National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa (Ontario)
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395 rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Our file / Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-19384-5

Canada

1

PM-1 3 1/2" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

SEQUENCE ANALYSIS AND CHARACTERIZATION
OF THE *mob* REGION OF A CONJUGATIVE PLASMID pCU1

by E. Suzanne Paterson, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy
Department of Biology

Carleton University

Ottawa, Ontario

Sept. 20, 1996

Copyright

1996, Suzanne Paterson

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395 rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / votre référence

Your file / votre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-19384-5

Canada

Name E. Suzanne Paterson

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Molecular

0307

U-M-I

SUBJECT TERM

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dances 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Literary Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
General 0289
Ancient 0290
Linguistics 0291
Modern 0401
Literature 0294
General 0295
Classical 0297
Comparative 0298
Medieval 0316
Modern 0591
African 0305
Asian 0352
Canadian (English) 0355
Canadian (French) 0593
English 0311
Germanic 0312
Latin American 0315
Middle Eastern 0313
Romance 0314
Slavic and East European

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion 0318
General 0321
Biblical Studies 0319
Clergy 0320
History of 0322
Philosophy of 0469
Theology

SOCIAL SCIENCES

American Studies 0323
Anthropology 0324
Archaeology 0326
Cultural 0327
Physical 0310
Business Administration 0272
General 0770
Accounting 0454
Banking 0338
Marketing 0385
Canadian Studies 0501
Economics 0503
General 0505
Agricultural 0508
Commerce-Business 0509
Finance 0510
History 0511
Labor 0358
Theory 0366
Folklore 0351
Geography 0578
Gerontology 0578
History 0578
General

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science 0615
General 0616
International Law and 0617
Relations 0814
Public Administration 0452
Recreation 0626
Social Work 0627
Sociology 0938
General 0631
Criminology and Penology 0628
Demography 0629
Ethnic and Racial Studies 0630
Individual and Family 0700
Studies 0344
Industrial and Labor 0709
Relations 0999
Public and Social Welfare 0453
Social Structure and 0453
Development 0453
Theory and Methods 0453
Transportation 0453
Urban and Regional Planning 0453
Women's Studies

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy 0475
Animal Culture and 0476
Nutrition 0359
Animal Pathology 0478
Food Science and 0479
Technology 0480
Forestry and Wildlife 0817
Plant Culture 0777
Plant Pathology 0746
Plant Physiology 0306
Range Management 0287
Wood Technology 0308
Biology 0309
General 0379
Anatomy 0329
Biostatistics 0353
Botany 0369
Cell 0793
Ecology 0410
Entomology 0307
Genetics 0317
Limnology 0433
Microbiology 0821
Molecular 0778
Neuroscience 0472
Oceanography 0786
Physiology 0760
Radiation 0425
Veterinary Science 0996
Zoology

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences 0566
General 0300
Audiology 0992
Chemotherapy 0567
Dentistry 0350
Education 0769
Hospital Management 0758
Human Development 0982
Immunology 0564
Medicine and Surgery 0347
Mental Health 0569
Nursing 0570
Nutrition 0380
Obstetrics and Gynecology 0354
Occupational Health and 0381
Therapy 0571
Ophthalmology 0419
Pathology 0572
Pharmacology 0382
Pharmacy 0573
Physical Therapy 0574
Public Health 0575
Radiology 0575
Recreation

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences 0485
Chemistry 0749
General 0486
Agricultural 0487
Analytical 0488
Biochemistry 0738
Inorganic 0490
Nuclear 0491
Organic 0494
Pharmaceutical 0495
Physical 0754
Polymer 0405
Radiation 0605
Mathematics 0986
Physics 0606
General 0608
Acoustics 0748
Astronomy and 0607
Astrophysics 0798
Atmospheric Science 0759
Atomic 0609
Electronics and Electricity 0752
Elementary Particles and 0756
High Energy 0611
Fluid and Plasma 0463
Molecular 0346
Nuclear 0984
Optics 0346
Radiation 0984
Solid State 0346
Statistics 0984
Applied Sciences 0346
Applied Mechanics 0984
Computer Science

Engineering 0537
General 0538
Aerospace 0539
Agricultural 0540
Automotive 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0348
Heat and Thermodynamics 0545
Hydraulic 0546
Industrial 0547
Marine 0794
Materials Science 0548
Mechanical 0743
Metallurgy 0551
Mining 0552
Nuclear 0549
Packaging 0765
Petroleum 0554
Sanitary and Municipal 0790
System Science 0428
Geotechnology 0796
Operations Research 0795
Plastics Technology 0994
Textile Technology

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Carleton University

Ottawa, Canada K1S 5B6

Thesis contains black & white
photographs &/or explanatory tables
which when microfilmed may lose their
significance. The hardcopy of the
thesis is available upon request from
Carleton University Library.

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis

SEQUENCE ANALYSIS AND CHARACTERIZATION
OF THE *mob* REGION OF A CONJUGATIVE PLASMID pCU1

submitted by
E. Suzanne Paterson, B.Sc., M.Sc.

in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

Chair, Department of Biology

Thesis Supervisor

External Examiner

Carleton University
September 20, 1996

ABSTRACT

The N type *oriT* of plasmid pMUR274 was cloned on a 474 bp *RsaI-SspI* fragment and the nucleotide sequence determined. A comparison of the pMUR274 *oriT* sequence and the sequence of the *oriTs* of IncN plasmid pCU1 and IncW plasmid R388 demonstrated 57% and 34.5% sequence identity, respectively. Site-specific recombination between the pCU1 *oriT* and the *oriT* of pMUR274 resulted in the formation of a hybrid *oriT* containing one half of each parental sequence. The junction point of the hybrid occurred within a 10 bp sequence GCTATACACC present in both sequences and represents the *nic* site of each *oriT*. Mutation of the first A residue within the 10 bp junction sequence had little effect on transfer while mutation of either the second or third A residue reduced transfer by at least 1000 fold. Site-specific recombination between a wildtype pCU1 *oriT* and each of these three mutant pCU1 *oriTs* demonstrated that *nic* lies on either the right or left side of the second T residue in the 10 bp junction sequence. Site-specific recombination between wildtype and mutant pCU1 *oriTs* also demonstrated that both initiation and termination of transfer are affected by mutation of either A residue to the right of *nic* within the 10 bp junction and by deletion of a portion of the AT rich region 42 bases to the right of *nic*. Deletion of 22 bp into the right end of the pCU1 *oriT*, removing the last inverted repeat sequence, affected initiation but not termination.

The *mob* region of pCU1 was mutagenized with *TnphoA* and complementation tests with Tn5 mutants of IncN plasmid pKM101 demonstrated three complementation

regions which correspond to three ORFs designated *traK*, *traJ* and *traHI* seen in the nucleotide sequence. The amino acid sequence of the TraK protein shows 26% identity with the TrwA protein of IncW plasmid R388. The TraJ protein shares 40% and 23% sequence identity with the TrwB protein of R388 and the TraD protein of the F plasmid, respectively. The TraHI protein shares 41% sequence identity with the TrwC of R388 and 25% identity with the TraI (helicase I/nickase) of the F plasmid. TraK, TraHI and the carboxyl portion of the TraJ protein were overexpressed and visualized by SDS-PAGE. High level induction of the *traK* gene reduced the amount of TraK protein observed, suggesting that the TraK protein may autoregulate its expression, possibly by binding to sequences within the *oriT* region which lies directly upstream of the *traK* gene. Both TraK and TraHI were shown to play important roles in site specific recombination at *nic* while TraJ plays only a minor role.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Iyer. Over the years, he has provided support and guidance as well as a dynamic environment in which to work and study. I would further like to express my thanks for his continued support after his retirement last year. I would like to wish him a long and satisfying retirement, filled with many interesting adventures equal to his long and distinguished academic career.

Many thanks are also extended to Dr. Wyndham and Dr. Johnson for their efforts as members of my committee, their suggestions and insight have proven instructive. I would like to add a special thanks to Dr. Wyndham for stepping into the role of my supervisor this past year and allowing me to finish my research with little disruption.

I have had the great pleasure of working with many wonderful people over the years. Their suggestions, instruction and advice have been greatly appreciated. In particular I would like to thank Martin Holcik for instruction in SDS-PAGE and Margarita Rodriguez for getting double-stranded sequencing to work. I would also like to thank Jussi Helava for his calm guidance through the intricacies of Unix, the net and other assorted computer problems. Over the past year, I have worked in the laboratory of Dr. Iain Lambert and Dr. Myron Smith, both whom I would like to thank for acceptance into their lab and for their interest in my research. I would also like to extend many thanks to Margret More and Steve Winans of Cornell University for sharing the *traHI* sequence with me.

I would like to thank my friends and family for their support and encouragement.

and finally, I would like to thank my husband Tony who has shared both the joys of successful experiments as well as the disappointment those gone awry. He has provided a constant source of strength and support throughout my graduate studies and I thank him for this.

This research was sponsored by a grant to Dr. Iyer from the Medical Research Council of Canada.

TABLE OF CONTENTS

ACCEPTANCE SHEET	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xiii
1. INTRODUCTION	1
2. LITERATURE REVIEW	12
2.1 RELATIONSHIPS BETWEEN TRANSFER SYSTEMS	12
2.2 THE ORIGIN OF TRANSFER (<i>oriT</i>)	16
2.3 INITIATION OF TRANSFER VIA SITE-SPECIFIC NICKING AT <i>criT</i>	25
2.4 OTHER PROTEINS THAT BIND TO THE <i>oriT</i> REGION	42
2.5 COVALENT ATTACHMENT OF PROTEIN TO THE 5' END OF NICKED DNA	43
2.6 UNWINDING THE DNA BY HELICASE	47
2.7 CIRCULARIZATION OF THE TRANSFERRED STRAND	50
2.8 DIRECTION AND MODE OF TRANSFER THROUGH THE MEMBRANE	56

2.9	COMPLEMENTARY STRAND SYNTHESIS	59
3.	METHODS AND MATERIALS	63
4.	RESULTS	88
4.1.	COMPARISON OF THE <i>oriT</i> REGIONS OF pCU1 AND pMUR274	88
4.1.1	Cloning the <i>oriT</i> region of pMUR274	88
4.1.2	Sequencing the pMUR <i>oriT</i>	95
4.1.3	Sequence alignment of the pMUR and pCU1 <i>oriT</i> regions	100
4.2	CONSTRUCTION OF A HYBRID <i>oriT</i> BY SITE-SPECIFIC RECOMBINATION	106
4.2.1	Construction of pMKN derivatives	107
4.2.2	Site-specific recombination	107
4.2.3	Sequencing the putative hybrid <i>oriTs</i>	111
4.2.4	Mobilization of the hybrid <i>oriT</i> derivatives	114
4.3	SITE-SPECIFIC AND DELETION MUTAGENESIS OF THE <i>oriT</i> OF pCU1	119
4.3.1	Generation of random point mutations by PCR	120
4.3.2	Site-specific mutation of <i>nic</i> by PCR cloning with degenerate oligonucleotides	121
4.3.3	Deletion mutagenesis of <i>oriT</i>	130
4.4	SITE-SPECIFIC RECOMBINATION AT <i>oriT</i>	136
4.4.1	Locating <i>nic</i> by intramolecular site-specific recombination	136
4.4.2	Identification of initiation and termination functions in the pCU1	

<i>oriT</i>	142
4.5 TRANSPOSON MUTAGENESIS AND SEQUENCING OF THE <i>mob</i>	
REGION	145
4.5.1 Construction of pSP34	146
4.5.2 Mutagenesis of pCU1 <i>mob</i> by Tn <i>phoA</i>	149
4.5.3 Complementation tests between pSP34::Tn <i>phoA</i> and	
pKM101::Tn5	149
4.5.4 Analysis of the nucleotide sequence of the <i>mob</i> region of pCU1	152
4.5.5 Expression of <i>mob</i> proteins	159
4.5.6 Amino acid sequence analysis of pCU1 <i>mob</i> proteins	165
4.6 IDENTIFICATION OF <i>mob</i> FUNCTIONS INVOLVED IN NICKING AT	
<i>oriT</i>	177
4.6.1 Intermolecular site-specific recombination assay with	
pSP34::Tn <i>phoA</i>	177
5. DISCUSSION	183
6. REFERENCES	200

LIST OF TABLES

TABLE	DESCRIPTION	PAGE
2.1	Comparative summary of genes involved in conjugal and F DNA transfer metabolism.	14
3.1	<i>E. coli</i> strains used in this study	64
3.2	Bacteriophages used in this study	65
3.3	Plasmids used in this study	66
3.4	Primers used in this study.	68
4.1	Mobilization frequency of pMUR <i>oriT</i> derivatives by pSP274	96
4.2	Formation of Km ^r transconjugants during mobilization of pMKN derivatives by pCU101 and pSP274	110
4.3	Mobilization of pCU1 and pMUR wildtype and hybrid <i>oriT</i> plasmids by pCU101 and pSP274	115
4.4	Mobilization and recombination frequencies of pNKN derivatives	143
4.5	Complementation of pSP34::Tn <i>phoA</i> mutants by pKM101. Tn <i>o</i> mutants.	150
4.6	Mobilization of pSP27 by pSP34 <i>tra</i> mutants from S17.1	178

LIST OF FIGURES

FIGURE	DESCRIPTION	PAGE
1.1	Plasmid map of pCU1	8
2.1	Schematic representation of the F and P type <i>oriT</i> region	19
4.1	Hybridization of pCU1 <i>oriT</i> probe to <i>Hpa</i> I, <i>Ssp</i> I and <i>Dra</i> I digests of pCU1, pSP274 and pSP545 DNA.	90
4.2	Construction of pMO2, pMO2S and pMO2R.	93
4.3	Sequencing strategy for the pMUR <i>oriT</i>	98
4.4	Nucleotide sequence of the 1375 bp <i>Dra</i> I <i>oriT</i> fragment from pSP274.	101
4.5	Nucleotide sequence alignment of the <i>oriT</i> regions of pCU1, pMUR274 and R388.	103
4.6	Construction of pMKN derivatives.	108
4.7	Formation of hybrid <i>oriT</i> via site-specific recombination.	112
4.8	PCR cloning of <i>nic</i> mutants.	122
4.9	Point and deletion mutations within the <i>nic</i> region of the pCU1 <i>oriT</i>	125
4.10	Deletion and point mutations lying outside the <i>nic</i> region.	132
4.11	Construction of pNKNE and pNKNB double <i>oriT</i> derivatives . . .	137
4.12	Restriction and functional map of pSP34.	147
4.13	Strategy for sequencing the <i>mob</i> region of pCU1.	153

4.14	DNA sequence and putative ORFs of pCU1 <i>traK</i> and <i>traJ</i> regions	155
4.15	5' end of pCU1 <i>tra</i> genes cloned into the pET23d expression vector.	160
4.16	SDS-PAGE of pCU1 Mob proteins.	163
4.17	Amino acid sequence alignment of pCU1 TraK, R388 TrwA and F TraY.	166
4.18	Amino acid sequence alignment of pCU1 TraJ, R388 TrwB and F TraD.	169
4.19	Comparison of hydrophobicity plots of pCU1 TraJ, R388 TrwB and F TraD.	171
4.20	Amino acid sequence alignment of pCU1 TraHI, R388 TrwC and F TraI	173
4.21	Schematic diagram comparing the results of mobilization of pSP27 by pSP34 from S17-1 by conduction	179
5.1	Comparison of <i>nic</i> sequence in pCU1, R388 and F plasmids. . . .	184

LIST OF ABBREVIATIONS

+/-	presence or absence of phenotype
<	less than
>	greater than
%	percent
A	Adenine
aa	amino acid
ala	Alanine
Ap	Ampicillin
AS	Acetosyringinone
ATP	Adenosine 5' triphosphate
bp	base pair(s)
C	Cytosine
°C	degrees celsius
CaCl₂	Calcium chloride
cfu	colony forming units
Cm	Chloramphenicol
cpm	counts per minute
CsCl	Cesium chloride
dNTP	deoxynucleotide triphosphate
ddNTP	dideoxynucleotide triphosphate
dH₂O	deionized water
DNA	deoxyribonucleic acid
DTT	Dithiothreitol
ds	double-stranded
EDTA	Ethylenediaminetetraacetic acid
F'	F plasmid carrying chromosomal DNA
G	Guanine
hr	hour(s)

HCl	Hydrochloric acid
his	Histidine
Inc	Incompatability
IHF	Integration host factor
IPTG	Isopropyl-β-D-thiogalactopyranoside
IR	inverted repeat
l	Litre
LB	L broth
LBA	L broth agar
leu	Leucine
KCl	Potassium chloride
Km	Kanamycin
Kb	Kilobase(s)
kDa	Kilo dalton(s)
mamp	milli Amp(s)
mg	milligram(s)
ml	millilitre(s)
mM	millimolar
M	molar
min	minute(s)
MgCl₂	Magnesium chloride
MgSO₄	Magnesium sulphate
<i>mob</i>	mobilization gene
msec	millisec(s)
mt	mutant
NaCl	Sodium chloride
NaI	Sodium iodide
Nal	Nalidixic acid
NaOH	Sodium hydroxide
<i>nic</i>	site of strand specific nicking within the oriT

ORF	open reading frame
<i>oriT</i>	origin of transfer
<i>oriV</i>	origin of vegetative replication
μg	picogram(s)
μmole	picomole(s)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pfu	plaque forming units
phe	Phenylalanine
<i>phoA</i>	alkaline phosphatase gene
r	resistant
<i>rec</i>	recombination gene
rep	Replicon
Rif	Rifampicin
RNA	ribonucleic acid
RPM	revolutions per minute
SDS	Sodium deodecyl sulfate
s	sensitive
sec	second(s)
ser	Serine
Sm	Streptomycin
SM	Sodium chloride Magnesium sulphate buffer
Sp	Spectinomycin
SOC	Electroporation media
ss	single-stranded
SSB	Single-strand binding protein
SSC	Sodium citrate sodium chloride
STET	Sucrose Tris EDTA Triton-X

T	Thymine
TA	Tris acetate restriction buffer
TAE	Tris acetate EDTA electrophoresis buffer
TBE	Tris borate EDTA electrophoresis buffer
Tc	Tetracycline
T-DNA	double-stranded DNA transferred in <i>Agrobacterium</i> virulence system
TE	Tris EDTA
Tn	transposon
<i>tra</i>	transfer gene/region
<i>trw</i>	transfer gene of IncW plasmid
trp	Tryptophan
T strand	single-stranded DNA transferred DNA of <i>Agrobacterium</i> virulence system
tyr	Tyrosine
UV	ultraviolet
<i>vir</i>	virulence gene
wt	wildtype
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YT	Yeast tryptone broth
μCi	microcurrie(s)
μg	microgram(s)
μl	microlitre(s)
μm	micrometre(s)
μM	micromolar

1. INTRODUCTION

Bacterial conjugation is a natural process by which DNA is transferred from one bacterial cell (the donor) to another (the recipient) during cell to cell contact. In Gram negative bacteria, this process is encoded by genes which are typically located on plasmids. Plasmids that are able to promote their own transfer are called conjugative plasmids. The genes which are responsible for transfer are referred to as transfer (*tra*) genes and typically comprise 25-50% of the plasmid.

Many different conjugative plasmids have been identified in both Gram negative and Gram positive bacteria. Due to the fact that many conjugative plasmids carry genes which encode resistance to antibiotics, conjugal transfer is responsible for much of the spread of antibiotic resistance in clinically important bacteria. Indeed, many of the conjugative plasmids studied to date were originally isolated from clinical specimens.

Conjugative plasmids from Gram negative bacteria are grouped primarily based on similarities in replication functions (Datta, 1974). Plasmids which encode similar replicative functions are unable to be stably maintained in the same cell and are placed in the same incompatibility or Inc group. Conjugative plasmids belonging to the same Inc group generally encode similar transfer systems.

Conjugal transfer can be broken down into two major events: the formation of stable mating pairs and conjugal DNA metabolism. The majority of *tra* genes are involved in the assembly of extracellular, proteinaceous appendages called conjugative pili which are used to establish cell-to-cell contact between mating cells. It has been

postulated that retraction of the conjugative pili, an event observed with the conjugative pili of the F plasmid (Jacobson, 1972), brings the two cell walls into direct contact (Clark and Warren, 1979). Once contact is made, other *tra* genes stabilize the mating-pair by an unknown mechanism (Frost *et al.*, 1994; Ippen-Ihler and Skurray, 1993). Conjugative pili also act as receptor sites for bacteriophage, making the host cell susceptible to infection by the phage. Pilus morphology, serology and specificity for bacteriophage have all been used to group conjugative transfer systems (Bradley, 1980). For a recent review of conjugative pili see Frost (1993).

The remaining *tra* genes act following the formation of stable mating pairs to prepare the DNA for transfer from the donor to the recipient in a process referred to as conjugal DNA metabolism or mobilization. The subset of *tra* genes involved in this process are also known as mobilization (*mob*) genes (Willetts and Wilkins, 1984). The activities involved in conjugal DNA metabolism centre around a small region of DNA carried by all conjugative plasmids known as the origin of transfer or *oriT*. The *oriT* is the only *tra* function which acts *in cis*; all of the other *tra* functions are able to act *in trans*. While all conjugative plasmids carry an *oriT*, only in very closely related plasmids is the sequence of this region conserved and even then, only limited identity is observed. Due to the variation in nucleotide sequence within the *oriT* of different plasmids, the *mob* proteins which interact with the *oriT* tend to show plasmid specificity (Willetts and Maule, 1985). Despite the plasmid specificity of *mob* functions, studies over the past decade have demonstrated that conjugal DNA metabolism occurs in a similar manner for all of the transfer systems in Gram negative bacteria studied to date.

This information has recently been reviewed and is outlined below (Lanka and Wilkins, 1995; Wilkins and Lanka, 1993).

During conjugal DNA metabolism, one small Tra protein binds to the *oriT* in a plasmid specific manner. The binding of this protein is required for the recognition of the *oriT* by a second, larger Tra protein either via protein-protein interaction with the smaller *oriT* bound protein or by recognition of a sequence which has been bent or made single stranded by the activities of the smaller bound protein. In some cases additional plasmid encoded Tra proteins or host encoded proteins such as Integration Host Factor (IHF) are incorporated in the DNA-protein complex at *oriT* referred to as the relaxosome.

Within the relaxosome, the endonucleolytic activity of the large protein introduces a nick into one strand of the DNA at a specific site referred to as *nic*. Following this, the large protein becomes covalently attached to the 5' end of the nicked strand. In some systems, the nicking protein also encodes a helicase activity which acts to unwind the nicked DNA in the 5'-3' direction. In the absence of a plasmid encoded helicase, a host encoded protein is believed to provide this function. Complementary strand synthesis on the unnicked strand (either via extension of the 3' end of the nicked strand or by other host encoded replication mechanisms used for lagging strand synthesis) is believed to occur concomitantly with the unwinding of the nicked strand, providing a reconstituted *oriT* sequence such that once the nicked strand is completely unwound, a second nick within the reconstituted *oriT* can be made. Following the second nick, the large protein is transferred to the new 5' end while the old 5' end is ligated to the newly

generated 3' end resulting in the release of a single-stranded, circular molecule which is transferred through the cell wall.

The route taken by the single-stranded circular molecule through the cell wall is as yet unknown. There is some evidence to support early speculation that the DNA travels through the hollow conjugative pilus (Ippen and Valentine, 1967; Harrington and Rogerson, 1990). It has also been suggested that retraction of the pilus into the donor leaves a pore through which the DNA can pass (Frost, 1993). Other theories suggest that the pili act only to ensure cell-to-cell contact and that other proteins, either plasmid or host encoded, form the pore through which the DNA passes (Lanka and Wilkins, 1995). In either scenario, there is some indication that a plasmid encoded protein acts to anchor the DNA to the membrane pore and that transfer may occur as the DNA is being unwound.

In addition to the many conjugative plasmids identified to date, a number of smaller plasmids have also been identified, which, while unable to promote their own transfer, are able to be transferred or 'mobilized' by a conjugative plasmid in the same donor cell. These mobilizable plasmids carry an *oriT* region as well as several *mob* genes but do not encode genes required for mating pair formation (Clark and Warren 1979). The Mob proteins of the smaller plasmid interact in a plasmid specific manner with their own *oriT* to prepare the DNA for transfer to the recipient cell which is brought into contact with the donor via stable mating pair functions encoded by the larger co-resident conjugative plasmid. This type of mobilization is referred to as donation and is characterized by high transfer frequencies comparable to those of the conjugative

plasmid. Mobilization by donation does not alter either plasmid and can result in the transfer of the non-conjugative plasmid to some transconjugants without co-transfer of the conjugative plasmid (Clark and Warren, 1979).

