gel stained with coomassie brilliant blue to detect the purified His-Erm(B) protein (Figure 3.7). A single protein species, which appeared to run at the expected molecular weight of His-Erm(B), was observed in all fractions, with the exception of the 200 mM imidazole fraction. Most of the His-Erm(B) protein appeared to be eluted in the 60 mM and 100 mM fractions, with small amounts being eluted at 20 mM and 140 mM (Figure 3.7). The 60 mM, 100 mM, and 140 mM imidazole fractions were pooled and, to remove the imidazole, were dialyzed against 20 mM Tris Buffer (pH 7.5) containing glycerol. The protein was then concentrated by centrifugation through a Millipore Ultrafree-15 Centrifugal Filter Device in a Beckman TJ-6 centrifuge to a final concentration of 875 μg/ml.

c) Production of Erm(B) antiserum

Antiserum to the purified His-Erm(B) protein was raised in two female New Zealand White rabbits (Chapter 2). The rabbits were initially injected intramuscularly with approximately 50 μg of His-Erm(B) protein emulsified in Freund's complete adjuvant and were given two boosts, consisting of approximately 50 μg of His-Erm(B) protein emulsified in Freund's incomplete adjuvant, administered at four weekly intervals. Two weeks following the second boost, approximately 100 ml of blood was collected from each rabbit *via* cardiac puncture.

Figure 3.7 : Purification of the His-Erm(B) protein.

A whole cell lysate was prepared from a culture of JIR5735 which had been induced by the addition of 2 mM IPTG and grown at 37°C for 3 h. The His-Erm(B) protein was purified from this lysate by metal affinity chromatography on a Talon™ column. The protein was eluted from the column in five 1 ml fractions using buffer containing 20 mM, 60 mM, 100 mM, 140 mM, or 200 mM imidazole. Proteins present in equivalent samples of each fraction were separated on a 12% SDS-PAGE gel alongside appropriate size standards, and the gel stained with coomassie brilliant blue to detect the purified His-Erm(B) protein.


Standards
20 mM Imidazole
60 mM Imidazole
100 mM Imidazole
140 mM Imidazole
200 mM Imidazole
94
67
kDa
43
30

Prior to use, the His-Erm(B) antiserum was adsorbed (Chapter 2) against the base strain, JIR5268, to remove non-specific antibodies. The adsorbed antiserum was then used at a 1 in 5000 dilution in all subsequent immunoblot analyses.

d) Detecting the Erm(B) proteins encoded by the mutated pJIR418 derivatives

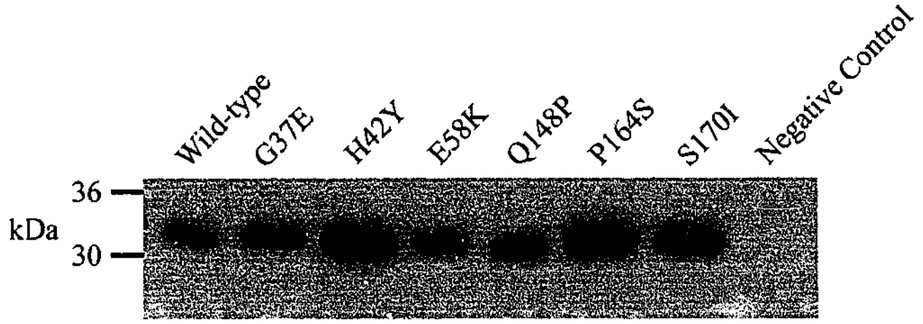
The ability to produce an Erm(B) protein was examined for each of the missense *erm*(B) mutants. Whole cell lysates were prepared from *E. coli* strains carrying plasmid derivatives encoding the wild-type Erm(B) protein (pJIR418), or the mutant Erm(B) proteins (Table 3.1) and also from a strain carrying a plasmid that did not encode the Erm(B) protein (pJIR750). The lysates were examined by Western blotting using the adsorbed His-Erm(B) polyclonal antiserum. The results showed that all six of the mutated derivatives that were tested produced similar levels of an immunoreactive protein that was the same size as the wild-type protein (Figure 3.8), suggesting that the erythromycin-sensitive phenotype exhibited by the *E. coli* and *C. perfringens* derivatives carrying these plasmids resulted from a difference in the structure or function of the Erm(B) protein rather than failure to produce Erm(B) or the production of an unstable Erm(B) protein.

e) Analysis of truncated Erm(B) mutants

Western blots were also carried out on cell extracts from strains carrying the *erm*(B) nonsense mutants. Whole cell lysates were prepared from *E. coli* strains carrying plasmid derivatives that encoded the wild-type Erm(B) protein, a selection of the nonsense mutants (Table 3.2), and from a strain carrying a plasmid that did not encode Erm(B), as before. The plasmids encoding Erm(B)9+2* and Erm(B)16+4* both contained a single base insertion in the *erm*(B) gene, resulting in a frameshift that would lead to the production of Erm(B) variants of only 11 (~1.5 kDa) and

Figure 3.8 : Western blot analysis of lysates from *E. coli* strains carrying plasmid derivatives encoding mutant Erm(B) proteins.

Proteins present in standardized amounts (10μg) of whole cell lysates prepared from *E. coli* strains carrying plasmids encoding wild-type Erm(B), G37E, H42Y, E58K, Q148P, P164S, S170I, or no Erm(B) protein (negative control), were separated by electrophoresis on a 12% SDS-PAGE gel and then transferred to nitrocellulose. The Erm(B) derivatives were detected using the adsorbed His-Erm(B)-specific polyclonal antiserum.



20 (~2.8 kDa) amino acids, respectively. The plasmids encoding Erm(B)87*, Erm(B)138*, Erm(B)147*, Erm(B)203* and Erm(B)229* contained point mutations that introduced stop codons and would result in the production of Erm(B) variants of 87 (~10.0 kDa), 138 (~16.0 kDa), 147 (~17.0 kDa), 203 (~23.8 kDa) and 229 (~26.9 kDa) amino acids, respectively.

Western blots showed that the Erm(B) antiserum detected most of these truncated variants (Figure 3.9). As expected, the antiserum was not able to detect the Erm(B)9+2* or the Erm(B)16+4* variants, which represent the two smallest truncated variants. The Erm(B)87*, Erm(B)138*, Erm(B)147*, Erm(B)203* and Erm(B)229* variants were detected, indicating that truncated Erm(B) variants that include as little as a third of the full length protein can be detected using this antiserum.

Two immunoreactive proteins were detected in extracts derived from strains carrying the plasmid encoding the Erm(B)138* variant. One band was the correct size (~16 kDa) for the predicted Erm(B) truncated variant, and the other corresponded in size to the full length Erm(B) protein. Two other mutated pJIR418 plasmid derivatives in this study, pJIR973 and pJIR1551, were found to contain the same point mutation as pJIR971 and would therefore also encode the Erm(B)138* variant. The production of Erm(B) was also examined in lysates from *E. coli* strains carrying these plasmids and the same two bands were produced (data not shown).

Structural modelling of the Erm(B) protein

The crystal structure of the rRNA methyltransferase Erm(C), formerly referred to as ErmC', from *B. subtilis* (Bussiere *et al.*, 1998) and the NMR structure

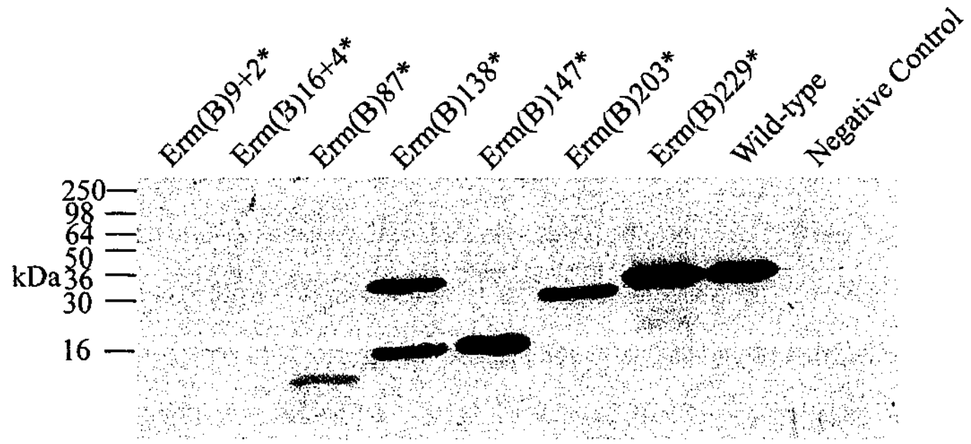


Figure 3.9 : Western blot analysis of lysates from *E. coli* strains carrying plasmid derivatives encoding truncated Erm(B) proteins.

Proteins present in standardized amounts (10µg) of whole cell lysates, prepared from *E. coli* strains carrying plasmids encoding wild-type Erm(B), Erm(B)9+2*, Erm(B)16+4*, Erm(B)87*, Erm(B)138*, Erm(B)147*, Erm(B)203*, Erm(B)229*, or no Erm(B) protein (negative control), were separated by electrophoresis on a 12% SDS-PAGE gel and then transferred to nitrocellulose. The truncated Erm(B) derivatives were detected using the adsorbed His-Erm(B) specific polyclonal antiserum.

of the closely related rRNA methyltransferase Erm(B), formerly referred to as ErmAM, from *S. pneumoniae* (Yu *et al.*, 1997), have been determined. The Erm(B) protein from *S. pneumoniae* is nearly identical to the Erm(B) protein from *C. perfringens*, differing in only three amino acid residues (at positions 100, 118, and 226, Figure 3.2), and is therefore a very good model for the *C. perfringens* Erm(B) protein. The Erm(B) and Erm(C) proteins have approximately 50% amino acid sequence identity and therefore the structure of Erm(C) also provides a reliable basis for a model of the *C. perfringens* Erm(B) protein.

Based on the NMR model for ErmAM (PDB entry 1YUB) and the crystal structure of Erm(C) (PDB entry 1QAO) the structure of the Erm(B) protein was modelled, by our collaborator, Dr. Galina Polekhina, using Program O (Jones *et al.*, 1991) (Figure 3.10). The predicted structure is bilobal, meaning that it consists of two main domains; the catalytic domain that contains the SAM-binding region and the catalytic region, and the RNA recognition domain. Based on this structure, as in all other methyltransferases, Motifs I-VIII and X would form the SAM-binding region and the catalytic region.

Each of the six mutations we obtained in the *erm*(B) gene result in a change to an amino acid either in or close to a conserved motif (Figure 3.2). We obtained two mutations within or close to Motif I (G37E and H42Y), two mutations in or close to Motif II (E58K and L63P), one mutation in Motif VII (Q148P), and two mutations in or close to Motif VIII (P164S and S170I). With the exception of L63P, the location of each of the mutated residues in relation to the structure of the Erm(B) protein, and the side chains of the normal residues at these positions, are shown in Figure 3.10. The fact that each of the mutations we obtained occurred within or

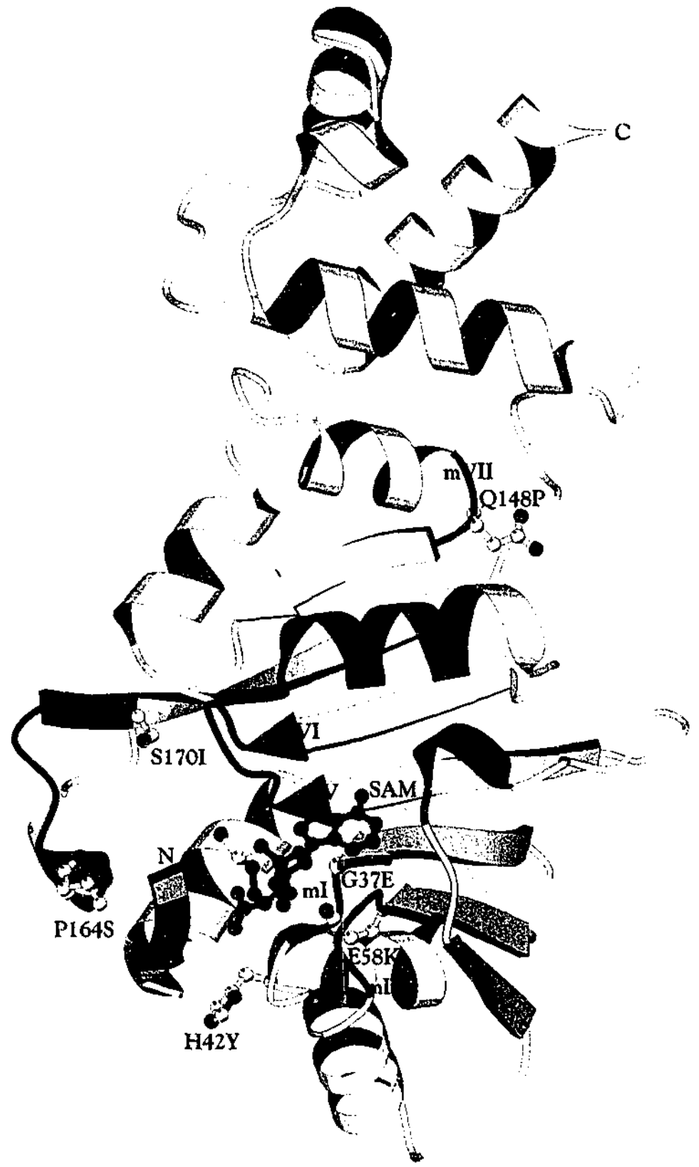


Figure 3.10 : Structural model of the *C. perfringens* Erm(B) protein.

The *C. perfringens* Erm(B) protein has a bilobate structure consisting of two domains. The upper domain, consisting mainly of α-helices, is the RNA recognition domain. The lower domain, containing the putative SAM-binding region and the putative catalytic region, is the catalytic domain. The conserved motifs, Motifs I-VIII and X are shown in green, and the positions of the residues mutated in this study are shown by indicating the side chain of the naturally occurring residue at these positions. A molecule of the substrate, SAM, is shown in complex with the structure.

close to a conserved methyltransferase motif suggests that derivatives producing these proteins exhibit an erythromycin-sensitive phenotype because the Erm(B) proteins are impaired in their ability to either bind SAM, to bind to the rRNA target, or to catalyze the transfer of the methyl group from SAM to the adenine residue in the RNA target.

SAM binding assays

Researchers working on DNA methyltransferases including the Dam methyltransferase (Wenzel *et al.*, 1991) and the EcoKI methyltransferase (O'Neill *et al.*, 1998) have been successful in showing specific SAM binding using a cross-linking assay. These assays involve incubating the methyltransferase protein and either ^{3}H- or ^{14}C-labelled SAM on ice, and then exposing the samples to UV radiation for varying lengths of time. During this procedure, labelled SAM is specifically crosslinked to proteins that have an affinity for the substrate, whereas SAM is not crosslinked to proteins to which it is unable to bind. The samples are then run on SDS-PAGE gels and proteins with bound SAM can be visualized by autoradiography.

To determine whether the Erm(B) mutant derivatives were able to bind SAM, purified wild-type *C. perfringens* Erm(B) protein (both His-Erm(B) and Erm(B) from which the His tag had been removed by enterokinase cleavage) and purified BSA as a negative control, were analyzed in a SAM binding assay using ^{3}H-labelled SAM. The experiment was conducted several times, varying conditions such as the concentration of Erm(B) protein and ^{3}H-labelled SAM, the length of time the samples were exposed to UV radiation, the distance of the sample from the UV source, the method of autoradiography used to detect the sample, and the length of

exposure. Despite these attempts it was not possible to detect the crosslinking of SAM to the wild-type Erm(B) protein.

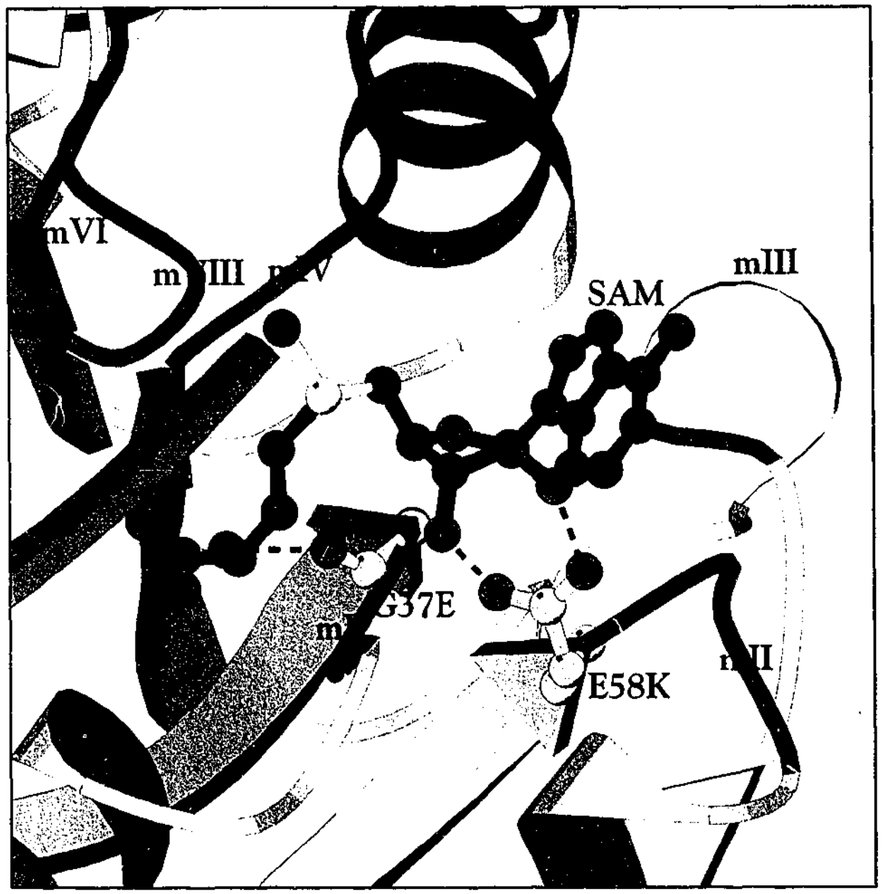
Discussion

To develop specific inhibitors of the Erm methyltransferases it is first necessary to determine which residues of these proteins are critical for their function. In this study seven pJIR418 derivatives that contained single point mutations in the *erm*(B) gene, were isolated after random mutagenesis. These mutations resulted in the production of Erm(B) proteins with single amino acid changes that either abolished or reduced erythromycin resistance in both *C. perfringens* and *E. coli* backgrounds. Each of the amino acid changes in the Erm(B) variants G37E, H42Y, E58K, L63P, Q148P, P164S and S170I, occurred within or close to conserved methyltransferase motifs that in other methyltransferases are either involved in SAM binding, RNA binding or catalysis of the methylation reaction.

Motif I creates part of the binding pocket for the methionine and ribose regions of SAM (Figure 3.11). The motif is centered around residues G-X-G (Figure 3.2), and is typically preceded by a D or E residue four residues toward the N-terminus (Bussiere *et al.*, 1998). It forms a secondary structure known as the G-loop, which binds the methionine moiety of SAM (Malone *et al.*, 1995). G37 and H42 are located within Motif I. Upon binding of SAM, a change in the backbone conformation of G37 allows the main chain carbonyl oxygen of this residue to accept a hydrogen bond from the α-amino nitrogen of SAM (Schluckebier *et al.*, 1999) (Figure 3.11), implying that flexibility in the structure is very important at this particular residue. In addition, the torsion angles of G37 lie in the area of the left-

Figure 3.11 : Interactions between SAM and residues in motifs I and II.

The structure of the Erm(B) protein in the region surrounding the SAM molecule is shown. The location of the conserved motifs I-VI and VIII are shown in green. The



handed helix region of the Ramachandran plot, which suggests that no other amino acid will be tolerated in this conformation. Therefore, the first mutant we obtained, G37E, is unlikely to be able to bind SAM either due to the lack of flexibility in Motif I, or due to the steric hindrance caused by the glutamate side chain partially occupying the region of the SAM binding site that is normally occupied by the methionine portion of SAM (Figure 3.12A). The second mutant we obtained in Motif I, H42Y, is not as easily explained. The mutation occurs in a histidine residue that is not well conserved among methyltransferases. The *C. perfringens* Erm(B) model (Figure 3.10) predicts H42 to be exposed, and the mutation to Y42 can easily be accommodated into the model without significant alterations to the structure (Figure 3.12A). However, structurally, H42 is in the vicinity of both Motif VIII and the SAM-binding site and it may influence the binding of RNA substrate either directly or indirectly through long distance structural changes.

Motif II, like Motif I, forms part of the SAM binding pocket. This motif contains a negatively charged residue that hydrogen bonds to the ribose hydroxyl groups of SAM, and is followed by a hydrophobic residue, which is within van der Waals contact of the adenine ring of SAM (Bussiere *et al.*, 1998). In Erm(B), E58 is the negatively charged residue and L59 is the hydrophobic residue (Figure 3.2, Figure 3.11). The mutation of the negatively charged E58 residue to K58, which is positively charged, is therefore highly likely to interfere with SAM binding (Figure 3.12B). A second mutant, L63P, was also obtained near Motif II. The Erm(B)L63P protein is partially functional, resulting in resistance to erythromycin, but not to wild-type levels (Table 3.1). L63 is not well conserved amongst methyltransferases and has not been shown to be involved in interactions with SAM. The reduction in erythromycin resistance exhibited by strains producing Erm(B)L63P is most likely

Figure 3.12 : Stereo views of the environment around mutations in motifs I and II.

Key residues are shown in ball-and-stick format. The top panel in each figure is a stereo view of the wild-type structure and the lower panel shows the mutation. (A) Motif I mutations (G37E and H42Y). (B) Motif II mutation (E58K). These figures were produced using BOBSCRIPT (Esnouf, 1999) and should be examined using the glasses provided at the back of the thesis.

due to minor structural changes to the SAM binding pocket, as proline residues commonly introduce bends in proteins.

Motifs IV, VI, VII and VIII are believed to be involved in the formation and folding of the catalytic site and in RNA binding. Motif IV is located in a region known as the P-loop (Malone *et al.*, 1995). The P-loop, along with Motif VI and Motif VIII, forms the active site of the protein. Motif IV of Erm(B) contains the sequence GSIPY, which corresponds to the catalytic sequences GNIPY in Erm(C) and GNPPY in M-TaqI (Figure 3.2) (Bussiere *et al.*, 1998). These sequences are thought to be involved in binding the target adenine residue *via* interactions between the N/S residue of the motif and the target adenine residue (Malone *et al.*, 1995). In addition, the Y residue in this motif, in conjunction with the F residue in Motif VIII and hydrophobic side-chains in Motif VI, could function in properly orientating the target DNA adenine (Malone *et al.*, 1995).

Motif VII is weakly conserved among methyltransferases, however, credible candidates can be found in most proteins (Malone *et al.*, 1995). This motif is believed to play a role in the folding of the catalytic region (Cheng, 1995). In this study we obtained a Motif VII mutant in which the Q148 residue was changed to P148 (Figure 3.13A). This change should not cause any steric clashes in the structure, however, as Motif VII is proposed to play a role in folding of the catalytic region, mutation of Q148 to P148 may influence the binding of RNA due to changes in the folding of the catalytic site.

Motif VIII is believed to be involved in recognition of the adenine residue in the target RNA. It contains a phenylalanine residue that has been proposed to

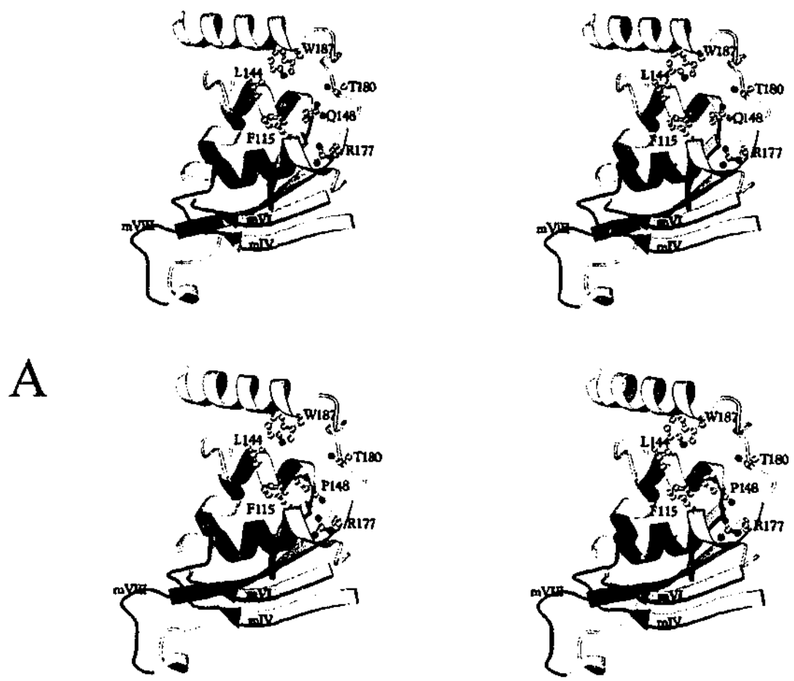


Figure 3.13 : Stereo views of the environment around mutations in motifs VII and VIII.

Key residues are shown in ball-and-stick format. The top panel in each figure is a stereo view of the wild-type structure and the lower panel shows the mutation. (A) Motif VII mutation (Q148P). (B) Motif VIII mutations (S70I and P164S). These figures were produced using BOBSCRIPT (Esnouf, 1999) and should be examined using the glasses provided at the back of the thesis.

interact with the target adenine (Schluckebier *et al.*, 1995) and to play a role in catalysis *via* cation-π interactions (Schluckebier *et al.*, 1998). This motif occurs in a loop that forms a prominent arched feature over the catalytic domain (Bussiere *et al.*, 1998) and is comprised of the residues 161-FHPKPKVNS-171 in Erm(B). This motif hangs over the active site and has been referred to as the adenine binding loop (Bussiere *et al.*, 1998). This motif is very well conserved amongst methyltransferase proteins and appears to have the consensus sequence FxPxPxVxS (Bussiere *et al.*, 1998). The first proline residue in this sequence is one of the key residues in the adenine-binding loop. This proline residue is conserved throughout the Erm [illegible] of methyltransferase enzymes and is found in the *cis* conformation (Schluc[illegible] *al.*, 1999), which is important for the conformation of the adenine binding lo[illegible] this study two mutations in Motif VIII, P164S and S170I, were obtained. The mutation of P164 to S164 would alter the conformation of Motif VIII and therefore is highly likely to disrupt RNA binding (Figure 3.13B). The effect of mutation of S170 to I170 is probably more difficult to understand. Based on the predicted Erm(B) structure two possible explanations exist to account for the loss of erythromycin resistance. S170 is in van der Waals contact with the conserved F162 residue, which is predicted to be involved in the correct positioning of the target adenine. Therefore, mutation of S170 to I170 could perturb the binding of the target adenine due to subtle changes in the binding pocket, which in turn would probably be sufficient to prevent transfer of the methyl group from SAM to the target adenine (Figure 3.13B). Secondly, S170 may form a hydrogen bond with E127, which in turn interacts with Y103, the conserved tyrosine residue in the catalytic sequence formed by Motif IV. Mutation of S170 to I170 may therefore destabilize the structure in the catalytic region and could result in these mutants being unable to

catalyze the transfer of the methyl group from SAM to the target adenine (Figure 3.13B).

Recent studies have reported several potential inhibitors of the Erm methyltransferases, which are essentially based on the ability of the end product of the methylation reaction, *S*-adenosyl-L- homocysteine (SAH), to inhibit the methylation reaction (Hajduk *et al.*, 1999; Hanessian and Sgarbi, 2000). These potential inhibitors bind to the active site of the Erm protein, thereby competing with the substrate of the methylation reaction, SAM. In this study we have identified several residues that are potentially involved in either binding of S ′M (G37 and E58) or binding of RNA (H42, Q148, P164 and S170), which are critical for function of the Erm(B) methyltransferase from *C. perfringens*. These residues, and the motifs they are part of, are generally well conserved among most Erm methyltransferase proteins and may represent good targets for the development of inhibitors to this important family of enzymes.

CHAPTER FOUR

IDENTIFICATION AND ANALYSIS OF THE *C. difficile* ELEMENT Tn*5398*

Introduction

Because *C. difficile* causes disease almost exclusively as a result of antimicrobial therapy, the antibiotic susceptibility of clinical isolates has been the subject of many studies. Generally, collections of *C. difficile* strains react fairly homogeneously to a given antimicrobial agent, with wide variations in susceptibility observed only against chloramphenicol, clindamycin, erythromycin, rifampicin, and tetracycline (Roberts *et al.*, 1994; Wüst and Hardegger, 1988). Resistance to erythromycin and clindamycin in *C. difficile* has always been of particular interest because clindamycin often seems to trigger the development of disease.

The Erm B resistance determinant from *C. difficile* was first detected by hybridization analysis (Berryman and Rood, 1989; Hächler *et al.*, 1987a). It has been shown to be transferred both intragenerically (Wüst and Hardegger, 1983) to *C. difficile* recipients, and intergenerically to both *S. aureus* (Hächler *et al.*, 1987a) and *B. subtilis* (Mullany *et al.*, 1995). In addition, *B. subtilis* transconjugants have been shown to be able to transfer the determinant back to *C. difficile* (Mullany *et al.*, 1995). Transfer occurs in the absence of detectable plasmid DNA. Integration of the determinant has been shown to occur at a specific site in *C. difficile* and in various chromosomal locations in *B. subtilis* (Mullany *et al.*, 1995). Because of these

observations it has been proposed that the *C. difficile* Erm B determinant resides on a conjugative transposon, which has been designated Tn*5398* (Mullany *et al.*, 1995).

The objectives of the research presented in this chapter were to identify and characterize the putative conjugative transposon Tn*5398*, and to determine the genetic organization of the Erm B determinant carried on this element.

Results

Cloning of Tn*5398* from the chromosome of *C. difficile* strain 630

Prior to this study, little was known about the size or genetic content of Tn*5398*, other than the fact that it encoded the Erm B determinant. Consequently, a shotgun approach was used to clone Tn*5398* from the genome of *C. difficile* strain 630. Preliminary Southern hybridization analysis was conducted on chromosomal DNA digested with a variety of enzymes. A DIG-labelled *C. perfringens erm*(B) probe was used to detect potential *erm*(B) genes. Southern hybridization analysis revealed that the *erm*(B) gene was located on approximately 9.7 kb *Hin*dIII and 20 kb *Xba*I fragments in the genome of strain 630 (data not shown). Accordingly, DNA from strain 630 was digested with *Hin*dIII or *Xba*I and ligated with similarly digested pWSK29 DNA. The ligated DNA was then used to transform *E. coli* DH5α cells to erythromycin resistance. Potential recombinants were selected on media containing erythromycin (150 μg/ml).

Two recombinant plasmids were isolated. The first plasmid, pJIR1594, contained an approximately 9.7 kb *Hin*dIII insert, the second, pJIR1790, contained the 20 kb *Xba*I fragment.

Sequence analysis of pJIR1594 and pJIR1790

Restriction endonuclease analysis of pJIR1594 and pJIR1790 revealed that the cloned fragments had common restriction fragments, which encompassed a region commencing at the *Xba*I site at the beginning of the pJIR1790 insert and at nucleotide position 1507 in the pJIR1594 insert, extending downstream to the *Hin*dIII site at nucleotide 8159 in the pJIR1790 insert and nucleotide 9666 in the pJIR1594 insert (Figure 4.1). Consequently, pJIR1594 was sequenced on both strands, firstly using the oligonucleotide primers UP and RP and then using a primer walking approach (Figure 4.1). The insert in plasmid pJIR1790 was sequenced on both strands, using a primer walking approach, across a region that extended from the *Asp*718 sites common to both plasmid inserts to approximately 5.8 kb downstream (Figure 4.1).

Analysis of the sequence using the BLASTX algorithm (Altschul *et al.*, 1997) revealed the presence of fifteen complete and one incomplete ORFs. The deduced amino acid sequence of each of these ORFs was analyzed using the BLASTP algorithm (Altschul *et al.*, 1997) to determine their similarity to known proteins in the databases.

***a)* ilvD**

The first ORF detected in pJIR1594 was incomplete (nucleotides 1 to 1040, Figure 4.2). No start codon, RBS, or promoter elements were found upstream of this ORF (Table 4.1). BLASTP analysis of the amino acid sequence showed homology to IlvD proteins from several organisms. IlvD is a dihydroxy acid dehydratase, that catalyzes the conversion of dihydroxyacids to branched-chain keto acids (Fink,

Figure 4.1 : Sequence analysis of pJIR1594 and pJIR1790.

A 14992 bp region from the recombinant plasmids was sequenced using the oligonucleotide primers shown. The extent of sequence obtained from each primer is indicated by the arrow below the number of the oligonucleotide primer. The location of the ORFs are shown as colored block arrows above the scale bar. The extent of the sequenced DNA in each recombinant plasmid, and the restriction enzyme profile of each DNA insert, is show below the scale bar. The dashed arrow in the pJIR1790 DNA restriction profile indicates DNA that was not sequenced.

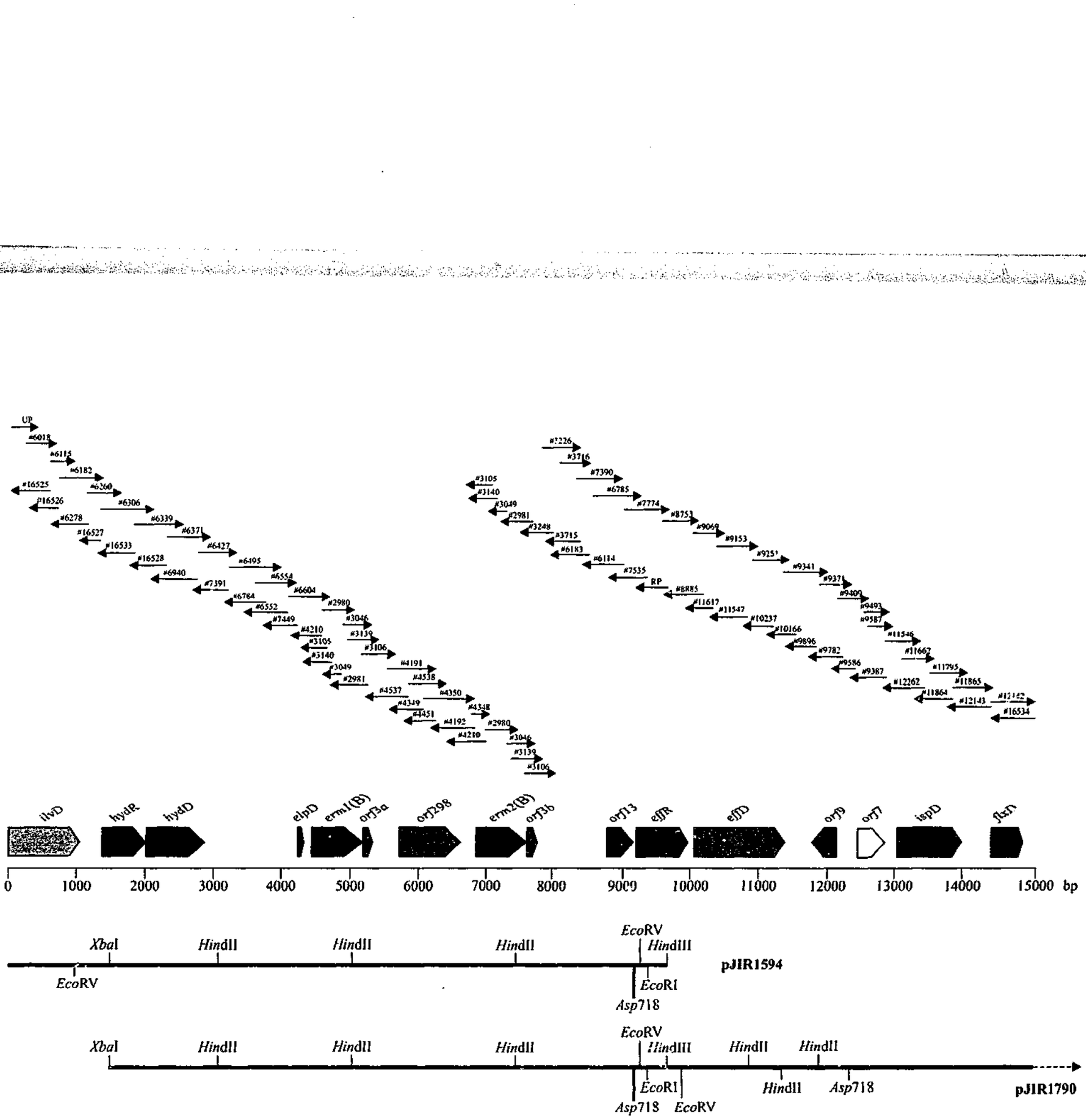


Figure 4.2 : Nucleotide sequence of Tn*5398* and its flanking regions.

Nucleotide positions are stated at the end of each line of sequence. The start and stop codons of each ORF are indicated in bold type face above the nucleotide sequence. The amino acid sequence of each predicted ORF is given below the appropriate nucleotide sequence. Potential RBS, -10 and -35 promoter sequences are underlined in bold. The potential right and left hand ends of Tn*5398* are indicated in bold above the nucleotide sequence. The beginning and end of regions encompassed by DR sequences are shown in bold above the nucleotide sequence. Palindromic sequences *palA* and *palB* are indicated by bold arrows below the nucleotide sequence. Potential *oriT* sites are shown as bold type face nucleotides within the sequence. The site at which a deletion event has removed the promoter sequences upstream of *erm2*(B) is marked by a black triangle (▲) beneath the nucleotide sequence. The GenBank accession number for the the Tn*5398* element and flanking sequence is AF109075.

```
TAGGCATGGGGATACCATATAATGGTACTGCTGCGTCACATTCTGGAGAAAGAAAAAGGA  60
   G  M  G  I  P  Y  N  G  T  A  A  S  H  S  G  E  R  K  R
TAGCAAAATATGCAGGTATGTATGTTATGGAGTTACTTAAGAACGACATAAAACCTAGAG  120
I  A  K  Y  A  G  M  Y  V  M  E  L  L  K  N  D  I  K  P  R
ATATTTTAACAATAGATGCTTTTAAAAATGCTATAGCTGTGGATATGGCAATGGCTGGTT  180
D  I  L  T  I  D  A  F  K  N  A  I  A  V  D  M  A  M  A  G
CTACAAATACAGTACTTCACTTACCTGCAATAGCTATATGAATCAGGAATAGAGCTTAACT  240
S  T  N  T  V  L  H  L  P  A  I  A  Y  E  S  G  I  E  L  N
TAGATTTTTTTGATGAAATAAGTGAAAAAACTCCATGTTTAACAAAATTAAGTCCAAGTG  300
L  D  F  F  D  E  I  S  E  K  T  P  C  L  T  K  L  S  P  S
GAAAACATCATATTGAAGATTTACATATGGCAGGAGGAATACCAGCTATAATGAACGAGC  360
G  K  H  H  I  E  D  L  H  M  A  G  G  I  P  A  I  M  N  E
TTTCAAAGATAAATGGAATAAATTTAGATTGCAAAACCGTAACAGGCAAGACTATAAGGG  420
L  S  K  I  N  G  I  N  L  D  C  K  T  V  T  G  K  T  I  R
AAAATATAAGAAATTGTGAAATAGAAAAT TAAGAAGTAATACATACATTAAAGAATCCAT  480
E  N  I  R  N  C  E  I  E  N  L  E  V  I  H  T  L  K  N  P
ATAGTAACCAAGGTGGGCTTGCAATATTGAAAGGAAATCTTGCTCTAAATGGAGCTGTTG  540
Y  S  N  Q  G  G  L  A  I  L  K  G  N  L  A  L  N  G  A  V
TAAAAAATCAGCAGTTGCAGAAGAATGTTAGTTCCATGAAGGACCCAGCAAGAATTTTTA  600
V  K  N  Q  Q  L  Q  K  N  V  S  S  M  K  D  P  A  R  I  F
ATTCAGAAGAAGAAGCTGTAAATGCTATTTTTGGTAAAAAAATAAACAAAGGTGATGTTA  660
N  S  E  E  E  A  V  N  A  I  F  G  K  K  I  N  K  G  D  V
TAGTTATAAGATATCAAGGTCCAAAGGGTGGTCCAGGAATGAAAGAAATGCTATCTCCTA  720
I  V  I  R  Y  E  G  P  K  G  G  P  G  M  K  E  M  L  S  P
CATCAGCAGTTGCAGGAATGGGACTTGATAAGCATGTAGCACTTCTTACTGATGGTCGTT  780
T  S  A  V  A  G  M  G  L  D  K  H  V  A  L  L  T  D  G  R
TTTCAGGGGCAACTACAGGAGCATCTATAGGCCATATTTCTCCAGAAGCTATGGAAGGTG  840
F  S  G  A  T  T  G  A  S  I  G  H  I  S  P  E  A  M  E  G
GTTTAATCGGACTAGTTGAAGAAGGAGATATAATTTCTATAAACATACCAGACAAAAAAT  900
G  L  I  G  L  V  E  E  G  D  I  I  S  I  N  I  P  D  K  K
```

```
TAGAGTTAAAAGTAGATGAAGTTGAAATAGAAAATAGAAAATTAAAATTTAAACCTTTAG   960

L  E  L  K  V  D  E  V  E  I  E  N  R  K  L  K  F  K  P  L

AACCGAAAATAAAGCATGGATACTTAAGTAGATATGCTAAATTGGTAACATCAGCAAATA  1020

E  E  L  K  V  D  E  V  E  I  E  N  R  K  L  K  F  K  P  L
                Stop ilvD |
CAGGAGCAGTTTTAAAATAGTAGACTATATTTATGTAAATTAGAAAAAGTTAAGAAAATA  1080

T  G  A  V  L  K
         EcoRV
TTAATTGGATATCTCAGGTGTTATATCACTAATGAGGTATCCTTTTTATTTTGTATGAAA  1140

AACACAGCAAAATCTTGATTATCTTTGAAAACAATACATCTCTATTATAAGGAATACGAG  1200

TACAGCATAGAGATGTATCTTTTAGACCTTGTTGAGGTAATACAACTTTTTAAAAGTTTG  1260

ATGTGGAATGCTTTTCGTTCCATCTGAACTATGAAGAAATTGAAAGTCAGTTAATTCTTG  1320
                                                      RBS
ACATAATAAATATTAAAATGTTTAATAAAAAACAAACAGACAGTCTGTTTTTAGGAGGGG  1380

    Start hydR
AATATGAATAGAGAAGAAAAAAGTAAAAATAGTAAAGAAAAAATTATCCAATCAGCATTT  1440

   M  N  R  E  E  K  S  K  N  S  K  E  K  I  I  Q  S  A  F

TCACTATTTTCGTCTAAAGGATATGATTCAACATCTACACAAGATATTATCAATTTATCT  1500

 S  L  F  S  S  K  G  Y  D  S  T  S  T  Q  D  I  I  N  L  S
         XbaI
GGTCTATCTAGAGGTGCAATGTATCATCACTTTAAAACTAAAGAAGATATACTGAGAAGT  1560

 G  L  S  R  G  A  M  Y  H  H  F  K  T  K  E  D  I  L  R  S

GTCACAAAAGAACTTTACTCACAAATGAATAATTTTTTAGAGTATCTTGTTGCTGATGAC  1620

 V  T  K  E  L  Y  S  Q  M  N  N  F  L  E  Y  L  V  A  D  D

ACCCTTACAGCAAATGAAAAAATAATAGAATTGGTTGTTCATAGTGCGAATGATTACACA  1680

 T  L  T  A  N  E  K  I  I  E  L  V  V  H  S  A  N  D  Y  T

CGTAGAAAAATGGTACATTGTAGCTGGTTAGAAAAAATCCCATTCGCTTTAATAGAGGAA  1740

 R  R  K  M  V  H  C  S  W  L  E  K  I  P  F  A  L  I  E  E

GTTCGTAATCTCAACAATGTAGTTGCACCCAATATTGCTAAGATAATTAAACAAGGTGTT  1800

 V  R  N  L  N  N  V  V  A  P  N  I  A  K  I  I  K  Q  G  V

GAAAATAAAGAATTTTCTTGTGAATATCCAGAGGAATTAGCTGAAATGCTCGTTTTTAGT  1860

 E  N  K  E  F  S  C  E  Y  P  E  E  L  A  E  M  L  V  F  S

ATTGATATTTTACTTGACCCTGTATTATTTAAACGTGAATACAGTGAAGTATGTAATAGG  1920

 I  D  I  L  L  D  P  V  L  F  K  R  E  Y  S  E  V  C  N  R
```

TTGGATTTTTTATTATTCATGTTAAAAAAAATGGATATTCCTCTAATTGATGAGTACGGT 1980

L D F L L F M L K K M D I P L I D E Y G

Stop *hydR* | **RBS** **Start *hydD***

ATTCAAAAATTTAAAGATTTATTTAAACAATGATAATAGGAAAAGTGAGGAATAGACTAT 2040

I Q K F K D L F K Q M

GTTTTATAATGCAAAAAGCTGTCAAATGATGATAGGAAGTAATACAGTGAATTATATTGA 2100

F Y N A K S C Q M M I G S N T V N Y I E

ATTTGGTAGTGGTAAGAAGACACTAATTATTTTACCAGGATTAGGAGATGGTTTATTTCC 2160

F G S G K K T L I I L P G L G D G L F P

TTTACATGGAAAAATACAGGCGATTGCTTTTGCTTTTAGATATAAGCAATTTGCAAAAGA 2220

L H G K I Q A I A F A F R Y K Q F A K D

TTATAAGGTTTATGTATTTAGTAGAAAAAATCAAATTACAGAAAAATATTCCACAAGGGA 2280

Y K V Y V F S R K N Q I T E K Y S T R D

TATGGCTAAAGACCAAGCAGATATAATGAAAAAACTTGGAATTATGAAAGCAGAAGTAAT 2340

M A K D Q A D I M K K L G I M K A E V M

GGGTGTTTCTCAAGGTGGAATGATAGCTCAATATCTGGCAATTGATTATCCTGAGTTAGT 2400

G V S Q G G M I A Q Y L A I D Y P E L V

TGAGAAACTTGTGTTGGCAGTTACTTCTTCAAAGCAAAATGATACTATCCAAAATGTTAT 2460

E K L V L A V T S S K Q N D T I Q N V I

TTGTAGCTGGATTGATATGGCGAAGAAGCAGAATTACAATGATTTAATGATTGACACAGC 2520

C S W I D M A K K Q N Y N D L M I D T A

AAAGAAATCCTATTCTGAAAGATATTTAAAAAAATACCAATTATTTATTCCTTTTTTAGG 2580

K K S Y S E R Y L K K Y Q L F I P F L G

AAAAGTAGGTAAACCAAAGGATTTTAAACGATTTATTATTCAAGCGACCTCTTGCATAGA 2640

K V G K P K D F K R F I I Q A T S C I E

ACATAATGCCTTTTCCGAATTAAATAAAATAACATGTCCAACCTTAATTATAGGAGGTGC 2700

H N A F S E L N K I T C P T L I I G G A

TAATGATAAAATTGTTGGAAATAATGCTTCTTTTCATTTAGCAGAAAAGATTAAAAAAAG 2760

N D K I V G N N A S F H L A E K I K K S

CGAAATATTTATCTATGAGGGATTGGGACACGCTACATATGAAGAAGCACAGGATTTTAA 2820

E I F I Y E G L G H A T Y E E A Q D F N

Stop *hydD* |

TGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCTTTGTTAAAAATATACATATGAAG 2880

E R V L E F L N K

ATTGGAAATTTAATGTTAAAAATAGAAACATGAAAATATGCTTAACTGGTATTTTTACTA 2940

TTCATAACCAATTTTTAATACATTATCTACTATAAATACAAATATAGCTTCAATGTGATT 3000

ATATATTGTTGTATTGGTAAAGCACTTATACAAACAGAGGAATTTTGTAAATTCAGATTA 3060

*Hin*dII | Left End of Tn*5398*

TATCCACATTT**GTTAAC**TTATGAAAATATAATCAAAATTTTTATGAGCTTTTTATATAAA 3120

AAAACGCCCTAAAAATCTGATTATCCCCATAAACACTGTATCTACAAGCATATTCAATAG 3180

GAAATAAATCGTGATATTACTACGAATTTACTACTAATTTACTACTAATGAATGAGCCTT 3240

GATACGTCTTATTTCCCAGATATGCAAAGATATGGCATGGCACATCAGTAAAAATTGAAT 3300

ACTTATATAGACTATGGAACGTACACTTTTGGCGTTCCTTTTCTATTTCCAGACGTTCTT 3360

TTCAGAGCGTCTTTTTTTCATACCCAAAATCGAAAGGAGAAAGAGAAAATATGAATAAGC 3420

TAGTAAAGCGATTGCTGACAGGGACGCTCGCCATTGCAACCATTCTTACCGCATTGCCTG 3480

TGACGGTGGTTCATGCTTCTGGCAATTAATACTGGACAGAATCAGCAGAACGTGTCGGCT 3540

ACATTGAACATGTTATGAATGATGGTTCTATCAAATCCAAATTAAATGAGGGACACATGA 3600

AAGTTGAGGGCGAAACTGCCTATTGCGTGAACATCAATACAAATTTCAAAAATGGATATA 3660

AAACAAGGTATGACGCAAGCTCCCATATGAGTAGCGATCAGATTGCGGACATTGCTCTTT 3720

CCTTAGAGTACGTCAAGTAATATACTGCTTCTCATACAAACTTGAATTACAAGCAGGGTT 3780

ACTTATTGGAACAGTGTGTTGTCTGGCAGAGATTGAGTGAACAGCTCGGCTGGCAATGTG 3840

ATAACGTCAGAGCCTCCTATAATGAAATCTCACAGGCGGTACAGAATAAAGTTTACGCTG 3900

GTGCGAAAGCATTTGTGAAAGCAAATAAGGGGTGCTATGAATGTGGCGGTTACATCTACA 3960

CTGGCGAAGGACAGGACATTGGACAGTTCTGGGCGAGTTGAATGTAGGAAATGAAAAGGT 4020

| Start DR sequence

CAAAAAGACTTCTTCCAAATCATAAAAATCGAAACAGCAAAGAATGGCGGAAACGTAAAA 4080

Stop DR sequence| -35

GAAGTTATGGAAATAAGACTTAGAAGCAAACTTAAGAGTGTGTTGATAGTGCATTATCTT 4140

-10 RBS

AAAATTTTGTATAATAGGAATTGAAGTTAAATTAGATGCTAAAAATTTGTAATTAAGAAG 4200

Start *elpD*

GAGGGATTCGTCATGTTGGTATTCCAAATGCGTTATCAAATGCGTTATGTAGATAAAACA 4260

M L V F Q M R Y Q M R Y V D K T

Stop *elpD*

TCTACTGTTTTGAAACAGACTAAAAACAGTGATTACGCAGATAAATAAATACGTTAGATT 4320

S T V L K Q T K N S D Y A D K

AATTCCTACCAGTGACTAATCTTATGACTTTTTAAACAGATAACTAAAATTACAAACAAA 4380

Start *erm1*(B)

TCGTTTAACTTCTGTATTTGTTTATAGATGTAATCACTTCAGGAGTGATTACATGAACAA 4440

M N K

AAATATAAAATATTCTCAAAACTTTTTAACGAGTGAAAAAGTACTCAACCAAATAATAAA 4500

N I K Y S Q N F L T S E K V L N Q I I K

ACAATTGAATTTAAAAGAAACCGATACCGTTTACGAAATTGGAACAGGTAAAGGGCATTT 4560

Q L N L K E T D T V Y E I G T G K G H L

AACGACGAAACTGGCTAAAATAAGTAAACAGGTAACGTCTATTGAATTAGACAGTCATCT 4620

T T K L A K I S K Q V T S I E L D S H L

ATTCAACTTATCGTCAGAAAAATTAAAACTGAATACTCGTGTCACTTTAATTCACCAAGA 4680

F N L S S E K L K L N T R V T L I H Q D

TATTCTACAGTTTCAATTCCCTAACAAACAGAGGTATAAAATTGTTGGGAGTATTCCTTA 4740

I L Q F Q F P N K Q R Y K I V G S I P Y

CCATTTAAGCACACAAATTATTAAAAAAGTGGTTTTTGAAAGCCGTGCGTCTGACATCTA 4800

H L S T Q I I K K V V F E S R A S D I Y

TCTGATTGTTGAAGAAGGATTCTACAAGCGTACCTTGGATATTCACCGAACACTAGGGTT 4860

L I V E E G F Y K R T L D I H R T L G L

GCTCTTGCACACTCAAGTCTCGATTAAGCAATTGCTTAAGCTGCCAGCTGAATGCTTTCA 4920

I L H T Q V S I K Q L L K L P A E C F H

TCCTAAACCAAAAGTAAACAGTGCCTTAATAAAACTTACCCGCCATACCACAGATGTTCC 4980

P K P K V N S A L I K L T R H T T D V P

AGATAAATATTGGAAACTATATACGTACTTTGTTTCAAAATGGGTCAATCGAGAATATCG 5040

D K Y W K L Y T Y F V S K W V N R E Y R

*Hind*III

TCAACTGTTTACTAAAAATCAGTTTCATCAAGCAATGAAATACGCCAAAGTAAACGATTT 5100

Q L F T K N Q F H Q A M K Y A K V N D L

-35 -10

AAGTACCGTTACTTATGAGCAAGTATTGTCTATTTTTAATAGTTATCTATTATTTAACGG 5160

S T V T Y E Q V L S I F N S Y L L F N G

Stop *erm1*(B) | Start *orf3a*

GAGGAAATAATTCTATGAGTCGCTTTTTTAAATTTGGAAAGTTACACGTTACTAAAGGGA 5220

RBS

R K M S R F F K F G K L H V T K G

ATGGAGATAAATTATTAGATATACTACTGACAGCTTCCAAGAAGCTAAAGAGGTCCCTAG 5280

N G D K L L D I L L T A S K K L K R S L

Stop *orf3a* |

CGCCTACGGGGAATTTGTATCGATAAGGGGTACAAATTCCCACTAAGCGCTCGGGACCCC 5340

A P T G N L Y R

TTGTAGGAAAATGTCCTAAGTGTGGCAACAATATTGTATTAAAAAAATCGTTTTATGGTT 5400

GTTCAAATTATCCTGAATGTAAGTTTACTTTAGCTGAACATTTTAGAAAGAAAAAACTAA 5460

CCAAAACGAATGTAAAAGAATTACTGGAGGGAAAAGAAACCCTGGTAAAAGGAATCAAAA 5520

| Start DR sequence

ACAAAGAGAAAAAGCCCTACAATGCCGTTGTAAAAATTGGGGAAAAGGGATATATTGATT 5580

TTATATCTTTCTCAAAATAAATATAAAAGCCCTTTAAAGAGGGCTTTTATATATTAATCA 5640

CAAATCACTTATCACAAATCACAAGTGATTAATCACAAATCACTTGTGATTTGTGATTCT 5700

palA

Start *orf298*

TAATGATACAATATTACTATACAAAAAAAGAATGGGGCGTAGTTATGGAGAAGGAAAAAC 5760

M I Q Y Y Y T K K E W G V V M E K E K

TAAAAATACTTGAAGAATTAAGACGTATTTTAAACAATAAAAATGAAGCAATTATT**ATCT** 5820

L K I L E E L R R I L N N K N E A I I I

oriT (1)

TGAATAATTACTTTAAAGGTGGTGTTGGAAAGTCCAAATTATCGACTATGTTTGCTTACT 5880

L N N Y F K G G V G K S K L S T M F A Y

TGACAGACAAATTGAATTTAAAAGTTTTAATGATCGATAAGGACTTACAGGCAACATTGA 5940

L T D K L N L K V L M I D K D L Q A T L

CAAAAGACTTAGCAAAAACTTTTGAGGTAGAATTGCCACGTGTCAATTTTTATGAAGGCT 6000

T K D L A K T F E V E L P R V N F Y E G

TGAAAAATGGAAACTTGGCTTCTTCTATTATTCATTTGACTGATAATTTAGACTTGATCC 6060

L K N G N L A S S I I H L T D N L D L I

CTGGCACGTTTGATTTGATGTTACTGCCAAAATTAACTCGCTCATGGACGTTTGAAAATG 6120

P G T F D L M L L P K L T R S W T F E N

AAAGTAGATTGCTTGCTACTCTTTTAGCACCTTTAAAAAGTGACTATGATCTTATTATTA 6180

E S R L L A T L L A P L K S D Y D L I I

TTGATACTGTACCAACGCCAAGCGTTTATACAAATAATGCAATCGTGGCGAGTGATTACG 6240

I D T V P T P S V Y T N N A I V A S D Y

TTATGATCCCTTTACAAGCAGAAGAAGAAAGTACAAACAACATTCAAAACTATATTTCCT 6300

V M I P L Q A E E E S T N N I Q N Y I S

ATTTGATTGATTTGCAAGAACAGTTTAACCCTGGACTAGATATGATTGGTTTTGTTCCTT 6360

Y L I D L Q E Q F N P G L D M I G F V P

ATTTAGTTGATACGGACAGCGCAACGATAAAATCAAACCTGGAAGAACTGTACAAAGAAC 6420

Y L V D T D S A T I K S N L E E L Y K E

ATAAAGAGGATAATTTGGTTTTCCGAAATATTATCAAGCGAAGTAATAAAGTAAGTACTT 6480

H K E D N L V F R N I I K R S N K V S T

GGTCTAAAAATGGCATTACAGAACACAAAGGCTATGACAAAAAAGTTTTGTCTATGTATG 6540

W S K N G I T E H K G Y D K K V L S M Y

Stop *orf298*

AGAACGTATTTTTTGAAATGATTGAGCGAATCATTCAATTAGAAAATGAAAAAGAATAGA 6600

E N V F F E M I E R I I Q L E N E K E

ATCACAAATCACAAGTGATTAATCACAAATCACTTGTGATTTGTGATAGGTGATGATAAA 6660

palB

ATAAATAGTAAGAAGAAATAGAAAGAAGTGAGCGATCGTGGGAAATTTAGGCGCACAGAA 6720

AGCAAAACGAAATGATACGCCAATCAGCGCAAAAAAAGATATAATGGGGGATAAGACGGT 6780

Stop DR Sequence

Start *erm2*(B)

TCGTGTTCGTGCCGACTTGCACCATATCATAAAAATCGAAACAGCTCAGGAGTGATTACA 6840

TGAACAAAAATATAAAATATTCTCAAAACTTTTTAACGAGTGAAAAAGTACTCAACCAAA 6900

M N K N I K Y S Q N F L T S E K V L N Q

TAATAAAACAATTGAATTTAAAAGAAACCGATACCGTTTACGAAATTGGAACAGGTAAAG 6960

I I K Q L N L K E T D T V Y E I G T G K

GGCATTTAACGACGAAACTGGCTAAAATAAGTAAACAGGTAACGTCTATTGAATTAGACA 7020

G H L T T K L A K I S K Q V T S I E L D

GTCATCTATTCAACTTATCGTCAGAAAAATTAAAACTGAATACTCGTGTCACTTTAATTC 7080

S H L F N L S S E K L K L N T R V T L I

ACCAAGATATTCTACAGTTTCAATTCCCTAACAAACAGAGGTATAAAATTGTTGGGAGTA 7140

H Q D I L Q F Q F P N K Q R Y K I V G S

TTCCTTACCATTTAAGCACACAAATTATTAAAAAAGTGGTTTTTGAAAGCCGTGCGTCTG 7200

I P Y H L S T Q I I K K V V F E S R A S

ACATCTATCTGATTGTTGAAGAAGGATTCTACAAGCGTACCTTGGATATTCACCGAACAC 7260

D I Y L I V E E G F Y K R T L D I H R T

TAGGGTTGCTCTTGCACACTCAAGTCTCGATTAAGCAATTGCTTAAGCTGCCAGCTGAAT 7320

L G L L L H T Q V S I K Q L L K L P A E

GCTTTCATCCTAAACCAAAAGTAAACAGTGCCTTAATAAAACTTACCCGCCATACCACAG 7380

C F H P K P K V N S A L I K L T R H T T

ATGTTCCAGATAAATATTGGAAACTATATACGTACTTTGTTTCAAAATGGGTCAATCGAG 7440

D V P D K Y W K L Y T Y F V S K W V N R

***Hin*dII**

AATATC**GTCAAC**TGTTTACTAAAAATCAGTTTCATCAAGCAATGAAATACGCCAAAGTAA 7500

E Y R Q L F T K N Q F H Q A M K Y A K V

-35 **-10**

ACGATTTAAGTACCGTTACTTATGAGCAAGTATTGTCTATTTTTAATAGTTATCTATTAT 7560

N D L S T V T Y E Q V L S I F N S Y L L

Stop *erm2*(B) | **Start *orf5b***

TTAACGGGAGCAAATAATTCTATGAGTCGCTTTTTTAAATTTGGAAAGTTACACGTTACT 7620

RBS

F N G R K M S R F F K F G K L H V T

AAAGGGAATGGAGATAAATTATTAGATATACTACTGACAGCTTCCAAGAAGCTAAAGAGG 7680

K G N G D K L L D I L L T A S K K L K R

Stop *orf3b* |

TCCCTAGCGCCTACGGGGAATTTGTATCGATAAGGGGTACAAATTCCCACTAAGCGCTCG 7740

S L A P T G N L Y R

GGACCCCTTGTAGGAAAATGTCCTAAGTGTGGCAACAATATTGTATTAAAAAAATCGTTT 7800

TATGGTTGTTCAAATTATCCTGAATGTAAGTTTACTTTAGCTGAACATTTTAGAAAGAAA 7860

AAACTAACCAAAACGAATGTAAAAGAATTACTGGAGGGAAAAGAAACCCTGGTAAAAGGA 7920

| **Start DR Sequence**

ATCAAAACGAAAGATAAAAAGCCCTACAATGCCGTTGTAAAAATTGGGGAAAAGGGATAT 7980

ATTGATTTTATATCTTTCTCAAAATAAACATAAAAGCCCTTTAAAGAGGGCTTTTATATA 8040

TTAATCACAAATCACTTATCACAAATCACAAGTGATTTGTGATTGTTGATGATAAAATAA 8100

palA ***palB***

GAATAAGAAGAAATAGAAAGAAGTGAGTGATTGTGGGAAATTTAGGCGCACAAAAAGAAA 8160

AACGAAATGATACACCAATCAGTGCAAAAAAAGATATAATGGGAGATAAGACGGTTCGTG 8220

Stop DR Sequence|

TTCGTGCTGACTTGCACCATTCCCACCCTCTACAGATGGCAACGTCAATTATTCCTTTGA 8280

GGGGGCAACATTTGGCGAATCCCATATTGTAGTCACAGATAAAAACGGTCAGTTCTCTAC 8340

oriT (2)

TGCTTCAAGTTGGGCTTCCCGTAAGATAAATACA**AATACTGGAAAATCCAGTGAGGATGG** 8400

TGTATGGTTTGGTACTTCTGAACCAGACGACAGTAAAGGTGCATTACTTTATGATACCTG 8460

TGTGATTGAGGAATTAAAGTGTGATTCCAACGCCGGATTTAAGCTGATTCCAGCTTTTGA 8520

GGTGCTCGTATCCAGAAAAAAAGTGACCGTAGATTTAGGGACACTTACTGATGAATACGA 8580

AAAAGAAATCACAATCCATACCACAGCTACCGACAAGAAAACAGGCGAAAAAATGATTGT 8640

TGCCGGAAAAGACATCAAGATCGTGGACAAAGTCACACTTGATGGCTTGGAAACTGGCAG 8700

Start *orf13*

TCACAGAAGATTACACCGTGACCGTTCATGTGTATACAGATGGTGCAATGGGTCATTACC 8760

M C I Q M V Q W V I T

CAGAATCCTACCCTTGCTCCAGTGGTACAGAAATCAAAGTATGAACCGAAAGCACAGGCA 8820

Q N P T L A P V V Q K S K Y E P K A Q G

GCAGATGTCAGCGTCAGCTCCGATACGGTCAAAGACGCTACTACTTTCTTGGAAACATTC 8880

A D V S V S S D T V K D A T T F L E T F

TTTAAACTCTATCCGACAGCTACAGAAAAAGAACTTGCCTATTATGTCAAAGA CGTGTG 8940

F K L Y P T A T E K E L A Y Y V K D G V

CTTGCTCCTGTTTCCGGCGACTACGTATTTTCGGAACTGGTAAATCCTGTCTTTACCAAA 9000

L A P V S G D Y V F S E L V N P V F T K

GATGGCGATAATCTCAAGGTCAGTGTGTCAGTGAAATATCTGGATAACAAGTCGAAAATG 9060

D G D N L K V S V S V K Y L D N K S K M

ACACAAATCTCACAGTATGAGCTTGTGCTTCACAAGGACGATAATTGGAAAATTGTAGGA 9120

T Q I S Q Y E L V L H K D D N W K I V G

Stop *orf13*

TAAATATTACAGCAGACCAGCTATGGCAATTTCAAATACATATACTAAGTTTTAAAAATA 9180

-35 *Asp*718 -10 *Asp*718

TACATGTACCCTATTGACTTTTTGGGAA**GGTACC**TGTATATTATTTGTTAAA**GGTACC**TG 9240

Start *effR*

*Eco*RV

TTGAAAAGGAGGTCATACTATGAATAATATTTATTCT**GATATC**TATGAGAAACTATCAAC 9300

RBS

M N N I Y S D I Y E K L S T

CTTGCAATGGCTTATGAAACGCCATCAGATGTTTTGTCAGGCAGAGTCTGGTCCATTTGC 9360

L Q W L M K R H Q M F C Q A E S G P F A

TGATACATCACGAGGACAAGGAAGAATTCTTGCTATGTTAAAAATTCAGCCAGAAATTGC 9420

D T S R G Q G R I L A M L K I Q P E I A

```
AACAAAAGAGTTGGCATATTTATTGGGAATACGCCAACAATCCCTAAATGAGTTGCTTAA  9480
  T  K  E  L  A  Y  L  L  G  I  R  Q  Q  S  L  N  E  L  L  N
TAAAATGGAGAAAAATGGATATGTAGAACGAAAACCATCTGAAAATGATAAACGCGTTAT  9540
  K  M  E  K  N  G  Y  V  E  R  K  P  S  E  N  D  K  R  V  M
GATTGTTCATTTGACAGAAAAGGGAAAGCAAGTTCAGCAACCGAAAACAGACTATCAAAA  9600
  I  V  H  L  T  E  K  G  K  Q  V  Q  Q  P  K  T  D  Y  Q  N
TATCTTTAATTGTCTACTTCCAGAAGAATTGCTACAAATGTCGCAATATTTAGATCGTAT  9660
  I  F  N  C  L  L  P  E  E  L  L  Q  M  S  Q  Y  L  D  R  I
                HindIII
TATTGAAGCTTTTCAATTGCAAAATGGAAATGCTTTAGAAGAAAACAATATGATTGACTG  9720
  I  E  A  F  Q  L  Q  N  G  N  A  L  E  E  N  N  M  I  D  W
GATGGCTCAGGCAAGAGAACGTATGGGTGATGAGCATTTTGAACAGTTAATGTCTATGCG  9780
  M  A  Q  A  R  E  R  M  G  D  E  H  F  E  Q  L  M  S  M  R
TGAAAGAGCTTTTGGACATATGAGACCACCCAAAGATATACCGGGAGCTGAACGCTTTTC  9840
  E  R  A  F  G  H  M  R  P  P  K  D  I  P  G  A  E  R  F  S
TGAAAACTATAACGGATATGTTCCAGATAGAGACGGATTTCAGCCAAGAAACTTTAGACC  9900
  E  N  Y  N  G  Y  V  P  D  R  D  G  F  Q  P  R  N  F  R  P

EcoRV Stop effR |
AGATATCAAATAACAAAAGGCACTGTAAAATTACAGCAATAAGCCCTATGGAGATTTTTT  9960
  D  I  K
                                          Start effD
GAACTCCATAGGGCTTATTTTTTGAAAAGGAGACGAACAAAATGAACCAGAAAAATGAAA  10020
                          RBS
                                        M  N  Q  K  N  E
ATTACTGGCTTTTAGACGCACCTGTCACAAAAGCTATATGGCATATGGCGATTCCAATGA  10080
  N  Y  W  L  L  D  A  P  V  T  K  A  I  W  H  M  A  I  P  M
TGCTTGGAATGTCAATAAACATTATTTACAATATTACAGATACATTCTTCATAGGCAGAT  10140
  M  L  G  M  S  I  N  I  I  Y  N  I  T  D  T  F  F  I  G  R
TAAATGACACAGCGGCTCTTGCCGCAATCTCACTGCTGTTACCTTTCACGACCATTTTAA  10200
  L  N  D  T  A  A  L  A  A  I  S  L  L  L  P  F  T  T  I  L
TGGCAATTGGGAATTTGTTTGGAACAGGTGGAAGCACTTTGTTTTCACGACTGTTAGGAA  10260
  M  A  I  G  N  L  F  G  T  G  G  S  T  L  F  S  R  L  L  G
GTGAAAATACGGACAGGACAAAACAATGTTCAGCTACGACATTATGGCTATCTTTCCTGT  10320
  S  E  N  T  D  R  T  K  Q  C  S  A  T  T  L  W  L  S  F  L
```

TTGGACTATTGACTGCAATAATATCCATCATTTTCAGTAACTATATTATCCGACTTCTCG 10380

F G L L T A I I S I I F S N Y I I R L L

GTGCTGATAGTAATACTTTTGCCTATGTCAAACAGTATCTTATTTTTTATGGAATGGGTG 10440

G A D S N T F A Y V K Q Y L I F Y G M G

CTCCGTTTATTATTGCCAATTTTACGCTAGAACAGTTAATTAGAGGTGACGGTAAATCTG 10500

A P F I I A N F T L E Q L I R G D G K S

TAGAATCTATGATTGGAATGATGATAAGCATTGGTGCTAATATCATTCTTGACCCAATTC 10560

V E S M I G M M I S I G A N I I L D P I

TGATGTTTGGATTACAGCATGGTATTCGTGGGGCAGCCATTGCTACAGTAATCGGAAATG 10620

L M F G L Q L G I R G A A I A T V I G N

CTTTCGCTGTTATCTATTATATTGTCTGTATACAACGAGCAGACAATCAGTTATCTGCTC 10680

A F A V I Y Y I V C I Q R A D N Q L S A

TTCCAAAATATTTCAGGCTTGAAAAACAAATGCTAAAAGAAATTTTTTTAGTTGGATTAT 10740

L P K Y F R L E K Q M L K E I F L V G L

CTGCAATGTTGTTAGATATTCTTTTGATTGTTTCAAGCCTTATGTTTAATTACTATGCAC 10800

S A M L L D I L L I V S S L M F N Y Y A

TAAAATATGGAGATTATGTGCTTGCCGGATTTGGGATTTCTCAAAAACTTGTGCAGATTG 10860

L K Y G D Y V L A G F G I S Q K L V Q I

***Hin*dII**

TCGACCTAATCGGCATGGGACTTTACATGGGAGTAATTCCACTTATCGCTGTTGCCTATG 10920

V D L I G M G L Y M G V I P L I A V A Y

GTGCAAGAAATGAACTTCGTATGAAAGAAATCATTAAAAAGACTGCTCTCTATTTAGCAC 10980

G A R N E L R M K E I I K K T A L Y L A

TAGTAATTACATGTTTGTTTGCTATTCTATTTACATGCAGAAACTTTATTGTTCATTGTT 11040

L V I T C L F A I L F T C R N F I V H C

TTTCAAATGATTCAGATGTAATTCGTATAGGTGCGTACATCTTAACCGTTCAACTCTCTT 11100

F S N D S D V I R I G A Y I L T V Q L C

CTTCTTTCTTTGCGGCAGGAGCGGGTCTTCTAACGGGAATTTTTCAGTCAAAAGGAGAAG 11160

S S F F A A G A G L L T G I F Q S K G E

GAACTCCGGCTGTTGTTATGTCCGTTATGAGAGGTCTTATGCTCATACCAGCAATTATCT 11220

G T P A V V M S V M R G L M L I P A I I

```
TCGGCAATTATCTCTTTAAAATGAATGGAGTTATTTTTTCTCTGCTTGTAGCCGAAGCTA  11280

F  G  N  Y  L  F  K  M  N  G  V  I  F  S  L  L  V  A  E  A

                                  Stop effD |
TTTCATGCATTACAGGGATTGTATTATATAAATTGAAAAAGTAATGCTGAACTCCATCAG  11340

I  S  C  I  T  G  I  V  L  Y  K  L  K  K

      HindII
ACTTGTTAACAAAGTTAGTGATGGTTATAAAAATCATCAACAATAAATTAAGAGATTTCT  11400

ATAGAATGAGTATAGATTGGTAGGAGCTGGAAAAATATGATTGGGCTTAAAAAAAGAAAT  11460

ATAAAAAAATGTTGAAAACTGGTAGATGTATCATTAGCTGATATTCGGGCTAATATAGAA  11520

GCCACTATTGATGAAGAAATGAACAGTCCAGCCCCAGAGGTACAGGCAAATTTCAAAAAG  11580

TATTTTGGCAATAAACGTCCTACACCAGAAGAATATATTTACAAGATTACAAAAAAACAA  11640

AAGTTTGATTTACGACTATTTGTTACCTGCGGAGTATATCAACAAATAAGCGTTGAAGAA  11700

                                  | Stop orf9
TCTGCACAGAAAAACAGAGAAAATAAGAAGTGAATCACTTCTTATTCTCTCTGCTGTTCA  11760

                                        K  K  N  E  R  S  N  L

AGATACCATCAATCGTACCTTGAATGATTTTTAATTCGTCATCACTCAATAAATCAAGTG  11820

   I  G  D  I  T  G  Q  I  I  K  L  E  D  D  S  L  L  D  L  S

ACGAATCTATCTGACGGCGAACTGTGCTTTTTTCCACATTCTTCGCTGGATAGAAATACT  11880

   S  D  I  Q  R  R  V  T  S  K  E  V  N  K  A  P  Y  F  Y  E

      HindIII
CGTCAACAGATATATTGAACATGGTAACTAAATCATGGAATAAATGAAAGCTAGGGTGTT  11940

   D  V  S  I  N  E  M  T  V  L  D  H  F  L  H  F  S  P  H  K

TCCCGATATTTTCAATATCTGCGATATGACGTTCTCCGTAAAAGACTTTATCGCCTAAAT  12000

   G  I  N  E  I  D  A  I  H  R  E  G  Y  F  V  K  D  G  L  D

CATTTC:      ACCCTGCTTTCTCTCTGGCTTCTCTGATTGCCAGTCCTAAAGGACGAA  12060

   N  F  S  F  G  A  K  E  R  A  E  R  I  A  L  G  L  P  R  F

                                  Start orf9 <-|          -10
AATCAAAGCTGTGCGTATCTTTCTTTTTTCTCATTTTTAATCACCTAGCATGAGTTTATA  12120

   D  F  S  H  T  D  K  K  K  R  M

                              -35
CTATTTTCACAGGTTTGCTCTGCTCAAGGAGCAATTTGTGATTCTAAGTAAAACAGGCAG  12180

CCGGCACCGACAACCGAATTTATTCCCTTTTTTTTTAGTTTATCATCAGCCCTTGAGCATG  12240

AATTAAAAACTCATATATCAGATGTTTTAGATACAGTCTATTTGTTATTAGTGCAGATAG  12300

GCTGTTTTCATATTTTCAAAAAGTTTTTCAATTCTGCAACAAATCACACCCTGCTGTACA  12360
```

Start *orf7*

GATATGGTGCGGAAAGAGGGATACATACTTTTCACTTCATAGCGAAAGGGGGTGTGATTT 12420

RBS

TGAAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGTAAGCTCAAAAAGG 12480

TTGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTAGCCAGACGACAGGCAAAGGAAG 12540

TACCCTTTTGTGAGCTTCCAGAAATTGTTATTGAGAAATTGATTGTCTGGGATGATTACG 12600

AAAGTGACTATACGACATTCGATGTGTGCAGTATGGAAATCCGTGTGCTTGATGAAGAAC 12660

TTGAAAAATACAGGATATATCGGCATTTACCAGCTTATGAAAAGATAATCAGAAATTTAG 12720

Right End Tn*5398* |

TGTATTTTTATAATAAAAATATAATGCTTGTATACAAAAATATTAAAGATATTTTAGTAA 12780

Stop *orf7* |

GTTTTGTATATAAGCAAACATGTATTTTTTAATATAGTTATCTAAATATTATTTTATATA 12840

CAATAAAATATGACTCCTAAAAATAAAAAATATATCATAAATATAAATAAGTAGATATAG 12900

-10

GTTTTAATTTGATTTATAAACCTATTATAAAATAAAATTCATGATTTTTATAAGCAATAA 12960

Start *ispD*

AATTTTTAGGAGGATATATGATGAATAAAAGAATGAAACTAATTCCGTATGAAATAAATG 13020

RBS

M N K R M K L I P Y E I N

AAAATCTAAGAGGTGCAAAAAATAAATTCCCATATGGAATAAAACAAATGAATGCTAGGG 13080

E N L R G A K N K F P Y G I K Q M N A R

GAATGTGGGATGAAGGTTATACTGGTAAAAATATTGTAGTTGGTATAATAGATACAGGTT 13140

G M W D E G Y T G K N I V V G I I D T G

GTGATATATCTCATCCTCTTTTAAAAGGAAAAATAATTGGTGGTGCAAATTTTAGTGATG 13200

C D I S H P L L K G K I I G G A N F S D

ACAGTAATGGAAATAAAAATATATATGAGGATTTTAATGGTCATGGAACTCATGTGGCGG 13260

D S N G N K N I Y E D F N G H G T H V A

GTATTATAGCTGCATCTAATTATAATAATGAAGTTATGGGAGTAGCTCCAGATTGTAAAT 13320

G I I A A S N Y N N E V M G V A P D C K

```
TATTAATAGCAAAAGCATTAAATAAAGATGGTACCGGAACATATCAAAGTATAATTAATG  13380
L  L  I  A  K  A  L  N  K  D  G  T  G  T  Y  Q  S  I  I  N
CTATTAACTTTGCTGTAAATAACAAGGTTGATATTATATCTATGTCTCTTGGGGGAAACA  13440
A  I  N  F  A  V  N  N  K  V  D  I  I  S  M  S  L  G  G  N
AAGATGATAAGAATTTAAAAAATGCTGTCATGCAAGCAGTAAAAAATAATATTTCTGTAG  13500
K  D  D  K  N  L  K  N  A  V  M  Q  A  V  K  N  N  I  S  V
TGTGTGCAGCAGGTAATAATGGAGATGGTGATTCTAGTACAAGTGAGTATAGTTATCCAG  13560
V  C  A  A  G  N  N  G  D  G  D  S  S  T  S  E  Y  S  Y  P
CCAGTTATGCTGAGGTAATAGAAGTAGGTGCAATAAATGAAAACTATTTGGTTGAAAAGT  13620
A  S  Y  A  E  V  I  E  V  G  A  I  N  E  N  Y  L  V  E  K
TTAGTAATTCAAATACTACAATAGATTTGGTGGCTCCAGCAAGAAATATTATATCGACTT  13680
F  S  N  S  N  T  T  I  D  L  V  A  P  G  R  N  I  I  S  T
ATATGGATAATAAACTTGCTATTATGAGTGGTACTAGTATGAGTGCACCATACGTATCAG  13740
Y  M  D  N  K  L  A  I  M  S  G  T  S  M  S  A  P  Y  V  S
GCTCATTAGCACTAATTAAAGAATGGGCAAGAGAGGAGTTTGAAAGAGATTTAGATGAAG  13800
G  S  L  A  L  I  K  E  W  A  R  E  E  F  E  R  D  L  D  E
CTGAACTGTATGCACAATTAATAAAATGTACGAGAGCGCTTGGAATACCTAGAACGGAAC  13860
A  E  L  Y  A  Q  L  I  K  C  T  R  A  L  G  I  P  R  T  E
                                                  Stop ispD
AAGGAAATGGATACTTATATTTAAATCTTTATAAATACAAGAATAATAGCAAAAGATAAT  13920
Q  G  N  G  Y  L  Y  L  N  L  Y  K  Y  K  N  N  S  K  R
TTTTGATTTGATGATAAAATAGCTATATTATATAGAGTCGAGACAAATAATAAAAATTAC  13980
TTAGGTGTAAGATTTTACATCAATGTATAAAGGTATTAAGTGATAAAATTTATAAACATA  14040
TTAGCTAGTTAGAATTGAAAAATAAATATACGATATTATAGATAGCACATCTGGAAAAGG  14100
TGTGTTTTTCTATGTACTTATACATATAAAAGATTTATAAGAGATGCAAAAGTACTATAA  14160
GTTAGAAATTTTTCTATGAAAGATTAGAAATACGATACAGTGTTTTGATATTATAAAATG  14220
AATATAAAGTATAAATGATAAATGTTATCAATAAGTATTGAATTTCAAAATCAATTGTGA  14280
TATTCTATATTTAGAAAAATAAAGAGATTTCAGATAAAGAAATCTTTTGTTTTTTAGATG  14340
                                                  Start flxD
                -10                                          
TTAGTTAGCTATAACTAACTATTGATTAATTAAAAAATACTACAGGAGGTAAAAATGAGT  14400
                                          RBS
                                                  M  S
```

```
AAAATATATATTGTTTATTGGAGTGGAACAGGAAATACTGAAAAAATGGCAAATTTTGTG  14460
 K  I  Y  I  V  Y  W  S  G  T  G  N  T  E  K  M  A  N  F  V
GCTGAAGGTGTGAAGTTAAAAGGTAAGACACCAGAAGTTTTAGATGTGAGCTTACTGAAA  14520
 A  E  G  V  K  L  K  G  K  T  P  E  V  L  D  V  S  L  L  K
CCAAGTGATTTAAAAGAAGAAGATAAATTTGCATTAGGTTGCCCATCTATGGGAGCAGAG  14580
 P  S  D  L  K  E  E  D  K  F  A  L  G  C  P  S  M  G  A  E
CAACTAGAAGAGGGGGATATGGAGCCATTTGTTTCAGAATTAGAATCTATGGTATCAGGT  14640
 Q  L  E  E  G  D  M  E  P  F  V  S  E  L  E  S  M  V  S  G
AAACAGATTGGATTATTTGGTTCATATGGATGGGGAAATTGTGAATGGATGAGAGATTGG  14700
 K  Q  I  G  L  F  G  S  Y  G  W  G  N  C  E  W  M  R  D  W
GAAGAACGTATGCAAAATGCTGGTGCTACAATTATTGGTGGAGAAGGAATTACAGCAATG  14760
 E  E  R  M  Q  N  A  G  A  T  I  I  G  G  E  G  I  T  A  M
GAAGACCCAAATGAAGAAGCAAAAGATGAGTGTATAGAATTAGGCAAAACGTTAGCTGAA  14820
 E  D  P  N  E  E  A  K  D  E  C  I  E  L  G  K  T  L  A  E
Stop flxD|
TAAATTTGTATATTATAAAAATAGTATAAATAGCAACTAATGATGATAGCAGTATATAAT  14880
AATAAGAGAGACTAATATTATGATACATAAGAAAATATCCTTAATAGAGATAGATAAAAT  14940
ATTAGTCTCTTTTAATATAAAATTAGATTTATAATTTAATGTTGTTTTTGAA          14992
```

Table 4.1 : Features of the ORFs detected in pJIR1594 and pJIR1790.

ORF	Promoter Sequences Identified			Start Codon	Stop Codon	Nucleotide Region	Size of ORF (nucleotides)	Size of Predicted Protein (amino acids)
	-10	-35	RBS					
ilvD	N	N	N	None	TAG	1-1040	>1035	>345
hydR	N	N	Y	ATG	TAG	1384-2013	630	209
hydD	N	N	Y	ATG	TAA	2039-2851	813	270
elpD	Y	Y	Y	ATG	TAA	4213-4210	96	31
erm1(B)	N	N	N	ATG	TAA	3433-5167	738	245
orf3a	Y	Y	Y	ATG	TAA	5175-5306	132	43
orf298	N	N	N	ATG	TAG	5703-6599	897	298
erm2(B)	N	N	N	ATG	TAA	6839-7574	738	245
orf3b	Y	Y	Y	ATG	TAA	7622-7703	132	43
orf13	N	N	N	ATG	TAA	8728-9123	395	131
effR	Y	Y	Y	ATG	TAA	9260-9913	653	217
effD	N	N	Y	ATG	TAA	10002-11324	1322	440
orf9	Y	Y	N	ATG	TGA	11735-12094	360	119
orf7	N	N	Y	GTG	TAA	12414-12812	398	132
ispD	Y	N	Y	ATG	TAA	12981-13919	938	312
flxD	Y	N	Y	ATG	TAA	14395-14823	428	142

Sequence analysis of pJIR1594 and pJIR1790

Restriction endonuclease analysis of pJIR1594 and pJIR1790 revealed that the cloned fragments had common restriction fragments, which encompassed a region commencing at the *Xba*I site at the beginning of the pJIR1790 insert and at nucleotide position 1507 in the pJIR1594 insert, extending downstream to the *Hin*dIII site at nucleotide 8159 in the pJIR1790 insert and nucleotide 9666 in the pJIR1594 insert (Figure 4.1). Consequently, pJIR1594 was sequenced on both strands, firstly using the oligonucleotide primers UP and RP and then using a primer walking approach (Figure 4.1). The insert in plasmid pJIR1790 was sequenced on both strands, using a primer walking approach, across a region that extended from the *Asp*718 sites common to both plasmid inserts to approximately 5.8 kb downstream (Figure 4.1).

Analysis of the sequence using the BLASTX algorithm (Altschul *et al.*, 1997) revealed the presence of fifteen complete and one incomplete ORFs. The deduced amino acid sequence of each of these ORFs was analyzed using the BLASTP algorithm (Altschul *et al.*, 1997) to determine their similarity to known proteins in the databases.

a) **ilvD**

The first ORF detected in pJIR1594 was incomplete (nucleotides 1 to 1040, Figure 4.2). No start codon, RBS, or promoter elements were found upstream of this ORF (Table 4.1). BLASTP analysis of the amino acid sequence showed homology to IlvD proteins from several organisms. IlvD is a dihydroxy acid dehydratase, that catalyzes the conversion of dihydroxyacids to branched-chain keto acids (Fink,

Figure 4.1 : Sequence analysis of pJIR1594 and pJIR1790.

A 14992 bp region from the recombinant plasmids was sequenced using the oligonucleotide primers shown. The extent of sequence obtained from each primer is indicated by the arrow below the number of the oligonucleotide primer. The location of the ORFs are shown as colored block arrows above the scale bar. The extent of the sequenced DNA in each recombinant plasmid, and the restriction enzyme profile of each DNA insert, is show below the scale bar. The dashed arrow in the pJIR1790 DNA restriction profile indicates DNA that was not sequenced.

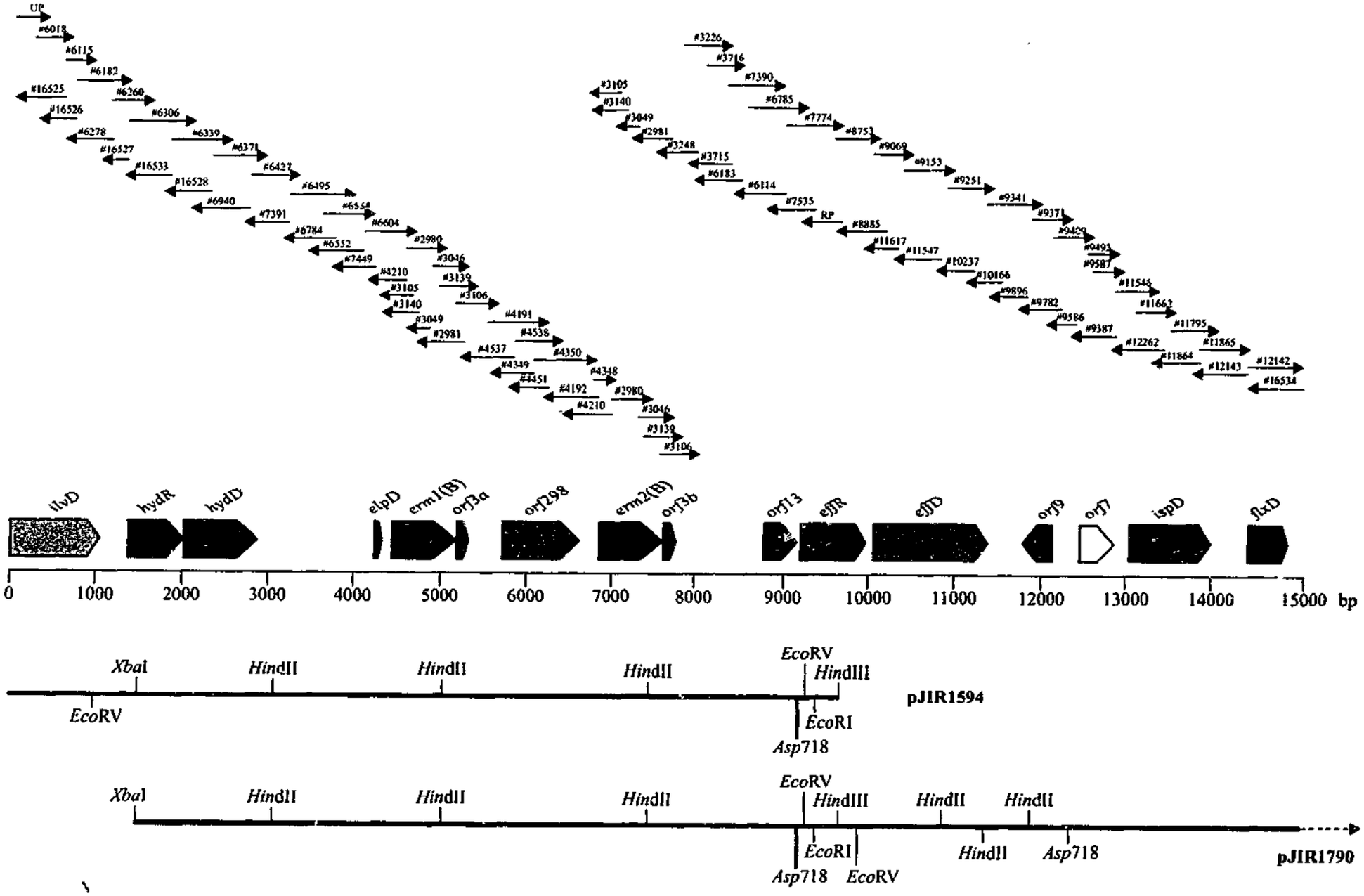

UP
#6018
#6115
#6182
#16525
#6260
#16526
#6306
#6278
#6339
#16527
#6371
#16533
#6427
#16528
#6495
#6940
#6554
#7391
#6604
#6784
#2980
#6552
#3046
#7449
#3139
#4210
#3106
#3105
#4191
#3140
#4538
#3049
#2981
#4350
#4537
#4349
#4348
#4451
#4192
#2980
#4210
#3046
#3139
#3106
#3226
#3716
#7390
#3105
#6785
#3140
#7774
#3049
#8753
#2981
#9069
#3248
#9153
#3715
#9251
#6183
#9341
#6114
#9371
#7535
#9429
RP
#9493
#8885
#9587
#11617
#11546
#11547
#11662
#10237
#11795
#10166
#11865
#9896
#12262
#9782
#11864
#9586
#12143
#12142
#9387
#16534
ilvD
hydR
hydD
elpD
erm1(B)
orf3a
orf298
erm2(B)
orf3b
orf13
ejfR
ejfD
orf9
orf7
ispD
fldD
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000 bp
XbaI
HindII
HindII
HindII
EcoRV
HindIII
pJIR1594
EcoRV
EcoRI
Asp718
XbaI
HindII
HindII
HindII
EcoRV
HindIII
HindII
HindII
EcoRI
HindII
Asp718
pJIR1790
Asp718
EcoRV

Figure 4.2 : Nucleotide sequence of Tn*5398* and its flanking regions.

Nucleotide positions are stated at the end of each line of sequence. The start and stop codons of each ORF are indicated in bold type face above the nucleotide sequence. The amino acid sequence of each predicted ORF is given below the appropriate nucleotide sequence. Potential RBS, -10 and -35 promoter sequences are underlined in bold. The potential right and left hand ends of Tn*5398* are indicated in bold above the nucleotide sequence. The beginning and end of regions encompassed by DR sequences are shown in bold above the nucleotide sequence. Palindromic sequences *palA* and *palB* are indicated by bold arrows below the nucleotide sequence. Potential *oriT* sites are shown as bold type face nucleotides within the sequence. The site at which a deletion event has removed the promoter sequences upstream of *erm2*(B) is marked by a black triangle (▲) beneath the nucleotide sequence. The GenBank accession number for the the Tn*5398* element and flanking sequence is AF109075.

```
TAGGCATGGGGATACCATATAATGGTACTGCTGCGTCACATTCTGGAGAAAGAAAAAGGA   60
   G  M  G  I  P  Y  N  G  T  A  A  S  H  S  G  E  R  K  R
TAGCAAAATATGCAGGTATGTATGTTATGGAGTTACTTAAGAACGACATAAAACCTAGAG  120
I  A  K  Y  A  G  M  Y  V  M  E  L  I  K  N  D  I  K  P  R
ATATTTTAACAATAGATGCTTTTAAAAATGCTATAGCTGTGGATATGGCAATGGCTGGTT  180
D  I  L  T  I  D  A  F  K  N  A  I  A  V  D  M  A  M  A  G
CTACAAATACAGTACTTCACTTACCTGCAATAGCTTATGAATCAGGAATAGAGCTTAACT  240
S  T  N  T  V  L  H  L  P  A  I  A  Y  E  S  G  I  E  L  N
TAGATTTTTTTGATGAAATAAGTGAAAAAACTCCTTGTTTAACAAAATTAAGTCCAAGTG  300
L  D  F  F  D  E  I  S  E  K  T  P  C  L  T  K  L  S  P  S
GAAAACATCATATTGAAGATTTACATATGGCAGGAGGAATACCAGCTATAATGAACGAGC  360
G  K  H  H  I  E  D  L  H  M  A  G  G  I  P  A  I  M  N  E
TTTCAAAGATAAATGGAATAAATTTAGATTGCAAAACCGTAACAGGCAAGACTATAAGGG  420
L  S  K  I  N  G  I  N  L  D  C  K  T  V  T  G  K  T  I  R
AAAATATAAGAAATTGTGAAATAGAAAATGAAGAAGTAATACATACATTAAAGAATCCAT  480
E  N  I  R  N  C  E  I  E  N  E  E  V  I  H  T  L  K  N  P
ATAGTAACCAAGGTGGGCTTGCAATATTGAAAGGAAATCTTGCTCTAAATGGAGCTGTTG  540
Y  S  N  Q  G  G  L  A  I  L  K  G  N  L  A  L  N  G  A  V
TAAAAAATCAGCAGTTGCAGAAGAATGTTAGTTCCATGAAGGACCCAGCAAGAATTTTTA  600
V  K  N  Q  Q  L  Q  K  N  V  S  S  M  K  D  P  A  R  I  F
ATTCAGAAGAAGAAGCTGTAAATGCTATTTTTGGTAAAAAAATAAACAAAGGTGATGTTA  660
N  S  E  E  E  A  V  N  A  I  F  G  K  K  I  N  K  G  D  V
TAGTTATAAGATATGAAGGTCCAAAGGGTGGTCCAGGAATGAAAGAAATGCTATCTCCTA  720
I  V  I  R  Y  E  G  P  K  G  G  P  G  M  K  E  M  L  S  P
CATCAGCAGTTGCAGGAATGGGACTTGATAAGCATGTAGCACTTCTTACTGATGGTCGTT  780
T  S  A  V  A  G  M  G  L  D  K  H  V  A  L  L  T  D  G  R
TTTCAGGGGCAACTACAGGAGCATCTATAGGCCATATTTCTCCAGAAGCTATGGAAGGTG  840
F  S  G  A  T  T  G  A  S  I  G  H  I  S  P  E  A  M  E  G
GTTTAATCGGACTAGTTGAAGAAGGAGATATAATTTCTATAAACATACCAGACAAAAAAT  900
G  L  I  G  L  V  E  E  G  D  I  I  S  I  N  I  P  D  K  K
```

```
TAGAGTTAAAAGTAGATGAAGTTGAAATAGAAAATAGAAAATTAAAATTTAAACCTTTAG  960

L  E  L  K  V  D  E  V  E  I  E  N  R  K  L  K  F  K  P  L

AACCGAAAATAAAGCATGGATACTTAAGTAGATATGCTAAATTGGTAACATCAGCAAATA  1020

E  E  L  K  V  D  E  V  E  I  E  N  R  K  L  K  F  K  P  L
                Stop ilvD |
CAGGAGCAGTTTTAAAATAGTAGACTATATTTATGTAAATTAGAAAAAGTTAAGAAAATA  1080

T  G  A  V  L  K
          EcoRV
TTAATTGGATATCTCAGGTGTTATATCACTAATGAGGTATCCTTTTTATTTTGTATGAAA  1140

AACACAGCAAAATCTTGATTATCTTTGAAAACAATACATCTCTATTATAAGGAATACGAG  1200

TACAGCATAGAGATGTATCTTTTAGACCTTGTTGAGGTAATACAACTTTTTAAAAGTTTG  1260

ATGTGGAATGCTTTTCGTTCCATCTGAACTATGAAGAAATTGAAAGTCAGTTAATTCTTG  1320
                                                      RBS
ACATAATAAATATTAAAATGTTTAATAAAAAACAAACAGACAGTCTGTTTTTAGGAGGGG  1380

     Start hydR
AATATGAATAGAGAAGAAAAAAGTAAAAATAGTAAAGAAAAAATTATCCAATCAGCATTT  1440

    M  N  R  E  E  K  S  K  N  S  K  E  K  I  I  Q  S  A  F

TCACTATTTTCGTCTAAAGGATATGATTCAACATCTACACAAGATATTATCAATTTATCT  1500

 S  L  F  S  S  K  G  Y  D  S  T  S  T  Q  D  I  I  N  L  S
          XbaI
GGTCTATCTAGAGGTGCAATGTATCATCACTTTAAAACTAAAGAAGATATACTGAGAAGT  1560

 G  L  S  R  G  A  M  Y  H  H  F  K  T  K  E  D  I  L  R  S

GTCACAAAAGAACTTTACTCACAAATGAATAATTTTTTAGAGTATCTTGTTGCTGATGAC  1620

 V  T  K  E  L  Y  S  Q  M  N  N  F  L  E  Y  L  V  A  D  D

ACCCTTACAGCAAATGAAAAAATAATAGAATTGGTTGTTCATAGTGCGAATGATTACACA  1680

 T  L  T  A  N  E  K  I  I  E  L  V  V  H  S  A  N  D  Y  T

CGTAGAAAAATGGTACATTGTAGCTGGTTAGAAAAAATCCCATTCGCTTTAATAGAGGAA  1740

 R  R  K  M  V  H  C  S  W  L  E  K  I  P  F  A  L  I  E  E

GTTCGTAATCTCAACAATGTAGTTGCACCCAATATTGCTAAGATAATTAAACAAGGTGTT  1800

 V  R  N  L  N  N  V  V  A  P  N  I  A  K  I  I  K  Q  G  V

GAAAATAAAGAATTTTCTTGTGAATATCCAGAGGAATTAGCTGAAATGCTCGTTTTTAGT  1860

 E  N  K  E  F  S  C  E  Y  P  E  E  L  A  E  M  L  V  F  S

ATTGATATTTTACTTGACCCTGTATTATTTAAACGTGAATACAGTGAAGTATGTAATAGG  1920

 I  D  I  L  L  D  P  V  L  F  K  R  E  Y  S  E  V  C  N  R
```

```
TTGGATTTTTATTATTCATGTTAAAAAAAATGGATATTCCTCTAATTGATGAGTACGGT  1980
 L  D  F  L  L  F  M  L  K  K  M  D  I  P  L  I  D  E  Y  G
                        Stop hydR |           RBS            |-> Start hydD
ATTCAAAAATTTAAAGATTTATTTAAACAATGATAATAGGAAAAGTGAGGAATAGACTAT  2040
 I  Q  K  F  K  D  L  F  K  Q                                M
GTTTTATAATGCAAAAAGCTGTCAAATGATGATAGGAAGTAATACAGTGAATTATATTGA  2100
  F  Y  N  A  K  S  C  Q  M  M  I  G  S  N  T  V  N  Y  I  E
ATTTGGTAGTGGTAAGAAGACACTAATTATTTTACCAGGATTAGGAGATGGTTTATTTCC  2160
  F  G  S  G  K  K  T  L  I  I  L  P  G  L  G  D  G  L  F  P
TTTACATGGAAAAATACAGGCGATTGCTTTTGCTTTTAGATATAAGCAATTTGCAAAAGA  2220
  L  H  G  K  I  Q  A  I  A  F  A  F  R  Y  K  Q  F  A  K  D
TTATAAGGTTTATGTATTTAGTAGAAAAAATCAAATTACAGAAAAATATTCCACAAGGGA  2280
  Y  K  V  Y  V  F  S  R  K  N  Q  I  T  E  K  Y  S  T  R  D
TATGGCTAAAGACCAAGCAGATATAATGAAAAAACTTGGAATTATGAAAGCAGAAGTAAT  2340
  M  A  K  D  Q  A  D  I  M  K  K  L  G  I  M  K  A  E  V  M
GGGTGTTTCTCAAGGTGGAATGATAGCTCAATATCTGGCAATTGATTATCCTGAGTTAGT  2400
  G  V  S  Q  G  G  M  I  A  Q  Y  L  A  I  D  Y  P  E  L  V
TGAGAAACTTGTGTTGGCAGTTACTTCTTCAAAGCAAAATGATACTATCCAAAATGTTAT  2460
  E  K  L  V  L  A  V  T  S  S  K  Q  N  D  T  I  Q  N  V  I
TTGTAGCTGGATTGATATGGCGAAGAAGCAGAATTACAATGATTTAATGATTGACACAGC  2520
  C  S  W  I  D  M  A  K  K  Q  N  Y  N  D  L  M  I  D  T  A
AAAGAAATCCTATTCTGAAAGATATTTAAAAAAATACCAATTATTTATTCCTTTTTTAGG  2580
  K  K  S  Y  S  E  R  Y  L  K  K  Y  Q  L  F  I  P  F  L  G
AAAAGTAGGTAAACCAAAGGATTTTAAACGATTTATTATTCAAGCGACCTCTTGCATAGA  2640
  K  V  G  K  P  K  D  F  K  R  F  I  I  Q  A  T  S  C  I  E
ACATAATGCCTTTTCCGAATTAAATAAAATAACATGTCCAACCTTAATTATAGGAGGTGC  2700
  H  N  A  F  S  E  L  N  K  I  T  C  P  T  L  I  I  G  G  A
TAATGATAAAATTGTTGGAAATAATGCTTCTTTTCATTTAGCAGAAAAGATTAAAAAAAG  2760
  N  D  K  I  V  G  N  N  A  S  F  H  L  A  E  K  I  K  K  S
CGAAATATTTATCTATGAGGGATTGGGACACGCTACATATGAAGAAGCACAGGATTTTAA  2820
  E  I  F  I  Y  E  G  L  G  H  A  T  Y  E  E  A  Q  D  F  N
```

Stop *hydD* |

TGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCTTTGTTAAAAATATACATATGAAG 2880

E R V L E F L N K

ATTGGAAATTTAATGTTAAAAATAGAAACATGAAAATATGCTTAACTGGTATTTTTACTA 2940

TTCATAACCAATTTTTAATACATTATCTACTATAAATACAAATATAGCTTCAATGTGATT 3000

ATATATTGTTGTATTGGTAAAGCACTTATACAAACAGAGGAATTTTGTAAATTCAGATTA 3060

Hin*dII** | **Left End of Tn*5398

TATCCACATTT**GTTAAC**TTATGAAAATATAATCAAAATTTTTATGAGCTTTTTATATAAA 3120

AAAACGCCCTAAAAATCTGATTATCCCCATAAACACTGTATCTACAAGCATATTCAATAG 3180

GAAATAAATCGTGATATTACTACGAATTTACTACTAATTTACTACTAATGAATGAGCCTT 3240

GATACGTCTTATTTCCCAGATATGCAAAGATATGGCATGGCACATCAGTAAAAATTGAAT 3300

ACTTATATAGACTATGGAACGTACACTTTTGGCGTTCCTTTTCTATTTCCAGACGTTCTT 3360

TTCAGAGCGTCTTTTTTTCATACCCAAAATCGAAAGGAGAAAGAGAAAATATGAATAAGC 3420

TAGTAAAGCGATTGCTGACAGGGACGCTCGCCATTGCAACCATTCTTACCGCATTGCCTG 3480

TGACGGTGGTTCATGCTTCTGGCAATTAATACTGGACAGAATCAGCAGAACGTGTCGGCT 3540

ACATTGAACATGTTATGAATGATGGTTCTATCAAATCCAAATTAAATGAGGGACACATGA 3600

AAGTTGAGGGCGAAACTGCCTATTGCGTGAACATCAATACAAATTTCAAAAATGGATATA 3660

AAACAAGGTATGACGCAAGCTCCCATATGAGTAGCGATCAGATTGCGGACATTGCTCTTT 3720

CCTTAGAGTACGTCAAGTAATATACTGCTTCTCATACAAACTTGAATTACAAGCAGGGTT 3780

ACTTATTGGAACAGTGTGTTGTCTGGCAGAGATTGAGTGAACAGCTCGGCTGGCAATGTG 3840

ATAACGTCAGAGCCTCCTATAATGAAATCTCACAGGCGGTACAGAATAAAGTTTACGCTG 3900

GTGCGAAAGCATTTGTGAAAGCAAATAAGGGGTGCTATGAATGTGGCGGTTACATCTACA 3960

CTGGCGAAGGACAGGACATTGGACAGTTCTGGGCGAGTTGAATGTAGGAAATGAAAAGGT 4020

| **Start DR sequence**

CAAAAAGACTTCTTCCAAATCATAAAAATCGAAACAGCAAAGAATGGCGGAAACGTAAAA 4080

Stop DR sequence| **-35**

GAAGTTATGGAAATAAGACTTAGAAGCAAACTTAAGAGTGTGTTGATAGTGC TTATCTT 4140

-10 **RBS**

AAAATTTTGTATAATAGGAATTGAAGTTAAATTAGATGCTAAAAATTTGTAATTAAGAAG 4200

Start *elpD*

GAGGGATTCGTCATGTTGGTATTCCAAATGCGTTATCAAATGCGTTATGTAGATAAAACA 4260

M L V F Q M R Y Q M R Y V D K T

Stop *elpD*

```
TCTACTGTTTTGAAACAGACTAAAAACAGTGATTACGCAGATAAATAAATACGTTAGATT   4320
 S  T  V  L  K  Q  T  K  N  S  D  Y  A  D  K

AATTCCTACCAGTGACTAATCTTATGACTTTTTAAACAGATAACTAAAATTACAAACAAA   4380
```

Start *erm1*(B)

```
TCGTTTAACTTCTGTATTTGTTTATAGATGTAATCACTTCAGGAGTGATTACATGAACAA   4440
                                                  M  N  K

AAATATAAAATATTCTCAAAACTTTTTAACGAGTGAAAAAGTACTCAACCAAATAATAAA   4500
  N  I  K  Y  S  Q  N  F  L  T  S  E  K  V  L  N  Q  I  I  K

ACAATTGAATTTAAAAGAAACCGATACCGTTTACGAAATTGGAACAGGTAAAGGGCATTT   4560
  Q  L  N  L  K  E  T  D  T  V  Y  E  I  G  T  G  K  G  H  L

AACGACGAAACTGGCTAAAATAAGTAAACAGGTAACGTCTATTGAATTAGACAGTCATCT   4620
  T  T  K  L  A  K  I  S  K  Q  V  T  S  I  E  L  D  S  H  L

ATTCAACTTATCGTCAGAAAAATTAAAACTGAATACTCGTGTCACTTTAATTCACCAAGA   4680
  F  N  L  S  S  E  K  L  K  L  N  T  R  V  T  L  I  H  Q  D

TATTCTACAGTTTCAATTCCCTAACAAACAGAGGTATAAAATTGTTGGGAGTATTCCTTA   4740
  I  L  Q  F  Q  F  P  N  K  Q  R  Y  K  I  V  G  S  I  P  Y

CCATTTAAGCACACAAATTATTAAAAAAGTGGTTTTTGAAAGCCGTGCGTCTGACATCTA   4800
  H  L  S  T  Q  I  I  K  K  V  V  F  E  S  R  A  S  D  I  Y

TCTGATTGTTGAAGAAGGATTCTACAAGCGTACCTTGGATATTCACCGAACACTAGGGTT   4860
  L  I  V  E  E  G  F  Y  K  R  T  L  D  I  H  R  T  L  G  L

GCTCTTGCACACTCAAGTCTCGATTAAGCAATTGCTTAAGCTGCCAGCTGAATGCTTTCA   4920
  L  L  H  T  Q  V  S  I  K  Q  L  L  K  L  P  A  E  C  F  H

TCCTAAACCAAAAGTAAACAGTGCCTTAATAAAACTTACCCGCCATACCACAGATGTTCC   4980
  P  K  P  K  V  N  S  A  L  I  K  L  T  R  H  T  T  D  V  P

AGATAAATATTGGAAACTATATACGTACTTTGTTTCAAAATGGGTCAATCGAGAATATCG   5040
  D  K  Y  W  K  L  Y  T  Y  F  V  S  K  W  V  N  R  E  Y  R
```

***Hin*dIII**

```
TCAACTGTTTACTAAAAATCAGTTTCATCAAGCAATGAAATACGCCAAAGTAAACGATTT   5100
  Q  L  F  T  K  N  Q  F  H  Q  A  M  K  Y  A  K  V  N  D  L

                                -35                -10
AAGTACCGTTACTTATGAGCAAGTATTGTCTATTTTTAATAGTTATCTATTATTTAACGG   5160
  S  T  V  T  Y  E  Q  V  L  S  I  F  N  S  Y  L  L  F  N  G
```

Stop *erm1*(B) | Start *orf3a*

GAGGAAATAATTCTATGAGTCGCTTTTTTAAATTTGGAAAGTTACACGTTACTAAAGGGA 5220

RBS

R K M S R F F K F G K L H V T K G

ATGGAGATAAATTATTAGATATACTACTGACAGCTTCCAAGAAGCTAAAGAGG[illegible] 5280

N G D K L L D I L L T A S K K L K R

Stop *orf3a* |

CGCCTACGGGGAATTTGTATCGATAAGGGGTACAAATTCCCACTAAGCGCTCGGG CC 5340

A P T G N L Y R

TTGTAGGAAAATGTCCTAAGTGTGGCAACAATATTGTATTAAAAAAATCGTTTTATGGTT 5400

GTTCAAATTATCCTGAATGTAAGTTTACTTTAGCTGAACATTTTAGAAAGAAAAAACTAA 5460

CCAAAACGAATGTAAAAGAATTACTGGAGGGAAAAGAAACCCTGGTAAAAGGAATCAAAA 5520

| Start DR sequence

ACAAAGAGAAAAAGCCCTACAATGCCGTTGTAAAAATTGGGGAAAAGGGATATATTGATT 5580

TTATATCTTTCTCAAAATAAATATAAAAGCCCTTTAAAGAGGGCTTTTATATATTAATCA 5640

CAAATCACTTATCACAAATCACAAGTGATTAATCACAAATCACTTGTGATTTGTGATTCT 5700

palA

Start *orf298*

TAATGATACAATATTACTATACAAAAAAAGAATGGGGCGTAGTTATGGAGAAGGAAAAAC 5760

M I Q Y Y Y T K K E W G V V M E K E K

TAAAAATACTTGAAGAATTAAGACGTATTTTAAACAATAAAAATGAAGCAATTATT**ATCT** 5820

L K I L E E L R R I L N N K N E A I I I

oriT (1)

TGAATAATTACTTTAAAGGTGGTGTTGGAAAGTCCAAATTATCGACTATGTTTGCTTACT 5880

L N N Y F K G G V G K S K L S T M F A Y

TGACAGACAAATTGAATTTAAAAGTTTTAATGATCGATAAGGACTTACAGGCAACATTGA 5940

L T D K L N L K V L M I D K D L Q A T L

CAAAAGACTTAGCAAAAACTTTTGAGGTAGAATTGCCACGTGTCAATTTTTATGAAGGCT 6000

T K D L A K T F E V E L P R V N F Y E G

TGAAAAATGGAAACTTGGCTTCTTCTATTATTCATTTGACTGATAATTTAGACTTGATCC 6060

L K N G N L A S S I I H L T D N L D L I

CTGGCACGTTTGATTTGATGTTACTGCCAAAATTAACTCGCTCATGGACGTTTGAAAATG 6120

P G T F D L M L L P K L T R S W T F E N

AAAGTAGATTGCTTGCTACTCTTTTAGCACCTTTAAAAAGTGACTATGATCTTATTATTA 6180

E S R L L A T L L A P L K S D Y D L I I

TTGATACTGTACCAACGCCAAGCGTTTATACAAATAATGCAATCGTGGCGAGTGATTACG 6240

I D T V P T P S V Y T N N A I V A S D Y

TTATGATCCCTTTACAAGCAGAAGAAGAAAGTACAAACAACATTCAAAACTATATTTCCT 6300

V M I P L Q A E E E S T N N I Q N Y I S

ATTTGATTGATTTGCAAGAACAGTTTAACCCTGGACTAGATATGATTGGTTTTGTTCCTT 6360

Y L I D L Q E Q F N P G L D M I G F V P

ATTTAGTTGATACGGACAGCGCAACGATAAAATCAAACCTGGAAGAACTGTACAAAGAAC 6420

Y L V D T D S A T I K S N L E E L Y K E

ATAAAGAGGATAATTTGGTTTTCCGAAATATTATCAAGCGAAGTAATAAAGTAAGTACTT 6480

H K E D N L V F R N I I K R S N K V S T

GGTCTAAAAATGGCATTACAGAACACAAAGGCTATGACAAAAAAGTTTTGTCTATGTATG 6540

W S K N G I T E H K G Y D K K V L S M Y

Stop *orf298*

AGAACGTATTTTTTGAAATGATTGAGCGAATCATTCAATTAGAAAATGAAAAAGAATAGA 6600

E N V F F E M I E R I I Q L E N E K E

ATCACAAATCACAAGTGATTAATCACAAATCACTTGTGATTTGTGATAGGTGATGATAAA 6660

palB

ATAAATAGTAAGAAGAAATAGAAAGAAGTGAGCGATCGTGGGAAATTTAGGCGCACAGAA 6720

AGCAAAACGAAATGATACGCCAATCAGCGCAAAAAAAGATATAATGGGGGATAAGACGGT 6780

Stop DR Sequence **Start *erm2*(B)**

TCGTGTTCGTGCCGACTTGCACCATATCATAAAAATCGAAACAGCTCAGGAGTGATTACA 6840

TGAACAAAAATATAAAATATTCTCAAAACTTTTTAACGAGTGAAAAAGTACTCAACCAAA 6900

M N K N I K Y S Q N F L T S E K V L N Q

TAATAAAACAATTGAATTTAAAAGAAACCGATACCGTTTACGAAATTGGAACAGGTAAAG 6960

I I K Q L N L K E T D T V Y E I G T G K

GGCATTTAACGACGAAACTGGCTAAAATAAGTAAACAGGTAACGTCTATTGAATTAGACA 7020

G H L T T K L A K I S K Q V T S I E L D

GTCATCTATTCAACTTATCGTCAGAAAAATTAAAACTGAATACTCGTGTCACTTTAATTC 7080

S H L F N L S S E K L K L N T R V T L I

ACCAAGATATTCTACAGTTTCAATTCCCTAACAAACAGAGGTATAAAATTGTTGGGAGTA 7140

H Q D I L Q F Q F P N K Q R Y K I V G S

TTCCTTACCATTTAAGCACACAAATTATTAAAAAAGTGGTTTTTGAAAGCCGTGCGTCTG 7200

I P Y H L S T Q I I K K V V F E S R A S

ACATCTATCTGATTGTTGAAGAAGGATTCTACAAGCGTACCTTGGATATTCACCGAACAC 7260

D I Y L I V E E G F Y K R T L D I H R T

TAGGGTTGCTCTTGCACACTCAAGTCTCGATTAAGCAATTGCTTAAGCTGCCAGCTGAAT 7320

L G L L L H T Q V S I K Q L L K L P A E

GCTTTCATCCTAAACCAAAAGTAAACAGTGCCTTAATAAAACTTACCCGCCATACCACAG 7380

C F H P K P K V N S A L I K L T R H T T

ATGTTCCAGATAAATATTGGAAACTATATACGTACTTTGTTTCAAAATGGGTCAATCGAG 7440

D V P D K Y W K L Y T Y F V S K W V N R

***Hin*dII**

AATATC**GTCAAC**TGTTTACTAAAAATCAGTTTCATCAAGCAATGAAATACGCCAAAGTAA 7500

E Y R Q L F T K N Q F H Q A M K Y A K V

-35 **-10**

ACGATTTAAGTACCGTTACTTATGAGCAAGTATTGTCTATTTTTAATAGTTATCTATTAT 7560

N D L S T V T Y E Q V L S I F N S Y L L

Stop *erm2*(B) | **Start *orf3b***

TTAACGGGAGGAAATAATTCTATGAGTCGCTTTTTTAAATTTGGAAAGTTACACGTTACT 7620

RBS

F N G R K M S R F F K F G K L H V T

AAAGGGAATGGAGATAAATTATTAGATATACTACTGACAGCTTCCAAGAAGCTAAAGAGG 7680

K G N G D K L L D I L L T A S K K L K R

Stop *orf3b* |

TCCCTAGCGCCTACGGGGAATTTGTATCGATAAGGGGTACAAATTCCCACTAAGCGCTCG 7740

S L A P T G N L Y R

GGACCCCTTGTAGGAAAATGTCCTAAGTGTGGCAACAATATTGTATTAAAAAAATCGTTT 7800

TATGGTTGTTCAAATTATCCTGAATGTAAGTTTACTTTAGCTGAACATTTTAGAAAGAAA 7860

AAACTAACCAAAACGAATGTAAAAGAATTACTGGAGGGAAAAGAAACCCTGGTAAAAGGA 7920

| **Start DR Sequence**

ATCAAAACGAAAGATAAAAAGCCCTACAATGCCGTTGTAAAAATTGGGGAAAAGGGATAT 7980

ATTGATTTTATATCTTTCTCAAAATAAACATAAAAGCCCTTTAAAGAGGGCTTTTATATA 8040

TTAATCACAAATCACTTATCACAAATCACAAGTGATTTGTGATTGTTGATGATAAAATAA 8100

→ → * → → ← ←

* → |

palA ***palB***

GAATAAGAAGAAATAGAAAGAAGTGAGTGATTGTGGGAAATTTAGGCGCACAAAAAGAAA 8160

AACGAAATGATACACCAATCAGTGCAAAAAAAGATATAATGGGAGATAAGACGGTTCGTG 8220

Stop DR Sequence|

TTCGTGCTGACTTGCACCATTCCCACCCTCTACAGATGGCAACGTCAATTATTCCTTTGA 8280

GGGGGCAACATTTGGCGAATCCCATATTGTAGTCACAGATAAAAACGGTCAGTTCTCTAC 8340

oriT (2)

TGCTTCAAGTTGGGCTTCCCGTAAGATAAATACA**AATACTGGAAAATCCAGTGAGGATGG** 8400

TGTATGGTTTGGTACTTCTGAACCAGACGACAGTAAAGGTGCATTACTTTATGATACCTG 8460

TGTGATTGAGGAATTAAAGTGTGATTCCAACGCCGGATTTAAGCTGATTCCAGCTTTTGA 8520

GGTGGTCGTATCCAGAAAAAAAGTGACCGTAGATTTAGGGACACTTACTGATGAATACGA 8580

AAAAGAAATCACAATCCATACCACAGCTACCGACAAGAAAACAGGCGAAAAAATGATTGT 8640

TGCCGGAAAAGACATCAAGATCGTGGACAAAGTCACACTTGATGGCTTGGAAACTGGCAG 8700

Start *orf13*

TCACAGAAGATTACACCGTGACCGTTCATGTGTATACAGATGGTGCAATGGGTCATTACC 8760

M C I Q M V Q W V I T

CAGAATCCTACCCTTGCTCCAGTGGTACAGAAATCAAAGTATGAACCGAAAGCACAGGGA 8820

Q N P T L A P V V Q K S K Y E P K A Q G

GCAGATGTCAGCGTCAGCTCCGATACGGTCAAAGACGCTACTACTTTCTTGGAAACATTC 8880

A D V S V S S D T V K D A T T F L E T F

TTTAAACTCTATCCGACAGCTACAGAAAAAGAACTTGCCTATTATGTCAAAGACGGTGTG 8940

F K L Y P T A T E K E L A Y Y V K D G V

CTTGCTCCTGTTTCCGGCGACTACGTATTTTCGGAACTGGTAAATCCTGTCTTTACCAAA 9000

L A P V S G D Y V F S E L V N P V F T K

GATGGCGATAATCTCAAGGTCAGTGTGTCAGTGAAATATCTGGATAACAAGTCGAAAATG 9060

D G D N L K V S V S V K Y L D N K S K M

ACACAAATCTCACAGTATGAGCTTGTGCTTCACAAGGACGATAATTGGAAAATTGTAGGA 9120

T Q I S Q Y E L V L H K D D N W K I V G

Stop *orf13*

TAAATATTACAGCAGACCAGCTATGGCAATTTCAAATACATATACTAAGTTTTAAAAATA 9180

-35 *Asp*718 -10 *Asp*718

TACATGTACCCTATTGACTTTTTGGGAA**GGTACC**TGTATATTATTTGTTAAA**GGTACC**TG 9240

Start *effR*

*Eco*RV

TTGAAAAGGAGGTCATACTATGAATAATATTTATTCT**GATATC**TATGAGAAACTATCAAC 9300

RBS

M N N I Y S D I Y E K L S T

CTTGCAATGGCTTATGAAACGCCATCAGATGTTTTGTCAGGCAGAGTCTGGTCCATTTGC 9360

L Q W L M K R H Q M F C Q A E S G P F A

TGATACATCACGAGGACAAGGAAGAATTCTTGCTATGTTAAAAATTCAGCCAGAAATTGC 9420

D T S R G Q G R I L A M L K I Q P E I A

```
AACAAAAGAGTTGGCATATTTATTGGGAATACGCCAACAATCCCTAAATGAGTTGCTTAA  9480
  T  K  E  L  A  Y  L  L  G  I  R  Q  Q  S  L  N  E  L  L  N
TAAAATGGAGAAAAATGGATATGTAGAACGAAAACCATCTGAAAATGATAAACGCGTTAT  9540
  K  M  E  K  N  G  Y  V  E  R  K  P  S  E  N  D  K  R  V  M
GATTGTTCATTTGACAGAAAAGGGAAAGCAAGTTCAGCAACCGAAAACAGACTATCAAAA  9600
  I  V  H  L  T  E  K  G  K  Q  V  Q  Q  P  K  T  D  Y  Q  N
TATCTTTAATTGTCTACTTCCAGAAGAATTGCTACAAATGTCGCAATATTTAGATCGTAT  9660
  I  F  N  C  L  L  P  E  E  L  L  Q  M  S  Q  Y  L  D  R  I
                HindIII
TATTGAAGCTTTTCAATTGCAAAATGGAAATGCTTTAGAAGAAAACAATATGATTGACTG  9720
  I  E  A  F  Q  L  Q  N  G  N  A  L  E  E  N  N  M  I  D  W
GATGGCTCAGGCAAGAGAACGTATGGGTGATGAGCATTTTGAACAGTTAATGTCTATGCG  9780
  M  A  Q  A  R  E  R  M  G  D  E  H  F  E  Q  L  M  S  M  R
TGAAAGAGCTTTTGGACATATGAGACCACCCAAAGATATACCGGGAGCTGAACGCTTTTC  9840
  E  R  A  F  G  H  M  R  P  P  K  D  I  P  G  A  E  R  F  S
TGAAAACTATAACGGATATGTTCCAGATAGAGACGGATTTCAGCCAAGAAACTTTAGACC  9900
  E  N  Y  N  G  Y  V  P  D  R  D  G  F  Q  P  R  N  F  R  P

EcoRV Stop effR
AGATATCAAATAACAAAAGGCACTGTAAAATTACAGCAATAAGCCCTATGGAGATTTTTT  9960
  D  I  K
                                          Start effD
GAACTCCATAGGGCTTATTTTTTGAAAAGGAGACGAACAAAATGAACCAGAAAAATGAAA  10020
                          RBS
                                        M  N  Q  K  N  E
ATTACTGGCTTTTAGACGCACCTGTCACAAAAGCTATATGGCATATGGCGATTCCAATGA  10080
  N  Y  W  L  L  D  A  P  V  I  K  A  I  W  H  M  A  I  P  M
TGCTTGGAATGTCAATAAACATTATTTACAATATTACAGATACATTCTTCATAGGCAGAT  10140
  M  L  G  M  S  I  N  I  I  Y  N  I  T  D  T  F  F  I  G  R
TAAATGACACAGCGGCTCTTGCCGCAATCTCACTGCTGTTACCTTTCACGACCATTTTAA  10200
  L  N  D  T  A  A  L  A  A  I  S  L  L  L  P  F  T  T  I  L
TGGCAATTGGGAATTTGTTTGGAACAGGTGGAAGCACTTTGTTTTCACGACTGTTAGGAA  10260
  M  A  I  G  N  L  F  G  T  G  G  S  T  L  F  S  R  L  L  G
GTGAAAATACGGACAGGACAAAACAATGTTCAGCTACGACATTATGGCTATCTTTCCTGT  10320
  S  E  N  T  D  R  T  K  Q  C  S  A  T  T  L  W  L  S  F  L
```

```
TTGGACTATTGACTGCAATAATATCCATCATTTTCAGTAACTATATTATCCGACTTCTCG  10380
F  G  L  L  T  A  I  I  S  I  I  F  S  N  Y  I  I  R  L  L
GTGCTGATAGTAATACTTTTGCCTATGTCAAACAGTATCTTATTTTTTATGGAATGGGTG  10440
G  A  D  S  N  T  F  A  Y  V  K  Q  Y  L  I  F  Y  G  M  G
CTCCGTTTATTATTGCCAATTTTACGCTAGAACAGTTAATTAGAGGTGACGGTAAATCTG  10500
A  P  F  I  I  A  N  F  T  L  E  Q  L  I  R  G  D  G  K  S
TAGAATCTATGATTGGAATGATGATAAGCATTGGTGCTAATATCATTCTTGACCCAATTC  10560
V  E  S  M  I  G  M  M  I  S  I  G  A  N  I  I  L  D  P  I
TGATGTTTGGATTACAGCTTGGTATTCGTGGGGCAGCCATTGCTACAGTAATCGGAAATG  10620
L  M  F  G  L  Q  L  G  I  R  G  A  A  I  A  T  V  I  G  N
CTTTCGCTGTTATCTATTATATTGTCTGTATACAACGAGCAGACAATCAGTTATCTGCTC  10680
A  F  A  V  I  Y  Y  I  V  C  I  Q  R  A  D  N  Q  L  S  A
TTCCAAAATATTTCAGGCTTGAAAAACAAATGCTAAAAGAAATTTTTTTAGTTGGATTAT  10740
L  P  K  Y  F  R  L  E  K  Q  M  L  K  E  I  F  L  V  G  L
CTGCAATGTTGTTAGATATTCTTTTGATTGTTTCAAGCCTTATGTTTAATTACTATGCAC  10800
S  A  M  L  L  D  I  L  L  I  V  S  S  L  M  F  N  Y  Y  A
TAAAATATGGAGATTATGTGCTTGCCGGATTTGGGATTTCTCAAAAACTTGTGCAGATTG  10860
L  K  Y  G  D  Y  V  L  A  G  F  G  I  S  Q  K  L  V  Q  I
HindIII
TCGACCTAATCGGCATGGGACTTTACATGGGAGTAATTCCACTTATCGCTGTTGCCTATG  10920
V  D  L  I  G  M  G  L  Y  M  G  V  I  P  L  I  A  V  A  Y
GTGCAAGAAATGAACTTCGTATGAAAGAAATCATTAAAAAGACTGCTCTCTATTTAGCAC  10980
G  A  R  N  E  L  R  M  K  E  I  I  K  K  T  A  L  Y  L  A
TAGTAATTACATGTTTGTTTGCTATTCTATTTACATGCAGAAACTTTATTGTTCATTGTT  11040
L  V  I  T  C  L  F  A  I  L  F  T  C  R  N  F  I  V  H  C
TTTCAAATGATTCAGATGTAATTCGTATAGGTGCGTACATCTTAACCGTTCAACTCTGTT  11100
F  S  N  D  S  D  V  I  R  I  G  A  Y  I  L  T  V  Q  L  C
CTTCTTTCTTTGCGGCAGGAGCGGGTCTTCTAACGGGAATTTTTCAGTCAAAAGGAGAAG  11160
S  S  F  F  A  A  G  A  G  L  L  T  G  I  F  Q  S  K  G  E
GAACTCCGGCTGTTGTTATGTCCGTTATGAGAGGTCTTATGCTCATACCAGCAATTATCT  11220
G  T  P  A  V  V  M  S  V  M  R  G  L  M  L  I  P  A  I  I
```

```
TCGGCAATTATCTCTTTAAAATGAATGGAGTTATTTTTTCTCTGCTTGTAGCCGAAGCTA  11280
F  G  N  Y  L  F  K  M  N  G  V  I  F  S  L  L  V  A  A  A
                                  Stop effD |
TTTCATGCATTACAGGGATTGTATTATATAAATTGAAAAAGTAATGCTGAACTCCATCAG  11340
I  S  C  I  T  G  I  V  L  Y  K  L  K  K
     HindII
ACTTGTTAACAAAGTTAGTGATGGTTATAAAAATCATCAACAATAAATTAAGAGATTTCT  11400
ATAGAATGAGTATAGATTGGTAGGAGCTGGAAAAATATGATTGGGCTTAAAAAAAGAAAT  11460
ATAAAAAAATGTTGAAAACTGGTAGATGTATCATTAGCTGATATTCGGGCTAATATAGAA  11520
GCCACTATTGATGAAGAAATGAACAGTCCAGCCCCAGAGGTACAGGCAAATTTCAAAAAG  11580
TATTTTGGCAATAAACGTCCTACACCAGAAGAATATATTTACAAGATTACAAAAAAACAA  11640
AAGTTTGATTTACGACTATTTGTTACCTGCGGAGTATATCAACAAATAAGCGTTGAAGAA  11700

                                  | Stop orf9
TCTGCACAGAAAAACAGAGAAAATAAGAAGTGAATCACTTCTTATTCTCTCTGCTGTTCA  11760
                                  K  K  N  E  R  S  N  L
AGATACCATCAATCGTACCTTGAATGATTTTTAATTCGTCATCACTCAATAAATCAAGTG  11820
   I  G  D  I  T  G  Q  I  I  K  L  E  D  D  S  L  L  D  L  S
ACGAATCTATCTGACGGCGAACTGTGCTTTTTTCCACATTCTTCGCTGGATAGAAATACT  11880
   S  D  I  Q  R  R  V  T  S  K  E  V  N  K  A  P  Y  F  Y  E
     HindII
CGTCAACAGATATATTGAACATGGTAACTAAATCATGGAATAAATGAAAGCTAGGGTGTT  11940
   D  V  S  I  N  F  M  T  V  L  D  H  F  L  H  F  S  P  H  K
TCCCGATATTTTCAATATCTGCGATATGACGTTCTCCGTAAAAGACTTTATCGCCTAAAT  12000
   G  I  N  E  I  D  A  I  H  R  E  G  Y  F  V  K  D  G  L  D
CATTTCTGGAAAACCCTGCTTTCTCTCTGGCTTCTCTGATTGCCAGTCCTAAAGGACGAA  12060
   N  R  S  F  G  A  K  E  R  A  E  R  I  A  L  G  L  P  R  F

                  Start orf9
                                                        -10
AATCAAAGCTGTGCGTATCTTTCTTTTTTCTCATTTTTAATCACCTAGCATGAGTTTATA  12120
   D  F  S  H  T  D  K  K  K  R  M
                  -35
CTATTTTCACAGGTTTGCTCTGCTCAAGGAGCAATTTGTGATTCTAAGTAAAACAGGCAG  12180
CCGGCACCGACAACCGAATTTATTCCCTTTTTTTTAGTTTATCATCAGCCCTTGAGCATG  12240
AATTAAAAACTCATATATCAGATGTTTTAGATACAGTCTATTTGTTATTAGTGCAGATAG  12300
GCTGTTTTCATATTTTCAAAAAGTTTTTCAATTCTGCAACAAATCACACCCTGCTGTACA  12360
```

Start *orf7*

GATATGGTGCGGAAAGAGGGATACATACTTTTCACTTCATAGCGAAAGGGGGTGTGATTT 12420

RBS

TGAAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGTAAGCTCAAAAAGG 12480

TTGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTAGCCAGACGACAGGCAAAGGAAG 12540

TACCCTTTTGTGAGCTTCCAGAAATTGTTATTGAGAAATTGATTGTCTGGGATGATTACG 12600

AAAGTGACTATACGACATTCGATGTGTGCAGTATGGAAATCCGTGTGCTTGATGAAGAAC 12660

TTGAAAAATACAGGATATATCGGCATTTACCAGCTTATGAAAAGATAATCAGAAATTTAG 12720

Right End Tn*5398* |

TGTATTTTTATAATAAAAATATAATGCTTGTATACAAAAATATTAAAGATATTTTAGTAA 12780

Stop *orf7* |

GTTTTGTATATAAGCAAACATGTATTTTTTAATATAGTTATCTAAATATTATTTTATATA 12840

CAATAAAATATGACTCCTAAAAATAAAAAATATATCATAAATATAAATAAGTAGATATAG 12900

-10

GTTTTAATTTGATTTATAAACCTATTATAAAATAAAATTCATGATTTTTATAAGCAATAA 12960

Start *ispD*

AATTTTTAGGAGGATATATGATGAATAAAAGAATGAAACTAATTCCGTATGAAATAAATG 13020

RBS

M N K R M K L I P Y E I N

AAAATCTAAGAGGTGCAAAAAATAAATTCCCATATGGAATAAAACAAATGAATGCTAGGG 13080

E N L R G A K N K F P Y G I K Q M N A R

GAATGTGGGATGAAGGTTATACTGGTAAAAATATTGTAGTTGGTATAATAGATACAGGTT 13140

G M W D E G Y T G K N I V V G I I D T G

GTGATATATCTCATCCTCTTTTAAAAGGAAAAATAATTGGTGGTGCAAATTTTAGTGATG 13200

C D I S H P L L K G K I I G G A N F S D

ACAGTAATGGAAATAAAAATATATATGAGGATTTTAATGGTCATGGAACTCATGTGGCGG 13260

D S N G N K N I Y E D F N G H G T H V A

GTATTATAGCTGCATCTAATTATAATAATGAAGTTATGGGAGTAGCTCCAGATTGTAAAT 13320

G I I A A S N Y N N E V M G V A P D C K

TATTAATAGCAAAAGCATTAAATAAAGATGGTACCGGAACATATCAAAGTATAATTAATG 13380

L L I A K A L N K D G T G T Y Q S I I N

CTATTAACTTTGCTGTAAATAACAAGGTTGATATTATATCTATGTCTCTTGGGGGAAACA 13440

A I N F A V N N K V D I I S M S L G G N

AAGATGATAAGAATTTAAAAAATGCTGTCATGCAAGCAGTAAAAAATAATATTTCTGTAG 13500

K D D K N L K N A V M Q A V K N N I S V

TGTGTGCAGCAGGTAATAATGGAGATGGTGATTCTAGTACAAGTGAGTATAGTTATCCAG 13560

V C A A G N N G D G D S S T S E Y S Y P

CCAGTTATGCTGAGGTAATAGAAGTAGGTGCAATAAATGAAAACTATTTGGTTGAAAAGT 13620

A S Y A E V I E V G A I N E N Y L V E K

TTAGTAATTCAAATACTACAATAGATTTGGTGGCTCCAGGAAGAAATATTATATCGACTT 13680

F S N S N T T I D L V A P G R N I I S T

ATATGGATAATAAACTTGCTATTATGAGTGGTACTAGTATGAGTGCACCATACGTATCAG 13740

Y M D N K L A I M S G T S M S A P Y V S

GCTCATTAGCACTAATTAAAGAATGGGCAAGAGAGGAGTTTGAAAGAGATTTAGATGAAG 13800

G S L A L I K E W A R E E F E R D L D E

CTGAACTGTATGCACAATTAATAAAATGTACGAGAGCGCTTGGAATACCTAGAACGGAAC 13860

A E L Y A Q L I K C T R A L G I P R T E

Stop *ispD*|

AAGGAAATGGATACTTATATTTAAATCTTTATAAATACAAGAATAATAGCAAAAGATAAT 13920

Q G N G Y L Y L N L Y K Y K N N S K R

TTTTGATTTGATGATAAAATAGCTATATTATATAGAGTCGAGACAAATAATAAAAATTAC 13980

TTAGGTGTAAGATTTTACATCAATGTATAAAGGTATTAAGTGATAAAATTTATAAACATA 14040

TTAGCTAGTTAGAATTGAAAAATAAATATACGATATTATAGATAGCACATCTGGAAAAGG 14100

TGTGTTTTTCTATGTACTTATACATATAAAAGATTTATAAGAGATGCAAAAGTACTATAA 14160

GTTAGAAATTTTTCTATGAAAGATTAGAAATACGATACAGTGTTTTGATATTATAAAATG 14220

AATATAAAGTATAAATGATAAATGTTATCAATAAGTATTGAATTTCAAAATCAATTGTGA 14280

TATTCTATATTTAGAAAAATAAAGAGATTTCAGATAAAGAAATCTTTTGTTTTTTAGATG 14340

Start *flxD*

-10

TTAGTTAGCTATAACTAACTATTGATTAATTAAAAAATACTACAGGAGGTAAAAATGAGT 14400

RBS

M S

```
AAAATATATATTGTTTATTGGAGTGGAACAGGAAATACTGAAAAAATGGCAAATTTTGTG  14460
 K  I  Y  I  V  Y  W  S  G  T  G  N  T  E  K  M  A  N  F  V
GCTGAAGGTGTGAAGTTAAAAGGTAAGACACCAGAAGTTTTAGATGTGAGCTTACTGAAA  14520
 A  E  G  V  K  L  K  G  K  T  P  E  V  L  D  V  S  L  L  K
CCAAGTGATTTAAAAGAAGAAGATAAATTTGCATTAGGTTGCCCATCTATGGGAGCAGAG  14580
 P  S  D  L  K  E  E  D  K  F  A  L  G  C  P  S  M  G  A  E
CAACTAGAAGAGGGGGATATGGAGCCATTTGTTTCAGAATTAGAATCTATGGTATCAGGT  14640
 Q  L  E  E  G  D  M  E  P  F  V  S  E  L  E  S  M  V  S  G
AAACAGATTGGATTATTTGGTTCATATGGATGGGGAAATTGTGAATGGATGAGAGATTGG  14700
 K  Q  I  G  L  F  G  S  Y  G  W  G  N  C  E  W  M  R  D  W
GAAGAACGTATGCAAAATGCTGGTGCTACAATTATTGGTGGAGAAGGAATTACAGCAATG  14760
 E  E  R  M  Q  N  A  G  A  T  I  I  G  G  E  G  I  T  A  M
GAAGACCCAAATGAAGAAGCAAAAGATGAGTGTATAGAATTAGGCAAAACGTTAGCTGAA  14820
 E  D  P  N  E  E  A  K  D  E  C  I  E  L  G  K  T  L  A  E
Stop flxD|
TAAATTTGTATATTATAAAAATAGTATAAATAGCAACTAATGATGATAGCAGTATATAAT  14880
AATAAGAGAGACTAATATTATGATACATAAGAAAATATCCTTAATAGAGATAGATAAAAT  14940
ATTAGTCTCTTTTAATATAAAATTAGATTTATAATTTAATGTTGTTTTTGAA          14992
```

Table 4.1 : Features of the ORFs detected in pJIR1594 and pJIR1790.

ORF	Promoter Sequences Identified			Start Codon	Stop Codon	Nucleotide Region	Size of ORF (nucleotides)	Size of Predicted Protein (amino acids)
	-10	-35	RBS					
ilvD	N	N	N	None	TAG	1-1040	>1035	>345
hydR	N	N	Y	ATG	TAG	1384-2013	630	209
hydD	N	N	Y	ATG	TAA	2039-2851	813	270
elpD	Y	Y	Y	ATG	TAA	4213-4210	96	31
ermI(B)	N	N	N	ATG	TAA	3433-5167	738	245
orf3a	Y	Y	Y	ATG	TAA	5175-5306	132	43
orf298	N	N	N	ATG	TAG	5703-6599	897	298
erm2(B)	N	N	N	ATG	TAA	6839-7574	738	245
orf3b	Y	Y	Y	ATG	TAA	7622-7703	132	43
orf13	N	N	N	ATG	TAA	8728-9123	395	131
effR	Y	Y	Y	ATG	TAA	9260-9913	653	217
effD	N	N	Y	ATG	TAA	10002-11324	1322	440
orf9	Y	Y	N	ATG	TGA	11735-12094	360	119
orf7	N	N	Y	GTG	TAA	12414-12812	398	132
ispD	Y	N	Y	ATG	TAA	12981-13919	938	312
flxD	Y	N	Y	ATG	TAA	14395-14823	428	142

1993). This reaction is the penultimate step in the synthesis of the amino acids isoleucine and valine, prior to transamination.

A CLUSTAL W alignment with IlvD proteins from *Pyrococcus abyssi*, *Aquifex aeolicus*, *Methanococcus jannaschii*, *B. subtilis*, and *Bacillus cereus* reveals 52-57% identity at the amino acid level across the sequenced portion of the IlvD protein (Figure 4.3). Based on this alignment the incomplete ORF was designated as *ilvD*. It seems likely that *ilvD* encodes the dihydroxyacid dehydratase (IlvD) homologue in *C. difficile*.

b) **hydR**

The second ORF detected, *hydR*, was a complete ORF (nucleotides 1384 to 2013, Figure 4.2, Table 4.1). A putative RBS was located upstream of the start codon (Figure 4.2). BLASTP analysis of the putative HydR protein revealed similarity to proteins belonging to the TetR family of transcriptional regulators.

The genetic control and mechanism of tetracycline resistance has been well characterized. The expression of *tetA*, which encodes an integral membrane protein that belongs to the major facilitator superfamily (MFS) and which exports tetracycline, is under the control of the TetR repressor. In the absence of tetracycline, transcription of *tetA* and the divergent *tetR* gene is repressed by TetR (Hillen and Berens, 1994). *tetA* is efficiently expressed only when TetR is released from its operator sites by its association with tetracycline or its analogues. Many members of the TetR family of regulatory proteins appear to be repressors that bind DNA through a helix-turn-helix motif. The helix-turn-helix motif is commonly used

Figure 4.3 : CLUSTAL W alignment of IlvD proteins.