Mobilization of a non-conjugative plasmid can occur by a second mechanism which does not require it to contain an *oriT* region. In this process, called conduction, the non-conjugative and conjugative plasmids recombine and transfer of the resultant co-integrate depends entirely upon the conjugative plasmid (Clark and Warren, 1979). Co-integration can occur by various mechanisms involving either RecA dependant homologous recombination or various RecA independent mechanisms including illegitimate recombination, site-specific recombination, transposition of transposable elements or recombination between replicative or transfer origins (for a recent review see Reinmunn and Hass, 1993). This type of transfer is characterized by very low transfer frequencies due to the relatively low frequency of the recombination event. Conduction can also be identified by the demonstration of the co-integrate plasmid or of separate conjugative and non-conjugative plasmids in the recipient cell which are different from either parental plasmid and by the inability to isolate transconjugants carrying only the non-conjugative plasmid (Clark and Warren, 1979).

The ability of conjugative plasmids to donate mobilizable plasmids demonstrates that the conjugal DNA metabolism functions of these mobilizable plasmids are able to interact successfully with the mating pair functions of the conjugative plasmids. However, in most instances, conjugative *tra* functions are not able to complement mutations in the *mob* functions of the smaller plasmid. This trend is also seen in the

interaction of *tra* functions among closely related conjugative plasmids. While *pili* functions are generally interchangeable between plasmids belonging to the same Inc group (Willettts and Skurray, 1980), *mob* functions show a much higher degree of plasmid specificity even among closely related plasmids (Willettts and Maule, 1985; Pansegrau *et al.*, 1988).

While genetic studies have identified many different transfer systems among Gram negative bacteria, the increase in DNA sequence data which has become available over the past few years has demonstrated previously unsuspected relationships between the different systems. One of the most interesting relationships which has been supported by sequence data is that of the *Agrobacterium* T-DNA transfer system with the conjugative plasmid transfer systems of the Gram negative bacteria. The T-DNA transfer system promotes the transfer of DNA on tumour inducing (Ti) or root inducing (Ri) plasmids in *A. tumefaciens* into a wide variety of mainly dicotyledonous plants where it is integrated into the plant genome and the genes contained on it are expressed. The DNA transferred (the T-DNA) is contained between two short directly repeated sequences which, like the *oriT* of the conjugative plasmids, are required *in cis*. Transfer of the T-DNA is dependant on the expression of a series of *trans* acting virulence or *vir* genes carried by the Ti or Ri plasmids in a manner analogous to the mobilization of *oriT* containing plasmids by the *tra* genes. For a recent review of the *Agrobacterium* virulence system, see Kado (1993).

The functional relationship between the *vir* system and the transfer systems of conjugative plasmids is illustrated by the ability of the *oriT* sequence from the small

mobilizable plasmid RSF1010 to substitute for the direct repeats of the T-DNA in the transfer of DNA from bacteria to plants (Buchanan-Wollaston *et al.* 1987). Conversely, the *vir* genes have been used to transfer RSF1010 between bacterial cells (Beijersbergen *et al.* 1992). In both instances, both the *vir* genes of the *Agrobacterium* plasmid and the *mob* genes of RSF1010 were required. Numerous examples of both functional and sequence identities between the *tra* genes of conjugative plasmids and the *vir* genes have been identified (Kado, 1993; Lessl *et al.* 1994; Pohlman *et al.* 1994; Farrand *et al.* 1996; Alt-Morbe *et al.* 1996).

Of the conjugative transfer systems identified to date, the most extensively studied are those of the IncF (reviewed by Ippen-Ihler and Skurray, 1993; Frost *et al.* 1994) and the IncP plasmids (reviewed by Guiney and Lanka, 1989; Guiney, 1993). The transfer systems of the IncW (Valentine and Kado, 1989), IncI and IncN (Iyer, 1989) plasmids have been investigated to a limited extent as has the *mob* system of the IncQ plasmids (reviewed by, Guiney, 1989).

The IncN plasmid pCU1 (for Carleton University) is a 39 Kb conjugative plasmid isolated in Canada from a clinical isolate of *Salmonella* (Konarska-Kozłowska and Iyer, 1981). pCU1 carries genes which confer resistance to the antibiotics ampicillin, streptomycin and spectinomycin.

The transfer system of pCU1 comprises approximately 16.5 Kb extending from coordinates 31.5 to 9 Kb on the pCU1 map seen in Figure 1.1. Initial transposon mutagenesis of the *tra* region identified 8 complementation regions, six of which (*traA-F*) are involved in pili production (Thatte *et al.* 1985). pCU1 encodes short, rigid pili

Figure 1.1 Plasmid map of pCUL. The eight complementation regions within the transfer region (Tra) identified by Thatte *et al* (1985) are indicated by the letters from A to H. The extent of complementation regions C, D, G and H are unknown. *oriT* is shown as a solid box at the right end of *tra*. The plasmid replicon (Rep) is also indicated together with genes which confer antibiotic resistance: Sp, spectinomycin. Sm streptomycin; Ap, ampicillin. Restriction sites relevant to this study are also indicated

similar to those encoded by the IncP and IncW plasmids (Bradley, 1979). Like other IncN pili they confer susceptibility to icosahedral double-stranded DNA bacteriophage PRD1 and filamentous, single-stranded DNA phage IKe (Khatoon *et al.*, 1972). The restriction map of pCU1 shares striking similarity to that of another IncN plasmid pKM101 (a deletion derivative of R46), particularly in the *tra* region (Iyer, 1989). The nucleotide sequence of the portion of the *tra* region of pKM101 which encodes conjugative pili has been determined and the amino acid sequence shown to share identity with the Vir proteins of the *Agrobacterium* T-DNA transfer system (Pohlman *et al.*, 1995).

The remaining two complementation regions (*traG* and *traH*) are believed to be involved in conjugal DNA metabolism. An earlier study located the *oriT* region of pCU1 to a 325 bp sequence at one end of the *tra* region. This sequence has been determined (Paterson and Iyer, 1992) and shown to be identical to the *oriT* sequence of pKM101 (Coupland *et al.*, 1987). These sequences show 34.5% identity to the *oriT* sequence of IncW plasmid R388 (Llosa *et al.*, 1991). While pCU1 can mobilize a plasmid carrying just the *oriT* of pKM101, it can only mobilize the *oriT* of R388 if the IncW *mob* functions are also provided *in trans* (Bolland *et al.*, 1990). Two plasmids, pMUR274 and pMUR545, from a collection of plasmids isolated prior to the antibiotic era (Hughes and Datta, 1983; Datta and Hughes, 1983) encode an N type transfer system as determined by susceptibility to the N specific phage IKe, but do not appear to belong to the IncN incompatibility group (Krishnan *et al.*, 1991). These plasmids are unable to mobilize the pCU1 *oriT* unless the pCU1 *mob* functions are provided (Paterson and Iyer, 1992).

While a number of other IncN plasmids have been identified only very limited work has been done on N3 (Ando and Arai, 1981; Glocker and Rached, 1990) and R15 (Dobritsa *et al.*, 1985).

This study is a continuation of earlier work done to locate and sequence the *oriT* of pCU1. In particular, it was the goal of this study to locate the *nic* site within *oriT*. Further effort was made to identify functionally important sequences within the *oriT* as well as the *tra* functions which interact specifically with the *oriT* to promote the transfer of pCU1. It was hoped that this information would provide further insight into the relationship between the *mob* regions of IncN plasmids and those of other conjugative plasmids.

2. LITERATURE REVIEW

2.1 RELATIONSHIPS BETWEEN TRANSFER SYSTEMS

After the initial discovery of the F plasmid, subsequent classification of other conjugative plasmids was based on a comparison to the F plasmid. The ability to co-inhabit the same cell or to inhibit transfer of a co-resident F plasmid provided a basis for initial grouping of the rapidly growing number of identified conjugative plasmids (Datta, 1974). It eventually became evident that there were numerous conjugative plasmids of which the F plasmid and its relatives were but a small minority. While it was believed that all of the transfer systems functioned in a similar manner, the inability of functions from one plasmid to complement *Tra*⁻ mutants in another suggested that many if not all of the transfer functions were plasmid specific. Mobilization of small non-conjugative plasmids such as ColE1, CloDF13 and RSF1010 by different conjugative plasmids provided the first differentiation of plasmid specific and non-specific transfer functions. That these small plasmids carried genes necessary for their own transfer suggested that the functions encoded by these genes probably represented the plasmid specific functions (Warren *et al*, 1979; Willetts and Crowther, 1981).

Plasmid specificity was further illustrated in the different levels of complementation observed for *tra* functions of the plasmids that belonged to the group of F like plasmids. Willetts and Maule (1985) demonstrated that the various F plasmids could be separated into four groups based on their ability to mobilize recombinant plasmids carrying the *oriT* regions of plasmids from the other groups. The ability to

mobilize the *oriT* of another plasmid usually corresponded to the plasmid's ability to complement mutations in the *traM*, *traI* and *traY* genes of that plasmid. These groupings also showed general agreement with the level of sequence identity observed between the *oriT* sequences of the different plasmids.

Similar plasmid specificity has been observed in the IncP group of plasmids where those genes which encode proteins which interact specifically with the *oriT* region, show greater plasmid specificity (Yakobson and Guiney, 1983). Weak DNA:DNA hybridization results between the *oriT* regions of the IncP α and IncP β correspond to the inability of plasmids from one group to mobilize recombinant plasmids carrying only the *oriT* region of the other. However, the ability of the IncP β plasmid R751 to mobilize a recombinant construct carrying a larger portion of the IncP α plasmid RP4 *tra* region including the genes involved in relaxosome formation, indicated that it was the *mob* genes which conferred plasmid specificity. Subsequent studies indicated that it is these plasmid specific functions that encode proteins which bind to the *oriT* (Pansegrau *et al*, 1988).

As discussed in the following sections, the transfer systems encoded by plasmids of the Gram negative bacteria possess many functional similarities. Table 2.1 lists the function of each of the *mob* genes in the different plasmids studied to date. Each encodes two proteins required for nicking within the *oriT*. While the smaller of the two binds specifically to a site close to *nic*, the larger contains the nicking activity. In most cases, the larger protein has been shown to be bifunctional with the scission activity located in the amino terminus. The carboxyl terminus appears to have divergent

Table 2.1 Comparative summary of genes involved in conjugal and T-DNA transfer metabolism.

Plasmid type	Nickase/ ligase	<i>oriT</i> binding and nicking	Membrane spanning protein	other <i>oriT</i> binding	other
IncP	<i>traI</i>	<i>traJ</i>	<i>traG</i>	<i>traK</i>	<i>traH</i> <i>traC</i> (primase)
IncI	<i>nikB</i>	<i>nikA</i>	?	?	<i>sog</i> (primase)
IncF	<i>traI'</i>	<i>traY</i>	<i>traD</i>	<i>traM</i>	ND
IncW	<i>trwC'</i>	<i>trwA</i>	<i>trwB</i>	ND	ND
IncQ	<i>mobA</i>	<i>mobB</i>	ND	<i>mobC</i>	ND
Ti/Ri virulence	<i>virD2</i>	<i>virD1</i>	<i>virD4</i>	ND	<i>virE</i> (ss binding)
Ti/Ri conjugative	<i>traA</i> ^(*)	?	<i>traG</i>	?	?

? unknown due to lack of study
 * (putative) helicase activity
 ss single-stranded DNA
 ND not detected in a well defined system

functions which can be used to separate the plasmids into two groups: those which encode a helicase activity (IncF and IncW plasmids) and those which do not (IncP, IncI and IncQ plasmids). The function encoded by the carboxyl terminus of the latter group is unknown except in the case of the IncQ plasmids which encodes a primase function. The fact that both the IncP and IncI plasmids encode separate primase functions, which are not part of the nicking protein, suggests that this group of primase encoding plasmids may be further divided into two groups.

Structural organization and nucleotide sequence conservation within the *oriT* region supports the grouping described above: the IncP-like group including both the IncP α and the IncP β plasmids, IncI plasmids, mobilizable plasmid pTF-FC2; the F-like group including the different IncF plasmids, the IncN plasmids and the IncW plasmids; and the IncQ like plasmids including the three IncQ plasmids and pTF1 as well as conjugative plasmids from the Gram positive bacteria (Pansegrau and Lanka, 1991; Lanka and Wilkins, 1995).

Protein sequence analysis also supports grouping of the conjugative plasmids into the F-like and the P-like groups. Plasmids belonging to both groups encode a protein with hydrophobic domains in the N terminus which could act as membrane spanning regions. All of these proteins show some sequence identity, however, the proteins of the F-like plasmids contain two highly conserved nucleotide binding sites while one of the sites in the corresponding P-like plasmids proteins is less conserved (Lessl *et al.*, 1992; Llosa *et al.*, 1994₂). Similarly, the nicking proteins of the IncP, IncI and IncQ share three conserved motifs which are only marginally conserved in endonucleases of the IncF and IncW plasmids (Pansegrau and Lanka, 1991; Llosa *et al.*, 1994₂; Balzer *et al.*, 1994).

While conjugative systems of most Gram negative bacteria studied to date can be placed into one of the three groups: IncP like, IncF like or IncQ like, the conjugative and virulence systems of the *Agrobacterium* Ti and Ri plasmids fit into more than one of these three groups suggesting that they may have arisen from chimeras of the other systems. For example, while the Vir proteins responsible for mating bridge formation show sequence identity with the pili proteins of the IncN plasmids (Pohlman *et al.*, 1995), the DNA processing proteins of the *vir* systems share identity with the IncP Mob proteins (Lessl *et al.*, 1992; Lessl *et al.*, 1994, Balzer *et al.*, 1994). The Ti border sequences also show identity to the *oriT* sequence of the IncP and IncI plasmids (Lessl *et al.*, 1994). In the case of the conjugative systems of the Ti plasmids, both the pili and hydrophobic *mob* proteins share sequence identity with the analogous proteins of the IncP like plasmids while the *oriT* sequence shows identity to the IncQ *oriT* sequence. In addition, while the protein involved in nicking at the Ti *oriT* shows sequence identity to the endonuclease portion of the IncQ scission protein, it also shares sequence identity to the helicase domain of the IncF and IncW proteins (Alt-Morbe *et al.*, 1996; Farrand *et al.*, 1996).

2.2. THE ORIGIN OF TRANSFER (*oriT*)

The origin of transfer (*oriT*) is the only *tra* function required in *cis* for the transfer of conjugative or mobilizable plasmids. As its name implies, the *oriT* is the site at which transfer is initiated. Initiation involves the interaction of *tra/mob* proteins which cleave one strand of the DNA at a unique site referred to as *nic*.

The nucleotide sequence for the *oriT* of several conjugative plasmids is known

including the F plasmid (Thompson *et al.*, 1984), its close relatives R1 (Ostermann *et al.*, 1984; Finlay *et al.*, 1986), R100 (McIntire and Dempsey, 1987), ColB4 (Finlay *et al.*, 1986), P307 (Goldner *et al.*, 1987), pSU316 (Lopez *et al.*, 1991) and pED208 (Di Laurenzio *et al.*, 1991), IncP plasmids RP4/RK2 (Guiney and Yakobson, 1983) and R751 (Furste *et al.*, 1989), IncN plasmids pCU1/R46 (Paterson and Iyer, 1992; Coupland *et al.*, 1987), IncW plasmid R388 (Llosa *et al.*, 1991), Inc11 plasmid R64-11 (Komano *et al.*, 1988), *Agrobacterium tumefaciens* plasmid pTiC58 (Cook and Farrand, 1992) as well as of mobilizable plasmids ColE1 (Bastia *et al.*, 1978), CloDF13 (Snidjers *et al.*, 1983), the IncQ plasmids RSF101G (Derbyshire *et al.*, 1987), R1162 (Brasch and Meyer, 1987) and pSC101 (Willett and Wilkins, 1984) and of *Thiobacillus ferrooxidans* plasmids pTF-FC2 (Rohrer and Rawlings, 1992) and pTF1 (Drolet and Lau, 1991). In addition, the nucleotide sequence of the *Agrobacterium* T-DNA border sequences which are functionally analogous to the *oriT* region of conjugative plasmids has been determined for plasmids pTiC58 (Durrenberger *et al.*, 1989), pTiA6 (Albright *et al.*, 1987), pTiAch5 (Gielen *et al.*, 1984) and RiA4 and pTi15955 (Jouanin *et al.*, 1989).

The minimal region required *in cis* for transfer ranges from the 25 bp direct repeats of the T-DNA in *Agrobacterium* Ti plasmids and the 38 bp sequence required for transfer of R1162 to the larger 200-300 bp regions required by the F plasmid, RP4, R388 and pCU1. While these regions show only limited if any sequence identity, all but the T-DNA borders of *Agrobacterium* Ti and Ri plasmids contain conserved structural and functional sequences such as inverted repeat sequences and sites which specifically bind plasmid encoded *tra* proteins.

The best studied *oriT* is that of the F plasmid and its close relatives. A schematic representation of the F *oriT* is presented in Figure 2.1. Both the nucleotide sequence and the structural features are conserved among the *oriT* regions of these plasmids. Deletion mutations defined the minimal *oriT* region in the F and R100 plasmids to between 280 and 330 bases (Fu *et al.*, 1991; Abo and Ohtsubo, 1995). Sequencing SDS-relaxed DNA demonstrated that *nic* lies 30 bases from the left end of *oriT* in both plasmids (where the bottom strand is nicked) (Thompson *et al.*, 1989; Inamoto *et al.*, 1991). In the region to the left of *nic*, there are two inverted repeat sequences (IR), to the right there are three additional IR, an AT rich sequence (Finlay *et al.*, 1986₂), one or two sites which bind the F and R100 TraY protein respectively (Lahue and Matson, 1990; Inamoto and Ohtsubo, 1990), one or two IHF binding sites in R100 and F respectively (Inamoto *et al.*, 1990; Tsai *et al.*, 1990) and two to four TraM binding sites identified in F, R1, R100 and pED208 (Schwabb *et al.*, 1991; Abo *et al.*, 1991; Di Laurenzio *et al.*, 1991; Di Laurenzio *et al.*, 1992). Two intrinsic bends centred within the right portion of the F *oriT* were identified by gel mobility studies (Tsai *et al.*, 1990). Among the F-like plasmids, the nucleotide sequence of the *oriT* region from the left end through the *nic* site, to the beginning of the TraY binding site shows almost 100% sequence identity. Sequence identity drops to 50-70% for the AT rich sequence and TraY binding sequence and to 40-82% for the TraM binding region (Finlay *et al.*, 1986₂; Frost *et al.*, 1994).

The nucleotide sequence of the *oriT* regions of IncN plasmids pCU1 and pKM101 are identical (Paterson and Iyer, 1992; Coupland *et al.*, 1987). The 325 bp minimal *oriT* sequence contains five IR, one at the left end, one in the middle and three in the right

Figure 2.1 Schematic representation of the F and P type *oriT* regions.

Upper figure. F *oriT*. Binding sites of TraY (dotted boxes), TraM (vertical hatched boxes) and IHF protein (shaded boxes) are shown above the line together with the AT rich sequence (slanted hatched box) and sites of localized bending (open boxes). Solid horizontal arrows and arrowheads below the line depict inverted repeat sequences. A vertical open arrow indicates the location of *nic* in the lower strand of DNA. The *traM* gene is indicated by a horizontal open arrow at far right side. The rest of the *tra* genes lie to the right of the *traM* gene.

Lower figure. P *oriT*. Binding sites of the Tra proteins are indicated by boxes above the line. The TraJ binding site (dotted box), TraI recognition site (solid box) and TraK binding sites at low concentrations (vertical hatched box) and at higher concentrations (open box). Inverted repeats are indicated below the line by solid horizontal arrows and arrowheads. A vertical open arrow indicates the location of *nic* in the upper strand of DNA. The divergent expression of the *traJ* and *traK* genes is depicted by open horizontal arrows on the line.

F oriT

P oriT

side. The sequence also contains two direct repeats and an AT rich sequence. The IncN *oriT* is the only *oriT* region studied to date that contains a large AT rich sequence comparable to that seen in the F plasmid. Prior to the studies described in this thesis, the location of *nic* had not yet been determined. To the immediate left of the *oriT* region there are thirteen, 11 bp direct repeats, however, deletion of these repeats from the pCU1 *oriT* region resulted in less than a 50% drop in transfer (Paterson and Iyer, 1992).

The *oriT* of the IncW plasmid R388 shares 34.5% sequence identity with that of the IncN plasmids pCU1 and pKM101 (Llosa *et al.* 1991). The *oriT* sequence contains four IR which align with four of the IR found in pCU1. The R388 *oriT* also contains a set of five, 11 bp direct repeats at the far left end of *oriT* however the sequence of these repeats is different from those observed in pCU1. As with pCU1, deletion of four of these direct repeats results in only a 3 fold reduction in transfer. Sequencing SDS-relaxed DNA demonstrated *nic* to be located in the middle of the *oriT*, nine bases to the left of the second IR (where nicking occurs on the lower strand) (de la Cruz, personal communication). Computer analysis of the *oriT* of R388 predicts an intrinsic bend in the sequence centred around the second IR.

The *oriT* of the IncP α plasmids RK2 and RP4 (which are considered to be the same plasmid) has been localized to a 368 bp *Xma*III-*Dde*I fragment (Guiney *et al.*, 1988). Like the IncF *oriT*, the IncP *oriT* has been studied extensively. The relevant features of the IncP *oriT* are depicted in Figure 2.1. The *oriT* of the IncP β plasmid R751 shares 71% sequence identity with that of RP4. In both plasmids *nic* is located 163 bases from the left side (in the upper strand) (Pansegrau *et al.*, 1988). There is a 19 bp

imperfect inverted repeat 8 bases to the left of *nic* in both plasmids showing identity at 14 of the 19 bases. The *traJ* product of RP4 binds to the right arm of the IR of the RP4 *oriT* but not to that of the R751 *oriT* presumably due to the sequence variation (Ziegelin *et al.*, 1989; Furste *et al.*, 1989). Point mutations within the 8 bp sequence between *nic* and the IR of RP4 reduce both transfer and relaxation significantly (Waters *et al.*, 1991). Seven of the eight bases are conserved between R751 and RP4.

The sequence to the right of *nic* in RP4 contains a curvature which is believed to result from short AT tracks that correspond to the periodicity of the DNA helix (Pansegrau *et al.*, 1990₁). This curvature is significantly enhanced by binding of the TraK protein to the right portion of the *oriT* including a second inverted repeat sequence 110 bases to the right of *nic* (Ziegelin *et al.*, 1992). Deletion of this region results in a 200 fold reduction in transfer (Furste *et al.*, 1989).

The *oriT* region of IncII plasmid R64 has been cloned on a 141 bp *RsaI* *HaeIII* fragment which contains two imperfect inverted repeats 8 and 17 bp in length. Deletion from the left side including the entire 8 bp IR and the left arm of the 17 bp IR leaving only 44 bp of the original 141 fragment, results in a drop in transfer to 8% of that of the 141 bp fragment. By contrast, deletion of 44 bp from the right, results in complete loss of transfer ability indicating that *nic* must lie within this 44 bp fragment. Sequencing SDS-relaxed DNA located *nic* within the 44 bp sequence, 8 nt to the right of the 17 bp IR (where *nic* is in the lower strand) (Furuya and Komano, 1991). The sequence surrounding *nic* shows identity at eight of nine bases to the sequence surrounding the *nic* site of IncP plasmid RP4. While the 17 bp IR in R64 is located the same distance from

nic as the 19 bp IR of RP4 is from its *nic*, the sequence of the two IR match at only four of the seventeen bases. Footprinting experiments demonstrated that purified NikA binds to the right side of the 17 bp IR of the R64 *oriT* in a manner similar to the binding of RP4 TraJ to its *oriT* (Furuya and Komano, 1995).

The *oriT* of *Thiobacillus ferrooxidans* mobilizable plasmid pTF-FC2 has been cloned on a 138 bp *HhaI* fragment. Structural organization of this *oriT* is similar to that of the IncP and IncI plasmids with a small 7 bp IR at the left end and a larger 18 bp imperfect IR in the middle. While the larger IR shares only limited sequence identity with that of the RP4 or R64-11, the sequence directly to the right matches at 6 of 8 and 7 of 8 bases respectively. (Rohrer and Rawlings, 1992). Also like the IncP plasmids, there is a natural bend in the pTF-FC2 *oriT* region 58 bp to the right of the large IR (Rohrer and Rawlings, 1993).

The *oriT* of IncQ plasmid R1162 has been cloned on overlapping 156 bp *HaeIII* and 153 bp *TaqI* fragments (Brasch and Meyer, 1986). The fact that the *TaqI* fragment is mobilized at a frequency 100 fold lower than the *HaeIII* fragment indicates the importance of a 10 bp imperfect inverted repeat, half of which is missing from the *TaqI* fragment. The *oriT* of a second IncQ plasmid RSF1010 has also been isolated on a 153 *TaqI* fragment which is mobilized at a frequency 20 fold lower than the parental plasmid (Derbyshire and Willetts, 1987). The nucleotide sequence of the *oriTs* of these two IncQ plasmids is identical over the 153 *TaqI* fragment at all but one base and these two plasmids are believed to be very similar if not identical to each other. Deletion of all but 38 bp of the R1162 *HaeIII* fragment results in less than a 10 fold drop in mobility

indicating that the critical sequences (including *nic*) must lie within this 38 bp sequence (Brasch and Meyer, 1987). Sequencing SDS-relaxed RSF1010 DNA placed *nic* within the 38 bp sequence, 8 bases to the left of the inverted repeat (where *nic* is in the upper strand) (Scherzinger *et al.*, 1992). The same site was identified in R1162 by recombination between mutated and wild type *oriTs* (Barlett *et al.*, 1990).

The T-DNA of *Agrobacterium* Ti and Ri plasmids is bordered by 25 bp direct repeats which are required *in cis* for transfer of the T-DNA in a manner analogous to the *oriT* region of conjugative plasmids. The nucleotide sequence of the 25 bp repeats is highly conserved among the 16 T-DNA borders studied to date with a minimum of 14 of 25 bases conserved (Pansegrau and Lanka, 1991). The site of nicking is between the third and forth bases from the left end of the direct repeat (wherein the bottom strand of DNA is nicked). This position is conserved in both borders of the same T-DNA and among the borders of different T-DNAs except in pTiA6, in which the cleavage site has been located 1 base to the right, between bases four and five (Albright *et al.*, 1989; Durrenberger *et al.*, 1989; Jasper *et al.*, 1994). As expected, deletion of the 6 bases containing the cleavage site abolishes transfer (Wang *et al.*, 1987). The sequence surrounding *nic* shows a high level of sequence identity to the *nic* site of RP4 and R64 (Waters *et al.*, 1991; Pansegrau and Lanka, 1991) however, the T-DNA repeats do not contain a comparable inverted repeat sequence. Despite the lack of identifiable structure within the right portion of the 25 bp direct repeat, the entire repeat is required for transfer as demonstrated by the loss of nicking accompanying the deletion of nine bases from the right end (Wang *et al.*, 1987).

2.3 INITIATION OF TRANSFER VIA SITE-SPECIFIC NICKING AT *oriT*

Nicking at *oriT* was first demonstrated with the small mobilizable plasmid ColEI. Supercoiled ColEI plasmid DNA isolated from *E. coli* was found to be associated with plasmid encoded proteins (Clewel and Helinski, 1969). Treatment of this DNA-protein complex with sodium dodecyl sulfate (SDS) resulted in relaxation of the DNA from the supercoiled form to the open circular or relaxed form which sediments more slowly on sucrose gradients. When run on a denaturing gel, the relaxed DNA demonstrated both single-stranded circular and single-stranded linear molecules complementary to each other, indicating that one strand had been nicked (Clewel and Helinski, 1970). The supercoiled DNA-protein complex was referred to as the relaxation complex. Genetic experiments identified the protein components as *mob* proteins and the location of nicking as the *oriT* (Lovett *et al*, 1974). Similar relaxation complexes have been identified for conjugative plasmids including the F plasmid (Kline and Helinski, 1971), the IncP plasmid RK2 (Guiney and Helinski, 1979), the IncX plasmid R6K (Kupersztoch-Portnoy *et al*, 1974₁) and the IncII plasmid R64 (Kupersztoch-Portnoy *et al*, 1974₂)

SDS-relaxed DNA can also be visualized by agarose gel electrophoresis where it appears as a band which migrates more slowly than expected for open circular DNA (DNA relaxed by topoisomerase I) (Thompson *et al*, 1989). Treatment of the relaxed DNA with proteinase K prior to electrophoresis results in a band which co-migrates with open circular DNA suggesting that the 5' end of the nicked strand is bound by protein. SDS-relaxation has been demonstrated by gel electrophoresis for the IncF plasmids F (Matson and Morton, 1991) and R100 (Inamoto *et al*, 1991), the IncP plasmid RP4

(Pansegrau *et al.*, 1988), the IncI1 plasmid R64 (Furuya and Komano, 1991) and IncQ plasmids RSF1010 and R1162 (Bhattacharjee and Meyer, 1991; Scherzinger *et al.*, 1992).

Demonstration that less than one half of the possible *oriT* containing substrate molecules can be isolated in the relaxed form upon treatment with SDS, even at high concentrations of enzyme relative to DNA substrate for prolonged incubation periods, suggests that the molecules actually exist in equilibrium between nicked and sealed (Pansegrau *et al.*, 1990, Matson and Morton, 1991). The addition of SDS is believed to denature the enzyme, trapping the molecules in the state they find themselves at the point of denaturation.

In vitro SDS-relaxation assays conducted on F *oriT*s demonstrated that purified F plasmid TraI protein, in the presence of Mg^{++} , is sufficient for relaxation of supercoiled DNA (Matson and Morton, 1991). These results conflict with results obtained *in vivo* with a lambda phage derivative carrying the F *oriT* (Everett and Willetts, 1980). Gel electrophoresis of denatured phage DNA demonstrated the presence of two small fragments in addition to the two larger bands representative of the full length single-stranded lambda derivatives. These two smaller bands were not seen in the absence of either *oriT* or the *tra* genes. Specifically, both *traY* and *traZ* (later determined to correspond to *tral*, Traxler and Minkley, 1987) were shown to be required for this nicking event. Other *in vivo* experiments demonstrate that both *tral* and *traY* functions are required for enhanced site-specific recombination involving the *oriT* of the F plasmid (Carter and Porter, 1991).

The *tral* gene of IncF plasmids encodes both the endonuclease activity responsible

for nicking at *oriT*, as well as a helicase activity which is believed to unwind the DNA strand to be transferred following nicking at *oriT* (Abdel-Monem *et al.*, 1983; Traxler and Minkley, 1988; Inamoto *et al.*, 1991). The F *tral* nucleotide sequence predicts a protein product of 192 kDa (Bradshaw *et al.*, 1990) corresponding to the 180 kDa size of helicase I protein initially identified on SDS-PAGE (Abdel-Monem and Hoffmann-Berling, 1976). Similar sizes were observed for the Tral protein of R100 (Inamoto *et al.*, 1991).

Preliminary studies indicate that the endonuclease activity of the Tral protein is located within the amino terminus of the protein while the helicase activity is determined by the carboxyl terminus. A derivative encoding a 96 kDa Tral protein from the amino terminus is fully functional with respect to nicking but is unable to unwind DNA (Traxler and Minkley, 1988). A 160 kDa protein containing the carboxyl terminus of the Tral protein is unable to nick at *oriT* but retains full unwinding activity (Benz and Muller, 1990). A smaller 94 kDa protein which cross reacts with antibodies prepared against the 180 kDa Tral protein has been observed and is believed to arise from an translational restart site within *tral*. This restart protein (Tral*) corresponds to the protein originally designated 2b and TraZ. It possesses neither nicking nor unwinding activity (Traxler and Minkley, 1988).

The nucleotide sequences of the *traY* genes of F (Fowler *et al.*, 1983), R100 (Inamoto *et al.*, 1988), ColB4-K94 (Finlay *et al.*, 1986₂) and R1-19 (Finlay *et al.*, 1986₁) have been determined. Except the F *traY*, all encode a 75 amino acid protein (MW 8.5-9 kDa). The *traY* gene of F encodes a 131 aa protein which is believed to have arisen from a duplication event (Inamoto *et al.*, 1988; Lahue and Matson, 1990). A

comparison of the amino acid sequences of these proteins demonstrates 96% amino acid identity between the TraY of ColB4 and R1 and 39-49% between these two proteins and the TraY of F and R100 (Finlay *et al.*, 1986₂). Sequence conservation is greatest in the carboxyl terminus. The TraY proteins demonstrate amino acid sequence identity, particularly in the carboxyl terminus, to the Arc and Mnt DNA binding proteins which act as repressors of phage 22 (Bowie and Sauer, 1990).

The TraY proteins of both F and R100 have been shown by gel retardation studies to bind to their respective *oriT*s as well as to the promoter regions of their *traY* genes (Lahue and Matson, 1990; Nelson *et al.*, 1993; Inamoto and Ohtsubo, 1990). The exact locations of TraY binding sites were determined by DNaseI protection studies. The TraY binding sites, *sbyA*, in the F and R100 *oriT*s are in approximately the same location in both plasmids, extending 33 and 34 bases respectively, approximately 64 and 59 bases from *nic* between two IHF binding sites and within the non-conserved portion of the AT-rich sequence (Inamoto and Ohtsubo, 1990; Nelson *et al.*, 1993). Binding to both DNA strands was observed.

A second TraY binding site in the F *oriT* (*sbyC*), not detected by footprinting, was suggested by the presence of two retardation complexes one of which was not observed when the *oriT* DNA carried a 22 bp deletion directly upstream of *sbyA*. This localized *sbyC* to the sequence shown by footprinting experiments to bind IHF (Luo *et al.*, 1994). The TraY protein has a lower binding affinity for *sbyC* than for *sbyA* (Luo *et al.*, 1994) which may explain its lack of detection with the footprinting experiments (Nelson *et al.*, 1993). In addition to the two strong bands observed on the retardation

gel attributed to binding at *sbyA* and *sbyC*, Lou *et al.*, (1994) observed two fainter bands, each migrating faster than one of the two more intense bands. They interpreted these faint bands as evidence that each of *sbyA* and *sbyC* binds two TraY molecules. The F TraY protein was shown to exist as a monomer in solution, suggesting that the two molecules must bind as monomers rather than a dimer (Nelson *et al.*, 1993).

In addition to the *sby* sites located within *oriT*, TraY binding sites have been identified within the promoter region of the *traY* gene of both F and R100. In the case of R100, two binding sites were observed with the one closest to the translational start of the gene (*sbyB*) having greater affinity for the TraY protein than the *sbyC* located 20 bases upstream (Inamoto and Ohtsubo, 1990). *sbyB* also shows greater homology to the *sbyA* located in the R100 *oriT*. While the two *oriT* F *sby* sites show 65% sequence identity to each other they show less identity to *sbyB* in the *traY* promoter (Luo *et al.*, 1994). There is no sequence similarity between the F and R100 *sby* sites, which agrees with the plasmid specificity exhibited by the *traY* function (Willettts and Maule, 1985). Demonstration that TraY binds to the promoter region of the *traY* suggests that it may regulate its own expression (Inamoto and Ohtsubo, 1990).

Lack of agreement in the role played by TraY as determined by *in vitro* SDS-relaxation assays in comparison to other indications of nicking and transfer, is demonstrated by mobilization of mutants containing deletions of *sbyA* is reduced by 10-300 fold despite the lack of apparent reduction in their ability to be relaxed by SDS *in vitro*. Fu *et al.* (1991) describe an F deletion derivative lacking one side of *oriT* including 16 nucleotides of *sbyA* which showed a 6 fold reduction in its ability to bind

TraY but no reduction in nicking activity in the SDS- relaxation assay. A point mutation within *sbyA*, shown to reduce TraY binding affinity by more than 200 fold, was still able to undergo relaxation *in vitro* (Luo *et al.*, 1994).

A recent study has resolved this conflicting data. Nelson *et al.*, (1995) demonstrated that while the Tral protein alone could relax double-stranded, supercoiled DNA under low salt conditions, NaCl concentrations of 75 mM or more abolished this activity unless both the TraY and the integrative host factor (IHF) protein were also present. TraY and IHF were also able to promote the relaxation of linear DNA by Tral. They suggest that the two proteins act to destabilize the *oriT*, providing a single stranded region capable of being recognized and nicked by the Tral protein. These results also explain earlier relaxation data obtained with plasmid R100. While initial SDS-relaxation assays conducted with purified Tral and TraY proteins from R100 demonstrated that nicking depended on both proteins (Inamoto *et al.*, 1991) subsequent experiments indicated that Tral could promote relaxation on its own, and that TraY together with the host encoded IHF protein only served to enhance the nicking activity of Tral (Inamoto *et al.*, 1994).

Interaction of the Tral, TraY and IHF proteins at *oriT* to form a relaxosome structure has been shown to occur in a defined order (Howard *et al.*, 1995). Tral was shown to bind to the *oriT* region only if both TraY and IHF were already bound to their respective binding sites. Binding of Tral extended the DNaseI protection footprint already produced by TraY and IHF leftwards to encompass the *nic* site. Electron microscopy has demonstrated that formation of this relaxosome complex is not

accompanied by shortening of the *oriT* region which would be expected if the DNA becomes wrapped around the proteins.

In the IncW plasmid R388, only the largest of the three R388 *mob* proteins, TwrC (MW 107 kDa) is required for *in vivo* intramolecular site-specific recombination between *oriTs* (described in Section 2.7) suggesting that it contains the nicking activity. This activity is maintained in a derivative carrying only the first 272 amino acids of TwrC (Llosa *et al.*, 1994₁). As discussed below, the TwrC protein of R388 possesses helicase activity similar to that associated with the Tral protein of the F plasmids (Grandosa *et al.*, 1994) and shows 33% amino acid sequence identity to the Tral protein of the F plasmid (Llosa *et al.*, 1994₂). R388 also encodes a small *mob* protein TrwA (MW 13.4 kDa) which is comparable in size to the IncF TraY protein. TrwA is acidic in nature and contains a leucine zipper motif suggesting that it forms a dimer in solution (Llosa *et al.*, 1994₂). Unlike the IncF TraY protein, mutations in *trwA* result in only a 200 fold reduction in transfer, indicating that it is unessential for transfer. Furthermore, recombination experiments suggest that it is not required for nicking (Llosa *et al.*, 1994₁).

In the IncP system, the relaxosome includes *traJ*, *tral* and *traH* gene products, however only TraJ and Tral are essential for nicking of supercoiled double-stranded DNA *in vitro*. These genes are transcribed divergently from the *oriT* on *tra* region 1 together with the *traG*, *traK*, *traL* and *traM* genes. *traG* and *traK*, while essential for mobilization, are not required for nicking at *oriT*, although the TraK protein has been shown to enhance the retrieval of molecules relaxed at *oriT* as described below. *traL* and *traM* are not essential for mobilization (Lessl *et al.*, 1993). For recent reviews of the

IncP system see Guiney and Lanka (1989) and Guiney (1993).

traJ encodes a 122 aa protein with an observed molecular weight of 11 kDa (Pansegrau *et al.*, 1988; Furste *et al.*, 1989). Footprinting experiments have shown that TraJ binds to the right arm of the 19 bp imperfect inverted repeat (IR) sequence 8 bp from *nic* (Pansegrau *et al.*, 1988). The protein binds to both strands of DNA with staggered footprints indicative of binding to only one side of the duplex DNA (Ziegelin *et al.*, 1989). Inability of the TraJ to bind to the left side of the inverted repeat is presumed to be the result of the 1 base pair mismatch in the IR. The TraJ protein of IncP α plasmid RP4 does not bind to the IR of IncP β plasmid R751 which exhibits four base differences within the IR. RP4 TraJ does bind non-specifically to the *oriT* of R751 at very high protein concentrations indicating the DNA binding character of the protein (Furste *et al.*, 1989).

The endonuclease activity responsible for nicking within the *oriT* of RP4 is encoded by the *tral* gene. The nucleotide sequence of *tral* predicts a protein with a molecular weight of 81.5 kDa which agrees well with the 82 kDa observed on SDS-PAGE. A smaller protein of 57 kDa cross-reacts with anti-TraJ, however such a protein has never been isolated (Pansegrau *et al.*, 1990₂). Gel retardation experiments suggest that TraJ does not bind to the DNA at *oriT* prior to relaxation by SDS. However, immunoblotting has demonstrated that the TraJ protein becomes covalently bound to the 5' end following nicking (Pansegrau *et al.*, 1990). It was therefore initially assumed that TraJ interacts with the *oriT*-TraJ complex prior to nicking (Pansegrau *et al.*, 1990₁). While both TraJ and TraI are required for nicking of double-stranded supercoiled DNA,

TraI can cleave single-stranded linear DNA alone (Pansegrau *et al.*, 1993₂). This observation has led to the suggestion that TraI may act to destabilize the DNA within the *oriT* providing sufficient single-stranded sequence for the TraI to nick one strand.

A study by Waters *et al.* (1991) demonstrated that point mutations within the 250 bp *oriT* which reduced mobilization more than 100 fold were localized to the 4 bp sequence directly to the left of *nic* and in the right arm of the IR. While all of these mutations affected SDS-relaxation *in vitro*, only the mutations in the IR affected binding of TraI protein as demonstrated by gel retardation assays. These results were taken to suggest that the mutations between the IR and *nic* may affect recognition of *nic* by the TraI protein. An additional mutant directly on the right side of *nic*, representing the 5' nucleotide of the nicked DNA, was shown to result in a 100 fold reduction in transfer and did not appear to affect SDS-relaxation significantly. Further, it was shown that the mutation did not alter the location of nicking and that TraI was still covalently bound to this altered nucleotide following relaxation. Hence it was concluded that only the region on the 3' side of *nic* is involved in relaxation.

As with the TraI endonuclease of the F plasmids, a truncated IncP TraI protein carrying only 417 amino acids from the amino terminus is fully functional at cleavage and ligation (Pansegrau *et al.*, 1993₂). A comparison of IncP TraI protein with analogous protein VirD2 from the Ti/Ri plasmids, NikB from the IncI plasmid R64-11 and MobA protein from pTF-FC2 identified three conserved motifs within the amino terminus (Pansegrau *et al.*, 1994). Mutations within each of the three motifs in TraI were shown to reduce transfer (Balzer, 1994).

The first motif $_{16}\text{FAELVKYITDE}_{25}$ contains the tyrosine (tyr) residue which binds to the 5' end of the DNA following nicking (see below). As expected, mutation of tyr22 to leucine (leu) eliminates binding as well as cleavage of the *nic* oligonucleotide. This motif and specifically tyr22 is considered the catalytic centre of the nickase (Balzer *et al.*, 1994; Pansegrau and Lanka, 1996).

The second motif $_{66}\text{DKTYHLLVSFRAGE}_{78}$ is believed to be involved in recognition of *nic* by Tral and non-covalent binding of the protein to the 3' end of the nicked strand allowing the nicked DNA within the relaxosome to remain supercoiled prior to SDS induced relaxation. Mutation of the serine (ser) residue in this motif to alanine (ala) reduces binding of Tral but not cleavage activity. The finding that this mutation also caused an increase in topoisomerase activity lead to the suggestion that motif II is involved in recognition of the sequence between *nic* and the IR and that mutations in motif II result in the occasional release of the 3' end and subsequent formation of partially relaxed topoisomers. The lack of specificity of the Tral protein between the IncP α and IncP β plasmids is consistent with the high degree of sequence conservation on the 3' side of *nic* in RP4 and R751 (Pansegrau *et al.*, 1988). Conservation of this motif in IncI NikB and Ti VirD2 proteins but not in the IncF or IncW nickases is also consistent with higher DNA sequence identity in the *nic* regions of the former plasmids (Lanka and Wilkins, 1995).

The third motif in the Tral protein $_{105}\text{QRVSAVHHDTDNLHHHIAINKI}_{124}$ is the only motif which is partially conserved in the IncF and IncW nickase proteins (Cabezen *et al.*, 1994). Only one of the three mutations made within this motif of the IncP Tral

affect binding, however all three affect cleavage and it was therefore proposed that this motif is also part of the catalytic centre. It was suggested that the third histidine residue (His116) interacts with motif I by proton abstraction of the tyr hydroxyl group to increase the efficiency of the nucleophilic attack by this residue (Pansegrau and Lanka, 1996).

The TraH protein of the IncP plasmid is a 22 kDa out-of frame translational restart protein encoded by the 5' end of the *tral* gene (Furste *et al.* 1989). While the TraH protein increases SDS-relaxation only 3% *in vitro*, formation of the relaxosome prior to SDS treatment was not observable via electron microscopy in the absence of TraH. This suggests that while TraH is required for stable relaxosome formation it is dispensable for actual nicking (Pansegrau *et al.* 1990). These same experiments indicated that while Mg^{++} is required for SDS induced relaxation, it is not necessary for assembly of the relaxosome proteins prior to relaxation by SDS. Since gel retardation assays demonstrate that TraH does not bind to DNA within the *oriT*, it is assumed that this stabilization occurs via protein-protein interactions. This role may be exclusive to *in vitro* conditions as inactivation of the *traH* gene (while maintaining the native amino acid sequence of the Tral protein) has no effect on transfer (Cole *et al.* 1993).

A fourth protein which has been shown to interact with the RP4 *oriT*, but which is not considered part of the relaxosome, is the *traK* product. *traK* encodes a 133 amino acid protein with a predicted molecular weight of 14,585 Da corresponding to the observed MW of 17 kDa (Ziegelin *et al.* 1992). It has been shown to behave as a tetramer in solution. *traK* is essential for mobilization and while it is not necessary for

nicking, it has been shown to enhance the yield of relaxed molecules (Furste *et al.*, 1989; Waters *et al.*, 1991). TraK has been shown to suppress the topoisomerase activity observed with the motif II ser74 mutants of TraI by acting as a DNA chaperon to reduce superhelical stress on the relaxosome (Pansegrau and Lanka, 1996).

Mutations in RP4 *traK* can not be complemented by the corresponding function of the IncP β plasmid R751 (Furste *et al.*, 1989). Footprinting experiments have demonstrated that the TraK protein binds to a 200 bp sequence to the right of *nic* and that helical spacing of the protected sequences suggests that the protein is associated with only one side of the helix. This region contains a second IR which is also present in the IncP β , however its sequence is not conserved. The TraK binding site also encompasses an intrinsic bend present in both RP4 and R751 *oriT* sequences. In RP4 this curvature has been attributed to the presence of short AT tracks, however, similar tracts are not observed in R751. Electron microscopy demonstrated that TraK binding results in the formation of a sharp bend in the DNA as well as significant reduction in length of the DNA fragment suggesting that the DNA is wrapped around the TraK tetramer forming a nucleosome type structure (Ziegelin *et al.*, 1992).

Preliminary studies of the *mob* region of mobilizable plasmid pTF-FC2 from *Thiobacillus ferrooxidans* demonstrate limited (25-30%) amino acid sequence conservation between the MobA and MobB proteins of this plasmid and the TraI and TraJ proteins of the IncP plasmid, suggesting that these two proteins may be involved in nicking at *oriT*. pTF-FC2 also encodes a protein with limited sequence identity to the TraK protein of RP4 (Rohrer and Rawlings, 1992).

SDS-relaxation experiments with IncII plasmid R64 deletion derivatives have implicated an operon containing two genes, *nikA* and *nikB*, in the nicking of R64 *oriT* (Furuya *et al.*, 1991). The nucleotide sequence of the two ORFs predict molecular weights of 12.6 and 104 kDa respectively. A 14 kDa protein corresponding to NikA was demonstrated in maxicell preparations run on SDS-PAGE, however a band corresponding to NikB could not be seen. Whether both genes are essential for nicking has not been determined as a derivative carrying a functional *nikB* without *nikA* has not been tested. A mutant carrying a truncated *nikB* encoding 495 of 899 amino acids representing the amino terminus of the protein lost mobility but retained full nicking activity, suggesting that like the IncP, IncF and Vir systems, the NikA has two functional domains, with nicking activity in the amino terminus. As with the P and F systems, gel retardation studies have demonstrated that the smaller NikA protein binds to the *oriT*. NikA binds to the right arm of the 17 bp imperfect IR located 8 bases to the left of *nic* in a manner similar to the IncP system. Binding occurs to both strands of DNA and causes bending as seen by altered mobility on gel electrophoresis (Furuya and Komano, 1995).