The amino acid sequence of IlvD from *C. difficile* strain 630 was aligned with the IlvD proteins from *P. abyssi* (F75045), *M. jannaschii* (Q58672), *B. subtilis* (P51785), *B. cereus* (Q9XB13), and *A. aeolicus* (O67009). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile    ------------------------------------------------------------
P. abyssi       --------MRSDVIKKGIERAPHRALFKAMGLTDEELDKPLIGIVNSFNELIPGHIHLRR
M. jannaschii   MNFMKREKMISDRVKKGLKRAPNRSLLKACGYTDEELERPFIGVVNSFTEVVPGHIHLRD
B. subtilis     -----MAELRSNMITQGIDRAPHRSLLRAAGVKEEDFGKPFIAVCNSYIDIVPGHVHLQE
B. cereus       ------------------------------------------------------------
A. aeolicus     ------MKFRSDKVKKGIERAPHRALLRACGLSDEDFDKPLIGIANSYIDIIPGHVHLRE

                        70        80        90       100       110       120
                        |         |         |         |         |         |
C. difficile    ------------------------------------------------------------
P. abyssi       IAEAVKTGVRMSGGTPLEFSTIGICDGIAMGHGGMKYSLPSRELIADSIEAVVRAYNFDG
M. jannaschii   IAEAVKKGIYANGGTAFEFNTMAICDGIAMGHEGMKYSLPSREIIADTVESMAKAHGFDG
B. subtilis     FGKIVKEAIREAGGVPFEFNTIGVDDGIAMGHIGMRYSLPSREIIADSVETVVSAHWFDG
B. cereus       ------------------------------------------------------------
A. aeolicus     FVEPIKEEVRKLGGVPIEFNVIGVDDGIAMGHEGMHYSLPSRELIADSIETVVNAHQLDA

                       130       140       150       160       170       180
                        |         |         |         |         |         |
C. difficile    ------------------------------------------------------------
P. abyssi       IVMIASCDKIIPGMLMAMARLDIPAIFISGGPMLPGRFK-GEYVDVKTVFEAVGAVKAGK
M. jannaschii   LVLIPSCDKIVPGMIMGAIRTGLPFIVVTGGPMFPGELR-GKKYDLISVFEGVGACAAGK
B. subtilis     MVCIPNCDKITPGMLMAAMRINIPTIFVSGGPMAAGRTSYGRKISLSSVFEGVGAYQAGK
B. cereus       ------------------------------------------------------------
A. aeolicus     LICIPNCDKIVPGMLMGALRVNVPTVFISGGPMLAGEVN-GQKVDLISVFEGIGKVKRGE

                       190       200       210       220       230       240
                        |         |         |         |         |         |
C. difficile    ------------------------------------GMGIPYNGTAASHSGERKRIAKYA
P. abyssi       MSEKELKLLEDFACPGCGSCAGMFTANTMNALTEALGISLPWNGTAPAVYAHRIRIAKQT
M. jannaschii   ITEEELKEIEDIACPGAGSCAGLFTANTMACLTEAMGLSLPYCATSHATTAEKIRIAKRS
B. subtilis     INENELQELEQFGCPTCGSCSGMFTANSMNCLSEALGLALPGNGTILATSPERKEFVRKS
B. cereus       ---------------------------MNCLCEVLGLALPGNGSILAIDPRREELIKQA
A. aeolicus     ISEQELKVIEASACPTCGSCSGMFTANSMNCLTEVLGLALPGNGTILAIDPRREILARNA

                       250       260       270       280       290       300
                        |         |         |         |         |         |
C. difficile    GMYVMELLKNDIKPRDILTIDAFKNAIAVDMAMAGSTNTVLHLPAIAYESGIE-LNLDFF
P. abyssi       GMQIMKLVEEDLKPSDILTPEAFEDAIAVDMALGGSTNTVLHLMAIAREAGVK-LTLDTF
M. jannaschii   GMRIVDLVRNNITPDKILTKEAFENAILVDLALGGSTNTTLHIPAIANEVKPKFITLDDF
B. subtilis     AAQLMETIRKDIKPRDIVTVKAIDNAFALDMALGGSTNTVLHTLALANEAGVE-YSLERI
B. cereus       AEKLKILIERDIKPRDIVTEEAIDDAFALDMAMGGSTNTVLHTLALAHEAGLD-YDMSRI
A. aeolicus     VKALFELLEKDVKPRDIVTEEALDDAFTVDIAMGGSSNTILHLLAIAREAGIE-YNLAKI

                       310       320       330       340       350       360
                        |         |         |         |         |         |
C. difficile    DEISEKTPCLTKLSPSGKHHIEDLHMAGGIPAIMNELSKIN-GINLDCKTVTGKTIRENI
P. abyssi       DEISEKTPTLVKISPAGKHFVLDLYEAGGVLAIMKRLSELG-LIHEDRITVSLKTVGELL
M. jannaschii   DRLSGEVPHIASLRPGGEHFIIDLHRAGGIPAVLKVLEEK---IRKECLTVSGKTIGEII
B. subtilis     NEVAERVPHLAKLAPASDVFIEDLHEAGGVSAALNELSKKEGALHLDALTVTGKTLGETI
B. cereus       DAVSRRVPHLCKVSPASNWHMEDIDRAGGISAILKEMSRKEGVLHLDRITATGQTLRENI
A. aeolicus     NEISKRTPTICKISPASHYHIEDLDRVGGIPTIMKELS-KLGLLHTERKTVSGKTIGEII

                       370       380       390       400       410       420
                        |         |         |         |         |         |
C. difficile    RNCEIENEEVIHTLKNPYSNQGGLAILKGNLALNGAVVKNQQLQKNVSSMKDPARIFNSE
P. abyssi       RDVSVLRDDVIRPVTRPYLSRGGLMILYGSLAPKGAVLKVSAI-PDIETFEGEARVFDCE
M. jannaschii   KEVKYIDYSVIRPVDNPVHETAGLRILKGSLAPNGAVVKIGAVNPKMYKHEGPARVFDSE
B. subtilis     AGHEVKDYDVIHPLDQPFTEKGGLAVLFGNLAPDGAIIKTGGVQNGITRHEGPAVVFDSQ
B. cereus       AHAEIKDKEVIHSLENPHSEEGGLRILKGNLAKDGAVIKSG--ATEVKRFEGPCVIFNSQ
A. aeolicus     SDAPDADGEVVRTIENPYSKDGGIAILFGNLAPEGAVVKTAGVDPKMLTFKGKAICFDSE
```

```
                          430        440        450        460        470        480
                           |          |          |          |          |          |
C. difficile      EEAVNAIFGKKINKGDVIVIRYEGPKGGPGMKEMLSPTSAVAGMGLDKHVALLTDGRFSG
P. abyssi         EDAVKAILSGDIEKGDVVVIRYEGPKGGPGMREMLAPTSAIAGMGLDRDVALVTDGRFSG
M. jannaschii     EEAVDAILGGDIERGDVVVIRYEGPAGGPGMREMLAPTSAICGMGLDDSVALITDGRFSG
B. subtilis       DEALDGIINRKVKEGDVVIIRYEGPKGGPGMPEMLAPTSQIVGMGLGPKVALITDGRFSG
B. cereus         DEALAGIMLGKVKKGDVVVIRYEGPRGGPGMPEMLAPTSAIAGMGLGADVALLTDGRFSG
A. aeolicus       EEAIEGILGGKVKPGHVVVIRYEGPKGGPGMREMLSPTSAIMGMGLGDKVALITDGRFSG

                          490        500        510        520        530        540
                           |          |          |          |          |          |
C. difficile      ATTGASIGHISPEAMEGGLIGLVEEGDIISINIPDKKLELKVDEVEIENRKLKFKPLEPK
P. abyssi         ATRGLSIGHVSPEAAEGGPIALVEDGDLIRIDVKAKRIDLLVDEEELKERKAKWK-PKVK
M. jannaschii     GSRGPCIGHVSPEAMAGGPIAIVEDGDIIKIDMINKKLDLALDEEEIKERLAKWKKPEPK
B. subtilis       ASRGLSIGHVSPEAAEGGPLAFVENGDHIIVDIEKRILDVQVPEEEWEKRKANWKGFEPK
B. cereus         ASRGISVGHISPEAAAGGTIALLEQGDIVCIDVEERLLEVRVSDEELDKRKKEWKRPEPK
A. aeolicus       GTRGACVGHISPEAAAGGPIGIVKDGDEILIDIPNRRIELLISEEEFNERMKNFKPKQKE

                                   550        560
                                    |          |
C. difficile      IKHGYLSRYAKLVTSANTGAVLK-----
P. abyssi         EVKGYLKRYSSLVTSANTGAVFRE----
M. jannaschii     VKKGYLARYAKLVSSADEGAVLRYD---
B. subtilis       VKTGYLARYSKLVTSANTGGIMKI----
B. cereus         VKTGWLGRYAQMVTSANTGAVLKVPNFD
A. aeolicus       IKSSWLRRYAKLVTSASKGAILEA----
```

as the signature sequence for identifying members of this family of bacterial regulatory proteins.

CLUSTAL W alignment of HydR and the amino acid sequences of the five most closely related protein sequences from the databases revealed 25% identity to a transcriptional regulator of the TetR family from *A. aeolicus*, 22% identity to a regulatory protein, IfeR, from *Agrobacterium tumefaciens*, 19% identity to probable transcriptional regulators from *Pseudomonas aeruginosa* and *Listeria innocua*, and 16% identity to a probable transcription regulator of the TetR family from *Streptomyces coelicolor*. Most of the identity between these proteins is across the region that is expected to form the helix-turn-helix motif (Figure 4.4). On the basis of this alignment it appears that *hydR* is likely to encode a repressor protein that belongs to the TetR family of bacterial repressors.

c) **hydD**

The *hydD* gene was located 26 bp downstream of the *hydR* stop codon (nucleotides 2039 to 2851, Figure 4.2, Table 4.1). A putative RBS was located upstream of the *hydD* start codon (Figure 4.2).

The putative HydD protein had strongest similarity to a cultivar specificity protein from *Rhizobium leguminosarum*, a lipolytic enzyme from *S. acidocaldarius*, and to a hydrolase, IpbD, from *Pseudomonas putida*, each of which had 22% identity to HydD. Identity (21%) was also observed to the PcbD hydrolase from *Archaeglobus fulgidus* and 19 % identity was found to a carboxyl esterase from *Acinetobacter calcoaceticus* (Figure 4.5).

Figure 4.4 : CLUSTAL W alignment of HydR from *C. difficile* and TetR-like transcriptional regulators.

The amino acid sequence of HydR from strain 630 was aligned with transcriptional regulator proteins from *A. aeolicus* (C70487), *A. tumefaciens* (AAC25692), *P. aeruginosa* (C83286), *L. innocua* (CAC19089), and *S. coelicolor* (T36792). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile    ----------------------------------------------------------MNRE
A. aeolicus     ----------------------------------------------------------MYILLFMG
A. tumefaciens  ----------------------------------------------------------MRRTK
P. aeruginosa   ---------------------------------------------------------MLELVATGQLTD
L. innocua      ----------------------------------------------------------
S. coelicolor   MTPAGAHAPARAALPDAPGTASGKILRREPRSHRHQDGHGTHGTDRATERQVTAVPRTTD

                        70        80        90        100       110       120
                        |         |         |         |         |         |
C. difficile    EKSKNSKEKIIQSAFSLFSSKGYDSTSTQDIINLSGLSRGAMYHHFKTKEDILRSVTKEL
A. aeolicus     EKRSDTKEKILSSALKLFSKKGFKETTIKDIAKEVGITEGAIYRHFTSKEEIIKSLLESI
A. tumefaciens  EQAAETGRQILQAAETLFLDKGYDNVSLEEIAALSGVTRGAIHWHFKNKHGLLLALRNEA
P. aeruginosa   PESARG--KLLQTAAHLFRSKGYERTTVRDLASAVGIQSGSIFHHFKSKDEILRSVMEET
L. innocua      --MKEKKQRIIKSAKEVFQKQGYLKTSVQDMVDAAGISKGTFYNYFTSKEELAIVIFKQE
S. coelicolor   GDSTPVPQRLLAAATRLFAEQGYDRTSVQEIVEAAGVTKGALYHYFGSKDDLLHEVYARV
                                  Helix-Turn-Helix Motif

                        130       140       150       160       170       180
                        |         |         |         |         |         |
C. difficile    YSQMNNFLEYLVADDTLTA------NEKIIELVVHSANDYTRRKMVH---CSWLEKIPFA
A. aeolicus     TKELRHKLEVALQRGET--------DEEILESIVDTLIDYAFSNPES---FRFLNLYHLL
A. tumefaciens  QEPFRQFADELSEGRGSAS------IEKLGDIITDTFKLLEQDPRQRGLLRVMMRLDIGL
P. aeruginosa   ILYNTALMRAALADAEDLR------ERVLGLIRCELQSIMGGTGEAM---AVLVYEWRSL
L. innocua      YSVLHQRLEYTMALDGTKKDNFTECLKIIIHFYTFNGEILNITFSQTMIDDDFNAFLQNV
S. coelicolor   LRLQQERL-DAFADADEPV------EKRVRDAAADVVVTTIENLDDA---SIFFRSMHQL

                        190       200       210       220       230       240
                        |         |         |         |         |         |
C. difficile    LIEEVR----NLNNVVAPNIAKIIKQGVEN--KEFSCEYPEELAEMLVFSIDILLDPVLF
A. aeolicus     KEYG------EVKNLPGELILKFLNGLYLK--RKTKTYP--EIALAVVTGSVERVFIFKE
A. tumefaciens  AEKEEGGEN-TFPEEMHALFVRIFRAVERSP-GMMKPWTPEKAASMVYAAMGGLITEWAL
P. aeruginosa   SAEGQAYIL-GLRDIYEQMWLDVLGEARLA--GYCQGDP--FILRRFLTGALSWTTTWFR
L. innocua      RLKNMEWVKNQLLEVYGKETEPYIIDITMLLSGMAAMYVFASGSKNVDSGLIERAIPYVV
S. coelicolor   SPEKNKQVR-AERRRYHERFRALIEEGQRT--GVFTKEIPADLVVDYHFGSIHHLSTWYR

                        250       260       270       280       290       300
                        |         |         |         |         |         |
C. difficile    KREYSEVCNRLDFLLFMLKKMDIPLIDEYGIQKFKDLFKQ--------------------
A. aeolicus     RNFLDYDEETIKKELKKVLKSAILA-----------------------------------
A. tumefaciens  RKTEFTLSNDGGLFIRTLLAGLQRRPTDEN-------------------------------
P. aeruginosa   PEGPMSLDQLAEEALALVIKNA---------------------------------------
L. innocua      RGLDALVKDILESGEIIFTEADTENLVPDQTMIRKKRLAKLREALEELNVGIENADKADS
S. coelicolor   PDGPLSPQEVADHLAGLLLRALRP---------------------------------------

C. difficile    --------
A. aeolicus     --------
A. tumefaciens  --------
P. aeruginosa   --------
L. innocua      DKWQYKES
S. coelicolor   --------
```

Figure 4.5 : CLUSTAL W alignment of HydD from *C. difficile* and hydrolase homologues from other organisms.

The amino acid sequence of HydD from strain 630 was aligned with a cultivar specificity protein from *R. leguminosarum* (AAF89759), a lipolytic enzyme from *S. acidocaldarius* (AAC67392), hydrolase proteins from *P. putida* (AAC03446) and *A. fulgidus* (A69463), and a carboxyl esterase from *A. calcoaceticus* (CAA61351). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile            -------------MFYNAKSCQMMIGSNTVNYIEFG-----------------------
R. leguminosarum        ------MN--TSVRYLRS-ATCYAEAPNLWINVGETP------------FAYRDLG----
S. acidocaldarius       ------MNRRDAIKIEGAGAIAGALATRIPLSISQTPPQINYSNLVPKEFGWAPVNNIHI
P. putida               --------------MANLEIGKSIVAAGIQTNYHDLG-----------------------
A. fulgidus             ------------MKKIQAMRFGFFDSDGLKLRYFECG-----------------------
A. calcoaceticus        MNSALQFSSSPYTSFMHETKVKLSNGLELHVEVGGNP-----------------------

                        70        80        90        100       110       120
                        |         |         |         |         |         |
C. difficile            -----SGKKTLIILPGLGDGLFPLHGK-IQAIAFAFRYKQFAKDYKVYVFSRKNQIT---
R. leguminosarum        ----LQGGVPVILLNHWGAVLDNFDP--RIVDGLATRHHVIATNYRGIGASGGKAP----
S. acidocaldarius       YYEIYGSGEPLIMIMGYLGNLESWGP--IIINGLASQYEVIIFDNRGTGRSGTVGQDPLH
P. putida               ------EGTPVLLVHGSGPGVSAYANWRLTMPALSKNFRVIAADIVGFGFTDRPEN----
A. fulgidus             ------EGEPLILIHGLGDCIEGWT---FQYDDFSKHFRVVALDLRGFGMSDVPES----
A. calcoaceticus        ------DHPTILLIMGLGAQMLFWPD-FFCKSLIDQGFYVIRFDNRDIGLSSKIRHKGKR

                        130       140       150       160       170       180
                        |         |         |         |         |         |
C. difficile            ------------------EKYSTRDMAKDQADIMKKLGIMKAEVMGVSQGGMIAQYLAID
R. leguminosarum        --------------------LTIDEMARDTIALIRALGFKKVDLLGFSLGGFVAQDITLK
S. acidocaldarius       D----------------ALTYTIPLYASDTIGLLNYLGYSNLNVLGWSMGGFVAQQIAID
P. putida               ------------------YTYSMDNWVKHIIGVMDALEIEKAHIVGNSFGGALAIAIAIR
A. fulgidus             ----------------------ISVEDFARDVKNLIDHLGIERANLLGLSMGGVVCMEFYRQ
A. calcoaceticus        LNTLKLMSRFTLGLGNQGAPYTLYDMAEDVSLLLEAMRIKKVNVIGASMGGMIAQIIAAK

                        190       200       210       220       230       240
                        |         |         |         |         |         |
C. difficile            YPELVEKLVLAVTSSKQNDTIQNVICSW--------ID-MAKKQNYNDLMIDTAKKSYSER
R. leguminosarum        APDLVRKLILAGTGPAGGKGIDKVGAVSWPLIIKGLLTLRDPKTYLFFTSTANGRQAAKA
S. acidocaldarius       YPSYVNKLVLLCTAPNIYLYPPKVSPQS---IITGFTASDPTVTVETIIPYLVPSDWLQA
P. putida               HPGRIDRMVLMGAAGTQFELTDGLDAVWG------YTPSIKNMRDLLDIFAYDRSLVSDEL
A. fulgidus             YPEMVKSLVLANTLHKLPDAGRAMFEQR--------LKLLESSPDMTQIAEFIAEMSIHQD
A. calcoaceticus        YPEKVEKLALMFTSNNQPLLPPPFPKQLFS--LIGKPKSSDEDGILNHSLKLFEIIGSPG

                        250       260       270       280       290       300
                        |         |         |         |         |         |
C. difficile            YLKKYQLFIPFLGKVGKPKDFKRFIIQATSCIEHNAFS-ELNKITCPTLIIGGANDKIVG
R. leguminosarum        FLDRLKERKAGRDKGPAPR-AFLRQLKAIKAWGRQAPQ-DLGRIDVPVLIANGDDDIMVP
S. acidocaldarius       HPDVAKYVLFTLEKYPISYTSVLKQTNALATFNSVG---QLQNITAPTLVIGGDSDLLLP
P. putida               ARLRYEASIQPGFQESFSRMFPAPRQRWVAALASSD-A-DIKGLSNETLIIHGREDQVVP
A. fulgidus             RPELKELVKTIIRKNNKEF--YTKVTAELGRINYED---LLPKIEVPTLVITAEFDVTTP
A. calcoaceticus        YVNQVEAIQTARKLYQRSYHPAGVLQQFLAILCTGSLLQLDKQISQPTLVIHGSRDRLLP

                        310       320       330       340       350
                        |         |         |         |         |
C. difficile            NNASFHLAEKIKKSEIFIY--EGLGHATYEEAQDFNERVLEFLNK----------
R. leguminosarum        TVNSADMARRIPGAEVVIY-QDAGHGGIFQNYADFVPKALSFLGA----------
S. acidocaldarius       PQNSQYLAENIPNAQLYIFSPDAGHGLIYQYPTQFINLVTSFLG-----------
P. putida               LSSSLRLVELIERAQLHVFG-RCGHWTQIEQADRFNRLVVEFFNEASVVPKLDVL
A. fulgidus             PALGREIASLIPKCDIPN-------------------------------------
A. calcoaceticus        PSHGKAVAKAIKGAKFELIQ-GMGHDIPPHFIPQLSYLFAHHFRS----------
```

P. putida is capable of using isopropylbenzene as a sole source of carbon and energy and IpbD is responsible for the conversion of 2-hydroxy-6-oxo-7-methylocta-2,4-dienoate to 2-hydroxypenta-2,4-dienoate and isobutyrate in this catabolic pathway (Eaton and Timmis, 1986). In *Pseudomonas* sp. the PcbD protein is a 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase, which is part of the pathway involved in the degradation of biphenyls and chlorinated biphenyls (Kim *et al.*, 1996). Both of these hydrolases are responsible for enabling their host to use unusual carbon compounds as energy sources. Perhaps the *C. difficile* HydD protein has a similar function.

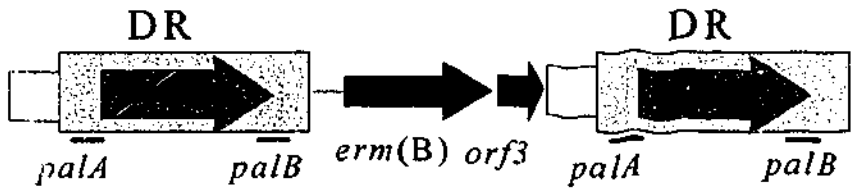
Based on the similarity to these hydrolase proteins this ORF was designated *hydD*, which stands for **hyd**rolase protein from *C. difficile*. The ORF upstream of *hydD*, *hydR*, encodes a putative transcriptional regulator of the TetR family and could potentially be regulating the transcription of *hydD*. Consequently this ORF was designated *hydR*, standing for **hyd**rolase gene repressor.

d) The C. difficile *Erm B determinant*

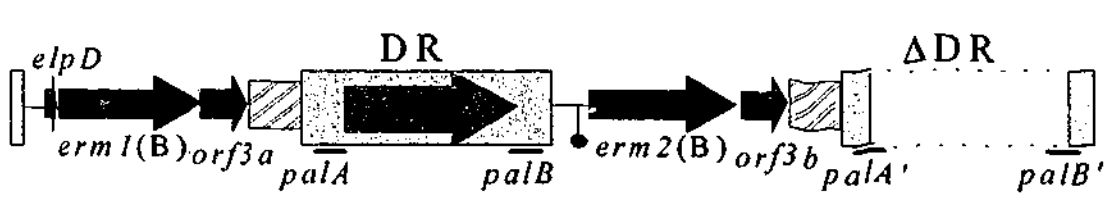
The *C. perfringens* Erm B determinant has been studied extensively (Berryman and Rood, 1995) and has been shown to be located on a large mobilizable plasmid, pIP402. The determinant consists of an *erm*(B) gene and a small ORF designated *orf3*, flanked by direct repeat (DR) sequences. Each DR contains an ORF, *orf298*, which is flanked by highly palindromic repeated sequences, *palA* and *palB* (Figure 4.6). Comparative nucleotide sequence analysis of the next 4 kb of *C. difficile* DNA sequenced in this study revealed that this region of the chromosome contained ORFs that appeared to constitute the *C. difficile* Erm B determinant.

Figure 4.6 : Genetic organization of the Erm B determinants from *C. perfringens* and *C. difficile* strain 630.

The approximate extent and organization of the Erm B determinants from *C. perfringens* (U18931) (Berryman and Rood, 1995) and *C. difficile* strain 630 are shown schematically and not necessarily to scale. Regions encompassing direct repeat sequences (DR) are shown in grey. Regions of nucleotide similarity are shown as pink hatched boxes. The approximate location of the palindromic sequences *palA*, *palB*, *palA'* and *palB'* are indicated by the lines below the DR sequences. The *erm*(B) structural genes are shown as blue solid arrows, the *orf3* genes as teal solid arrows, *orf298* as red solid arrows, and the *elpD* gene is shown as a turquoise solid arrow. The deletion of the promoter sequences upstream of *erm2*(B) is depicted by a solid black oval.



C. perfringens



C. difficile

i) ***elpD***

The first ORF detected in this region was very small, consisting of only 96 bp (nucleotides 4213 to 4210, Figure 4.2, Table 4.1). Putative RBS, -10 and -35 promoter sequences were identified (Figure 4.2, Table 4.1). When translated it encoded a putative protein of 31 amino acids. BLASTP analysis of the predicted amino acid sequence revealed similarity with the leader peptide sequences commonly found upstream of MLS resistance determinants. A CLUSTAL W alignment showed that this *C. difficile* Erm leader peptide had significant identity at the amino acid level to the leader peptides found upstream of the *erm*(B) genes from *E. coli* (97%), *E. faecalis* (81%), *Lactobacillus reuteri* (77%), and the *erm*(B) gene present on Tn*917* (65%) (Figure 4.7). Based on the high level of identity at the amino acid level this ORF was designated as *elpD*, standing for the Erm leader peptide from *C. difficile*.

The most well characterized leader peptide gene region, that upstream of the inducible *erm* gene, *erm*(C), contains a number of inverted repeats and leads to the regulation of *erm* expression by translational attenuation (Mayford and Weisblum, 1990). Based on the similarity of the upstream region of other genes to the leader peptide sequence upstream of *erm*(C), several other *erm* genes, including some *erm*(B) genes, have been proposed to be regulated by translational attenuation. Examination of constitutively expressed *erm* genes has shown that the leader peptide sequence was either absent or was mutated and non-functional (Kamimiya and Weisblum, 1988; Mayford and Weisblum, 1990).

An alignment of the *elpD* sequence and the nucleotide sequences of the leader peptides found upstream of other inducibly and constitutively expressed

Figure 4.7 : Amino acid and nucleotide sequence alignments of Erm leader peptide regions.

The amino acid sequence of the Erm leader peptide, ElpD, from *C. difficile* strain 630 was aligned with the inducible Erm leader peptide sequences from *E. faecalis* (AAC71782), *L. reuteri* (AAC31203), Tn*917* (*E. faecalis*) (P23130) and *S. aureus* (Erm(C)) (NC001386), and the constitutive Erm leader peptide sequence from *E. coli* (P10739). The nucleotide sequence of the *elpD* gene from *C. difficile* was aligned with the nucleotide sequences of the Erm leader peptide genes found upstream of the inducible *erm*(B) genes from *S. agalactiae* plasmid pIP501 (X72021), *E. faecalis* transposon Tn*917* (M11180) and the inducible *erm*(C) gene from *S. aureus* plasmid pE194 (NC001386) and the constitutively expressed *erm*(B) gene from *E. coli* (M19270). Identical amino acids and nucleotide bases are shown in red, highly similar amino acids are shown in green and weakly similar amino acids are shown in blue. Gaps introduced into the alignments are depicted as dashes.

Amino Acid Sequence Alignment

```
                              10        20        30        40
                              |         |         |         |
C. difficile          MLVFQMRYQMRYVDKTSTVLKQTKNSDYADK---------
E. coli               MLVFQMRYQMRYVDKTSTVLKQTKKSDYADK---------
E. faecalis           MLVFQMCN----VDKTSTVLKQTKNSDYADK---------
L. reuteri            MLVFQIRN----VDKTSTGLKQTKNSDYADK---------
E. faecalis Tn917     MLVFQMRN----VDKTSTVLKQTKNSDYADKYVRLIPTSD
S. aureus             MGIFSIFV------ISTVHYQPNKK---------------
```

Nucleotide Sequence Alignment

```
                                  10        20        30        40        50        60
                                  |         |         |         |         |         |
C. difficile          ATGTTGGTATTCCAAATGCGTTATCAAATGCGTTATGTAGATAAAACATCTACTGTTTTG
E. coli               ATGTTGGTATTCCAAATGCGTTATCAAATGCGTTATGTAGATAAAACATCTACTGTTTTG
S. agalactiae         ATGTTGGTATTCCAAATGCGTAAT----------GTAGATAAAACATCTACTGTTTTG
E. faecalis           ATGTTGGTATTCCAAATGCGTAAT----------GTAGATAAAACATCTACTGTTTTG
S. aureus             ATG--GGCATTTTTAGTAT-TTTT----------GTAATCAGCACAGTT--CATTATC

                                  70        80        90        100       110       120
                                  |         |         |         |         |         |
C. difficile          AAACAGACTAAAAACAGTGATTACGCAGATAA-------------------------
E. coli               AAACAGACTAAAAAAAGTGATTACGCAGATAAATAA-------------------------
S. agalactiae         AAACAGACTAAAAACAGTGATTACGCAGATAAATAA-------------------------
E. faecalis           AAACAGACTAAAAACAGTGATTACGCAGATAAATACGTTAGATTAATTCCTACCAGTGAC
S. aureus             AACCAAAC-AAAAAA----------TAA-------------------------

C. difficile          ---
E. coli               ---
S. agalactiae         ---
E. faecalis           TAA
S. aureus             ---
```

erm(B) genes shows that the Erm leader peptide from *C. difficile* is most similar to the leader peptides that are associated with constitutively expressed *erm* genes and is therefore likely to be non functional (Figure 4.7).

Induction experiments were carried out on *C. difficile* strain 630 to determine whether MLS resistance was inducibly or constitutively expressed. The results (Figure 4.8) showed that when the cells were subcultured from medium that did not contain erythromycin, the same growth rate was observed in the presence or absence of erythromycin. It is concluded that in *C. difficile* strain 630 MLS resistance is constitutively expressed and therefore it appears likely that the Erm leader peptide is not functional.

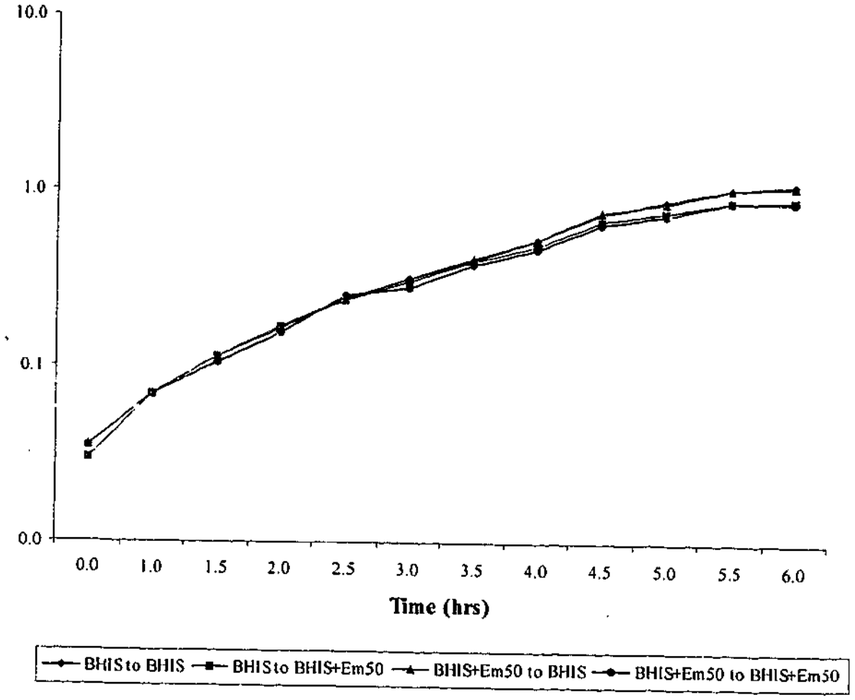
ii) The *erm*(B) genes

Sequencing and analysis of the next region of the recombinant plasmids proved difficult due to double priming of the oligonucleotide primers as a result of gene duplications. Consequently, three independent PCR products spanning the region from nucleotides 4954 to 7012 (Figure 4.2) were generated using oligonucleotides #3139 and #4210 and then sequenced to give reliable sequence data in this region.

Two identical ORFs, *erm1*(B) (nucleotides 3433 to 5167, Figure 4.2, Table 4.1) and *erm2*(B) (nucleotides 6839 to 7574, Figure 4.2, Table 4.1), were detected downstream of *elpD*. These ORFs encoded proteins that showed high levels of identity to 23S rRNA methyltransferases of the Erm(B) family. The resultant Erm(B) proteins had 98% identity to the Erm(B) protein from *C. perfringens*, 97% identity to Erm(B) from *S. agalactiae*, 96% identity to Erm(B) from *L. reuteri* and

Figure 4.8 : Expression of erythromycin resistance in *C. difficile* strain 630.

Growth of *C. difficile* cells that had not been pre-exposed to erythromycin (BHIS), or had been pre-exposed to erythromycin (BHIS+Em_{50}), was monitored after subculture into medium that either contained erythromycin (BHIS to BHIS+Em_{50}, BHIS+Em_{50} to BHIS+Em_{50}) or did not contain erythromycin (BHIS to BHIS, BHIS+Em_{50} to BHIS), over a six hour period by measuring the turbidity at 600 nm.



E. faecium, and 84% identity to Erm(B) from *E. faecalis* (Figure 4.9). Previously, *erm*(B) genes from *C. difficile* have been referred to as both *ermZ* (Hächler *et al.*, 1987a) and *ermBZ* (Rood and Cole, 1991), however, in accordance with the most recent nomenclature system for the Erm family of resistance determinants (Roberts *et al.*, 1999), these ORFs will be referred to as *erm1*(B) and *erm2*(B).

No promoter sequences were identified upstream of either *erm1*(B) or *erm2*(B) (Figure 4.2, Table 4.1). The fact that *erm1*(B) is preceded by *elpD* suggests that this ORF may be transcribed from the promoter sequences upstream of *elpD*. When the sequence upstream of *erm2*(B) was compared to the sequence upstream of the *C. perfringens erm*(B) gene, it was evident that a deletion event had removed the *erm2*(B) promoter sequences and a small portion of the upstream DR sequence (see later discussion on DR sequences). It is therefore unlikely that the *erm2*(B) gene is expressed, however, expression from an upstream promoter, such as the *elpD*, *orf3a*, or *orf298* promoters, cannot be ruled out.

iii) *orf3a* and *orf3b*

Two identical ORFs, *orf3a* (nucleotides 5175 to 5306, Figure 4.2, Table 4.1) and *orf3b* (nucleotides 7622 to 7703, Figure 4.2, Table 4.1), were detected downstream of the *erm1*(B) and *erm2*(B) genes, respectively. As before, due to the duplication of the genes in this region, reliable sequence data were obtained from the PCR products described for the *erm*(B) genes.

Analysis of the amino acid sequence derived from both *orf3a* and *orf3b* revealed identity to several ORF3 proteins from other bacteria. Identical ORF3 proteins are encoded by the *orf3* genes found downstream of the *erm*(B) genes from

Figure 4.9 : CLUSTAL W alignment of Erm(B) proteins.

The amino acid sequence of the Erm(B) protein from *C. difficile* strain 630 was aligned with the Erm(B) proteins from *C. perfringens* (P12038), *S. agalactiae* (NP053005), *L. reuteri* (AAC31204), *E. faecium* (AAF64431), and *E. faecalis* (pAMβ1) (A27507). Identical amino acids are shown in red, highly similar amino acids are shown in blue, weakly similar amino acids are shown in green.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile            MNKNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLAKISKQVTSIELD
C. perfringens          MNKNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLAKISKQVTSIELD
S. agalactiae           MNKNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLAKISKQVTSIELD
L. reuteri              MNKNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLAKISKQVTSIELD
E. faecium              MNKNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLAKISKQVTSIELD
E. faecalis             MNKNIKYSQNFLTSEKVLNQIIKQLNLKETDTVYEIGTGKGHLTTKLAKISKQVTSIELD

                                70        80        90       100       110       120
                                |         |         |         |         |         |
C. difficile            SHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKIVGSIPYHLSTQIIKKVVFESRAS
C. perfringens          SHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKIVGSIPYHLSTQIIKKVVFESHAS
S. agalactiae           SHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKIVGNIPYHLSTQIIKKVVFESRAS
L. reuteri              SHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKIVGNIPYHLSTQIIKKVVFESRAS
E. faecium              SHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKIVGSIPYNLSTQIIKKVVFESRAS
E. faecalis             SHLFNLSSEKLKLNTRVTLIHQDILQFQFPNKQRYKIVGSIPYHLSTQIIKKVVFESHAS

                                130       140       150       160       170       180
                                |         |         |         |         |         |
C. difficile            DIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIKQLLKLPAECFHPKPKVNSALIKLTRHTT
C. perfringens          DIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTT
S. agalactiae           DIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTT
L. reuteri              DIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTT
E. faecium              DIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTT
E. faecalis             DIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIQQLLKLPAECFHPKPKVNSVLIKLTRHTT

                                190       200       210       220       230       240
                                |         |         |         |         |         |
C. difficile            DVPDKYWKLYTYFVSKWVNREYRQLFTKNQFHQAMKYAKVNDLSTVTYEQVLSIFNSYLL
C. perfringens          DVPDKYWKLYTYFVSKWVNREYRQLFTKNQFHQAMKHAKVNNLSTVTYEQVLSIFNSYLL
S. agalactiae           DVPDKYWKLYTYFVSKWVNREYRQLFTKNQFHQAMKHAKVNNLSTITYEQVLSIFNSYLL
L. reuteri              DVPDKYWKLYTYFVSKWVNREYRQLFTKNQFHQAMKHAKVNNLSTITYEQVLSIFNSYLL
E. faecium              DVPDKYWKLYTYFVSKWVNREYRQLFTKNQFHQAMKHAKVNNLSTITYEQVLSIFNSYLL
E. faecalis             DVPDKYWKLYTYFVSKWVNREYRQLFTKNQFHQAMKHAKVNNLSTVTYEQVLSIFNSYLL

                                        250       260       270       280
                                        |         |         |         |
C. difficile            FNGRK-------------------------------------------
C. perfringens          FNGRK-------------------------------------------
S. agalactiae           FNGRK-------------------------------------------
L. reuteri              FNGRKQLPIF--------------------------------------
E. faecium              FNGRKLIL----------------------------------------
E. faecalis             FNGRKMSRFCKFGKLHVTKGNVDKLLGILLTASKELKRSLAPTGNL
```

S. agalactiae, *L. reuteri*, and *Staphylococcus intermedius*. In addition, the *C. difficile* protein had 91% identity to the *orf3* products distal to the *erm*(B) genes from *E. faecalis* and *C. perfringens* (Figure 4.10).

The predicted amino acid sequences of the ORF3 proteins have no homologues in the database and their function is unknown. The fact that *orf3* is so well conserved, and is nearly always found in association with *erm*(B) genes, suggests that it may have some function in MLS resistance, however, no such role has been elucidated.

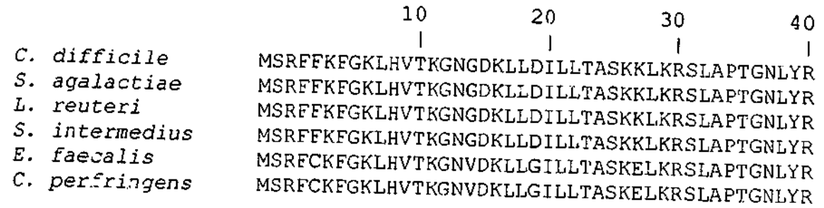
iv) DR sequences and *orf298*

In *C. perfringens* the *erm*(B) gene is flanked both upstream and downstream by DR sequences (Figure 4.6) that each contain an internal ORF, *orf298*, which is flanked by highly palindromic sequences, *palA* and *palB* (Berryman and Rood, 1995). Analysis of the nucleotide sequence of the *erm*(B) gene region from strain 630 revealed that the duplicated *erm*(B) and *orf3* ORFs are separated by a single copy of the DR sequence found in *C. perfringens*. However, a deletion event appears to have removed the last 51 bp of this DR homologue and the promoter sequences upstream of *erm2*(B). Further analysis of this DR revealed an ORF, *orf298* (nucleotides 5703 to 6599, Figure 4.2, Table 4.1), which was flanked by *palA* (nucleotides 5636 to 5698, Figure 4.2) and *palB* (nucleotides 6600 to 6655, Figure 4.2) sequences.

BLASTP analysis of the putative protein encoded by *orf298* revealed similarity to the ORF298 protein from *C. perfringens* and also to other ORF298 homologues that are found in association with *erm*(B) genes. A CLUSTAL W

Figure 4.10 : CLUSTAL W alignment of ORF3 proteins.

The amino acid sequence of the ORF3 protein encoded by *orf3a* and *orf3b* from *C. difficile* strain 630 was aligned with the ORF6 protein from *S. agalactiae* (NP053006), the ORF3 proteins from *L. reuteri* (AAB86540) and *S. intermedius* (AAG42228), the hypothetical erythromycin resistance protein 3 from Tn*917* (*E. faecalis*) (C25028), and ORF3 from *C. perfringens* (I40879). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.



alignment showed that ORF298 has 97% identity to the putative ORF298 proteins found within DR1 and DR2, which flank the *erm*(B) gene in *C. perfringens* (Berryman and Rood, 1995), and 97% identity to the hypothetical protein delta, which is found downstream of the *erm*(B) gene in *S. pyogenes* (Figure 4.11).

The function of these ORF298 homologues has not been elucidated. They contain two ATPase domains, an ArsA family ATPase domain and a ParA family ATPase domain. Besides the ORF298 homologues from *C. perfringens* and *S. pyogenes*, the *C. difficile* ORF298 protein is most closely related to replication associated proteins, and proteins from the ParA or Soj families of plasmid partitioning proteins. ParA and Soj proteins are generally involved in the partitioning of plasmids and chromosomes during replication (Easter *et al.*, 1998; Sharpe and Errington, 1996). However, the level of identity between either replication associated proteins or Soj/ParA proteins and ORF298 is low, approximately 26% and 20%, respectively (Figure 4.12, Figure 4.13).

Nucleotide sequence analysis of the *C. difficile erm*(B) gene region also detected two other segments that showed similarity to DR sequences. The first of these regions was detected downstream of the *erm2*(B) and *orf3b* genes (nucleotides 7941 to 8240, Figure 4.2, Figure 4.6). This DR variant contained a deletion that had removed *orf298*. This deletion appears to have occurred *via* homologous recombination between the palindromic sequences *palA* and *palB*, leaving vestigial *palA'* and *palB'* sequences. This variant also contained a second deletion, which had removed the last 70 bp at the 3' end of the DR sequence. The second variant DR region was upstream of the *elpD* gene (nucleotides 4039 to 4109, Figure 4.2, Figure 4.6). It consisted of a 70 bp region that corresponded in nucleotide sequence to the

Figure 4.11 : CLUSTAL W alignment of putative ORF298 proteins.

The amino acid sequence of ORF298 from *C. difficile* strain 630 was aligned with the hypothetical ORF298 proteins from *C. perfringens* DR1 (I40877) and DR2 (I40880), and hypothetical protein delta from *S. pyogenes* (S45079). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue. The ArsA and ParA family ATPase domains are underlined.

```
                                10        20        30        40        50        60
                                |         |         |         |         |         |
C. difficile                    MIQYYYTKKEWGVVMEKEKLKILEELRRILNNKNEAIIILNNYFKGGVGKSKLSTMFAYL
C. perfringens (DR1)            MIQYYYTKKEWGVVMEKEELKILEELRRILNSKNEAIVILNNYFKGGVGKSKLSTMFAYL
C. perfringens (DR2)            MIQYYYTKKEWGVVMEKEELKILEELRRILSNKNEAIVILNNYFKGGVGKSKLSTMFAYL
S. pyogenes                     MIQYYYTKKEWGVVMEKEKLKILEELRRILNSKNEAIVILNNYFKGGVGKSKLSTMFAYL

                                70        80        90        100       110       120
                                |         |         |         |         |         |
C. difficile                    TDKLNLKVLMIDKDLQATLTKDLAKTFEVELPRVNFYEGLKNGNLASSIIHLTDNLDLIP
C. perfringens (DR1)            TDKFNLKVLMIDKDLQATLTKDLAKTFKVELPRVNFYEGLKNGNLASSIVHLTDNLDLIP
C. perfringens (DR2)            TDKLNLKVLMIDKDLQATLTKDLAKTFKVELPRVNFYEGLKNGNLASSIVHLTDNLDLIP
S. pyogenes                     TDKFNLKVLMIDKDLQATLTKDLAKTFKVELPRVNFYEGLKNGNLASSIVHLTDNLDLIP
                                                ArsA family ATPase domain

                                130       140       150       160       170       180
                                |         |         |         |         |         |
C. difficile                    GTFDLMLLPKLTRSWTFENESRLLATLLAPLKSDYDLIIIDTVPTPSVYTNNAIVASDYV
C. perfringens (DR1)            GTFDLMLLPKLTRSWTFENESRLLATLLAPLKSDYDLIIIDTVPTPSVYTNNAIVASDYV
C. perfringens (DR2)            GTFDLMLLPKLTPSWTFENESRLLATLLAPLKSDYDLIIIDTVPTPSVYTNNAIVASDYV
S. pyogenes                     GTFDLMLLPKLTRSWTFENESRLLATLLAPLKSDYDLIIIDTVPTPSVYTNNAIVASDYV
                                ParA family ATPase domain

                                190       200       210       220       230       240
                                |         |         |         |         |         |
C. difficile                    MIPLQAEEESTNNIQNYISYLIDLQEQFNPGLDMIGFVPYLVDTDSATIKSNLEELYKEH
C. perfringens (DR1)            MIPLQAEEESTNNIQNYISYLIDLQEQFNPGLDMIGFVPYLVDTDSATIKSNLEELYKQH
C. perfringens (DR2)            MIPLQAEEESTNNIQNYISYLIDLQEQFNPGLDMIGFVPYLVDTDSATIKSNLEELYKQH
S. pyogenes                     MIPLQAEEESTNNIQNYISYLIDLQEQFNPGLDMIGFVPYLVDTDSATIKSNLEELYKQH

                                        250       260       270       280       290
                                        |         |         |         |         |
C. difficile                    KEDNLVFRNIIKRSNKVSTWSKNGITEHKGYDKKVLSMYENVFFEMIERIIQLENEKE
C. perfringens (DR1)            KEDNLVFQNIIKRSNKVSTWSKNGITEHKGYDKKVLSMYENVFFEMLERIIQLENEKE
C. perfringens (DR2)            KEDNLVFQNIIKRSNKVSTWSKNGITEHKGYDKKVLSMYKNVFFEMLERIIQLENEKE
S. pyogenes                     KEDNLVFQNIIKRSNKVSTWSKNGITEHKGYDKKVLSMYENVFFEMLERIIQLENEKE
```

Figure 4.12 : CLUSTAL W alignment of the *C. difficile* ORF298 protein with replication associated proteins.

The amino acid sequence of the ORF298 protein from *C. difficile* strain 630 was aligned with the replication associated protein from *E. faecalis* plasmid pAM373 (NP071998), the Rep63B protein from *B. thuringiensis* (CAB43193), the RepB protein from *E. faecalis* plasmid pAD1 (B47092) and the RepB protein from *L. reuteri* (AAC02983). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

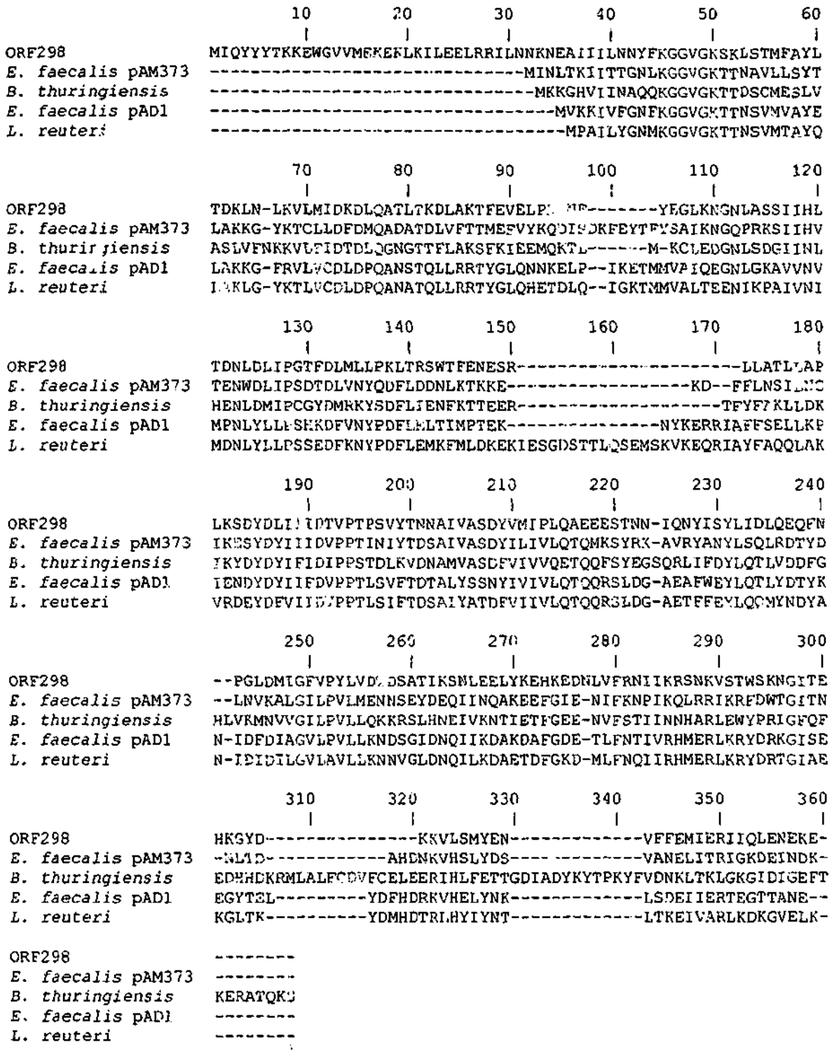


Figure 4.13 : CLUSTAL W alignment of the *C. difficile* ORF298 protein with Soj and ParA proteins.

The amino acid sequence of the ORF298 protein from *C. difficile* strain 630 was aligned with an ATPase involved in chromosome partitioning from *Thermoplasma volcanium* (NP111999), the Soj protein homologues from *P. abyssi* (NP126969) and *Pyrococcus horikoshii* (NP142704), and putative plasmid partitioning proteins from *B. burgdorferi* (NP051238) and *S. coelicolor* (T36875). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                    10        20        30        40        50        60
                     |         |         |         |         |         |
ORF298              ------------------------------------------------MIQYYYTKKEW
T. volcanium        ------------------------------------------------------------
P. abyssi           ------------------------------------------------------------
B. burgdorferi      ------------------------------------------------------------
P. horikoshii       ------------------------------------------------------------
S. coelicolor       MSMDGQHVNAMAGDGSGAPRNHFADYDELPEGHFYDPDAEYEPDPEYAATLAPDAARQRR

                    70        80        90        100       110       120
                     |         |         |         |         |         |
ORF298              GVVMEKEKLKILEELRRILNNKNEAIIILNNYFKGGVGKSKLSTMFAYLTDKLNLKVLMI
T. volcanium        ------------------------MIISIANQKGGCGKTTTAVNLGSVLARK-HKVLLI
P. abyssi           ------------------------MGTVISIANQKGGVGKTTIALNLGYSLSRLGKKVLLV
B. burgdorferi      -------------------MDNKKPKIITIASIKGGVGKSTSAIILATLLSKN-NKVLLI
P. horikoshii       ------------------------MGVVISIANQKGGVGKTTIALNLGYSLSKFGKKVLLV
S. coelicolor       ERIGPTGRPLPYFPIPGPLTDHGPAKIIAMCNQKGGVGKTTSTINLGAALAEYGRRVLLV

                    130       140       150       160       170       180
                     |         |         |         |         |         |
ORF298              DKDLQATLTKDL---AKTFEVELPRVNFYEGLKNGNLASSIIHLT--DNLDLIPGTFDLM
T. volcanium        DIDPQGNLTTS----FGVNKGELNRTMYDVMLDGGLEK-AILRK---DSIDIVPSIIDLA
P. abyssi           DVDPQFNLTFAL---IGMDVVNYEDKNVGTLMTRESSVEDVLIEVE-ENLHLIPSHLTLS
B. burgdorferi      DMDTQASITSYFYEKIEKLGINFTKFNIYEILKENVDIDSTIINVD-NNLDLIPSYLTLH
P. horikoshii       DIDPQFNLTFAL---IGMDVINYENKNVGTLMTKESTVEDVLIEIN-ENIHLIPSHLTLS
S. coelicolor       DFDPQGALSVG----LGVNPMELDLTVYNLLMERGMAADEVLLKTAVPNMDLLPSNIDLS

                    190       200       210       220       230       240
                     |         |         |         |         |         |
ORF298              LLPKLTRSWTFENESRLLATLLAPLKSDYDLIIIDTVPTPSVYTNNAIVASDYVMIPLQA
T. volcanium        GAEVQLSGRMGR--EYILANELSKLSRRYDFIIIDTPPSLGVFTINALVASDYVLIPVQA
P. abyssi           AKEIEIMNTYNR--ERRLEKALKPIFPDYDYIIIDNPPSMGIFLVNSLTASDYVLIPLEL
B. burgdorferi      NFSEDKIEHK----DFLLKTSLGTLYYKYDYIVIDTNPSLDVTLKNALLCSDYVIIPMTA
P. horikoshii       AKEIEILNTYNR--ERRLEKALKPVYPEYDYIIIDNPPSMGIFLVNSLTASDYVLIPLEL
S. coelicolor       AAEVQLVSEVAR--ESTLQRALKPLMDDYDYIVIDCQPSLGLLTVNALTAAHKVIVPLEC

                    250       260       270       280       290       300
                     |         |         |         |         |         |
ORF298              EEESTNNIQNYISYLIDLQEQFNPGLDMIGFVPYLVDTDSATIKSNLEELYKEHKEDNLV
T. volcanium        EFFALEGLTQLLSVVDLVNTRLGRTLKILGMVVTMFNSRTKSSNEVLEDVRKHYSKH--L
P. abyssi           SYFGVIGMQLMFNLMSMIREETNEGLRLLGIVPNKFTKQTKVPQARLKELKELYPDA--P
B. burgdorferi      EKWAVESLDLFN----FFVRKLNLFLPIFLIITRFKKNRTHKTLFEILKTKDR-------
P. horikoshii       SYFGVIGMQLMFNLMKMIKDETNEGLRLLGIVPNKFTRQTKIPEKRLKELKELYPEA--P
S. coelicolor       EFFALRGVALLTETIEKVQERLNPDLELDGILATMYDSRTVHSREVLARVVEAFDDH--V

                    310       320       330       340       350
                     |         |         |         |         |
ORF298              FRNIIKRSNKVSTWSKNGIT-EHKGYDKKVLSMYENVFFEMIERIIQLENEKE
T. volcanium        FRTIIPRNVTVTDSTMTGEPVVIYRKDASASKSYVELAKEVENRLRVKR----
P. abyssi           ILTTIPKAIAIEKAQAEGKSIFEYEPNGKASKAFEKLAREVISIVEGQ-----
B. burgdorferi      FLGTISEREDLNRRIAENN---NFDLNKDYIKEYENILEIFLKKI--------
P. horikoshii       ILTTIPKTVTIEKAQAEGKSILEYDPNGKASRAFEKLAREVISLVEG------
S. coelicolor       YHTVIGRTVRFPETTVAGEPITTYASNSVGAAAYRQLAREVLARCHAE-----
```

70 bp of the DR sequence that had been deleted from the DR variant downstream of *erm2*(B).

v) Comparative analysis of the *C. difficile* Erm B determinant

With the exception of the ORFs and DR sequence variants described above, no further similarity was detected between this region of the recombinant plasmids and any other Erm B determinants. Therefore, it appears that the *C. difficile* strain 630 Erm B determinant consists of two identical *erm*(B) genes, *erm1*(B) and *erm2*(B), which are separated by a single DR sequence that contains *orf298* (Figure 4.6). The Erm B determinant is bounded by variants of the DR sequence, and includes an Erm leader peptide upstream of *erm1*(B), and a deletion of the promoter sequences upstream of *erm2*(B) (Figure 4.6).

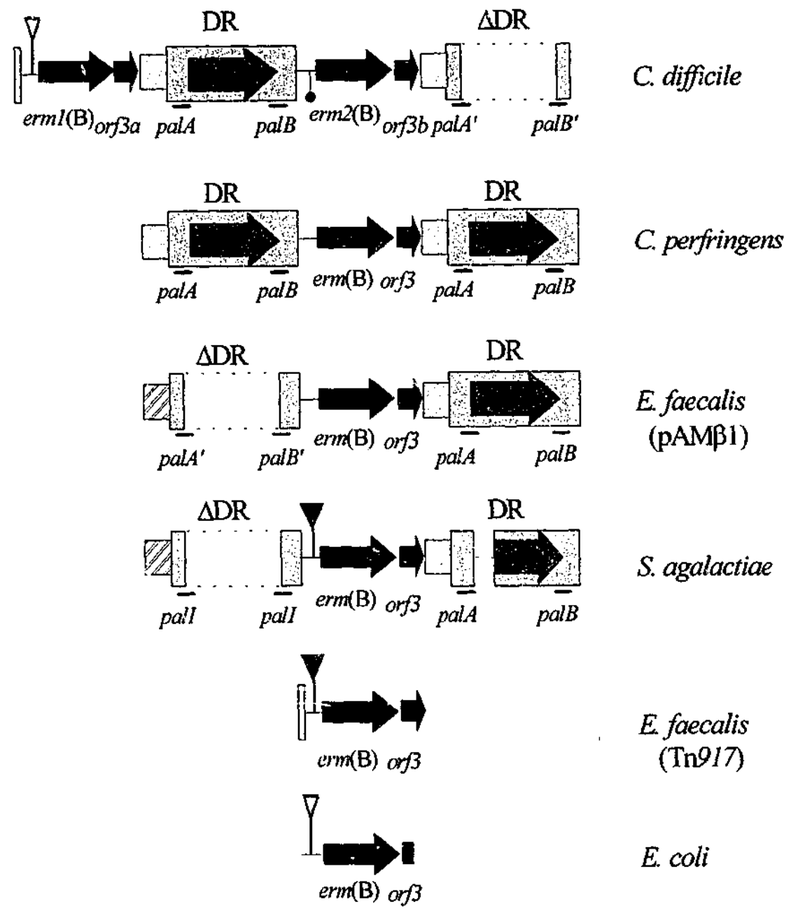
Comparative analysis of the various Erm B determinants for which sequence flanking the *erm*(B) gene is available (Figure 4.14) reveals that the *C. difficile* determinant is the only member of this class that has two *erm* structural genes. Almost all of the *erm*(B) genes are flanked by complete or deleted (ΔDR) variants of the DR sequence. None of the DR variants are identical, with each deletion apparently having occurred at slightly different locations within the *palA* and *palB* sequences. This finding supports the postulate that homologous recombination events involving the *palA* and *palB* sequences are responsible for the deletions, rather than site-specific recombination events.

***e)* orf13**

The ORF identified downstream of the *C. difficile* Erm B determinant was *orf13* (nucleotides 8728 to 9123, Figure 4.2, Table 4.1). The putative ORF13 protein

Figure 4.14 : Comparative genetic organization of the Erm B determinants.

The approximate extent and organization of the Erm B determinants from *C. difficile*, *C. perfringens* (U18931) (Berryman and Rood, 1995), *E. faecalis* (pAMβ1) (Berryman and Rood, 1995), *S. agalactiae* (pIP501) (U00453), *E. faecalis* (Tn*917*) (M11180), and *E. coli* (M19270) are shown schematically and are not necessarily to scale. Regions of nucleotide sequence similarity are indicated by the same coloring or shading. The solid arrows indicate the individual ORFs and their respective direction of transcription. The approximate location of the palindromic sequences (*palA* and *palB*) are indicated by the lines below the colored boxes. The *palA'*, *palB'*, and *palI* sequences represent the portions of the *C. perfringens erm*(B)-derived *palA* and *palB* homologues that are present at the ends of the deletion in these variants of the DR sequences. Functional and non-functional Erm leader peptide sequences are indicated by solid and open blue triangles respectively. The promoter deletion upstream of the *C. difficile erm2*(B) gene is indicated by the solid oval. The region of pIP501 for which no sequence data is available is indicated by a single broken line. This comparison has been modified from Berryman and Rood (1995).



showed significant similarity to only two other proteins when analyzed using BLASTP. The first of these proteins was ORF13 from Tn*916* (Flannagan *et al.*, 1994), and the second was a conserved hypothetical protein from *Thermoplasma acidophilum* (Ruepp *et al.*, 2000). A CLUSTAL W alignment revealed that ORF13 only had similarity to these proteins in the C-terminal region (Figure 4.15). The sequence upstream of the predicted start codon of *orf13* was consequently examined for potential deletions or insertions that may have caused a frameshift, but none were found. Therefore, it seems likely that ORF13 represents a truncated and presumably non-functional variant of these proteins.

ORF13 had 20% identity at the amino acid level to the complete ORF13 protein from Tn*916* (Flannagan *et al.*, 1994), however, if only the C-terminal end of ORF13 is used in the alignment this percentage was increased to 40%. ORF13 has no known function in the conjugative transposition of Tn*916* (Clewell and Flannagan, 1993).

f) effR

BLASTP analysis of the amino acid sequence encoded by the next ORF, *effR*, (nucleotides 9260 to 9913, Figure 4.2, Table 4.1) revealed low level similarity to the MarR family of transcriptional regulators. In *E. coli*, *marR* encodes a repressor of the *marRAB* operon, which regulates multiple antibiotic resistance by controlling the expression of at least 10 unlinked genes (Sulavik *et al.*, 1997). Mutations in *marR* lead to derepression of the *marRAB* operon and result in the increased expression of *marA*, which encodes the positive transcriptional regulator of the unlinked resistance genes (Cohen *et al.*, 1993).

Figure 4.15 : CLUSTAL W alignment of the ORF13 protein from *C. difficile* and its homologues.

The amino acid sequence of ORF13 from *C. difficile* strain 630 was aligned with the ORF13 protein from Tn*916* (*E. faecalis*) (AAB60020) and a conserved hypothetical protein from *T. acidophilum* (CAC11941). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile    ------------------------------------------------------------
E. faecalis     MRKEDLMMKFRKNQNKEKQIPKEKKPRVYKVNPHKKVVIALWVLLGLSFSFAIFKHFTAI
T. acidophilum  MKVSELMTTDPITVSIDDTFSKVMSKMNETGIHQLPVMDGDRYAGMITYSDLLKKRSIQV

                        70        80        90        100       110       120
                        |         |         |         |         |         |
C. difficile    ------------------------------------------------------------
E. faecalis     DTHTIHETTIIEKEYVDTHHVENFVENFAKVYYSWEQSDKSIDNRMESLKGYLTDELQAL
T. acidophilum  KSKISNYTISTPTLNADDDVLEAVRLIKDTGLSALPVFQKGKLVGIISRTDIIKNLPQIV

                        130       140       150       160       170       180
                        |         |         |         |         |         |
C. difficile    ------------------------------MCIGMVQWVIT------------------
E. faecalis     NVDTVRKDIPVSSSVRGFQIWTVEPTGDNEFNVTYSVDQLITEGENTKT-----VHSAYIV
T. acidophilum  DVRDVRIFQIMSSDPIYVYEDDDIEEAFDSMRMLNEVEIPVASRDEKLSGIIRLNSILNI

                        190       200       210       220       230       240
                        |         |         |         |         |         |
C. difficile    -------------QNPTLAPVVQKSKYEPKAQGADVSVSSDTVKDATTFLETFFKLYPTA
E. faecalis     SVYVDGSGNMVLVKNPTITNIPKKSSYKPKAIESEGTVDSITTNEINEFLTTFFKLYPTA
T. acidophilum  LYRQKEKIKYGGYGEKEPVEIKCKSLMDPPVSVDRYAGIDEAVKLMMQYSLHIMPVTDSG

                        250       260       270       280       290       300
                        |         |         |         |         |         |
C. difficile    TEKELAYYVKDG-VLAPVSGDYVFSELVNP-VFTKDGDNLKVSVSVKYLDNKSKMTQISQ
E. faecalis     TASELSYYVNDG-ILKPIGKEYIFQELVNP-IHNRKDNQVTVSLTVEYIDQQTKATQVSQ
T. acidophilum  KIVGIVDFSDLINMIKTESKEGILIEISGLDVYDEDLYDIAFELSERFLDKFSKMTDVSQ

                        310       320       330       340       350       360
                        |         |         |         |         |         |
C. difficile    YELVLHKD--------------------------DNWKIVG------------------
E. faecalis     FDLVLEKNG-------------------------SNWKIIE------------------
T. acidophilum  GKLLIHVMKYKTQGSTKYSIRTRISAPPLFLVQNGSGWNFAEVLGEIFDRYEERIKKMKE

C. difficile    --
E. faecalis     --
T. acidophilum  KQ
```

CLUSTAL W alignment of the EffR amino acid sequence with the five most closely related homologues from the database shows 20% identity at the amino acid level with a transcriptional regulator from *Lactococcus lactis*, 18% identity with MarR family homologues from *S. coelicolor* and *Rhodobacter capsulatus*, 16% identity with ORF145 from *Staphylococcus sciuri*, and 15% identity with a MarR family homologue from *B. subtilis* (Figure 4.16). The identity is primarily localized to the helix-turn-helix motif, which is part of the DNA binding portion of these proteins. Based on these data, it is possible that *effR* encodes a transcriptional repressor.

g) effD

The next ORF, *effD*, located downstream of *effR*, was the largest ORF detected in this study (nucleotides 10002 to 11324, Figure 4.2, Table 4.1). This ORF encodes a predicted protein of 440 amino acids. Using BLASTP, homology to several conserved, hypothetical, integral membrane proteins was observed. A CLUSTAL W alignment of the EffD amino acid sequence and the five most closely related proteins shows that EffD has 34% identity to a hypothetical protein from *L. lactis*, 25% identity to an unknown conserved protein from *Bacillus halodurans*, 24% identity to a hypothetical protein from *P. horikoshii* and to a DinF related pro·in from *P. abyssi*, and 22% identity to a conserved, hypothetical integral membrane protein from *B. burgdorferi* (Figure 4.17). BLASTP analysis also indicated that EffD had two regions of similarity to the consensus sequence for an uncharacterized membrane protein family (Pfam protein family, UPF0013 (Bateman *et al.*, 2000)), which includes hypothetical and proven integral membrane proteins, damage inducible proteins and some multidrug efflux proteins (data not shown).

Figure 4.16 : CLUSTAL W alignment of EffR with MarR family homologues.

The amino acid sequence of EffR from *C. difficile* strain 630 was aligned with a transcriptional regulator from *L. lactis* (AAK04806), a MarR family protein from *S. coelicolor* (CAB56677), the PetP protein from *R. capsulatus* (P31078), ORF145 from *S. sciuri* (CAA73494), and a transcription regulator MarR family homologue, YfiV, from *B. subtilis* (H69804). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue. The helix-turn-helix motif is underlined.

```
                          10        20        30        40        50        60
                          |         |         |         |         |         |
C. difficile   ----------MNNIYSDIYEKLSTLQWLMKRHQMFCQAESGPFADTS---RGQGRILAM
L. lactis      -------MSEQTNNLLNLFSKLLHNPNILFALRADGISKQMKNRGNRN---GAQGLLVEL
S. coelicolor  -------MSENESPSPSAVQASREVRAVISRLRRRILKASEVEDLTLG-----QSSVLAR
R. capsulatus  ---MADTGAPGGETLLFLTDEQLRKGIEAMFFAYRGFTADPDRILDQHDYGRAHHRAIHF
S. sciuri      ----------MDHNNMKLANPVCFSPYNVSRLFAQFYEKELKSFGLTY---SQYLVLLSL
B. subtilis    MEYNLHDTTVLNENILSSEEREIWVLYMKVLTSAGLGDVSEWMKLDMS---MPQMKVLML

                          70        80        90       100       110       120
                          |         |         |         |         |         |
C. difficile   LKIQPEIATKELAYLLGIRQQSLNELLNKMEKNGYVERKPSENDKRVMIVHLTEKGKQ-V
L. lactis      WNKD-GITNAEIAELLDIKPSSVTAQVKQLEEAEMVIRKQDENDKRVSRIFLTDKGREAQ
S. coelicolor  LSDRGSITASELAAIEGVRHQSMTATVTALASMGLVERRPDPDDGRRLLITLTPEGVRRV
R. capsulatus  INREPGLTVTTLISVLGVTKQSLNRVLRTLIDDGLVESRVGRRDKRERHLHLTEKGAVLE
S. sciuri      WEEN-PQTLHSIGQKLDLSSNTLTPLLKRLEQSGWVTRQKSDQDKRQLIVSLTQKGK---
B. subtilis    LNNHGTLKVSDIAEKMGASLSNTTGLLDRLEKSGFVKRSHSEEDRRSVVVQLTENAK---
                                  Helix-turn-helix Motif

                         130       140       150       160       170       180
                          |         |         |         |         |         |
C. difficile   QQPKTDYQNIFNCLLPEELLQMSQYLDRIIEAFQLQN----------G-----NA
L. lactis      ETRDTMHNDISETIFGNLTDEEQEQLAFLMEKLVENNDARNDQDFQRMAQFFGNPFMENM
S. coelicolor  EEGRQVRTEWLAERLQDRCTE--EERRTVIAAMAVL----------
R. capsulatus  RELSEAQRVRMPAAYRAAGPQAVAGFRQVLEAMMDP----------
S. sciuri      EQKQPIYDAISKCVSEDMNLELYKQTKDIMDQLETT----------
B. subtilis    KIFRGLYEKGHLKLKRSLELLSPEEKQAVYEGLSILS----------

                         190       200       210       220       230       240
                          |         |         |         |         |         |
C. difficile   LEENNMIDWMAQARERMGDEHFEQLMSMRERAFGHMRPPKDIPGAERFSENYNGYVPDRD
L. lactis      MNAHDRHAFGNHMRREMQNWQREMRRSSQEMKRGQNRMRRDERSQAPFGEDWKGFGKQLK
S. coelicolor  --------------ERLNRD----------
R. capsulatus  -----------AMRRHYQMLKDAE----------
S. sciuri      ------------LRKQLNK----------
B. subtilis    --------------RALENAKKE----------

                         250       260       270       280       290       300
                          |         |         |         |         |         |
C. difficile   GFQPRNFRPDIK----------
L. lactis      DSFRNNFGGRDQADFNEHWDHFSDEMKNRFGDFDGNRRPDFDSKPKTDKNKEPKKPENWD
S. coelicolor  ----------
R. capsulatus  ----------
S. sciuri      ----------
B. subtilis    ----------

C. difficile   --
L. lactis      DF
S. coelicolor  --
R. capsulatus  --
S. sciuri      --
B. subtilis    --
```

Figure 4.17 : CLUSTAL W alignment of the EffD amino acid sequence from *C. difficile* with conserved, hypothetical and integral membrane proteins.

The amino acid sequence of EffD from *C. difficile* strain 630 was aligned with a hypothetical protein from *L. lactis* (AAK05582), an unknown conserved protein from *B. halodurans* (BAB05882), a hypothetical protein from *P. horikoshii* (C71172), a DinF related protein from *P. abyssi* (B75053), and a conserved hypothetical integral membrane protein from *B. burgdorferi* (H70158). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile    ----------------MNQKNENYWLLDAPVTKAIWHMAIPMMLGMSINIIYNITDTFFI
L. lactis       ----------------MSKKSNSYYLKEAPIRKAIAHLSIPMMIGISATTIYNLINAYFI
B. halodurans   ----------------MEQKQQSERLGTEAIPKLLRSLSIPAMIGMFVMALYNVVDTIFI
P. horikoshii   MVLKGELDEEDSGMSIEKAKMMRKEILEGPIEGTLFKLAWPIIINNLIQVLYNITDTFWL
P. abyssi       -------------MSMEKVKAMRREILEGPIEKTLLKLAWPIVVNNLIQVLYNITDTFWL
B. burgdorferi  -------------MSTDKSKTRELILNGNLYKVLFLISFPIVITNIIQAFYDLTDMFYV

                        70        80        90       100       110       120
                        |         |         |         |         |         |
C. difficile    GRLNDTAALAAISLLLPFTTILMAIGNLFGTGGSTLFSRLLGSENTDRTKQCSATTLWLS
L. lactis       GLVHDTNMMSTITLGLTITIVFMAIGNLFGVGGGTYITRLLGSGDKEKGKKIAGYSFYMS
B. halodurans   SYAVGIEGVAGVTIAFPIMMIMMSMAGALGIGGASVISRRLGERRGEEANQVFGNILTVI
P. horikoshii   GKLG-RAALSAPGVSWPIIGTLMALGIGFTMAGFSIVGQYIGAGNFEKANRSAGALFSLI
P. abyssi       GKLG-REALSAPGVAWPIIGTLMALGMGFTMAGFSIVGQYIGAGDFKKANRSAGALFSLV
B. burgdorferi  GKLG-AMPLSALSLAGPVNFFIIAIAMGMATGSISLMSKCIGEGNFSRFSRYAGQLIVLN

                       130       140       150       160       170       180
                        |         |         |         |         |         |
C. difficile    FLFGLLTAIISIIFSNYIIRLLGADSNTFAYVKQYLIFYGMGAPFIIANFTLEQLIRGDG
L. lactis       LIIGLLIGLVSLFFLSPIAHILGSDAATLTYTKEYAAVLLGGGFSMILNYGLEQLVRAEG
B. halodurans   LVLSVIGFISAFTLLGPALQLFGATSVTQGYATDYLFPILLGSIFFFFAFAANNIIRSEG
P. horikoshii   IFFSTASAILSILILPYALSFMHVTSSIRPYAYKYSLIIFAGVPMSFLFMTFNVLMRAIG
P. abyssi       LFFSTLSALISILILPLALDFMNVTPSIRPYAYRYAIVIFAGVPVSFLFMVFNALMRATG
B. burgdorferi  FVLSLFVTICAFFFIDHLLDLLGVKGELKELSRVYFYVTIFAIPIMFLSISITYILNAQG

                       190       200       210       220       230       240
                        |         |         |         |         |         |
C. difficile    KSVESMIGMMISIGANIILDPILMFGLQLGIRGAAIATVIGNAFAVIYYIVCIQRADNQL
L. lactis       ASKESMYGMFVGVATSIVLDALFILVLNLHVIGAGLSMVIANLASTAYYVWFLENKSESL
B. halodurans   NATFAMVTMIVPAVLNILLDVLFIFGLNMGVLGASIATVIAQASVTGLVLRYFLTGKSTL
P. horikoshii   DTRTPVKISLITVFLNIILDPIFIFILNLGVEGAALATVLSNSVGTVIGVRILTSGRAGL
P. abyssi       DTKTPVKISALTVFLNIILDPIFIFIFKLGVLGAALATVLSNAVGTIIGVKILMSGKAGL
B. burgdorferi  ETILSMTIVLFANIVNFILDPILIFSFNMGITGAAWATLFSKLLTVVFYLFLTYRLNYGL

                       250       260       270       280       290       300
                        |         |         |         |         |         |
C. difficile    SALPKYFRLEKQMLKEIFLVGLSAMLLDILLIVSSLMFNYYALKYG-DYVLAGFGISQKL
L. lactis       KGFLKHFKISLKDQFEIYKIGIAEVFKSAFMIVTTLLLNHFSVQYG-DNVLASFGIALRI
B. halodurans   SLHWSDLRMKGSVIKEVCLVGLPAFVQQSSASLMMIAINSMLLRFGSDFYVGVFGLVQRI
P. horikoshii   KISRENLKPDWKFYSKIVRIGLPSSIGQSANSFGFVVLTRIIMGFG-DVTYAAYVITTRI
P. abyssi       KITLEDLKPDWKFYSKIVRVGLPSSIGQSANSFGFVVLTRIIMGFG-DVTYAAYVITTRL
B. burgdorferi  KIHLKDLVIDIRSIKEIVNLGLPSTFGQIMVSLSFFIFNYIVIEIS-PKFLAAYGLTNTI

                       310       320       330       340       350       360
                        |         |         |         |         |         |
C. difficile    VQIVDLIGMGLYMGVIPLIAVAYGARNELRMKEIIKKTALYLALVITCLFAILFTCRNFI
L. lactis       TQLPEFLTMGLFLGAIPLFAYSFASNNTQRLKEALKELTLWIGGISILFALIVYLFRVPV
B. halodurans   MMFVMMPMMGIMQAMQPIVGYNYGAKQYSRLRETVMLGFKVATIFSIGIFALLMLFPEAL
P. horikoshii   VNFITSISRGISMAMSTMVAQNIGAENYERAKVIAERTMVINFLIAGIAVLIIGIFRVPI
P. abyssi       VNFITSISRGISMAMGTMVAQNVGAEKYERAKVIAERTMVVNFAIAGFAVLVIGVFRVPI
B. burgdorferi  ISFLFLPAMGIGTGIISIVGQNLGAKKVNRVEEVLKKGFFISLAILLIINSIVIFNKQFI

                       370       380       390       400       410       420
                        |         |         |         |         |         |
C. difficile    VHCFSNDSDVIRIGAYILTVQLCSS-FFAAGAGLLTGIFQSKGEGTPAVVMSVMR-----
L. lactis       MQLFTDDSAVIKVGLIVLAAQLVSS-VFNGFSGLFTGIFEASGLGAQTGVMSTIQ-----
B. halodurans   LRVFTADREVIQAGVSAMHILFCVT-FLIGAQIVAGGLYQSLGKPKQALILSLSR-----
P. horikoshii   FKVFLNDPMVIAESKIVLEYFLISVPFFNGIFVVVTGAFSSSGHTVKSMIMSMLRLWGFR
P. abyssi       FKVFLDDPEVIAESKIVLEYFLISVPFFNGIFVVVTGVFSSAGHTVKSMVMSMLRLWGFR
B. burgdorferi  LRLFTNDLEVLNYANNYLLLTTIGT-FGYGLQQVFFGGLIGSGRTKIAMIVIFIR-----

                       430       440       450       460       470       480
                        |         |         |         |         |         |
C. difficile    -----GLMLIP--------AIIFGNYLFKMNGVIFSLLVAEAISCITG---IVLYKLKK-
L. lactis       -----GILYIP--------IVIILHHFYGLTGLIWSMTLTEGIAFVAGLIMLIPYLLKLN
B. halodurans   ----QIIFLIP--------LVLILPHIFGLSGVWWAFPIADVLSFILT--VVLLYRDRNV
P. horikoshii   IPLSYVFGYVP---------TFAKLFGFTSRGVFFGMGLSNFLAAIVG----LAWFLKGS
P. abyssi       IPLSYAFGYIPAITFLGIRIPMAELFNFTSRGVFFGMGLSNFLAAIIG----LAWFLKGS
B. burgdorferi  ----LWLIRLP-------VVFIFQYFGIMENSLGYAFIISNYLAIIIL----IAFTCTRY

                       490       500
                        |         |
C. difficile    --------------------
L. lactis       NGKLMENQPELAVAE-----
B. halodurans   FFLKTKEERELDLVKTASST
P. horikoshii   WMRKIIE--------------
P. abyssi       WMRRIID--------------
B. burgdorferi  WVKPILIKKYK---------
```

Given that the levels of identity between EffD and these proteins was only moderate, EffD was also analyzed using SOSUI (Takatsugu *et al.*, 1998), which is a program used for the classification and secondary structure prediction of membrane proteins. This analysis predicted that EffD was a membrane protein that had twelve membrane spanning domains (Figure 4.18). Based on this analysis of EffD, and the fact that *effD* is directly downstream of *effR*, it is possible that *effD* may encode a putative efflux protein and that *effR* may encode its associated regulator protein. Based on the above analysis these two ORFs are referred to as *effD* and *effR*, representing **eff**lux protein from *C. **difficile*** and **eff**lux protein **r**egulator. However, it is recognized that the designation of these proteins as being involved in efflux must remain speculative.

***h)* orf9**

The ORF detected downstream of *effD* was unique among all of the ORFs detected in this region of the chromosome, in that it was detected on the complementary DNA strand of. This ORF, *orf9* (complement of nucleotides 11735 to 12094, Figure 4.2, Table 4.1) consists of 352 nucleotides and, when translated, encodes a putative protein of 119 amino acids.

Analysis of the predicted amino acid sequence of ORF9 returned only four homologues with statistically significant identity. A CLUSTAL W alignment showed 43% identity to ORF9 from Tn*916*, 40% identity to an ORF9 homologue from the *L. lactis* plasmid pK214, and 28% identity to an ORF9 homologue from the *E. faecalis* transposon Tn*1549* (Figure 4.19). In addition, ORF9 also had 43% identity to ORF9 from the tetracycline resistance transposon Tn5397 from *C. difficile* (Roberts *et al.*, 2001). In the conjugative transposon Tn*916*, *orf9* is predicted to

Figure 4.18 : Secondary structure prediction of the EffD protein.

The secondary structure predicted using SOSUI (http://sosui.proteome.bio.tuat.ac.jp/) consists of twelve membrane spanning regions (green boxes), which results in both the N-terminal and C-terminal ends of the protein being located in the cytoplasm (IN) of the cell. Polar and positively charged residues are in shown in blue with a blue circle surrounding the residue. Negatively charged residues are shown in red with a red circle surrounding the residue. The membrane surrounding the cell is represented by the yellow region.

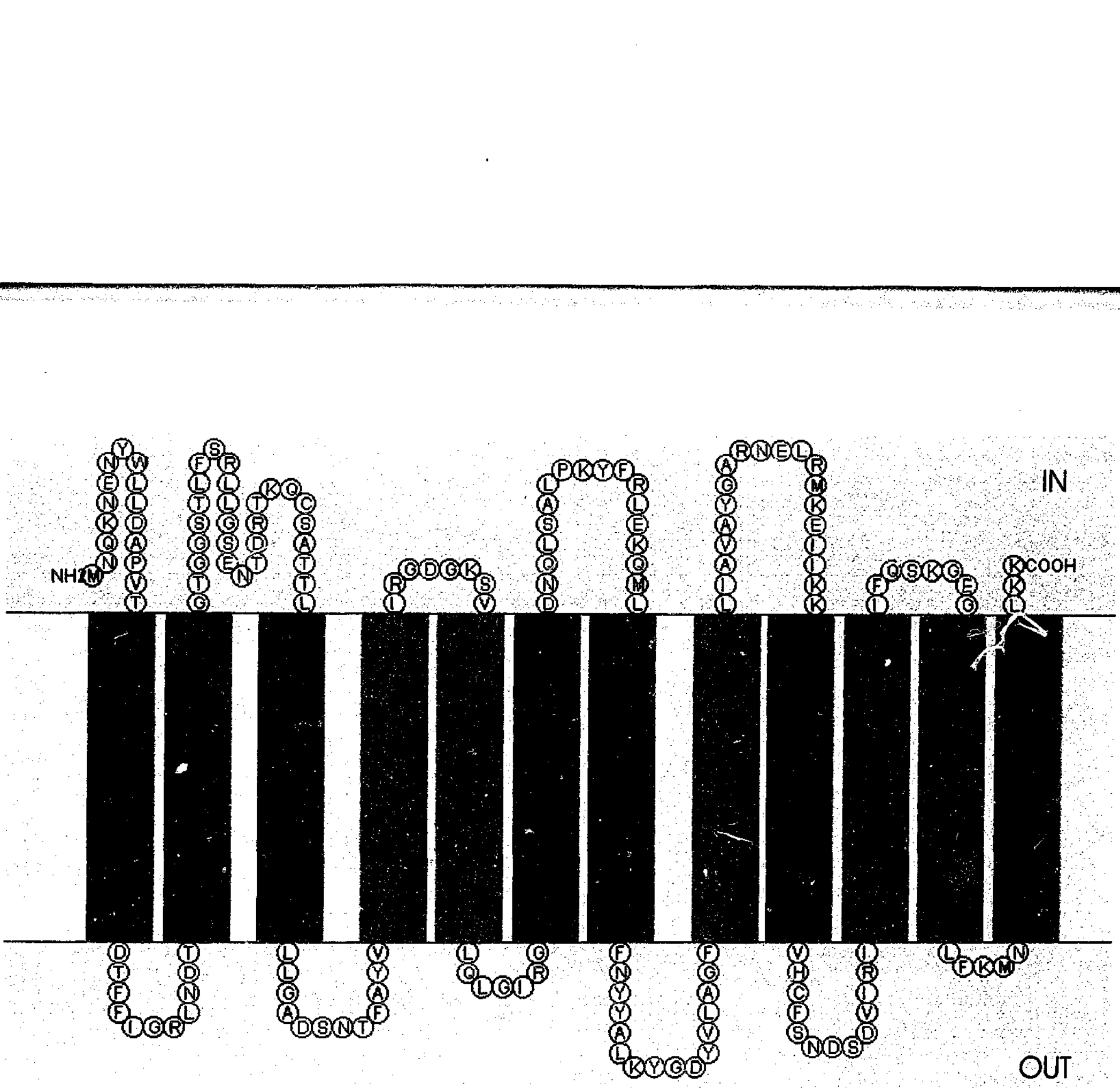


Figure 4.19 : CLUSTAL W alignment of ORF9 homologues.

The amino acid sequence of ORF9 from *C. difficile* strain 630 was aligned with the ORF9 proteins from Tn*916* (*E. faecalis*) (AAB60024) and Tn*5397* (Roberts *et al.*, 2001), a hypothetical protein from the *L. lactis* plasmid pK241 (CAA63526), and an unknown protein from the *E. faecalis* conjugative transposon Tn*1549* (AAF72357). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile    --MRKKKDTHSFDFRPLGLAIREAREKAGFSRNDLGDKVFYGERHIADIENIGKHPSFHL
Tn916           --MRKKEDK--YDFRAFGLAIKEARLQRGLTREQVGALIEIDPRYLTNIENKGQHPSIQV
Tn5397          --MRKKEDK--FDFRAFGLAIKEARMKRGLTCEQVGALIEIDPRYLTNIENKGQHPSIQV
L. lactis       MMDRKKHRT--FDFKPLGLAIKEARLSKGLTREQVGSMIQIAPRYLTNIENKGQQPSFHV
Tn1549          --MANIEDCP--GFETFGADVKAARQAKRVSRKAMAEKINIDWRYLANIENDGAIPSLPV

                        70        80        90       100       110       120
                        |         |         |         |         |         |
C. difficile    FHDLVTMFNISVDEYFYPAKNVEKSTVRRQIDSSLDLLSDDELKIIQ-GTIDGILNSR--
Tn916           LYDLVSLLHVSVDEFFLPANNLVKSTRRLQIEKYMDSFTDKELSLME-SLASGINEARNI
Tn5397          LYDLVSLLDVSVDESFLPHNDLTKSTRRLQVEKLLDNFTDKELSLIE-ALASGINESRNI
L. lactis       VYDLVTLLNISLDGFILGEENSHKSSKRLQVEAQLDCLDEDFLVTIG-KVIQAIYGEKCK
Tn1549          VIQLIKACGLPVERYFNPEIMREESELRQRVSHKLKLCPEEYLPIIEGAIDSAILIEKPK

C. difficile    EN-----
Tn916           ED-----
Tn5397          KNSS---
L. lactis       FN-----
Tn1549          KNETEDV
```

encode a putative transcriptional repressor, however, the exact role of this protein in the mobility of Tn*916*, or in gene regulation, remains to be elucidated (Celli and Trieu-Cuot, 1998).

i) **orf7**

A second ORF with similarity to an ORF from Tn*916*, *orf7* (nucleotides 12414 to 12812, Figure 4.2, Table 4.1), was located downstream of *orf9*. *orf7* was the only ORF detected in the study that appeared to utilize an alternate, GTG, start codon.

A CLUSTAL W alignment of homologues of the predicted amino acid sequence of ORF7 revealed that it had 20% identity to the ORF7 proteins from Tn*916*, Tn*1549* and an additional *E. faecium* transposon Tn*5382*, and 16% identity to an ORF7 homologue from the *L. lactis* plasmid pK214 (Figure 4.20). In addition, ORF7 had 20% identity to ORF7 from Tn*5397* (Roberts *et al.*, 2001).

The ORF7 protein from the prototype conjugative transposon Tn*916* shows limited identity to various prokaryotic RNA polymerase sigma factors (Flannagan *et al.*, 1994). Recent work (Celli and Trieu-Cuot, 1998) has suggested that ORF7 has a regulatory role in the mobility of Tn*916*, in that increased expression of ORF7 leads to increased transcription of transfer genes, *orf7, orf8, xis, int* and adjacent chromosomal genes.

j) **ispD**

BLASTP analysis of the predicted amino acid sequence encoded by the penultimate ORF detected in this study, *ispD* (nucleotides 12981 to 13919,

Figure 4.20 : CLUSTAL W alignment of ORF7 homologues.

The amino acid sequence of ORF7 from *C. difficile* strain 630 was aligned with the ORF7 proteins from Tn*916* (*E. faecalis*) (AAB60026) and Tn*5397* (Roberts *et al.*, 2001), and the ORF7 homologues from Tn*1549* (*E. faecalis*) (AAF72365), Tn*5382* (*E. faecium*) (AAC34795), and *L. lactis* plasmid pK214 (CAA63525). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile    ----------------VILKPSDFQKTIQCQLDCKLKKVVKGSVRNYCKELARRQAKE
Tn916           --MNEQGSLPFLD-RKGVRNMKPSSFQTTIENQFDYICKRAMEDERKNYMLYLSRIAKRE
Tn5397          KEIKEQSSLPFQ--RKGVRNMKPSSFQERIEHQFDFICMRAMDDERKNYFLYLSRLAKRE
Tn1549          --------------------MDAIPRDYEARCMFDAFCKTVLRNEAKSYLAEMKRRRDRE
Tn5382          --------------------MDAIPRDYEARCMFDAFCKTVLRNEAKSYLAEMKRRRDRE
L. lactis       ---MFIGIIFFLNGKGGVNDMSSSLFQAAIEMQFDYICKRSIDDERKDYLKSLSRISKKE

                        70        80        90       100       110       120
                        |         |         |         |         |         |
C. difficile    VPFCELPEIVIEKLIVWDDYESDYTTFDVCSMEIRVLDEELEKYRIYRHLPAYEKIIRNL
Tn916           VSFSDVGDYLVSQFATTDNYSTDFQIFTLNGLSVGVENDLLSEA--LRELPDKKREILLL
Tn5397          VSFSDIGDYLVNQFATTDSYSSDFQIFTLNDISVGIENDLLSEA--LKELPDKKREILLL
Tn1549          VSLSSLSQADLDKLCTVDHYPSDTFTFSSHGYDLHINNELVAEA--FAALPSMEQSILIL
Tn5382          VSLSSLSQADLDKLCTVDHYPSDSFTFSSHGYDLHINNELVAEA--FAALPSMEQSILIL
L. lactis       VAFSELDDYVVEQFASIDQ------------------------------------------

                       130       140       150       160
                        |         |         |         |
C. difficile    VYFYNKNIMLVYKNIKDILVSFVYKQTCIF---------------
Tn916           FYFMDMSDSEIADLLKLNRSTVYRHRTSGLALIKKFMEEFEE---
Tn5397          FYFMDMSDSEIADLLKLNRSTVYRHRTSGLDLIKKFMEEENEE--
Tn1549          HCVLDMADGEIGGLVGMSRSAVQRHRTNTLSELRKQLKALMPKGG
Tn5382          HCVLDMADGEIGGLVGMSRSAVQRHRTNTLSELRKQLKALMPKGG
L. lactis       ------------------------------------------------
```

Figure 4.2, Table 4.1), revealed high levels of identity to intracellular proteases. CLUSTAL W alignment of the amino acid sequence of the putative IspD protein and the five most closely related proteins from the database showed that it had 52% identity to the major intracellular serine protease, ISP-1, from *B. subtilis*, 48% identity to the intracellular serine proteases from *Paenibacillus (Bacillus) polymyxa* and *Bacillus amyloliquefaciens*, and 44% identity to the intracellular alkaline serine proteinases from *B. halodurans* and *Thermoactinomyces* sp. (Figure 4.21). Based on the high level of identity between IspD and the intracellular serine proteases, it seems likely that *ispD* may encode an intracellular serine protease, which may or may not be functional in *C. difficile*. Accordingly this ORF is referred to as *ispD*, which stands for intracellular serine protease from *C. difficile*.

In *B. subtilis* ISP-1 has been postulated to have a critical role in sporulation, possibly through the turnover of intracellular proteins, in the processing of spore coat protein precursors and in the inactivation of transcarbamylase and several other enzymes (Koide *et al.*, 1986). As *C. difficile* is also a spore producing bacterium, IspD may have a similar role in this organism.

***k)* flxD**

The last ORF detected was *flxD* (nucleotides 14395 to 14823, Figure 4.2, Table 4.1). Analysis of the predicted amino acid sequence of the FlxD protein revealed high levels of identity to flavodoxin proteins from several organisms. Alignment of these homologues showed that FlxD had similarity to flavodoxin proteins from *Clostridium beijerinckii* (45% identity), *Megasphaera elsdenii* (43% identity), *Treponema pallidum* (39% identity), *Desulfovibrio salexigens* (31% identity), and *Desulfovibrio desulfuricans* (28% identity) (Figure 4.22). These high

Figure 4.21 : CLUSTAL W alignment of IspD with intracellular serine protease homologues.

The amino acid sequence of IspD from *C. difficile* strain 630 was aligned with the intracellular serine proteases from *B. subtilis* (P11018), *B. amyloliquefaciens* (AAB33888), and *P. polymyxa* (P29139), and the intracellular alkaline serine proteinases from *B. halodurans* (BAB05902) and *Thermoactinomyces* sp. (JC5460). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue.

```
                                10        20        30        40        50        60
                                 |         |         |         |         |         |
C. difficile                    MNKRMKLIPYEINENLRGAKNKFPYGIKQMNARGMWDEGYTGKNIVVGIIDTGCDISHPL
B. subtilis                     MNGEIRLIPYVTNE-QIMDVNELPEGIKVIKAPEMWAKGVKGKNIKVAVLDTGCDTSHPD
B. amyloliquefaciens            ------------------DVNELPEGIKVIKAPELWAKGFKGKDIKIAVLDTGCDVSHPN
P. polymyxa                     MERKVHIIPYQVIK-QEQQVNEIPRGVEMIQAPAVWNQ-TRGRGVKVAVLDTGCDADHPD
B. halodurans                   MG-QVRLIPYEVTS-IQDDTKKIPPGIEMIEAPDLWQQGYKGKGIVVAVLDTGCDVEHYE
Thermoactinomyces sp.           MG-QVRLIPYEVTS-IQDDTKKIPPGIEMIEAPDLWQQGYKGKGIVVAVLDTGCDVEHYE

                                70        80        90       100       110       120
                                 |         |         |         |         |         |
C. difficile                    LKGKIIGGANFSDDSNGNKNIYEDFNGHGTHVAGIIAASNYNNEVMGVAPDCKLLIAKAL
B. subtilis                     LKNQIIGGKNFTDDDGKEDAISDYNGHGTHVAGTIAANDSNGGIAGVAPEASLLIVKVL
B. amyloliquefaciens            LKNRIIGGKNFTDDDGKEDAFSDYNGHGTHVSGTIAANDSNGGISGVAPEASLLIVKVL
P. polymyxa                     LKARIIGGRNFTDDDEGDPEIFKDYNGHGTHVAGTIAATENENGVVGVAPEADLLIIKVL
B. halodurans                   LRDRIIGKHNVTSEDGNDPEIVSDQNGHGTHVCGTIAATENDRGVIGVAPECQLLVVKVL
Thermoactinomyces sp.           LRDRIIGKHNVTSDDGNDPEIVSDQNGHGTHVCGTIAATENDR-AIGVAPECQLLVVKVL

                               130       140       150       160       170       180
                                 |         |         |         |         |         |
C. difficile                    NKD-GTGTYQSIINAINFAVN----N--KVDIISMSLGGNKDDKNLKNAVMQAVKNNISV
B. subtilis                     GGENGSGQYEWIINGINYAV----EQ--KVDIISMSLGGPSDVPELKEAVKNAVKNGVLV
B. amyloliquefaciens            GGQDGSGKYEWIINGINYAV----EQ--KADIISMSLGGPSDVPELKEAVTNAVKVGSLV
P. polymyxa                     NKQ-GSGQYDWIIQGIYYAI----EQ--KVDIISMSLGGPEDVPELHEAVKKAVASQILV
B. halodurans                   SNR-GFGTTEWVVEGIRHAINWEGPNGEKVQVLSMSLGGKENDPRLHDAIKEAVASGRLV
Thermoactinomyces sp.           SNR-GFGTTEWVVEGIRHAINWEGPNGEKVQVLSMSLGGKENDPRLHDAIKEAVASGRLV

                               190       200       210       220       230       240
                                 |         |         |         |         |         |
C. difficile                    VCAAGNNGDGDSSTSEY-SYPASYAEVIEVGAINENYLVEKFSNSNTTIDLVAPGRNIIS
B. subtilis                     VCAAGNEGDGDERTEEL-SYPAAYNEVIAVGSVSVARELSEFSNANKEIDLVAPGENILS
B. amyloliquefaciens            VCAAGNEGDGDERTEELLSYPAAYNEVIAVGSVSVARKSSEFSNAGKEIDLVAPGENILS
P. polymyxa                     MCAAGNEGDGDDRTDEL-GYPGCYNEVISVGAINFDRHASEFSNSNNEVDLVAPGEDILS
B. halodurans                   VCAAGNDGDGNEETDEF-AYPGAYPEVVQVGSVSLSGEISRFSNSNCKIDLVAPGEKILS
Thermoactinomyces sp.           VCAAGNDGDGNEETDEF-AYPGAYPEVVQVGSVSLSGEISRFSNSNCKIDLVAPGEKILS

                               250       260       270       280       290       300
                                 |         |         |         |         |         |
C. difficile                    TYMDNKLAIMSGTSMSAPYVSGSLALIKEWAREEFERDLDEAELYAQLIKCTRALGIPRT
B. subtilis                     TLPNKKYGKLTGTSMAAPHVSGALALIKSYEEESFQRKLSESEVFAQLIRRTLPLDIAKT
B. amyloliquefaciens            TLPNHKYGKLTGTSMAAPHVSGALALIKGLEQASFQRTLSEAEVYAQLVRRTLPLDIAKT
P. polymyxa                     TVPGGKYATFSGTSMATPHVAGALALIKQLANASFERDLTEPELYAQLIKRTIPLGNSPK
B. halodurans                   TYPGDKFATLTGTSMATPHVTGAAALLIEKFEREFERKITEPELFAQLIKRTVSLSYSRK
Thermoactinomyces sp.           TYPGDKFATLTGTSMATPHVTGAAALLIEKFEREFERKITEPELFAQLIKRTVSLSYSRK

                               310       320       330
                                 |         |         |
C. difficile                    EQGNGYLYLNLYK---YKNNSKR----------
B. subtilis                     LAGNGFLYLTAPD---ELAEKAEQSHLLTL------
B. amyloliquefaciens            LAGNGFLYLDAPD---VLMEKAEQA----------
P. polymyxa                     MEGNGLLYLTAVEELSRIFDTQRVAGILSTASLKVK
B. halodurans                   LQGNGLLKLTSGG--SRISETEDR---IETA-----
Thermoactinomyces sp.           LQGNGLLKLTSGG--SRISETEDR---IETA-----
```

Figure 4.22 : CLUSTAL W alignment of FlxD with flavodoxin proteins.

The amino acid sequence of FlxD from *C. difficile* was aligned with the flavodoxin proteins from *C. beijerinckii* (P00322), *M. elsdenii* (P00321), *T. pallidum* (O83895), *D. salexigens* (P18086) and *D. desulfuricans* (P86312). Identical amino acids are shown in red, highly similar amino acids are shown in green, weakly similar amino acids are shown in blue. Residues 6 to 22, which are part of the binding site for the flavin molecule are underlined.

```
                        10        20        30        40        50        60
                        |         |         |         |         |         |
C. difficile            MSKIYIVYWSGTGNTEKMANFVAEGVKLKGKTPEVLDVSLLKPSDLKE-EDKFALGCPSM
C. beijerinkii          ---MKIVYWSGTGNTEKMAELIAKGIIESGKDVNTINVSDVNIDELLN-EDILILGCSAM
M. elsdenii             --MVEIVYWSGTGNTEAMANEIEAAVKAAGADVESVRFEDTNVDDVAS-KDVILLGCPAM
T. pallidum             MAKVAVIFWSGTGHTETMARCIVEGLNVGGAKADLFSVMDFDVGTFDS-YDRFAFGCSAA
D. salexigens           MSKSLIVYGSTTGNTETAAEYVAEAFENKEIDVELKNVTDVSVADLGNGYDIVLFGCSTW
D. desulfuricans        MSKVLILFGSSTGNTESIAQKLEELVAAGGHEVTLLNAAEASADNLADGYDAVLMGCSAW

                        70        80        90        100       110       120
                        |         |         |         |         |         |
C. difficile            GAEQLE-EGDMEPFVSELESM-VSGKQIGLFGSYGWGN----CEWMRDWEERMQNAGATI
C. beijerinkii          GDEVLE-ESEFEPFIEEISTK-ISGKKVALFGSYGWGD----GKWMRDFEERMNGYGCVV
M. elsdenii             GSEELE-DSVVEPFFTDLAPK-LKGKKVGLFGSYGWGS----GEWMDAWKQRTEDTGATV
T. pallidum             GSEELE-SSEFEPFFTSIEGR-LSGKKVALFGSYEWAGEGEGGEWMVNWVERCKAAGADV
D. salexigens           GEEEIELQDDFIPLYDSLENADLKGKKVSV[illegible]GDSDYTY-FCGAVDAIEEKLEKMGAVV
D. desulfuricans        GMEDLELQDDFAPLFDEMENMGLKGKKLA[illegible]ASGDMEYEH-YCGAVPAIEEKAYGLGAEV

                                130       140       150
                                |         |         |
C. difficile            IGGEGITAMEDPNEEAKDECIELGKTLAE--
C. beijerinkii          VETP-LIVQNEP-DEAEQDCIEFGKKIANI-
M. elsdenii             IGTA--IVNEMP--DNAPECKELGEAAAKA-
T. pallidum             FEGKGEIAYDDPSEEAQASCKAFGERFAR--
D. salexigens           IGDSLKIDGDP----ERDEIVSWGSGIADKI
D. desulfuricans        IPEGLKIEGDAS--SDPDAVSAFAEDVLK--
```

levels of identity suggest that FlxD is likely to be a flavodoxin protein and therefore the ORF which encodes this protein is referred to as *flxD*, which stands for **fl**avodoxin from *C. difficile*.

Flavodoxins proteins function as electron carriers between other oxidation-reduction enzymes (Geoghegan *et al.*, 2000). These proteins are small and acidic and in many reactions they substitute efficiently for ferrodoxin. However, in contrast to the ferrodoxins, which contain iron and acid-labile sulfides, flavodoxins utilize a molecule of flavin mononucleotide (FMN) as their redox-active component (Mayhew and Ludwig, 1975). From comparative analysis of the amino acid sequences, it appears the region nearer the N-terminus of these proteins, which is part of the FMN binding site), is highly conserved during evolution, but farther along in the C-terminal region of these proteins similarities are harder to find (Mayhew and Ludwig, 1975). Residues 6 through to 22 are also highly conserved, as this region forms part of the binding site for the phosphate portion of the flavin molecule (Figure 4.22).

Delineation and genetic analysis of the putative conjugative transposon Tn*5398*

The sequence data obtained did not resolve the question as to whether Tn*5398* was a conjugative transposon or a mobilizable element. While the sequenced region did contain three ORFs with similarity to ORFs from the conjugative transposon Tn*916*, *orf13*, *orf9* and *orf7*, it did not appear to contain ORFs that would encode transposase, integrase, resolvase or mobilization proteins, which are proteins normally involved in the movement of conjugative or mobilizable elements. The aim of the experiments presented in the following section was to

obtain more information about the precise region that was transferred from the donor to the recipient, and about the way in which the element may be transferred.

*a) Transfer of Tn*5398 *to* C. difficile *and* B. subtilis

To obtain transconjugants that could be used to determine the precise DNA region that was transferred during conjugation, Tn*5398* was transferred from the erythromycin and tetracycline resistant strain 630 to the rifampicin resistant strain CD37 by filter mating (Chapter 2). Briefly, aliquots of strain 630 and strain CD37 cells, which had been grown to mid-exponential phase under antibiotic selection, were mixed together on nitrocellulose filters placed on BHIS agar. The cells were incubated anaerobically at 37°C overnight to allow transfer of genetic material. The cells were then resuspended in fresh BHIS medium and plated on to selective media to isolate potential transconjugants. In initial mating studies the transconjugants, with one exception, were resistant to rifampicin and erythromycin (e.g. JIR1162). The exception was a single transconjugant, JIR1164, which was resistant to rifampicin, erythromycin and tetracycline. In subsequent experiments all of the transconjugants were resistant to rifampicin, erythromycin and tetracycline (e.g. JIR1181 and JIR1184).

The association between the transfer of erythromycin resistance and tetracycline resistance, which in strain 630 involves the *tet*(M) gene carried on the conjugative transposon Tn*5397* (Mullany *et al.*, 1990), was an interesting finding and suggested that Tn*5398* may be dependent upon Tn*5397* for its conjugative transfer. To determine whether Tn*5397* was required for the conjugative transfer of the Tn*5398* element, attempts were made to construct a system in which the conjugative transfer of Tn*5398* could be examined in the presence and absence of Tn*5397*. If

Tn*5398* transfer depended on the presence of Tn*5397*, strains containing only Tn*5398* would be unable to transfer erythromycin resistance to a suitable recipient. However, those strains that carried both Tn*5398* and Tn*5397* should be able to transfer erythromycin resistance either independently or together with the transfer of tetracycline resistance. Although isogenic CD37-derived Tn*5398* donors that did or did not carry Tn*5397* were available for use in mating experiments, there was no suitable recipient strain available. Despite extensive attempts to mutate CD37 to generate a second selectable marker in this strain, and searches of the large *C. difficile* strain collection in our laboratory, no suitable recipient strain could be constructed or identified.

Since it had been reported that Tn*5398* could be transferred from *C. difficile* to appropriate *B. subtilis* recipient strains (Mullany *et al.*, 1995), attempts were made to use this organism as a recipient. The advantage of using *B. subtilis* is that no additional antibiotic resistance marker is required because the ability to grow aerobically can be used as the selective marker. Preliminary conjugative transfer experiments were performed as described previously (Mullany *et al.*, 1995), using *C. difficile* strain 630 as the donor and *B. subtilis* strain CU2189 as the recipient. These experiments were performed numerous times but unfortunately no transconjugants were obtained.

b) Conjugative mobilization using* E. coli *strain S17-1

Most conjugative and mobilizable transposons excise from the donor chromosome and form a non-replicating circular intermediate, which is then transferred from the donor to the recipient. A functional *mob* gene is generally required, the product of which nicks the circular molecule at an *oriT* site. Once

nicked a single strand of the circular intermediate can be transferred to the recipient cell by conjugation.

The possibility that Tn*5398* was a mobilizable transposon rather than a conjugative transposon was, to some extent, investigated in *E. coli*. The *E. coli* strain S17-1 contains a copy of the broad host range plasmid RP4 integrated into the chromosome (Simon *et al.*, 1983). This strain is capable of mobilizing co-resident mobilizable plasmids with a compatible *oriT* site, including elements of clostridial origin (Crellin and Rood, 1998; Lyras *et al.*, 1998). If Tn*5398* contains a compatible *oriT* site or *mob* gene, then when the recombinant plasmid pJIR1790 is introduced into strain S17-1 it should be possible to mobilize the plasmid to an appropriate *E. coli* recipient strain. To this end, three isogenic strains were constructed, the negative control strain S17-1(pWSK29), the test strain S17-1(pJIR1790), and a second test strain S17-1(pJIR1790, pJIR1537). The latter plasmid carries a copy of the site-specific recombinase gene, *tndX*, from Tn*5397* (Wang *et al.*, 2000a). This strain was included to allow for the possibility that the addition of *tndX in trans* may excise Tn*5398*, and allow mobilization of the resultant circular form. The positive control strain used in the experiment was S17-1(pJIR1377), which contains a copy of the chloramphenicol resistance conjugative transposon Tn*4453a* from *C. difficile* (Lyras *et al.*, 1998).

These strains were used as donors in conjugation experiments with the recipient strain LT101, a rifampicin resistant derivative of HB101. Transconjugants that were resistant to both chloramphenicol and rifampicin were observed from the positive control mating at a level of approximately 1×10^{-2} chloramphenicol resistant colonies per donor cell. No transconjugants resistant to erythromycin were obtained

from any of the matings, indicating that it is unlikely that Tn*5398* carries *oriT* or *mob* genes that are compatible with the RP4 mobilization system. This result does not, however, exclude the possibility that the transposon could be mobilized using an *oriT* site or *mob* genes associated with a different mobilization system.

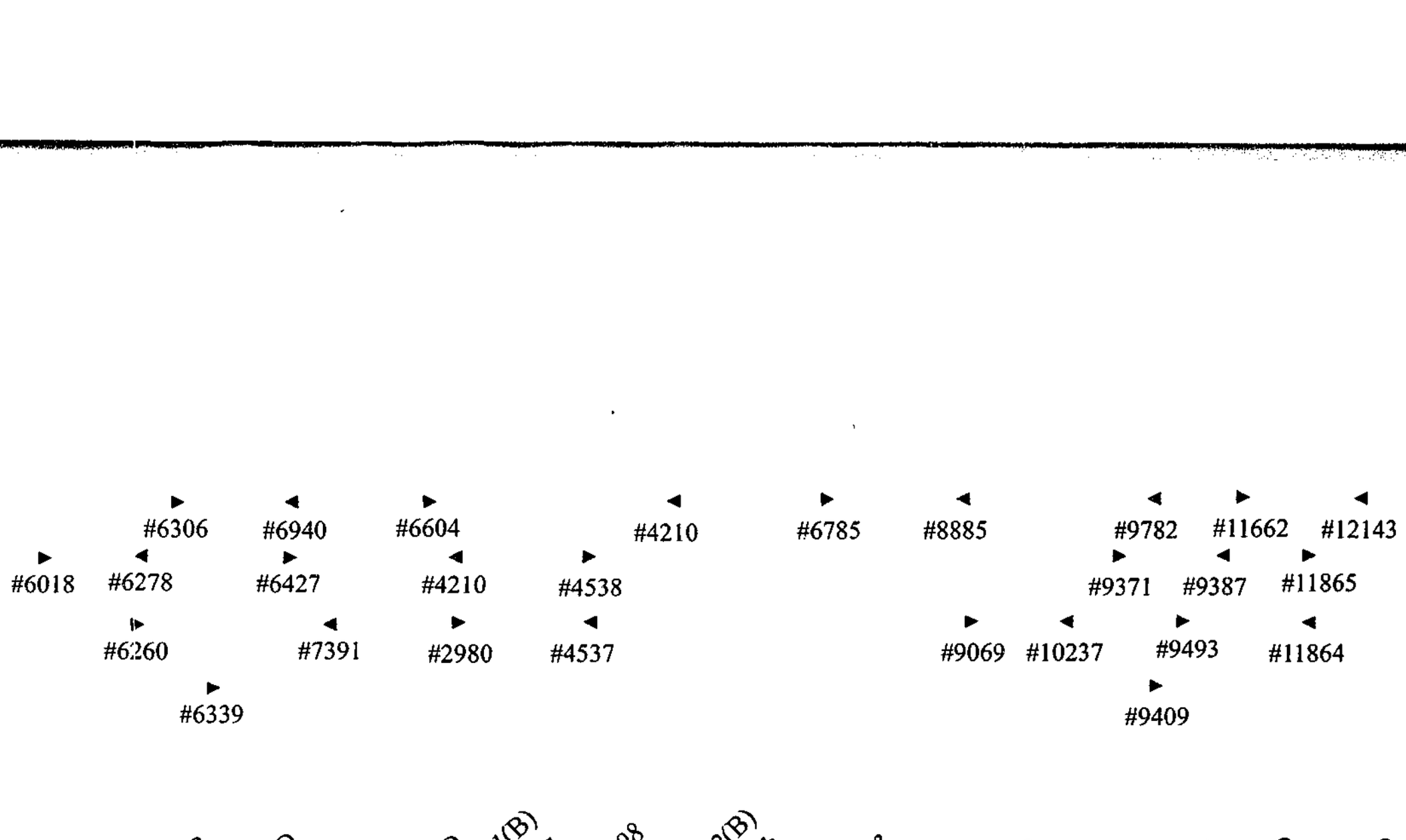
c) Delineation of the putative transposon using dot blot hybridization analysis

To delineate the extent of the Tn*5398* element, a series of DNA-DNA hyrbridization experiments were performed using probes generated against a selection of genes identified in the earlier part of this chapter. Each of the hybridization experiments included DNA extracted from *C. difficile* strain 630 (positive control), from *C. difficile* strain CD37 (which is MLS sensitive and therefore represents the negative control), and from four independently derived, MLS resistant, transconjugant strains JIR1162, JIR1164, JIR1181 and JIR1184. These transconjugants were derived by conjugative transfer of erythromycin resistance from strain 630 to strain CD37.

DIG-labelled DNA probes that contained internal portions of the *ilvD*, *hydD*, *erm*(B), *orf13*, *effD*, and *ispD* genes were generated by PCR using the oligonucleotide primer pairs #6018 and #6278, #6339 and #6940, #2980 and #2981, #6019 and #6785, #9069 and #10237, and #11546 and #11864, respectively (Figure 4.23). The rationale behind this approach was that if a gene was of chromosomal or housekeeping (i.e. non-transposon) origin the probe would be expected to bind to the DNA from all of the strains tested. However, if the gene was of transposon origin, we would expect the probe to bind only to the positive control and the transconjugant strains.

Figure 4.23 : Oligonucleotide primers used to delineate Tn*5398*.

The number and position of each oligonuclotide primer is shown above a schematic representation of the Tn*5398* region from *C. difficile* strain 630. The polarity of each oligonucleotide primer is indicated by the small, black arrow heads. Colored block arrows indicate each of the ORFs and their respective direction of transcription. Grey shaded boxes represent regions encompassing DR sequences.



The results (Figure 4.24) showed that the *ilvD*-specific, *hydD*-specific and *ispD*-specific probes hybridized to DNA from the positive control, the negative control and the transconjugant strains, indicating that these genes were likely to be of a chromosomal (i.e. non-transposon) origin. The *erm*(B)-specific, *orf13*-specific, and *effD*-specific probes bound only to the DNA from the positive control and the transconjugants, indicating that these genes are likely to be of transposon origin. This analysis therefore allowed us to delineate the transposon as encompassing the region between the *hydD* and *ispD* genes.

d) Delineation of the putative transposon using PCR analysis

To more precisely delineate the putative transposon, a PCR based strategy was used to determine the presence or absence of each of the genes previously identified in these strains. Once more, the rationale was based on the fact that a PCR product should be amplified from the DNA of all strains tested if the gene was of a non-transposon origin, and should only be amplified from strain 630 and the transconjugant strains if the gene was of transposon origin.

A series of twelve PCRs were performed, which gave products spanning the entire region previously sequenced (Table 4.2, Figure 4.23). The binding positions of each of the oligonucleotide primers used are shown (Figure 4.23). The results (Table 4.2), showed that a PCR product of the expected size was amplified from all six strains for the *ilvD*, *hydR-hydD*, *ispD* and *ispD-flxD* reactions, which indicates that these gene regions are chromosomally located and outside the region occupied by the putative transposon. PCR products of the expected sizes for the *elpD-erm1*(B), *erm1*(B)-*orf298*, *orf298-erm2*(B), *orf13*, *effR*, *effD*, *orf9*, and *orf7*

Figure 4.24 : Dot blot hybridization analysis of the Tn*5398* region.

Dot blot hybridizations were conducted on chromosomal DNA from *C difficile* strains 630 (A2), CD37 (A3) and the transconjugants JIR1162 (B1), JIR1164 (B2), JIR1181 (B3) and JIR1184 (B4) using *ilvD*, *hydD*, *erm*(B), *orf13*, *effD*, and *ispD* specific probes.

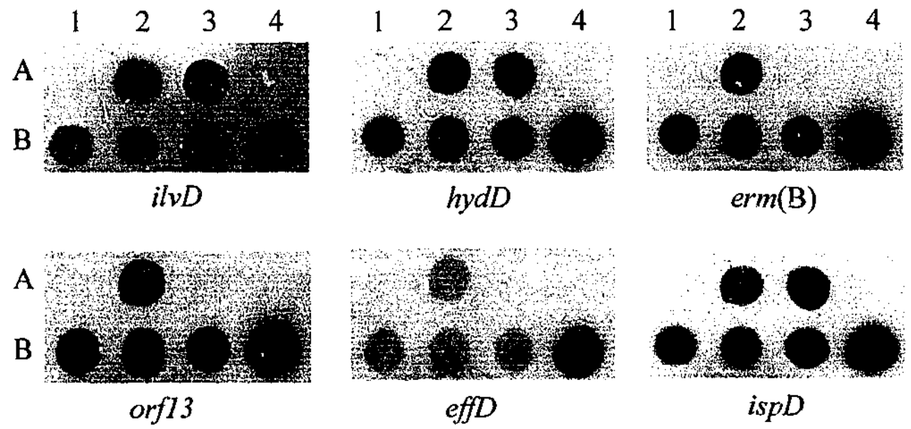


Table 4.2 : Delineation of the conjugative transposon Tn*5398* by PCR analysis*.

Gene/s Amplified	Oligonucleotides Used in PCR	630	CD37	Transconjugants**
ilvD	#6018 and #6278	+	+	+
hydR and *hydD*	#6306 and #6940	+	+	+
elpD and *erm1*(B)	#6604 and #4210	+	-	+
erm1(B) and *orf298*	#2980 and #4537	+	-	+
orf298 and *erm2*(B)	#4538 and #4210	+	-	+
orf13	#6785 and #6019	+	-	+
effR	#7774 and #8885	+	-	+
effD	#9069 and #10237	+	-	+
orf9	#9371 and #9782	+	-	+
orf7	#9493 and #9387	+	-	+
ispD	#11662 and #11864	+	+	+
ispD to *flxD*	#11865 and #12143	+	+	+

* (+) indicates a PCR product of the expected size was obtained.
(-) indicates that no PCR product was detected.

** *C. difficile* strains JIR1162, JIR1164, JIR1181 and JIR1184.

reactions were only amplified from strain 630 and the transconjugants, indicating that these genes or regions are likely to be contained within the putative transposon.

These results allowed a more precise delineation of the putative transposon. The left end of the element appeared to lie in the intergenic space between the *hydD* and *elpD* genes; upstream of the Erm leader peptide, but downstream of the #6940 oligonucleotide primer binding site (Figure 4.23), since a PCR product was obtained for the *hydR-hydD* PCR from strain CD37. The right end of the element appeared to lie upstream of the *ispD* gene, and upstream of the #9387 oligonucleotide primer binding site, since an *orf7* PCR product was not amplified from strain CD37 using the oligonucleotide primers #9493 and #9387, but an *ispD* PCR product was amplified from the same strain using the oligonucleotide primers #11662 and #11864 (Figure 4.23, Table 4.2). These results suggested that Tn*5398* was less than 10 kb in size.

*e) Precise delineation of the ends of Tn*5398

To precisely define the ends of Tn*5398*, and to determine the target or insertion site, the regions encompassing both ends were amplified from chromosomal DNA from strain 630, and from each of the transconjugant strains, using the oligonucleotide primers #9409 and #9387 for the right end, and oligonucleotide primers #7391 and #6427 for the left end. The potential target region from strain CD37 was amplified from chromosomal DNA using the oligonucleotide primers #6260 and#12143. The resultant PCR products were then sequenced to determine the nucleotide sequence of the ends of the transposon and the point of insertion in strain CD37.

i) The left end

Comparison of the nucleotide sequences obtained from the left end junction PCR products from strain 630 and the transconjugant strains with the sequence obtained in the same region from strain CD37 showed that the left end of the element was located 272 bp downstream of the *hydD* stop codon (Figure 4.25). Prior to this point the nucleotide sequences of all of the strains were almost identical, and, unexpectedly, the nucleotide sequence of the transconjugants more closely resembled the nucleotide sequence of strain 630. However, past this point the nucleotide sequences of strain 630 and the transconjugants diverged from that of strain CD37, indicating that at this point the flanking sequence ends and the transposon sequence begins. Interestingly, the sequence of the left hand end of the putative transposon was a palindromic sequence consisting of the nucleotides TTTTTATATAAAAA (Figures 4.25 and 4.27), which may be of significance in the transposition or mobilization of this putative element.

ii) The right end

Comparison of the nucleotide sequences obtained from the equivalent PCR products from the other end of the element revealed that the right end point was located within *orf7*, 84 bp upstream of the stop codon (Figure 4.26). The right end was an imperfect palindromic sequence that consisted of the nucleotides TTTTTATAATAAAAA (Figures 4.26 and 4.27) This sequence was identical to the palindromic sequence found at the left hand end of the element, except for the insertion of an A nucleotide in the central region.

The delineation of the right end of the transposon was not as clear as that of the left end. There was considerable similarity upstream of the proposed right end of

Figure 4.25 : Delineation of the left end of Tn*5398*.

CLUSTAL W alignment of the nucleotide sequences obtained from PCR products encompassing the left end of the transposon amplified from chromosomal DNA from strain 630 and each of the transconjugants using the oligonucleotide primers #7391 and #6427, and the potential target region of the transposon from strain CD37 amplified from chromosomal DNA using the oligonucleotide primers #6260 and #12143. Identical nucleotides are shown in red. The boxed region indicates the palindromic sequence at the junction point. The position of the *hydD* stop codon is shown.

```
                    10        20        30        40        50        60
                     |         |         |         |         |         |
                                                   hydD stop |
630          TATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCT
JIR1162      TATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCT
JIR1164      TATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCT
JIR1181      TATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCT
JIR1184      TATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCT
CD37         TATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTTAAATAAGTAAACACCT

                    70        80        90       100       110       120
                     |         |         |         |         |         |
630          TTGTTAAAAATATACATATGAAGATTGGAAATTTAATGTTAAAAATAGAAACATGAAAAT
JIR1162      TTGTTAAAAATATACATATGAAGATTGGAAATTTAATGTTAAAAATAGAAACATGAAAAT
JIR1164      TTGTTAAAAATATACATATGAAGATTGGAAATTTAATGTTAAAAATAGAAACATGAAAAT
JIR1181      TTGTTAAAAATATACATATGAAGATTGGAAATTTAATGTTAAAAATAGAAACATGAAAAT
JIR1184      TTGTTAAAAATATACATATGAAGATTGGAAATTTAATGTTAAAAATAGAAACATGAAAAT
CD37         TTGTTAAAAATATACATATGAAGATTGGAAATTTAATGTTAAAAATAGAAACATGAAAAT

                   130       140       150       160       170       180
                     |         |         |         |         |         |
630          ATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTAATACATTATCTACTATAAAT
JIR1162      ATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTAATACATTATCTACTATAAAT
JIR1164      ATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTAATACATTATCTACTATAAAT
JIR1181      ATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTAATACATTATCTACTATAAAT
JIR1184      ATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTAATACATTATCTACTATAAAT
CD37         ATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTGATACATTATCTACTATAAAT

                   190       200       210       220       230       240
                     |         |         |         |         |         |
630          ACAAATATAGCTTCAATGTGATTATATATTGTTGTATTGGTAAAGCACTTATACAAACAG
JIR1162      ACAAATATAGCTTCAATGTGATTATATATTGTTGTATTGGTAAAGCACTTATACAAACAG
JIR1164      ACAAATATAGCTTCAATGTGATTATATATTGTTGTATTGGTAAAGCACTTATACAAACAG
JIR1181      ACAAATATAGCTTCAATGTGATTATATATTGTTGTATTGGTAAAGCACTTATACAAACAG
JIR1184      ACAAATATAGCTTCAATGTGATTATATATTGTTGTATTGGTAAAGCACTTATACAAACAG
CD37         ACAAATATAGCTTCAATGTGATTATATATTAGTGTATTGGTAAAGCACTTATACAAACAG

                   250       260       270       280       290       300
                     |         |         |         |         |         |
630          AGGAATTTTGTAAATTCAGATTATATCCACATTTGTTAACTTATGAAAATATAATCAAAA
JIR1162      AGGAATTTTGTAAATTCAGATTATATCCACATTTGTTAACTTATGAAAATATAATCAAAA
JIR1164      AGGAATTTTGTAAATTCAGATTATATCCACATTTGTTAACTTATGAAAATATAATCAAAA
JIR1181      AGGAATTTTGTAAATTCAGATTATATCCACATTTGTTAACTTATGAAAATATAATCAAAA
JIR1184      AGGAATTTTGTAAATTCAGATTATATCCACATTTGTTAACTTATGAAAATATAATCAAAA
CD37         AGGAATTTTGTAAATTCATATTATACCCACATTTGTTAACTTATGAAAATATAATCAAAA

                   310       320       330       340       350       360
                     |         |         |         |         |         |
630          TTTTTATGAGCTTTTTTATATAAAAAAACGCCCTAAAAATCTGATTATCCCCATAAACACT
JIR1162      TTTTTATGAGCTTTTTTATATAAAAAAACGCCCTAAAAATCTGATTATCCCCATAAACACT
JIR1164      TTTTTATGAGCTTTTTTATATAAAAAAACGCCCTAAAAATCTGATTATCCCCATAAACACT
JIR1181      TTTTTATGAGCTTTTTTATATAAAAAAACGCCCTAAAAATCTGATTATCCCCATAAACACT
JIR1184      TTTTTATGAGCTTTTTTATATAAAAAAACGCCCTAAAAATCTGATTATCCCCATAAACACT
CD37         TTTTTATGAGATTTTTTATATAAAAAATATAATGTTTGTATACAAAAATATTGAAGATATT

                   370       380       390
                     |         |         |
630          GTATCTACAAGCATATTCAATAGGAAATAA
JIR1162      GTATCTACAAGCATATTCAATAGGAAATAA
JIR1164      GTATCTACAAGCATATTCAATAGGAAATAA
JIR1181      GTATCTACAAGCATATTCAATAGGAAATAA
JIR1184      GTATCTACAAGCATATTCAATAGGAAATAA
CD37         TTAGTAAGTTTTGTATATAAGCAAACATGT
```

Figure 4.26 : Delineation of the right end of Tn*5398*.

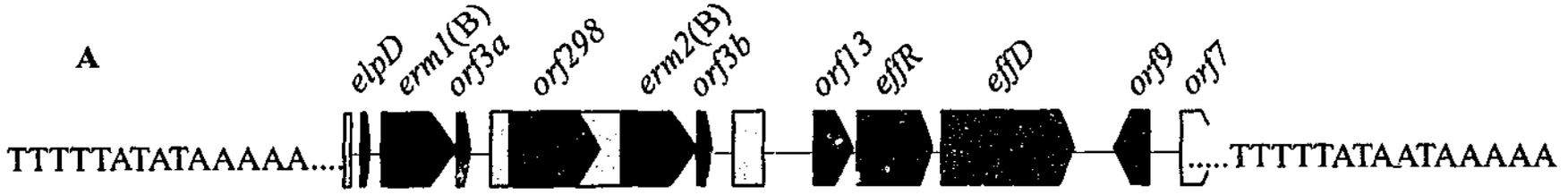
CLUSTAL W alignment of the nucleotide sequences obtained from PCR products encompassing the right end of the transposon amplified from chromosomal DNA from strain 630 and each of the transconjugants using the oligonucleotide primers #9409 and #9387, and the potential target region of the transposon from strain CD37 amplified from chromosomal DNA using the oligonucleotide primers #6260 and #12143. Identical nucleotides are shown in red. The boxed region indicates the palindromic sequence at the junction point. Gaps introduced in the alignment are shown as dashed lines. The *orf7* stop codon is indicated.

```
                  10         20         30         40         50         60
                   |          |          |          |          |          |
630          ---AAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGT----AAGCTC--
1162         ---AAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGT----AAGCTC--
1164         ---AAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGT----AAGCTC--
1181         ---AAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGT----AAGCTC--
1184         ---AAACCATCAGACTTCCAAAAGACGATACAGTGTCAGTTGGACTGT----AAGCTC--
CD37         ATTGGGACACGCTACATATGAAGAAGCACAGGATTTTAATGAAAGAGTTTTAGAGTTTTT

                  70         80         90        100        110        120
                   |          |          |          |          |          |
630          AAAAAGGT---------TGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTA--GCCA
1162         AAAAAGGT---------TGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTA--GCCA
1164         AAAAAGGT---------TGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTA--GCCA
1181         AAAAAGGT---------TGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTA--GCCA
1184         AAAAAGGT---------TGTAAAAGGCAGTGTCCGTAACTACTGCAAGGAATTA--GCCA
CD37         AAATAAGTAAACACCTTTGTTAAAAATA-TACATATGAAGATTGGAA--ATTTAATGTTA

                 130        140        150        160        170        180
                   |          |          |          |          |          |
630          GACGACAGGCAAA-GGAAGTACCCTTTTGTGA-GCTTCCAGAAATTGTTATTGAGA---A
1162         GACGACAGGCAAA-GGAAGTACCCTTTTGTGA-GCTTCCAGAAATTGTTATTGAGA---A
1164         GACGACAGGCAAA-GGAAGTACCCTTTTGTGA-GCTTCCAGAAATTGTTATTGAGA---A
1181         GACGACAGGCAAA-GGAAGTACCCTTTTGTGA-GCTTCCAGAAATTGTTATTGAGA---A
1184         GACGACAGGCAAA-GGAAGTACCCTTTTGTGA-GCTTCCAGAAATTGTTATTGAGA---A
CD37         AA-AATAGAAACATGAAAATATGCTTAACTGGTATTTTTACTATTCATAACCAATTTTTG

                 190        200        210        220        230        240
                   |          |          |          |          |          |
630          ATTGATTGTCTGGGATGATTACGAAAGTGACT---ATACGACATTCGATGTGTGCAGTAT
1162         ATTGATTGTCTGGGATGATTACGAAAGTGACT---ATACGACATTCGATGTGTGCAGTAT
1164         ATTGATTGTCTGGGATGATTACGAAAGTGACT---ATACGACATTCGATGTGTGCAGTAT
1181         ATTGATTGTCTGGGATGATTACGAAAGTGACT---ATACGACATTCGATGTGTGCAGTAT
1184         ATTGATTGTCTGGGATGATTACGAAAGTGACT---ATACGACATTCGATGTGTGCAGTAT
CD37         ATACATTATCTACTATAAATACAAATATAGCTTCAATGTGATTATATATTAGTGTATTGG

                 250        260        270        280        290        300
                   |          |          |          |          |          |
630          GGAAATCCGTGTGCTTGATGAAGAACTTGAAAAATACAGGATATATCGGCATTTACCAGC
1162         GGAAATCCGTGTGCTTGATGAAGAACTTGAAAAATACAGGATATATCGGCATTTACCAGC
1164         GGAAATCCGTGTGCTTGATGAAGAACTTGAAAAATACAGGATATATCGGCATTTACCAGC
1181         GGAAATCCGTGTGCTTGATGAAGAACTTGAAAAATACAGGATATATCGGCATTTACCAGC
1184         GGAAATCCGTGTGCTTGATGAAGAACTTGAAAAATACAGGATATATCGGCATTTACCAGC
CD37         TAAAGCACTTATACAAACAGAGGAATTTTGTAAATTCATATTATACCCACATTTGTTAAC

                 310        320        330        340        350        360
                   |          |          |          |          |          |
630          TTATGAAAAGATAATCAGAAATTTAGTGT-ATTTTTATAATAAAAATATAATGCTTGTAT
1162         TTATGAAAAGATAATCAGAAATTTAGTGT-ATTTTTATAATAAAAATATAATGCTTGTAT
1164         TTATGAAAAGATAATCAGAAATTTAGTGT-ATTTTTATAATAAAAATATAATGCTTGTAT
1181         TTATGAAAAGATAATCAGAAATTTAGTGT-ATTTTTATAATAAAAATATAATGCTTGTAT
1184         TTATGAAAAGATAATCAGAAATTTAGTGT-ATTTTTATAATAAAAATATAATGCTTGTAT
CD37         TTATGAAAATATAATCAAAATTTTTATGAGATTTTTATATAAAAAATATAATGTTTGTAT