By all indications, mobilizable IncQ plasmids RSF1010 and R1162 are essentially the same plasmid. The complete sequence of RSF1010 has been determined (Schloz *et al.*, 1989). These plasmids encode three genes required for their mobilization by a conjugative plasmid which are transcribed divergently from the *oriT*. *mobB* overlaps *mobA* however, the two protein products are translated from different reading frames (Schloz *et al.*, 1989). The molecular weights of the Mob proteins predicted from the nucleotide sequences are 77.9 kDa for MobA, 15.1 kDa for MobB and 10.9 kDa for

MobC (Scherzinger *et al.*, 1992).

As demonstrated by a deletion derivative carrying only the first 243 of the 708 amino acids of the RSF1010 MobA protein, the endonuclease activity responsible for nicking at *oriT* is encoded by the amino terminus of this protein (Scherzinger *et al.*, 1992). The carboxyl terminus of the MobA protein has been shown to encode primase activity (Scholz *et al.*, 1989). As with the IncP and Ti endonucleases, MobA can nick single-stranded DNA without the aid of other Mob proteins (Bhattacharjee and Meyer, 1991; Scherzinger *et al.*, 1992) however MobB does stimulate cleavage of ssDNA (Scherzinger *et al.*, 1992). Similarly, *in vitro* nicking of double-stranded DNA at *oriT* requires both MobA and MobC and the presence of MobB in the reaction increases nicking from 55% to 85% of available molecules. While both linear and supercoiled DNA are nicked *in vitro*, the former required excess MobC. These results are consistent with demonstration that the MobA protein forms a stable complex with single-stranded DNA containing the R1162 *oriT* strand which is nicked. Footprinting with P1 nuclease demonstrated a footprint corresponding to the sequence between the IR and *nic*. Since deletion of either the IR or the sequence between the IR and *nic* results in reduced binding, it was concluded that the intervening sequence was sufficient for binding while the IR stabilized the complex (Bhattacharjee and Meyer, 1993).

A study by Zhang and Meyer (1995) demonstrates that the region between *nic* and the IR is sensitive to permanganate, suggesting that this region undergoes localized melting to produce a single-stranded region which is subsequently nicked. Permanganate sensitivity is reduced by mutations in both *mobA* and *mobC* as well as by two point

mutations within the sensitive region. These findings are similar to those in which the sequence between *nic* and the IR in the RP4 *oriT* are believed to be recognized by the TraI protein following destabilization by the TraJ protein.

Nicking within the 25 bp direct repeat sequence of the T-DNA borders was first demonstrated by RNA protection studies (Albright *et al*, 1987). Radiolabelled RNA spanning the border sequence was hybridized to denatured DNA containing the T-DNA region. When the resultant hybrid was digested with RNase and run on a denaturing polyacrylamide gel, two bands representing shorter than unit length RNA molecules could be seen on the gel only when the radiolabelled RNA fragment was complementary to the bottom strand. When the radiolabelled RNA probe was complementary to the top strand, the bands seen on the gel were primarily full length molecules. Involvement of the *vir* genes in nicking was demonstrated by the lack of smaller RNA bands in the absence of *vir* induction by acetosyringone (AS). Single-stranded nicking within the borders has also been demonstrated by S1 nuclease cleavage of double-stranded DNA containing T-DNA sequence from *Agrobacterium* induced with AS. Subsequent digestion with restriction enzymes produced fragments consistent with cleavage within the border sequence by S1 nuclease opposite a single-stranded nick (Wang *et al*, 1987; Stachel *et al*, 1987). *vir* dependant double-stranded cleavage has also been demonstrated by Veluthambi *et al* (1987) as shorter than expected double-stranded restriction fragments consistent with cleavage within the border contained on the restriction fragment. While these results indicate that both double-stranded and single stranded nicking occurs within the T-DNA borders, a site-specific single-stranded nick within the bottom strand is

predominant (Yanofsky *et al.*, 1986; Albright *et al.*, 1987).

Sequential nicking within the T-DNA borders is believed to result in the production of a linear, ss DNA molecule referred to as the T strand. The T strand was first visualized as a small DNA band from AS induced *Agrobacterium* genomic DNA blotted on a nitrocellulose filter which was homologous to T-DNA sequences. The single-stranded nature of the DNA comprising this band was demonstrated by its ability to be transferred to nitrocellulose paper without prior denaturation and its sensitivity to S1 nuclease. Based on a comparison of the intensity of the T strand DNA with that of the double-stranded T-DNA region in the Ti plasmid, it was estimated that each cell contained only one T strand molecule (Stachel *et al.*, 1986; Stachel *et al.*, 1987).

Generation of T strands containing adjacent T-DNA sequences was demonstrated in a strain carrying Ti plasmid pTiA6 which contains four border sequences (Stachel *et al.*, 1987). Single-stranded, linear DNA fragments homologous to the T-DNA region the size expected for each of the possible six T strands from such an arrangement were identified, indicating that each border is able to act as either an initiation or termination site. Except for the full length T strand which would be produced from initiation at the first repeat and termination at the fourth, all of the possible T strands were produced in approximately equal quantities, indicating that each border sequence is nicked with equal frequency. The isolation of T strands resulting from nicking at non-adjacent borders indicates that all borders are not nicked at same the time and suggests that the individual T strands are produced independently rather than as the result of subsequent nicking of the full length T strand.

Involvement of the *virD* operon in virulence and specifically in nicking within the 25 bp direct repeats of the T-DNA borders and subsequent production of T strand DNA *in vivo* is well documented. The nucleotide sequence of the *virD* operons from *Agrobacterium* plasmids pTiA6, pTiC58 and pRi4b has been determined and found to contain four open reading frames (ORFs) (Yanofsky *et al*, 1986; Porter *et al*, 1987; Jaysawal *et al*, 1987; Wang *et al*, 1990; Hirayama *et al*, 1988). While *virD1*, *virD2* and *virD4* are required for virulence (Porter *et al*, 1987; Vogel and Das, 1992), only *virD1* and *virD2* are required for *in vivo* nicking (Yanofsky *et al*, 1986; Veluthambi *et al*, 1987; Jayswal *et al*, 1987) and T strand formation (Stachel *et al*, 1987). A molecular weight of 16.2 kDa predicted for the protein encoded by *virD1* matches the 16 kDa size observed on SDS-PAGE, however, the 56-57 kDa VirD2 protein observed on SDS-PAGE is slightly larger than the 47.4 kDa size predicted by the *virD2* nucleotide sequence (Yanofsky *et al*, 1986; Jaysawal *et al*, 1987; Ghai and Das, 1989; Jasper *et al*, 1994).

Like the Tral protein of the IncP plasmid RK2, purified VirD2 protein alone is unable to nick supercoiled DNA containing T-DNA, however, it is able to nick a 17 mer ss oligo containing the right border sequence of pTiA6 T-DNA (Pansegrau *et al*, 1993). Jasper *et al* (1994) confirmed that purified VirD2 could cleave ss T-DNA alone, but demonstrated that it could not nick ds linear T-DNA alone or in the presence of VirD1. Subsequent experiments demonstrated that purified VirD1 and VirD2 could nick supercoiled DNA containing the right border in a reaction which required Mg^{++} (Scheifele *et al*, 1995). As seen with the P and F conjugative plasmids, only 35-40% of

the molecules were relaxed. This percentage doubled for plasmids containing both right and left borders. Also like the IncP and IncF systems, only the amino terminus of the VirD2 protein is required for nicking (Yanofsky *et al.*, 1986; Wang *et al.*, 1990).

2.4 OTHER PROTEINS THAT BIND TO THE *oriT* REGION

An additional *tra* function encoded by the F-like plasmids which is required for conjugation but not for mating pair formation is the *traM* gene. While it is clear that *traM* is required for transfer, it is equally clear that it is not involved in nicking. The finding that TraM is associated with the inner membrane has led to the suggestion that it acts as the signal that promotes initiation of transfer or that it anchors the DNA to the cell wall during transfer (Atchmann *et al.*, 1979; DiLaurenzio *et al.*, 1991; Abo *et al.*, 1991) .

The nucleotide sequence is known for the *traM* gene of the F plasmid (Finlay *et al.*, 1986₂), R1/R1-19 (Finlay *et al.*, 1986₁; Schwabb *et al.*, 1991), R100 (Abo *et al.*, 1991), pED208 (DiLaurenzio, 1991) and ColB4-K98 (Finlay *et al.*, 1986₂). The gene encodes a 127 amino acid protein with an estimated molecular weight of 14.5 kDa, which is slightly larger than the observed MW of 10.9 kDa for pED208 but very close to the 14 kDa observed for TraM of R1.

Each of the TraM proteins of F, pED208, R100 and R1 have been shown to bind to their respective *oriT* regions over an extended sequence in the non-conserved portion of the *oriT* to the right of the TraY binding sites up to and including the promoter of the *traM* gene. Footprinting has demonstrated three TraM binding sites (*sbmA-C*) between

the TraY binding sites and the start of the *traM* gene in the F *oriT* (Di Lorenzo *et al.*, 1992). Four distinct TraM binding sites (*sbmA-D*) have been identified in the R100 *oriT* with a 15 bp consensus sequence (Abo *et al.*, 1991). Similar experiments identified up to 14 separate binding sites in the pED208 *oriT* corresponding to 14 of the 16 *HinfI* restriction sites repeated in this sequence (DiLaurenzio *et al.*, 1991). The TraM binding sites in the R1 *oriT* have only been divided into two large regions of 65 and 110 bases, the size of which is suggestive of multiple binding sites as well (Schwabb *et al.*, 1991). Binding of the TraM protein to the promoter of the *traM* gene suggests autoregulation which has been demonstrated for the *traM* gene of R1 (Schwabb *et al.*, 1993).

2.5 COVALENT ATTACHMENT OF PROTEIN TO THE 5' END OF NICKED DNA

In the SDS-relaxation assay, the observation that relaxed DNA migrates more slowly than expected for open circular DNA unless pretreated with proteinase K suggested that the 5' end is bound by protein. This assumption was supported by the demonstration that the 5' end of nicked DNA is protected from T7 exonuclease digestion (5' → 3' exonuclease activity) or to labelling with ³²P via polynucleotide kinase following extensive treatment with proteinase K and alkaline phosphatase. Covalent binding of a protein to the 5' end of the nicked DNA strand has been demonstrated for ColE1 (Lovett and Helinski, 1975; Guiney and Helinski, 1975), the IncF plasmids F (Matson and Morton, 1991) and R100 (Inamoto *et al.*, 1991), the IncP plasmid RP4 (Pansegrau *et al.*, 1988), the IncII plasmid R64 (Furuya and Komano, 1991) and IncQ plasmids RSF1010

and R1162 (Bhattacharjee and Meyer, 1991; Scherzinger *et al*, 1992). In all cases, the 3' end of the nicked DNA remains unaltered, the evidence being its ability to be digested by *E. coli* exonucleases I and III (3' → 5' exonuclease activity) or to be extended by DNA polymerase I.

Similar experiments by Durrenberger *et al* (1989) demonstrated that the 5' end of the lower strand of T-DNA was protected from digestion by lambda exonuclease (5' → 3' exonuclease) but not T4 DNA 3' → 5' exonuclease activity. Sensitivity of T strand DNA to exonuclease III and T4 polymerase, both 3' → 5' exonucleases, indicates that it has a free 3' end (Stachel *et al*, 1986).

Further evidence of a DNA-protein complex in the T-DNA system of *Agrobacterium* similar to that observed in conjugative plasmid systems was provided by the finding that both ds T-DNA and ss T strand DNA were found primarily in the phenol interface during phenol extract procedures (Young and Nester, 1988; Ward and Barnes, 1988; Herrera-Estrella *et al*, 1988; Howard *et al*, 1989). Treatment of the DNA with either proteinase K or pronase resulted in its extraction in the aqueous layer. Similar proof was obtained by Herrera-Estrella *et al* (1988) by using the fact that under low salt conditions proteins but not DNA can be transferred to nitrocellulose paper and that under high salt conditions, protein association prevents the transfer of DNA. Hybridization of a T-DNA probe was not detected with DNA from *Agrobacterium* cells treated with pronase and blotted onto nitrocellulose paper under low salt concentrations. Omission of the pronase treatment resulted in strong hybridization signals, confirming the presence of ss DNA on the filter. Double-stranded T-DNA nicked by AS induced proteins,

digested with a restriction enzyme, then run on a denaturing electrophoresis gel and blotted under high salt concentrations, demonstrated hybridization to probes complementary to the 5' side of the lower strand of the T-DNA only if the DNA was treated with proteinase K prior to electrophoresis.

Since purified Tral protein of the F plasmid has been shown to cause relaxation *in vitro* in the absence of any other proteins, it is assumed to be the protein which remains covalently bound to the 5' end (Matson *et al.*, 1993). In studies with the IncQ plasmids, the extent of the gel retardation of the nicked DNA suggests that the largest of the three *mob* products (MobA protein) remains covalently bound to the 5' end of nicked RSF1010 and R1162 DNA (Scherzinger *et al.*, 1992). In the IncP system, immunoblots of relaxed RP4 *oriT* DNA demonstrate cross reactivity with anti-Tral (Pansegrau *et al.*, 1990₂). Immunoblotting, immunoprecipitation and gel retardation have demonstrated that the protein bound to the 5' end of nicked T-DNA is the VirD2 protein (Durrenberger *et al.*, 1989; Howard *et al.*, 1989; Filichkin and Gelvin, 1993; Jasper *et al.*, 1994).

Covalent attachment of protein to the 5' end of the nicked DNA is believed to result from a nucleophilic attack of the phosphodiester bond of the DNA by the aromatic hydroxyl group of a tyrosine residue in the nicking protein. This type of linkage has been implicated in the rolling circle replication of bacteriophage ϕ X174 by protein A (van Mansfield *et al.*, 1986). The ϕ X174 protein A has been shown to not only nick the DNA at *oriV* but also to ligate it following the unwinding of the nicked strand. It has been proposed that the energy conserved in the tyrosyl-phosphate ester bond is used for

the subsequent sealing reaction. Protein A contains two tyr residues separated by three amino acid residues which are proposed to be used for alternate nicks at *oriV*. After a first nick at *oriV* and covalent binding of protein A to the 5' end via one tyr residue, the nicked strand is unwound from the 5' end and extended via host proteins from the 3' OH end. After a full length strand has been unwound, the protein nicks again at the reconstituted *oriV* and the second tyr residue is covalently linked to the new 5' end. The resultant 3' OH undergoes a transesterification with the first phosphotyrosyl linkage forming a new phosphodiester linkage and recircularizing the unwound strand (Hanai and Wang, 1993). A similar nicking and sealing process has been suggested to occur during conjugal transfer.

Evidence of the involvement of a tyrosine in nicking and resealing at the *P oriT* was provided by mutation of the tyr22 of the TraI protein of RK2 to phe (phenylalanine). This mutation reduced nicking activity to 10%. N terminal sequencing of the peptide remaining covalently attached to the 5' end following digestion of nicked DNA with proteinase K demonstrated a tetrapeptide extending from amino acid 21-24 of TraI with no signal identified for the second residue corresponding to the tyr22 residue presumably due to its covalent linkage to the DNA (Pansegrau *et al.*, 1993).

A tyrosine residue has also been implicated in the attachment of the VirD2 protein to the 5' end of the nicked ds T-DNA and free ss T strand DNA of the *Agrobacterium* T-DNA system. Of the 10 tyr residues in the Ti plasmid pTiA6 VirD2 protein, seven which are conserved in VirD2 proteins from other Ti and Ri plasmids were mutated to phe residues. Only mutation of residues tyr29 and tyr121 had a significant effect on

nicking. Of these two, tyr29 could not be replaced by any other residue while replacement of tyr121 with tryptophan (trp) resulted in partial activity. From these results, Vogel and Das (1992₁) concluded that tyr29 most likely facilitates binding of the VirD2 to the nicked T-DNA. Migration of the DNA-protein complex following digestion with different proteases is consistent with attachment of the protein via tyr29 (Panscgrau *et al*, 1993).

2.6 UNWINDING THE DNA BY HELICASE

Helicase I is encoded by the carboxyl terminus of the F plasmid *tral* gene (Abdel-Monem *et al*, 1983). It is a DNA dependant ATPase which requires the presence of Mg^{++} or Ca^{++} ions (Abdel-Monem and Hoffmann-Berling, 1976). Two type A ATP binding sites have been identified in the *tral* sequence (Bradshaw *et al*, 1990) and it is proposed that the energy released from the hydrolysis of ATP is used to translocate the helicase along the DNA in the 5' -> 3' direction (Kuhn *et al*, 1978; Kuhn *et al*, 1979). Studies show that the ATPase activity continues after circular dsDNA is completely denatured but ceases upon complete unwinding of linear DNA. These results suggest that in the case of the circular DNA substrate, the helicase continues to translocate along the single-stranded molecule with concomitant ATPase activity, while it stops at the end of the linear molecule where it remains bound rather than attaching to another linear dsDNA substrate molecule (Lahue and Matson, 1988). This is consistent with the finding that helicase I does not dissociate quickly from DNA (the half life of the DNA-helicase complex being more than 10 min) (Abdel-Monem *et al*, 1976; Khun *et al*, 1979; Fukuda

and Ohtsuba, 1995).

Helicase I exists as a monomer in salt concentrations greater than 0.15 M NaCl but tends to form aggregates at lower ionic conditions. ATPase activity is greatest under conditions where aggregates form (Abdel-Monem and Hoffmann-Berling, 1976). Unlike other helicases, the protein:DNA ratio which allows maximum activity depends on the number of binding sites rather than the length of the DNA to be unwound, indicating that helicase I acts in a processive manner (Kuhn *et al.*, 1978; Kuhn *et al.*, 1979). Early studies suggested that 70-90 molecules of helicase I are required to unwind one DNA molecule (Abdel-Monem *et al.*, 1976; Kuhn *et al.*, 1978; Kuhn *et al.*, 1979; Benz and Muller, 1990). Other studies indicate that as few as 1-10 helicase molecules are able to unwind DNA (Lahue and Watson, 1988; Wessel *et al.*, 1990; Dash *et al.*, 1992). Possible explanations for the discrepancies between these observations include differences in enzyme activity resulting from variations in extraction procedures (Lahue and Matson, 1988) or difference in ionic condition of the helicase storage buffer causing aggregation of the protein molecules (Dash *et al.*, 1992).

Single-stranded DNA is the preferred substrate molecule for helicase I (Abdel Monem and Hoffmann-Berling, 1976). Helicase I binds ssDNA cooperatively as demonstrated by electron microscopy studies in which ssDNA molecules are either completely coated with helicase I -protein or associated with very little protein (Benz and Muller, 1990). Only dsDNA containing a single-stranded region of at least 200 bp is able to bind helicase I (Kuhn *et al.*, 1979). Electron micrographs of the fork produced by helicase I activity on dsDNA in the presence of single-stranded binding protein (SSB)

demonstrates the presence of a short region behind the fork unbound by SSB protein equivalent to the distance which would be occupied by at least four helicase molecules (Benz and Muller, 1990). These results provide evidence that helicase I moves along the DNA molecule as a tract of protein molecules. The truncated 160 kDa carboxyl terminus protein has been shown to retain both the helicase and ATPase activity of the full protein but does not bind ssDNA cooperatively. On dsDNA substrates, this protein demonstrates a smaller gap behind the fork which is not bound by SSB protein, suggesting that even a single helicase is capable of unwinding dsDNA (Benz and Muller, 1990).

The *TwrC* protein of IncW plasmid R388 has also been shown to encode a ssDNA dependant ATPase which requires Mg^{++} (Grandosa *et al.*, 1994). The nucleotide sequence of *twrC* predicts a protein size of 107,688 Da which corresponds well to the 105 kDa protein observed on SDS-PAGE of maxicell extracts (Llosa *et al.*, 1994₂). There is 33% identity between the amino acid sequence of *TwrC* and *Tral* of the F plasmid (Llosa *et al.*, 1994₂). Like *Tral*, *TwrC* is also a helicase which unwinds DNA in the 5' -> 3' direction relative to the ssDNA (Grandoso *et al.*, 1994). Although *TwrC* has been shown to exist as a dimer at high salt concentrations (>500 mM NaCl), it is believed to form aggregates at lower concentrations. Due to the fact that the ATPase activity is maximal at low NaCl concentrations (<100 mM) and not detected at higher concentrations (>250 mM NaCl), it is assumed that this protein, like helicase I encoded by the F plasmid, acts as a multimer (Grandoso *et al.*, 1984). The main difference observed between *TwrC* and *Tral* of F and R100 is that the former has very low processivity and it has been suggested that it may require additional proteins to increase

the processivity of the unwinding reaction (Grandoso *et al.*, 1994)

Sequence identity has been observed between the TraA protein of the F⁺ conjugative plasmids and the Tral protein of F and TrwC protein of R388, suggesting that it too may encode a helicase activity (Alt-Morbe *et al.*, 1996; Farrand *et al.*, 1996)

2.7 CIRCULARIZATION OF THE TRANSFERRED STRAND

Site-specific recombination involving the *oriT* region was first described in Tra deletion mutants of the F plasmid from Hfr strains in which one end of the deletion mapped to the *oriT* region (Guyer *et al.*, 1977; Horowitz and Deonier, 1985). Based on observations of site-specific recombination between two directly repeated phage origins (Dotto *et al.*, 1982), Brasch and Meyer (1987) demonstrated that a similar recombination event could occur between two directly repeated copies of the 38 bp R1162 *oriT*. Mobilization of plasmids carrying two directly repeated copies of the R1162 *oriT* separated by the gene for kanamycin resistance resulted in kanamycin sensitive transconjugants carrying plasmids with a single functional *oriT*. This *oriT* was presumably a hybrid of the two separate albeit identical *oriTs* on the original plasmid formed by site-specific recombination between the two *oriTs*. Demonstration that the recombination event was dependant on both *trans*-acting *mob* functions and *oriT* indicated that this site-specific recombination event could be used to study events at *oriT* during conjugal transfer.

In the R1162 double-*oriT* plasmid construct described above, recombinants were identified by screening transconjugants for loss of the intervening antibiotic marker

following mobilization. Meyer (1989) also tested ssDNA phage double-*oriT* constructs. Isolation of recombinant phage by transfection demonstrated that while the recombination event was dependant on *tra* functions, it could occur in the absence of actual transfer. Transfer independent recombination was also observed for double-*oriT* plasmid constructs involving the *oriT* of IncW plasmid R388 by screening for loss of the intervening antibiotic marker after extended growth without antibiotic selection (Llosa *et al*, 1994).

Site-specific recombination between border sequences of T-DNA has also been observed and shown to be dependant on induction of the *vir* genes, functional analogues of the *tra* genes. Koukoilikova-Nicola *et al* (1985) demonstrated that T-DNA containing a copy of *E. coli* plasmid pBR322 could be isolated and transformed into *E. coli* resulting in circular molecules containing a single border sequence. Similarly, T-DNA containing lambda *cos* sites could be packaged *in vitro*, resulting in cosmids of the exact size expected for joining at the borders. In a variation of Meyer's approach, a recombinant plasmid was constructed containing two T-DNA borders separated by an ampicillin gene which was cloned into a kanamycin gene such that the promoter and coding regions of the kanamycin gene were separated. Recombination between the two borders resulted in loss of ampicillin resistance and acquisition of kanamycin resistance due to replacement of the Km promoter directly upstream from its coding region. Km^r, Ap^s isolates were only obtained when both borders were present in directly repeated orientation and only upon induction of the *vir* genes by co-cultivation of the cells with protoplasts (Machida *et al*, 1986).

In vivo site-specific recombination implies that in addition to nicking at *oriT*,

resealing of the nick to form a circular molecule must also occur. Evidence that the proteins involved in nicking at *oriT* also act to seal the nick was provided by demonstrating that recombination could occur *in vitro* between directly repeated R1162 *oriT*s on single stranded DNA by the action of an R1162 *mobA* fusion protein (Bhattacharjee and Meyer, 1993). Pansegrau *et al* (1993,) also demonstrated that the TraI protein of RP4 could transfer the 5' end of one oligonucleotide containing *oriT* sequences to the 3' OH of a second previously nicked oligonucleotide. Similar activity has been demonstrated with the VirD2 protein and oligonucleotides containing the T-DNA border sequence (Pansegrau *et al*, 1993,). Demonstration of resealing ability is consistent with SDS-relaxation experiments in which only a portion of the plasmid molecules are isolated in the relaxed form.