                 370        380        390        400        410        420
                   |          |          |          |          |          |
                                                              orf7 Stop |
630          ACAAAAATATTAAAGATATTTTAGTAAGTTTTGTATATAAGCAAACATGTATTTTTTAA
1162         ACAAAAATATTAAAGATATTTTAGTAAGTTTTGTATATAAGCAAACATGTATTTTTTAA
1164         ACAAAAATATTAAAGATATTTTAGTAAGTTTTGTATATAAGCAAACATGTATTTTTTAA
1181         ACAAAAATATTAAAGATATTTTAGTAAGTTTTGTATATAAGCAAACATGTATTTTTTAA
1184         ACAAAAATATTAAAGATATTTTAGTAAGTTTTGTATATAAGCAAACATGTATTTTTTAA
CD37         ACAAAAATATTGAAGATATTTTAGTAAGTTTTGTATATAAGCAAACATGTATTTTTTAA
```

Figure 4.27 : Schematic depiction of Tn*5398*.

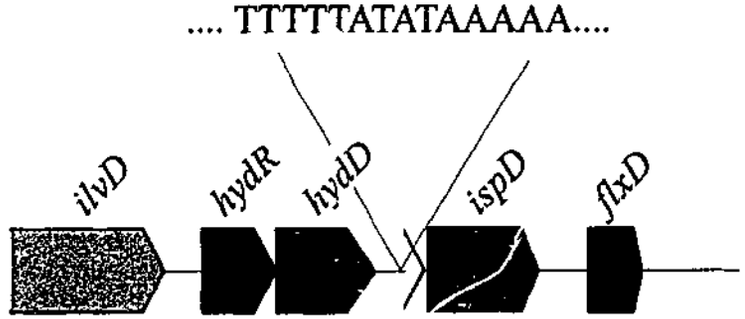
The schematic representations include (A) the putative conjugative transposon Tn*5398*, (B) the target site of Tn*5398* in the recipient strain and (C) Tn*5398* as present in strain 630 and the transconjugant strains. The palindromic sequences at the ends of the element are shown.

A



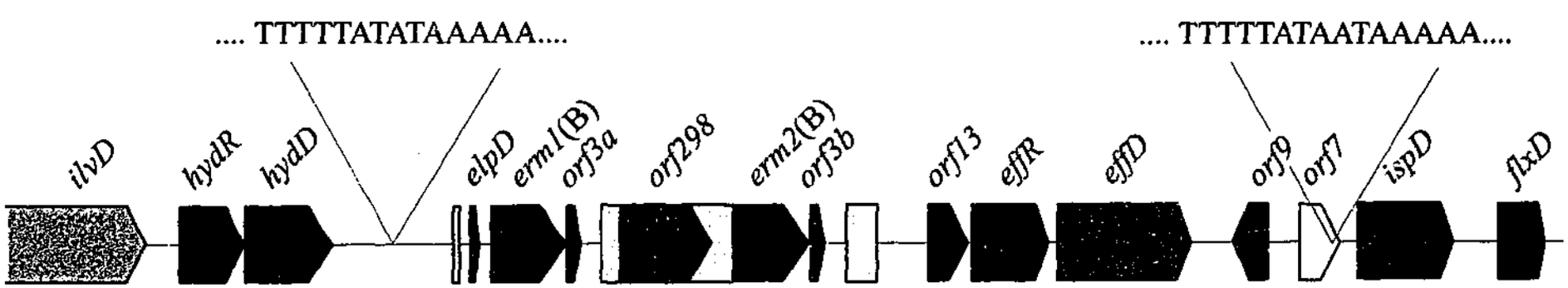
Putative Transposon

B



Recipient Strain

C



Donor and Transconjugant Strains

the transposon (Figure 4.26), which extended back through *orf7*, indicating that there was sequence identity between the target region in the recipient, strain CD37, and the right end of the putative transposon. RecA mediated homologous recombination events may be possible between these two sites, and may provide a method for the insertion of this putative element into the recipient chromosome. This hypothesis will be discussed further later in this chapter. This alignment also showed that, unexpectedly, downstream of the proposed right end of the element the nucleotide sequences of the transconjugants were more closely related to strain 630 than to strain CD37, which was also observed upstream of the proposed left end of the element.

Based on this putative delineation of the right and left ends of the element, Tn*5398* is 9630 bp in length. In addition to the *C. difficile* Erm B determinant it carries four complete ORFs, *orf13*, *effR*, *effD*, and *orf9*, and an incomplete ORF, *orf7*.

iii) The target region in strain CD37

Comparison of the nucleotide sequence obtained from the PCR product from strain CD37 and the sequences from the left and right end PCR products from strain 630 and the transconjugants, led to the conclusion that the target region was located in the intergenic space between the *hydD* gene and the *ispD* gene of the recipient (Figure 4.27). The target site appeared to be a palindromic sequence that was 227 bp downstream of the *hydD* stop codon and 208 bp upstream of the *ispD* start codon. This palindromic sequence consisted of the nucleotides TTTTTATATAAAAA, and was identical in sequence to the palindromic sequence found at the left end of the putative transposon and almost identical to the palindromic sequence found at the

right end of the transposon. The location of this target region, between the *hydD* and *ispD* genes, also leaves open the possibility of a homologous recombination event being involved in the insertion of the putative element, as large amounts of homologous DNA are present between the target region of the recipient strain and the regions both upstream and downstream of the putative transposon in the strain 630 genome.

Discussion

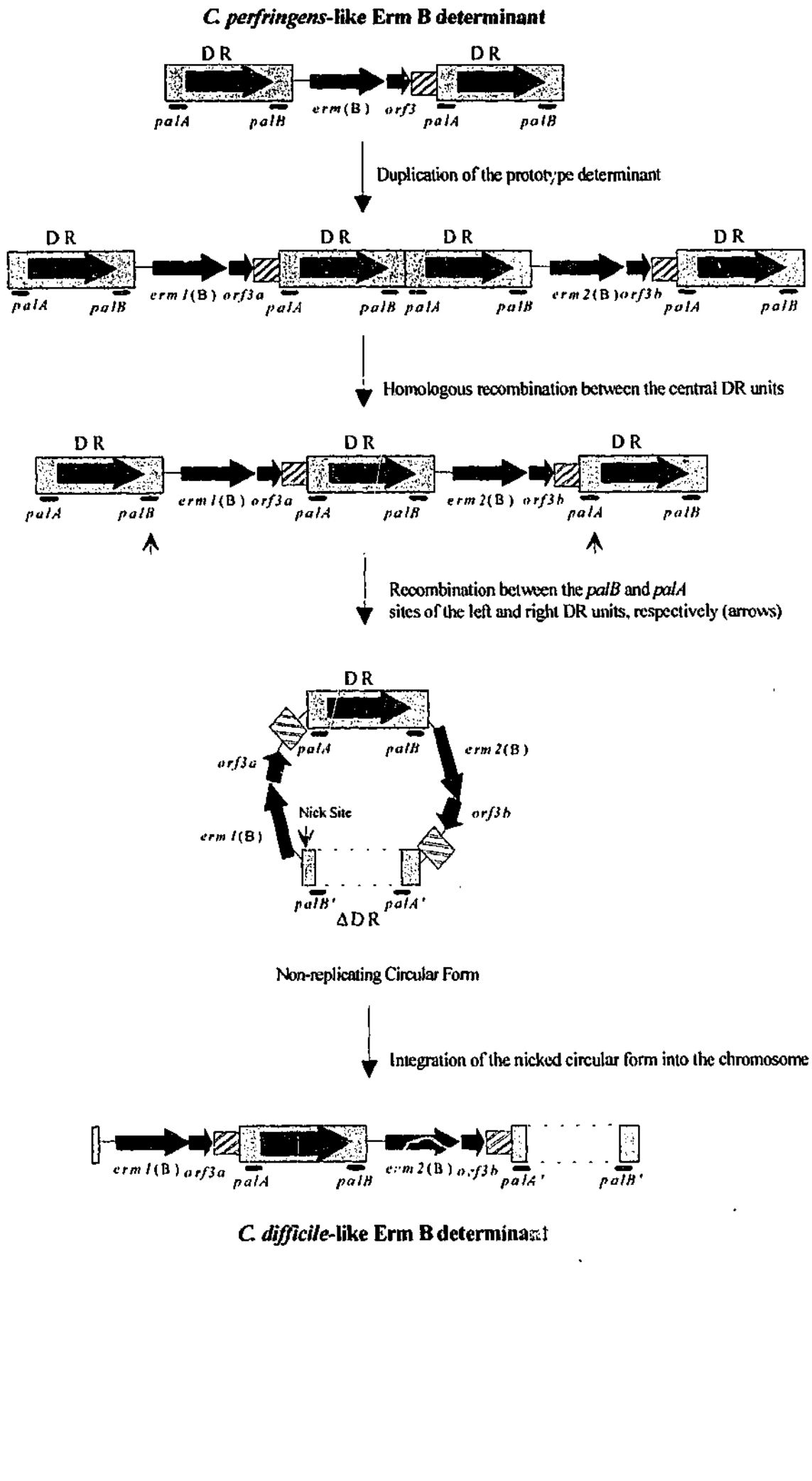
The data reported in this chapter reveals that the Erm B determinant found in *C. difficile* strain 630 and carried on Tn*5398*, is the first member of the Erm B class of MLS resistance determinants to have two *erm*(B) structural genes. Previous workers in this laboratory postulated that the Erm B determinant found in *C. perfringens* represents the Erm B progenitor as it is the only Erm B determinant in which the *erm*(B) structural gene is flanked by two complete copies of the DR sequence (Berryman and Rood, 1995). Furthermore, it was proposed that other Erm B determinants have evolved from this progenitor by homologous recombination events that removed part of the DR sequences. This analysis also showed that most of the *erm*(B) genes are flanked by complete or deleted (ΔDR) variants of the DR sequence. The fact that each of the deleted variants of the DR sequence is different, with each deletion event having occurred at slightly different locations in the palindromic sequences *palA* and *palB*, suggests that homologous recombination is responsible for these deletions rather than site-specific recombination events (Berryman and Rood, 1995).

The evolution of the Erm B determinant from *C. difficile* strain 630 can be explained in terms of this progenitor hypothesis (Figure 4.28). A duplication of the Erm B determinant from *C. perfringens* would first be necessary to create two copies of the progenitor, arranged in a directly repeated orientation. A homologous recombination event between the two central DR sequences would effectively remove one of these DR sequences, leaving two *erm*(B) structural genes separated by a single complete copy of the DR sequence. If recombination was then to occur between the *palB* sequence from the DR sequence upstream of the first *erm*(B) structural gene, and the *palA* sequence from the DR sequence downstream of the second *erm*(B) structural gene, the intervening DNA could be removed as a non-replicating circular molecule. Nicking of this circular molecule, at a site 70 bp from the end of the DR sequence downstream of the second *erm*(B) structural gene, and integration of the nicked circular form into the chromosome, could result in the formation of an Erm B determinant like that found in Tn*5398*. The loss of the promoter sequences upstream of *erm2*(B) and the acquisition of the Erm leader peptide upstream of *erm1*(B) would then be all that was required to complete the evolution of the strain 630 Erm B determinant.

Homologous recombination could also account for the loss of the promoter sequences upstream of *erm2*(B). Inducibly expressed *erm*(B) genes, which possess leader peptide sequences, do not usually have promoter sequences associated with the *erm*(B) gene as the promoter sequences are located upstream of the Erm leader peptide. However, because of the high level of homology between the upstream regions of both inducibly and constitutively expressed *erm*(B) genes (other than across the leader peptide region) and between the *erm*(B) genes themselves, it is possible that homologous recombination events between inducible and constitutive

Figure 4.28 : Proposed model for the evolution of the *C. difficile* strain 630 Erm B determinant.

Regions of nucleotide sequence similarity are indicated by similar coloring or shading. The positions of ORFs and their respective direction of transcription are shown as solid colored arrows. The approximate positions of the palindromic sequences *palA* and *palB* are shown as lines under the DR sequences.


C. perfringens-like Erm B determinant
DR
DR
ermB
orf3
palA
palB
palA
palB
Duplication of the prototype determinant
DR
DR
DR
DR
erm1(B)
orf3a
erm2(B)
orf3b
Homologous recombination between the central DR units
DR
DR
DR
Recombination between the palB and palA
sites of the left and right DR units, respectively (arrows)
DR
palA
palB
erm2(B)
orf3a
orf3b
Nick Site
erm1(B)
palB'
palA'
ΔDR
Non-replicating Circular Form
Integration of the nicked circular form into the chromosome
erm1(B)
orf3a
palA
palB
erm2(B)
orf3b
palA'
palB'
C. difficile-like Erm B determinant

Erm B determinants could lead to the deletion of the promoter sequences upstream of a constitutively expressed determinant.

The Erm leader peptide upstream of *erm1*(B) could have been acquired *via* transfer of genetic material between *C. difficile* and other organisms present in the intestinal flora. For example, *E. coli* is a common inhabitant of the gastrointestinal tract, and the Erm B determinant from this organism contains an Erm leader peptide sequence that is almost identical (one nucleotide change) to the Erm leader peptide sequence found upstream of *erm1*(B) in strain 630. *C. difficile* may have acquired this segment of DNA from *E. coli* by conjugative transfer, and subsequent homologous recombination events may have then allowed the integration of the DNA upstream of the *erm1*(B) structural gene. Alternatively, the original progenitor may have had a leader peptide sequence upstream of the *erm*(B) gene. This sequence may have been duplicated along with the *erm*(B) gene and the copy of the leader peptide upstream of *erm2*(B) may have been subsequently lost.

The results reported in this chapter showed that Tn*5398* was 9.6 kb in size and carried two *erm*(B) genes in addition to *orf3a*, *orf3b*, *orf298*, *effR*, *effD*, *orf13* and *orf9* and an incomplete ORF, *orf7*. The proteins encoded by *effR* and *effD* are unlikely to be involved in the movement of the putative transposon, as they show no homology to any proteins shown to have a role in transposition, mobilization, or conjugative transfer of known conjugative transposons.

Three ORFs, *orf13*, *orf9* and *orf7*, have similarity to ORFs from Tn*916*. While these ORFs have been studied in some detail, their functions are still unknown. However, the proteins encoded by both *orf9* and *orf7* are postulated to

have a regulatory role. The Tn*5398* homologues of these ORFs, if they have any role at all, therefore may encode proteins that are involved in regulation rather than encoding proteins that are involved in excision, mobilization, transposition, or integration.

The only other Tn*5398* ORFs that may encode proteins involved in the mobility of the element are *orf3a*, *orf3b* and *orf298*. The proteins encoded by *orf3a* and *orf3b* have no homology to other proteins in the database, other than to ORF3 homologues. ORFs that encode the ORF3 protein are commonly found in association with *erm*(B) genes, however, the function of the ORF3 protein is unknown. The *orf298* ORF encodes a protein that has very weak similarity to both replication proteins and proteins from the ParA and Soj families, as already discussed. ParA and Soj proteins generally have a role in the partitioning of plasmids and chromosomes during the replication cycle (Easter *et al.*, 1998; Sharpe and Errington, 1996). It appears unlikely that either ORF3 or ORF298 have a role in either the excision or integration of Tn*5398* but this possibility cannot be completely eliminated.

Therefore, on the basis of database homology searching, none of the proteins encoded by Tn*5398* appear likely to be involved in the basic processes associated with conjugative transposition. No ORFs with similarity to known *mob* genes, or transposase, integrase, resolvase or excisionase genes were detected. Based on these results, it appears that Tn*5398* is not a classical conjugative transposon.

The process of conjugative transfer of a chromosomally located genetic element involves three main steps; the excision of the element from the donor

chromosome to form a circular intermediate, the transfer of one strand of that intermediate from the donor cell to the recipient cell, and the integration of the genetic element into the genome of the recipient. Conjugative transposons generally encode proteins that facilitate each of these steps.

Excision of a conjugative element from the donor chromosome involves the action of site-specific recombinases of the integrase or resolvase families. However, analysis of Tn*5398* did not reveal the presence of any ORFs that would encode homologues of these proteins. Perhaps the excision of this element is mediated *in trans* by other conjugative elements present in the cell, or by other proteins encoded on the genome. In addition to Tn*5398*, strain 630 contains the conjugative tetracycline resistance transposon Tn*5397*. Tn*5397* is excised from the donor chromosome through the action of the large resolvase, TndX (Wang and Mullany, 2000), which is encoded on Tn*5397* by the *tndX* gene. It is possible that the TndX protein is also responsible for the excision of Tn*5398* from the donor chromosome, especially since in most instances both Tn*5397* and Tn*5398* appear to be co-transferred to the recipient cell. However, the sequences at the ends of these elements are different, which makes it unlikely that the TndX protein could recognize the ends of Tn*5398*, since large resolvases generally recognize specific target sequences. Note that the genome of *C. difficile* strain 630 does encode as many as five other large resolvases (M. Smith, personal communication), one of which may potentially be responsible for the excision of Tn*5398* from the chromosome.

Once excised from the chromosome, most classical conjugative transposons form a non-replicating circular intermediate. It is this circular molecule that is then

nicked and subsequently transferred by conjugation from the donor to the recipient cell. This process is generally dependent on the action of proteins encoded by transfer (*tra*) and mobilization (*mob*) genes present on the transposon. The process of conjugative transfer of the circular intermediate has been well studied in Tn*916*. It has been shown that this transposon encodes many genes that have a role in the conjugation process (Senghas *et al.*, 1988). Tn*916*, once excised from the donor chromosome, forms a circular intermediate that is then transferred to the recipient cell (Scott *et al.*, 1988). It is postulated that, during transfer, the circular intermediate is nicked at a functional *oriT* site that is located on the transposon (Jaworski and Clewell, 1995), followed by transfer of a single strand of the circular intermediate to the recipient cell, where it is replicated and then integrated into the recipient genome (Scott *et al.*, 1994).

In addition to being capable of catalyzing its own conjugative transposition Tn*916* has been shown to be capable of enhancing the transfer of another homologous conjugative transposon that is co-resident in the cell (Flannagan and Clewell, 1991) and also of mobilizing non-transferable plasmids (Jaworski and Clewell, 1995; Showsh and Andrews Jnr., 1999). Mobilization of non-transferable plasmids does not appear to be dependent on the plasmid possessing a functional *mob* gene, but does require the presence of a sequence similar to the *oriT* sequence present on Tn*916*. It is postulated that the same protein or proteins involved in nicking of the Tn*916* circular intermediate nicks similar sequences present on co-resident plasmids. Once nicked at the *nic* site, the plasmid then assumes a relaxed form that is capable of being transferred during conjugation (Showsh and Andrews Jnr., 1999). ORF23 of Tn*916* shows homology to the MbeA mobilization protein of ColE1 (Flannagan *et al.*, 1994) and is likely to be the Tn*916*-encoded mobilization

protein (Showsh and Andrews Jnr., 1999). This system appears to be dependent only on the co-resident plasmid possessing an *oriT* site that is homologous to that found on Tn*916* (Showsh and Andrews Jnr., 1999).

If the Tn*5398* element was able to be excised to form a circular intermediate, the resultant molecule would resemble a non-conjugative plasmid that lacks *mob* genes. Tn*916* is not present in strain 630, but this strain does contain Tn*5397*. Tn*5397* is closely related to Tn*916* and comparative analysis of the two elements has revealed that they have very similar conjugation regions (ORFs 15 to 23) but different insertion and excision modules, with *xis* and *int* in Tn*916* being replaced by *tndX* in Tn*5397* (Roberts *et al.*, 2001) (Figure 1.7). Furthermore, the *oriT* sites present on both elements are identical and the amino acid sequences of the ORF23 homologues have greater than 90% identity. It is therefore possible that if Tn*5398* is excised from the chromosome to form a circular intermediate, that intermediate could be nicked and transferred to the recipient through the action of proteins, such as ORF23, provided *in trans* by Tn*5397*.

Sequence analysis reveals that there are two potential *oriT* sites on Tn*5398*, each of which has similarity to the *oriT* sites on Tn*916* and Tn*5397* (Figure 4.29). The first of the two sites was located within the coding sequence of *orf298* (nucleotides 5817 to 5856, Figure 4.2). Although it had limited similarity across the *oriT* region, the *nic* site was completely conserved (Figure 4.29). The second *oriT* site was located in the intergenic space between *orf3b* and *orf13* (nucleotides 8375 to 8414, Figure 4.2). Again this possible *oriT* site had limited similarity but the *nic* site was completely conserved (Figure 4.29).

Figure 4.29 : CLUSTAL W alignment of the *oriT* sites from Tn*5398*, Tn*916*, and Tn*5397*.

The Tn*5398 oriT* (1) and *oriT* (2) sites are aligned with the *oriT* sites present on Tn*916* (Jaworski and Clewell, 1995) and Tn*5397* (Roberts *et al.*, 2001). Identical nucleotides are shown in red. The *nic* site is indicated by a black triangle below the nucleotide alignment.

Tn*916*	AGGTGTGGACTAATCTTGCTGGCTGGTGTGGCGACAGCCA
Tn*5397*	AGGTGTGGACTAATCTTGCTGGCTGGTGTGGCGACAGCCA
Tn*5398*(1)	ATCTTGAATAATTACTTTAAAGGTGGTGTTGGAAAGTCCA
Tn*5398*(2)	AATACTGGAAAATCCAGTGAGGATGGTGTATGGTTTGGTA

The last step in the transfer of a conjugative transposon to a recipient cell is the integration of the element into the recipient cell genome. This process is generally dependent on the action of integrases, transposases or resolvases. A site-specific cut is made by one of these proteins in both the circular intermediate and in the target genome and the element is then integrated. The mechanism by which Tn*5398* may integrate is unclear, as no ORFs that appear to encode the appropriate proteins are present. It is possible that, as postulated for excision, the element is integrated by means of the TndX protein or another large resolvase present on the chromosome of the recipient cell. However, because of the large amount of sequence identity that was observed between the target region of the transposon in strain CD37 and the sequence upstream of the right hand end of the putative Tn*5398* element (Figure 4.26), integration of the putative element by a RecA-dependent homologous recombination process is also possible.

It is possible that the region excised from the chromosome to form the circular intermediate could be much larger than that proposed and may include genes upstream of the proposed left end (*ilvD*, *hydR*, and *hydD*) and downstream of the proposed right end (*ispD*, *flxD*). After transfer of this region, which contains Tn*5398*, it could be integrated into the recipient chromosome by RecA-dependent homologous recombination. The end result would be the integration of Tn*5398* and some of the genes flanking the element. This hypothesis is supported by the fact that upstream and downstream of the putative element nucleotide sequence differences between the donor and recipient were observed and that the DNA sequence of the transconjugant strains at these positions was the same as that of the donor rather than the recipient (Figures 4.25 and 4.26). The position of the proposed right end of Tn*5398* also agrees with this hypothesis. This end lies within the coding sequence of

orf7, which is unusual as excision and subsequent transfer of the proposed element would interfere with the transcription and expression of this ORF. If we presume that Tn*5398* was once a fully functional conjugative transposon that may have contained genes downstream of *orf7* that were involved in conjugative transfer of the element, a single homologous recombination event between the *orf7* gene region and the recipient genome may have resulted in a truncated form of Tn*5398*, which is no longer capable of catalyzing its own transfer.

It should be possible to distinguish between the two hypotheses outlined above by using PCR with outward firing oligonucleotide primers to detect the circular form of Tn*5398*. This experiment was performed but no product was obtained (data not shown). Failure to obtain this product may mean that Tn*5398* does not form a circular molecule, or is forming a circular molecule that is very different to that predicted. The oligonucleotide primers that were chosen to amplify this region would only give a product if the region between the priming sites was less than approximately 2 kb. If a much larger amount of DNA flanking the putative transposon was excised from the chromosome to form a circular intermediate, the oligonucleotide primers used for PCR would have been too far apart to generate a product.

Based on the above discussion it is concluded that Tn*5398* is not a classical conjugative transposon but is a mobilizable genetic element that is capable of disseminating the Erm B resistance determinant using the transfer functions of other elements or proteins present in the donor and recipient cells.

CHAPTER FIVE

GENOMIC ANALYSIS OF THE ERM B DETERMINANTS FROM *C. difficile* STRAINS OF DIFFERENT GEOGRAPHICAL ORIGINS

Introduction

The previous chapter reported the identification and delineation of a novel mobile genetic element, Tn*5398*, which carries the Erm B determinant. The prevalence and distribution of this element is unknown, with the Erm B determinant from strain 630 being the only Erm determinant from *C. difficile* that has been sequenced.

To determine if Tn*5398* was common in *C. difficile*, 27 *C. difficile* isolates were examined for the presence of *erm*(B) genes. The strains chosen for this study were clinical isolates and were obtained from different geographical locations (Table 2.1). Included in the study were five Australian isolates, ten American isolates, five Japanese isolates, two French isolates, two British isolates, a Belgian isolate and a Swiss isolate. The positive control was the Swiss strain 630, from which Tn*5398* was isolated. The negative control was the MLS sensitive strain, CD37, from the U.S.A.

Each of the strains found to contain an *erm*(B) gene was further examined for the presence of duplicated *erm*(B) genes and for other genes carried on the Tn*5398*. The objective was to determine whether the strain 630 Erm B determinant was

representative of the Erm B determinants carried by most MLS resistant strains of *C. difficile.*

Results

The *erm*(B) gene is not present in all MLS resistant *C. difficile* isolates

Preliminary studies were conducted on 11 isolates that were provided by Dr. Stuart Johnson from the U.S.A. These strains consisted of eight erythromycin resistant isolates collected from outbreaks of *C. difficile* diarrhoea that occurred between 1989 and 1992 in four hospitals in different parts of the U.S.A (Johnson *et al.*, 1999). They also included two erythromycin sensitive control isolates (Y4 and K12p) and an erythromycin resistant control isolate (B1/832) from another source.

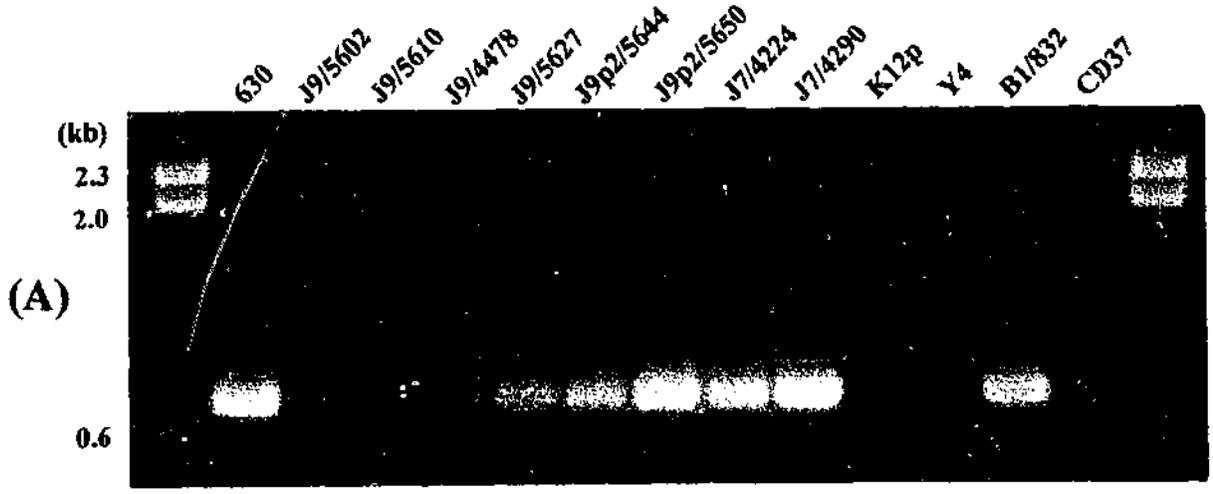
Initial studies involved PCR and DNA hybridizations with the aimed at determining if the resistant strains carried *erm*(B) genes. PCR was performed on chromosomal DNA extracted from each of the *C. difficile* isolates and the positive (strain 630) and negative (strain CD37) controls. The oligonucleotide primers #2980 and #2981 were used to amplify a 688 bp product encompassing most of the *erm*(B) and *orf3* genes. The results (Figure 5.1A) showed that a PCR product was amplified from all of the erythromycin resistant strains, indicating that an *erm*(B) gene was present in these isolates.

A dot blot hybridization experiment was also performed on chromosomal DNA from each strain. The 688 bp *erm*(B) PCR product from strain 630 was used as the *erm*(B)-specific probe. The results (Figure 5.1B) were in agreement with the

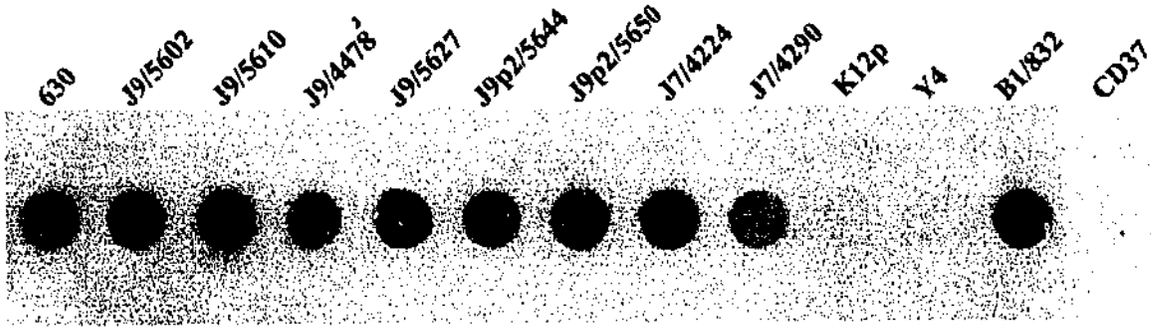
Figure 5.1 : Detection of *erm*(B) genes in *C. difficile* isolates from U.S.A. hospital diarrhoeal outbreaks.

(A) PCR was performed on chromosomal DNA extracted from each of the isolates indicated, using the oligonucleotide primers #2980 and #2981. To detect the presence of the 688 bp *erm*(B) product, samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

(B) DNA hybridization using a 688 bp *erm*(B)-specific probe was conducted on chromosomal DNA extracted from the strains indicated.


630
J9/5602
J9/5610
J9/4478
J9/5627
J9p2/5644
J9p2/5650
J7/4224
J7/4290
K12p
Y4
B1/832
CD37
(kb)
2.3
2.0
0.6

(A)


630
J9/5602
J9/5610
J9/4478
J9/5627
J9p2/5644
J9p2/5650
J7/4224
J7/4290
K12p
Y4
B1/832
CD37

(B)

PCR data and indicated that erythromycin resistance in these isolates was likely to be mediated by an *erm*(B) gene.

Based on these results it was decided to expand the study to include erythromycin resistant strains from a wider variety of geographical locations. As in the preliminary study, the presence of *erm*(B) genes in each isolate was detected by PCR and dot blot hybridization analysis. The PCR results (Figure 5.2A) showed that a PCR product was amplified from most of the isolates, indicating that an *erm*(B) gene was present in these strains. The strains from which a product could not be amplified included the Australian isolates AM480, AM1182 and 24/5-507, the Japanese isolates KZ1604, KZ1610, KZ1614, KZ1623 and KZ1655, and a British isolate R5948.

To ensure that differences in nucleotide sequence were not preventing the oligonucleotide primers from binding during the PCR, thereby leading to a false negative result, a dot blot hybridization experiment was performed on chromosomal DNA from each of the strains. The results (Figure 5.2B) concurred with the PCR results. It was concluded that the erythromycin resistant strains that were negative in these experiments did not carry an *erm*(B) gene.

Are the *erm*(B) genes all associated with DR sequences?

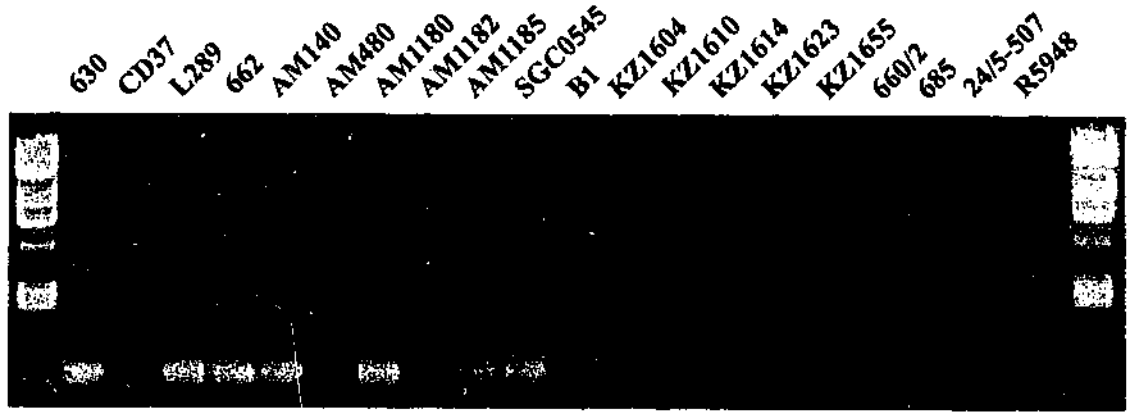
Previous analysis of Erm B determinants, including the strain 630 determinant, had indicated that most *erm*(B) genes were associated with complete DR sequences, or with variants of the DR sequence that are found in association with the *erm*(B) gene from *C. perfringens* (Berryman and Rood, 1995). To see if this association also applied to the other *erm*(B)-positive C. *difficile* isolates, further dot

Figure 5.2 : Detection of *erm*(B) genes in *C. difficile* isolates from different geographical locations.

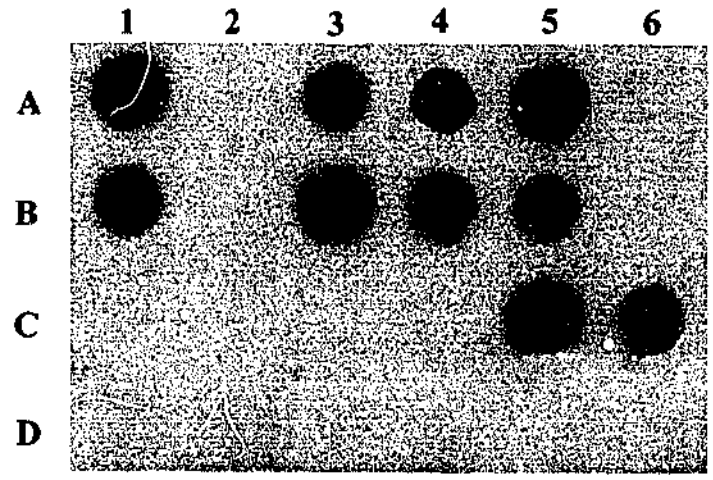
(A) PCR was performed on chromosomal DNA extracted from each of the isolates indicated, using the oligonucleotide primers #2980 and #2981. Samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

(B) DNA hybridization using a 688 bp *erm*(B)-specific probe was conducted on chromosomal DNA extracted from strains 630 (A1), CD37 (A2), L289 (A3), 662 (A4), AM140 (A5), AM480 (A6), AM1180 (B1), AM1182 (B2), AM1185 (B3), SGC0545 (B4), B1 (B5), KZ1604 (B6), KZ1610 (C1), KZ1614 (C2), KZ1623 (C3), KZ1655(C4), 660/2 (C5), 685 (C6), 24/5-507 (D1) and R5948 (D2). See Table 2.1 for strain descriptions.

(A)


630
CD37
L289
662
AM140
AM480
AM1180
AM1182
AM1185
SGC0545
B1
KZ1604
KZ1610
KZ1614
KZ1623
KZ1655
660/2
685
24/5-507
R5948
(kb)
23.1
9.4
6.6
4.4
2.3
2.0
0.6

(B)


1
2
3
4
5
6
A
B
C
D

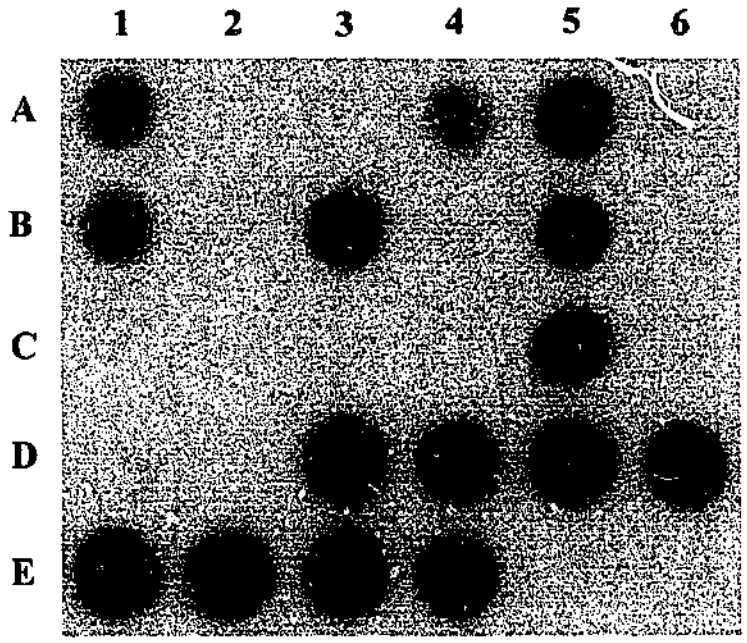
blot hybridizations were carried out using a 398 bp *orf298*-specific DIG labelled probe generated by PCR from strain 630 with the oligonucleotide primers #4538 and #4451. The results (Figure 5.3A) showed that the British strain L289, the Belgian strain SGC0545, the French isolate 685 and the American isolate B1/832 did not have a complete DR sequence associated with the *erm*(B) structural gene, since no hybridization was observed between these isolates and the *orf298*-specific probe. Note that the strains that did not have an *erm*(B) gene also did not have a DR sequence.

It was possible that the *orf298*-negative isolates may have contained incomplete (or variant) copies of the DR sequence. Since incomplete copies of the DR sequence appear to arise through deletion of *orf298* by homologous recombination between the palindromic sites *palA* and *palB* (Berryman and Rood, 1995), a 339 bp *palA*-specific DIG-labelled PCR product was amplified from strain 630 using the oligonucleotide primers #4191 and #4537. This probe should detect variants of the DR sequence that have lost *orf298* but still retain *palA*-like sequences. The results (Figure 5.3B) concurred with the data obtained with the *orf298*-specific probe, in that all of the isolates that contained a complete DR sequence also contained *palA*-like sequences. In addition, the French isolate 685, which lacked *orf298*, also appeared to contain *palA*-like sequences, indicating that although the *erm*(B) gene in this strain was not likely to be associated with a complete DR sequence, it may be associated with a DR variant since *palA*-like sequences were detected. Following this analysis, no further studies were carried out on isolates that either did not have an *erm*(B) gene, or did not have an *erm*(B) gene that was associated with either a complete DR or a DR variant.

Figure 5.3 : Detection of DR sequences in *C. difficile* strains from different geographical locations.

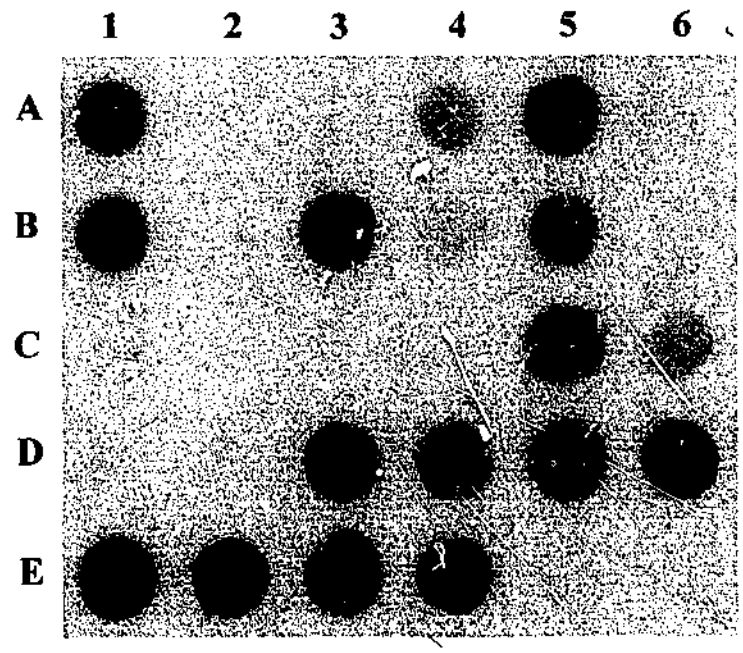
Dot blot hybridizations using a 398 bp *orf298*-specific probe **(A)** and a 339 bp *palA*-specific probe **(B)** were conducted on chromosomal DNA extracted from strains 630 (A1), CD37 (A2), L289 (A3), 662 (A4), AM140 (A5), AM480 (A6), AM1180 (B1), AM1182 (B2), AM1185 (B3), SGC0545 (B4), B1 (B5), KZ1604 (B6), KZ1610 (C1), KZ1614 (C2), KZ1623 (C3), KZ1655(C4), 660/2 (C5), 685 (C6), 24/5-507 (D1), R5948 (D2), J9/5602 (D3), J9/5610 (D4), J9/5627 (D5), J9/4478 (D6), J9p2/5644 (E1), J9p2/5650 (E2), J7/4224 (E3), J7/4290 (E4) and B1/832 (E5).

(A)



orf298-specific probe

(B)



palA-specific probe

Detailed PCR analysis of the Erm B determinants

To study the approximate extent and arrangement of the Erm B determinants that were present in the remaining isolates, a series of six PCRs were employed to examine the region surrounding the *erm*(B) gene. The PCR primers were designed based on the genetic organization of the Erm B determinants from *C. difficile* strain 630 and *C. perfringens* strain CP592 (Figure 4.6). Each reaction was designed to determine the presence or absence of specific regions of the Erm B determinant. Therefore, in combination, they would give an overall picture of the *erm*(B) gene region in each of these isolates.

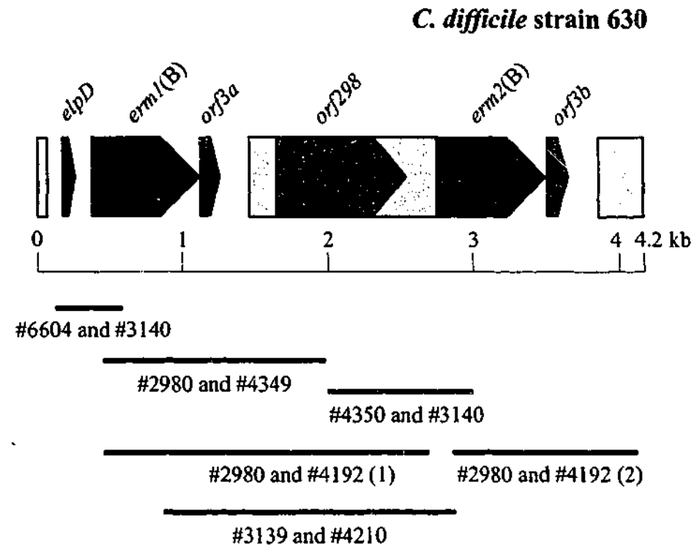
a) The* elpD *to* erm*(B) region

The first PCR in the series was designed to detect the presence of an Erm leader peptide upstream of an *erm*(B) structural gene. The oligonucleotide primers #6604, which binds upstream of the promoter sequences for the *elpD* leader peptide gene, and #3140, which binds within the 5' end of the *erm*(B) structural gene, were used in this reaction (Figure 5.4). If the gene arrangement was the same as in strain 630, PCR should generate a product of 610 bp (Figure 5.4). Only the two Australian isolates, AM1180 and AM1185, and the two French isolates, 660/2 and 685, yielded this product (Figure 5.5), indicating that these isolates contain a leader peptide sequence upstream of an *erm*(B) gene.

If the genetic organization was the same as that found in *C. perfringens* strain CP592, PCR should generate a product of 388 bp, which encompasses only the promoter sequences upstream of the *erm*(B) structural gene and the 5' end of *erm*(B) (Figure 5.4). A product of this size was amplified from the remaining isolates

Figure 5.4 : Schematic representations of the genetic organization of the Erm B determinants from *C. difficile* strain 630 and *C. perfringens* strain CP592.

ORFs are indicated by colored arrows and regions encompassing DR sequences are represented by grey shaded boxes. The approximate extent of PCR products that would be amplified using the oligonucleotide primers indicated are shown as black lines.



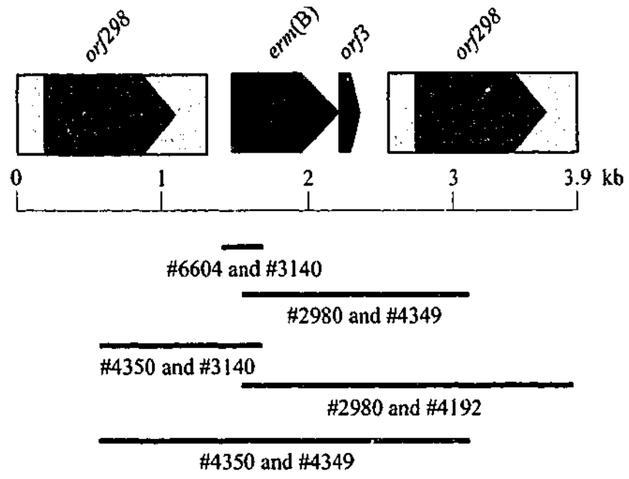
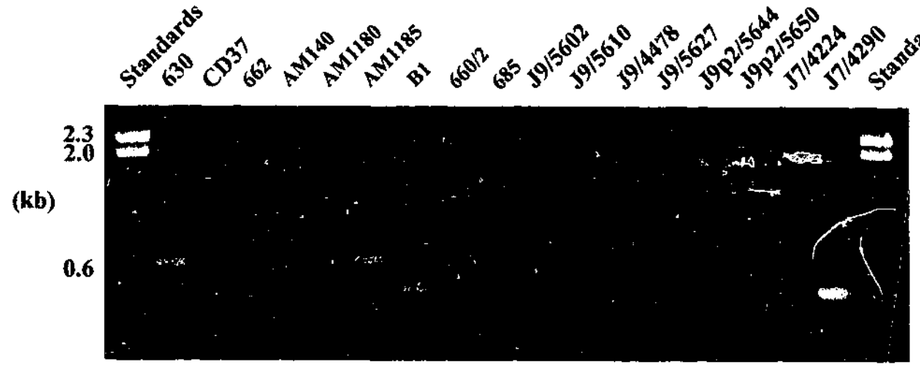


Figure 5.5 : PCR analysis of the *elpD* to *erm*(B) region.

PCR was performed on chromosomal DNA extracted from each of the *C. difficile* isolates indicated, using the oligonucleotide primers #6604 and #3140. To detect the presence of either 610 bp or 388 bp PCR products samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.



(Figure 5.5), which indicates that these isolates do not contain the leader peptide sequence upstream of an *erm*(B) structural gene.

To confirm that the two different PCR products did encompass the regions expected, the 610 bp product from strains 630, AM1185 and 685 and the 388 bp product from strains 662 and J9p2/5644 were sequenced using oligonucleotide primer #6604. The results showed that the region encompassed by these PCR products was exactly as predicted (data not shown).

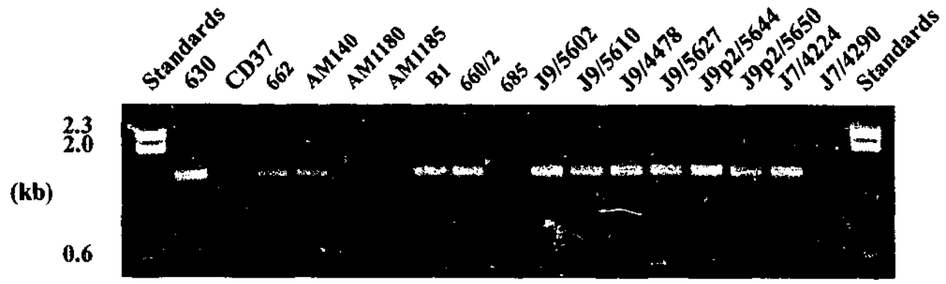
b) The* erm*(B) to* orf298 *region

To detect the presence of a complete DR sequence downstream of an *erm*(B) gene, the oligonucleotide primers #2980, which binds within the 5' end of *erm*(B), and #4349, which binds in the central region of *orf298*, were used in PCRs (Figure 5.4). If the genetic arrangement is the same as in either *C. difficile* strain 630 or *C. perfringens* strain CP592 (both organisms are identical in this region and contain a complete DR sequence downstream of an *erm*(B) structural gene), PCR using these two oligonucleotide primers should amplify a 1529 bp PCR product (Figure 5.4).

A product of this size was amplified from all but one of the isolates (Figure 5.6), which indicates these isolates contain *orf298* downstream of *erm*(B). This result makes it highly likely that the *erm*(B) structural gene in these isolates is associated with a downstream, complete DR sequence. As expected, no PCR product was amplified from the French isolate 685. Previous hybridization analysis had indicated that this isolate did not contain sequences homologous to *orf298* (Figure 5.3) and was therefore unlikely to contain a complete DR sequence. If

Figure 5.6 : PCR analysis of the *erm*(B) to *orf298* region.

PCR was performed on chromosomal DNA extracted from each of the *C. difficile* isolates indicated, using the oligonucleotide primers #2980 and #4349. To detect the presence of a 1529 bp PCR product samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.



orf298 was not present in this isolate, the binding site for #4349 would not be present, thereby preventing a PCR product from being amplified.

c) *The* orf298 *to* erm*(B) region*

The previous PCR detected *erm*(B) genes that were upstream of a complete DR sequence while the next experiment was designed to detect the presence of an *erm*(B) gene located downstream of a complete DR sequence. The oligonucleotide primers #4350, which binds in the central region of *orf298*, and #3140, which binds within the 5' end of *erm*(B), were used in this reaction (Figure 5.4). If the genetic organization of the *erm*(B) gene region was the same as in strain 630, PCR should amplify a 1044 bp product, which encompasses the 3' end of *orf298* and the 5' end of the *erm2*(B) gene (Figure 5.4). If, however, the arrangement is the same as in *C. perfringens* strain CP592, PCR should amplify a 1181 bp product, which encompasses the 3' end of *orf298* and the 5' end of *erm*(B) (Figure 5.4). The difference in the sizes of the PCR products obtained from the *C. difficile* and *C. perfringens* determinants across this region is due to the deletion in the former of the last 51 bp of the DR sequence and the promoter sequences found upstream of the *erm2*(B) gene.

A product of the same size as that of strain 630 (1044 bp) was amplified from AM1180, AM1185, and 660/2 (Figure 5.7), indicating that these isolates contain an *erm*(B) gene downstream of a complete DR sequence. Moreover, the same deletion as in strain 630 appeared to be present, since the smaller 1044 bp PCR product was obtained. PCR products were not detected from any of the remaining isolates, indicating that they are unlikely to contain an *erm*(B) gene downstream of a complete DR sequence.

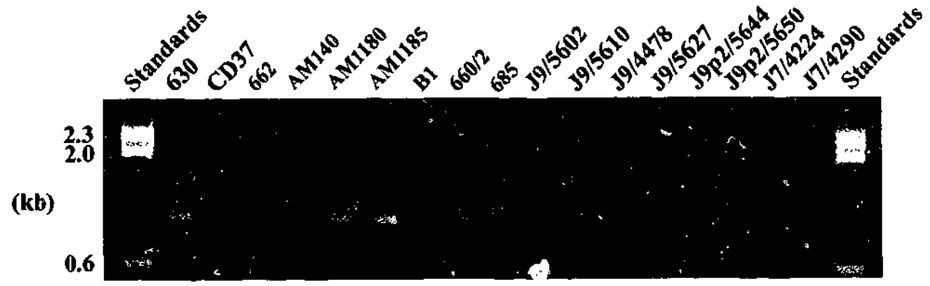


Figure 5.7 : PCR analysis of the *orf298* to *erm*(B) region.

PCR was performed on chromosomal DNA extracted from each of the *C. difficile* isolates indicated, using the oligonucleotide primers #4350 and #3140. To detect the presence of a 1044 bp or a 1181 bp PCR product samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

d) The erm*(B) to DR region*

In *C. perfringens* strain CP592 the *erm*(B) gene is associated with complete copies of the DR sequence (Figure 5.4). However, in *C. difficile*, and in other organisms (Figure 4.14), the *erm*(B) genes are associated with both complete and incomplete (or variant) copies of the DR sequence. The next PCR was designed to detect the presence of either a complete or variant DR sequence downstream of *erm*(B). This reaction used the oligonucleotide primers #2980, which binds within the 5' end of *erm*(B), and #4192, which binds at the 3' end of the DR sequence (downstream of *orf298*) (Figure 5.4). This reaction was therefore capable of detecting complete DR sequences and variants of the DR sequence from which *orf298* has been deleted. The previous PCR experiment was only able to detect complete DR sequences downstream of *erm*(B) because the target sequence of one of the oligonucleotide primers (#4349) was within *orf298*.

PCR using oligonucleotide primers #2980 and #4192 should yield two products (Figure 5.4) if the genetic organization is the same as that in strain 630. The first product should be 2219 bp in size and encompass most of *erm1*(B), *orf3a*, *orf298* and most of the complete DR sequence. The second product should be 1247 bp in size and encompass most of the *erm2*(B) structural gene, *orf3b* and the DR variant from which *orf298* has been deleted. However, if the arrangement of the *erm*(B) gene region is the same as in *C. perfringens* strain CP592, a single product of 2219 bp would be amplified (Figure 5.4).

PCR products of 2219 bp and 1247 bp were amplified from isolates AM1180, AM1185, and 660/2 (Figure 5.8), indicating that these isolates appear to have two *erm*(B) genes, one upstream of a complete DR sequence and the other upstream of a

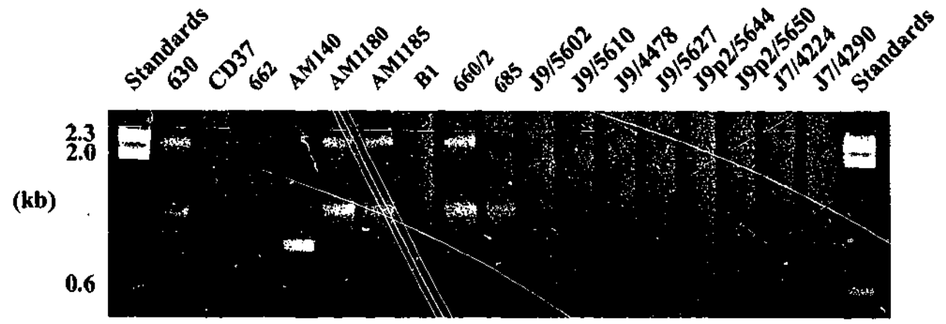


Figure 5.8 : PCR analysis of the *erm*(B) to DR region.

PCR was performed on chromosomal DNA extracted from each of the *C. difficile* isolates indicated, using the oligonucleotide primers #2980 and #4192. To detect the presence of 2219 bp and 1247 bp PCR products samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

variant of the DR sequence from which *orf298* has been deleted. These isolates appeared to have the same genetic arrangement as strain 630.

A single product of 1247 bp was amplified from isolate 685 (Figure 5.8). Previous analysis had already shown that this isolate did not contain a complete DR sequence. These results confirm that this isolate contains an *erm*(B) gene upstream of a DR sequence from which *orf298* has been deleted.

Two products, 1247 bp and 800 bp, were amplified from the Australian isolate AM140 (Figure 5.8). The presence of the 1247 bp product would appear to indicate that the *erm*(B) gene in this isolate is upstream of a deleted DR sequence. However, previous analysis of this isolate showed that the *erm*(B) gene was upstream of a complete DR sequence, so the result obtained here is somewhat contradictory. The second PCR product amplified did not correlate in size to either of the products expected using these oligonucleotide primers. Attempts were made to sequence this product to determine what region of DNA was encompassed, however, they were unsuccessful. This product may represent the result of non-specific binding of the oligonucleotide primers to unrelated sequences on the genome of this isolate.

No products were amplified from the remaining isolates, which was also unexpected. Previous analysis (*erm*(B) to *orf298* PCR) had indicated that the *erm*(B) gene in each of these isolates was upstream of a complete DR sequence. If these results were correct, it would be expected that at least a 2219 bp product should be amplified from each of these isolates, since the #4192 binding site should be present at the end of the complete DR sequence. The failure of this reaction to amplify products from these isolates could be due to two reasons. Deletion events are

reasonably common in the DR sequence region, so it is possible that a deletion event has removed the 3' end of the DR sequence, thereby removing the binding site for oligonucleotide primer #4192. However, it is more likely that sequence differences in these isolates may prevent oligonucleotide #4192 from binding to the template.

e) The erm1*(B) to* erm2*(B) region*

Since the previous results indicated that at least three of the isolates appeared to be similar to strain 630, the next experiment was designed to detect duplicated *erm*(B) genes that are arranged in a directly repeated orientation, as in strain 630. The oligonucleotide primers #3139, which binds in the 3' end of *erm*(B), and #4210, which binds in the 5' end of *erm*(B), were used in this reaction (Figure 5.4). A 2059 bp product would be amplified if the strain contains two *erm*(B) genes and the two genes are organized in a directly repeated orientation (Figure 5.4).

A PCR product of this size was amplified from AM1180, AM1185 and 660/2 (Figure 5.9), indicating that these isolates contain two *erm*(B) structural genes that are arranged in a directly repeated orientation. PCR products were not amplified from any of the remaining isolates indicating that these isolates contain either only one *erm*(B) gene, two widely separated *erm*(B) genes, or two *erm*(B) genes that are not arranged in a directly repeated orientation.

f) The orf298 *to* orf298 *region*

The final PCR was designed to detect the presence of two complete DR sequences flanking an *erm*(B) gene, as is observed in *C. perfringens* strain CP592 (Figure 5.4). The oligonucleotide primers #4349 and #4350, which are complementary to each other and bind in the central region of *orf298*, were used in

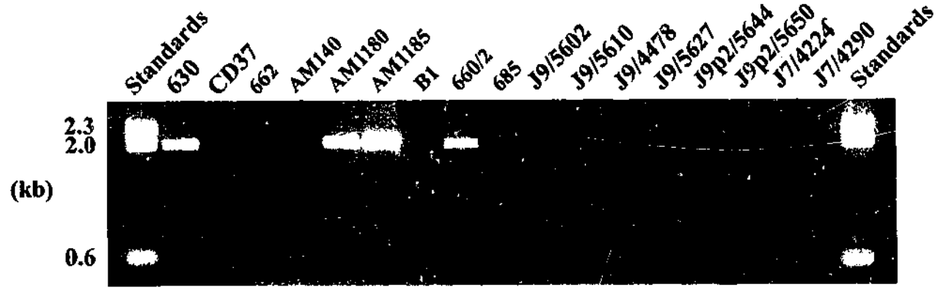


Figure 5.9 : PCR analysis of the *erm*(B) to *erm*(B) region.

PCR was performed on chromosomal DNA extracted from each of the *C. difficile* isolates indicated, using the oligonucleotide primers #3139 and #4210. To detect the presence of a 2059 bp PCR product samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

this reaction (Figure 5.4). If the arrangement of the *erm*(B) gene region in the *C. difficile* isolates was the same as in *C. perfringens*, PCR should amplify a 2565 bp product (Figure 5.4). In addition to conducting this PCR experiment on chromosomal DNA extracted from the *C. difficile* strains, the reaction was also performed on chromosomal DNA extracted from *C. perfringens* strain CP592 as a positive control.

No products were amplified from any of the *C. difficile* isolates, indicating that they did not contain two complete DR sequences. A product of the expected size was amplified from *C. perfringens* strain CP592 (data not shown).

The results from this systematic series of PCR experiments are summarized in Table 5.1. The data show that the genetic organization of the *erm*(B) gene region varies in the different erythromycin resistant *C. difficile* isolates.

Do AM1180, AM1185 and 660/2 carry Tn*5398*?

In the previous chapter it was shown that the *erm*(B) genes in strain 630 were carried on a novel mobilizable genetic element, Tn*5398*. Tn*5398* has been detected in as many as six other strains of *C. difficile* (Mullany *et al.*, 1995). The PCR results reported in this chapter indicate that three isolates, AM1180, AM1185 and 660/2, have the same *erm*(B) genetic arrangement as found in strain 630. Therefore, it was of interest to determine if these isolates, and any of the other isolates, also carried Tn*5398*.

Two PCR experiments were performed to detect the presence of Tn*5398*-specific sequences in each of the *C. difficile* isolates. The first reaction was designed

Table 5.1 : Summary of PCR results*.

C. difficile isolate	PCR Product Encompassing:					
	elpD to *erm*(B)	*erm*(B) to *orf298*	*orf298* to *erm*(B)	*erm*(B) to DR	*erm*(B) to *erm*(B)	*orf298* to *orf298***
630	+(610)	+	+	+(2219,1247)	+	-
CD37	-	-	-	-	-	-
AM1180	+(610)	+	+	+(2219,1247)	+	-
AM1185	+(610)	+	+	+(2219,1247)	+	-
660/2	+(610)	+	+	+(2219,1247)	+	-
685	+(610)	-	-	+(1247)	-	-
AM140	+(388)	+	-	+(1247,~800)	-	-
662	+(388)	+	-	-	-	-
B1	+(388)	+	-	-	-	-
J9/5602	+(388)	+	-	-	-	-
J9/5610	+(388)	+	-	-	-	-
J9/5627	+(388)	+	-	-	-	-
J9/4478	+(388)	+	-	-	-	-
J9p2/5644	+(388)	+	-	-	-	-
J9p2/5650	+(388)	+	-	-	-	-
J7/4224	+(388)	+	-	-	-	-
J7/4290	+(388)	+	-	-	-	-

(+) indicates that a PCR product of the expected size was observed

(-) indicates that a PCR product of the expected size was not observed.

* Where more than one PCR product could be expected from a PCR reaction, the size/s of the product/s observed is given in brackets following the (+).

** A product of 2565 bp was obtained from chromosomal DNA extracted from *C. perfringens* strain CP592 for this reaction.

to detect the presence of the *effD* gene, which is carried on Tn*5398*. The second reaction was used to detect an association between *effD* and *erm*(B). A positive result for both reactions would suggest the presence of a Tn*5398*-like element.

PCR using the *effD*-specific oligonucleotide primers #9069 and #10237, was conducted on chromosomal DNA extracted from all of the *C. difficile* strains used in the study. If an *effD* gene homologous to that found in strain 630 was present a 1166 bp PCR product would be detected. A PCR product of the expected size was amplified from AM1180, AM1185, 660/2, 685, AM140, and B1 (Figure 5.10). Sequence analysis of the resultant PCR products, using the oligonucleotide primer #9153, revealed that all of the products contained the same *effD* sequence (data not shown). No PCR products were amplified from the remaining strains, indicating that it was unlikely that these isolates carried Tn*5398*.

To determine if the *effD* gene present in these isolates was located close to *erm*(B), PCR using the oligonucleotide primers #3106, which binds at the end of the *erm*(B) coding sequence, and #11617, which binds within *effD*, was conducted. If a Tn*5398*-like element was present, and the arrangement of the element was the same as in strain 630, a 2759 bp product, which encompasses *orf3b*, *orf13*, *effR* and the 5'end of *effD*, would be detected. A PCR product of the expected size was amplified from AM1180, AM1185, 660/2, and 685 (Figure 5.11), indicating that it is highly likely that these strains carry a Tn*5398*-like element. Although the *effD* PCR had amplified a homologue of this gene from AM140 and B1, no *erm*(B)-*effD* PCR product was amplified from these strains, indicating that these isolates, while possessing an *effD* homologue, do not appear to have a Tn*5398*-like element. No PCR products were amplified from the remaining isolates, as expected.

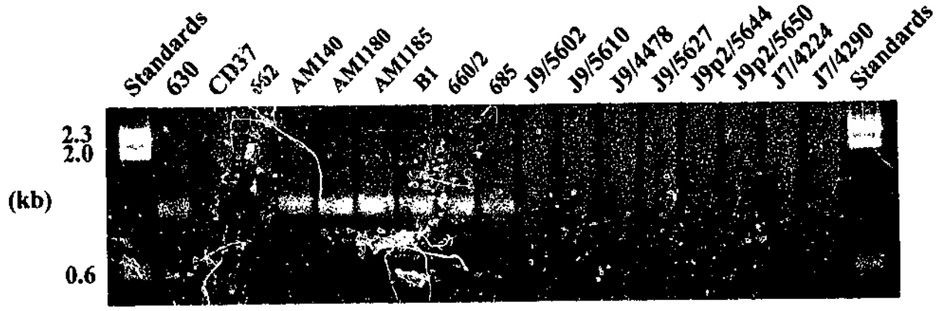


Figure 5.10 : Detection of Tn*5398* in *C. difficile* isolates using an *effD*-specific PCR.

PCR was performed on chromosomal DNA extracted from each of the isolates indicated, using the oligonucleotide primers #9069 and #10237. To detect the presence of a 1166 bp PCR product samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

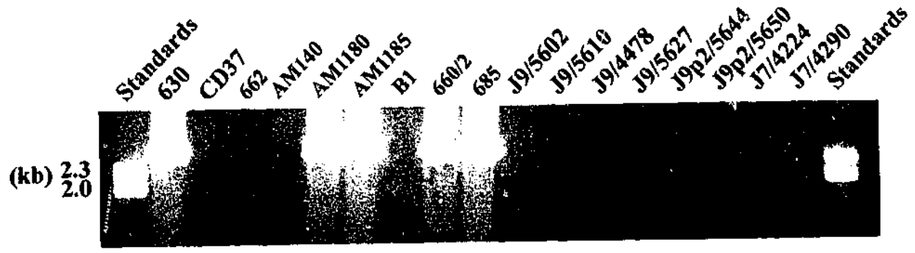


Figure 5.11 : Detection of Tn*5398* in *C. difficile* isolates using an *erm*(B)-*effD* PCR. PCR was performed on chromosomal DNA extracted from each of the isolates indicated, using the oligonucleotide primers #3106 and #11617. To detect the presence of a 2759 bp PCR product samples of each reaction were run on a 0.8% agarose gel alongside λ*Hin*dIII standards.

Discussion

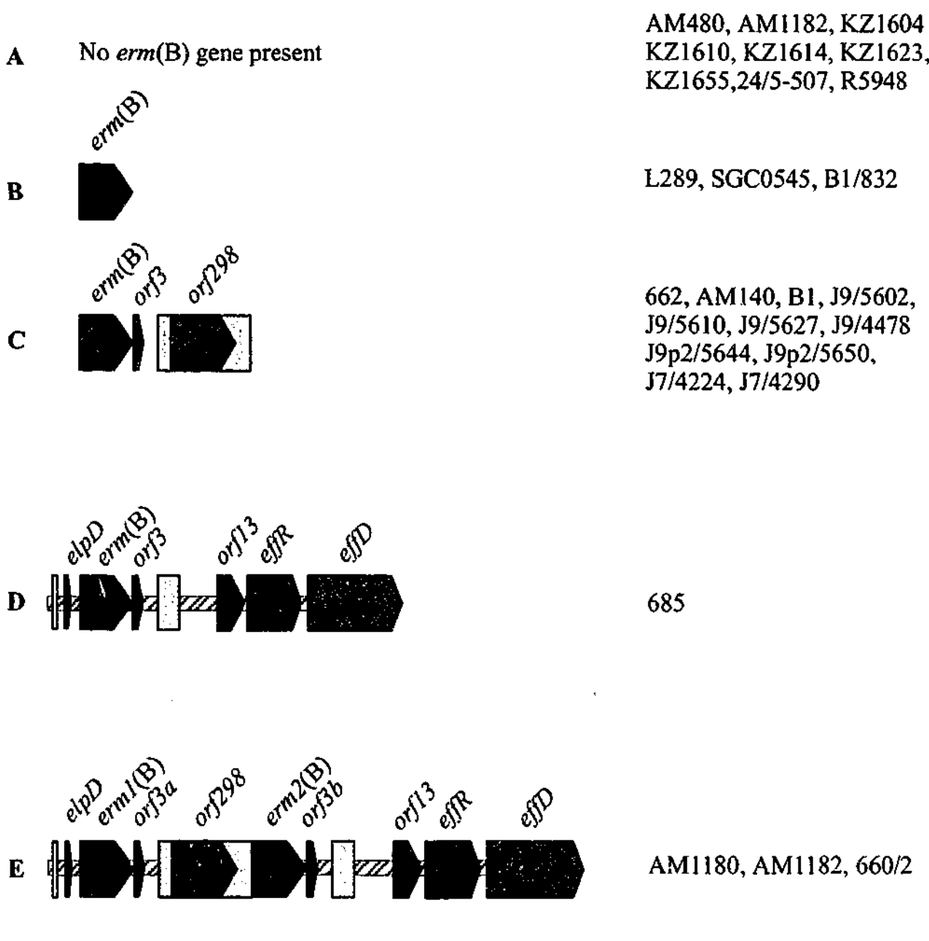
The results presented in this chapter reveal that there is considerable heterogeneity amongst the erythromycin resistance determinants carried by *C. difficile* isolates from different geographical locations. The PCR results (Table 5.1) allowed the various *C. difficile* isolates to be divided into five groups based on the type of erythromycin resistance gene encoded and the arrangement of the *erm*(B) gene region (Figure 5.12). The first group of isolates, which includes all five Japanese isolates, three Australian isolates, AM480, AM1182 and 24/5-507, and a British isolate, R5948, were resistant to erythromycin but did not contain an *erm*(B) gene. In addition to being resistant to erythromycin, each of these isolates was also resistant to clindamycin, suggesting that another *erm* gene is responsible for the resistance phenotype observed in these isolates. The isolation and characterization of the gene responsible for MLS resistance in these isolates would be most worthwhile since *erm* genes from classes other than Erm B have not yet been characterized from *C. difficile*. All of the Japanese isolates were contained in this group of isolates, suggesting a geographical association with this unknown *erm* gene.

The British isolate L289, the Belgian isolate SGC0545, and the American isolate B1/832 comprised the second group of isolates. These strains carried an *erm*(B) gene but did not appear to have any DR sequences associated with this gene (Figure 5.12B). Although only three isolates were included in this group, there does not appear to be any geographical focus for this type of determinant.

The largest group of isolates had an *erm*(B) gene located upstream of a complete DR sequence (Figure 5.12C). These strains included the American isolates AM140, J9/5602, J9/5610, J9/5627, J9/4478, J9p2/5644, J9p2/5650, J7/4224 and

Figure 5.12 : The *erm*(B) gene regions in *C. difficile* isolates from different geographical locations.

Schematic representations of each of the different *erm*(B) gene regions are shown. ORFs are depicted as colored block arrows, and regions encompassing DR sequences are shown as shaded grey boxes. Regions encompassing Tn*5398*-like elements are backed by a hatched box. Isolates that contain each arrangement are listed next to the diagrams.



J74290, the Swiss isolate 662, and the British isolate B1. While the PCR results were somewhat contradictory for AM140, it has been placed in this group because, with the exception of the *erm*(B) to DR PCR, all of the results agree with the arrangement found for the other members. This group contained all of the American isolates included in the study, with the exception of B1/832, which was included in the previous group of isolates.

All of the American isolates in this group, except for AM140, were clinical isolates that were isolated from large outbreaks of diarrhoea, which occurred in four hospitals located in different parts of the United States between 1989 and 1992 (Johnson *et al.*, 1999). Analysis of these strains by restriction digestion, PCR and PFGE had determined that these isolates were derivatives of the same strain (Johnson *et al.*, 1999), which was referred to as the epidemic strain. They were all highly resistant to clindamycin and were shown to account for between 30 and 66% of the *C. difficile* strains isolated at the four hospitals during the indicated period (Johnson *et al.*, 1999). The results presented in this chapter showed that each of these isolates contained an *erm*(B) gene region with the same genetic arrangement, lending support to the conclusion that these isolates are derivatives of a single epidemic strain. Johnson *et al.* (1999) suggested that the *erm*(B) gene present in these isolates may be associated with Tn*5398*. This study reveals that this hypothesis is unlikely as the Tn*5398*-specific gene, *effD*, was not detected in these isolates. The *effD* gene was detected in two other isolates in this group, the American isolate AM140 and the British isolate B1, however, no association between the *effD* gene and the *erm*(B) gene present in these isolates was observed. These isolates may represent strains in which rearrangement of Tn*5398* has led to the separation of the *erm*(B) gene from the mobile element or to the deletion of the *effD* gene.

The PCR results indicated that the four remaining *C. difficile* isolates examined in this study appeared to contain a Tn*5398*-like element. These isolates comprised the remaining two groups. The first group contained only the French isolate 685 (Figure 5.12D). The genetic arrangement in this strain consisted of the *elpD* leader peptide gene and the *erm*(B) gene, which were located upstream of a variant of the DR sequence from which *orf298* had been deleted. The remainder of the Tn*5398* element appeared to be present (Figure 5.12D). Further genetic analysis of this element may contribute to our understanding of the mechanism of transfer of Tn*5398* because one of the two proposed *oriT* sites was not present.

The Australian isolates AM1180 and AM1185, and the French isolate 660/2, comprised the final group of isolates. These strains were shown to contain two *erm*(B) genes arranged as in strain 630, in association with a Tn*5398*-like element (Figure 5.12E). The two Australian isolates were both isolated in the 1980's from two different Victorian hospitals (AM1180 from the Latrobe Valley Hospital, and AM1185 from the Royal Melbourne Hospital).

The results presented in this chapter have implications for the distribution and spread of erythromycin resistance in *C. difficile*. The *erm*(B) genes that were not associated with Tn*5398* (Figure 5.12B and C) appeared to have a greater geographic focus, indicating that global spread of these isolates may be reduced. However, *erm*(B) genes that were associated with Tn*5398* (Figure 5.12D and E) seem to be isolated from a wider variety of geographic sources, perhaps as a result of the transfer of the Tn*5398* elements between different isolates of *C. difficile*. These results suggest that the conjugative mobilization of Tn*5398* may contribute to the global distribution of MLS resistance in *C. difficile*.

CHAPTER SIX

CONCLUSIONS AND FUTURE DIRECTIONS

MLS resistance in *C. perfringens* and *C. difficile* has been shown to be mediated in many resistant strains by the presence of *erm*(B) genes (Berryman and Rood, 1989; Berryman and Rood, 1995; Hächler *et al.*, 1987a; Mullany *et al.*, 1995). In *C. perfringens* the *erm*(B) gene is located on the large, non-conjugative but mobilizable plasmid, pIP402. The work described in Chapter Three of this thesis examined the structure and function of the Erm(B) protein that is encoded by this *erm*(B) gene. Six pJIR418 derivatives that contained single point mutations in the *erm*(B) gene were isolated by random mutagenesis. These mutations resulted in the production of Erm(B) proteins with single amino acid changes that either abolished or reduced erythromycin resistance in both *C. perfringens* and *E. coli* backgrounds. Each of the amino acid changes in the Erm(B) variants occurred within or close to the conserved methyltransferase motifs that in other methyltransferases are either involved in SAM binding, RNA binding or catalysis of the methylation reaction.

Because of the clinical importance of bacteria that exhibit the MLS resistance phenotype, which most likely arise from the intergeneric spread of elements, such as Tn*5398*, carrying *erm* genes, it has become important to develop inhibitors of the Erm methyltransferases to be able to treat infections caused by MLS resistant bacteria. These inhibitors could then be administered in combination with an MLS antibiotic to overcome the resistance created by the Erm methyltransferase, analogous to the administration of clavulanic acid (a beta-lactamase inhibitor) in

combination with amoxicillin (Augmentin®) to overcome the resistance created by beta-lactamases (Parker and Eggleston, 1987). Several potential inhibitors of the Erm methyltransferases are being developed, which are essentially based on the ability of the end product of the methylation reaction, SAH, to inhibit the methylation reaction (Hajduk *et al.*, 1999; Hanessian and Sgarbi, 2000). These inhibitors bind to the active site of the Erm protein, thereby competing with the substrate of the methylation reaction, SAM.

The residues that were identified during this study as abolishing or reducing erythromycin resistance are potentially involved in either binding of SAM or binding of RNA, which are critical for function of the Erm(B) methyltransferase from *C. perfringens*. These residues, and the motifs they are part of, are generally well conserved among most Erm methyltransferase proteins and may represent good targets for the development of inhibitors to this important family of enzymes.

There is a third region of the Erm methyltransferase proteins that is involved in recognition of the RNA target. The residues that are important to the function of this region of the Erm proteins are not well characterized, but generally tend to be positively charged residues in the C-terminal domain that are predicted to have an exposed location in the structure of the protein. The *C. perfringens* Erm(B) protein has several lysine (K185, K188 and K196) and arginine (R200, R203 and R204) residues, which, based on the predicted structure of the protein, would have a surface exposed location. Site-directed mutagenesis of these residues to neutral or negatively charged residues may determine whether they are involved in binding of the RNA target. While the development of inhibitors has so far focused on inhibiting

the binding of SAM to the methyltransferase protein, it might also be worthwhile to develop inhibitors that prevent the recognition of the RNA target.

The Erm B determinant from the *C. perfringens* plasmid pIP402 consists of the *erm*(B) gene and a downstream ORF of unknown function, *orf3*, flanked by two copies of the DR sequence (Berryman and Rood, 1995) (Figure 4.6). An analysis of the MLS resistance determinant from *C. difficile* strain 630, presented in Chapter Four of this thesis, reveals that this Erm B determinant is novel when compared to either the *C. perfringens* Erm B determinant or similar determinants from other organisms. The *C. difficile* Erm B determinant contains two identical *erm*(B) and *orf3* genes, *erm1*(B) and *erm2*(B), and *orf3a* and *orf3b*, respectively, which are separated by a single complete copy of the DR sequence and are flanked by variants of the DR sequence (Figure 4.6). The genetic organization of this region in *C. difficile* strain 630 is the first known example of an Erm B determinant that contains a duplicated *erm* gene. The encoded Erm(B) protein has 98% identity at the amino acid level with the Erm(B) protein from *C. perfringens* and greater than 80% identity to the other members of the Erm B class of determinants, thereby supporting the classification of these *erm* genes as belonging to the Erm B class.

Like many Erm B determinants, the *C. difficile* determinant is flanked by variants of the DR sequence. Downstream of *erm2*(B) is an incomplete copy of the DR sequence from which *orf298* has been deleted (ΔDR). Examination of the sequence of ΔDR indicated that recombination between the palindromic sites *palA* and *palB* was the likely cause for this deletion, as is the case for ΔDR variants from other Erm B determinants. Examination of the sequences of the ΔDR variants from *C. difficile* and other Erm B class determinants has shown that the deletion endpoint

in these variants is located within the *palA* and *palB* sequences, but they appear to have arisen from separate deletion events because the exact points of divergence are different. This suggests that homologous recombination events are responsible for the deletions rather than site-specific recombination events.

It has been postulated that the Erm B determinant from *C. perfringens* represents the progenitor and that other Erm B class determinants have evolved from this determinant *via* homologous recombination events (Berryman and Rood, 1995). As discussed in Chapter Four, the structure of the *C. difficile* Erm B determinant is consistent with this hypothesis. A duplication of the progenitor determinant followed by a series of homologous recombination events, the acquisition of an Erm leader peptide gene and the loss of the promoter sequences upstream of the *erm2*(B) gene could result in the formation of the *C. difficile* strain 630 Erm B determinant (Figure 4.28).

Further support for this hypothesis can be obtained from the heterogeneity observed amongst Erm B determinants found in *C. difficile* strains from different geographical locations (Chapter Five). Four genetic variants of the Erm B determinant were detected. The simplest variant consisted of only a single *erm*(B) gene while the Erm B determinant found in strains 630, AM1180, AM1185 and 660/2 (Figure 5.12) represented the most complex. The different genetic arrangements of the various *C. difficile* Erm B determinants could all have evolved from the common progenitor *via* homologous recombination events in a similar manner to the strain 630 determinant.

In *C. perfringens* the Erm B determinant is not widespread, as indicated by hybridization analysis of erythromycin-resistant *C. perfringens* strains in which only five out of 40 erythromycin-resistant *C. perfringens* strains hybridized with the *erm*(B) probe (Berryman and Rood, 1989). In *C. perfringens* the most common MLS resistance determinant appears to be Erm Q, which was present by in 30 of 38 erythromycin-resistant *C. perfringens* strains (Berryman *et al.*, 1994). In *C. difficile*, the Erm B determinant appears to be the most prevalent MLS resistance determinant as indicated by the results presented in Chapter Five. In addition to strain 630, 17 of the 27 erythromycin-resistant *C. difficile* isolates that were examined hybridized to the *erm*(B) probe. MLS resistance determinants from the Erm Q and Erm F classes have also been reported to be present in erythromycin-resistant *C. difficile* isolates (Roberts *et al.*, 1994), however, the presence of these genes has not been confirmed by cloning or sequence analysis. In the current study, ten *C. difficile* isolates, which were resistant to erythromycin and clindamycin, did not hybridize to the *erm*(B) probe. These isolates may contain Erm resistance genes from classes other than Erm B and therefore represent good candidates for the isolation and analysis of other *C. difficile* Erm determinants.

The *C. difficile* Erm B determinant has been shown to be transferred by conjugation, in the absence of plasmid DNA, to *C. difficile* (Wüst and Hardegger, 1983), *S. aureus* (Hächler *et al.*, 1987a) and *B. subtilis* (Mullany *et al.*, 1995) recipients. Furthermore, the Erm B determinant could be transferred from *B. subtilis* transconjugants back to *C. difficile* (Mullany *et al.*, 1995). The element carrying the Erm B determinant had also been shown to integrate into the recipient chromosome site-specifically in *C. difficile*, and without site specificity in *B. subtilis* (Mullany *et al.*, 1995). Transfer behaviour as described above is typical of conjugative

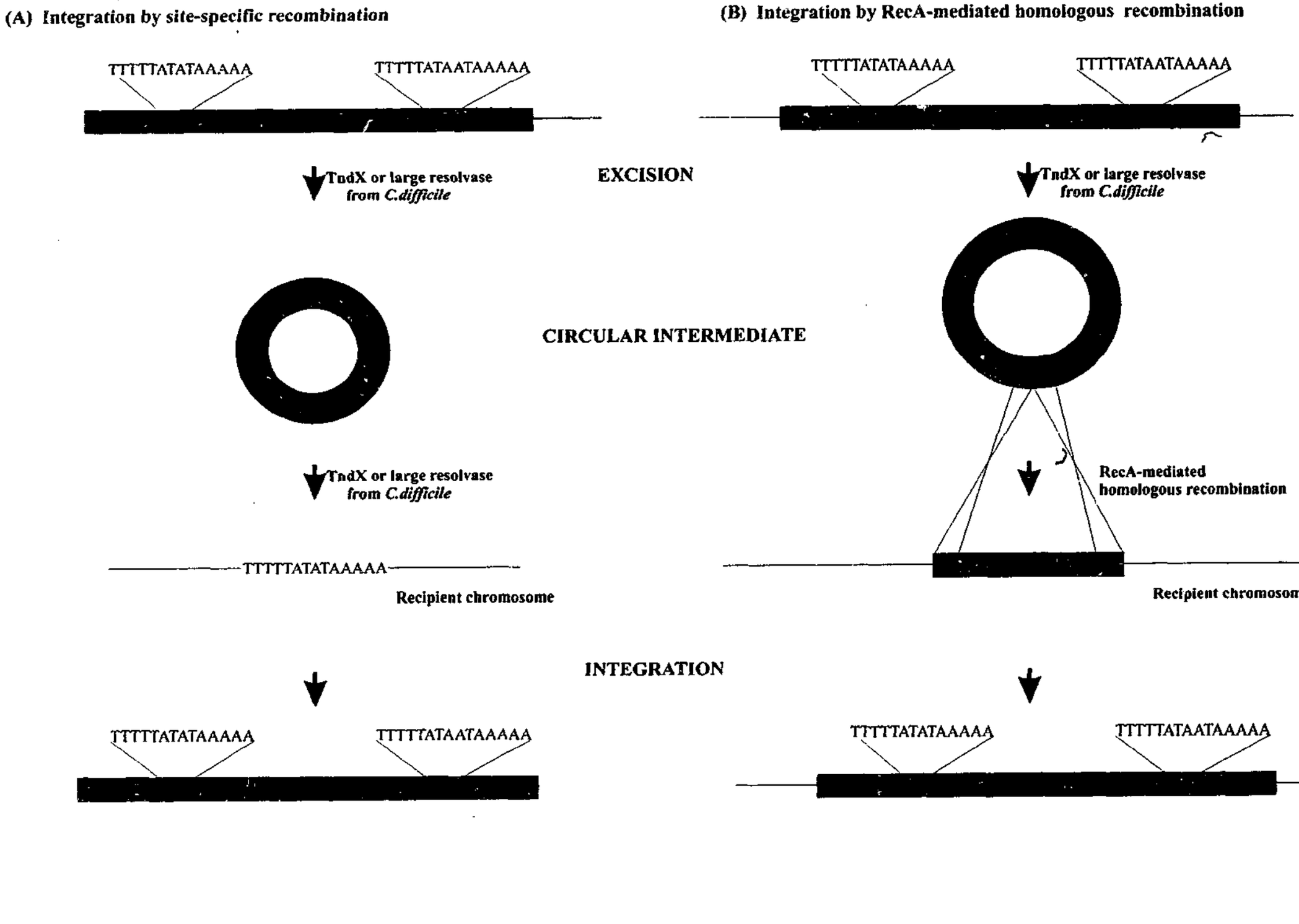
transposons and, accordingly, the genetic element carrying the *C. difficile* Erm B determinant was designated Tn*5398* (Mullany *et al.*, 1995).

This study involved a physical and genetic analysis of the Tn*5398* element, and revealed that rather than being a classical conjugative transposon, Tn*5398* was more likely to be a mobilizable genetic element. The results presented in Chapter Four revealed that Tn*5398* is approximately 9.6 kb in length and that, in addition to the Erm B determinant, it carries four complete ORFs, *orf5*, *ffR*, *effD* and *orf9*, and one incomplete ORF, *orf7*. The ends of the element were shown to consist of highly A-T rich palindromic sequences, which were almost identical to the target sequence in the *C. difficile* recipient strain CD37. In addition, the ends of Tn*5398* showed no similarity to the ends of other transposable elements in the database. The genes encoded on Tn*5398* do not appear to have any similarity either to genes known to be involved in conjugative transposition or to known *mob* genes.