Transfer of larger-than-unit-length molecules (Erickson and Meyer, 1993) has provided evidence that conjugal transfer resembles rolling circle replication of plasmids and phage that replicate through a single-stranded intermediate where the 3' end of the nicked replication origin is used to prime replacement strand synthesis (Gruss and Ehrlich, 1989). As indicated earlier, nicking and sealing at the *oriV* of the single-stranded phage ϕ X174 is carried out by two closely spaced tyr residues. In a recent study in which an oligonucleotide carrying the IncP *nic* site was immobilized and reacted with TraI protein, the resultant oligo-protein complex was shown to be unable to cleave an incoming oligo containing a complete *nic* sequence, indicating that a single molecule of TraI is insufficient to carry out both the cleavage and sealing reaction. It was proposed that separate TraI molecules are required for the first and second cleavage events suggesting that TraI acts in a dimeric form (Pansegrau and Lanka, 1996).

The rolling circle replication model of conjugation suggests that in the recombination event observed with double-*oriT* constructs, transfer is initiated at one *oriT* and terminated at the second, resulting in the transfer of the DNA between *oriT1* and *oriT2* and the deletion of the DNA between *oriT2* and *oriT1*. By creating double-*oriT* constructs in which one of the *oriT* carries a mutation, Meyer and his colleagues (Brasch and Meyer, 1987; Kim and Meyer, 1989, Meyer, 1989) demonstrated that mutations affect recombination differently depending on whether they are present in the *oriT* nicked first or second, suggesting that the events at *oriT* could be separated into initiation (presumably the initial nicking) and termination (presumably the second nicking and ligation) events. Two point mutations surrounding *nic* of R1162 affected initiation of transfer while deletion of the inverted repeat 8 bp from *nic* reduced termination. Correction of the one base mismatch within the inverted repeat resulted in an increase in the recombination event, suggesting that the inverted repeat acts to trap the MobA protein providing an opportunity for it to reseal the scission (Battacharjee *et al.*, 1992).

Intramolecular recombination has also been used to identify initiation and termination domains in the *F oriT* (Gao *et al.*, 1994). Deletion of the TraM binding sites reduced initiation 5 fold while further deletion of the TraY binding site resulted in a further 10 fold reduction. Two point mutations 4 and 8 nt to the right of *nic* also reduced initiation significantly. Termination was found to be unaffected by deletion of the TraM or TraY binding sites but was reduced significantly by the two point mutations as well as by deletions up to 30 bases to the right of *nic* which include one of two GCAAAAAC direct repeats and corresponds to the region of sequence conservation

among the F like plasmid *oriT*s.

Site-specific recombination between one wild type and one mutated *oriT* carried on the same plasmid has also been used to locate the exact site of *nic* in plasmid R1162. Because the resultant hybrid *oriT* contains one half of each of the two parental *oriT*s (with *nic* representing the junction point) only mutations located to one side of *nic* will be found in the hybrid *oriT*. An R1162 double-*oriT* construct in which one *oriT* contained two point mutations separated by a single base produced hybrids with only one of the point mutations indicating that *nic* must lie between the two mutations (Brasch and Meyer, 1987). A second double-*oriT* construct in which one of the *oriT*s carried a mutation of the intervening base was used to determine whether *nic* lay to the right or left of the intervening base (Barlett *et al.*, 1990).

In the recombination experiments involving R1162 double-*oriT* plasmids in which one of the *oriT*s contained a mutation, isolation of identical hybrid *oriT*s from a given construct indicates that the location of *nic* remains constant (Barlett *et al.*, 1990). However, when the same *oriT* sequences were carried on a single-stranded phage, not all of the hybrid *oriT*s from one double-*oriT* construct contained the same sequence indicating indeterminacy in the *nic* site. When the mutation was in the *oriT* nicked first, the *nic* site remained the same in all of the recombinant phage, however, when the mutation was in the *oriT* nicked second, the *nic* site became indeterminant, producing recombinants with wildtype *oriT*s as well as recombinants with mutant *oriT* sequences rather than recombinants all with the same sequence. In the phage system, the DNA lost from the hybrid was a short sequence, the deletion of which allowed for the expression

of the *lacZ* gene. Recombinants were identified by the production of blue plaques on X-gal. That none of the blue plaques tested contained a hybrid with a deletion or insertion suggested that both *oriT*s were nicked at the same site despite the fact that hybrids containing both original *oriT* sequences were obtained. A second derivative containing a +4 frameshift in the intervening DNA was constructed to detect 1 base deletions which could arise from different cleavage sites within the first and second *oriT*. In this assay, none of the 40,000 plaques isolated were blue, indicating that indeterminacy resulted from nicking the same but wrong site in both *oriT*s and that the location of *nic* depended on the site nicked in the first *oriT*.

Differences in the results obtained with ssDNA phage as compared to those obtained with similar plasmid double-*oriT* constructs were also observed in experiments to identify initiation and termination functions in R1162. The point mutations which resulted in reduced recombination in the plasmid system had no effect on the phage system while the IR deletion mutation had the same effect in both systems (Barlett *et al.*, 1990). The fact that mutations in sequences necessary for initiation do not appear to have an effect on recombination in the ssDNA phage system while mutations in termination functions do is consistent with previous results which suggested that the phage system imitates termination functions and possibly a ssDNA intermediated formed during conjugation (Meyer, 1989).

2.8 DIRECTION AND MODE OF TRANSFER THROUGH THE MEMBRANE

The direction of transfer of the nicked and unwound DNA through the membrane has been determined for the F plasmid (Rupp and Ihler, 1970; Ohki and Tomizawa, 1970), the IncP plasmid RP4 (Al-Doori *et al.*, 1982), the IncII plasmid Collb-P9 (Howland and Wilkins, 1988), IncQ plasmid R1162 (Kim and Meyer, 1989) and mobilizable plasmid ColE1 (Boyd and Sherratt, 1986). In all cases, transfer is unidirectional with *tra* or *mob* genes (or the bulk of these genes in the case of RP4) entering last. While transfer in the 5' → 3' direction has only been clearly demonstrated for the F plasmid (Rupp and Ihler, 1970; Ohki and Tomizawa, 1970) the general consensus that this is the case for all plasmids studied to date is supported by the combination of directional studies with those identifying the specific strand nicked. Transfer of the nicked strand in the 5'-3' direction is consistent with rolling circle replication where the unwound strand is replaced by DNA synthesis from the 3' end of the nicked DNA.

It is generally accepted that a single strand of DNA is transferred from the donor to the recipient (Vielmetter *et al.*, 1968; Lanka and Wilkins, 1995). Whether this strand is linear or circular has not yet been determined. One possible scenario suggests that the 5' end of the nicked strand, which is bound by the nicking enzyme, actually remains in the donor cell and the subsequent DNA is passed into the recipient as a loop. After a complete strand of DNA is unwound, a second nick at the reconstituted *oriT* followed by ligation releases the last portion of a circular molecule into the recipient (Willetts and Wilkins, 1984; Wilkins and Lanka, 1993; Lanka and Wilkins, 1995). This model allows

for the recircularization of the nicked strand by the endonucleolytic enzyme without necessitating the transfer of this enzyme, an event which has not been successfully demonstrated (Rees and Wilkins, 1990).

The route of transfer through the membrane is unknown. While the belief that transfer occurs through the retracted pili is supported by the demonstration of transfer between donor and recipient separated by a membrane filter (Harrington and Rogerson, 1990), this possibility is not fully accepted. The possibility that other *tra* proteins act as a pore through the membrane has also been proposed (Jalajakumari and Manning, 1989).

Most of the conjugative systems studied to date encode a protein with hydrophobic domains suggestive of association with the membrane, which is required for transfer but not for nicking at *oriT* or for stable mating pair formation. The sequence of such proteins has been determined for the *traD* gene of the F plasmid (Jalajakumari *et al.*, 1989) and R100 (Yoshioka *et al.*, 1990), the *traG* gene of IncP plasmids RP4 and R751 (Ziegelin *et al.*, 1991), the IncW plasmid R388 *trwB* gene (Llosa *et al.*, 1994₂), the *virD4* gene of *Agrobacterium* plasmids pTiA6 (Jayaswal *et al.*, 1987) and pRiA4 (Hirayama *et al.*, 1988), the *traG* genes of nopaline- and octopine-type pTiC58 plasmids (Alt-Morbe *et al.*, 1996; Farrand *et al.*, 1996) and the *mobB* gene of mobilizable plasmid CloDF13 (Nijkamp *et al.*, 1986; Lanka and Wilkins, 1995). In all cases there is some degree of sequence conservation including conservation of nucleotide binding sites. The F, W CloDF13 proteins contain a well conserved A type nucleotide binding site with consensus sequence: hydrophobic stretch (2 of 5 residues) (G/A)xx(G)xGKS/T (where the residues in brackets are less conserved). This sequence is less conserved in the IncP and Ti TraG

proteins and VirD4 proteins. All of the proteins contain a B type nucleotide binding site with consensus sequence: hydrophobic stretch (3 of 5 residues) D(E/D) (Gorbalenya and Koonin, 1990; Lessl *et al.*, 1992; Cabezon *et al.*, 1994).

In the F plasmid, *traD* acts prior to SDS sensitive extension of the *pil* and concurrently with the inactivation of transfer by nalidixic acid inhibition of gyrase (Panicker and Minkley, 1985). Demonstration of non-specific binding of TraD to nucleic acids (Panicker and Minkley, 1992), together with its potential NTP binding domains and clear demonstration that the TraD protein is associated with the inner membrane (Achtman *et al.*, 1979; Panicker and Minkley, 1992) suggests that it may anchor the relaxosome to the membrane and provide the energy required for transfer of DNA through the membrane via ATP hydrolysis (Panicker and Minkley, 1985).

Unlike the mobilizable plasmid CloDF13, which encodes its own TraD like protein, mobilization of ColE1 requires the F *traD* for transfer (Willettts, 1980). The F-*traD* is the only function other than those involved in stable mating pair formation, required for the transfer of ColE1. Similar mobilization experiments demonstrate that RSF1010 requires the *traG* function of the conjugative plasmid RP4 or the *twrB* of IncW R388 in addition to the mating pair functions. Surprisingly, while *traG* can complement mutations in *twrB* in the mobilization of RSF1010, it can not complement mutations in *twrB* in the transfer of R388 (Lessl *et al.*, 1993; Cabezon *et al.*, 1994).

2.9 COMPLEMENTARY STRAND SYNTHESIS

While it is generally accepted that complementary strand synthesis in the donor occurs by extension of the 3' end of the nicked strand in a manner comparable to rolling circle replication (Willetts and Wilkins, 1984; Waters and Guiney, 1993; Lanka and Wilkins, 1993), several other mechanisms of complementary strand synthesis which could be used in either donor or recipient have been postulated. Some conjugative plasmids encode their own primase protein which is transferred to the recipient and could be used to prime synthesis on the incoming ssDNA. In the absence of a plasmid encoded primase, it is assumed that host encoded proteins are used to prime complementary synthesis on the incoming strand. The fact that mutations in plasmid encoded primases do not severely reduce transfer ability would suggest that host encoded proteins may be used in either case (Merryweather *et al.*, 1986₁; Merryweather *et al.*, 1986₂).

Two mechanisms of DNA priming by host encoded primase have been described in studies with ssDNA phage. In complementary strand synthesis of the ssDNA phage ϕ X174, a specific sequence (AAGCGG) is recognized by the PriA protein (previously the π protein) followed by the assembly of five additional host proteins including RNA polymerase and primase. This complex, known as the primosome, translocates downstream to generate RNA primers used for DNA initiation. Plasmid ColE1 contains a primosome assembly site (PAS) with 5 of 6 matches to the ϕ X174 PAS site, which is used in priming the synthesis of the lagging strand during plasmid replication and which may also be used to synthesize the complementary strand of the transferred DNA in the recipient (Zavitz and Marians, 1991). The F plasmid also contains a ϕ X174 PAS

sequence with 5 of 6 bases conserved however the priming capabilities of this site have not been tested (Lanka and Wilkins, 1995). A study by Nomura *et al* (1991) identified eleven different sites from several different plasmids able to initiate complementary strand synthesis on M13 ssDNA (referred to as *ssi* sites). Seven of these sites use the ϕ X174 mechanism, however in each case, the PAS is much larger than the ϕ X174 consensus sequence. One of the sites identified in this study used a similar mechanism, but the primosome consisted of fewer host proteins.

Host encoded primase can also initiate complementary strand synthesis without the aid of the other components of the primosome from a hairpin loop structure as occurs in the replication of the single-stranded RNA phage G4 (Sims *et al*, 1980). Of the 11 *ssi* sites identified by Nomura *et al* (1991), three functioned in this manner. One of these sites is located on the leading strand of the F plasmid (Lanka and Wilkins, 1995).

The *oriT* of mobilizable plasmid pTF-FC2 has been shown to contain an *ssi* site which increases the replication efficiency of an M13 mutant impaired in the priming of DNA synthesis. While it was demonstrated that the pTF-FC2 encoded primase was not involved in this process, no indication was found as to which of the host encoded priming mechanisms was used (Rohrer and Rawlings, 1993).

The two best studied plasmid encoded primases are the products of the *sog* gene of the IncI plasmids and the *traC* (previously *pri*) of the IncP plasmids. While not essential for transfer, the *traC* gene has been shown to increase the transfer frequency of the IncP plasmids to some bacterial strains (Lanka and Barth, 1981). Derepression of transfer functions in the IncI plasmids also results in an increase in primase activity

suggesting a relationship between primase activity and conjugation (Lanka and Barth, 1981). Both mobilizable plasmids RSF1010 and pTF-FC2 encode a primase activity which is used in replication, however only in the former case has the primase also been implicated in mobilization (Scholz *et al.*, 1989; Rohrer and Rawlings, 1993). Both the IncP and IncI primases are able to initiate DNA synthesis on single-stranded phage DNA (Lanka *et al.*, 1979; Lanka and Barth, 1981; Yakoboson *et al.*, 1990) and complement *E. coli* temperature sensitive primase (*dnaG*) mutants (Wilkins, 1975; Lanka *et al.*, 1979; Lanka *et al.*, 1984), suggesting that they act to initiate complementary strand synthesis in the recipient following transfer. Both proteins contain a E-G-Y-A-T-A motif also present in the α protein encoded by the phage P4 (Strack *et al.*, 1992). Mutation of the E residue to Q eliminates primase activity suggesting that this conserved motif is part of the active centre of the protein. Notably, the IncF, IncW and IncN plasmids do not encode primase activity (Lanka and Barth, 1981).

The *traC* gene of the IncP α plasmid RP4 encodes two antigenically related proteins 118 and 80 kDa in size, both of which show primase activity (Lanka *et al.*, 1984). The IncP β plasmids encode three primase proteins, all of which are antigenically related to each other. DNA hybridization experiments demonstrate that all three are encoded by the same DNA fragment and therefore the two smaller proteins are assumed to be translational restart proteins of the largest. Both the 135 and 70 kDa proteins have been shown to have primase activity while the largest 185 kDa protein has not yet been tested. While the primases of the IncP β plasmids are antigenically related to those of the IncP α plasmids, the former have 10 fold greater activity than the latter (Lanka *et al.*,

1985).

The IncI *sog* gene also encodes two antigenically related proteins with molecular weights of 240 and 180 kDa. The smaller is a translational restart protein however, primase activity is only associated with the larger protein (Wilkins *et al.*, 1981). Neither the IncI or the IncP plasmid encoded primases show antigenic identity to each other or to the *E. coli dnaG* product.

Demonstration that recipients with mutations in the chromosomal *dnaG* gene recovered following conjugal transfer of plasmid Collb-P9 despite repression of transcription of the incoming *sog* gene in the recipient by rifampicin, suggested that the primase was transferred from the donor to the recipient (Chatfield *et al.*, 1982; Chatfield and Wilkins, 1984). These results are supported by physical demonstration of both IncI primase proteins in recipient minicells or recipient cells following lysis of donor cells with T6 phage following mating (Rees and Wilkins, 1989). Similar techniques demonstrated the transfer of the IncP primase, however in this case, only the larger protein was observed in the recipient (Rees and Wilkins, 1990).

An interesting study by Merryweather *et al.* (1986,) demonstrated that while the presence of the *sog* gene in the donor could not complement mutations in a co resident *traC* gene, a functional *sog* gene in the recipient could enhance transfer frequencies of *traC* mutants from the donor. These results suggest that while the *sog* primase can initiate complementary strand synthesis of the IncP plasmid in the recipient, it can not transfer from a donor carrying only the IncP plasmid, suggesting that it is unable to interact with the export machinery.

3. METHODS AND MATERIALS

3.1 BACTERIAL STRAINS, BACTERIOPHAGES, PLASMIDS AND PRIMERS

See Tables 3.1, 3.2, 3.3 and 3.4.

3.2 GROWTH AND MAINTENANCE OF BACTERIAL CULTURES

Cultures were grown in L broth (LB) (1.0% tryptone, 1.0% NaCl, 0.5% yeast extract). Solid media (LBA) contained 1.5% agar. Antibiotics were added from filter sterilized stock solutions to solid media in the following concentrations: chloramphenicol, Cm, 30 $\mu\text{g/ml}$; spectinomycin, Sp, 50 $\mu\text{g/ml}$; ampicillin, Ap, 30 $\mu\text{g/ml}$; kanamycin, Km, 50 $\mu\text{g/ml}$; tetracycline, Tc, 10 $\mu\text{g/ml}$; nalidixic acid, Nal, 50 $\mu\text{g/ml}$; rifampicin, Rif, 100 $\mu\text{g/ml}$. For Cm, Km, Ap and Sp, one half the concentration was used in liquid media. Tc was added to liquid media in the same concentration. Neither Nal or Rif was used in liquid media. Antibiotic selection for plasmid markers was maintained at all times, unless otherwise stated, to ensure plasmid stability.

Stock cultures were maintained at -80°C in L broth containing 20% glycerol.

3.3 BACTERIAL MATINGS

a) Large scale filter matings (for determining transfer frequencies)

Overnight cultures of donor cells were diluted 0.1 ml into 5 ml LB with the appropriate antibiotics and grown for 3 hr (approximately 5×10^8 cells per ml). 1 ml of over night recipient culture and 1 ml of 3 hour donor culture were filtered together

Table 3.1 *E. coli* strains used in this study

Strain	Genotype/Phenotype	Reference
BL21(DE3)	F' <i>ompT r_{pm} p_{lac}</i> λDE3	Studier <i>et al.</i> , 1990
BW11397	Δ <i>lac</i> -169 Δ <i>phoA</i> 8 <i>creB</i> 510 <i>supF</i> 58 <i>supE</i> 44 <i>hsdR</i> 514 <i>galK</i> 2 <i>galT</i> 22 <i>trpR</i> 55 <i>metB</i> 1 <i>tonA</i>	Wilmes-Riesenberg and Wanner, 1992
CC118	<i>araD</i> 139, Δ(<i>ara.leu</i>)7697, Δ <i>lacX</i> 74, <i>phoA</i> Δ20, <i>galE</i> , <i>galK</i> , <i>thi</i> , <i>rpsE</i> , <i>rpoB</i> , <i>argE_{am}</i> , <i>recA</i> 1	Manoil and Beckwith, 1985
C600Nal	F' <i>thi</i> 1 <i>thr</i> 1 <i>leuB</i> 6 <i>lacY</i> 1 <i>tonA</i> 21 <i>supE</i> 44 λ Nal ^r	this laboratory
DH5α	<i>endA</i> 1 <i>hsdR</i> 17 <i>supE</i> 44 <i>thi</i> 1 <i>recA</i> 1 <i>gyrA</i> <i>relA</i> 1 φ80 <i>dlacZ</i> M15 Δ (<i>lacZYA-argF</i>)U169	Jessee, 1986
DH5αF'	DH5α containing F' plasmid	Liss, 1987
HB101	F' <i>hsdS</i> 20 <i>recA</i> 13 <i>ara</i> 14 <i>proA</i> 2 <i>lacY</i> 1 <i>galK</i> 2 <i>rpsL</i> 20 <i>xy</i> 15 <i>met</i> 1 <i>supE</i> 44 λ	Sambrook, <i>et al.</i> , 1989
HB101Rif	spontaneous Rif ^r mutant of HB101	this laboratory
PN1	Tn <i>phoA</i> in chromosome of CC118	this lab.
S17-1	<i>thi</i> <i>pro</i> <i>hsdR</i> <i>hsdM</i> ^r <i>recA</i> RP4 in chromosome (Tp ^r Sm ^r)	Simon <i>et al.</i> , 1983

Table 3.2 Bacteriophages used in this study

Bacteriophage		Reference
<hr/>		
M13K07	Km ^r	Vieira and Messing, 1987
λ ::Tn <i>phoA</i>	<i>cll1857 b221 Pam3</i> <i>rex</i> ::Tn <i>phoA</i> (Km ^r)	Manoil and Beckwith, 1985

Table 3.3 Plasmids used in this study

Plasmid	Phenotype	Comments or pCU1 co-ordinates	Reference or Source
pCU1	Sp ^r Ap ^r Sm ^r Tra ⁺		Konarska-Kozłowska and Iyer, 1981
pCU101	Cm ^r Tra ⁺	27.5-9.74	Thatte and Iyer, 1983
pCU56	Cm ^r Tra ⁺ Mob ⁺ ^b	6.7-11.5	Thatte <i>et al.</i> , 1985
pCU57	Cm ^r Tra ⁺ Mob ⁺	0-9.74	"
pCU57Δ14	Cm ^r Tra ⁺ Mob ⁺	3.58-9.74	Paterson and Iyer, 1992
pSP34	Cm ^r Mob ⁺	3.58-9.74	this study
pET-traHI	Ap ^r Mob ⁻	3.58-6.78	"
pET-traJ	Ap ^r Mob	6.78-7.97	"
pET-traK	Ap ^r Mob ⁻	7.97-8.8	"
pET23d	Ap ^r		Novogen
pKan/oriT	Ap ^r Km ^r	pUC1318 + Tn5 Km + P <i>oriT</i>	Hengen and Iyer, 1992
pMK2004	Ap ^r Km ^r Tc ^r Mob ⁻		Kahn <i>et al.</i> , 1979
pSP6Km	Km ^r	pUC18 + Tn5 Km	this study
pUC18	Ap ^r		Norlander <i>et al.</i> , 1983
pUC128	Ap ^r		Keen <i>et al.</i> , 1988
pSP274	Tc ^r Tra ⁺	pMUR275::Tn10	Paterson and Iyer, 1992
pSP545	Tc ^r Tra ⁺	pMUR545::Tn10	"
pMO2	Ap ^r Mob ⁺	pMUR274 <i>oriT</i>	this study
pMO2Km	Ap ^r Km ^r Mob ⁺	"	this study
pMO2S	Ap ^r Mob ⁺	"	"
pMO2R	Ap ^r Mob	"	"
pKM101 <i>traH1143</i>	Ap ^r Km ^r Tra ⁺ Mob ⁻	N type <i>tra</i>	Winans and Walker, 1985.
pKM101 <i>traI1136</i>	Ap ^r Km ^r Tra ⁺ Mob ⁻	N type <i>tra</i>	"
pKM101 <i>traJ35</i>	Ap ^r Km ^r Tra ⁺ Mob ⁻	N type <i>tra</i>	"
pKM101 <i>traK1217</i>	Ap ^r Km ^r Tra ⁺ Mob ⁻	N type <i>tra</i>	"

Plasmid	Antibiotic marker	Nucleotide co-ordinates of pCU1 <i>oriT</i>	Reference or Source
pBP1	Ap Km	1-325	Paterson and Iyer, 1992
pBP11	Ap Km	1-283	"
pSP6	Km	118-288	this study
pSP7	Km	1-117	"
pSP8	Km	1-266	"
pSP9	Km	1-266	"
pSP10	Km	1-243	"
pSP12	Ap	32-325	Paterson and Iyer, 1992
pSP19	Ap	68-325	"
pSP24	Km	1-288	"
pSP25	Ap Km	1-325	Paterson and Iyer, 1992
pSP27	Ap Tc	1-325	this study
pSP28	Ap	1-325	"
pSP28Km	Ap Km	1-325	"
pSP29	Km	1-157;187-288	this study
nic3	Km	1-288	this study
nic11	Km	1-288	"
nic14	Km	1-288	"
nic15	Km	1-288	"
nic16	Km	1-288	"
nic45	Km	1-288	"
nic59	Km	1-288	"
nic109	Km	1-288	"
nic121	Km	1-288	"
nic120	Km	1-288	"
nic136	Km	100-288	"
nic137	Km	1-288	"
nic139	Km	1-288	"
nic144	Km	112-288	"
nic188	Km	1-288	"
nic211	Km	1-288	"
nic224	Km	1-288	"
nic238	Km	1-288	"
pNKN9A	Ap, Km	double <i>oriT</i>	this study
pNKN9B	Ap, Km	double <i>oriT</i>	"
pNKN10A	Ap, Km	double <i>oriT</i>	"
pNKN10B	Ap, Km	double <i>oriT</i>	"
pNKN29A	Ap, Km	double <i>oriT</i>	"
pNKN29B	Ap, Km	double <i>oriT</i>	"
pNKN3A	Ap, Km	double <i>oriT</i>	"
pNKN3B	Ap, Km	double <i>oriT</i>	"
pNKN238A	Ap, Km	double <i>oriT</i>	"
pNKN238B	Ap, Km	double <i>oriT</i>	"
pNKN109A	Ap, Km	double <i>oriT</i>	"

* nucleotides 1-325 corresponds to 9047-8722 Kb on the pCU1 map.