In Chapter Four it was postulated that Tn*5398* may be a mobilizable element that is excised from the donor, conjugatively transferred, and integrated into the recipient chromosome, using proteins that are either encoded on the co-resident transposon Tn*5397*, or elsewhere on the *C. difficile* chromosome (Figure 6.1). At this time, testing this hypothesis is difficult due to the lack of genetic methods for the introduction and manipulation of DNA in *C. difficile* strains, and also due to the non-availability of *C. difficile* strains with appropriate resistance markers that could be used as recipients. It may be possible to construct a system in a heterologous host, such as *E. coli*, that contains the required components either on plasmids or integrated into the chromosome. This approach might not prove successful,

Figure 6.1 : Models for the transfer of Tn*5398* in *C. difficile*.

Schematic representations of two alternative models for the transfer of the Tn*5398* element are shown. Tn*5398* is represented by the blue rectangle and DNA flanking Tn*5398*, which is homologous to that in the recipient strain, is represented by the green rectangles. The ends of Tn*5398* are represented by a string of nucleotides, as is the target site in the recipient chromosome. In both models Tn*5398* is excised from the donor chromosome by TndX, or by a large resolvase encoded on the *C. difficile* chromosome, and forms a circular molecule. The circular intermediate is then mobilized to the recipient strain. In Model A, only Tn*5398* is excised from the donor chromosome, and it is integrated into the recipient chromosome *via* site-specific recombination at the target site by TndX or by a large resolvase encoded on the *C. difficile* chromosome. In Model B, a larger element, including Tn*5398* and its flanking DNA, are excised from the donor chromosome to form a circular molecule. The larger element is then integrated into the recipient genome *via* RecA-mediated homologous recombination.


(A) Integration by site-specific recombination
TTTTTATATAAAAA
TTTTTATAATAAAAA
TndX or large resolvase from C.difficile
EXCISION
CIRCULAR INTERMEDIATE
TndX or large resolvase from C.difficile
TTTTTATATAAAAA
Recipient chromosome
INTEGRATION
(B) Integration by RecA-mediated homologous recombination
TTTTTATATAAAAA
TTTTTATAATAAAAA
TndX or large resolvase from C.difficile
RecA-mediated homologous recombination
Recipient chromosome
TTTTTATATAAAAA
TTTTTATAATAAAAA

however, as it is possible that one or more unknown *C. difficile* proteins or factors are required.

If we were to assume that there were appropriate genetic systems in *C. difficile* that would allow the introduction of DNA at frequencies high enough to obtain homologous recombination then several exciting options would be presented. To determine if transfer of Tn*5398* is dependent on the presence of Tn*5397* attempts could be made to transfer Tn*5398* from a strain background that lacks Tn*5397*. If Tn*5398* transfer depended on the presence of Tn*5397*, strains containing only Tn*5398* would be unable to transfer this element to a suitable recipient. However, those strains that carried both Tn*5398* and Tn*5397* should be able to transfer erythromycin resistance either independently or together with the transfer of tetracycline resistance. CD37-derived Tn*5398* strains that did not contain Tn*5397*, and CD37-derived Tn*5398*/Tn*5397* strains, which could potentially be used as donors for this type of experiment, were obtained during this study. Ideally the recipient strain should be a derivative of CD37 with an additional resistance marker, other than rifampicin. Repeated attempts to mutate CD37 to obtain a derivative that contained a second selectable marker were unsuccessful, as was screening our extensive *C. difficile* collection for an appropriate recipient. Another option may be to use either a *B. subtilis* or *S. aureus* recipient strain, where aerobic growth could be used as the second selective marker. Transfer of Tn*5398* from *C. difficile* strain 630 into an appropriate *B. subtilis* recipient was attempted, however, no transconjugants were obtained. These studies highlight another inherent problem in studying Tn*5398*, the very low level of transfer that is generally observed.

If transfer of Tn*5398* does prove to be dependent on the presence of Tn*5397*, mutation of several genes on both Tn*5397* and Tn*5398* would then be necessary. To determine if the TndX protein is responsible for the excision and/or integration of Tn*5398* it would be necessary to introduce a Tn*5397*-derivative with a non-functional *tndX* gene into a strain carrying Tn*5398* and to observe if transfer of Tn*5398* to an appropriate recipient still occurs. While TndX is an obvious choice for the protein responsible for the excision and integration of Tn*5398*, this protein is a site-specific recombinase and generally cuts only at a specific target sequence. In *C. difficile* strain CD37 this target sequence consists of the sequence 5'-TCCTTTTAGTGATGGTAATGGA-3', which resembles the ends of Tn*5397* (Wang *et al.*, 2000a). TndX cleaves this target sequence and the ends of Tn*5397* at the central GA dinucleotide. TndX cannot promote the excision of Tn*4451*-derivatives with a non-functional *tnpX* gene, which suggests specificity of the TndX enzyme in regard to the sequences of the transposon ends and insertion site (Wang *et al.*, 2000a). The sequence of the ends of Tn*5397* and Tn*5398* are unrelated, which therefore suggests that it would be unlikely to be the TndX protein that was responsible for the excision and integration of Tn*5398*. However, since limited information about the exact specificity of this enzyme is available, the possibility that TndX is performing these functions can not been eliminated.

As previously mentioned, there are at least five other large resolvases encoded on the genome of *C. difficile* strain 630 (M. Smith, personal communication). It may be one of these proteins that is responsible for the excision and integration of Tn*5398*, and it may prove necessary to inactivate each of the genes encoding these large resolvases individually to determine if they play a role in the conjugative transfer of Tn*5398*.

If Tn*5398* excises from the chromosome and forms a circular intermediate, its subsequent transfer to a recipient cell would then be dependent on the action of a Mob protein. Tn*5398* is not predicted to contain any ORFs that encode proteins with similarity to Mob proteins. In this study we have proposed that the protein encoded by *orf23* of Tn*5397*, whose homologue from Tn*916* has similarity to the MbeA mobilization protein of ColE1 (Flannagan *et al.*, 1994), may be performing this function. Two potential Tn*5398 oriT* sites that had significant similarity with the *oriT* sites found on Tn*916* (Jaworski and Clewell, 1995) and Tn*5397* (Roberts *et al.*, 2001) were identified. The *nic* sites in these putative *oriT* sites were conserved. To test whether these sites are involved in the transfer of Tn*5398*, site-directed mutagenesis of the *nic* site could be performed on plasmids containing the Tn*5398* *oriT* sites. The mutated *oriT* sequences would then need to be introduced into a *C. difficile* strain carrying Tn*5398* by homologous recombination, thereby replacing the wild-type *oriT* sequences on the element. If either of the *oriT* sites were functional it would be expected that the transfer frequency of Tn*5398* would be greatly reduced or abolished when the *nic* site is mutated. The reverse experiment of insertionally inactivating *orf23* on Tn*5397* should also be carried out to determine whether the protein encoded by this ORF is necessary for mobilization of Tn*5398*.

Although the Tn*5397*-mediated mobilization of Tn*5398* seems like the most likely explanation, it is possible that the proteins encoded by either *orf298*, *orf13*, *orf9* or *orf7* may play a role in transfer of the element. In Tn*916*, ORF7 and ORF9, which are similar to ORF7 and ORF9 from Tn*5398*, are postulated to have regulatory roles in the transposition process. The Tn*916* ORF7 protein shows limited similarity to sigma factors (Flannagan *et al.*, 1994) and has been proposed to have a regulatory role in the mobility of Tn*916* because, in the presence of tetracycline, increased

ORF7 expression leads to increased transcription of *orf7*, *orf8*, *xis* and *int*, and other genes (Celli and Trieu-Cuot, 1998). The finding that the right end of Tn*5398* is internal to *orf7* may have implications for the level of excision, transfer and integration of the element. After excision from the donor chromosome, Tn*5398* would leave behind part of *orf7*, resulting in an incomplete *orf7* gene in the circular intermediate. The result could be altered levels of transcription of other genes involved in transfer of the element. The end of *orf7* also appears to be the target sequence for the element in recipient strains such as CD37. It is this region of the ORF7 protein that has identity to the sigma-factor helix-turn-helix motif, which is involved in DNA binding. Fusion with this region may provide a selective advantage for recombination of the circular intermediate at the target site.

The Tn*916* ORF9 protein has been predicted to be a putative transcriptional repressor, however, the role of this repressor in the mobility of Tn*916* has not been determined (Celli and Trieu-Cuot, 1998), while the ORF13 protein has no known role in the mobility of Tn*916*. Mutational analysis of the *orf7*, *orf9* and *orf13* homologues on Tn*5398* may prove necessary to investigate the regulation of transfer of this element, but will probably not provide further insight into the mechanism of transfer.

The only other known ORF that could encode a protein involved in Tn*5398* mobility is *orf298*. The putative ORF298 protein has some similarity to replication proteins and proteins from the ParA and Soj families. ParA and Soj proteins generally have a role in the partitioning of plasmids and chromosomes during the replication cycle (Easter *et al.*, 1998; Sharpe and Errington, 1996). It appears

unlikely that ORF298 has a role in either the excision or integration of Tn*5398* but this possibility cannot be completely eliminated. The *C. difficile* strain 685 genetic variant of the Erm B determinant presented in Chapter Five consisted of an Erm leader peptide gene, the *erm*(B) gene and *orf3* flanked by two variants of the DR sequence, neither of which contained *orf298*. This genetic variant of the Erm B determinant was also shown to be associated with a Tn*5398*-like element. One way to determine if ORF298 is required for the transfer of Tn*5398* might be to introduce this genetic variant of Tn*5398* into a *C. difficile* strain carrying Tn*5397* and then to look at differences between the transfer frequency of this genetic variant and that of Tn*5398*. This experiment could not be performed during this study due to the non-availability of a *C. difficile* strain that contained only Tn*5397*.

It is tempting to speculate that ORF298 is the protein responsible for the mobilization of Tn*5398*, as the *C. perfringens* Erm B determinant, which is located on the large non-conjugative but mobilizable plasmid, pIP402, contains two copies of *orf298*. Little is known about pIP402 other than that it carries the Erm B MLS resistance determinant. It is possible that the reason that this plasmid is mobilizable is due to the presence of *orf298*. It would also be worthwhile to inactivate one or both of the *orf298* copies on pIP402, or the *orf298* open reading frame on Tn*5398*, and to determine if there was any effect on the transfer of either of these elements.

A second hypothesis relating to the integration of Tn*5398* was discussed in Chapter Four. It is possible that the region that is excised from the *C. difficile* chromosome to form the circular intermediate is much larger than the 9.6 kb element and that it includes genes upstream of the proposed left end (*ilvD*, *hydR*, and *hydD*) and downstream of the proposed right end (*ispD*, *flxD*). After transfer of this region,

which contains Tn*5398*, it could be integrated into the recipient chromosome by RecA-dependent homologous recombination (Figure 6.1). The end result would be the integration of Tn*5398* and some of the genes flanking the element.

The results presented in Chapter Four are consistent with this integration hypothesis, since the sequence of the region outside the putative element in the transconjugants was more similar to the donor than to the recipient strain (Figures 4.25 and 4.26). The position of the proposed right end of Tn*5398*, within *orf7*, is also unusual as excision and subsequent transfer of the proposed element would interfere with the transcription and expression of this ORF.

Additional supportive evidence was obtained by an examination of the transfer properties of the MLS resistance determinant from *C. difficile* (Hächler *et al.*, 1987a). In addition to showing that the MLS resistance determinant from *C. difficile* was transferable to *S. aureus*, these researchers also reported that a further 12.8 kb of DNA was transferred from the *C. difficile* donor strain to the *S. aureus* recipient. They partially characterized this additional DNA and showed that not only was it present in the *C. difficile* donor and the *S. aureus* transconjugants, but that it hybridized to DNA that was present in the *C. difficile* recipient. This result indicated that part of the additional DNA that was transferred to *S. aureus* from the *C. difficile* donor was also present in the *C. difficile* recipient strain (Hächler *et al.*, 1987a). Unfortunately, these *S. aureus* transconjugants are not available for further analysis and it has not been possible to demonstrate intergeneric transfer of Tn*5398* in this study.

The MLS resistance marker in these experiments was localized to an approximately 9.5 kb *Hin*dIII fragment in the donor and transconjugants strains (Hächler *et al.*, 1987a), which is in good agreement with the 9.665 kb *Hin*dIII fragment that was cloned to form pJIR1594 in this study. The additional DNA was on a second *Hin*dIII fragment of approximately 12.8 kb. This fragment would correspond to the next *Hin*dIII fragment downstream of that cloned in pJIR1594. Part of this region was sequenced in this study from pJIR1790 and revealed sequences that were present in both the donor and recipient *C. difficile* strains. Likewise, upstream of the proposed left end of Tn*5398*, sequence analysis revealed DNA that was present in both the donor and recipient. It would appear that in conjugation to *S. aureus*, the region transferred included Tn*5398* and additional DNA encoding *C. difficile* housekeeping genes. Genetic analysis of *S. aureus* MLS resistant transconjugants would therefore be advantageous to determine the exact nature of the DNA that is integrated following transfer. This type of analysis is nearly impossible in *C. difficile* transconjugants because of the high similarity between the donor and recipient DNA in the regions outside the currently delineated Tn*5398* element. Analysis of transconjugants resulting from interspecies conjugation may help to define the larger region of DNA that is excised from the *C. difficile* chromosome, forms a circular intermediate and is mobilized to the recipient strain.

The observation that Tn*5398* appears to integrate site-specifically in *C. difficile* but shows no site-specificity in either *B. subtilis* or *S. aureus* (Hächler *et al.*, 1987a; Mullany *et al.*, 1995) could also be explained in terms of the RecA homologous recombination theory. If homologous recombination between the genes flanking Tn*5398* and the *C. difficile* recipient genome is responsible for the

integration of Tn*5398*, it would be logical to assume that the target genes for recombination are only present in one location on the genome, hence the site-specificity in *C. difficile*. In *B. subtilis* and *S. aureus* recipients, the genome structure is different and the homologous recombination event may not occur as efficiently since the target may not be well conserved. It may be that several different recombination events occur in these recipients as individual recombination events between different genes flanking Tn*5398* and homologous genes in the recipient may be possible. Hence, there could be many recombination sites in the *B. subtilis* and *S. aureus* genomes, none of which are as conserved as the *C. difficile* target, which could be the reason for the low frequency of transfer of Tn*5398* to these organisms (Hächler *et al.*, 1987a; Mullany *et al.*, 1995) as well as the lack of site-specific integration.

Although the analyses of the Tn*5398* element that are presented in this thesis have raised many questions and obviously more experimentation is required, the work presented here makes a significant contribution to our understanding of what appears to be a very complex genetic element. This work has also improved the current state of knowledge of the *erm* determinants present in *C. difficile* and how they are likely to be disseminated. Further studies on this interesting element await the development of defined genetic methods for genetic analysis in *C. difficile*.

BIBLIOGRAPHY

Aagaard, C., Phan, H., Trevisanato, S. and Garrett, R.A. (1994). A spontaneous point mutation in the single 23S rRNA gene of the thermophilic arachaeon *Sulfolobus acidocaldarius* confers multiple drug resistance. *J. Bacteriol.* **176,** 7744-7747.

Abraham, L.J., Berryman, D.I. and Rood, J.I. (1988). Hybridization analysis of the class P tetracycline resistance determinant from the *Clostridium perfringens* R-plasmid, pCW3. *Plasmid* **19**, 113-120.

Abraham, L.J. and Rood, J.I. (1985a). Cloning and analysis of the *Clostridium perfringens* tetracycline resistance plasmid, pCW3. *Plasmid* **13**, 155-162.

Abraham, L.J. and Rood, J.I. (1985b). Molecular analysis of transferable tetracycline resistance plasmids from *Clostridium perfringens*. *J. Bacteriol.* **161**, 636-640.

Abraham, L.J. and Rood, J.I. (1987). Identification of Tn*4451* and Tn*4452*, chloramphenicol resistance transposons from *Clostridium perfringens*. *J. Bacteriol.* **169**, 1579-1584.

Aktories, K., Selzer, J., Hofmann, F. and Just, I. (1997). Molecular mechanisms of action of *Clostridium difficile* toxins A and B. In Rood, J.I., McClane, B.A., Songer, J.G. and Titball, R.W. (eds.), *The Clostridia: molecular biology and pathogenesis.* Academic Press, San Diego, pp. 393-407.

Allen, S.P. and Blaschek, H.P. (1988). Electroporation-induced transformation of intact cells of *Clostridium perfringens*. *Appl. Environ. Microbiol.* **54**, 2322-2324.

Allignet, J. and El Solh, N. (1995). Diversity among the gram-positive acetyltransferases inactivating streptogramin A and structurally related compounds and characterization of a new staphylococcal determinant, *vatB*. *Antimicrob. Agents Chemother.* **39**, 2027-2036.

Allignet, J. and El Solh, N. (1997). Characterization of a new staphylococcal gene, *vgaB*, encoding a putative ABC transporter conferring resistance to streptogramin A and related compounds. *Gene* **202**, 133-138.

Allignet, J., Liassine, N. and El Solh, N. (1998). Characterization of a staphylococcal plasmid related to pUB110 and carrying two novel genes, *vatC* and *vgbB*, encoding resistance to streptogramins A and B and similar antibiotics. *Antimicrob. Agents Chemother.* **42**, 1794-1798.

Allignet, J., Loncle, V. and El Solh, N. (1992). Sequence of a staphylococcal plasmid gene, *vga*, encoding a putative ATP-binding proteins involved in resistance to virginiamycin A-like antibiotics. *Gene* **117**, 45-51.

Allignet, J., Loncle, V., Mazodier, P. and El Solh, N. (1988). Nucleotide sequence of a staphylococcal plasmid gene, *vgb*, encoding a hydrolase inactivating the B components of virginiamycin-like antibiotics. *Plasmid* **20**, 271-275.

Allignet, J., Loncle, V., Simenel, C., Delepierre, M. and El Solh, N. (1993). Sequence of staphylococcal gene, *vat*, encoding an acetyltransferase inactivating the A-type compounds of virginiamycin-like antibiotics. *Gene* **130**, 91-98.

Alonso, A., Sanchez, P. and Martínez, J.L. (2000). *Stenotrophomonas maltophila* D457R contains a cluster of genes from Gram-positive bacteria involved in antibiotic and heavy metal resistance. *Antimicrob. Agents Chemother.* **44**, 1778-1782.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Argoudelis, A.D. and Coats, J.H. (1969). Microbial transformation of antibiotics. II. Phosphorylation of lincomycin by *Streptomyces* species. *J. Antibiot.* **22**, 341-343.

Argoudelis, A.D., Coats, J.H. and Mizsak, S.A. (1977). Microbial transformation of antibiotics. Clindamycin ribonucleotides. *J. Antibiot.* **30**, 474-487.

Arthur, M., Autissier, D. and Courvalin, P. (1986). Analysis of the nucleotide sequence of the *ereB* gene encoding the erythromycin esterase type II. *Nucleic Acids Res.* **14**, 4987-4999.

Arthur, M., Brisson-Noël, A. and Courvalin, P. (1987). Origin and evolution of genes specifying resistance to macrolide, lincosamide and streptogramin antibiotics: data and hypotheses. *J. Antimicrob. Chemother.* **20**, 783-802.

Arthur, M., Molinas, C., Mabilat, C. and Courvalin, P. (1990). Detection of erythromycin resistance by the polymerase chain reaction using primers in conserved regions of *erm* rRNA methylase genes. *Antimicrob. Agents Chemother.* **34**, 2024-2026.

Awad, M.M., Ellemor, D.M., Bryant, A.E., Matsushita, O., Boyd, R.L., Stevens, D.L., Emmins, J.J. and Rood, J.I. (2000). Construction and virulence testing of a collagenase mutant of *Clostridium perfringens*. *Microb. Pathog.* **28**, 107-117.

Awad, M.M. and Rood, J.I. (1997). Isolation of α-toxin, θ-toxin and κ-toxin mutants of *Clostridium perfringens* by Tn*916* mutagenesis. *Microb. Pathog.* **22**, 275-284.

Bannam, T.L., Crellin, P.K. and Rood, J.I. (1995). Molecular genetics of the chloramphenicol-resistance transposon Tn*4451* from *Clostridium perfringens*: the TnpX site-specific recombinase excises a circular transposon molecule. *Mol. Microbiol.* **16**, 535-551.

Bannam, T.L. and Rood, J.I. (1991). Relationship between the *Clostridium perfringens catQ* gene product and chloramphenicol acetyltransferases from other bacteria. *Antimicrob. Agents Chemother.* **35**, 471-476.

Bannam, T.L. and Rood, J.I. (1993). *Clostridium perfringens-Escherichia coli* shuttle vectors that carry single antibiotic resistance determinants. *Plasmid* **229**, 233-235.

Bannam, T.L. and Rood, J.I. (1999). Identification of structural and functional domains of the tetracycline efflux protein TetA(P) from *Clostridium perfringens*. *Microbiology* **145**, 2947-2955.

Banno, Y., Kobayashi, T., Kono, H., Watanabe, K., Ueno, K. and Nozawa, Y. (1984). Biochemical characterization and biologic actions of two toxins (D-1 and D-2) from *Clostridium difficile*. *Rev. Infect. Dis.* **6**, S11-S20.

Bartlett, J.G., Onderdonk, A.B., Cisneros, A.B. and Kasper, D.L. (1977). Clindamycin-associated colitis due to a toxin-producing species of *Clostridium* in hamsters. *J. Infect. Dis.* **136**, 701-705.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L.L. (2000). The Pfam protein families database. *Nucleic Acids Res.* **28**, 263-266.

Bedzyk, L.A., Shoemaker, N.B., Young, K.E. and Salyers, A.A. (1992). Insertion and excision of *Bacteroides* conjugative chromosomal elements. *J. Bacteriol.* **174**, 166-172.

Bennett, P.M. (1999). Integrons and gene cassettes: a genetic construction kit for bacteria. *J. Antimicrob. Chemother.* **43**, 1-4.

Berry, A.P. and Levett, P.N. (1986). Chronic diarrhoea in dogs associated with *Clostridium difficile* infection. *Vet. Rec.* **118**, 102-103.

Berryman, D.I., Lyristis, M. and Rood, J.I. (1994). Cloning and sequence analysis of *ermQ*, the predominant macrolide-lincosamide, streptogramin B resistance gene in *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **38**, 1041-1046.

Berryman, D.I. and Rood, J.I. (1989). Cloning and hybridization analysis of *ermP*, a macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **33**, 1346-1353.

Berryman, D.I. and Rood, J.I. (1995). The closely related *ermB-ermAM* genes from *Clostridium perfringens, Enterococcus faecalis* (pAMβ1), and *Streptococcus agalactiae* (pIP501) are flanked by variants of a directly repeated sequence. *Antimicrob. Agents Chemother.* **39**, 1830-1834.

Bibb, M.J., Janssen, G.R. and Ward, J.M. (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene* **38**, 215-226.

Borriello, S.P. (1998). Pathogenesis of *Clostridium difficile* infection. *J. Antimicrob. Chemother.* **41**, 13-19.

Borriello, S.P., Davies, H.A. and Barclay, F.E. (1988). Detection of fimbriae amongst strains of *Clostridium difficile*. *FEMS Microbiol. Lett.* **49**, 65-67.

Bozdogan, B., Berrezouga, L., Kuo, M.S., Yurek, D.A., Farley, K.A., Stockman, B.J. and Leclercq, R. (1999). A new resistance gene, *linB*, conferring resistance to lincosamides by nucleotidylation in *Enterococcus faecium* HM1025. *Antimicrob. Agents Chemother.* **43**, 925-929.

Bradshaw, M., Goodnough, M. and Johnson, E. (1998). Conjugative transfer of the *E. coli-C. perfringens* shuttle vector pJIR1457 to *Clostridium botulinum* type A strains. *Plasmid* **40**, 233-237.

Brantl, S., Kummer, C. and Behnke, D. (1994). Complete nucleotide sequence of plasmid pGB3631, a derivative of the *Streptococcus agalactiae* plasmid pIP501. *Gene* **142**, 155-156.

Braun, M., Herholz, C., Straub, R., Choisat, B., Frey, J., Nicolet, J. and Kuhnert, P. (2000). Detection of the ADP-ribosyltransferase toxin gene (*cdtA*) and its activity in *Clostridium difficile* isolates from equidae. *FEMS Microbiol. Lett.* **184**, 29-33.

Brazier, J.S. (1998). The diagnosis of *Clostridium difficile*-associated disease. *J. Antimicrob. Chemother.* **41**, 29-40.

Brefort, G., Magot, M., Ionesco, H. and Sebald, M. (1977). Characterization and transferability of *Clostridium perfringens* plasmids. *Plasmid* **1**, 52-66.

Brehm, J., Salmond, G. and Minton, N. (1987). Sequence of the adenine methylase gene of the *Streptococcus faecalis* plasmid pAMβ1. *Nucleic Acids Res.* **15**, 3177.

Brisson-Noël, A. and Courvalin, P. (1986). Nucleotide sequence of gene *linA* encoding resistance to lincosamides in *Staphylococcus haemolyticus*. *Gene* **43**, 247-253.

Brisson-Noël, A. and Courvalin, P. (1988). Evidence for natural gene transfer from Gram-positive cocci to *Escherichia coli*. *J. Bacteriol.* **170**, 1739-1745.

Brisson-Noël, A., Delrieu, P., Samain, D. and Courvalin, P. (1988). Inactivation of lincosaminide antibiotics in *Staphylococcus*. Identification of lincosaminide *O*-nucleotidyltransferases and comparison of the corresponding resistance genes. *J. Biol. Chem.* **263**, 15880-15887.

Brown, H.J., Stokes, H.W. and Hall, R.M. (1996). The integrons In0, In2, and In5 are defective transposon derivates. *J. Bacteriol.* **178**, 4429-4437.

Burman, W.J., Stone, B.L., Brown, B.A., Wallace Jnr., R.J. and Böttger, E.C. (1998). AIDS-related *Mycobacterium kansasii* infection with initial resistance to clarithromycin. *Diagn. Microbiol. Infect. Dis.* **31**, 369-371.

Bussiere, D.E., Muchmore, S.W., Dealwis, C.G., Schluckebier, G., Nienaber, V.L., Edalji, R.P., Walter, K.A., Ladror, U.S., Holzman, T.F. and Abad-Zapatero, C. (1998). Crystal structure of ErmC', an rRNA methyltransferase which mediates antibiotic resistance in bacteria. *Biochemistry* **37**, 7103-7112.

Caparon, M.G. and Scott, J.R. (1989). Excision and insertion of the conjugative transposon Tn*916* involves a novel recombination mechanism. *Cell* **59**, 1027-1034.

Catchpole, I. and Dyke, K.G.H. (1990). A *Staphylococcus aureus* plasmid that specifies constitutive macrolide-lincosamide-streptogramin B resistance contains a novel deletion in the *ermC* attenuator. *FEMS Microbiol. Lett.* **69**, 43-48.

Catchpole, I., Thomas, C., Davies, A. and Dyke, K.G. (1988). The nucleotide sequence of *Staphylococcus aureus* plasmid pT48 conferring inducible macrolide-lincosamide-streptogramin B resistance and comparison with similar plasmids expressing constitutive resistance. *J. Gen. Microbiol.* **134**, 697-709.

Ceglowski, P. and Alonso, J.C. (1994). Gene organization of the *Streptococcus pyogenes* plasmid pDB101: sequence analysis of the *orfη-copS* region. *Gene* **145**, 33-39.

Ceglowski, P., Boitsov, A., Chai, S. and Alonso, J.C. (1993). Analysis of the stabilization system of pSM19035-derived plasmid pBT233 in *Bacillus subtilis*. *Gene* **136**, 1-12.

Celli, J. and Trieu-Cuot, P. (1998). Circularization of Tn*916* is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol. Microbiol.* **28**, 103-117.

Cheng, Q., Paszkiet, B.J., Shoemaker, N., Gardner, J.F. and Salyers, A. (2000). Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. *J. Bacteriol.* **182**, 4035-4043.

Cheng, X. (1995). Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293-318.

Cheng, X., Kumar, S., Pósfai, J., Pflugrath, J.W. and Roberts, R.J. (1993). Crystal structure of the *Hha*I DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* **74**, 299-307.

Chopra, I. and Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232-260.

Clancy, J., Petitpas, J.W., Dib-Hajj, F., Yuan, W., Cronan, M., Kamath, A., Bergeron, J. and Retsema, J. (1996). Molecular cloning and functional analysis of a novel macrolide-resistance determinant *mefA* from *Streptococcus pyogenes*. *Mol. Microbiol.* **22**, 867-879.

Clewell, D.B. and Flannagan, S.E. (1993). The conjugative transposons of gram-positive bacteria. In Clewell, D.B. (ed.) *Bacterial Conjugation.* Plenum Press, New York.

Clewell, D.B., Flannagan, S.E. and Jaworski, D.D. (1995). Unconstrained bacterial promiscuity: the Tn*916*-Tn*1545* family of conjugative transposons. *Trends Microbiol.* **3**, 229-236.

Cocito, C. (1979). Antibiotics of the virginiamycin family, inhibitors which contain synergistic components. *Microbiological Reviews* **43**, 145-198.

Cocito, C., Di Giambattista, M., Nyssen, E. and Vannuffel, P. (1997). Inhibition of protein synthesis by streptogramins and related antibiotics. *J. Antimicrob. Chemother.* **39**, 7-13.

Cohen, S.P., Hächler, H. and Levy, S.B. (1993). Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli. J. Bacteriol.* **175**, 1484-1492.

Collignon, P.J. (1997). Antibiotic resistance: is it leading to the re-emergence of many infections from the past? *Recent Advances in Microbiology* **5**, 203-256.

Cooper, A.J., Kalinowski, A.P., Shoemaker, N. and Salyers, A. (1997). Construction and characterization of a *Bacteroides thetaiotaomicron recA* mutant: transfer of *Bacteroides* integrated conjugative elements is RecA independent. *J. Bacteriol.* **179**, 6221-6227.

Courvalin, P. and Carlier, C. (1986). Transposable multiple antibiotic resistance in *Streptococcus pneumoniae. Mol. Gen. Genet.* **205**, 291-297.

Courvalin, P. and Carlier, C. (1987). Tn*1545*: a conjugative shuttle transposon. *Mol. Gen. Genet.* **206**, 259-264.

Crellin, P.K. and Rood, J.I. (1997). The resolvase/invertase domain of the site-specific recombinase TnpX is functional and recognizes a target sequence that resembles the junction of the circular form of the *Clostridium perfringens* transposon Tn*4451. J. Bacteriol.* **179**, 5148-5156.

Crellin, P.K. and Rood, J.I. (1998). Tn*4451* from *Clostridium perfringens* is a mobilizable transposon that encodes the functional Mob protein, TnpZ. *Mol. Microbiol.* **27**, 631-642.

Dam, M., Douthwaite, S., Tenson, T. and Mankin, A.S. (1996). Mutations in domain II of 23 S rRNA facilitate translation of a 23 S rRNA-encoded pentapeptide conferring erythromycin resistance. *J. Mol. Biol.* **259**, 1-6.

Davies, H.A. and Borriello, S.P. (1990). Detection of capsule in strains of *Clostridium difficile. Microb. Pathog.* **9**, 141-146.

Depardieu, F. and Courvalin, P. (2001). Mutation in 23S rRNA responsible for resistance to 16-membered macrolides and streptogramins in *Streptococcus pneumoniae. Antimicrob. Agents Chemother.* **45**, 319-323.

Devriese, L.A. (1980). Two new types of resistance to lincomycin in pathogenic staphylococci from animals. *Ann. Inst. Pasteur (Paris).* **131B**, 261-266.

Dhillon, N. and Leadlay, P.F. (1990). A repeated decapeptide motif in the C-terminal domain of the ribosomal RNA methyltransferase from the erythromycin producer *Saccharopolyspora erythraea. FEBS Lett.* **262**, 189-193.

Docherty, A., Grandi, G., Grandi, R., Gryczan, T.J., Shivakumar, A.G. and Dubnau, D. (1981). Naturally occurring macrolide-lincosamide-streptogramin B resistance in *Bacillus licheniformis. J. Bacteriol.* **145**, 129-137.

Douthwaite, S., Prince, J. and Noller, H.F. (1985). Evidence for functional interaction between domains II and V of 23S ribosomal RNA from an erythromycin resistant mutant. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8330-8334.

Dubnau, D. and Monod, M. (1986). The regulation and evolution of MLS resistance. *Banbury Rep.* **24**, 369-387.

Dutta, G.N. and Devriese, L.A. (1981). Degradation of macrolide-lincosamide-streptogramin antibiotics by *Lactobacillus* strains from animals. *Ann. Inst. Pasteur (Paris).* **132A**, 51-57.

Dutta, G.N. and Devriese, L.A. (1982). Resistance to macrolide, lincosamide and streptogramin antibiotics and degradation of lincosamide antibiotics in streptococci from bovine mastitis. *J. Antimicrob. Chemother.* **10**, 403-408.

Eady, E.A., Ross, J.I., Tipper, J.L., Walters, C.E., Cove, J.H. and Noble, W.C. (1993). Distribution of genes encoding erythromycin ribosomal methylases and an erythromycin efflux pump in epidemiologically distinct groups of staphylococci. *Antimicrob. Agents Chemother.* **31**, 211-217.

Easter, C.L., Schwab, H. and Helinski, D.R. (1998). Role of the *parCBA* operon of the broad-host range plasmid RK2 in stable plasmid maintenance. *J. Bacteriol.* **180**, 6023-6030.

Eaton, R.W. and Timmis, K.N. (1986). Characterization of a plasmid -specified pathway for catabolism of isopropylbenzene in *Pseudomonas putida* RE204. *J. Bacteriol.* **168**, 123-131.

Epp, J.K., Burgett, S.G. and Schoner, B.E. (1987). Cloning and nucleotide sequence of a carbomycin-resistance gene from *Streptomyces thermotolerans*. *Gene* **53**, 73-83.

Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 938-940.

Eveillard, M., Fourel, V., Barc, M.-C., Kerneis, S., Coconnier, M.H., Karjalainen, T., Bourlioux, P. and Servin, A. (1993). Identification and characterization of adhesive factors of *Clostridium difficile* involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture. *Mol. Microbiol.* **7**, 371-381.

Farrow, K.A. (1995). Genetic analysis in *Clostridium difficile*. *Department of Microbiology*. Monash University, Clayton.

Farrow, K.A., Lyras, D. and Rood, J.I. (2000). The macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium difficile* 630 contains two *erm*(B) genes. *Antimicrob. Agents Chemother.* **44**, 411-413.

Fink, P.S. (1993). Biosynthesis of the branched-chain amino acids. In Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds.), *Bacillus subtilis and other Gram positive bacteria: biochemistry, physiology and molecular genetics.* American Society for Microbiology, Washington D.C.

Flannagan, S.E. and Clewell, D.B. (1991). Conjugative transfer of Tn*916* in *Enterococcus faecalis*: *trans* activation of homologous transposons. *J. Bacteriol.* **173**, 7136-7141.

Flannagan, S.E., Zitzow, L., Su, Y.A. and Clewell, D.B. (1994). Nucleotide sequence of the 18-kb conjugative transposon Tn*916* from *Enterococcus faecalis*. *Plasmid* **32**, 350-354.

Fons, M., Hege, T., Ladire, M., Raibaud, P., Ducluzeau, R. and Maguin, E. (1997). Isolation and characterization of a plasmid from *Lactobacillus fermentum* conferring erythromycin resistance. *Plasmid* **37**, 199-203.

Fraimow, H. and Knob, C. (1997). Amplification of macrolide efflux pumps *msr* and *mef* from *Enterococcus faecium* by polymerase chain reaction, abstr. A-125. *Abstracts of the 98th General Meeting of the American Society for Microbiology*. American Society for Microbiology, Washington, D.C., p. 22.

Franke, A.E. and Clewell, D.B. (1981). Evidence for a chromosome-borne resistance transposon (Tn*916*) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. *J. Bacteriol.* **145**, 494-502.

Frazier, K.S., Herron, A.J., Hines, M.E., Gaskin, J.M. and Altman, N.H. (1993). Diagnosis of enteritis and enterotoxemia due to *Clostridium difficile* in captive ostriches (*Struthio camelus*). *J. Vet. Diagn. Invest.* **5**, 623-625.

Fujisawa, Y. and Weisblum, B. (1981). A family of r-determinants in *Streptomyces* spp. that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibiotics. *J. Bacteriol.* **146**, 621-631.

Furneri, P.M., Rappazzo, G., Musumarra, M.P., Di Pietro, P., Catania, L.S. and Roccasalva, L.S. (2001). Two new point mutations at A2062 associated with resistance to 16-membered macrolide antibiotics in mutant strains of *Mycoplasma hominis*. *Antimicrob. Agents Chemother.* **45**, 2958-2960.

Gale, E.F., Cundliffe, E., Reynolds, P.E., Rickmond, M.K. and Waring, M.J. (1981). The molecular basis of antibiotic action. Wiley, London, pp. 462-468.

Gauthier, A., Turmel, M. and Lemieux, C. (1988). Mapping of chloroplast mutations conferring resistance to antibiotics in *Chlamydomonas*: evidence for a novel site of streptomycin resistance in the small subunit rRNA. *Mol. Gen. Genet.* **214**.

Geoghegan, S.M., Mayhew, S.G., Yalloway, G.N. and Butler, G. (2000). Cloning, sequencing and expression of the gene for flavodoxin from *Megasphaera elsdenii* and the effects of removing the protein negative charge that is closest to N(10) of the bound FMN. *Eur. J. Biochem.* **267**, 4434-4444.

Giuliano, M., Piemonte, F. and Gianfrilli, M. (1988). Production of an enterotoxin different from toxin A by *Clostridium difficile*. *FEMS Microbiol. Lett.* **50**, 191-194.

Gorbach, S. (1999). Antibiotics and *Clostridium difficile*. *N. Engl. J. Med.* **341**, 1690-1691.

Gülke, I., Pfeifer, G., Liese, J., Fritz, M., Hofmann, F., Aktories, K. and Barth, H. (2001). Characterization of the enzymatic component of the ADP-ribosyltransferase toxin CDTa from *Clostridium difficile*. *Infect. Immun.* **69**, 6004-6011.

Hächler, H., Berger-Bächi, B. and Kayser, F.H. (1987a). Genetic characterization of a *Clostridium difficile* erythromycin-clindamycin resistance determinant that is transferable to *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **31**, 1039-1045.

Hächler, H. and Kayser, F.H. (1993). Genetics and molecular biology of antibiotic resistance in *Clostridium difficile*: general and specific overview. In Sebald, M. (ed.) *Genetics and Molecular Biology of Anaerobic Bacteria.* Springer-Verlag, New York, pp. 156-173.

Hächler, H., Kayser, F.H. and Berger-Bächi, B. (1987b). Homology of a transferable tetracycline resistance determinant of *Clostridium difficile* with *Streptococcus* (*Enterococcus*) *faecalis* transposon Tn*916*. *Antimicrob. Agents Chemother.* **31**, 1033-1038.

Hajduk, P.J., Dinges, J., Schkeryantz, J.M., Janowick, D., Kaminski, M., Tufano, M., Augeri, D.J., Petros, A., Nienaber, V.L., Zhong, P., Hammond, R., Coen, M., Beutel, B., Katz, L. and Fesik, S.W. (1999). Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J. Med. Chem.* **42**, 3852-3859.

Hall, R.M. and Collis, C.M. (1995). Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.* **15**, 593-600.

Halula, M.C., Manning, S. and Macrina, F.L. (1991). Nucleotide sequence of *ermFU*, a macrolide-lincosamide-streptogramin (MLS) resistance gene encoding an RNA methylase from the conjugal element of *Bacteroides fragilis* V503. *Nucleic Acids Res.* **19**, 3453.

Hanahan, D. (1985). Techniques for transformation of *E. coli*. In Glover, D.M. (ed.) *DNA cloning: a practical approach.* IRL Press Ltd., Oxford, England, pp. 109-135.

Hanessian, S. and Sgarbi, P.W.M. (2000). Design and synthesis of mimics of *S*-adenosyl-L-Homocysteine as potential inhibitors of erythromycin methyltransferases. *Bioorg. Med. Chem. Lett.* **10**, 433-437.

Hara, Q. and Hutchinson, C.R. (1990). Cloning of midecamycin (MLS)-resistance genes from *Streptomyces mycarofaciens*, *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J. Antibiot.* **43**, 977-991.

Haroche, J., Allignet, J., Buchrieser, C. and El Solh, N. (2000). Characterization of a variant of *vga*(A) conferring resistance to Streptogramin A and related compounds. *Antimicrob. Agents Chemother.* **44**, 2271-2275.

Harris, E.H., Burkhart, B.D., Gillham, N.W. and Boynton, J.E. (1989). Antibiotic resistance mutations in the chloroplast 16S and 23S rRNA genes of *Chlamydomonas reinhardtii*: correlation of genetic and physical maps of the chloroplast genome. *Genetics* **123**, 281-292.

Hayter, P.M. and Dale, J.W. (1984). Detection of plasmids in clinical isolates of *Clostridium difficile*. *Microbios Letters* **27**, 151-156.

Hecht, G., Pothoulakis, C., LaMont, J.T. and Madara, J.L. (1988). *Clostridium difficile* toxin A perturbs cytoskeletal structure and junction permeability in cultured human epithelial cells. *J. Clin. Invest.* **82**, 1516-1524.

Hillen, W. and Berens, C. (1994). Mechanisms underlying expression of Tn*10* encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48**, 345-369.

Hinerfeld, D. and Churchward, G. (2001). Xis protein of the conjugative transposon Tn*916* plays dual opposing roles in transposon excision. *Mol. Microbiol.* **41**, 1459-1467.

Hodgson, A.L.M., Krywult, J. and Radford, A.J. (1990). Nucleotide sequence of the erythromycin resistance gene from the *Corynebacterium* plasmid pNG2. *Nucleic Acids Res.* **18**, 1891.

Horinouchi, S., Byeon, W.-H. and Weisblum, B. (1983). A complex attenuator regulates inducible resistance to macrolides, lincosamides, and streptogramin type B antibiotics in *Streptococcus sanguis*. *J. Bacteriol.* **154**, 1252-1262.

Horinouchi, S. and Weisblum, B. (1982). Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide and streptogramin type B antibiotics. *J. Bacteriol.* **150**, 804-814.

Hou, C.T., Perlman, D. and Schallock, M.R. (1970). Microbial transformation of peptide antibiotics. VI. Purification and properties of a peptide lactonase hydrolyzing dihydrostaphylomycin S. *J. Antibiot.* **23**, 35-42.

Hue, K.K. and Bechhofer, D.H. (1992). Regulation of the macrolide-lincosamide-streptogramin B resistance gene *ermD*. *J. Bacteriol.* **174**, 5860-5868.

Humphreys, G.O., Willshaw, G.A., Smith, H.R. and Anderson, E.S. (1976). Mutagenesis of plasmid DNA with hydroxylamine: isolation of mutants of multi-copy plasmids. *Mol. Gen. Genet.* **145**, 101-108.

Inouye, M., Morohoshi, T., Horinouchi, S. and Beppu, T. (1994). Cloning and sequences of two macrolide-resistance-encoding genes from mycinamicin-producing *Micromonospora griseorubida*. *Gene* **141**, 39-46.

Jaworski, D.D. and Clewell, D.B. (1995). A functional origin of transfer (*oriT*) on the conjugative transposon Tn*916*. *J. Bacteriol.* **177**, 6644-6651.

Jaworski, D.D., Flannagan, S.E. and Clewell, D.B. (1996). Analyses of *traA, int-Tn,* and *xis-Tn* mutations in the conjugative transposon Tn*916* in *Enterococcus faecalis*. *Plasmid* **36**, 201-208.

Jenkins, G. and Cundliffe, E. (1991). Cloning and characterization of two genes from *Streptomyces lividans* that confer inducible resistance to lincomycin and macrolide antibiotics. *Gene* **108**, 55-62.

Johanesen, P.A., Lyras, D., Bannam, T.L. and Rood, J.I. (2001). Transcriptional analysis of the *tet*(P) operon from *Clostridium perfringens*. *J. Bacteriol.* **183**, In press.

Johnson, S., Samore, M.H., Farrow, K.A., Killgore, G.E., Tenover, F.C., Lyras, D., Rood, J.I., DeGirolami, P.C., Baltch, A.L., Rafferty, M.E., Pear, S.M. and Gerding, D.N. (1999). Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. *N. Engl. J. Med.* **341**, 1645-1651.

Jones, T.A., Cowan, S., Zou, J.Y. and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A.* **47**, 110-119.

Just, I., Selzer, J., Wilm, M., von Eichel-Streiber, C., Mann, M. and Aktories, K. (1995). Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375**, 500-503.

Justus, P.G., Martin, J.L., Goldberg, D.A., Taylor, N.S., Bartlett, J.G., Alexander, R.W. and Mathias, J.R. (1982). Myoelectric effects of *Clostridium difficile*: motility-altering factors distinct from its cytotoxin and enterotoxin in rabbits. *Gastroenterology* **83**, 836-843.

Kamimiya, S. and Weisblum, B. (1988). Translational attenuation control of *ermSF*, an inducible resistance determinant encoding rRNA *N*-methyltransferase from *Streptomyces fradiae*. *J. Bacteriol.* **170**, 1800-1811.

Kamimiya, S. and Weisblum, B. (1997). Induction of *ermSV* by 16-membered-ring macrolide antibiotics. *Antimicrob. Agents Chemother.* 41, 530-534.

Karjalainen, T., Barc, M.-C., Collignon, A., Trollé, S., Boureau, H., Cotte-Laffitte, J. and Bourlioux, P. (1994). Cloning of a genetic determinant from *Clostridium difficile* involved in adherence to tissue culture cells and mucus. *Infect. Immun.* **62**, 4347-4355.

Karlsson, M., Fellstrom, C., Heldtander, M.U., Johansson, K.E. and Franklin, A. (1999). Genetic basis of macrolide and lincosamide resistance in *Brachyspira* (*Serpulina*) *hyodysenteriae*. *FEMS Microbiol. Lett.* 172, 255-260.

Kataja, J., Seppälä, H., Skurnik, M., Sarkkinen, H. and Huovinen, P. (1998). Different erythromycin resistance mechanisms in group C and group G Streptococci. *Antimicrob. Agents Chemother.* **42**, 1493-1494.

Kawata, T., Takeoka, A., Takumi, K. and Masuda, K. (1984). Demonstration and preliminary characterisation of a regular array in the cell wall of *Clostridium difficile*. *FEMS Microbiol. Lett.* **24**, 323-328.

Kelemen, G.H., Zalacain, M., Culebras, E., Seno, E.T. and Cundliffe, E. (1994). Transcriptional attenuation control of the tylosin-resistance gene *tlrA* in *Streptomyces fradiae*. *Mol. Microbiol.* 14, 833-842.

Kelly, C.P., Pothoulakis, C. and LaMont, J.T. (1994). *Clostridium difficile* colitis. *N. Engl. J. Med.* **330**, 257-262.

Kennan, R.M., McMurray, L.M., Levy, S.B. and Rood, J.I. (1997). Glutamate residues located within putative transmembrane helices are essential for TetA(P)-mediated tetracycline efflux. *J. Bacteriol.* 179, 7011-7015.

Kim, C.H., Otake, N. and Yonehara, Y. (1974). Studies on mikamycin B lactonase. Part I. Degradation of mikamycin B by *Streptomyces mitakaensis*. *J. Antibiot.* **27**, 903-908.

Kim, E., Kim, Y. and Kim, C.K. (1996). Genetic structures of the genes encoding 2,3-dihydroxybiphenyl 1,2-dioxygenase and 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid hydrolase from biphenyl- and 4-chlorobiphenyl-degrading *Pseudomonas* sp. strain DJ-12. *Appl. Environ. Microbiol.* **62**, 262-265.

Kim, H.-S., Choi, E.-C.C. and Kim, B.-K. (1993). A macrolide-lincosamide-streptogramin B resistance determinant from *Bacillus anthracis* 590: cloning and expression of *ermJ*. *J. Gen. Microbiol.* 139, 601-607.

Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369.

Koide, Y., Nakamura, A., Uozumi, T. and Beppu, T. (1986). Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. *J. Bacteriol.* **167**, 110-116.

Koutsis, K. (1993). Studies on the replication of plasmids from *Clostridium perfringens*. *Department of Microbiology*. Monash University, Clayton.

Kovalic, D., Giannattasio, R.B., Jin, H.-J. and Weisblum, B. (1994). 23S rRNA domain V, a fragment that can be specifically methylated in vitro by the ErmSF(TlrA) methyltransferase. *J. Bacteriol.* **176**, 6992-6998.

Kwak, J.-H., Choi, E.-C.C. and Weisblum, B. (1991). Transcriptional attenuation control of *ermK*, a macrolide-lincosamide-streptogramin B resistance determinant from *Bacillus licheniformis*. *J. Bacteriol.* **173**, 4725-4735.

Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. and Saenger, W. (1994). Three-dimensional structure of the adenine specific DNA methyltransferase M.*Taq*I in complex with the cofactor *S*-adenosylmethionine. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10957-10961.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lai, C.J. and Weisblum, B. (1971). Altered methylation of ribosomal RNA in an erythromycin-resistant strain of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 856-860.

Lampson, B.C. and Parisi, J.T. (1986). Nucleotide sequence of the constitutive macrolide-lincosamide-streptogramin B resistance plasmid pNE131 from *Staphylococcus epidermidis* and homologies with *Staphylococcus aureus* plasmids pE194 and pSN2. *J. Bacteriol.* **167**, 888-892.

Leclercq, R., Brisson-Noël, A., Duval, J. and Courvalin, P. (1987). Phenotypic expression and genetic heterogeneity of lincosamide inactivation in *Staphylococcus* spp. *Antimicrob. Agents Chemother.* **31**, 1887-1891.

Leclercq, R., Carlier, C., Duval, J. and Courvalin, P. (1985). Plasmid-mediated resistance to lincomycin by inactivation in *Staphylococcus haemolyticus*. *Antimicrob. Agents Chemother.* **28**, 421-424.

Leclercq, R. and Courvalin, P. (1991). Bacterial resistance to macrolide, lincosamide, and streptogramin antibiotics by target modification. *Antimicrob. Agents Chemother.* **35**, 1267-1272.

Levy, S.B. (1984). Resistance to the tetracyclines. In Bryan, L.E. (ed.) *Antimicrobial Drug Resistance.* Academic Press, Orlando, pp. 191-240.

Levy, S.B. (1992). Active efflux mechanisms for antimicrobial resistance. *Antimicrob. Agents Chemother.* **36**, 695-703.

Levy, S.B., McMurray, L.M., Barbosa, T.M., Burdett, V., Courvalin, P., Hillen, W., Roberts, M.C., Rood, J.I. and Taylor, D.E. (1999). Nomenclature for new tetracycline resistance determinants. *Antimicrob. Agents Chemother.* **43**, 1523-1524.

Levy, S.B., McMurray, L.M., Burdett, V., Courvalin, P., Hillen, W., Roberts, M.C. and Taylor, D.E. (1989). Nomenclature for tetracycline resistance determinants. *Antimicrob. Agents Chemother.* **33**, 1373-1374.

Lima, A.A.M., Lyerly, D.M., Wilkins, T.D., Innes, D.J. and Geurrant, R.L. (1988). Effects of *Clostridium difficile* toxins A and B in rabbit small and large intestine *in vivo* and on cultured cells *in vitro*. *Infect. Immun.* **56**, 582-588.

Lina, G., Quaglia, A., Reverdy, M.-E., Leclercq, R., Vandenesch, F. and Etienne, J. (1999). Distribution of genes encoding resistance to macrolides, lincosamides, and streptogramins among Staphylococci. *Antimicrob. Agents Chemother.* **43**, 1062-1066.

Liyanage, H., Kashket, S., Young, M. and Kashket, E.R. (2001). *Clostridium beijerinckii* and *Clostridium difficile* detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase. *Appl. Environ. Microbiol.* **67**, 2004-2010.

Lodder, G., Schwarz, S., Gregory, P. and Dyke, K. (1996). Tandem duplication in *ermC* translational attenuator of the macrolide-lincosamide-streptogramin B resistance plasmid pSES6 from *Staphylococcus equorum*. *Antimicrob. Agents Chemother.* **40**, 215-217.

Lodder, G., Werckenthin, C., Schwarz, S. and Dyke, K. (1997). Molecular analysis of naturally occurring *ermC*-encoding plasmids in staphylococci isolated from animals with and without previous contact with macrolide/lincosamide antibiotics. *FEMS Immunol. Med. Microbiol.* **18**, 7-15.

Lu, F. and Churchward, G. (1995). Tn*916* target DNA sequences bind the C-terminal domain of integrase protein with different affinities that correlate with transposon insertion frequency. *J. Bacteriol.* **177**, 1938-1946.

Lucier, T.S., Heitzman, K., Liu, S.K. and Hu, P.C. (1995). Transition mutations in the 23S rRNA of erythromycin-resistant isolates of *Mycoplasma pneumoniae*. *Antimicrob. Agents Chemother.* **39**, 2770-2773.

Luna, V.A., Coates, P., Eady, E.A., Cove, J.H., Nguyen, T.T.H. and Roberts, M.C. (1999). A variety of Gram-positive bacteria carry mobile *mef* genes. *J. Antimicrob. Chemother.* **44**, 19-25.

Lyerly, D.M., Lockwood, D.E., Richardson, S.H. and Wilkins, T.D. (1982). Biological activities of toxins A and B of *Clostridium difficile*. *Infect. Immun.* **35**, 1147-1150.

Lyras, D. and Rood, J.I. (1996). Genetic organisation and distribution of tetracycline resistance determinants from *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **40**, 2500-2504.

Lyras, D. and Rood, J.I. (1997). Transposable genetic elements and antibiotic resistance determinants from *Clostridium perfringens* and *Clostridium difficile*. In Rood, J.I., McClane, B., Songer, J. and Titball, R. (eds.), *The Clostridia : molecular biology and pathogenesis.* Academic Press, San Diego, California, pp. 73-92.

Lyras, D. and Rood, J.I. (1998). Conjugative transfer of RP4-*oriT* shuttle vectors from *Escherichia coli* to *Clostridium perfringens*. *Plasmid* **39**, 160-164.

Lyras, D. and Rood, J.I. (2000). Transposition of Tn*4451* and Tn*4453* involves a circular intermediate that forms a promoter for the large resolvase, TnpX. *Mol. Microbiol.* **38**, 588-601.

Lyras, D., Storie, C., Huggins, A.S., Crellin, P.K., Bannam, T.L. and Rood, J.I. (1998). Chloramphenicol resistance in *Clostridium difficile* is encoded on Tn*4453* transposons that are closely related to Tn*4451* from *Clostridium perfringens*. *Antimicrob. Agents Chemother.* **42**, 1563-1567.

Lyristis, M., Bryant, A.E., Sloan, J., Awad, M.M., Nisbet, I.T., Stevens, D.L. and Rood, J.I. (1994). Identification and molecular analysis of a locus that regulates extracellular toxin production in *Clostridium perfringens*. *Mol. Microbiol.* **12**, 761-777.

Mabilat, C. and Courvalin, P. (1988). Gene heterogeneity for resistance to macrolides, lincosamides and streptogramins in *Enterobacteriaceae. Ann. Inst. Pasteur. Microbiol.* **139**, 677-681.

Macrina, F.L. (1993). Conjugal transfer in anaerobic bacteria. In Clewell, D.B. (ed.) *Bacterial Conjugation*. Plenum Press, New York, pp. 331-348.

Madewell, B.R., Tang, Y.J., Jang, S., Madigan, J.E., Hirsch, D.C., Gumerlock, P.H. and Silva, J. (1995). Apparent outbreaks of *Clostridium difficile*-associated diarrhea in horses in a veterinary medical teaching hospital. *J. Vet. Diagn. Invest.* 7, 343-346.

Malone, T., Blumenthal, R.M. and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, 618-632.

Manganelli, R., Ricci, S. and Pozzi, G. (1996). Conjugative transposon Tn*916*: evidence for excision with formation of 5'-protruding termini. *J. Bacteriol.* **178**, 5813-5816.

Manganelli, R., Romano, L., Ricci, S., Zazzi, M. and Pozzi, G. (1995). Dosage of Tn*916* circular intermediates in *Enterococcus faecalis*. *Plasmid* **34**, 48-57.

Marra, D., Pethel, B., Churchward, G. and Scott, J.R. (1999). The frequency of conjugative transposition of Tn*916* is not determined by the frequency of excision. *J. Bacteriol.* **181**, 5414-5418.

Marshall, V.P., Liggett, W.F. and Cialdella, J.I. (1989). Enzymic inactivation of lincosamide and macrolide antibiotics: divalent cation and coenzyme specificities. *J. Antibiot.* **42**, 826-830.

Martin, B., Alloing, G., Mejean, V. and Claverys, J.P. (1987). Constitutive expression of erythromycin resistance mediated by the *ermAM* determinant of plasmid pAMβ1 results from deletion of 5' leader peptide sequences. *Plasmid* **18**, 250-253.

Matsuoka, M., Endou, K., Kobayashi, H., Inoue, M. and Nakajima, Y. (1998). A plasmid that encodes three genes for resistance to macrolide antibiotics in *Staphylococcus aureus*. *FEMS Microbiol. Lett.* **167**, 221-227.

Matsuoka, M., Jánosi, L., Endou, K. and Nakajima, Y. (1999). Cloning and sequences of inducible and constitutive macrolide resistance genes in *Staphylococcus aureus* that correspond to an ABC transporter. *FEMS Microbiol. Lett.* **181**, 91-100.

Mayford, M. and Weisblum, B. (1990). The *ermC* leader peptide: amino acid alterations leading to differential efficiency of induction by macrolide-lincosamide-streptogramin B antibiotics. *J. Bacteriol.* **172,** 3772-3779.

Mayhew, S.G. and Ludwig, M.L. (1975). Flavodoxins and electron transferring flavoproteins. In Boyer, P.D. (ed.) *The Enzymes.* Academic Press, New York,Vol. 12, pp. 57-118.

McCoubrey, J. and Poxton, I.R. (2001). Variation in the surface layer proteins of *Clostridium difficile. FEMS Immunol. Med. Microbiol.* **31,** 131-135.

McFarland, L.V., Mulligan, M.E., Kwok, R.Y. and Stamm, W.E. (1989). Nosocomial acquisition of *Clostridium difficile* infection. *N. Engl. J. Med.* **320,** 204-210.

Meier, A., Kirschner, P., Springer, B., Steingrub, V.A., Brown, B.A., Wallace Jnr., R.J. and Böttger, E.C. (1994). Identification of mutations in the 23S ribosomal RNA gene of clarithromycin resistant *Mycobacterium intracellulare. Antimicrob. Agents Chemother.* **38,** 381-384.

Miller, J.H. (1972). *Experiments in molecular genetics.* Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Milton, I.D., Hewitt, C.L. and Harwood, C.R. (1992). Cloning and sequencing of a plasmid-mediated erythromycin resistance determinant from *Staphylococcus xylosus. FEMS Microbiol. Lett.* **97,** 141-148.

Mitchell, T.J., Ketley, J.M., Burdon, D.W., Candy, D.C.A. and Stephen, J. (1987). Biological mode of action of *Clostridium difficile* toxin A: a novel enterotoxin. *J. Med. Microbiol.* **23,** 211-219.

Moncrief, J.S., Lyerly, D.M. and Wilkins, T.D. (1997). Molecular biology of the *Clostridium difficile* toxins. In Rood, J.I., McClane, B.A., Songer, J.G. and Titball, R.W. (eds.), *The Clostridia: molecular biology and pathogenesis.* Academic Press, San Diego, pp. 369-392.

Monod, M., Mohan, S. and Dubnau, D. (1987). Cloning and analysis of *ermG*, a new macrolide-lincosamide-streptogramin B resistance element from *Bacillus sphaericus. J. Bacteriol.* **169,** 340-350.

Morelle, G. (1989). A plasmid extraction procedure on a miniprep scale. *Focus* **11,** 7-8.

Mullany, P., Pallen, M., Wilks, M., Stephen, J.R. and Tabaqchali, S. (1996). A group II intron in a conjugative transposon from the gram-positive bacterium, *Clostridium difficile. Gene* **174,** 145-150.

Mullany, P., Wilks, M., Lamb, I., Clayton, C., Wren, B.W. and Tabaqchali, S. (1990). Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*. *J. Gen. Microbiol.* **136**, 1343-1349.

Mullany, P., Wilks, M., Puckey, L. and Tabaqchali, S. (1994). Gene cloning in *Clostridium difficile* using Tn*916* as a shuttle conjugative transposon. *Plasmid* **31**, 320-323.

Mullany, P., Wilks, M. and Tabaqchali, S. (1995). Transfer of macrolide-lincosamide-streptogramin B (MLS) resistance in *Clostridium difficile* is linked to a gene homologous with toxin A and is mediated by a conjugative transposon, Tn*5398*. *J. Antimicrob. Chemother.* **35**, 305-315.

Murphy, E. (1985). Nucleotide sequence of *ermA*, a macrolide-lincosamide-streptogramin B (MLS) determinant in *Staphylococcus aureus*. *J. Bacteriol.* **162**, 633-640.

Naglich, J.G. and Andrews Jnr., R.E. (1988). Tn*916*-dependent conjugal transfer of pC194 and pUB110 from *Bacillus subtilis* into *Bacillus thuringiensis* subsp. *israelensis*. *Plasmid* **20**, 113-126.

Nakamura, S., Yamakawa, K., Nakashio, S., Kamiya, S. and Nishida, S. (1987). Correlation between susceptibility to chloramphenicol, tetracycline and clindamycin, and serogroups of *Clostridium difficile*. *Med. Microbiol. Immunol.* **176**, 79-82.

Nash, K.A. and Inderlied, C.B. (1995). Genetic basis of macrolide resistance in *Mycobacterium avium* isolated from patients with disseminated disease. *Antimicrob. Agents Chemother.* **39**, 2625-2630.

Nawaz, M.S., Khan, S.A., Khan, A.A., Khambaty, F.M. and Cerniglia, C.E. (2000). Comparative molecular analysis of erythromycin-resistance determinants in staphylococcal isolates of poultry and human origin. *Mol. Cell. Probes* **14**, 311-319.

Needham, C., Noble, W.C. and Dyke, K.G. (1995). The staphylococcal insertion sequence IS257 is active. *Plasmid* **34**, 198-205.

Nikolich, M.P., Shoemaker, N.B., Wang, G.-R. and Salyers, A.A. (1994). Characterization of a new type of *Bacteroides* transposon, Tc^R Em^R 7853. *J. Bacteriol.* **176**, 6606-6612.

Noguchi, N., Katayama, J. and O'Hara, K. (1996). Cloning and nucleotide sequence of the *mphB* gene for macrolide 2'-phosphotransferase II in *Escherichia coli*. *FEMS Microbiol. Lett.* **144**, 197-202.

Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. and Waese, C.R. (1981). Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.* **9**, 6167-6189.

Oh, T.G., Kwon, A.R. and Choi, E.C. (1998). Induction of *ermAMR* from a clinical strain of *Enterococcus faecalis* by 16-membered-ring macrolide antibiotics. *J. Bacteriol.* **180**, 5788-5791.

Olano, C., Rodriguez, A.M., Méndez, C. and Salas, J.A. (1995). A second ABC transporter is involved in oleandomycin resistance and its secretion by *Streptomyces antibioticus*. *Mol. Microbiol.* **16**, 333-343.

Oliveira, S.S., Murphy, E., Gamon, M.R. and Bastos, M.C. (1993). pRJ5: a naturally occurring *Staphylococcus aureus* plasmid expressing constitutive macrolide-lincosamide-streptogramin B resistance contains a tandem duplication in the leader region of the *ermC* gene. *J. Gen. Microbiol.* **139**, 1461-1467.

O'Neill, M., Dryden, D.T.F. and Murray, N.E. (1998). Localization of a protein-DNA interface by random mutagenesis. *EMBO J.* **17**, 7118-7127.

Ounissi, H. and Courvalin, P. (1985). Nucleotide sequence of the gene *ereA* encoding the erythromycin esterase in *Escherichia coli*. *Gene* **35**, 271-278.

Palombo, E.A., Yusoff, K., Stanisich, V.A., Krishnapillai, V. and Willetts, N.S. (1989). Cloning and genetic analysis of *tra* cistrons of the Tra2/Tra3 region of plasmid RP1. *Plasmid* **22**, 59-69.

Parker, R.H. and Eggleston, M. (1987). Beta-lactamase inhibitors: another approach to overcoming antimicrobial resistance. *Infect. Control* **8**, 36-40.

Perelle, S., Gibert, M., Bourlioux, P., Corthier, G. and Popoff, M.R. (1997). Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect. Immun.* **65**, 1402-1407.

Pernodet, J.L., Blondelet-Rouault, M.H. and Guerineau, M. (1993). Resistance to spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at least two different mechanisms. *J. Gen. Microbiol.* **139**, 1003-1011.

Pernodet, J.L., Gourmelen, A., Blondelet-Rouault, M.H. and Cundliffe, E. (1999). Dispensable ribosomal resistance to spiramycin conferred by *srmA* in the spiramycin producer *Streptomyces ambofaciens*. *Microbiology* **145**, 2355-2364.

Peschke, U., Schmidt, H., Zhang, H.Z. and Piepersberg, W. (1995). Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78-11. *Mol. Microbiol.* **16**, 1137-1156.

Poilane, I., Karjalainen, T., Barc, M.-C., Bourlioux, P. and Collignon, A. (1998). Protease activity of *Clostridium difficile* strains. *Can. J. Microbiol.* **44**, 157-161.

Popoff, M.R., Rubin, E.J., Gill, D.M. and Boquet, P. (1988). Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect. Immun.* **56**, 2299-2306.

Portillo, A., Ruiz-Larrea, F., Zarazaga, M., Alonso, A., Martinez, J.L. and Torres, C. (2000). Macrolide resistance genes in *Enterococcus* spp. *Antimicrob. Agents Chemother.* **44**, 967-971.

Poyart-Salmeron, C., Trieu-Cuot, P., Carlier, C. and Courvalin, P. (1989). Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn*1545*: homologies with other site-specific recombinases. *EMBO J.* **8**, 2425-2433.

Projan, S.J., Monod, M., Narayanan, C.S. and Dubnau, D. (1987). Replication properties of pIM13, a naturally occurring plasmid found in *Bacillus subtilis*, and of its close relative pE5, a plasmid native to *Staphylococcus aureus*. *J. Bacteriol.* **169**, 5131-5139.

Pujol, C., Ehrlich, S. and Janniere, L. (1994). The promiscuous plasmids pIP501 and pAMβ1 from gram-positive bacteria encode complementary resolution functions. *Plasmid* **31**, 100-105.

Rasmussen, J.L., Odelson, D.A. and Macrina, F.L. (1986). Complete nucleotide sequence and transcription of *ermF*, a macrolide-lincosamide-streptogramin B resistance determinant from *Bacteroides fragilis*. *J. Bacteriol.* **168**, 523-533.

Rasmussen, J.L., Odelson, D.A. and Macrina, F.L. (1987). Complete nucleotide sequence of insertion element IS4351 from *Bacteroides fragilis*. *J. Bacteriol.* **169**, 3573-3580.

Recchia, G.D. and Hall, R.M. (1995). Gene cassettes: a new class of mobile element. *Microbiology* **141**, 3015-3027.

Recchia, G.D. and Hall, R.M. (1997). Origins of the mobile gene cassettes found in integrons. *Trends Microbiol.* **5**, 389-394.

Reig, M., Galan, J.-C., Baquero, F. and Perez-Diaz, J.C. (2001). Macrolide resistance in *Peptostreptococcus* spp. mediated by *ermTR*: possible source of macrolide-lincosamide-streptogramin B resistance in *Streptococcus pyogenes*. *Antimicrob. Agents Chemother.* **45**, 630-632.

Rende-Fournier, R., Leclercq, R., Galimand, M., Duval, J. and Courvalin, P. (1993). Identification of the *satA* gene encoding a streptogramin A acetyltransferase in *Enterococcus faecium* BM4145. *Antimicrob. Agents Chemother.* **37**, 2119-2125.

Rice, L.B. (1998). Tn*916* family conjugative transposons and dissemination of antimicrobial resistance determinants. *Antimicrob. Agents Chemother.* **42**, 1871-1877.

Richardson, M.A., Kuhstoss, S., Solenberg, P., Schaus, N.A. and Rao, R.N. (1987). A new shuttle cosmid vector, pKC505, for streptomycetes: its use in the cloning of three different spiramycin-resistance genes from a *Streptomyces ambofaciens* library. *Gene* **61**, 231-241.

Riley, T., Codde, J. and Rouse, I. (1995). Increased length of hospital stay due to *Clostridium difficile* associated diarrhoea. *Lancet* **345**, 455-456.

Roberts, A.N., Hudson, G.S. and Brenner, S. (1985). An erythromycin-resistance gene from an erythromycin-producing strain of *Arthrobacter* sp. *Gene* **35**, 259-270.

Roberts, A.P., Johanesen, P.A., Lyras, D., Mullany, P. and Rood, J.I. (2001). Comparison of Tn*5397* from *Clostridium difficile*, Tn*916* from *Enterococcus faecalis* and the CW459 *tet*(M) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* **147**, 1243-1251.

Roberts, M.C., McFarland, L.V., Mullany, P. and Mulligan, M.E. (1994). Characterization of the genetic basis of antibiotic resistance in *Clostridium difficile*. *J. Antimicrob. Chemother.* **33**, 419-429.

Roberts, M.C., Sutcliffe, J., Courvalin, P., Jensen, L.B., Rood, J.I. and Seppälä, H. (1999). Nomenclature for macrolide and macrolide-lincosamide-streptogramin B antibiotic resistance determinants. *Antimicrob. Agents Chemother.* **43**, 2823-2830.

Rodriguez, A.M., Olano, C., Vilches, C., Méndez, C. and Salas, J.A. (1993). *Streptomyces antibioticus* contains at least three oleandomycin resistance determinants, one of which shows similarity with proteins of the ABC transporter superfamily. *Mol. Microbiol.* **8**, 571-582.

Roe, D.E., Weinberg, A. and Roberts, M.C. (1996). Mobile rRNA methylase genes in *Actinobacillus actinomycetemcomitans*. *J. Antimicrob. Chemother.* **37**, 457-464.

Rood, J.I. (1983). Transferable tetracycline resistance in *Clostridium perfringens* strains of porcine origin. *Can. J. Microbiol.* **29**, 1241-1246.

Rood, J.I. (1997). Genetic analysis in *Clostridium perfringens*. In Rood, J.I., McClane, B., Songer, J. and Titball, R. (eds.), *The Clostridia: molecular biology and pathogenesis.* Academic Press Limited, San Diego, California.

Rood, J.I. (1998). Virulence genes of *Clostridium perfringens*. *Annu. Rev. Microbiol.* **52**, 333-360.

Rood, J.I., Buddle, J.R., Wales, A.J. and Sidhu, R. (1985). The occurrence of antibiotic resistance in *Clostridium perfringens* from pigs. *Aust. Vet. J.* **62**, 276-279.

Rood, J.I. and Cole, S.T. (1991). Molecular genetics and pathogenesis of *Clostridium perfringens*. *Microbiological Reviews* **55**, 621-648.

Rood, J.I., Jefferson, S., Bannam, T.L., Wilkie, J.M., Mullany, P. and Wren, B.W. (1989). Hybridization analysis of three chloramphenicol resistance determinants from *Clostridium perfringens* and *Clostridium difficile*. *Antimicrob. Agents Chemother.* **33**, 1569-1574.

Rood, J.I., Maher, E.A., Somers, E.B., Campos, E. and Duncan, C.L. (1978a). Isolation and characterization of multiply antibiotic-resistant *Clostridium perfringens* strains from porcine feces. *Antimicrob. Agents Chemother.* **13**, 871-880.

Rood, J.I., Scott, V.N. and Duncan, C.L. (1978b). Identification of a transferable resistance plasmid (pCW3) from *Clostridium perfringens*. *Plasmid* **1**, 563-570.

Ross, J.I., Eady, E.A., Cove, J.H., Cunliffe, W.J., Baumberg, S. and Wootton, J.C. (1990). Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super-gene family. *Mol. Microbiol.* 4, 1207-1214.

Ross, J.I., Eady, E.A., Cove, J.H., Jones, C.E., Ratyal, A.H., Miller, Y.W., Vyakrnam, S. and Cunliffe, W.J. (1997). Clinical resistance to erythromycin and clindamycin in cutaneous Propionibacteria isolated from acne patients is associated with mutations in 23S rRNA. *Antimicrob. Agents Chemother.* 41, 1162-1165.

Rosteck Jnr, P.I., Reynolds, P.A. and Hershberger, C.L. (1991). Homology between proteins controlling *Streptomyces fradiae* tylosin resistance and ATP-binding transport. *Gene* 102, 27-32.

Rowe-Magnus, D.A., Guérout, A.-M. and Mazel, D. (1999). Super-integrons. *Res. Microbiol.* 150, 641-651.

Rudy, C.K. and Scott, J.R. (1994). Length of the coupling sequence of Tn*916*. *J. Bacteriol.* 176, 3386-3388.

Rudy, C.K., Taylor, K.L., Hinerfeld, D., Scott, J.R. and Churchward, G. (1997). Excision of a conjugative transposon *in vitro* by the Int and Xis proteins of Tn*916*. *Nucleic Acids Res.* 25, 4061-4066.

Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* 407, 508-513.

Saksena, N.K. and Truffaut, N. (1992). Cloning of tetracycline-resistance genes from various strains of *Clostridium perfringens* and expression in *Escherichia coli*. *Can. J. Microbiol.* 38, 215-221.

Salyers, A. and Shoemaker, N. (1995). Conjugative transposons: the force behind the spread of antibiotic resistance genes among *Bacteroides* clinical isolates. *Anaerobe* 1, 143-150.

Salyers, A., Shoemaker, N., Bonheyo, G. and Frias, J. (1999). Conjugative transposons: transmissible resistance islands. In Kaper, J.B. and Hacker, J. (eds.), *Pathogenicity Islands and Other Mobile Virulence Elements.* American Society for Microbiology, Washington, D.C.

Salyers, A., Shoemaker, N. and Li, L.-Y. (1995a). In the driver's seat: the *Bacteroides* conjugative transposons and the elements they mobilize. *J. Bacteriol.* **177**, 5727-5731.

Salyers, A., Shoemaker, N., Stevens, A.M. and Li, L.-Y. (1995b). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiological Reviews* **59**, 579-590.

Salyers, A.A., Shoemaker, N.B. and Stevens, A.M. (1995c). Tetracycline regulation of conjugal transfer genes. In Hoch, J.A. and Silhavy, T.J. (eds.), *Two-component signal transduction.* American Society for Microbiology, Washington D.C., pp. 393-400.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sasaki, Y., Yamamoto, K., Tamura, Y. and Takahashi, T. (2001). Tetracycline-resistance genes of *Clostridium perfringens*, *Clostridium septicum* and *Clostridium sordellii* isolated from cattle affected with malignant edema. *Vet. Microbiol.* **83**, 61-69.

Schluckebier, G., Labahn, J., Granzin, J. and Saenger, W. (1998). TaqI: possible catalysis via cation-pi interactions in N-specific DNA methyltransferases. *Biol. Chem.* **379**, 389-400.

Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X. (1995). Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16-20.

Schluckebier, G., Zhong, P., Stewart, K.D., Kavanaugh, T.J. and Abad-Zapatero, C. (1999). The 2.2 Å structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: implications for the reaction mechanism. *J. Mol. Biol.* **289**, 277-291.

Schmitz, F.J., Petridou, J., Fluit, A.C., Hadding, U., Peters, G. and von Eiff, C. (2000). Distribution of macrolide-resistance genes in *Staphylococcus aureus* blood-cultures isolated from fifteen German university hospitals. M.A.R.S. study group multicentre study on antibiotic resistance in Staphylococci. *Eur. J. Clin. Microbiol. Infect. Dis.* **19**, 385-387.

Schoner, B., Geistlich, M., Rosteck Jnr, P.I., Rao, R.N., Seno, E., Reynolds, P., Cox, K., Burgett, S. and Hershberger, C. (1992). Sequence similarity between macrolide-resistance determinants and ATP-binding transport proteins. *Gene* **115**, 93-96.

Scott, J.R. (1992). Sex and the single circle: conjugative transposition. *J. Bacteriol.* **174**, 6005-6010.

Scott, J.R., Bringel, F., Marra, D., van Alstine, G. and Rudy, C.K. (1994). Conjugative transposition of Tn*916*: preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. *Mol. Microbiol.* **11**, 1099-1108.

Scott, J.R. and Churchward, G. (1995). Conjugative transposition. *Annu. Rev. Microbiol.* **49**, 367-397.

Scott, J.R., Kirchman, P.A. and Caparon, M.G. (1988). An intermediate in transposition of the conjugative transposon Tn*916*. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4809-4813.

Scott, P.T. and Rood, J.I. (1989). Electroporation-mediated transformation of lysostaphin-treated *Clostridium perfringens*. *Gene* **82**, 327-333.

Sebald, M., Bouanchaud, D. and Bieth, G. (1975). Nature plasmidique de la resistance à plusiers antibiotiques chez *C. perfringens* type A, souche 659. *C. R. Seances Acad. Sci. D.* **280**, 2401-2404.

Seddon, S.V. and Borriello, S.P. (1992). Proteolytic activity of *Clostridium difficile*. *J. Med. Microbiol.* **36**, 307-311.

Senghas, E., Jones, J.M., Yamamoto, M., Gawron-Burke, C. and Clewell, D.B. (1988). Genetic organization of the bacterial conjugative transposon Tn*916*. *J. Bacteriol.* **170**, 245-249.

Soane, A. and García Lobo, J.M. (2000). Identification of a streptogramin A acetyltransferase gene in the chromosome of *Yersinia enterocolitica*. *Antimicrob. Agents Chemother.* **44**, 905-909.

Seppälä, H., Skurnik, M., Soini, H., Roberts, M.C. and Huovinen, P. (1998). A novel erythromycin resistance methylase gene (*ermTR*) in *Streptococcus pyogenes*. *Antimicrob. Agents Chemother.* **42**, 257-262.

Serwold-Davis, T.M. and Groman, N.B. (1988). Identification of a methylase gene for erythromycin resistance with the sequence of a spontaneously deleting fragment of *Corynebacterium diphtheriae* plasmid pNG2. *FEMS Microbiol. Lett.* **46**, 7-14.

Sharp, J. and Poxton, I.R. (1988). The cell wall proteins of *Clostridium difficile*. *FEMS Microbiol. Lett.* **55**, 99-104.

Sharpe, M.E. and Errington, J. (1996). The *Bacillus subtilis soj-spoOJ* locus is required for a centromere-like function involved in prespore chromosome partitioning. *Mol. Microbiol.* **21**, 501-509.

Shaw, J.H. and Clewell, D.B. (1985). Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn*917* in *Streptococcus faecalis*. *J. Bacteriol.* **164**, 782-796.

Shaw, W.V. and Leslie, A.G.W. (1991). Chloramphenicol acetyltransferase. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 363-386.

Shoemaker, N., Wang, G.-R. and Salyers, A. (2000). Multiple gene products and sequences required for excision of the mobilizable integrated *Bacteroides* element NBU1. *J. Bacteriol.* **182**, 928-936.

Shortridge, V.D., Flamm, R.K., Ramer, N., Beyer, J. and Tanaka, S.K. (1996). Novel mechanism of macrolide resistance in *Streptococcus pneumoniae*. *Diagn. Microbiol. Infect. Dis.* **26**, 73-78.

Showsh, S.A. and Andrews Jnr., R.E. (1992). Tetracycline enhances Tn*916*-mediated conjugal transfer. *Plasmid* **28**, 213-224.

Showsh, S.A. and Andrews Jnr., R.E. (1999). Analysis of the requirement for a pUB110 *mob* region during Tn*916*-dependent mobilization. *Plasmid* **41**, 179-186.

Simon, R., Priefer, U. and Puhler, A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Biotechnology* **1**, 37-45.

Skinner, R., Cundliffe, E. and Schmidt, F.J. (1983). Site of action of a ribosomal RNA methylase responsible for resistance to erythromycin and other antibiotics. *J. Biol. Chem.* **258**, 12702-12706.

Sloan, J., McMurray, L.M., Lyras, D., Levy, S.B. and Rood, J.I. (1994). The *Clostridium perfringens* Tet P determinant comprises two overlapping genes: *tetA*(P), which mediates active tetracycline efflux, and *tetB*(P), which is related to the ribosomal protection family of tetracycline-resistance determinants. *Mol. Microbiol.* **11**, 403-415.

Sloan, J., Warner, T.A., Scott, P.T., Bannam, T.L., Berryman, D.I. and Rood, J.I. (1992). Construction of a sequenced *Clostridium perfringens-Escherichia coli* shuttle plasmid. *Plasmid* **27**, 207-219.

Smith, M., Jessee, J., Landers, T. and Jordan, J. (1990). High efficiency bacterial electroporation: 1 x 10^{10} *E. coli* transformants/mg. *Focus* **12**, 38-40.

Smith, C.J. (1987). Nucleotide sequence analysis of Tn*4551*: use of *ermFS* operon fusions to detect promoter activity in *Bacteroides fragilis*. *J. Bacteriol.* **169**, 4589-4596.

Smith, C.J., Markowitz, S.M. and Macrina, F.L. (1981). Transferable tetracycline resistance in *Clostridium difficile*. *Antimicrob. Agents Chemother.* **19**, 997-1003.

Smith, C.J. and Parker, A.C. (1993). Identification of a circular intermediate in the transfer and transposition of Tn*4555*, a mobilizable transposon from *Bacteroides*. *J. Bacteriol.* **175**, 2682-2691.

Smith, C.J., Tribble, G.D. and Bayley, D.P. (1998). Genetic elements of *Bacteroides* species: a moving story. *Plasmid* **40**, 12-29.

Somkuti, G.A., Solaiman, D.K. and Steinberg, D.H. (1997). Molecular properties of the erythromycin resistance plasmid pPV141 from *Staphylococcus chromogenes*. *Plasmid* **37**, 119-127.

Somkuti, G.A., Solaiman, D.K. and Steinberg, D.H. (1998). Molecular characterization of the erythromycin resistance plasmid pPV142 from *Staphylococcus simulans*. *FEMS Microbiol. Lett.* **165**, 281-288.

Songer, J.G. (1996). Clostridial enteric diseases of domestic animals. *Clin. Microbiol. Rev.* **9**, 216-234.

Sor, F. and Fukuhara, H. (1984). Erythromycin and spiramycin resistance mutations of yeast mitochondria: nature of the *rib2* locus in the large ribosomal RNA gene. *Nucleic Acids Res.* **12**, 8313-8318.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Spencer, R. (1998a). Clinical impact and associated costs of *Clostridium difficile*-associated disease. *J. Antimicrob. Chemother.* **41**, 5-12.

Spencer, R. (1998b). The role of antimicrobial agents in the aetiology of *Clostridium difficile*-associated disease. *J. Antimicrob. Chemother.* **41**, 21-27.

Stackebrandt, E. and Rainey, F.A. (1997). Phylogenetic relationships. In Rood, J.I., McClane, B.A., Songer, J.G. and Titball, R.W. (eds.), *The Clostridia: molecular biology and pathogenesis*. Academic Press, San Diego, pp. 3-19.

Steffen, C. and Matzura, H. (1989). Nucleotide sequence analysis and expression studies of a chloramphenicol-acetyltransferase-coding gene from *Clostridium perfringens*. *Gene* **75**, 349-354.

Stevens, A.M., Shoemaker, N.B., Li, L.-Y. and Salyers, A.A. (1993). Tetracycline regulation of genes on *Bacteroides* conjugative transposons. *J. Bacteriol.* **175**, 6134-6141.

Storrs, M.J., Poyart-Salmeron, C., Trieu-Cuot, P. and Courvalin, P. (1991). Conjugative transposition of Tn*916* requires the excisive and integrative activities of the transposon-encoded integrase. *J. Bacteriol.* **173**, 4347-4352.

Stubbs, S., Rupnik, M., Gibert, M., Brazier, J., Duerden, B. and Popoff, M. (2000). Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiol. Lett.* **186**, 307-312.

Su, Y.A. and Clewell, D.B. (1993). Characterization of the left 4 kb of conjugative transposon Tn*916*: determinants involved in excision. *Plasmid* **30**, 234-250.

Sulavik, M.C., Dazer, M. and Miller, P.F. (1997). The *Salmonella typhimurium mar* locus: molecular and genetic analyses and assessment of its role in virulence. *J. Bacteriol.* **179**, 1857-1866.

Sutcliffe, J., Grebe, T., Tait-Kamradt, A. and Wondrack, L. (1996). Detection of erythromycin-resistant determinants by PCR. *Antimicrob. Agents Chemother.* **40**, 2562-2566.

Tait-Kamradt, A., Clancy, J., Cronan, M., Dib-Hajj, F., Wondrack, L., Yuan, W. and Sutcliffe, J. (1997). *mefE* is necessary for the erythromycin-resistant M phenotype in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **41**, 2251-2255.

Tait-Kamradt, A., Davies, T., Cronan, M., Jacobs, M.R., Appelbaum, P.C. and Sutcliffe, J. (2000). Mutations in 23S rRNA and L4 ribosomal protein account for resistance in pneumococcal strains selected in vitro by macrolide passage. *Antimicrob. Agents Chemother.* **44**, 2118-2125.

Takatsugu, H., Boon-Chieng, S. and Mitaku, S. (1998). SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378-379.

Tannock, G.W., Luchansky, J.B., Miller, L., Connell, H., Thode-Andersen, S., Mercer, A.A. and Klaenhammer, T.R. (1994). Molecular characterization of a plasmid-borne (pGT633) erythromycin resistance determinant (*ermGT*) from *Lactobacillus reuteri* 100-63. *Plasmid* **31**, 60-71.

Tasteyre, A., Barc, M.-C., Karjalainen, T., Dodson, P., Hyde, S., Bourlioux, P. and Borriello, S.P. (2000a). A *Clostridium difficile* gene encoding flagellin. *Microbiology* **146**, 957-966.

Tasteyre, A., Karjalainen, T., Avesani, V., Delmeé, M., Collignon, A., Bourlioux, P. and Barc, M.-C. (2000b). Phenotypic and genotypic diversity of the flagellin gene (*fliC*) among *Clostridium difficile* isolates from different serogroups. *J. Clin. Microbiol.* 38, 3179-3186.

Tasteyre, A., Karjalainen, T., Avesani, V., Delmeé, M., Collignon, A., Bourlioux, P. and Barc, M.-C. (2001). Molecular characterization of *fliD* gene encoding flagellar cap and its expression among *Clostridium difficile* isolates from different serogroups. *J. Clin. Microbiol.* **39**, 1178-1183.

Tauch, A., Kassing, F., Kalinowski, J. and Puhler, A. (1995). The *Corynebacterium xerosis* composite transposon Tn*5432* consists of two identical insertion sequences, designated IS*1249*, flanking the erythromycin resistance gene *ermCX*. *Plasmid* **34**, 119-131.

Taylor, K.L. and Churchward, G. (1997). Specific DNA cleavage mediated by the integrase of conjugative transposon Tn*916*. *J. Bacteriol.* **179**, 117-1125.

Tenson, T., DeBlasio, A. and Mankin, A.S. (1996). A functional peptide encoded in the *Escherichia coli* 23S rRNA. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5641-5646.

Thakker-Varia, S., Jenssen, W.D., Moon-McDermott, L., Weinstein, M.P. and Dubin, D.T. (1987). Molecular epidemiology of macrolides-lincosamides-streptogramin B resistance in *Staphylococcus aureus* and coagulase-negative Staphylococci. *Antimicrob. Agents Chemother.* **31**, 735-743.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tribble, G.D., Parker, A.C. and Smith, C.J. (1997). The *Bacteroides* mobilizable transposon Tn*4555* integrates by a site-specific recombination mechanism similar to that of the gram-positive bacterial element Tn*916*. *J. Bacteriol.* **179**, 2731-2739.

Tribble, G.D., Parker, A.C. and Smith, C.J. (1999). Transposition genes of the *Bacteroides* mobilizable transposon Tn*4555*: role of a novel targeting gene. *Mol. Microbiol.* **34**, 385-394.

Trieu-Cuot, P., Poyart-Salmeron, C., Carlier, C. and Courvalin, P. (1990). Nucleotide sequence of the erythromycin resistance gene of the conjugative transposon Tn*1545*. *Nucleic Acids Res.* **18**, 3660.

Vannuffel, P., Di Giambattista, M., Morgan, E.A. and Cocito, C. (1992). Identification of a single base change in ribosomal RNA leading to erythromycin resistance. *J. Biol. Chem.* **267**, 8377-8382.

Vedantam, G., Novicki, T.J. and Hecht, D.W. (1999). *Bacteroides fragilis* transfer factor Tn*5520*: the smallest bacterial mobilizable transposon containing single integrase and mobilization genes that function in *Escherichia coli*. *J. Bacteriol.* **181**, 2564-2571.

Vester, B. and Douthwaite, S. (2001). Macrolide resistance conferred by base substitutions in 23S rRNA. *Antimicrob. Agents Chemother.* **45**, 1-12.

Waligora, A.-J., Barc, M.-C., Bourlioux, P., Collignon, A. and Karjalainen, T. (1999). *Clostridium difficile* cell attachment is modified by environmental factors. *Appl. Environ. Microbiol.* **65**, 4234-4238.

Waligora, A.-J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A. and Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infect. Immun.* **69**, 2144-2153.

Wallace Jnr., R.J., Meier, A., Brown, B.A., Zhang, Y., Sander, P., Onyi, G.O. and Böttger, E.C. (1996). Genetic basis for clarithromycin resistance among isolates of *Mycobacterium chelonae* and *Mycobacterium abscessus*. *Antimicrob. Agents Chemother.* **40**, 1676-1681.

Wang, G. and Taylor, D.E. (1998). Site-specific mutations in the 23S rRNA gene of *Helicobacter pylori* confer two types of resistance to macrolide-lincosamide-streptogramin B antibiotics. *Antimicrob. Agents Chemother.* **42**, 1952-1958.

Wang, H. and Mullany, P. (2000). The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn*5397*. *J. Bacteriol.* **182**, 6577-6583.

Wang, H., Roberts, A.P., Lyras, D., Rood, J.I., Wilks, M. and Mullany, P. (2000a). Characterization of the ends and target sites of the novel conjugative transposon Tn*5397* from *Clostridium difficile*: excision and circularization is mediated by the large resolvase, TndX. *J. Bacteriol.* **182**, 3775-3783.

Wang, H., Roberts, A.P. and Mullany, P. (2000b). DNA sequence of the insertional hot spot of Tn*916* in the *Clostridium difficile* genome and discovery of a Tn*916*-like element in an environmental isolate integrated in the same hot spot. *FEMS Microbiol. Lett.* **192**, 15-20.

Wang, R.F. and Kushner, S.R. (1991). Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* **100**, 195-199.

Wasteson, Y., Roe, D.E., Falk, K. and Roberts, M.C. (1996). Characterization of antibiotic resistance in *Actinobacillus pleuropneumoniae*. *Vet. Microbiol.* **48**, 41-50.

Waters, V.L. (1999). Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance. *Front. Biosci.* **4**, 416-439.

Weisblum, B. (1995). Insights into erythromycin action from studies of its activity as inducer of resistance. *Antimicrob. Agents Chemother.* **39**, 797-805.

Weisblum, B. (1998). Macrolide resistance. *Drug Resistance Updates* **1**, 29-41.

Wells, C.L. and Wilkins, T.D. (1991). Clostridia: spore-forming anaerobic bacilli. In Brown, S. (ed.) *Medical Microbiology*. Churchill Livingstone, New York, pp. 279-296.

Wenzel, C., Moulard, M., Løbner-Olesen, A. and Guschlbauer, W. (1991). Crosslinking of Dam methyltransferase with *S*-adenosyl-methionine. *FEBS Lett.* **280**, 147-151.

Werner, G. and Witte, W. (1999). Characterization of a new enterococcal gene, *satG*, encoding a putative acetyltransferase conferring resistance to streptogramin A compounds. *Antimicrob. Agents Chemother.* **43**, 1813-1814.

Whittle, G., Hund, B.D., Shoemaker, N.B. and Salyers, A.A. (2001). Characterization of the 13-kilobase *ermF* region of the *Bacteroides* conjugative transposon CTnDOT. *Appl. Environ. Microbiol.* **67**, 3488-3495.

Wilcox, M.H. (1998). *Clostridium difficile* - setting the scene. *J. Antimicrob. Chemother.* **41**, 1-3.

Wondrack, L., Massa, M., Yang, B.V. and Sutcliffe, J. (1996). Clinical strain of *Staphylococcus aureus* inactivates and causes efflux of macrolides. *Antimicrob. Agents Chemother.* **40**, 992-998.

Wren, B.W., Mullany, P., Clayton, C. and Tabaqchali, S. (1988). Molecular cloning and genetic analysis of a chloramphenicol acetyltransferase determinant from *Clostridium difficile*. *Antimicrob. Agents Chemother.* **32**, 1213-1217.

Wüst, J. and Hardegger, U. (1983). Transferable resistance to clindamycin, erythromycin, and tetracycline in *Clostridium difficile*. *Antimicrob. Agents Chemother.* **23**, 784-786.

Wüst, J. and Hardegger, U. (1988). Studies on the resistance of *Clostridium difficile* to antimicrobial agents. *Zentralbl Bakteriol Mikrobiol Hyg [A]* **267**, 383-394.

Xiong, L., Shah, S., Mauvais, P. and Mankin, A.S. (1999). A ketolide resistance mutation in domain II of 23S rRNA reveals the proximity of hairpin 35 to the peptidyl transferase centre. *Mol. Microbiol.* **31**, 633-639.

Yamamoto, M., Jones, J.M., Senghas, E., Gawron-Burke, C. and Clewell, D.B. (1987). Generation of Tn*5* insertions in streptococcal conjugative transposon Tn*916*. *Appl. Environ. Microbiol.* **53**, 1069-1072.

Yu, L., Petros, A.M., Schnuchel, A., Zhong, P., Severin , J.M., Walter, K., Holzman, T.F. and Fesik, S.W. (1997). Solution structure of an rRNA methyltransferase (ErmAM) that confers macrolide-lincosamide-streptogramin antibiotic resistance. *Nat. Struct. Biol.* **4**, 483-489.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D. and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883-896.

Zaidenzaig, Y., Fitton, J.E., Packman, L.C. and Shaw, W.V. (1979). Characterization and comparison of chloramphenicol acetyltransferase variants. *Eur. J. Biochem.* **100**, 609-618.

Zalacain, M. and Cundliffe, E. (1991). Cloning of *tlrD*, a fourth resistance gene, from the tylosin producer, *Streptomyces fradiae*. *Gene* **97**, 137-142.

APPENDIX

Publications

EPIDEMICS OF DIARRHEA CAUSED BY A CLINDAMYCIN-RESISTANT STRAIN OF *CLOSTRIDIUM DIFFICILE* IN FOUR HOSPITALS

STUART JOHNSON, M.D., MATTHEW H. SAMORE, M.D., KYLIE A. FARROW, B.SC., GEORGE E. KILLGORE, DR.P.H., FRED C. TENOVER, PH.D., DENA LYRAS, PH.D., JULIAN I. ROOD, PH.D., PAOLA DEGIROLAMI, M.D., ALDONA L. BALTCH, M.D., MARY ELLEN RAFFERTY, R.N., SUZANNE M. PEAR, R.N., AND DALE N. GERDING, M.D.

ABSTRACT

Background Large outbreaks of diarrhea caused by a newly recognized strain of *Clostridium difficile* occurred in four hospitals located in different parts of the United States between 1989 and 1992. Since frequent use of clindamycin was associated with the outbreak in one of the hospitals, we examined the resistance genes of the epidemic-strain isolates and studied the role of clindamycin use in these outbreaks.

Methods Case–control studies were performed at three of the four hospitals to assess the relation of the use of clindamycin to *C. difficile*–associated diarrhea. All isolates of the epidemic strain and representative isolates of other strains identified during each outbreak were tested for susceptibility to clindamycin. Chromosomal DNA from these representative isolates was also analyzed by dot blot hybridization and amplification with the polymerase chain reaction (PCR) with the use of probes and primers from a previously described determinant of erythromycin resistance — the erythromycin ribosomal methylase B (*ermB*) gene — found in *C. perfringens* and *C. difficile*.

Results In a stratified analysis of the case–control studies with pooling of the results according to the Mantel–Haenszel method, we found that the use of clindamycin was significantly increased among patients with diarrhea due to the epidemic strain of *C. difficile*, as compared with patients whose diarrhea was due to nonepidemic strains (pooled odds ratio, 4.35; 95 percent confidence interval, 2.02 to 9.38; $P<0.001$). Exposure to other types of antibiotics or hospitalization in a surgical ward was not significantly associated with the risk of *C. difficile*–associated diarrhea due to the epidemic strain. All epidemic-strain isolates were highly resistant to clindamycin (minimal inhibitory concentration, >256 μg per milliliter). DNA hybridization and PCR analysis showed that all these isolates had an *ermB* gene, which encodes a 23S ribosomal RNA methylase that mediates resistance to macrolide, lincosamide, and streptogramin antibiotics. Only 15 percent of the nonepidemic strains were resistant to clindamycin.

Conclusions A strain of *C. difficile* that is highly resistant to clindamycin was responsible for large outbreaks of diarrhea in four hospitals in different states. The use of clindamycin is a specific risk factor for diarrhea due to this strain. Resistance to clindamycin further increases the risk of *C. difficile*–associated diarrhea, an established complication of antimicrobial use. (N Engl J Med 1999;341:1645-51.)

Reprinted from THE NEW ENGLAND JOURNAL OF MEDICINE (ISSN 0028–4793) Vol. 341:1645-1651 (November 25, 1999).

Printed in the U.S.A. Fax: (781) 893-8103 www.nejm.org

SINCE its etiologic role in pseudomembranous colitis was discovered 21 years ago,[1] *Clostridium difficile* has been recognized as a major nosocomial pathogen throughout the world.[2] A wide variety of strains are capable of causing disease,[3,4] and outbreaks or epidemics of *C. difficile*–associated diarrhea are often linked to a single strain, but the relatedness of these strains among different institutions and geographic regions is not clear. A recent collaborative typing study demonstrated that a newly recognized strain of *C. difficile* was responsible for outbreaks of diarrhea in four hospitals in different parts of the United States that occurred between 1989 and 1992.[5] We evaluated the association of diarrhea from this strain with the use of clindamycin, the resistance of this strain to clindamycin, and the genetic basis for resistance to clindamycin. Three of these outbreaks were reported previously as unrelated events,[6-8] but we now know that the outbreaks were caused by one strain with an apparent propensity to cause epidemics.

METHODS

Outbreaks of Diarrhea Associated with *C. difficile* Infection

The clinical aspects of the previously reported outbreaks in New York, Arizona, and Massachusetts[6-8] and the outbreak in Florida are summarized in Table 1. Criteria for case definitions varied between investigations but were based on clinical symptoms of diarrhea and the detection of *C. difficile* cytotoxin in the stool of affected patients in each instance. In the Arizona outbreak diarrhea was defined as four or more loose or unformed stools in a period of 24 to 36 hours, but it was not defined on the basis of frequency or a specific period in the other outbreaks.

New York

In 1989 there was an abrupt increase in cases of *C. difficile*–associated diarrhea in a 460-bed Veterans Affairs facility in upstate New York.[6] The incidence of *C. difficile*–associated diarrhea

From the Infectious Disease Section, Department of Medicine, Veterans Affairs Chicago Health Care System, Lakeside Division, and Northwestern University Medical School, Chicago (S.J., D.N.G.); the Infectious Disease Section, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston (M.H.S., P.D.); the Department of Microbiology, Monash University, Clayton, Victoria, Australia (K.A.F., D.L., J.I.R.); the Centers for Disease Control and Prevention, Atlanta (G.E.K., F.C.T.); the Stratton Veterans Affairs Medical Center and Albany Medical College, Albany, N.Y. (A.L.B., M.E.R.); and the Veterans Affairs Medical Center, Tucson, Ariz. (S.M.P.). Address reprint requests to Dr. Johnson at the Veterans Affairs Chicago Health Care System, Lakeside Division, Medicine Service, 333 East Huron, Chicago, IL 60611, or at stu-johnson@nwu.edu.

TABLE 1. OUTBREAKS OF DIARRHEA ASSOCIATED WITH *CLOSTRIDIUM DIFFICILE* INFECTION IN FOUR HOSPITALS.

LOCATION	SIZE AND TYPE OF FACILITY*	NO. OF CASES	DATE OF REPORTED OUTBREAK	DURATION OF REPORTED OUTBREAK (MO)	INCIDENCE	COMMENTS
New York[6]	460-Bed VA hospital	174	1989–1990	18	20/1000 admissions	Incidence 10 times as high as in previous 2 years; outbreak continued through 1993
Arizona[7]	300-Bed VA hospital	101	1990–1991	13	15.8/1000 discharges	Incidence 5 times as high as in previous 21 months; outbreak resolved abruptly with restriction of clindamycin use
Florida	786-Bed community hospital	106	1990–1991	2.5	19/1000 discharges	Incidence decreased to 7/1000 discharges 2 months after the end of the outbreak
Massachusetts[8]	431-Bed teaching hospital	98	1992	6	16/1000 discharges	Overall incidence unchanged from previous year, but focal outbreaks occurred on two wards for 2 months

*VA denotes Veterans Affairs.

during this period (20 per 1000 admissions) was 10 times as high as in the previous two years. The reason for the marked increase in the number of cases was not reported, but two case–control studies conducted early in the outbreak (from December 1988 to May 1989) identified antimicrobial therapy, particularly with second- and third-generation cephalosporins, as the chief risk factor. Criteria for the use of antimicrobial therapy were adopted by the hospital, but the epidemic continued at least through the spring of 1993.

Arizona

In July 1990 a Veterans Affairs facility in Arizona noted an abrupt increase in cases of *C. difficile*–associated diarrhea.[7] The incidence of disease during this outbreak (15.8 per 1000 discharges) was five times as high as in the previous 21 months. However, three months after clindamycin was removed from the hospital formulary, the incidence decreased to rates documented before the outbreak.

Florida

A 786-bed community hospital in southwest Florida documented 106 cases of *C. difficile*–associated diarrhea between November 12, 1990, and January 28, 1991. The incidence during the outbreak was 19 per 1000 discharges, and it had decreased to 7 per 1000 discharges by March 1991. At the time, this decrease was attributed to a change in housekeeping procedures.

Massachusetts

A 431-bed tertiary-care teaching hospital in a large city in eastern Massachusetts documented 98 cases of *C. difficile*–associated diarrhea between June and December 1992.[8] The overall incidence during this period (16 per 1000 discharges) was unchanged from the previous year, but focal outbreaks were recognized on two hospital wards over a two-month period. These focal outbreaks resolved without specific intervention.

Identification of Strains Associated with the Outbreaks

Isolates from all four outbreaks were systematically compared by three methods[5]: restriction-endonuclease analysis of whole-cell DNA with the use of *Hin*dIII,[7,8] a polymerase-chain-reaction (PCR) assay with the use of arbitrary primers,[9] and pulsed-field gel electrophoresis with *Sma*I restriction-enzyme analysis.[8] The predominant, epidemic-associated strain at each hospital was either a single strain (on the basis of PCR analysis) or two highly related types (J7 and J9) that were only distinguished by one *Hin*dIII-derived genomic fragment on the basis of restriction-endonuclease analysis.[5] Type J7 isolates were recovered only from the Arizona outbreak. Since this difference between J7 and J9 was most likely the result of a single genetic event,[10] these strains were determined to be part of a single genetic lineage,[5] which we refer to as the epidemic strain.

The epidemic strain accounted for 66 percent of isolates (27 of 41) typed at the New York hospital,[9] 52 percent of isolates (33 of 63) at the Arizona hospital,[7] 33 percent of isolates (6 of 18) at the Florida hospital (unpublished data), and 33 percent of isolates (30 of 90) at the Massachusetts hospital.[8] During the two-month outbreak in the medical and surgical wards at the Massachusetts hospital, the epidemic strain accounted for 62 percent of the isolates (16 of 26).

Patients

Investigators at three of the hospitals reviewed data bases and patients' charts for clindamycin use in the patients with *C. difficile*–associated diarrhea. Data linking patients to the *C. difficile* typing results were not available from the Florida hospital. For patients who had more than one episode of *C. difficile*–associated diarrhea, only the first episode was analyzed. The records of all patients for whom the recovered *C. difficile* isolate was typed were analyzed to determine whether clindamycin had been given at any time during the two months before the illness. Patients were classified as having *C. difficile*–associated diarrhea due to the epidemic strain or due to nonepidemic strains. In the New York outbreak there were 29 episodes of *C. difficile*–associated diarrhea for which antibiotic histories and typing data were available; the epidemic strain was recovered from 20 patients and other strains were recovered from 9. These data were available for 63 episodes of *C. difficile*–associated diarrhea in the Arizona hospital (33 related to the epidemic strain and 30 related to other strains) and for 90 episodes in the Massachusetts hospital (30 related to the epidemic strain and 60 to other strains). Data on exposure to antibiotics other than clindamycin and the type of ward (surgical or other) the patient was in at the time of the episode were available for 160 of the 183 episodes: 28 episodes in New York (21 related to the epidemic strain and 7 related to other strains), 42 episodes in Arizona (25 related to the epidemic strain and 17 to other strains), and 90 episodes in Massachusetts (30 related to the epidemic strain and 60 to other strains). Odds ratios and confidence intervals for three variables (clindamycin use, use of other antibiotics, and hospitalization in a surgical ward) were calculated for individual institu-

tions and combined according to the Mantel–Haenszel method.[11] All P values are two-sided.

Susceptibility Testing of *C. difficile* Isolates

We used the E test (AB Biodisk, Solna, Sweden) to assess all 85 epidemic-strain isolates for susceptibility to clindamycin, including 16 isolates from New York, 33 from Arizona, 6 from Florida, and 30 from Massachusetts. Two representative isolates of the epidemic strain from each of the four outbreaks were identified by restriction-endonuclease analysis and selected for additional testing for susceptibility to erythromycin, ciprofloxacin, ampicillin, and tetracycline and were identified as follows: type J9 (isolates 5602 and 5610) from New York, type J7 (isolates 4224 and 4290) from Arizona, type J9p2 (isolates 5644 and 5650) from Florida, and type J9 (isolates 4478 and 5627) from Massachusetts.

Three toxigenic isolates, identified by restriction-endonuclease analysis, served as controls; two strains were susceptible to clindamycin (type K12p [isolate 5672], an endemic strain from Cook County Hospital, Chicago,[12] and type Y4 [isolate 1323], an endemic strain from the Minneapolis Veterans Affairs Medical Center[13]), and one strain was resistant to clindamycin (type B1 [isolate 832], an epidemic-associated strain from the Minneapolis Veterans Affairs Medical Center[14]).

In addition, representative *C. difficile* isolates of the nonepidemic strains from each of the four outbreaks were also tested for susceptibility to clindamycin. One isolate of each nonepidemic strain was chosen from each outbreak for analysis. In New York, 3 of the 9 types identified on PCR (from 9 nonepidemic cases of *C. difficile*–associated diarrhea) were available for susceptibility testing, whereas 1 isolate of each of the 17 identified by restriction-endonuclease analysis (from 30 nonepidemic cases) was available from Arizona, 1 isolate of 6 of the 7 types identified by PCR (from 12 nonepidemic cases) was available from Florida, and 1 isolate of each of the 20 strains identified by pulsed-field gel electrophoresis (from 60 nonepidemic cases) was available from Massachusetts. In brief, we performed the E test as directed by the manufacturer, using reduced brucella agar plates supplemented with 5 percent defibrinated sheep's blood, 1 mg of vitamin K per liter, and 5 mg of hemin per liter (Remel, Lenexa, Kans.).[15] The isolates were incubated overnight in reduced tryptic soy broth, the amount of the inoculum of *C. difficile* was standardized, and the bacteria were inoculated onto plates and grown to confluency. Antibiotic-impregnated strips were then placed on the inoculated plates, and the plates were incubated anaerobically at 37°C for 48 hours. The minimal inhibitory concentration was measured at the intercept of the inhibition ellipse.

Genetic Analysis of Strains' Resistance to Clindamycin

For the following analyses *C. difficile* strains were grown in brain–heart infusion medium with iron sulfate[16] in an anaerobic glove chamber in an atmosphere of 80 percent nitrogen, 10 percent hydrogen, and 10 percent carbon dioxide at 37°C. When appropriate, the medium was supplemented with erythromycin (50 μg per milliliter).

DNA was extracted from 100-ml broth cultures of *C. difficile* that had been grown to the late log phase. The cells were harvested and lysed according to the sarkosyl lysis procedure,[17] and chromosomal DNA purified by dye buoyant density–gradient ultracentrifugation at 260,000×*g* for 20 hours at 20°C. The chromosomal DNA was extracted from the gradient, dialyzed against TRIS–EDTA buffer (0.01 mM EDTA and 0.1 mM TRIS, pH 7.5), and concentrated by evaporation.

PCR assays were conducted with a GenAmp2400 thermal cycler (Perkin–Elmer Cetus, Norwalk, Conn.) in volumes of 100 μl that contained approximately 100 ng of template DNA. Two oligonucleotide primers specific for the erythromycin ribosomal methylase B (*ermB*) gene of *C. difficile* strain 630[18] — 2980 (5'AATAAGTAAACAGGTAACGTT3') and 2981 (5'GCTCCTTGGAAGCTGTCAGTAG3') — were included at a concentration of approximately 0.7 μM per reaction. The PCR assay consisted of 30 cycles of amplification at 95°C for one minute, two minutes of annealing at 55°C, and three minutes of extension at 72°C. The products were separated by electrophoresis on 0.8 percent agarose gels.

Samples of chromosomal DNA (10 μg) from each strain were blotted onto nylon membranes (Hybond N+, Amersham, Arlington Heights, Ill.), and cross-linked to the membrane by exposure to ultraviolet light for five minutes at 312 nm. A 688-bp *ermB*-specific probe labeled with digoxigenin-11-deoxyuridine triphosphate was prepared by PCR with use of the primers 2980 and 2981 and allowed to hybridize to DNA immobilized on the membrane at 65°C overnight. The membrane was washed twice at room temperature in 2× sodium citrate buffer (SSC) (300 mM sodium chloride and 30 mM sodium citrate), pH 7.5, containing 0.1 percent sodium dodecyl sulfate, and twice at 65°C in 0.2× SSC, containing 0.1 percent sodium dodecyl sulfate. Bound probe was detected with use of an anti-digoxigenin–specific, chemiluminescent substrate (CDP-Star, Roche Diagnostics Australia, Castle Hill, Australia) according to the manufacturer's specifications.

RESULTS

Association of the Epidemic Strain of *C. difficile* with Clindamycin Use

Case–control studies were performed at the New York, Arizona, and Massachusetts hospitals to evaluate the relation between exposure to clindamycin and diarrhea due to the epidemic strain of *C. difficile* (Fig. 1). The frequency of exposure to clindamycin among patients with diarrhea due to nonepidemic strains ranged from 7 percent in Massachusetts to 23 percent in Arizona, which is indicative of variation in the overall frequency of the use of clindamycin among the institutions. Yet, within each institution, clindamycin use was a more frequent cause of diarrhea due to the epidemic strain than of diarrhea due to nonepidemic strains. In the New York hospital, 11 of 20 cases of diarrhea due to the epidemic strain were associated with clindamycin use (55 percent), as compared with 1 of 9 cases of diarrhea due to nonepidemic strains (11 percent); the respective values for the Arizona hospital were 15 of 33 (45 percent) and 7 of 30 (23 percent), and the respective values for the Massachusetts hospital were 9 of 30 (30 percent) and 4 of 60 (7 percent). Overall, 35 of the 83 cases of diarrhea due to the epidemic strain were associated with clindamycin use (42 percent), as compared with 12 of 99 cases due to nonepidemic strains (12 percent, $P<0.001$). The odds ratio for the use of clindamycin ranged from 2.74 to 9.78 (Fig. 1). The pooled odds ratio for the association between clindamycin use and diarrhea due to the epidemic strain was 4.35 (95 percent confidence interval, 2.02 to 9.38; $P<0.001$) (Fig. 1).

In contrast, the use of other antibiotics was not associated with diarrhea due to the epidemic strain. The pooled odds ratio was 1.13 (95 percent confidence interval, 0.53 to 2.41; $P=0.74$) for cefazolin, 1.02 (95 percent confidence interval, 0.45 to 2.32; $P=0.95$) for third-generation cephalosporins (ceftazidime, ceftriaxone, and cefotaxime), 0.43 (95 percent confidence interval, 0.16 to 1.20; $P=0.10$) for ampicillin, 1.09 (95 percent confidence interval, 0.49 to 2.45; $P=0.83$) for vancomycin, and 1.10 (95 percent con-

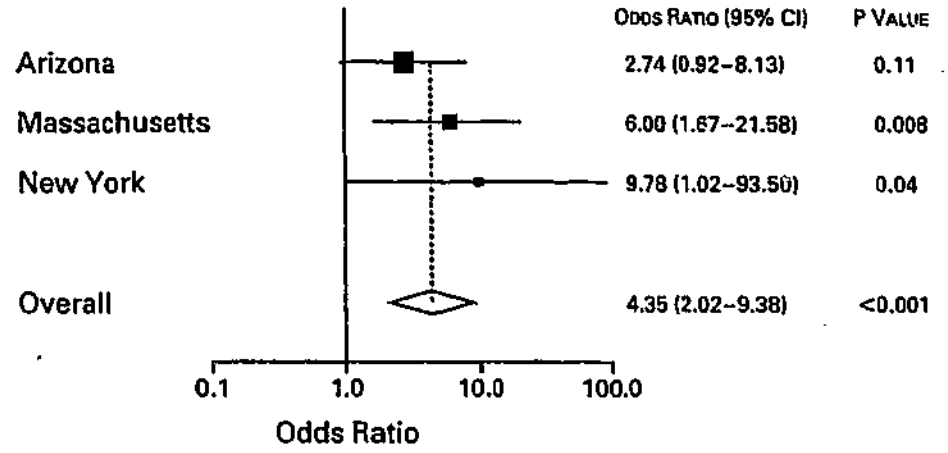


Figure 1. Odds Ratio for the Use of Clindamycin before Becoming Ill among Patients with Diarrhea Due to the Epidemic Strain of *Clostridium difficile* as Compared with Patients with Diarrhea Due to Nonepidemic Strains.

The size of each symbol is proportional to the weight of the corresponding study. The overall odds ratio was obtained with use of the Mantel–Haenszel method; the dotted line indicates the point estimate of this odds ratio. P=0.5 for the test of homogeneity. CI denotes confidence interval.

fidence interval, 0.45 to 2.71; P=0.83) for aminoglycosides (gentamicin and tobramycin). Similarly, hospitalization in a surgical ward was not a risk factor for diarrhea due to the epidemic strain (pooled odds ratio, 0.75; 95 percent confidence interval, 0.38 to 1.49; P=0.42).

Susceptibility of Epidemic and Nonepidemic Strains of *C. difficile* to Clindamycin

All 85 isolates of the epidemic strain of *C. difficile* were highly resistant to clindamycin (minimal inhibitory concentration of clindamycin, >256 μg per milliliter). The representative isolates of epidemic strains (type J9, J7, or J9p2) from each hospital outbreak were also highly resistant to erythromycin, as was the clindamycin-resistant strain (type B1) that was used as a control (Table 2). Both clindamycin-susceptible control strains (types K12p and Y4) were susceptible to clindamycin and erythromycin. The majority of nonepidemic strains from each outbreak were susceptible to clindamycin. High-level resistance to clindamycin (minimal inhibitory concentration, >256 μg per milliliter) was present in only 15 percent of the nonepidemic strains (7 of 46 strains; 1 of 3 in New York, 3 of 17 in Arizona, 0 of 6 in Florida, and 3 of 20 in Massachusetts). The minimal inhibitory concentration of clindamycin for the remaining isolates of nonepidemic strains was 4 μg per milliliter or less in the case of 34 strains and 6 μg per milliliter in the case of the other 5 strains.

Genetic Basis of Clindamycin Resistance in the Epidemic Strain

Resistance to macrolide–lincosamide–streptogramin (MLS) antimicrobial agents such as erythromycin and clindamycin is often mediated by a 23S ribosomal RNA methylase encoded by one of a group of highly related *erm* genes that have been found in gram-positive and gram-negative organisms. Two of these genes, one from *C. perfringens* and one from *C. difficile*, belong to the ErmB–ErmAM hybridization class and have been referred to as the *ermBP* and *ermZ* genes, respectively.[18,19] However, in accordance with a newly proposed nomenclature for the *erm* genes (unpublished data), these genes are both referred to here as *ermB* genes. DNA dot blot hybridizations were carried out on chromosomal DNA prepared from the epidemic strain of *C. difficile* and control strains under highly stringent conditions, with use of an *ermB*-specific probe derived from *C. difficile* strain 630. DNA from all the representative isolates of the clindamycin-resistant epidemic strain at each hospital showed strong hybridization with the probe, indicating that the isolates contained an *ermB* gene (Fig. 2). The control strains K12p and Y4, which are susceptible to MLS antibiotics, did not hybridize to the probe.

Next, we conducted PCR assays with primers 2980 and 2981 and each of the isolates analyzed by dot blot hybridization to confirm that the gene present in the epidemic strains was closely related to that of strain 630. Analysis of each of the MLS-resistant epidemic strains revealed PCR products of the expected size (688 bp), indicating that the gene present in the epidemic strains was an *ermB*-like gene (data not shown). No PCR products were obtained from the MLS-susceptible control strains. In addition, Southern blots carried out on DNA from each of the resistant strains indicated that there may have been more than one copy of the *ermB* gene in each of those strains (data not shown).

DISCUSSION

This study demonstrates that large outbreaks of diarrhea in four hospitals in separate regions of the Unit-

TABLE 2. ANTIMICROBIAL-RESISTANCE PROFILES OF REPRESENTATIVE ISOLATES OF THE EPIDEMIC STRAIN OF *CLOSTRIDIUM DIFFICILE* AT EACH HOSPITAL AND CONTROL STRAINS.*

DRUG	EPIDEMIC STRAIN				CONTROL STRAIN		
	J9 (NEW YORK)	J7 (ARIZONA)	J9p2 (FLORIDA)	J9 (MASSACHUSETTS)	CLINDAMYCIN-SUSCEPTIBLE (K12p)	CLINDAMYCIN-SUSCEPTIBLE (Y4)	CLINDAMYCIN-RESISTANT (B1)
	minimal inhibitory concentration (micrograms per milliliter)						
Clindamycin	>256	>256	>256	>256	0.75	0.75	>256
Erythromycin	>256	>256	>256	>256	0.5	0.50	>256
Ciprofloxacin	>32	>32	>32	>32	>32	>32	>32
Ampicillin	0.75	0.75	0.75	0.75	3.0	1.0	1.5
Tetracycline	0.06	0.06	0.06	0.06	0.09	0.05	12

*Each isolate was identified by restriction-endonuclease analysis; P indicates the presence of plasmids.[5] Groups of closely related *C. difficile* strains are designated by uppercase letters, whereas unique types are designated by numbers. The minimal inhibitory concentrations of each drug are shown for the epidemic strain from each of the four hospitals (two isolates were tested from each outbreak) and for the control strains.

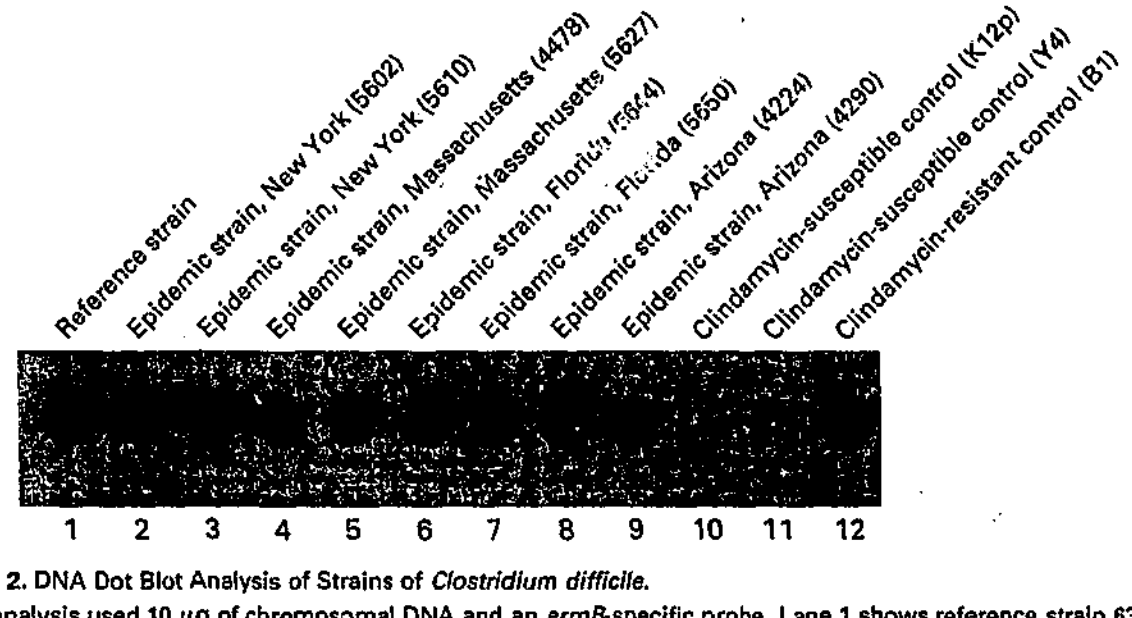


Figure 2. DNA Dot Blot Analysis of Strains of *Clostridium difficile*.

Each analysis used 10 μg of chromosomal DNA and an *ermB*-specific probe. Lane 1 shows reference strain 630, lanes 2 and 3 epidemic-strain isolates from New York, lanes 4 and 5 epidemic-strain isolates from Massachusetts, lanes 6 and 7 epidemic-strain isolates from Florida, lanes 8 and 9 epidemic-strain isolates from Arizona, lanes 10 and 11 clindamycin-susceptible control strains, and lane 12 a clindamycin-resistant control strain.

ed States were all caused by a specific, highly clindamycin-resistant strain of *C. difficile* and that the use of clindamycin was a specific risk factor. The relation between clindamycin use and infection with the epidemic strain was consistent among the institutions, which justifies our pooled analysis and strengthens our findings. Although other virulence factors associated with this particular strain may affect its epidemic potential, resistance of specific *C. difficile* strains to clindamycin may partially explain the well-known propensity of this agent to precipitate outbreaks and epidemics of diarrhea.

These results cast new light on the relation between antibiotic use and *C. difficile*–associated diarrhea. The role of the antimicrobial agent has been assumed to be to disrupt the normal intestinal flora, particularly anaerobes, of the host, which is an important resistance factor with respect to infection with *C. difficile*. The antimicrobial agent precipitating a particular episode of *C. difficile*–associated diarrhea has been thought to have no direct association with the pattern of resistance of the infecting strain.[2] For example, most strains of *C. difficile* are susceptible to ampicillin, yet historically, this agent has com-

monly been implicated in episodes of diarrhea. Although *C. difficile* isolates are routinely resistant to cephalosporins such as cefoxitin, resistance to clindamycin is less common. High-level resistance to clindamycin was present in only 15 percent of the nonepidemic strains in our study. Our results indicate that the relatively high likelihood of *C. difficile*–associated diarrhea after exposure to clindamycin is not just a consequence of effects on the resident flora; it may also be linked to the susceptibility profile of the organism.

Hospital-wide use of clindamycin has been identified as the chief factor responsible for the outbreak of *C. difficile*–associated diarrhea at the Arizona hospital,[7] but this association was not apparent or was not assessed in the initial investigations of the other outbreaks.[6,8] The outbreak in Arizona was abruptly terminated by the removal of clindamycin from the hospital formulary.[7] The use of cephalosporins was identified as the chief risk factor for *C. difficile*–associated diarrhea early in the New York outbreak, on the basis of multivariate analyses of two case–control studies in which ward controls and diarrhea controls, respectively, were used.[6] The use of clindamycin, however, was also a risk factor in the ward study and showed a trend in the diarrhea study.[6] Stool culture, with typing of the recovered *C. difficile* isolates, was not performed in New York until one year after the original case–control studies.[9] We used the later cohort of cases (identified between March and October 1990) and found that clindamycin use was a specific risk factor for diarrhea due to the epidemic strain of *C. difficile*. A subsequent comparative typing study of *C. difficile* isolates from this same hospital documented persistence of the epidemic strain two years later (January to November 1992).[20] The risk of *C. difficile*–associated diarrhea associated with the use of specific antibiotics had not been reported previously for the outbreak at the Massachusetts hospital,[8] and data on antibiotic use were not available from the Florida outbreak.

There is evidence that this strain or genetically related strains may have a [illegible]en broader geographic distribution than is suggested by the distribution of these four outbreaks. Preliminary results obtained with use of PCR ribotyping indicate that the epidemic strain from the Massachusetts hospital designated as type J9 on the basis of restriction-endonuclease analysis or type D1 on the basis of pulsed-field gel electrophoresis is PCR ribotype 1.[21] PCR ribotype 1 was the most common strain among hospitalized patients in England and Wales, accounting for 7 percent of all isolates in one survey,[21] and was responsible for a large outbreak in northwest England involving 175 patients and 17 deaths at one hospital.[22] A formal comparison of restriction-endonuclease analysis and PCR ribotyping methods should clarify whether these European epidemic strains are related to the epidemic strains we studied. We have also used restriction-endonuclease analysis to analyze two *C. difficile* isolates of the clonal strain associated with another clindamycin-related epidemic of diarrhea that was recently reported in Virginia.[23] Neither of these isolates (kindly provided by Michael Climo and Edward Wong) was type J9 or J7.

Each of the epidemic-strain isolates contained an *erm* gene[19] which, on the basis of its ability to hybridize under highly stringent conditions with an *ermB*-specific probe, belongs to the ErmB class of erythromycin-resistance determinants. This conclusion was supported by PCR analysis, which showed that a product of the same size as the *ermB* determinant from *C. difficile* strain 630 was amplified from each of the epidemic isolates with use of *ermB*-specific primers. These data provide evidence that the resistance to MLS antibiotics of each of the epidemic-strain isolates results from the presence of an *ermB* gene.

MLS-resistance genes from the ErmB hybridization class have been detected in both *C. perfringens* and *C. difficile*.[24-26] The *ermB* gene from *C. perfringens* is located on a large nonconjugative but mobilizable plasmid, pIP402.[27] By contrast, the *ermB* gene from *C. difficile* strain 630, which is 99 percent homologous to the *C. perfringens* gene (unpublished data), appears to be located on the chromosome.

In strain 630, transfer of erythromycin resistance occurs in the absence of detectable plasmids. The *ermB* gene has been postulated to reside on the as yet uncharacterized conjugative transposon Tn*5398*.[28] It is possible that the *ermB* gene detected in the epidemic strain that we studied is also associated with Tn*5398*, or with a related mobile genetic element located on the chromosome. Such elements are likely to represent an important method for the dissemination of resistance to MLS antibiotics among clinical isolates of *C. difficile*, especially in hospitals. It is also possible that the *ermB* gene in the epidemic strain is located on the chromosome but is not associated with a transposable element or that *ermB* is located on a plasmid. However, no antibiotic-resistance plasmids have ever been reported in *C. difficile*.

Taken together, these observations suggest that a single erythromycin–clindamycin resistance gene, present in specific strains of *C. difficile*, is associated with a significantly increased risk of *C. difficile*–associated diarrhea in widely dispersed U.S. hospitals, especially in association with clindamycin use. *C. difficile*–associated diarrhea is virtually unknown in the absence of use of antimicrobial agents, and the risk of this illness among hospitalized patients increases with the use of clindamycin and the presence of clindamycin-resistant strains of *C. difficile*. *C. difficile*–associated diarrhea is yet another example of the increasing number of nosocomial infections caused by organisms resistant to antimicrobial agents. It is encouraging that in two well-described outbreaks

caused by clindamycin-resistant *C. difficile*, there was rapid resolution of the epidemic with restriction of the use of clindamycin.[7,23]

Supported by a grant from the Department of Veterans Affairs Research Service (to Drs. Johnson and Gerding) and by a grant from the Australian National Health and Medical Research Council (to Dr. Rood).

We are indebted to Lata Venkataraman, John Galgiani, Tom Minnick, Melanie Hall, Jeffery Stahl, Janet K. Shim, and Susan P. Sambol for their contributions to this investigation.

REFERENCES

1. Bartlett JG, Chang TW, Gurwith M, Gorbach SL, Onderdonk AB. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. N Engl J Med 1978;298:531-4.
2. Gerding DN, Johnson S, Peterson LR, Mulligan ME, Silva J Jr. *Clostridium difficile*-associated diarrhea and colitis. Infect Control Hosp Epidemiol 1995;16:459-77.
3. Clabots CR, Johnson S, Olson MM, Peterson LR, Gerding DN. Acquisition of *Clostridium difficile* by hospitalized patients: evidence for colonized new admissions as a source of infection. J Infect Dis 1992;166:561-7.
4. Samore MH, Bettin KM, DeGirolami PC, Clabots CR, Gerding DN, Karchmer AW. Wide diversity of *Clostridium difficile* types at a tertiary referral hospital. J Infect Dis 1994;170:615-21.
5. Samore MH, Killgore G, Johnson S, et al. Multicenter typing comparison of sporadic and outbreak *Clostridium difficile* isolates from geographically diverse hospitals. J Infect Dis 1997;176:1233-8.
6. Nelson DE, Auerbach SB, Baltch AL, et al. Epidemic *Clostridium difficile*-associated diarrhea: role of second- and third-generation cephalosporins. Infect Control Hosp Epidemiol 1994;15:88-94. [Erratum, Infect Control Hosp Epidemiol 1994;15:366.]
7. Pear SM, Williamson TH, Bettin KM, Gerding DN, Galgiani JN. Decrease in nosocomial *Clostridium difficile*-associated diarrhea by restricting clindamycin use. Ann Intern Med 1994;120:272-7.
8. Samore MH, Venkataraman L, DeGirolami PC, Arbeit RD, Karchmer AW. Clinical and molecular epidemiology of sporadic and clustered cases of nosocomial *Clostridium difficile* diarrhea. Am J Med 1996;100:32-40.
9. Killgore GE, Kato H. Use of arbitrary primer PCR to type *Clostridium difficile* and comparison of results with those by immunoblot typing. J Clin Microbiol 1994;32:1591-3.
10. Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Infect Control Hosp Epidemiol 1997;18:426-39.
11. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, 1998:270-2.
12. Mekonen E, Johnson S, Pottinger J, Kocka F, Pulvirenti J, Gerding D. Predominance of a single restriction endonuclease analysis (REA) group of *Clostridium difficile* (CD) at two Chicago hospitals. Infect Control Hosp Epidemiol 1996;17:Suppl:P32. abstract.
13. Clabots CR, Pearson AC, Bettin K, et al. Typing of *Clostridium difficile* (CD) strains responsible for epidemics in a hospital with a high endemic incidence of CD-associated diarrhea. In: Program and abstracts of the 31st Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago, September 29–October 2, 1991. Washington, D.C.: American Society for Microbiology, 1991:352. abstract.
14. Johnson S, Clabots CR, Linn FV, Olson MM, Peterson LR, Gerding DN. Nosocomial *Clostridium difficile* colonisation and disease. Lancet 1990;336:97-100.
15. Methods for antimicrobial susceptibility testing of anaerobic bacteria: approved standard. 4th ed. NCCLS document M11-A4. Wayne, Pa.: National Committee for Clinical Laboratory Standards, 1997.
16. Smith CJ, Markowitz SM, Macrina FL. Transferable tetracycline resistance in *Clostridium difficile*. Antimicrob Agents Chemother 1981;19:997-1003.
17. Aim LJ, Rood JI. Molecular analysis of transferable tetracycline resistance plasmids from *Clostridium perfringens*. J Bacteriol 1985;161:636-40.
18. Wüst J, Hardegger U. Transferable resistance to clindamycin, erythromycin, and tetracycline in *Clostridium difficile*. Antimicrob Agents Chemother 1983;23:784-6.
19. Arthur M, Brisson-Noël A, Courvalin P. Origin and evolution of genes specifying resistance to macrolide, lincosamide and streptogramin antibiotics: data and hypotheses. J Antimicrob Chemother 1987;20:783-802.
20. Rafferty ME, Baltch AL, Smith RP, et al. Comparison of restriction enzyme analysis, arbitrarily primed PCR, and protein profile analysis for epidemiologic investigation of an ongoing *Clostridium difficile* outbreak. J Clin Microbiol 1998;36:2957-63.
21. Brazier JS. The epidemiology and typing of *Clostridium difficile*. J Antimicrob Chemother 1998;41:Suppl C:47-57.
22. Cartmill TDI, Panigrahi H, Worsley MA, McCann DC, Nice CN, Keith E. Management and control of a large outbreak of diarrhoea due to *Clostridium difficile*. J Hosp Infect 1994;27:1-15.
23. Climo MW, Israel DS, Wong ES, Williams D, Coudron P, Markowitz SM. Hospital-wide restriction of clindamycin: effect on the incidence of *Clostridium difficile*-associated diarrhea and cost. Ann Intern Med 1998;128:989-95.
24. Berryman DI, Lyristis M, Rood JI. Cloning and sequence analysis of *ermQ*, the predominant macrolide-lincosamide-streptogramin B resistance gene in *Clostridium perfringens*. Antimicrob Agents Chemother 1994;38:1041-6.
25. Berryman DI, Rood JI. The closely related *ermB-ermAM* genes from *Clostridium perfringens, Enterococcus faecalis* (pAM beta1), and *Streptococcus agalactiae* (pIP501) are flanked by variants of a directly repeated sequence. Antimicrob Agents Chemother 1995;39:1830-4.
26. Lyras D, Rood JI. Transposable genetic elements and antibiotic resistance determinants from *Clostridium perfringens* and *Clostridium difficile*. In: Rood JI, McClane BA, Songer JG, Titball RW, eds. The clostridia: molecular biology and pathogenesis. San Diego: Academic Press, 1997:73-92.
27. Brefort G, Magot M, Ionesco H, Sebald M. Characterization and transferability of *Clostridium perfringens* plasmids. Plasmid 1977;1:52-66.
28. Mullany P, Wilks M, Tabaqchali S. Transfer of macrolide-lincosamide-streptogramin B (MLS) resistance in *Clostridium difficile* is linked to a gene homologous with toxin A and is mediated by a conjugative transposon, Tn*5398*. J Antimicrob Chemother 1995;35:305-15.

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Feb. 2000, p. 411–413 Vol. 44, No. 2
0066-4804/00/$04.00+0

The Macrolide-Lincosamide-Streptogramin B Resistance Determinant from *Clostridium difficile* 630 Contains Two *erm*(B) Genes

KYLIE A. FARROW, DENA LYRAS, AND JULIAN I. ROOD*

Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia

Received 26 April 1999/Returned for modification 4 May 1999/Accepted 29 October 1999

The ErmB macrolide-lincosamide-streptogramin B (MLS) resistance determinant from *Clostridium difficile* 630 contains two copies of an *erm*(B) gene, separated by a 1.34-kb direct repeat also found in an Erm(B) determinant from *Clostridium perfringens*. In addition, both *erm*(B) genes are flanked by variants of the direct repeat sequence. This genetic arrangement is novel for an ErmB MLS resistance determinant.

Clostridium difficile is the causative agent of antibiotic-associated diarrhea and pseudomembranous colitis, diseases generally associated with exposure to antibiotics. The antibiotics most commonly involved include clindamycin, cephalosporins, and ampicillin (2); however, virtually all antibacterial agents have been implicated.