Table 3.4 Primers used in this study.

Primer	Sequence 5' -> 3'	Location
ORIT1	CCTCACATTGTGCATTTC	pCU1 <i>oriT</i>
ORIT2	GGGGATCCAACTCATTATAAAACA	pCU1 <i>oriT</i>
ORIT3	GGGGATCCGTATTACCAAAGTAATAAA	pCU1 <i>oriT</i>
ORIT4	ATCCCTTAACCTGCTATACACCTAACGCA	pCU1 <i>oriT</i>
ORIT4op	CTGCGTTAGGTGTATAGCAGGTTAAGGGAT	pCU1 <i>oriT</i>
MOB1	CAGTCTGTGCGCCAGATTAA	pCU1 <i>traHI</i>
MOB2	CACAACGCGAGTGACGTTCT	pCU1 <i>traHI</i>
MOB3	AGGAGGATTAGCGTTTC	pCU1 <i>traJ</i>
MOB4	GGCAATAACCGAAACGC	pCU1 <i>traJ</i>
MOB5	GAGTTCACTAAGCGCTC	pCU1 <i>traJ</i>
MOB6	GGTTCTTAAAGAATTTC	pCU1 <i>traJ</i>
MUR1	TCCGCTTCATTCATGACT	pMO2 <i>oriT</i>
MUR2	TGCGCTTCGGTCACTCT	pMO2 <i>oriT</i>
MUR3	GTGTCCTTTTCTTGTCTC	pMO2 <i>oriT</i>
pUC forward	CCCAGTCACGACGTTGT	pUC plasmids
pUC reverse	TAACAATTTTCACACAGG	pUC plasmids
Tn5-30	AACGGGAAAGGTTCCGTT	Tn5
PHOA	TGCAGTAATATCGCCCTG	Tn <i>phoA</i>

through a 47 mm, 0.45 μ m pore membrane filter. The filters were incubated on LBA media without antibiotics for 2 hours at 37°C then washed off into 1 ml saline (0.85 % NaCl). Dilutions were prepared in saline and 0.1 ml of selected dilutions were plated on LBA media containing antibiotic selection for both the recipient and the plasmid to be transferred. Where donor cells contained more than one plasmid (*ie* mobilization experiments), selection for each plasmid marker was applied separately so that the transfer frequency of individual plasmids could be calculated. Plates were incubated at 37°C for 24 hours.

Transconjugant counts/ml of mating mixture were calculated as the number of colony forming units $\times 10 \times$ the dilution factor. Where ever possible, plates containing 30-300 colonies were used preferentially over those containing more than 300 or less than 30 colonies.

Viable counts of donor cultures were obtained by the procedure outlined above except that 2 ml of donor culture was filtered in place of 1 ml of each of donor and recipient. The diluted cultures were plated on LBA containing the antibiotic(s) used to select for the plasmid(s) in the donor cell. Donor count per ml of culture was calculated as the number of colony forming units $\times 10 \times 0.5 \times$ the dilution factor.

To ensure that neither donor nor recipient could grow on the media used to select for transconjugants, 0.1 ml of the recipient culture or donor saline suspension was plated on each selective media used to grow the transconjugants of the corresponding mating.

Transfer frequencies were determined by dividing the transconjugant count by the donor count to give the number of transconjugants receiving a particular marker per

donor cell carrying that marker. In mobilization experiments, in order to compensate for the variation of transfer ability of the various helper plasmids, mobilization frequencies are given in terms of the number of transconjugants receiving the marker of the mobilizable plasmid per transconjugant receiving the marker of the conjugative plasmid.

b) Mini-filter matings (preparative matings)

50 μ l of overnight recipient culture was mixed with 50 μ l of 3 hour donor culture on a 27 mm, 0.45 μ m pore size filter disc on an LBA plate and incubated for 2 hours. The filter was rinsed in 1 ml saline and 0.1 ml of selected dilutions plated on LBA media containing antibiotics to select for recipients receiving the conjugative or mobilizable plasmid. Plates were incubated for 24-48 hours. Controls were run on both the donor and recipient cultures.

c) Spot matings (for screening point mutants)

Donor cultures were grown overnight on an LBA plate containing antibiotics and transferred by toothpick to 20 μ l of overnight LB recipient culture spotted onto the bottom of a sterile petri dish. Following mixing, 5 μ l of the mixture was spotted onto a sterile 45 μ m pore filter on an LBA plate. A maximum of 15 spot matings were conducted on one 47 mm diameter filter disc. The filter was incubated for 2 hours at 37°C and then transferred to an LBA plate containing antibiotics selective for both the recipient and the plasmid marker. In biparental matings where the donor contained both the *oriT* plasmid and the conjugative helper plasmid, two filters were prepared from each

mating mixture and transferred to separate plates containing antibiotics for the recipient and either the helper or *oriT* plasmid. In triparental matings, equal volumes of overnight recipient and 3 hour helper cultures were mixed and 20 μ l spotted onto the petri dish for mixture with culture containing the *oriT* plasmid.

For more quantitative screening by spot mating, filters were incubated on LBA at 37°C for 6 hours, following which a toothpick was used to remove a tip full of mating growth. This was resuspended in 50 μ l of saline spotted onto a sterile petri dish bottom. Following mixing with a pipette tip, serial dilutions were prepared by mixing 10 μ l in 100 μ l of saline in a microtitre plate. 5 μ l of each dilution was spotted onto LBA plates containing antibiotics selective for the recipient and either plasmid. The degree of growth within the spot at each of the dilutions was compared from the plates selecting for the transfer of the conjugative plasmid and that selective for the *oriT* plasmid.

3.4. PLASMID DNA EXTRACTION

a) Large scale plasmid DNA extraction by CsCl gradient

5 ml of overnight culture was diluted into 800 ml LB plus antibiotics and incubated overnight at 37°C with shaking. The cells were pelleted in a Sorval GSA rotor spun at 5000 RPM for 5 min. The pellets were resuspended in a total of 30 ml sucrose buffer (25% sucrose in 0.05 M Tris, pH 8.0) and transferred to two SS-34 centrifuge tubes. To each tube, 0.5 ml of lysozyme (10 mg/ml) and 3 ml 0.25 M EDTA was added. The mixture was left on ice for 5 min and 4 ml of Triton X-100 lytic solution (Triton X-100, 2 %, Tris, 50 mM, EDTA, 60 mM, pH 8.0) added. If necessary, cell

lysis was promoted by incubation at 50°C for 5-10 min. Once lysis occurred, the solution was centrifuged at 15,000 RPM for 30-60 min at 4°C. The supernatant was collected and 7 ml of a 40% PEG 6000 solution in 2 M NaCl was added. Precipitation was allowed to proceed overnight at 4°C followed by centrifugation at 6,000 RPM for 5 min at 4°C. The two pellets were resuspended in a total of 3.2 ml TE buffer (Tris HCl, 0.01 M, EDTA, 0.001 M pH 8.0) to which 4.000 g CsCl and 0.5 ml ethidium bromide (10 mg/ml) was added. The DNA was centrifuged overnight at 15°C in a VTi80 rotor at 60,000 RPM. The plasmid DNA band was removed from the tube by side puncture using a needle and syringe (Sambrooke *et al.*, 1989). The ethidium bromide was extracted with CsCl saturated isopropanol and the CsCl removed by dialysis in TE buffer. The DNA was precipitated with one fifth volume of 10 M ammonium acetate and two volumes of absolute ethanol at -20°C overnight followed by a 1 hour spin at 8,000 RPM in an HB4 rotor in the Sorval centrifuge. The DNA pellet was dried at 50°C under vacuum, resuspended in 25-100 μ l of dH₂O and stored at 4°C.

b) Wizard miniprep extraction

WizardTM (Promega) miniprep plasmid extraction procedure was conducted following manufacturers instructions. Briefly, 3 mls overnight culture was pelleted in a bench top centrifuge and the cells resuspended in 200 μ l cell resuspension solution (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 100 μ g/ml RNaseA). Following transfer to a 1.5 ml microfuge tube, 200 μ l cell lysis solution (0.2 M NaOH, 1% SDS) was added followed by 200 μ l neutralization solution (1.32 M K acetate, pH 4.8). This lysate was

centrifuged for 5 min. The resultant supernatant was mixed with 800 μ l resin and subsequently passed through a Wizard column attached to a 3 ml syringe. The column was washed with 2 ml of buffered ethanol (90 mM NaCl, 9 mM Tris-HCL, pH 7.5, 2.25 M EDTA, 55% ethanol) and spun for 20 sec to drain. 50 μ l dH₂O was added to the column and the DNA solution collected by a 20 sec spin.

b) Boiling 'mini-prep' plasmid DNA extraction

The method used is a variation of the boiling method described by Holmes and Quigley (1981). A 4 ml overnight culture was pelleted in a bench top centrifuge and resuspended in 400 μ l STET buffer (sucrose, 8%, Triton-X-100, 5%, EDTA, 50 mM, Tris, 50 mM, pH 8.0). The suspension was transferred to 1.5 ml Eppendorf tube containing 20 μ l lysozyme (10 mg/ml) and placed in a boiling bath for 40 sec. The tube was then spun in the microcentrifuge for 5 min at 4°C and the gelatinous pellet removed with a sterile toothpick. One fifth volume 10 M ammonium acetate and two volumes ethanol was added to the supernatant and precipitated by spinning 30 min at 4°C. The pellet was washed with 70% ethanol and dried at 50°C under vacuum. The dried pellet was resuspended in 20 μ l of dH₂O.

3.5. RESTRICTION DIGEST OF DNA

Restriction endonuclease digests were carried out in 10 or 20 μ l reactions using conditions recommended by the supplier. Digests were incubated for 1-2 hours at temperature recommended by the supplier except *Bst*YI digests which were incubated for

a maximum of 30 min. Digests typically contained 1 μ l of CsCl purified plasmid DNA or 2 μ l of mini-prep DNA and 1 μ l of enzyme (4-20 units/ μ l) except *Bst*YI digests in which no more than 2 units of enzyme was used. Restriction enzymes were from New England BioLabs.

3.6 AGAROSE GEL ELECTROPHORESIS OF RESTRICTION DIGESTS

A Bio-Rad horizontal, submerged mini-gel apparatus was used to analyze restriction digests. Bio-Rad agarose (0.7-2% w/v) was dissolved in Tris-acetate electrophoresis buffer, TAE (80 mM Tris, 40 mM sodium acetate, 4 mM EDTA, pH 8.3) by heating in a microwave oven. The gels were left to set for 30-60 min.

For mini-prep digests, 2 μ l of stop solution (30% glycerol, 0.05 % bromophenol blue, 5 mM EDTA) was added to a 10 μ l digest and loaded onto the gel; for CsCl DNA digests, 1-2 μ l of a 10 μ l digest was mixed with 5 μ l of stop solution and loaded onto the gel.

Gels were run at 60-100 volts for 1-3 hours. Gels were stained in ethidium bromide solution (0.4 mg/ml) 5-10 min and destained in dH₂O for 10-60 min. Gels were photographed using a Polaroid Land camera and Polaroid type 57 or 55 film.

BRL 1 Kb DNA ladder (0.4 μ g/lane) was run on each gel as a DNA size standard. The size of the fragments, in base pairs (bp) were: 12216, 11198, 10180, 9162, 8144, 7126, 6108, 5090, 4072, 3054, 2036, 1636, 1018, 517/ 506. Fragment sizes of digested DNA were determined by computer program DNA Frag (Schaffer and Sederoff, 1981)

3.7. CLONING RESTRICTION FRAGMENTS INTO VECTOR DNA

a) Shotgun cloning

1-2 μ g of CsCl or Wizard isolated insert and vector DNA were digested with the appropriate restriction enzymes in separate reactions. The restriction enzymes in each reaction were heat inactivated or removed by Gene Clean only if they would cause unwanted digestion in subsequent ligation reactions. 1-5 μ l of insert DNA digest was mixed with 1-3 μ l of vector DNA digest such that the insert DNA was present in at least two fold excess with respect to the vector DNA. 4 μ l of 5x BRL ligation buffer (330 mM Tris-HCl, pH 7.6, 33 mM $MgCl_2$, 50 mM DTT, 330 μ M ATP) was added and the volume brought to 19 μ l with dH_2O . To this 1 μ l of BRL T4 DNA ligase and the reaction was incubated at 15°C overnight.

b) Extraction of DNA restriction fragments from agarose gels by Gene Clean procedure for subcloning

1-2 μ g of CsCl or Wizard isolated insert DNA was digested and run on a TAE electrophoresis gel. The gel was stained and viewed under long wave U.V. light and the desired fragment cut from the gel. The agarose fragment was weighed and 2-3 volumes 6 M NaI solution was added. The agarose was dissolved at 50°C for 5 min. 5 μ l of Gene Clean Glass Milk[™] was added and the preparation held on ice for 5 min. The DNA coated Glass Milk was pelleted by spinning in a microfuge for 5 sec and the supernatant discarded. The pellet was washed in 300 μ l of NEW solution three times and the final pellet resuspended in 10 μ l of dH_2O . Following incubation at 50°C for 2-5 min.

the Glass Milk matrix was pelleted 30 sec in a microfuge and the supernatant containing the DNA was retrieved. Up to 10 μ l of this insert DNA preparation was ligated to 1 μ l of vector DNA digestion, in a 20 μ l reaction as outlined above for shot gun cloning

3.8. BACTERIAL TRANSFORMATION

a) CaCl_2 transformation.

An overnight culture of the strain to be transformed was diluted 0.5 ml in 30 ml LB and reincubated for 2.5-3 hours at 37°C. The cells were chilled on ice for 20 min then pelleted at 5000 RPM for 5 min at 4°C in an SS-34 rotor and Sorval centrifuge. The cells were resuspended in 7.5 ml of ice cold 100 mM MgCl_2 and repelleted. The cells were then resuspended in 1-1.5 ml of ice cold 100 mM CaCl_2 and held on ice for a minimum of 30 min, however cells were most competent after overnight incubation on ice and remained competent for several days. Glycerol was added to 20% to cells to be frozen at -80°C.

50-200 μ l aliquots of competent cells were mixed with DNA in chilled glass tubes and left on ice for one hour. Generally, 1 μ l of CsCl DNA, 2.5 μ l of min prep DNA or 5-10 μ l of ligation mixtures diluted to 50 μ l in cold sterile dH_2O was used to transform cells. After one hour on ice, the transformed cells were heat pulsed at 37°C for 5 min then transferred to ice for 10 min. 1 ml of LB was added to the cells which were then incubated at 37°C for 45-90 min with gentle shaking. The cells were plated on LBA containing the antibiotic to select for the plasmid marker. When pUC18 or pUC128 recombinant DNA was transformed into DH5 α , X gal (40 μ g/ml) was added

to the agar medium along with ampicillin (30 $\mu\text{g}/\text{ml}$) to allow the identification of recombinant colonies. Plates were incubated at 37°C overnight and isolated colonies streaked for purity on selective plates.

b) Transformation by electroporation

5 ml of overnight culture was diluted into 1 L of LB and incubated at 37°C with shaking until an OD_{600} of 0.5-1 was obtained. The cells were chilled on ice 30 min then pelleted in a Sorval GSA rotor at 5000 RPM for 10 min at 4°C. The pellets were washed in a total of 1 L cold dH_2O and pelleted as above. The cells were then washed in 500 ml cold, sterile dH_2O and pelleted again. After a final wash in 20 ml cold, sterile dH_2O the cells were resuspended in 2-3 ml 10% cold glycerol and frozen at -80°C until required.

Electroporation was carried out by mixing 40 μl of frozen cells in 10% glycerol prepared as outlined above with 1-10 μl of DNA in H_2O . Generally 1 μl of CsCl DNA or 2 μl of miniprep was electroporated. Prior to electroporation, ligation mixtures were either desalted by ethanol precipitation or Gene Clean. 2-10 μl of ligation DNA was used. After a brief incubation on ice for 30-60 sec, the mixture was transferred to a chilled 0.2 cm electroporation cuvette and a pulse of 2.5 volts applied to provide a time constant of approximately 5 msec. Immediately following the pulse, 1 ml of SOC (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 25 mM KCl, 10 mM MgCl_2 , 10 mM MgSO_4 , 20 mM glucose) was added to the cuvette and the cells were transferred to a sterile tube. The cells were incubated at 37°C, with shaking, for 1 hour prior to plating

on LBA containing antibiotics to select for the plasmid introduced with or without *Xba*I (40 μ g/ml).

3.9. SEQUENCING

a) Preparation of single stranded template DNA

Single-stranded template DNA was extracted from pUC128 derivatives by the method of Vieira and Messing (1987). 2.5 ml of 2X YT broth (1.6% tryptone, 1.0% yeast extract, 0.5% NaCl) was inoculated with individual white DH5 α pUC128 colonies and incubated at 37°C with vigorous shaking. After 3-4 hours, 0.1 ml M13KO7 helper phage (10^7 PFU/ml) was added. After 1 hour further incubation kanamycin was added to 70 μ g/ml and the culture was incubated overnight with vigorous shaking. 1.5 ml of this culture was pelleted in the microfuge and 1 ml of supernatant was precipitated with 115 μ l 10 M ammonium acetate and 185 μ l 40% PEG for 1 hour on ice or overnight at 4°C. The phage were pelleted at 4°C for 5 min in a microfuge and resuspended in 120 μ l TE, extracted with phenol-phenol and chloroform (1:1) and chloroform and precipitated with ammonium acetate and ethanol. The resultant pellet was washed with 70% ethanol, dried and resuspended in 50 μ l of dH $_2$ O.

b) Preparation of double-stranded template

Double stranded DNA to be sequenced was prepared either by λ ST1 extraction, Wizard plasmid extraction or boiling miniprep method outlined above. DNA isolated by boiling miniprep method was Gene Cleaned and resuspended in 10 μ l dH $_2$ O prior to

Figure 4.1. Hybridization of pCU1 *oriT* probe to *HpaI*, *SspI* and *DraI* digests of pCU1, pSP274 and pSP545 DNA. a) restriction digests b) autoradiograph after wash in 1.5X SSC 1., 1 Kb ladder. 2. pSP274; 1. pCU1; 5. pSP545; H, *HpaI*; D, *DraI*; S, *SspI*.

2

PM-1 3 1/2" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

sequencing.

c) Sequencing reactions

The Sanger, dideoxy sequencing procedure was carried out on single stranded DNA template as outlined in the US Biochemical Sequenase product pamphlet. 7 μ l of single stranded DNA was mixed with 1 μ l (0.5 μ moles/ μ l) primer DNA and 2 μ l 5X sequencing buffer (200 mM Tris-HCl, pH 7.5, 100 mM MgCl₂, 250 mM NaCl) at 65°C for 5 min and allowed to cool slowly in the same beaker of water to 37°C. To the annealed template was added: 1 μ l 0.1 M DTT, 2 μ l diluted labelling mix - either 1:4 or 1:9 dilution was used (1.5 or 0.75 μ M dGTP, dCTP and dTTP) depending on whether the region of the DNA to be sequenced was close to the primer or further away. 1 μ l ³²S- α dATP (10 μ Ci) and 2 μ l of dilute Sequenase[™] enzyme. The tubes were incubated 5 min at room temperature after which time 3.2 μ l was transferred to each of 4 prewarmed tubes containing 2.5 μ l of one of the 4 termination mixtures (50 mM NaCl, 80 μ M dGTP, 80 μ M dATP, 80 μ M dCTP, 80 μ M dTTP, 8 μ M of either ddGTP, ddATP, ddCTP or ddCTP) and incubated 5 mins at 37°C. The reaction was terminated with the addition of 4 μ l STOP solution (95% formamide, 20 mM EDTA, 0.05% Bromophenol blue, 0.05% xylene cyanol FF).

Sequencing of double-stranded DNA was carried out using a modification of the US Biochemical Sequenase protocol. 1 μ l of CsCl DNA or 7 μ l of Wizard or Gene Cleaned miniprep DNA was boiled for 5 min with 12-20 μ moles of primer and immediately quick chilled in an ice slurry. After quick spinning at 4°C to transfer the

liquid to the bottom of the tube, ice cold reagents were added as outlined above for single-stranded template. Immediately following addition of the enzyme, the tube was mixed, quick spun and then 3.2 μ l of the reaction transferred directly to tubes containing 2.5 μ l of termination mixture and incubated at 37°C for 5 mins. 4 μ l STOP solution was added to terminate the reaction.

d) Sequencing acrylamide gel electrophoresis

8% Acrylamide gels (Sequagel 8, Diamed) containing 7 M urea in TBE (89 mM Tris, 89 mM boric acid, 25 mM EDTA, pH 8.3) buffer were used for sequencing electrophoresis. The sequencing apparatus was from BRL. Gels were prerun at 1000 volts for 45 min. Samples were boiled for 2-3 min and 2-3 μ l loaded. Gels were run at 60 watts (1200-1700 volts, 30-45 mamps) for 30-60 min after the bromophenol blue ran off the gel. The rest of the sample was denatured and loaded for the second run until the bromophenol blue ran off the gel. Gels were fixed in a solution of 10% acetic acid and 10% methanol for 30 min and dried at 80°C under vacuum for 1 hour. X-Omat AR film was exposed for 24-48 hours at room temperature and developed in a film processor.

3.11 DNA:DNA HYBRIDIZATION

a) Southern Blot

Following electrophoresis, the gels were soaked in 0.2 M HCl acid for 30 min, rinsed with dH₂O then soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 30

mins. Following rinsing in dH₂O, the gels were soaked in neutralizing solution (1 M Tris-HCl, pH 7.4, 1.5 M NaCl). The gels were placed on a piece of Whatman #3 filter paper wicked in 20X SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0). A piece of Hybond N (Amersham) was soaked in 2X SSC and placed on the gels followed by two pieces of Whatman #3 filter paper also soaked in 20X SSC. 10 cm of paper towel was placed on top and a 500 g weight added. After blotting overnight, filters were washed in 2X SSC, dried and exposed to UV light for 5 min. Blots were stored at room temperature until used for hybridization.

b) Labelling the probe

6 μ l of CsCl pSP25 DNA was digested with *Eco*RI and *Bam*HI and run on a TAE electrophoresis gel. The 325 bp *oriT* DNA band was extracted from the gel by Gene Clean and resuspended in 30 μ l dH₂O. The 3-10 Kb fragments of 1 μ g of 1 Kb ladder (BRL) were also extracted from the gel by Gene Clean. Both the probe and the 1 Kb ladder were labelled using a nick translation kit from BRL. Each DNA was mixed with 10 μ Ci ³²P α dATP, 0.2 mM each of dTTP, dCTP and dGTP, 200 μ g DNaseI and 2.0 units DNA polymerase I in a 50 μ l reaction containing 50 mM Tris-HCl (pH 7.8), 5 mM MgCl, 10 mM 2-mercaptoethanol and 10 μ g/ml BSA. This reaction was incubated 1 hour at 15°C. The unincorporated nucleotides and enzymes were removed by Gene Clean and the labelled DNA resuspended in 100 μ l. Scintillation of the pSP25 probe demonstrated a count of 1.7×10^5 cpm/ μ l. The 1 Kb probe gave a count of 2.3×10^6 cpm/ μ l.

c) DNA:DNA hybridization

The southern blot filter were wet in 2xSSC and placed in a hybridization tube with 40 mls hybridization solution (6X SSC, 5X Denhardt solution, 0.5% SDS, 100 $\mu\text{g/ml}$ denatured salmon sperm DNA). The filter was incubated at 30°C for 1 hour after which the solution was replaced with 40 ml of fresh hybridization solution. 5 μl 1 Kb ladder probe and 60 μl pSP25 probe denatured by boiling 5 min was added and hybridization was run at 30°C overnight. The filter was washed 5 times with 25 ml of 1.5X SSC + 0.5% SDS for 5 min each at room temperature (20-25°C) followed by 5 washes with 25 ml 1X SSC + 0.1% SDS at 30°C. The filter was wrapped in Saran wrap and exposed to X-Omat AR film at room temperature for 24-48 hours.

3.11 POLYMERASE CHAIN REACTION (PCR)

a) Introduction of random point mutations by PCR

100 μl PCR reactions were prepared containing 0.5 μl plasmid DNA prepared by Wizard miniprep (approximately 0.1 μg), 5 mM MgCl, 0.25 mM each of dATP, dTTP, dCTP and dGTP (unless otherwise stated), 10 μl Promega 10 *Taq* buffer (10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1% Triton X), 20 μmole each of pUC forward and reverse primers, and 2 units of Promega *Taq* DNA polymerase. The PCR consisted of an initial 1 min denaturation step at 94°C followed by 25 cycles of 30 sec at 94°C, 30 sec at 55°C and 1 min at 72°C and finished with a single 5 min extension period at 72°C.

b) PCR with degenerate oligonucleotides

The degenerate oligonucleotides were synthesized on a PCR mate synthesizer using chemicals from Applied Biosystems Inc. They were synthesized on 0.2 mmole columns using the low cycle. Degeneracy was introduced by allowing the machine to deliver a mixture of 90% wild type phosphoramidite (in this case A) and 3.3% of each of the other three phosphoramidites. The oligonucleotides were cleaved from the column in ammonium hydroxide at room temperature and deprotected by overnight incubation in ammonium hydroxide at 55°C. The oligonucleotides were then purified on an oligonucleotide purification column (OPC, Applied Biosystems Inc) following the manufacturer's instruction.

The percentage of oligo expected to contain a single number of mutations was calculated using the following equation: $n! \times (P)^{n-r} \times (1-P)/(n-r)! \times r!$ where n is the number of sites to be mutated (7 for ORIT4, 6 for ORIT4op), P is the ratio of wildtype base to contaminating bases (in both cases 0.9) and r is the number of mutations per oligo (ideally 1) (Sambrooke *et al.* 1989).

The PCR was conducted in two separate reactions. Both reactions contained 1/10th the reaction volume Promega *Taq* buffer, 0.2 mM MgCl, 0.5 mM of each dNTP, 2 units of Promega *Taq* DNA polymerase. The first round reactions were 100 μ l reactions containing either 67 μ mole of both pUC forward and ORIT4 primers or 67 μ mole of pUC reverse and ORIT4op primers. The reaction conditions consisted of an initial 1 min denaturation step at 94°C followed by 5 cycles of 30 sec at 94°C, 30 sec at 50°C and 1 min at 72°C. A final 5 min elongation period at 72°C completed the reaction.

The PCR products of the first reactions were purified using a Centricon 100 column following the manufacturer's instructions. 40-50 μ l of purified PCR product was collected. In the second round of PCR, 10 μ l of each of the Centricon purified PCR products from the first round PCR were mixed together with 10 μ mole of each of the pUC forward and reverse primers in a 50 μ l reaction. The conditions of this PCR consisted of an initial 1 min denaturation step at 94°C, followed by 15 cycles of 30 sec at 94°C, 30 sec at 50°C and 1 min at 72°C. A single 5 min elongation step at 72°C finished the reaction.

c) Deletion by PCR

Deletion PCR was conducted in 100 μ l reactions containing 10 μ l Promega *Taq* buffer, 0.2 mM MgCl₂, 0.5 mM of each dNTP, 50 μ mole of either ORIT2 or ORIT3 primer and 50 μ mole of pUC reverse primer and 2 units of Promega *Taq* DNA polymerase. Reaction conditions were the same as for random PCR mutagenesis.

3.12 INTRODUCTION OF *TnphoA* INSERTIONS

a) Preparation of λ ::*TnphoA* phage suspension

0.5 ml of λ *TnphoA* (titer 10^4 pfu/ml) was mixed with 1 ml of 3 hr BW11397 and incubated at 37°C for 20 min without shaking. 10 ml of molten λ top agar (1% tryptone, 0.25% NaCl, 0.6% agar) was added to the mixture and 2.5 ml spread on each of 5 λ bottom agar plates (as for λ top agar except 1.2% agar). After overnight incubation at 37°C the top agar was scraped off the plates and vortexed with 5 ml SM buffer (0.1 M

NaCl, 0.07 M Tris pH 7.5, 0.01 M MgSO_4 , 0.01% gelatin). The suspension was centrifuged at 10,000g for 5 min at 4°C in a SS34 Sorval rotor. The supernatant was collected and filtered through a 0.45 μm pore membrane filter. The phage suspension was titred by mixing 0.1 ml of phage dilutions prepared in SM buffer with 0.2 ml 3 hr BW11397, incubated for 20 min at 37°C and plated with 2.5 ml λ top agar on λ bottom agar plates. Plaques were counted after 24 hours at 37°C.

b) Mutagenesis of pSP34

1 ml log phase CC118/pSP34 was incubated with 1 ml of λTnphoA suspension (titer 10^8 - 10^9) overnight at 30°C with gentle shaking. The culture was pelleted and resuspended in 1 ml LB. 0.2 ml of this suspension was spread onto each of 5 LBA+Km+Cm plates to dryness. 500 μl of sterile Km solution (10^4 $\mu\text{g/ml}$) was then added to the centre of the plate to create a large puddle. The plates were incubated upright overnight at 37°C. Isolated colonies within the zone of clearing on the lawn of bacterial growth were selected for further analysis by miniprep DNA extraction and restriction digests.

c) Mutagenesis of pET-traHI by TnphoA

pET-traHI was introduced into *E. coli* strain PN1 which carries TnphoA on the chromosome by CaCl_2 transformation. Rather than plating the transformation mixture onto LBA plates, the culture was incubated in LB overnight at 37°C and plated the following day on LBA+Cm+Km. After overnight incubation at 37°C, the confluent

growth was scraped from the plate and plasmid DNA extracted by the Wizard miniprep procedure. 5 μ l of this DNA was electroporated into DH5 α and plated on LB+ Cm+Km. Cm'Km' DH5 α colonies were selected for further analysis by restriction digestion of boiling miniprep DNA.

3.14 SDS-PAGE OF *mob* PROTEIN SAMPLES

a) Preparation of protein sample

pET derivatives were introduced into *E. coli* strain BL21(DE3) via CaCl₂ transformation. Either a single freshly transformed colony or 20 μ l of an overnight culture was inoculated into 2 ml of LB+Ap (or LB+Ap+Km for pET-traHI::phoA isolates) and incubated at 37°C for 2 hours with vigorous shaking. Cultures were induced by adding IPTG to a concentration of 0.05 - 0.4 mM followed by further incubation for 2-3 hours (Studier *et al.*, 1990). 1 ml of cells was transferred to a microfuge tube and pelleted. The pellet was resuspended in 100 μ l of protein loading buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 0.01% Bromophenol blue, 10% glycerol, 100 mM DTT) and boiled 5 min. After a brief vortex, the boiled preparation was spun for 5 min in the microfuge. 90 μ l of supernatant was recovered and stored at -20°C until required.

b) SDS-PAGE

Discontinuous SDS-PAGE were prepared as described by Hames (1981). A Hoefer vertical gel system was used. 1 μ l (approximately 2 μ g of each protein) of broad

range protein standard from Bio Rad was mixed with 5 μ l of protein loading buffer. Protein standard and protein preparations were boiled for 5 min prior to loading and 5 μ l of protein sample was loaded per well. Gels were run at 100 volts until sample entered the resolving gel and then at 200 volts until the bromophenol dye ran off the gel. The gel was stained 30 min in Coomassie blue solution (0.25% Coomassie Brilliant Blue R250, 10% acetic acid, 45% methanol, filtered through a Whatman #1 filter) followed by overnight destaining in several changes of 10% acetic acid and 12.5% isopropanol. The gels were soaked briefly in dH₂O and dried overnight using Biowrap^{1m} as instructed.

4. RESULTS

4.1. COMPARISON OF THE *oriT* REGIONS OF pCU1 AND pMUR274

Plasmids pMUR274 and pMUR545 (referred to as the pMUR plasmids) encode an N type transfer system as defined by their sensitivity to I_{ke} bacteriophage (Paterson and Iyer, 1992). Early studies indicated that while they are able to mobilize pCU57, which carries the entire pCU1 *mob* region, the pMUR plasmids are unable to mobilize pBP1 which carries just the *oriT* of pCU1 (Paterson and Iyer, 1992). This intermediate level of interaction between the *tra* systems of pCU1 and the pMUR plasmids suggested that they might also exhibit an intermediate level of sequence identity. By contrast, N plasmid pKM101 which is able to mobilize both pCU57 and pBP1 has been shown to carry an *oriT* sharing 100% sequence identity to that of pCU1 (Coupland *et al.* 1987).

The initial intent of cloning and sequencing the *oriT* region of the pMUR plasmids was to allow a comparison of the sequence with that of pCU1. This comparison would help to identify functionally important sequences based on the belief that essential sequences, such as *nrc*, would be highly conserved while protein binding sites, believed to cause plasmid specificity of *mob* functions, would be less conserved.

4.1.1 Cloning the *oriT* region of pMUR274

Prior to this study, there were no restriction or functional maps of the pMUR plasmids. Initial restriction analysis was conducted on pSP274 and pSP545 (Tra⁺ Tn10 derivatives of pMUR274 and pMUR545 respectively) to determine whether the two

plasmids showed any similarities in restriction pattern to pCU1 which might help to locate the *oriT*. Digestion of the three plasmids with several restriction enzymes known to cut pCU1 indicated that while the two pMUR plasmids were similar (but not identical) to each other, they bore little similarity to pCU1 (see Figure 4.1a). This is in contrast to the well studied N plasmid pKM101 which shares many restriction fragments with pCU1, particularly within the *tra* region (Iyer, 1989). The size of the pMUR plasmids is approximately 40 Kb.

Since no information concerning the location of the pMUR *oriT* could be discerned directly from the restriction digests, hybridization experiments were conducted using the pCU1 *oriT* as a probe. It was hoped that there would be sufficient sequence homology between the two *oriT* regions to identify a restriction fragment carrying the pMUR *oriT*. Restriction enzymes which cut plasmids pSP274 and pSP545 the greatest number of times were used in hopes of locating the pMUR *oriT* on a small restriction fragment which could be sequenced quickly. A Southern blot of a gel containing *SspI*, *HpaI* and *DraI* digests of pSP274, pSP545 and pCU1 DNA (see Figure 4.1a) was probed with the 325 bp pCU1 *oriT* *EcoRI/BamHI* fragment from pSP25 labelled with ^{32}P .

Initial hybridization reactions conducted at 50°C produced strong signals with the pCU1 control DNA but no signals with the pMUR DNA (data not shown). Subsequent hybridizations were conducted at 30°C. After initial washes at 30°C in 1.5X SSC (Figure 4.1b), the pCU1 digests show intense signals with one or two small fragments depending on the digest, as well as some less intense signals with several

Figure 4.1. Hybridization of pCU1 *oriT* probe to *HpaI*, *SspI* and *DraI* digests of pCU1, pSP274 and pSP545 DNA. a) restriction digests b) autoradiograph after wash in 1.5X SSC. 1, 1 Kb ladder. 2, pSP274: 1, pCU1; 5, pSP545; H, *HpaI*; D, *DraI*; S, *SspI*.

HpaI			DraI			SspI		
2	1	5	2	1	5	2	1	5
1	5	L	1	5	L	1	5	L
5	L	2	5	L	2	5	L	2

A

HpaI			DraI			SspI		
2	1	5	2	1	5	2	1	5
1	5	L	1	5	L	1	5	L
5	L	2	5	L	2	5	L	2

B

larger fragments while the pSP274 and pSP545 digests show weak signals with several fragments in each of the three digests. A second 30°C wash in 1X SSC eliminated most of the background signals with the larger pCU1 fragments, leaving only the strong signals with the smaller fragments which correlate to the known location of the pCU1 *oriT* (data not shown). After the 1X SSC wash, only very faint signals with the pMUR DNA remained which corresponded to the strongest signals observed after the initial 1.5X SSC wash (see Figure 4.1b). In the pSP545 DNA these bands were a 2 Kb *HpaI*, a 1.4 Kb *DraI* and a 5.5 Kb *SspI* fragment. In the pSP274 DNA the bands were a 5 Kb *HpaI*, a 1.3 Kb *DraI* and a 4.2 Kb *SspI* fragment. Note that the signals corresponds to different size restriction fragments in pSP274 and pSP545 in each of the three digests emphasizing that the two plasmids are not identical. The smaller *DraI* fragments of pSP274 and pSP545 which produced weak signals after the 1X SSC wash were chosen for subcloning.

The cluster of pSP274 and pSP545 *DraI* fragments around the 1.3 Kb size were removed from an agarose gel by Gene Clean and ligated with pUC128 cut with *SmaI* (see Figure 4.2). To isolate recombinant molecules containing the *oriT* fragment from those containing one of the other small *DraI* fragments extracted from the gel, the transformation mixture was mated with HB101Rif in the presence of a helper strain carrying pSP274 or pSP545. Rif^rAp^r transconjugants (those carrying recombinant plasmids able to be mobilized due to the presence of *oriT*) were only obtained from the mating involving the pSP274 transformation. Miniprep DNA was extracted from Rif^rAp^r transconjugants and purified by transformation into DH5 α . The recombinant plasmid

Figure 4.2 Construction of pMO2, pMO2S and pMO2R. pMO2 was constructed by shotgun cloning a group of pSP274 *Dra*I fragments, 1.1-1.5 Kb in size, extracted from an agarose gel by Gene Clean, into the phagemid pUC128 cut with *Sma*I. To construct pMO2S, the pMUR *oriT* from pMO2 was isolated on a *Bam*HI-*Hind*III fragment by Gene Clean, digested with *Ssp*I and ligated into pUC128 cut with *Hinc*II-*Bam*HI. To construct pMO2R, the pMUR *oriT* region was isolated from pMO2S on a *Kpn*I-*Sac*I fragment, digested with *Rsa*I and ligated to pUC128 cut with *Kpn*I-*Sph*I (the *Sph*I was made blunt with T4 DNA polymerase). The dotted arrow represents pSP274 DNA with the direction depicting the orientation of the *oriT* relative to that presented in all subsequent figures.

was named pMO2. pSP274 was introduced into DH5 α /pMO2 and the resultant strain was mated with HB101Rif. The results seen in Table 4.1 indicate that pMO2 is mobilized efficiently confirming the presence of the pMUR *oriT* in pMO2.

A comparison of the pMUR *oriT* sequence with that of pCU1 showed sequence identity between the two plasmids within the internal *SspI* to *RsaI* fragment of the pMUR *oriT* (an alignment of the two sequences is presented in Figure 4.5 and discussed below). A derivative carrying only this *SspI*-*RsaI* fragment was constructed by first subcloning the *SspI*-*Bam*HI fragment to produce pMO2S as described in Figure 4.2. The *RsaI*-*KpnI* fragment from pMO2S was then subcloned into pUC128 to create pMO2R. Sequencing either end of both pMO2S and pMO2R confirmed the proper fragment had been cloned in each case.

Mobilization experiments demonstrated that both deletion derivatives pMO2S and pMO2R were mobilized by pSP274 at frequencies comparable to pMO2 (see Table 4.1). The 474 bp *RsaI*-*SspI* fragment carried by pMO2R represents the minimal pMUR *oriT* region necessary for efficient transfer.

4.1.2 Sequencing the pMUR *oriT*

While it was possible to sequence several hundred bases into both ends of the 1.3 *DraI* *oriT* region carried by pMO2 from the primer sites at either end of the polylinker of the pUC128 cloning vector, in order to determine the entire sequence it was necessary to subclone the internal portions. pMO2 was digested with 15 different restriction enzymes with 6 bp recognition sequences including *SspI* and *SnaBI* which are present in

Table 4.1 Mobilization frequency of pMUR *oriT* derivatives by pSP274.

<i>oriT</i> derivative	Mobilization frequency ^a
pMO2	1.1×10^0
pMO2S	9.4×10^{-1}
pMO2R	1.3×10^0

^a calculated as the number of transconjugants receiving the *oriT* derivative per recipient receiving pSP274. Values are the average of at least two trials. Replica trial values varied less than 2 fold from each other.

the pCU1 *oriT* as well as those enzymes which cut within the pUC128 polylinker and would therefore facilitate deletion of portion of the 1.3 Kb *oriT*. Of these enzymes only *SspI* cut within the pMUR *oriT*. Further digests with enzymes which recognize 4 bp sequences indicated that both *HpaII* and *RsaI* each cut at least twice within the pMUR *oriT*.

HpaII produces a 1 bp overhang compatible with *Clal* found in the polylinker of pUC128. The *EcoRI* fragment from pMO2 containing the pMUR *oriT* was Gene Cleaned from an agarose gel and digested with *HpaII*. Only two fragments, approximately 700 and 200 bp, were observed indicating that there were additional, smaller fragments which could not be seen on the gel. These fragments were subcloned back into pUC128 cut with either *Clal* (for internal fragments) or *Clal/EcoRI* (for 'end' fragments).

RsaI produces blunt ends which can be ligated to the blunt end generated in pUC128 by *HincII* alone or with linkers for a sticky end restriction enzyme. The *BamHI/HincII* fragment of pMO2 containing the pMUR *oriT* was digested with *RsaI* and fragments of approximately 800 and 300 bp were observed again, suggesting the presence of at least one internal *RsaI* fragment. The *RsaI/BamHI* fragments were ligated together with *BamHI* linkers into pUC128 cut with *BamHI*.

Two clones (HA3 and HA6) were obtained from the *HpaII* subcloning and one (RA1) from the *RsaI* subcloning (see Figure 4.3). Single-stranded DNA was generated from each of these four clones and pMO2 and sequenced using either the forward or reverse sequencing primers. This provided the sequence 600 bp from the *HindIII* end of

Figure 4.3. Sequencing strategy for the pMUR *oriT*. The 1375 bp *DraI* fragment from pSP274 is represented by the middle line. The minimal *oriT* region contained on the internal *SspI-RsaI* fragment is indicated by the dotted arrow on the line. The direction of the arrow corresponds to the orientation depicted in all subsequent figures. Arrows above the line represent the direction (strand) and extent of the sequence data obtained from each plasmid-primer pair. Primers F and R are the forward and reverse pUC primers. Primers 1-3 are the MUR primers generated from the known sequence. The extent of the three pMO2 deletion derivatives is indicated by the lines below the pMO2 line. Restriction sites are indicated.

the pUC128 polylinker and 400 bp from the *Bam*HI end. Three synthetic primers, MUR1, MUR2 and MUR3 (see Figure 4.4), were generated from the determined sequence to allow sequencing of the internal portion of pMO2. A summary of the clones used and the sequencing strategy is presented in Figure 4.3 and the nucleotide sequence of the entire 1375 bp *Dra*I fragment cloned from pSP274 is presented in Figure 4.4.

4.1.3 Sequence alignment of the pMUR and pCU1 *oriT* regions

A sequence alignment of the pCU1 and pMUR *oriT* regions together with the *oriT* sequence of the IncW plasmid R388 is presented in Figure 4.5. The identity between the pCU1 and R388 *oriT* sequences has been described previously (Llsoa *et al.*, 1991). The overall sequence identity between the pCU1 and pMUR *oriT* regions is 57% over the 266 bp sequence carried by the deletion derivative pSP9 described in Section 4.3.4 as the minimal pCU1 *oriT* region. Identity over this region is slightly higher between pCU1 and R388 at 34.5% than between pMUR and R388 at 28%. As described in Section 4.2.3, *nic* lies within the 30 bp sequence showing 86.7% and 80% sequence identity with pMUR274 and R388 respectively and which has been shown to contain the *nic* site of R388 (de la Cruz, personal communication).

All three plasmids contain several inverted repeats (IR). There is a 10 bp IR at the left most end of each of the *oriT* sequences. While there is some sequence conservation in the pCU1 and pMUR IR₁ (7/10 bases), the corresponding IR₁ in R388 shares only 2 and 4 of 10 bases with the pCU1 and pMUR sequence respectively. Each of the three plasmids have additional IRs to the right of *nic*. pCU1 has four IR, 7, 7,

Figure 4.4 Nucleotide sequence of the 1375 bp *Dra*I *oriT* fragment from pSP274. Restriction sites are in bold. MUR primer sites are indicated by arrows below the sequence (direction of the arrow depicts 5'-3').