Erythromycin is a member of the macrolide-lincosamide-streptogramin B (MLS) group of protein synthesis inhibitors (11, 16). In many bacterial species (4, 6, 11, 12), MLS resistance is mediated by *erm* genes, which encode 23S RNA methylases. Numerous *erm* genes have been characterized and divided into distinct classes based on their sequence similarity (19). The most widely distributed of these classes of Erm determinants is the Erm B/AM class, which has recently been renamed as the ErmB class (19), the *erm* genes belonging to this class now being referred to as *erm*(B) genes (19).

ErmB determinants have been detected in both *Clostridium perfringens* (3) and *C. difficile* (9, 21). The *C. perfringens* determinant is located on a large mobilizable plasmid, pIP402, and consists of an *erm*(B) gene (previously *ermBP*) flanked by 1.34-kb direct repeat (DR) sequences (4) (Fig. 1). Each DR contains an open reading frame (ORF), ORF298, the putative product of which has similarity to ParA and Soj proteins, which are involved in plasmid and chromosomal partitioning (8, 23). ORF298 is flanked by the highly palindromic repeated sequences of *palA* and *palB* (4).

Hybridization analysis of erythromycin-resistant *C. difficile* strains has also revealed the presence of *erm*(B) genes (previously *ermZ* or *ermBZ*) (9, 21). The objective of our studies was to examine the genetic organization of the ErmB determinant from *C. difficile* 630. This strain (28) was grown at 37°C in an anaerobic glove chamber (Coy Laboratories; 80% N_2, 10% H_2, 10% CO_2) in BHIS medium (25) supplemented with erythromycin (50 μg/ml) or rifampin (20 μg/ml). *C. perfringens* CP592 (5) was grown anaerobically on nutrient agar (20) containing erythromycin (50 μg/ml). Recombinant strains were derivatives of *Escherichia coli* DH5α (Bethesda Research Laboratories, Inc.) and were grown in 2YT medium (17) containing erythromycin (150 μg/ml).

Cloning experiments (22) utilized the low-copy-number *E. coli* plasmid vector pWSK29 (27). Small-scale plasmid DNA isolation was performed by a modified mini alkaline-lysis–polyethylene glycol precipitation procedure (Applied Biosystems). DNA sequencing was carried out with the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit on an Applied Biosystems 373 DNA sequencer. DNA was prepared from both *C. difficile* and *C. perfringens* by dye buoyant density gradient ultracentrifugation at 260,000 × *g* for 20 h at 20°C (1).

To determine the size of the *C. difficile* 630 fragment carrying the *erm*(B) gene, chromosomal DNA samples (10 μg) were digested with *Sau*3A or *Hin*dIII and separated by electrophoresis on 0.8% agarose. Southern blots (26) were probed at high stringency with a 688-bp *erm*(B)-specific digoxigenin-labelled probe prepared by PCR with the primers 2980 (5′AAT AAGTAAACAGGTAACGTT 3′) and 2981 (5′GCTCCTTG GAAGCTGTCAGTAG 3′). A single hybridizing 9.7-kb *Hin*dIII band was observed (Fig. 2) after washing at high stringency and probe detection with CDP-Star (Boehringer-Mannheim). However, after *Sau*3A digestion, two hybridizing bands of 2.0 and 2.3 kb were evident (Fig. 2). In contrast, with DNA from *C. perfringens* CP592, only single hybridizing bands were detected with each enzyme (Fig. 2). The presence of two hybridizing *Sau*3A bands in strain 630 DNA suggested that either there were two *erm*(B) genes separated by less than 9.7 kb, or there was a single *erm*(B) gene which contained an internal *Sau*3A site that was not present in the *erm*(B) gene from *C. perfringens*.

The 9.7-kb *Hin*dIII fragment from strain 630 was cloned into pWSK29, and the *erm*(B) gene region of the recombinant plasmid, pJIR1594, was completely sequenced on both strands across all restriction sites. Sequence analysis revealed that this ErmB determinant had a novel genetic organization. Two identical copies of the *erm*(B) gene were present, which we have designated as *erm1*(B) and *erm2*(B) (Fig. 1). The genes had 99% sequence identity to the *erm*(B) gene from *C. perfringens* and greater than 97% sequence identity to all other *erm*(B) genes. In addition, the two genes were separated by a single complete copy of the DR sequence that is found on either side of the *C. perfringens erm*(B) gene and in association with most of the other *erm*(B) genes (Fig. 1).

Upstream of *erm2*(B) was an apparent deletion that removed the *erm*(B) promoter. It is therefore unlikely that the *erm2*(B) gene is expressed; however, expression from an upstream promoter such as the *erm1*(B), ORF3, or ORF298 promoters cannot be ignored.

Upstream of *erm1*(B) was a potential *erm* leader peptide sequence, a potential promoter, and 75 bp of the DR sequence

* Corresponding author. Mailing address: Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia. Phone: 61 [redacted]

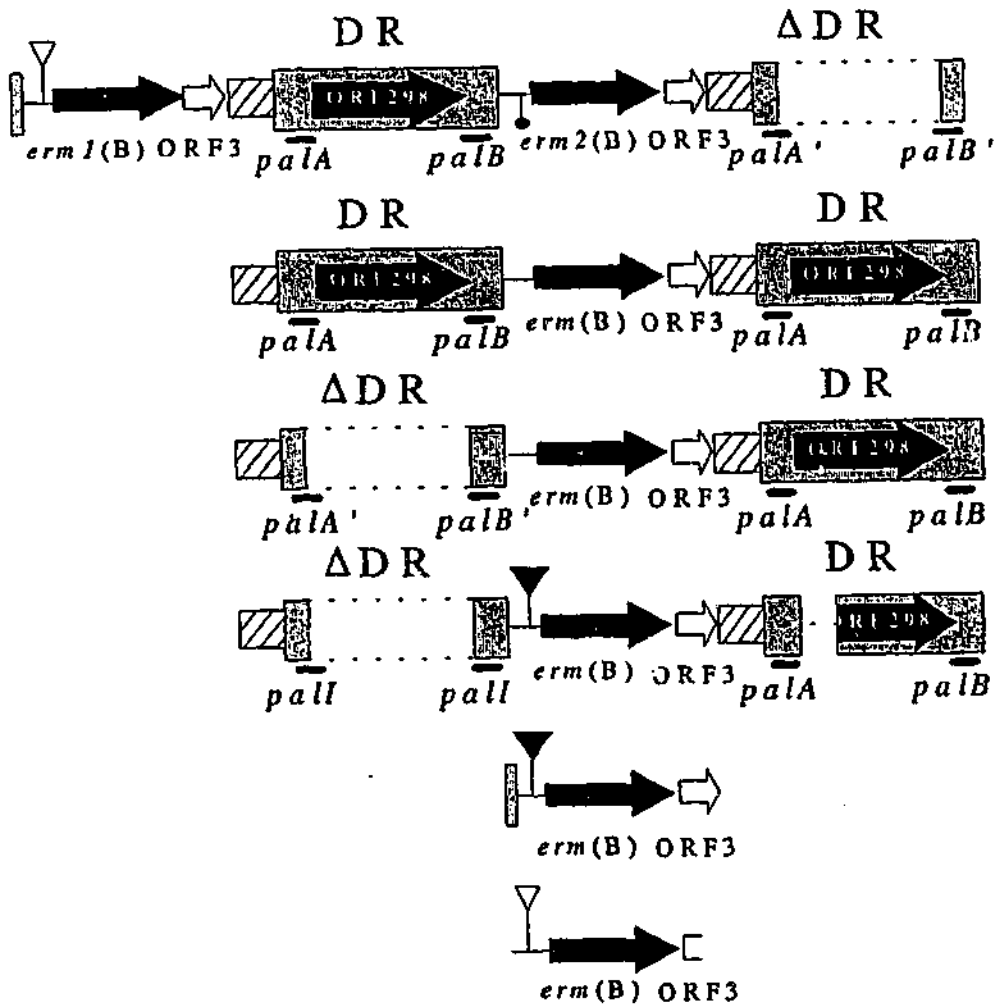


FIG. 1. Comparative genetic organization of the Erm(B) determinants from *C. difficile*, *C. perfringens* (4), *Enterococcus faecalis* (pAMβ1) (14), *Streptococcus agalactiae* (pIP501) (18), *E. faecalis* (Tn*917*) (24), and *Escherichia coli* (6). The approximate extent and organization of the determinants are shown schematically and are not necessarily to scale. Regions of nucleotide sequence similarity are indicated by the same shading. The solid arrows indicate the individual ORFs and their respective direction of transcription. The approximate location of the palindromic sequences (*palA* and *palI*) is indicated by the boldface lines below the shaded boxes. The *palA'*, *palB'*, and *palI* sequences represent the portions of the *C. perfringens erm*(B)-derived *palA* and *palB* homologues that are present at the ends of the deletion in these variants of the DR sequence. Functional and nonfunctional leader peptide sequences are indicated by solid and open triangles respectively. The promoter deletion upstream of the *C. difficile erm2*(B) gene is indicated by the solid oval. The region of pIP501 for which no sequence data are available is indicated by a single broken line. This comparison varies slightly from the previously published figure (Fig. 2 in reference 4).

(Fig. 1). Leader peptide sequences are commonly found upstream of inducible *erm* genes (7). The leader peptide gene region contains a number of inverted repeats and leads to the regulation of *erm* expression by translational attenuation (15). Based on the similarity of the upstream region of other *erm* genes to the leader peptide sequence upstream of *erm*(C), several other *erm* genes, including some *erm*(B) genes, have been proposed to be regulated by translational attenuation. Examination of constitutive *erm* genes showed that the leader peptide sequence was either absent or was mutated and nonfunctional (10, 15). Analysis of the leader peptide sequence upstream of *erm1*(B) indicated that it was similar to nonfunctional leader peptides. Therefore, induction experiments were carried out to determine whether erythromycin resistance was constitutively or inducibly expressed in strain 630. The results showed that when the cells were subcultured from medium that did not contain erythromycin, the same growth rate was observed in the presence or absence of erythromycin (data not shown), suggesting that the *erm1*(B) leader peptide is not functional in strain 630 and that erythromycin resistance is constitutively expressed.

Downstream of both *erm1*(B) and *erm2*(B) was ORF3, which is found in the same position in virtually all ErmB determinants (13) (Fig. 1). Further downstream of *erm2*(B)

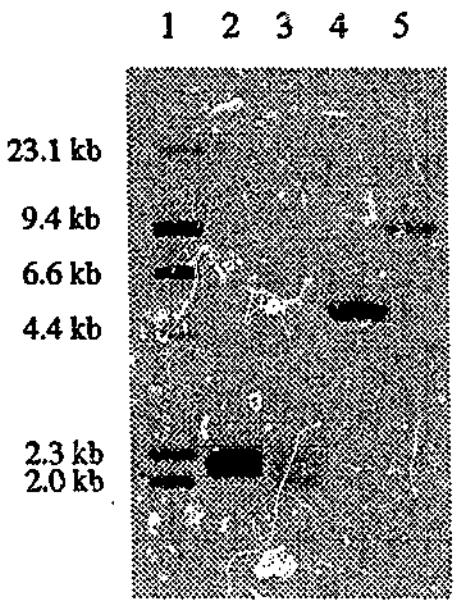


FIG. 2. Southern hybridization analysis. Analysis of *C. difficile* 630 (lanes 3 and 5) and *C. perfringens* CP592 (lanes 2 and 4) DNA with an *erm*(B)-specific probe. DNA was digested with either *Sau*3A (lanes 2 and 3) or *Hin*dIII (lanes 4 and 5). Digoxigenin-labelled λ*cI857Hin*dIII standards are shown (lane 1).

was a variant of the DR sequence. This variant contained a deletion that had removed ORF298 and another deletion that appeared to have removed the last 75 bp of the DR sequence. This region was identical to the 75 bp of the DR found upstream of *erm1*(B), suggesting that other recombination events may also have occurred.

Comparative analysis (Fig. 1) of the various ErmB determinants revealed that the *C. difficile* 630 determinant is the only member of this class which has two *erm* structural genes. In addition, almost all of the *erm*(B) genes are flanked by complete or deleted (ΔDR) variants of the DR sequence. None of these variants are identical, each deletion apparently having occurred at a slightly different location. Therefore, it is likely that homologous recombination events involving the *palA* and *palB* sequences are responsible for the deletions, rather than site-specific recombination events.

The only *erm*(B) gene that is flanked by two complete copies of the DR sequence is from *C. perfringens*. We previously postulated that this determinant represents the ErmB progenitor and that the other determinants have arisen through homologous recombination events which remove part of the DR sequences (4). We propose that the evolution of the *C. difficile* determinant may have involved a duplication of the putative progenitor determinant with subsequent recombination events, which resulted in two *erm*(B) genes separated by a complete copy of the DR sequence.

Nucleotide sequence accession number. The GenBank accession number of the DNA sequence of the *C. difficile* ErmB determinant is AF109075.

We thank the Australian National Health and Medical Research Council for research support.

K.A.F. was the recipient of an Australian Postgraduate Award.

REFERENCES

1. **Abraham, L. J., and J. I. Rood.** 1985. Molecular analysis of transferable tetracycline resistance plasmids from *Clostridium perfringens*. J. Bacteriol. 161:636–640.
2. **Bartlett, J. G.** 1981. Antimicrobial agents implicated in *Clostridium difficile* toxin associated diarrhea or colitis. Johns Hopkins Med. J. 149:6–9.
3. **Berryman, D. I., and J. I. Rood.** 1989. Cloning and hybridization analysis of *ermI*, a macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium perfringens*. Antimicrob. Agents Chemother. 33:1346–1353.
4. **Berryman, D. I., and J. I. Rood.** 1995. The closely related *ermB-ermAM* genes from *Clostridium perfringens*, *Enterococcus faecalis* (pAMβ1), and *Streptococcus agalactiae* (pIP501) are flanked by variants of a directly repeated sequence. Antimicrob. Agents Chemother. 39:1830–1834.
5. **Brefort, G., M. Magot, H. Ionesco, and M. Sebald.** 1977. Characterization and transferability of *Clostridium perfringens* plasmids. Plasmid 1:52–56.
6. **Brisson-Noël, A., M. Arthur, and P. Courvalin.** 1988. Evidence for natural gene transfer from gram-positive cocci to *Escherichia coli*. J. Bacteriol. 170: 1739–1745.
7. **Dubnau, D., and M. Moued.** 1986. The regulation and evolution of MLS resistance. Banbury Rep. 24:369–387.
8. **Easter, C. L., H. Schwab, and D. R. Helinski.** 1998. Role of the *parCBA* operon of the broad-host-range plasmid RK2 in stable plasmid maintenance. J. Bacteriol. 180:6023–6030.
9. **Hächler, H., B. Berger-Bächi, and F. H. Kayser.** 1987. Genetic characterization of a *Clostridium difficile* erythromycin-clindamycin resistance determinant that is transferable to *Staphylococcus aureus*. Antimicrob. Agents Chemother. 31:1039–1045.
10. **Kamimiya, S., and B. Weisblum.** 1988. Translational attenuation control of *ermSF*, an inducible resistance determinant encoding rRNA *N*-methyltransferase from *Streptomyces fradiae*. J. Bacteriol. 170:1800–1811.
11. **LeClerq, R., and P. Courvalin.** 1991. Bacterial resistance to macrolide, lincosamide, and streptogramin antibiotics by target modification. Antimicrob. Agents Chemother. 35:1267–1272.
12. **LeClerq, R., and P. Courvalin.** 1991. Intrinsic and unusual resistance to macrolide, lincosamide, and streptogramin antibiotics in bacteria. Antimicrob. Agents Chemother. 35:1273–1276.
13. **Lyras, D., and J. I. Rood.** 1997. Transposable genetic elements and antibiotic resistance determinants from *Clostridium perfringens* and *Clostridium difficile*, p. 73–92. *In* J. Rood, B. McClane, J. Songer, and R. Titball (ed.), The clostridia: molecular biology and pathogenesis. Academic Press, San Diego, Calif.
14. **Martin, B., G. Alloing, V. Mejean, and J.-P. Claverys.** 1987. Constitutive expression of erythromycin resistance mediated by the ermAM determinant of plasmid pAM-beta-1 results from deletion of 5′ leader peptide sequences. Plasmid 18:250–253.
15. **Mayford, M., and B. Weisblum.** 1990. The *ermC* leader peptide: amino acid alterations leading to differential efficiency of induction by macrolide-lincosamide-streptogramin B antibiotics. J. Bacteriol. 172:3772–3779.
16. **Menninger, J. R.** 1985. Functional consequences of binding macrolides to ribosomes. J. Mol. Biol. 16:23–24.
17. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Pujol, C., S. Ehrlich, and L. Janniere.** 1994. The promiscuous plasmids pIP501 and pAMβ1 from Gram-positive bacteria encode complementary resolution functions. Plasmid 31:100–105.
19. **Roberts, M. C., J. Sutcliffe, P. Courvalin, L. B. Jensen, J. I. Rood, and H. Seppala.** 1999. Nomenclature for macrolide and macrolide-lincosamide Streptogramin B antibiotic resistance determinants. Antimicrob. Agents Chemother. 43:2823–2830.
20. **Rood, J. I.** 1983. Transferable tetracycline resistance in *Clostridium perfringens* strains of porcine origin. Can. J. Microbiol. 29:1241–1246.
21. **Rood, J. I., and S. T. Cole.** 1991. Molecular genetics and pathogenesis of *Clostridium perfringens*. Microbiol. Rev. 55:621–648.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
23. **Sharpe, M. E., and J. Errington.** 1996. The *Bacillus subtilis soj-spoOJ* locus is required for a centromere-like function involved in prespore chromosome partitioning. Mol. Microbiol. 21:501–509.
24. **Shaw, J. H., and D. B. Clewell.** 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn*917* in *Streptococcus faecalis*. J. Bacteriol. 164:782–796.
25. **Smith, C. J., S. M. Markowitz, and F. L. Macrina.** 1981. Transferable tetracycline resistance in *Clostridium difficile*. Antimicrob. Agents Chemother. 19:997–1003.
26. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
27. **Wang, R. F., and S. R. Kushner.** 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene 100:195–199.
28. **Wüst, J., and U. Hardegger.** 1983. Transferable resistance to clindamycin, erythromycin, and tetracycline in *Clostridium difficile*. Antimicrob. Agents Chemother. 23:784–786.

Microbiology (2001), 147, 2717–2728 Printed in Great Britain

Genomic analysis of the erythromycin resistance element Tn*5398* from *Clostridium difficile*

Kylie A. Farrow, Dena Lyras and Julian I. Rood

Author for correspondence: Julian I. Rood. [redacted]

Bacterial Pathogenesis Research Group, Department of Microbiology, PO Box 53, Monash University, Victoria 3800, Australia

Clostridium difficile is a nosocomial pathogen that causes a range of chronic intestinal diseases, usually as a result of antimicrobial therapy. Macrolide-lincosamide-streptogramin B (MLS) resistance in _C. difficile_ is encoded by the Erm B resistance determinant, which is thought to be located on a conjugative transposon, Tn*5398*. The 9630 bp Tn*5398* element has been cloned and completely sequenced and its insertion site determined. Analysis of the resultant data reveals that Tn*5398* is not a classical conjugative transposon but appears to be a mobilizable non-conjugative element. It does not carry any transposase or site-specific recombinase genes, nor any genes likely to be involved in conjugation. Furthermore, using PCR analysis it has been shown that isolates of _C. difficile_ obtained from different geographical locations exhibit heterogeneity in the genetic arrangement of both Tn*5398* and their Erm B determinants. These results indicate that genetic exchange and recombination between these determinants occurs in the clinical and natural environment.

Keywords: Erm determinants, conjugative transposons, mobilization

INTRODUCTION

Clostridium difficile is a Gram-positive, spore-forming, anaerobic bacterium that causes a range of chronic gastrointestinal syndromes, including antibiotic-associated diarrhoea and colitis. The most severe form of these infections is pseudomembranous colitis, a potentially lethal infection that primarily occurs in hospital patients that have been treated with antimicrobial agents such as cephalosporins, penicillins or macrolides (Kelly & LaMont, 1998). *C. difficile* is recognized as the major cause of nosocomial diarrhoea in the USA (Gorbach, 1999) and is a significant pathogen in both British (Wilcox, 1998) and Australian hospitals (Riley *et al.*, 1995).

The *C. difficile* isolates that cause antibiotic-associated diarrhoea are usually not resistant to the antibiotic responsible for the onset of infection, suggesting that an important factor in pathogenesis is the elimination of the normal flora of the intestine by antibiotic therapy. The end result is the provision of an ecological niche for the germination and growth of *C. difficile* spores that originate in the hospital environment (Banerjee & LaMont, 2000). However, because of the strong association between antimicrobial therapy and the onset of *C. difficile* disease, antibiotic resistance in this organism has been extensively studied (Chow *et al.*, 1985; Levett, 1988; Roberts *et al.*, 1994; Wüst & Hardegger, 1988). In addition, studies have shown that resistance to clindamycin increases the risk of *C. difficile*-associated diarrhoea (Johnson *et al.*, 1999).

Clindamycin and erythromycin, both of which are members of the macrolide-lincosamide-streptogramin B (MLS) group of antibiotics, have often been implicated in the onset of *C. difficile*-associated disease. The most common mechanism of resistance to these antibiotics involves N^6-dimethylation of a specific adenine residue of the 23S rRNA molecule (Leclercq & Courvalin, 1991). This alteration of the antibiotic target site is invariably catalysed by an rRNA methyltransferase that is encoded by an *erm* gene. Numerous *erm* genes have been characterized and divided into distinct classes based on their level of sequence similarity (Roberts *et al.*, 1999). In general, each of the classes is loosely

Abbreviations: DR, direct repeat; MLS, macrolide-lincosamide-streptogramin B.

The GenBank accession number for the Tn*5398* element and flanking sequence is AF109075.

Table 1. Bacterial strains and plasmids

Bacterial strain	Relevant characteristics	Source/reference
E. coli		
DH5α	F⁻ ϕ80 d*lacZ*ΔM15Δ(*lacZYA–argF*)U169 *endA1* *recA1* *hadR17* (r_k^- m_k^-) *deoR* *thi-1* *supE44*⁻ *gyrA96* *relA1*	BRL
C. difficile		
630	Em^R	Wüst & Hardegger (1983) (Switzerland)
CD37	Em^S Rif^R	Smith *et al.* (1981) (USA)
JIR1162, JIR1164, JIR1181, JIR1184	CD37 Em^R Rif^R	Transconjugant: 630 × CD37
L289	Em^R	Hayter & Dale (1984) (Surrey, UK)
662	Em^R	Wüst & Hardegger (1983) (Switzerland)
AM140	Em^R	R. Wilkinson, unpublished (USA)
AM480	Em^R	Institute of Medical and Veterinary Science (Adelaide, Australia)
AM1180	Em^R	LaTrobe Valley Hospital (Sale, Australia)
AM1182, AM1185	Em^R	Royal Melbourne Hospital (Melbourne, Australia)
SGC0545	Em^R	Wren *et al.* (1988) (Brussels, Belgium)
B1	Em^R	Borriello *et al.* (1988) (UK)
KZ1604, KZ1610, KZ1614, KZ1623, KZ1655	Em^R	Nakamura *et al.* (1987) (Japan)
660/2, 685	Em^R	Pasteur Institute (France)
24/5-507	Em^R	Monash Medical Centre (Melbourne, Australia)
R5948	Em^R	PHLS (Cardiff, UK)
J9/5602, J9/5610	Em^R	Johnson *et al.* (1999) (New York, USA)
J9/5627, J9/4478	Em^R	Johnson *et al.* (1999) (Massachusetts, USA)
J9p2/5644, J9p2/5650	Em^R	Johnson *et al.* (1999) (Florida, USA)
J7/4224, J7/4290	Em^R	Johnson *et al.* (1999) (Arizona, USA)
B1/832	Em^R	Johnson *et al.* (1999) (Minneapolis, USA)

associated with a particular bacterial genus, with the exception of the Erm B class of determinants, which have been detected in a wide variety of bacterial genera, indicating their potential for intergeneric transfer.

Hybridization analysis has indicated that MLS-resistant strains of *C. difficile* carry *ermB* genes (Farrow *et al.*, 2000; Hächler *et al.*, 1987). The Erm B determinant carried by *C. difficile* strain 630 has been shown to be transferred by a conjugation-like mechanism to *C. difficile* (Wüst & Hardegger, 1983), *Staphylococcus aureus* (Hächler *et al.*, 1987) and *Bacillus subtilis* (Mullany *et al.*, 1995). Transfer has been shown to occur in the absence of detectable plasmid DNA. The *B. subtilis* transconjugants could transfer the Erm B determinant back to *C. difficile*, with integration of the determinant occurring at a specific site on the *C. difficile* chromosome. By contrast, integration was not site-specific in *B. subtilis*. Because of these observations it was proposed that the Erm B determinant from *C. difficile* resides on a conjugative transposon, Tn*5398* (Mullany *et al.*, 1995). This element has not been completely sequenced or characterized, although we have shown (Farrow *et al.*, 2000) that it carries two identical *ermB* genes that are separated by a copy of the direct repeat (DR) sequence that is found on either side of the *ermB* gene from *Clostridium perfringens* (Berryman & Rood, 1995). These *C. perfringens* DR sequences are two directly repeated segments of DNA that primarily consist of an ORF, ORF298, which shows low levels of identity at the amino acid level to Soj and ParA proteins, flanked by highly palindromic sequences, *palA* and *palB*. In *C. difficile* strain 630 the two *ermB* genes are separated by a copy of the DR sequence and are flanked downstream by a variant of this DR sequence from which ORF298 has been deleted and flanked upstream by a DR variant that contains only the sequence downstream of *palB* (Farrow *et al.*, 2000).

The aim of this research was to characterize the putative transposon and to examine the arrangement and distribution of both the *ermB* genes and Tn*5398* in *C.*

difficile strains from diverse geographical locations. The results showed that there was considerable genetic heterogeneity in the organization of the *erm*(B) gene region and that Tn*5398* was an unusual genetic element in that it did not contain any discernible recombinase or mobilization genes.

METHODS

Strains, growth conditions and molecular methods. All C. *difficile* strains (Table 1) were grown in BHIS medium (Smith *et al.*, 1981), supplemented with 50 μg erythromycin ml^{-1} or 20 μg rifampicin ml^{-1}, at 37 °C in an anaerobic glove chamber (Coy Laboratories) in an atmosphere of 80% N_2, 10% H_2, 10% CO_2. DNA was prepared from C. *difficile* strains by dye buoyant density gradient ultracentrifugation at 260000 *g* for 20 h at 20 °C (Abraham & Rood, 1985). Recombinant strains were derivatives of *Escherichia coli* DH5α (BRL) and were grown in 2× YT medium (Miller, 1972). Plasmid DNA was purified from *E. coli* cells by a modified mini alkaline lysis/polyethylene glycol precipitation procedure (Applied Biosystems). Unless otherwise stated molecular manipulations were carried out as described by Sambrook *et al.* (1989).

Table 2. Sequences of oligonucleotide pr

Oligonucleotide no.	Sequence (5′–3′)
2980	AATAAGTAAACAGGTAACGTCT
2981	GCTCCTTGGAAGCTGTCAGTAG
3106	CGGGAGGAAATAATTCTATGAG
3139	ACTTACCCGCCATACCACAGAT
3140	ATTTTATACCTCTGTTTGTTAG
4191	CGTTGTAAAAATTGGGGAAAAG
4192	CAAGTCGGCACGAACACGAACC
4210	TCAATAGACGTTACCTGTTTAC
4349	CATGAGCGAGTTAATTTTGGCA
4350	TGCCAAAATTAACTCGCTCATG
4451	CTGCTTGTAAAGGGATCATAAC
4537	GTCAAGTAAGCAAACATAGTCG
4538	CGACTATGTTTGCTTACTTGAC
6018	AATGGCTGGTTCTACAAATACA
6019	ACTCTGCCTGACAAAACATCTG
6260	GTATGAAAAACACAGCAAAATC
6278	GATTTTGCTGTGTTTTTCATAC
6306	CATTTTCACTATTTTCGTCTAA
6339	ATGCTCGTTTTTAGTATTGAT
6427	AGGGATTGGGACACGCTACATA
6604	TAAGAGTGTGTTGATAGTGC
6785	TTAGGGACACTTACTGATGAAT
6940	TAGCGTGTCCCAATCCCTCATA
7391	ATCAAGGCTCATTCATTAGTAG
7774	ATAATCTCAAGGTCAGTGTGTC
8885	TGGTTCATTTTGTTCGTCTCC
9069	TACTGGCTTTTAGACGCACCTG
9371	GATAGAAATACTCGTCAACAGA
9387	ATTTTTTATTTTTAGGAGTCAT
9409	TACTATTTTCAGAGGTTTGCTC
9493	AACCATCAGACTTCCAAAA
9782	CAAGGGCTGATGATAAACTA
10237	CATAACGGACATAACAACAGCC
11617	CCAAACAGGAAAGATAGCCATA
11662	TGTGGGATGAAGGTTAT
11864	AGTATCCATTTCCTTGTTC
11865	GAACAAGGAAATGGATACT
12143	GTATTTCCTGTTCCACTCC

Cloning of Tn*5398* from C. *difficile*. Chromosomal DNA from *C. difficile* strain 630 was digested overnight with *Xba*I and ligated at 15 °C to *Xba*I-digested DNA from the low-copy-number *E. coli* plasmid vector pWSK29 (Wang & Kushner, 1991). *E. coli* DH5α transformants were selected on medium containing erythromycin (150 μg ml^{-1}).

DNA sequencing and computer analysis. DNA sequencing was carried out using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit and analysed using an Applied Biosystems 373 DNA sequencer. Both DNA strands were sequenced using sequence-specific oligonucleotide primers. Nucleotide and amino acid comparisons were accomplished using the National Center for Biotechnology Information BLAST server at http://www.ncbi.nlm.nih.gov/BLAST/. The SOSUI program developed by the Mitaku Group, Department of Biotechnology, Tokyo University of Agriculture and Technology, Japan (http://sosui.proteome.bio.tuat.ac.jp/cgi-bin/sosui.cgi?/sosui_submit.html) was used to predict the structure of putative transmembrane proteins.

DNA dot blots. DNA dot blots were performed by transferring 8 μg chromosomal DNA from each respective strain to a Hybond-N+ (Amersham Pharmacia Biotech) nylon membrane using a Minifold I Dot Blotter (Schleicher & Schuell). The DNA was cross-linked to the membrane by exposure to UV light at 312 nm for 3 min. The DNA was denatured by prehybridization for a minimum of 3 h in a solution containing SDS. The blots were then probed at high stringency with digoxigenin (DIG)-labelled probes prepared by PCR with the oligonucleotide primers listed in Table 2 as follows. The probes included a 688 bp *erm*(B)-specific probe (PCR primers: 2980 and 2981), a 399 bp ORF298-specific probe (4538 and 4451), a 339 bp *palA*-specific probe (4191 and 4537), a 984 bp *ilvD*-specific probe (6018 and 6278), a 933 bp *hydD*-specific probe (6339 and 6940), a 792 bp ORF13-specific probe (6019 and 6785), a 1166 bp *effD*-specific probe (9069 and 10237) and a 1124 bp *ispD*-specific probe (11546 and 11864). Following high stringency washes the blots were developed using the chemiluminescent substrate CDP-Star (Roche Molecular Biochemicals).

PCR conditions. PCR analysis of *C. difficile* strains was carried out on a Perkin Elmer GeneAmp PCR System 2400. Each reaction contained 0·24 μg chromosomal DNA, 50 μM mixed deoxynucleotide triphosphates, 1× PCR buffer (Roche Molecular Biochemicals), 0·2 μM each oligonucleotide primer and 2·5 units *Taq* polymerase in a 100 μl final volume. Reactions were incubated for 1 cycle of 95 °C for 3 min, 70 °C for 1 min; 30 cycles of 95 °C for 1 min, 50 °C for 2 min, 72 °C for 3 min; 1 cycle of 50 °C for 2 min, 72 °C for 5 min and were then held at 4 °C. Following PCR, 10 μl each reaction was run on a 0·8% agarose gel to detect the PCR products.

Filter matings. *C. difficile* cultures were grown overnight on BHIS agar at 37 °C. Single colonies of the donor and recipient were separately used to inoculate 20 ml BHIS broth and the cultures were grown until mid-exponential phase (OD_{600} ~ 0·45). The cells were then harvested by centrifugation at 1500 *g* for 10 min at room temperature and the cell pellets

resuspended in 1 ml BHIS broth. Aliquots (100 µl) of the donor and recipient suspensions were mixed together on 0·45 µm pore-size nitrocellulose filters on BHIS agar. After incubation for 24 h at 37 °C, the filters were then removed, vigorously washed with 1 ml BHIS broth and 100 µl aliquots spread onto BHIS agar supplemented with the appropriate antibiotics and incubated for 48 h.

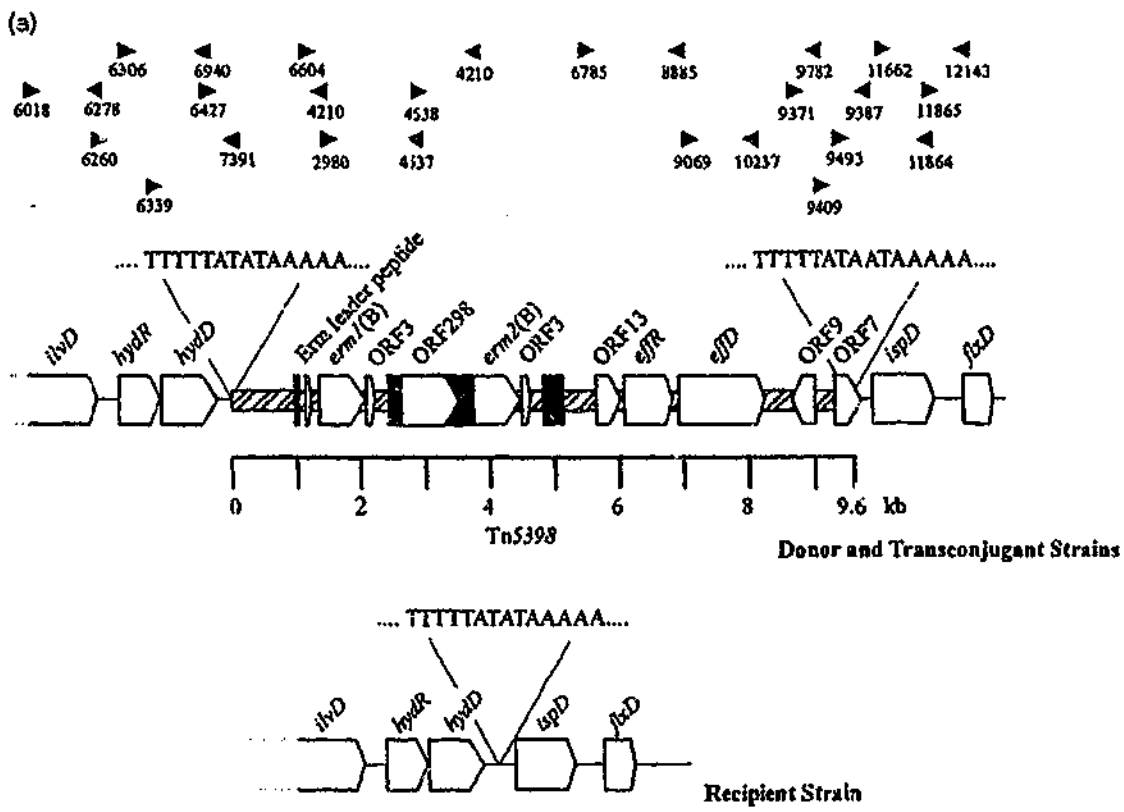
RESULTS

Delineation and comparative analysis of Tn*5398*

To clone the entire Tn*5398* element, *Xba*I-digested chromosomal DNA from *C. difficile* strain 630 was cloned into the low-copy-number vector pWSK29 and an erythromycin-resistant DH5α transformant was isolated and analysed. This strain carried a recombinant plasmid, pJIR1790, which contained a 19·5 kb insert. Approximately 13·5 kb of pJIR1790 was sequenced on both strands using a primer walking approach, beginning with primers within the Erm leader peptide and *erm2*(B) genes (Farrow *et al.*, 2000). Analysis of the resultant data indicated that pJIR1790 carried a potentially novel genetic element.

In addition to duplicated *erm*(B)–ORF3 genes and ORF298, which were previously identified by Farrow *et al.* (2000), the gene region contained nine other potential genes (Fig. 1). Upstream of the Erm B determinant two complete ORFs and one partial ORF were detected. The latter appeared to encode a protein with 52% amino acid sequence identity to IlvD from *B. subtilis* (Sorokin *et al.*, 1996) and had greater than 50% identity to IlvD proteins from many other organisms. IlvD is a dihydroxy-acid dehydratase that is involved in the synthesis of the amino acids isoleucine and valine.

The potential gene product of the ORF located downstream of *ilvD* had 25% identity to a transcription regulator of the TetR family from *Aquifex aeolicus*



(b)

Left End	AAAATTTTTATGAGC**TTTTTATATAAAAA**
Right End	**TTTTTATAATAAAAA**ATATAATGCTTGTA
Target Region	TTATGAGA***TTTTTATATAAAAA***ATATAAT

Fig. 1. Genetic organization of Tn*5398*. (a) Schematic representations of Tn*5398*, as observed in the donor and transconjugant strains, shown to scale. ORFs and their direction of transcription, are represented as blocked arrows. The region encompassed by Tn*5398* is represented by a cross-hatched box and is further indicated by the scale below the diagram. Regions encompassing DR sequences are indicated by black boxes. The location and direction of oligonucleotides used in the delineation of Tn*5398* are shown above the diagram. The target site in the recipient strain CD37 is also shown. The location of each of the ends of the transposon and the target sequence is indicated. (b) The nucleotide sequences of the left and right ends of Tn*5398* and of the target region. Nucleotides included in Tn*5398* are indicated in bold and are underlined. The nucleotides that represent the Tn*5398* target site in *C. difficile* strain CD37 are represented in bold italic type.

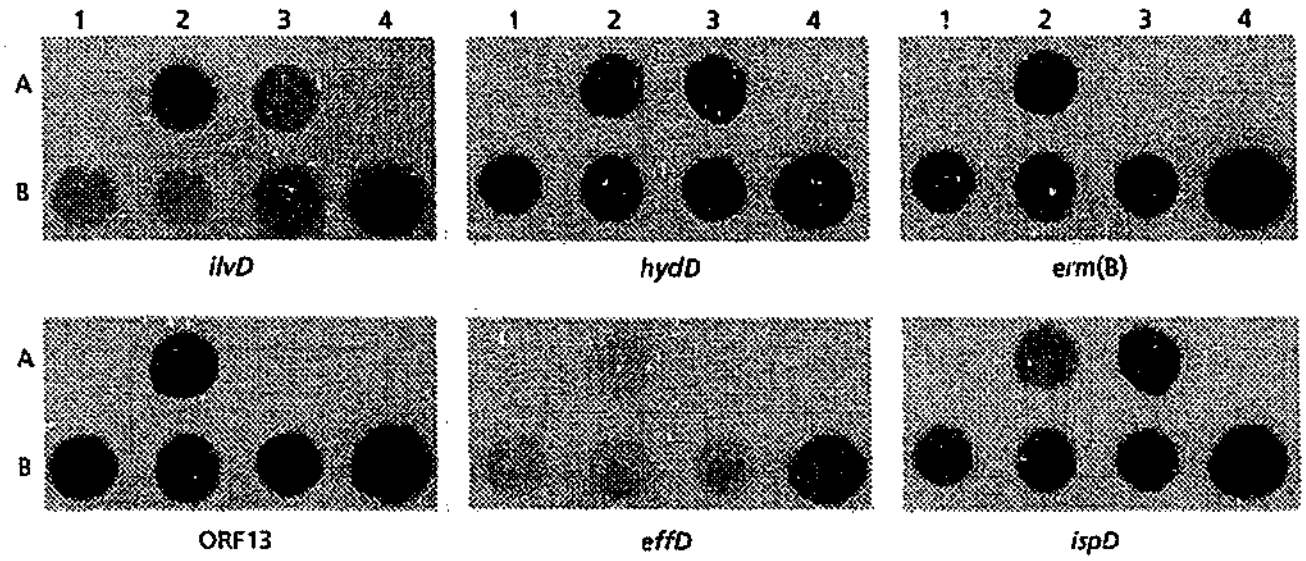


Fig. 2. DNA dot blots of Tn*5398* derivatives. Hybridization analysis was carried out on chromosomal DNA from strains 630 (A2), CD37 (A3) and the transconjugant strains JIR1162 (B1), JIR1164 (B2), JIR1181 (B3) and JIR1184 (B4) using *ilvD*-, *hydD*-, *erm*(B)-, ORF13-, *effD*-, *ispD*-specific probes.

(Deckert *et al.*, 1998), 22% identity to a regulatory protein, IfeR, from *Agrobacterium tumefaciens* (Palumbo *et al.*, 1998) and 19% identity to probable transcriptional regulators from *Pseudomonas aeruginosa* and *Listeria innocua* (Stever *et al.*, 2000). The next ORF encoded a putative product with 22% identity to the hydrolase IpbD from *Pseudomonas putida* (Eaton & Timmis, 1986) and 21% to the PcbD hydrolase from *Archaeoglobus fulgidus* (Klenk *et al.*, 1997). Therefore, these two ORFs were designated as *hydR* and *hydD*, respectively.

Downstream of the Erm B determinant there are three ORFs, designated ORF13, ORF9 and ORF7 (Fig. 1) because their putative products share 20, 43 and 20% amino acid sequence identity, respectively, to the equivalent ORFs from the conjugative transposon Tn*916* from *Enterococcus faecalis* (Flannagan *et al.*, 1994). The function of these ORFs in Tn916 is not known.

The *effR* and *effD* genes were also located downstream of the Erm B determinant (Fig. 1). EffR shows low-level (15–20%) identity to several repressor genes, including the repressor of the *marRAB* operon from *Salmonella typhimurium*, which is involved in multiple antibiotic resistance (Sulavik *et al.*, 1997). EffD showed some similarity to integral membrane proteins from a variety of organisms, including the hyperthermophile *Pyrococcus horikoshii* (Kawarabayasi *et al.*, 1998) and the alkaliphilic *Bacillus* sp. C-125 (Takami *et al.*, 1999). Computer analysis using the membrane prediction program SOSUI suggested that EffD was a membrane protein with 12 membrane-spanning helices.

The putative protein encoded by the gene located 3′ of ORF7, *ispD*, had 48–52% identity to intracellular serine proteases from *B. subtilis* (Koide *et al.*, 1986), *Bacillus amyloliquefaciens* and *Bacillus polymyxa* (Surova *et al.*, 1994; Takekawa *et al.*, 1991). In *B. subtilis* the homologous protease ISP-1 has been postulated to have a critical role in sporulation, possibly through the turnover of intracellular proteins, the processing of spore coat protein precursors and the inactivation of transcarbamylase and several other enzymes (Koide *et al.*, 1986).

The last ORF identified in this study was designated *flxD* because it appeared to encode a protein with 39–45% identity to flavodoxin proteins from *Clostridium beijerinckii* (Tanaka *et al.*, 1974a), *Megasphaera elsdenii* and *Treponema pallidum* (Fraser *et al.*, 1998; Tanaka *et al.*, 1974b). Flavodoxins are low-molecular-mass proteins that function as electron transfer agents in a variety of microbial metabolic processes (Simondsen & Tollin, 1980).

Since *ilvD*, *hydR*, *hydD*, *ispD* and *flxD* were probably housekeeping genes we postulated that Tn*5398* extended from a region downstream of *hydD* to the intergenic ORF7–*ispD* region (Fig. 1). To examine this hypothesis a series of DIG-labelled probes specific for *ilvD*, *hydD*, *erm*(B), ORF13, *effD* and *ispD* were used in dot blots to examine chromosomal DNA from the wild-type *C. difficile* erythromycin-resistant strain 630, the susceptible *C. difficile* recipient CD37 and four independently derived transconjugants. The latter were derivatives of CD37 and were the result of conjugative transfer of erythromycin resistance from strain 630. If a particular gene was of a generic or housekeeping nature, we would expect a gene-specific probe to bind to DNA from all of the strains tested. If the gene was located on Tn*5398* we would expect the probe to hybridize with DNA from the wild-type and the transconjugants but not the recipient. The results showed that each of the predicted housekeeping genes hybridized to all of the strains tested (Fig. 2). By contrast, the *erm*(B), ORF13

Table 3. **Delineation of the conjugative transposon Tn*5398* by PCR analysis***

\+ indicates a PCR product of the expected sized was obtained; − indicates that no PCR product was detected.

Gene(s) amplified	Oligonucleotides	630	CD37	Transconjugants
ilvD	6018 and 6278	+	+	+
hydR and *hydD*	6306 and 6940	+	+	+
Erm leader peptide and *erm1*(B)	6604 and 4210	+	−	+
erm1(B) and ORF298	2980 and 4537	+	−	+
ORF298 and *erm2*(B)	4538 and 4210	+	−	+
ORF13	6785 and 6019	+	−	+
effR	7774 and 8885	+	−	+
effD	9069 and 10237	+	−	+
ORF9	9371 and 9782	+	−	+
ORF7	9493 and 9387	+	−	+
ispD	11662 and 11864	+	+	+
ispD to *flxD*	11865 and 12143	+	+	+

and *effD* probes hybridized only to the wild-type and transconjugant strains, indicating that these genes are likely to be located within Tn*5398*.

A PCR-based strategy was used to more precisely delineate the potential transposon by sequentially moving across the sequenced gene region (Fig. 1, Table 3). Amplification of chromosomal DNA from the wild-type strain 630, the recipient strain CD37 and the four transconjugant strains, confirmed (Table 3) the results of the DNA dot blots and indicated that the *erm*(B) genes, ORF13, *effR*, *effD*, ORF9 and ORF7 were located within Tn*5398*. Based on the products that were amplified from strain 630, but not from strain CD37, the left end was localized to the region between oligonucleotides 6940 and 6604 and the right end was localized to the region between oligonucleotides 9493 and 11662 (Fig. 1).

To precisely define the ends of the putative element, the regions encompassing the ends of the transposon were amplified from strain 630 and the transconjugants and sequenced. The target region from C. *difficile* strain CD37 was amplified using oligonucleotides 6260 and 12143 and was also sequenced. The results showed that the left end of Tn*5398* was located in the intergenic space between *hydD* and the region encoding the Erm leader peptide, at a site 272 bp downstream from the *hydD* stop codon. The right end was shown to be within the coding sequence of ORF7, 84 bp upstream of the stop codon (Fig. 1). Both the right and left ends of the element were extremely AT-rich. The left end consisted of the palindromic sequence TTTTTATATAAAAA, while the right hand consisted of the imperfect palindrome TTTATAATAAAAA. These ends had no significant similarity to the ends of any known conjugative transposon in the databases. The target site in strain CD37 was in the intergenic space between the *hydD* and *ispD* genes, 227 bp downstream of the *hydD* stop codon and 208 bp upstream of the *ispD* start codon (Fig. 1). The target site was also extremely AT-rich and consisted of the palindromic sequence TTTTTATA-TAAAAA (Fig. 1), which is identical to the palindromic sequence located at the left end of the element and differs from the right end by the omission of one of the central A nucleotides. Based on these data it was concluded that Tn*5398* was 9639 bp in length.

Variation in the genetic organization of *erm*(B) gene regions of C. *difficile* strains from different geographical locations

Prior to this study we had reported that the duplicated arrangement of the *erm*(B) genes in C. *difficile* strain 630 was novel when it was compared to Erm B MLS resistance determinants from other bacterial species (Farrow *et al.*, 2000). To determine if this arrangement was common in C. *difficile* we analysed 27 erythromycin-resistant C. *difficile* strains from a range of geographical locations and clinical sources (Table 1). DNA dot blots were carried out on chromosomal DNA from each isolate using an *erm*(B)-specific probe. The results showed that nine of these strains, including all five of the Japanese isolates, did not carry an *erm*(B) gene (Fig. 3). Probing with an ORF298-specific probe showed that an additional four isolates contained an *erm*(B) gene but did not contain the complete DR sequence that contains ORF298 (Fig. 3). Subsequent analysis, using a probe specific for the *palA*-like sequence that is located on either side of ORF298, revealed that the French isolate, 685, which did not contain ORF298, had a DR sequence with an internal deletion (Fig. 3), as found previously in other bacteria (Berryman & Rood, 1995; Farrow *et al.*, 2000).

Therefore, in addition to strain 630, these studies revealed 15 isolates that could carry variants of Tn*5398*.

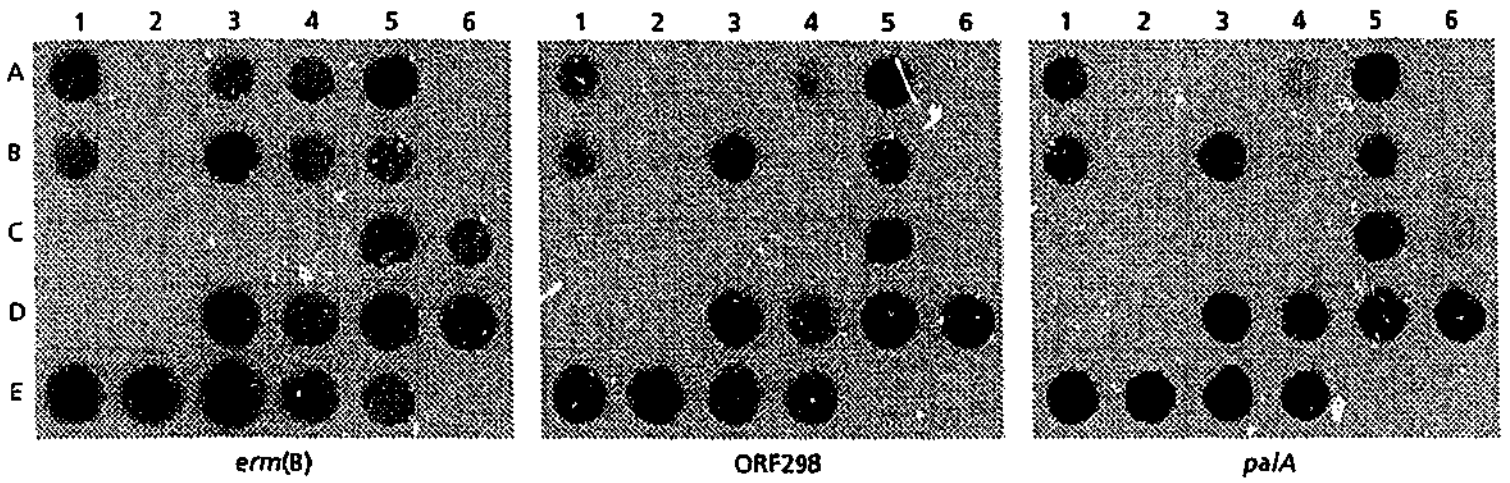


Fig. 3. DNA dot blots on *C. difficile* isolates from different geographical locations. Chromosomal DNA from the strains indicated was probed with *erm*(B)-, ORF298- or *palA*-specific probes. Strain key: A1, 630; A2, CD37; A3, B289; A4, 662; A5, AM140; A6, AM480; B1, AM1180; B2, AM1182; B3, AM1185; B4, SGC0545; B5, B1; B6, KZ1604; C1, KZ1610; C2, KZ1614; C3, KZ1623; C4, KZ1655; C5, 660/2; C6, 685; D1, 24/5-507; D2, R5948; D3, J9/5602; D4, J9/5610; D5, J9/5627; D6, J9/4478; E1, J9p2/5644; E2, J9p2/5650; E3, J7/4224; E4, J7/4290; E5, B1/832.

Table 4. PCR analysis of the genetic organization of the Erm B determinants from *C. difficile* isolates from different geographical locations

\+ indicates that a PCR product of the anticipated size was obtained; − indicates that no PCR product was obtained. Where more than one PCR product was possible, the size of the observed product(s) is indicated in parentheses. The oligonucleotides used, their positions and the region expected to be amplified in each reaction are shown in Fig. 4.

Isolate	PCR 1	PCR 2	PCR 3	PCR 4	PCR 5	PCR 6	PCR 7	PCR 8
630	+(610)	+	+	+(2219 and 1247)	+	−	+	+
CD37	−	−	−	−	−	−	−	−
AM1180	+(610)	+	+	+(2219 and 1247)	+	−	+	+
AM1185	+(610)	+	+	+(2219 and 1247)	+	−	+	+
660/2	+(610)	+	+	+(2219 and 1247)	+	−	+	+
685	+(610)	−	−	+(1247)	−	−	+	+
AM140	+(388)	+	−	+(1247)	−	−	+	−
662	+(388)	+	−	−	−	−	−	−
B1	+(388)	+	−	−	−	−	+	−
J9/5602	+(388)	+	−	−	−	−	−	−
J9/5610	+(388)	+	−	−	−	−	−	−
J9/5627	+(388)	+	−	−	−	−	−	−
J9/4478	+(388)	+	−	−	−	−	−	−
J9p2/5644	+(388)	+	−	−	−	−	−	−
J9p2/5650	+(388)	+	−	−	−	−	−	−
J7/4224	+(388)	+	−	−	−	−	−	−
J7/4290	+(388)	+	−	−	−	−	−	−

A series of eight PCR amplifications were conducted on each isolate to determine the arrangement of the Erm B determinant in these strains. The first reaction was designed to detect the presence of an *erm* leader peptide upstream of an *erm*(B) gene. If the arrangement was the same as in strain 630, we would expect a fragment of 610 bp. Four of the isolates had this profile (Table 4). The remaining isolates had a 388 bp product, which as revealed by sequence analysis, consisted of the same region but without the leader peptide sequence, as is found upstream of the *C. perfringens erm*(B) gene (Berryman & Rood, 1995) (Fig. 4).

The remaining PCR reactions were designed to step sequentially across the *erm*(B) gene region and to detect the presence and location of the Tn*5398* gene, *effD*. The

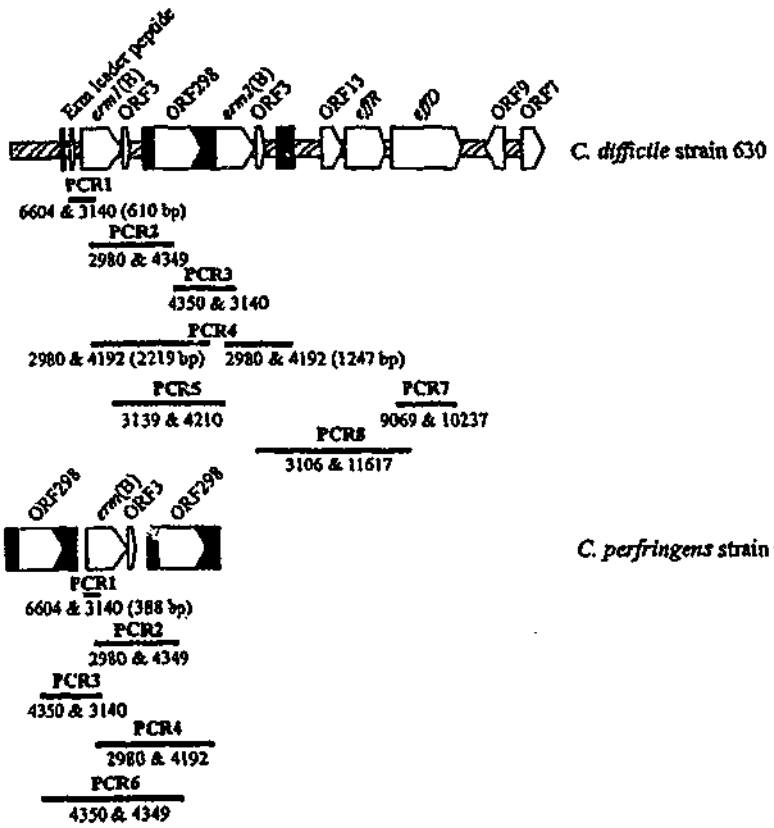


Fig. 4. PCR analysis of the genetic organization of the Erm B determinants from *C. difficile* isolates from different geographical locations. The regions encompassed by each reaction are shown in relation to the arrangements of the Erm B determinants from *C. difficile* strain 630 (Fig. 1) and *C. perfringens* strain CP592 (Berryman & Rood, 1995). Results of the PCR analysis are shown in Table 4.

combined results of these PCR experiments (Table 4) allowed us to divide the strains into five groups based on the arrangement of the *erm*(B) gene region (Fig. 5). The first group of nine isolates were resistant to erythromycin but did not contain an *erm*(B) gene (Fig. 5a). This group included all of the Japanese isolates, three of the Australian isolates and a British isolate. The second group of three isolates, from the UK, Belgium and the USA, had an *erm*(B) gene but did not have either complete or incomplete DR sequences (Fig. 5b). Eleven strains, nine of which were from the USA, had a complete DR sequence that was located downstream of the *erm*(B) gene (Fig. 5c). The two non-USA strains in this group also carried an *effD* gene but it was not associated with the *erm*(B) gene. Strain 685 was very similar except that it had an *erm* leader peptide, the *erm*(B) gene was followed by an incomplete DR sequence and the *effD* gene was associated with the *erm*(B) gene (Fig. 5d).

PCR primers designed to detect the presence of two complete DR sequences flanking an *erm*(B) gene, as is observed in the arrangement of the Erm B determinant from *C. perfringens*, were included in these experiments (Fig. 4). No product was amplified from any of the *C. difficile* strains, although a product was observed when DNA from *C. perfringens* strain CP592 was included as a positive control.

Three *C. difficile* isolates had the same arrangement as observed in strain 630 and appeared to have a complete copy of Tn*5398*. That is, they had two *erm*(B) genes, one located upstream of a complete DR sequence and the other upstream of an incomplete DR sequence, and had a genetically linked ORF13–*effD* gene region (Fig. 5e). This group was geographically diverse as it included two Australian isolates from different hospitals and a French isolate, as well as the prototype Swiss isolate, strain 630.

DISCUSSION

Tn*5398* appears to be a mobilizable but non-conjugative genetic element

Previous studies suggested that the erythromycin resistance determinant in *C. difficile* strain 630 was located on a conjugative transposon, Tn*5398* (Mullany *et al.*, 1995). Conjugative transposons are discrete DNA elements that are normally integrated into the bacterial chromosome and are characterized by their ability to encode their own movement from one bacterial cell to another by a process requiring cell to cell contact. Conjugative transposition involves excision of the element from the chromosome to form a non-replicating covalently closed circular intermediate, which can either integrate elsewhere in the genome or transfer by conjugation to another cell where it integrates into the recipient's genome. To carry out these reactions conjugative transposons generally encode site-specific recombinases that are responsible for the excision and integration of the element and other proteins that are required for conjugation (Salyers *et al.*, 1995).

Tn*5398* does not appear to encode genes that are involved in either excision, integration or conjugation. We have shown that it is 9·6 kb in size and encodes two copies of *erm*(B)–ORF3 and one copy of ORF298 (Farrow *et al.*, 2000). In addition, there is only one incomplete and four complete ORFs located within the putative transposon. The proteins encoded by the *effR* and *effD* genes are unlikely to be involved in the

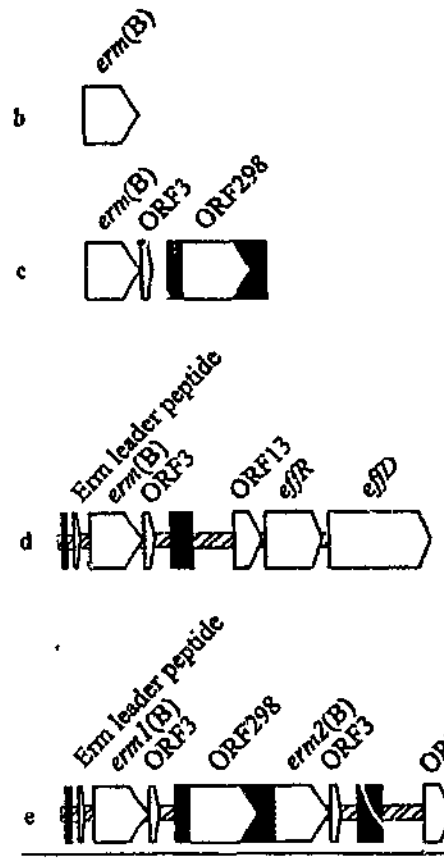


Fig. 5. The different arrangements of the Erm B determinants. A diagrammatic representation of each genetic organization is shown to scale. The isolates that have this organization are listed beside each diagram.

transposition of the putative transposon, as they appear to encode a potential efflux protein and its associated regulator. The remaining ORFs, ORF13, ORF9 and ORF7, have similarity to equivalent ORFs from the prototype conjugative transposon, Tn*916*. The ORF7 protein shows limited homology to sigma factors (Flannagan *et al.*, 1994). It has been postulated to have a regulatory role in the mobility of Tn*916* because in the presence of tetracycline, increased ORF7 expression leads to increased transcription of ORF7, ORF8, *xis*, *int* and other genes (Celli & Trieu-Cuot, 1998). The finding that the right end of Tn*5398* is internal to ORF7 may have implications for the level of excision, transfer and integration of the element. After excision from the donor chromosome, Tn*5398* would leave behind part of ORF7, resulting in an incomplete ORF7 gene in the circular intermediate. The result could be altered levels of transcription of other genes involved in transfer of the element. The end of ORF7 also appears to be the target sequence for the element in recipient strains such as CD37. It is this region of the ORF7 protein that has identity to the helix–turn–helix motif of sigma factors, which is involved in DNA binding. Fusion with this region may provide a selective advantage for recombination of the circular intermediate at the target site.

In Tn*916* ORF9 has been predicted to be a putative transcriptional repressor; however, the role of this repressor in the mobility of Tn*916* has not been determined (Celli & Trieu-Cuot, 1998). The ORF13 protein has no known role in the mobility of Tn*916*. If these Tn*916* homologues have any role at all in the movement of the Tn*5398* element, it would appear that they are most likely to encode proteins that are involved in the regulation of transposition events, rather than proteins that are involved in excision, mobilization, transposition or integration.

ORF298 is the only other ORF that could encode a protein involved in Tn*5398* mobility. The putative ORF298 protein has some similarity to replication proteins and proteins from the ParA and Soj families. ParA and Soj proteins generally have a role in the partitioning of plasmids and chromosomes during the replication cycle (Easter *et al.*, 1998; Sharpe & Errington, 1996). It appears unlikely that ORF298 has a role in either the excision or integration of Tn*5398* but this possibility cannot be completely eliminated.

In addition to being capable of catalysing its own conjugative transposition, Tn*916* is capable of enhancing the transfer of another homologous conjugative transposon that is co-resident in the cell (Flannagan & Clewell, 1991) and of mobilizing non-conjugative plasmids (Jaworski & Clewell, 1995; Showsh & Andrews, 1999). Based on its small size and our comparative analysis of the genes carried on Tn*5398*, we postulate that it is more likely to be a non-conjugative but mobilizable element rather than a conjugative transposon.

Mobilization of non-conjugative plasmids by Tn*916* does not appear to be dependent on the presence of a functional mobilization, or *mob*, gene on the plasmid, but does require the presence of an origin of transfer, or *oriT*, sequence. It is postulated that the same protein or

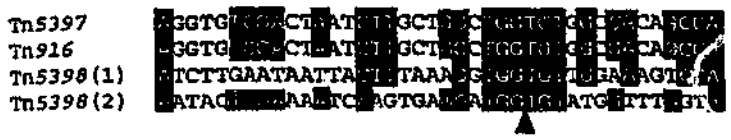


Fig. 6. Alignment of the *oriT* sites from Tn*916*, Tn*5397* and the two potential *oriT* sites located on Tn*5398*. Regions of sequence identity are boxed in black. The *nic* site is marked with a black triangle.

proteins involved in the nicking of the Tn*916* circular intermediate at its *oriT* site nicks similar sequences present on co-resident mobilizable plasmids. Once nicked, the plasmid then assumes a relaxed form, which is capable of being transferred during conjugation (Showsh & Andrews, 1999).

In addition to Tn*5398*, strain 630 carries a second conjugative transposon, Tn5397 (Mullany *et al.*, 1990), which is closely related to Tn*916* (Roberts *et al.*, 2001). Comparison of Tn*5397* and Tn*916* reveals that they have very similar conjugation regions (ORF15–ORF23), but different insertion and excision modules (Roberts *et al.*, 2001). In addition, the *oriT* sites present on both elements are identical and the putative ORF23 mobilization proteins have greater than 90% identity. Analysis of Tn*5398* reveals the presence of two potential *oriT* sites, which are located within the coding sequence of ORF298 and in the intergenic space between ORF3 and ORF13 (Fig. 6). The *nic* sites (TGGTGT) of these two potential *oriT* sites are identical to the *nic* sites found on Tn*916* and Tn*5397*. We postulate that Tn*5398* is excised from the chromosome either by TndX, the site-specific recombinase responsible for the excision of Tn*5397*, or by another large resolvase encoded on the chromosome of strain 630. The resultant circular intermediate of Tn*5398* would then be nicked at one of the *oriT* sites by the same protein responsible for nicking the Tn*5397* circular intermediate and subsequently transferred to a recipient cell by a Tn*5397*-dependent process. Once in the recipient the element may be either integrated into the chromosome by means of the TndX protein or by another large resolvase encoded on the chromosome of the recipient cell. Attempts to verify this hypothesis by using outward firing PCR primers to amplify the putative circular intermediate were unsuccessful. However, in the absence of a positive chromosomal control, little can be concluded from this experiment.

Note that there is an alternative explanation for these results. It is possible that the region excised from the chromosome is much larger than that proposed and includes genes upstream of the proposed left end (*ilvD*, *hydR* and *hydD*) and downstream of the proposed right end (*ispD*, *flxD*). After transfer of this region, which contains Tn*5398*, it could be integrated into the recipient chromosome by RecA-dependent homologous recombination. The end result would be the integration of Tn*5398* and some of the genes flanking the element. This hypothesis is consistent with the experimental data.

Tn*5398* is present in *C. difficile* strains from diverse sources

To determine if Tn*5398*, with its duplicated *erm*(B) genes, was common in *C. difficile* we carried out a series of comparative PCR and dot blot studies. The results showed that the arrangement of the Erm B determinants in 15 *C. difficile* isolates from diverse sources was not the same, with four major genetic variants being detected. The simplest variant was that of a single *erm*(B) gene and the most complex was represented by the Erm B determinant carried by Tn*5398* (Fig. 5). Tn*5398* was present in strain 630, which was originally from Switzerland, and three other *C. difficile* isolates, AM1180 and AM1185, isolated from different Australian hospitals, and the French isolate 660/2. The identification of these strains provides evidence that the duplicated *erm*(B) gene organization carried by Tn*5398* is widespread and is therefore likely to be transferred between *C. difficile* isolates.

In general, each of the genetic variants was represented by isolates from a wide variety of geographical sources, with the exception of the group that carried a single *erm*(B) gene flanked by a downstream complete DR sequence (Fig. 5c). This group of isolates, with the exception of strains 662 and B1, were all isolated in the USA. With one exception, AM140, these USA isolates were isolated from large outbreaks of diarrhoea that occurred in four hospitals located in different parts of the country (Johnson *et al.*, 1999). Analysis of these strains by restriction digestion, PCR and PFGE had previously determined that these isolates were actually derivatives of the same strain (Johnson *et al.*, 1999), which was referred to as the epidemic strain. The isolates were all highly resistant to clindamycin and were shown to account for approximately 30–66% of the *C. difficile* strains isolated at these hospitals during the period 1989 to 1992 (Johnson *et al.*, 1999). We previously showed that each of these isolates contained an *erm*(B) gene (Johnson *et al.*, 1999). Our finding that they contain Erm B determinants with the same genetic organization supports the conclusion that they are derivatives of an epidemic strain. It was previously suggested (Johnson *et al.*, 1999) that the *erm*(B) gene present in these isolates was associated with Tn*5398*. This study reveals that this is unlikely as the Tn*5398*-specific gene, *effD*, was not detected in these isolates.

We previously proposed that the *C. perfringens* Erm B determinant, which consists of an *erm*(B) gene flanked by two complete DR sequences, represents the Erm B progenitor and that Erm B determinants in other bacteria probably evolved by homologous recombination events that removed part of the DR sequences (Berryman & Rood, 1995). The different genetic arrangements of the various *C. difficile* Erm B determinants observed in this study are all consistent with this hypothesis. However, despite the fact that many *erm*(B) genes are located on mobile genetic elements that can move freely between different bacterial species, *C. perfringens* still appears to be the only species that

contains an *erm*(B) gene associated with two intact copies of the DR sequence.

ACKNOWLEDGEMENTS

We thank the Australian National Health and Medical Research Council for its research support and Peter Mullany for helpful discussions. K.A.F. was the recipient of an Australian Postgraduate Award.

REFERENCES

Abraham, L. J. & Rood, J. I. (1985). Molecular analysis of transferable tetracycline resistance plasmids from *Clostridium perfringens*. *J Bacteriol* 161, 636–640.

Banerjee, S. & LaMont, J. T. (2000). Non-antibiotic therapy for *Clostridium difficile* infection. *Curr Opin Infect Dis* 13, 215–219.

Berryman, D. I. & Rood, J. I. (1995). The closely related *ermB-ermAM* genes from *Clostridium perfringens*, *Enterococcus faecalis* (pAMβ1), and *Streptococcus agalactiae* (pIP501) are flanked by variants of a directly repeated sequence. *Antimicrob Agents Chemother* 39, 1830–1834.

Borriello, S. P., Davies, H. A. & Barclay, F. E. (1988). Detection of fimbriae amongst strains of *Clostridium difficile*. *FEMS Microbiol Lett* 49, 65–67.

Celli, J. & Trieu-Cuot, P. (1998). Circularization of Tn*916* is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol Microbiol* 28, 103–117.

Chow, A. W., Cheng, N. & Bartlett, K. H. (1985). *In vitro* susceptibility of *Clostridium difficile* to new β-lactam and quinolone antibiotics. *Antimicrob Agents Chemother* 28, 842–844.

Deckert, G., Warren, P. V., Gaasterland, T. & 12 other authors (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392, 353–358.

Easter, C. L., Schwab, H. & Helinski, D. R. (1998). Role of the *parCBA* operon of the broad-host range plasmid RK2 in stable plasmid maintenance. *J Bacteriol* 180, 6023–6030.

Eaton, R. W. & Timmis, K. N. (1986). Characterization of a plasmid-specified pathway for catabolism of isopropylbenzene in *Pseudomonas putida* RE204. *J Bacteriol* 168, 123–131.

Farrow, K. A., Lyras, D. & Rood, J. I. (2000). The macrolide-lincosamide-streptogramin B resistance determinant from *Clostridium difficile* 630 contains two *erm*(B) genes. *Antimicrob Agents Chemother* 44, 411–413.

Flannagan, S. E. & Clewell, D. B. (1991). Conjugative transfer of Tn*916* in *Enterococcus faecalis*: *trans* activation of homologous transposons. *J Bacteriol* 173, 7136–7141.

Flannagan, S. E., Zitzow, L., Su, Y. A. & Clewell, D. B. (1994). Nucleotide sequence of the 18-kb conjugative transposon Tn*916* from *Enterococcus faecalis*. *Plasmid* 32, 350–354.

Fraser, C. M., Norris, S. J., Weinstock, G. M. & 30 other authors (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281, 375–388.

Gorbach, S. (1999). Antibiotics and *Clostridium difficile*. *N Engl J Med* 341, 1690–1691.

Hächler, H., Berger-Bächi, B. & Kayser, F. H. (1987). Genetic characterization of a *Clostridium difficile* erythromycin-clindamycin resistance determinant that is transferable to *Staphylococcus aureus*. *Antimicrob Agents Chemother* 31, 1039–1045.

Hayter, P. M. & Dale, J. W. (1984). Detection of plasmids in clinical isolates of *Clostridium difficile*. *Microbios Lett* 27, 151–156.

Jaworski, D. D. & Clewell, D. B. (1995). A functional origin of transfer (*oriT*) on the conjugative transposon Tn*916*. *J Bacteriol* 177, 6644–6651.

Johnson, S., Samore, M. H., Farrow K. A. & 9 other authors (1999). Epidemics of diarrhea caused by a clindamycin-resistant strain of *Clostridium difficile* in four hospitals. *N Engl J Med* 341, 1645–1651.

Kawarabayasi, Y., Sawada, M., Horikawa, H. & 22 other authors (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res* 5, 55–76.

Kelly, C. & LaMont, J. T. (1998). *Clostridium difficile* infection. *Annu Rev Med* 49, 375–390.

Kolde, Y., Nakamura, A., Uozumi, T. & Beppu, T. (1986). Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. *J Bacteriol* 167, 110–116.

Klenk, H. P., Clayton, R. A., Tomb, J. F. & 48 other authors (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390, 364–370.

Leclercq, R. & Courvalin, P. (1991). Bacterial resistance to macrolide, lincosamide, and streptogramin antibiotics by target modification. *Antimicrob Agents Chemother* 35, 1267–1272.

Levett, P. N. (1988). Antimicrobial susceptibility of *Clostridium difficile* determined by disc diffusion and breakpoint methods. *J Antimicrob Chemother* 22, 167–173.

Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Mullany, P., Wilks, M., Lamb, I., Clayton, C., Wren, B. W. & Tabaqchali, S. (1990). Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*. *J Gen Microbiol* 136, 1343–1349.

Mullany, P., Wilks, M. & Tabaqchali, S. (1995). Transfer of macrolide-lincosamide-streptogramin B (MLS) resistance in *Clostridium difficile* is linked to a gene homologous with toxin A and is mediated by a conjugative transposon, Tn*5398*. *J Antimicrob Chemother* 35, 305–315.

Nakamura, S., Yamakawa, K., Nakashio, S., Kamiya, S. & Nishida, S. (1987). Correlation between susceptibility to chloramphenicol, tetracycline and clindamycin, and serogroups of *Clostridium difficile*. *Med Microbiol Immunol* 176, 79–82.

Palumbo, J. D., Kado, C. I. & Phillips, D. A. (1998). An isoflavonoid-inducible efflux pump in *Agrobacterium tumefaciens* is involved in competitive colonization of roots. *J Bacteriol* 180, 3107–3113.

Riley, T., Codde, J. & Rouse, I. (1995). Increased length of hospital stay due to *Clostridium difficile* associated diarrhoea. *Lancet* 345, 455–456.

Roberts, A. P., Johanesen, P. A., Lyras, D., Mullany, P. & Rood, J. I. (2001). Comparison of Tn*5397* from *Clostridium difficile*, Tn*916* from *Enterococcus faecalis*, and the CW459 *tet*(M) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* 147, 1243–1251.

Roberts, M. C., McFarland, L. V., Mullany, P. & Mulligan, M. E. (1994). Characterization of the genetic basis of antibiotic resistance in *Clostridium difficile*. *J Antimicrob Chemother* 33, 419–429.

Roberts, M. C., Sutcliffe, J., Courvalin, P., Jensen, L. B., Rood, J. I. & Seppala, H. (1999). Nomenclature for macrolide and macrolide-

lincosamide-streptogramin B antibiotic resistance determinants. *Antimicrob Agents Chemother* 43, 2823–2830.

Salyers, A., Shoemaker, N., Stevens, A. M. & Li, L. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* 59, 579–590.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sharpe, M. E. & Errington, J. (1996). The *Bacillus subtilis soj-spoOJ* locus is required for a centromere-like function involved in prespore chromosome partitioning. *Mol Microbiol* 21, 501–509.

Showsh, S. A. & Andrews, R. E., Jr (1999). Analysis of the requirement for a pUB110 *mob* region during Tn*916*-dependent mobilization. *Plasmid* 41, 179–186.

Simondsen, R. P. & Tollin, G. (1980). Structure–function relations in flavodoxins. *Mol Cell Biochem* 33, 13–24.

Smith, C. J., Markowitz, S. M. & Macrina, F. L. (1981). Transferable tetracycline resistance in *Clostridium difficile*. *Antimicrob Agents Chemother* 19, 997–1003.

Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S. D. & Serror, P. (1996). Sequence analysis of the *Bacillus subtilis* chromosome region between the *serA* and *kdg* loci cloned in a yeast artificial chromosome. *Microbiology* 142, 2005–2016.

Stover, C. K., Pham, X. Q., Erwin, A. L. & 28 other authors (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406, 959–964.

Sulavik, M. C., Dazer, M. & Miller, P. F. (1997). The *Salmonella typhimurium mar* locus: molecular and genetic analyses and assessment of its role in virulence. *J Bacteriol* 179, 1857–1866.

Surova, I. A., Revina, L. P., Ianonis, V. V., Kolesnikova, L. A. & Stepanov, V. M. (1994). Primary structure of the intracellular serine proteinase from *Bacillus amyloliquefaciens*. III. Amino acid sequence of peptides obtained by hydrolysis with a Glu, Asp-specific proteinase. Reconstruction of the entire amino acid sequence of the proteinase. *Bioorg Khim* 20, 1310–1326.

Takami, H., Nakasone, K., Hirama, C., Takaki, Y., Masui, N., Fuji, F., Nakamura, Y. & Inoue, A. (1999). An improved physical and genetic map of the genome of alkaliphilic *Bacillus* sp. C-125. *Extremophiles* 3, 21–28.

Takekawa, S., Uozumi, N., Tsukagoshi, N. & Udaka, S. (1991). Proteases involved in generation of β- and α-amylases from a large amylase precursor in *Bacillus polymyxa*. *J Bacteriol* 173, 6820–6825.

Tanaka, M., Haniu, M., Yasunobu, F. & Mayhew, S. G. (1974a). The amino acid sequence of the *Clostridium* MP flavodoxin. *J Biol Chem* 249, 4393–4396.

Tanaka, M., Haniu, M., Yasunobu, F., Mayhew, S. G. & Massey, V. (1974b). Correction of the amino acid sequence of *Peptostreptococcus elsdenii* flavodoxin. *J Biol Chem* 249, 4397.

Wang, R. F. & Kushner, S. R. (1991). Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* 100, 195–199.

Wilcox, M. H. (1998). Treatment of *Clostridium difficile* infection. *J Antimicrob Chemother* 41, 41–46.

Wren, B. W., Mullany, P., Clayton, C. & Tabaqchali, S. (1988). Molecular cloning and genetic analysis of a chloramphenicol acetyltransferase determinant from *Clostridium difficile*. *Antimicrob Agents Chemother* 32, 1213–1217.

Wüst, J. & Hardegger, U. (1983). Transferable resistance to clindamycin, erythromycin, and tetracycline in *Clostridium difficile*. *Antimicrob Agents Chemother* 23, 784–786.

Wüst, J. & Hardegger, U. (1988). Studies on the resistance of *Clostridium difficile* to antimicrobial agents. *Zentbl Bakteriol Mikrobiol Hyg (A)* 267, 383–394.

Received 21 May 2001; revised 2 July 2001; accepted 3 July 2001.