DraI
 AAATCGGTAT CGGATAAAAA TTCTTTATTG CCGCCCATAT CTTCAACCAG CTTATCTTGT 60
 TTTGGGTTGT GCAATCCATA ATCCGCTTGC *RsaI*
 TGTACTTCCT CTCTCCTTTC TGGATTGTGC 120
 AATCCACTTT CTATCTGGCT TTTATCTTCC TGTTTTGTG AATTGTGCAA TCTACTTTTC 180
 AGTTCTTTCC TTGCCCTATA CAGAGCCTTA TCGAAGGCCG TTAATGTGAT AGAACACGTA 240
 GCGCTCAGTC AATACGCTCG CAATCTCATC ACGACTAACG CCATTTTTC CAAATATCTA 300
 TGTACGCCAG TGACGCTCTA ACAACAGCAA TTACCGTTCT AGGCCAAGAT TCGCACTAGT 360
 CATTATAACA GCCCTTACTT GTCTTAACAC AGTCCTAATA *SspI* TTGTCCATTT CTAGTCTCAA 420
 GTCAATGTTT TTTTCTAAAA AATGTCTTAT TTGTGTCCTT TTCTTGTCTC AATAACTTAC 480
 TCATCGTGTC TTATTTTGT CTAATTCCTG TCCGTTTCCT TTCATGCAAG CATGAAAGGT 540
 GTGAATGTGA ACGCTTTTTC AGCGATTAC TTGCACGTAA *HpaII* CACCGTTTC TCCCTTAAAA 600
 AAGGCGTGGT ATGGTATCCC TTAACGGCTA TACACCATAC GCAATTACGA TTTTCATCGT 660
 AAATGCTTGA AAACAAAACA AATTATTCTA ATTGTGTTT TAATATGTAA TACACAACGT 720
 ATTACATATG TGGAATTAAG CCGCATCTAT GCTAGATTTG TATGTATTAC ATCGGTATTA 780
 CACAAAGGAA ACCTATGCCA TCTATCAGTT TAAGAGTGAC CGAAGCGCAA AAAATCGACC 840
 TTGAATACCG TTCGGAGGGC AACGTATCGG *RsaI* AGTACATCAA GCAACAGCTT TTTGGACAAA 900
 AAGAGAGTGC CAGTGAAGTG ATCGCCCGC *HpaII* TTGATAAATT GCTGGAACGT TCATCTTCAG 960
 AGCAGCAAGC *HpaII* CGCCCGACGA *RsaI* GATCGGTACA TCGAGGCAAT GTTGATTGAG GTGTTATTGA 1020
 TACTTCGCCA GGTGAGTAAA GCAGATGCGC AGAAAGTTGC *HpaII* CGGTGGTGAA *HpaII* ATTCGCCGT 1080
 TAGGGCTGGA GCCGTGGTCA GCAGAGCTGG CCCAGGAGAG AGAGGAACAG AGTCATGAAT 1140
 GAAGCGGATA AGGGTAAAT TTGCCTGGGG GCATTGCTTT TTACCCCCCT TGTTACGTGG 1200
 TTTATCGCGG CGAAACCTT GTATGGTTTT ACGCCAAAAG ATGCGCGTGA ACGTCTGGTT 1260
 TTACTTATTC AGCAGACTTT TGATTTATGG CCTTTATGGG GTGCTTTATT ACGTCGGTGA 1320
 AATTATTGCC ATCATAGGCT TAATCATATT TTTCAAATTT AATCACCAGA *DraI* TATTT 1375

Figure 4.5 Nucleotide sequence alignment of the *oriT* regions of pCU1, pMUR274 and R388. * indicate nucleotide matches with the pCU1 sequence; - indicate gaps introduced to allow alignment; direct repeats (open arrows) and inverted repeats (solid arrows, numbered 1-5) are indicated above each sequence; the AT rich regions of pCU1 and pMUR274 are in bold; the 10 bp junction of hybrid pMUR274-pCU1 *oriT* (see Section 4.2.3) and corresponding sequence in R388 is shaded. Nucleotide numbering for the pMUR274 DNA is such that nt 1 corresponds to the *SspI* site in Figure 4.4; nucleotide numbering for R388 corresponds to that found in Llosa *et. al.*, (1991). The alignment was prepared by Clustal V version 3.52 (Higgins and Sharp, 1989).

pMUR274	***** GTTTCCTTTCATGCA-AGCATGAAAGGTGTGAATG-TGAACG-CTTTTGT	163
pCU1	ACTGGTATTCATGCA-AGCATGAATAC-G-GAATATTGAACG-CTTTTTA	46
R388	ATAACCCAATGCGCATAGCGCATTGGGAGCAG-TTTTGACCTCTTTAGG * * * * *	64
pMUR274	* * * * * * * * * * * * AGCGATTCACTTGACGTAACACCGGTTTCTCCCTTAAAAAGGCGTGGT	211
pCU1	GTCAGTTAGTCCTTAACCTACGTAGATCGCTCCTACCAAAAAG-ATGAT	95
R388	GTCACGCTGGCGTTATCGCATGG---TTTCCGTCTTAAAAAGCCGGGTT- ***** * * *****	110
pMUR274	* ***** * ***** * * * * * * ATGGTATCCCTTAA--C--GCTATACACCATACGCA---ATTACGATT	253
pCU1	TTTTTATCCCTTAA--CCT-CTATACACCTAACGCAGCGGCATAAATCA	142
R388	GG--TAT-CCTTAAATCTGGCTATACGCAATACGCACCTTTCGGTGCGC *** * * * * *	157
pMUR274	* * * * * * * * * * * * TTCATCGTAAATGCTTGAAAACAAAACAAATTATTCATTTGTGTTTTA	303
pCU1	CGCTATCACACTGATTTAAAAAATAAAAAAATA---AATAGCGTCAATA	188
R388	G-----	158
pMUR274	* ***** * ***** * * * * * * ATATGTAATACACAACGTATTACATATGTGGAATTAAGCCGCATCTATGC	353
pCU1	TTCCGGTAATAATTA-GGTATTAAAAACACCTAATAACTCATTGTTTTATA	237
R388	-----TAAGTCATTGATTTACA *** * * * * *	175
pMUR274	* * * * * * * * * * * * TAGATTTGTA-----TGTATTACATCGGTATTACACAAA	387
pCU1	ATGAGTTTGC-----TTTATTACTTTGGTAATACTAAGA	271
R388	ATGACTTCCACGCCAAGGACGAAACCTGTGTAGTGTTATGCCACTACAA ***** * * * * *	225

8 and 6 bases in length, pMUR has three, 11, 11 and 8 bases long, and R388 has three, 6, 9 and 7 bases long. Unlike IR₁, the IR to the right of *nic* (shaded region) do not show any clear pattern of conservation. Only the last IR is in the same position in all three plasmids. The position of IR₂ next to *nic* is conserved in R388 and pMUR but not pCU1, while the IR₃ and IR₄ of pCU1 each align with similar structures in only one of the other two sequences. Noteworthy is the high sequence conservation between IR₄ of pCU1 and IR₃ of R388. Equally interesting is the apparent sequence conservation in the region of IR₃ in pCU1 and pMUR despite the fact that the two IRs are not in comparable positions. This latter observation may suggest that there is the potential for a smaller IR to form in pCU1 in the region corresponding to IR₁ of pMUR.

Other features conserved in all three plasmids are two directly repeated sequences, 9 bp long in pCU1 and 8 bp in both pMUR274 and R388 all containing the sequence CCTTAA. While these two direct repeats are in comparable locations in R388 and pMUR, only the second repeat directly to the left of *nic* aligns in all three plasmids.

A major difference between the N and W *oriT* sequences is the presence of a large AT rich region (99 bp in pCU1, 70 bp in pMUR) in the both the pCU1 and pMUR sequences but missing from the IncW *oriT*. Not only is there no corresponding AT rich sequence in the R388 *oriT*, it was necessary to introduce a large gap in the R388 sequence in this region in order to maintain the sequence alignment.

One final observation which can be made with respect to the pMUR sequence is the lack of direct repeats directly to the right of the *oriT* sequence shown in Figure 4.5. Both pCU1 and R388 have a series of 11 bp iterons (13 in pCU1 and 5 in R388) to the

left of *oriT* however the sequence of these iterons do not show any identity between the two plasmids. The lack of similar sequences to the left of the pMUR *oriT* supports an earlier conclusion that the iterons do not play a significant role in the transfer of pCU1 (Paterson and Iyer, 1992).

4.2 CONSTRUCTION OF A HYBRID *oriT* BY SITE-SPECIFIC RECOMBINATION

Once it was clear that the *oriT* regions of pMUR274 and pCU1 contained regions of sequence identity, attempts were made to create a hybrid *oriT* between the two similar *oriT* regions by site-specific recombination. This process has been described for two identical *oriT* sequences (Meyer, 1989; Llosa *et al.*, 1993; Gao *et al.*, 1994). In each case, plasmids were constructed containing the two *oriT* regions separated by a selectable (antibiotic resistance) or identifiable (*lacZ*) marker. When site-specific recombination occurs between the two *oriT* sequences, the marker gene is deleted resulting in an altered phenotype.

Attempts to create such a hybrid between the *oriT* regions of pCU1 and R388 were unsuccessful (data not shown, Llosa *et al.*, 1993) suggesting that neither sequence can be processed by the heterologous *mob* proteins. It was hoped that the higher degree of sequence identity between the two N plasmids would be sufficient to produce a hybrid *oriT* while still allowing for the identification of *nic* based on the sequence of the hybrid junction.

4.2.1 Construction of pMKN derivatives

Figure 4.6 outlines the construction of four pMKN derivatives which contain both the pCU1 and pMUR *oriT*s separated by the Km gene from Tn5. In each of the four derivatives the two *oriT*s are arranged in opposite orientations with respect to each other (1 vs 2) and to the Ap gene and replication origin (A vs B). The only other differences between these four derivatives is that the A versions were constructed in pUC128, which has a higher copy number than pUC18 used for the B versions (personal observation) and the A versions contain pCU1 DNA to the left of the minimal *oriT* including the 13, 11 bp iterons. Neither feature appears to have had any observable effect on the following experiments.

4.2.2 Site-specific recombination

Separate DH5 α derivatives were constructed containing each of the four pMKN derivatives together with either pSP274, pCU101 (a Tra⁺ derivative of pCU1) or both conjugative plasmids. These strains were mated with HB101Rif and the resultant Rif^rAp^r transconjugants were replica plated onto separate plates containing either Rif+Ap or Rif+Km to screen for Km^r transconjugants. The results are presented in Table 4.2. Km^r transconjugants, representing potential hybrid *oriT* derivatives, were obtained only when the two *oriT* sequences are oriented as direct repeats (versions 2A and 2B). This is consistent with a strand-specific, single-stranded nicking event. While Km^r transconjugants were obtained from pMKN2A regardless of the conjugative plasmid present, pMKN2B produced Km^r transconjugants only if pCU101 was present.

Figure 4.6 Construction of pMKN derivatives.

pMKN1A and 2A were constructed from pMO2Km, a derivative constructed by cloning the Km gene from Tn5 (heavy line) on a *Sma*I fragment from pKan/oriT into the *Eco*RV site of pMO2 which carries the pMUR *oriT* (dotted arrow). The pCU1 *oriT* (open arrow) was cloned from pBP1 on a *Cl*aI-*Xho*I fragment into pMO2Km cut with the same enzymes to create pMKN1A. pMKN2A was constructed from pMKN1A by inverting *Eco*RI fragment carrying the pMUR *oriT*.

pMKN1B and 2B were constructed from pSP28, a pUC18 derivative carrying the 325 bp *R*saI-*B*amHI pCU1 *oriT* fragment from pBP1. An *Eco*RI linker was ligated to the *R*saI site prior to ligation with the pUC18 DNA. To create pMKN1B, the pMUR *oriT* and Tn5 Km gene were cloned simultaneously on a *B*amHI-*H*indIII fragment from pMO2 and a *B*amHI fragment from pKan/oriT into the *B*amHI-*H*indIII sites of pSP28. pMKN2B was constructed from pMKN1B by inverting the *H*indIII fragment carrying the pMUR *oriT*.

The orientation of the arrows representing the two *oriT*s depicts the orientation of the *oriT* relative to the alignments seen in Figures 4.5 and 4.6. Ap, ampicillin; Km, kanamycin; Rep, plasmid replicon. Restriction sites are indicated. *R*saI* represents 5 closely spaced *R*saI sites.

Table 4.2 **Formation of Km^r transconjugants during mobilization of pMKN derivatives by pCU101^a and pSP274^b**

oriT Plasmid	Conjugative Plasmid(s)	Km ^r /Ap ^r transconjugants	
		Actual ^c	Percentage
pMKN1A	pCU101	0/200	<0.5
pMKN1A	pSP274	0/200	<0.5
pMKN1A	pCU101, pSP274	0/200	<0.5
pMKN2A	pCU101	6/300	2.0
pMKN2A	pSP274	4/200	2.0
pMKN2A	pCU101, pSP274	14/300	4.7
pMKN1B	pCU101	0/200	<0.5
pMKN1B	pSP274	0/200	<0.5
pMKN1B	pCU101, pSP274	0/200	<0.5
pMKN2B	pCU101	50/200	25.0
pMKN2B	pSP274	0/200	<0.5
pMKN2B	pCU101, pSP274	10/200	5.0

^a pCU101 is a Cm^r, Tra⁺ derivative of pCU1.

^b pSP274 is a Tc^r, Tra⁺, Tn10 derivative of pMUR274.

^c values are totals from at least two separate trials. Comparable results were obtained on each trial

Surprisingly, pCU101 alone produced more pNKN2BKm^r transconjugants than when it was present together with pSP274. These results are explained below (Section 4.2.4).

4.2.3 Sequencing the putative hybrid *oriT*s.

Single-stranded DNA was isolated from 11 pMKN2AKm^r derivatives (from 5 independent matings involving either or both conjugative helper plasmids) and sequenced from the pCU1 ORT1 primer (see Figure 4.10). Double-stranded DNA was isolated from eight pMKN2BKm^r derivatives (from 4 independent matings involving either pCU101 or both pCU1 and pSP274) and sequenced from the pUC reverse primer. All of the 2AKm^r *oriT* sequences were identical to each other regardless of the conjugative plasmid responsible for their formation, as were all of the 2BKm^r *oriT* sequences. Both the 2AKm^r and 2BKm^r *oriT* sequences were hybrids of the two parental *oriT*s however each contained opposite halves of the parental *oriT*s due to their position relative to the Ap gene and plasmid replicon (see Figure 4.7a for clarification). In both cases the junction between the two sequences occurred within the same 10 bp sequence GCTATACACC found in both parental plasmids (see Figure 4.7b). It can be concluded that *nic* lies within this 10 bp junction sequence. This 10 bp sequence shows 8/10 bp sequence identity with the sequence surrounding the R388 *nic* site GCTATAGACA where *nic* lies between the fifth and six nucleotide as determined by an *in vitro* nicking assay with purified TwrC protein (de la Cruz, personal communication).

The fact that all of the 19 Km^r hybrid *oriT*s are composed of opposite halves of each of the parental *oriT*s with the same junction indicates that the recombination event

Figure 4.7 Formation of hybrid *oriT* via site-specific recombination.

A). Schematic representation of the formation of hybrid *oriT* molecules. pMKN is a recombinant plasmid carrying two *oriT*s in directly repeated orientation. Assuming transfer occurs counterclockwise (see text), formation of a viable, Km^r hybrid *oriT* plasmid can only occur as depicted for hybrid to the left (L) where the initial nick occurs at *oriT*1 (open arrow) followed by unwinding in the counterclockwise orientation through the Rep region and Ap gene and termination at *oriT*2 (shaded arrow). Initiation at *oriT* 2 followed by unwinding in the counterclockwise direction through the Km gene and termination at *oriT* 1 results in a non-viable hybrid shown to the right (R) due to the deletion of the Rep and Ap genes.

B. Comparison of the hybrid *oriT* sequences of the pMKN2A and 2B Km^r plasmids and the parental plasmids pMUR274 and pCU1. Nucleotide matches are indicated by |. The 10 bp junction sequence (underlined) proposed to contain *nic* is defined by the T/G mismatch on the left side and the A/T mismatch to the right.

A.

B.

pMUR274	GTGGTATGGTATCCCTTAACGGCTATACACCATACGCAATTACGATTTT
pMKN2AKm ^o	GTGGTATGGTATCCCTTAACGGCTATACACCTAACGCAGCGGCATAAAT
pCU1	TGATTTTTTATCCCTTAACCTGCTATACACCTAACGCAGCGGCATAAAT
pMUR274	GTGGTATGGTATCCCTTAACGGCTATACACCATACGCAATTACGATTTT
pMKN2BKm ^o	TGATTTTTTATCCCTTAACCTGCTATACACCATACGCAATTACGATTTT
pCU1	TGATTTTTTATCCCTTAACCTGCTATACACCTAACGCAGCGGCATAAAT

is specific for this 10 base sequence. In addition to these experiments being conducted in a *RecA* host, the lack of hybrids with different junctions, despite the presence of other highly conserved sequences within the two parental *oriT* sequences, confirms that the recombination event is not due to homologous recombination. Furthermore, the fact that the 10 bp junction is identical in all of the hybrid *oriT*s indicates that there is a unique *nic* site in each *oriT* and that it is the same in both plasmids. Multiple *nic* sites or different sites within the two *oriT*s would produce a small (1 to several bases) deletion or duplication within the 10 bp junction.

4.2.4 Mobilization of the hybrid *oriT* derivatives

As indicated in Table 4.3, neither the pCU1 nor pMUR274 transfer systems is able to mobilize a plasmid carrying the *oriT* of the other, suggesting that the *mob* proteins of each are unable to interact with the *oriT* of the other. To test whether either parental plasmid is capable of mobilizing the hybrid *oriT* plasmids, representatives of both the pMKN2AKm⁺ and pMKN2BKm⁺ hybrid *oriT* derivatives were introduced into strains carrying either pCU101, pSP274 or both conjugative plasmids. These strains were mated with HB101Rif. The results shown in Table 4.3 indicate that each hybrid is mobilized by only one of the transfer systems. pMKN2AKm⁺ is efficiently mobilized by the pCU1 *tra* system while pMKN2BKm⁺ is efficiently mobilized by the pMUR *tra* system.

The ability of each conjugative plasmid to mobilize only one of the two hybrid *oriT* plasmids provides insight into the source of plasmid specificity. In each instance, the hybrid plasmids are mobilized only by the conjugative plasmid corresponding to the

Table 4.3 Mobilization of pCU1 and pMUR wildtype and hybrid *oriT* plasmids by pCU101^a and pSP274^b

<i>oriT</i> Plasmid	Conjugative Plasmid	Mobilization Frequency ^c	
pSP28Km	pCU101	1.3x10 ⁰	
pSP28Km	pSP274	8.2x10 ⁻⁷	
pMO2Km	pCU101	1.4x10 ⁻⁶	
pMO2Km	pSP274	1.0x10 ⁰	
pMKN2AKm ^s	pCU101	2.2x10 ⁰	
pMKN2AKm ^s	pSP274	< 3.2x10 ⁻⁵	
pMKN2AKm ^s	pCU101, pSP274	2.0x10 ⁰	8.8x10 ⁻¹
pMKN2BKm ^s	pCU101	< 4.5x10 ⁻⁴	
pMKN2BKm ^s	pSP274	5.0x10 ⁻²	
pMKN2BKm ^s	pCU101, pSP274	1.1x10 ⁰	1.5x10 ⁻¹

^a pCU101 is a Cm^r, Tra⁺ derivative of pCU1.

^b pSP274 is a Tc^r, Tra⁺, Tn10 derivative of pMUR274.

^c Mobilization frequencies are given as the number of transconjugants receiving the hybrid *oriT* plasmid per transconjugant receiving the conjugative plasmid. Where two frequencies are given, each corresponds to each of the conjugative plasmids respectively. Results are the average of at least two trials. Replica trial values varied less than 10 fold from each other.

right hand portion of the hybrid *oriT*. Since the degree of sequence identity between the two sequences is approximately equal on both side of *nic*, it is likely that the right hand portion of the *oriT* region contains the sequence(s) recognized by the *mob* proteins and by comparison to other conjugative systems, it is most likely that one or more of the *mob* proteins bind to this region.

The inability of each parental plasmid to mobilize one of the hybrid *oriT*s also raises the question of how recombination between the *oriT* regions occurs in the presence of only one of the conjugative plasmids. Site-specific recombination in the presence of only one parental *tra* system indicates that each system is able to nick and seal both *oriT*s. This would suggest that plasmid specificity is a function of the ability of the *mob* protein(s) to bind to the *oriT* and not one of nicking/sealing ability. The nicking protein(s) would bind to and initiate transfer from the corresponding *oriT*. By comparison to other systems, the nicking protein would remain bound to the 5' end of the nicked DNA and unwind the nicked strand in the 5' -> 3' direction (Pansegrau *et al.*, 1990, Inamoto *et al.*, 1991, Matson *et al.*, 1993). The bound protein would then nick the second, non-self *oriT* due to the similarity in the DNA sequence surrounding the *nic* site and seal the two ends to terminate the recombination event.

As indicated in Figure 4.7a, production of a hybrid *oriT* from the pMKN derivatives can occur only when transfer is initiated at one *oriT* on one side of the plasmid replicon and the DNA is unwound through the replicon and ampicillin gene followed by termination in the second *oriT* on the other side of the replicon. Initiation of transfer in one *oriT* followed by unwinding of the DNA through the Km gene and

termination in the other *oriT* would produce a non-viable molecule due to the lack of a functional replicon. Although the direction of transfer of pCU1 has not been determined, Coupland (in Lanka and Wilkins, 1995) demonstrated that transfer of IncN plasmid pKM101 proceeds with the *tra* genes last as seen with F plasmid (Frost *et al.*, 1994), and plasmids RP4 (Al-doori *et al.*, 1982), ColE1 (Boyd and Sherratt, 1986), RSF1010 (Kim and Meyer, 1986) and R64 (Howland and Wilkins, 1988). pCU1 and pKM101 have been shown to be very similar if not identical on both the restriction site and sequence level (Iyer, 1989; Paterson and Iyer, 1992; this study Section 4.5). Therefore, for transfer to occur in the 5' → 3' direction as seen in other transfer systems, *nic* would have to occur in the lower strand of the pCU1 *oriT* (complementary to the sequence illustrated in Figure 4.5) which agrees with data obtained for plasmid R388 (de la Cruz, personal communication). In the case of the pMKN constructs, transfer would occur in the counterclockwise direction.

This arrangement is consistent with the recombination frequencies presented in Table 4.2. In the case of pMKN2B, viable recombinants would be observed when transfer is initiated in the pCU1 *oriT* followed by transfer in the counter-clockwise direction and termination in the pMUR274 *oriT*. pCU101 is a derivative of pACYC184. As such, pCU101 has a higher copy number than pSP274 which is a *Tn10* insertion derivative of naturally occurring pMUR274. As a result, pCU101 would produce a higher concentration of *mob* protein than pSP274 and hence transfer initiation could proceed at a higher rate. This would explain the high recombination frequency (25%) of pMKN2B in the presence of pCU101 (Table 4.2). In the pMKN2A, recombinant

formation would be initiated by nicking within the pMUR274 *oriT* and due to the lower copy number of plasmid pSP274, fewer *mob* protein molecules would be available for initiation and site-specific recombination would occur less frequently (5%) (Table 4.2).

This arrangement also explains why pMKN2A recombinants were observed in the presence of either parental transfer system alone but pMKN2B recombinants were observed only if pCU101 was present in the donor. If the *mob* proteins of each transfer system are only able to bind with the corresponding *oriT*, then pMKN2A recombinants should readily occur in the presence of the pMUR274 transfer system by initiation within the pMUR274 *oriT* and termination within the pCU1 *oriT*. For pMKN2AKm^r recombinants to occur in the presence of pCU101 alone, the *mob* proteins would first have to bind to the pCU1 *oriT*, nick it, unwind the DNA in the counterclockwise direction, terminate transfer by nicking the pMUR274 *oriT* and produce a non-viable circular molecule containing the Km gene and no replicon. The pCU101 *mob* proteins could then continue unwinding the DNA from the nick in the pMUR274 *oriT* and proceed counterclockwise to terminate transfer within the pCU1 *oriT* producing a viable recombinant plasmid containing the replicon and Ap gene. Due to the increased chance of dissociation of the *mob* protein from the plasmid, this double initiation event would presumably occur at a reduced frequency, however due to the increased number of pCU101 *mob* protein molecules, we observe comparable recombination frequencies for pMKN2A in the presence of either conjugative plasmid. By comparison, in the case of pMKN2BKm^r, recombinant molecules would most easily be formed by the initiation of transfer within the pCU1 *oriT* followed by termination in the pMUR274 *oriT* by the

pCU101 *mob* proteins. In the presence of the pMUR274 *mob*, recombinant pMKN2B molecules would require the less efficient double initiation process, however due to the low copy number of pSP274, this event is not observed.

4.3 SITE-SPECIFIC AND DELETION MUTAGENESIS OF THE *oriT* OF pCU1

Alignment of the *oriT* sequences of pMUR274, pCU1 and R388 identified conservation on both the sequence and structural level which can be used as a basis for examining the importance of different regions of the *oriT* and suggest a logical approach to mutagenesis of the area. Initially, two approaches to mutagenesis of the *oriT* were considered: random mutagenesis via misincorporation by *Taq* polymerase during PCR and site-specific mutagenesis using degenerate oligonucleotides and PCR. The sequences that are the most highly conserved between the pCU1 and pMUR *oriT*s represent good targets for site-specific mutagenesis by PCR. On the other hand, using PCR to introduce random point mutations would allow for the mutation of sequences which might otherwise be overlooked due to their lack of identifiable structure or sequence conservation. Since both of these approaches had merit, both were attempted. Once it was demonstrated that point mutations had relatively little effect on transfer, small deletions of 20-50 bp were introduced using restriction sites either already present within the *oriT* or introduced by point mutations or by PCR deletion mutagenesis.

4.3.1 Generation of random point mutations by PCR

In the first attempts to introduce random mutations via PCR, the MgCl_2 concentration was increased to 5 mM while the dNTP concentration was decreased to 0.25 mM each, 1 mM total. This high ratio of MgCl_2 to dNTP has been shown to increase incorporation errors during PCR (Eckert and Kunkel, 1990). A PCR reaction was conducted using the forward and reverse pUC primers and pSP24 (see Figure 4.10) as the DNA template. The resultant PCR product was digested with *EcoRI-SalI* and ligated to pSPKm6 (pUC18 with the Km gene from pKan/*oriT* on a *HincII* fragment cloned into the *SspI-DraI* sites, see Figure 4.8) cut with the same enzymes. The ligation mixture was transformed into DH5 α /pCU1. The mobility of the recombinant plasmids in 83 white transformants growing on Xgal Km Sp plates was tested by spot mating with HB101Rif. All of the isolates showing no growth on RifKm plates were checked by restriction digest of miniprep DNA and were found to contain no insert. No isolates showing significant reduction in mobility of the Km marker were identified. To increase the sensitivity of the mobilization assay, 23 colonies which showed possible reductions in Km marker mobility by spot matings were tested by mini-filter mating, however none of these showed a significant reduction in mobility (more than a 5 fold reduction). On the chance that there were silent mutations, nine randomly selected isolates were sequenced and no mutations were identified.

In the second attempt to introduce mutations by PCR, in addition to the increased MgCl_2 to dNTP ratio as described above, the concentration of one of the four dNTPs was reduced from 0.25 mM to 0.05 or 0.10 mM to enhance the chance of misincorporation

of an abundant nucleotide for the limiting nucleotide. Four separate reactions were set up, two with limiting dATP and two with limiting dCTP. The resultant PCR products were cut with *EcoRI-SalI* and ligated to pSPKm6 cut with the same enzymes. The ligations were transformed into DH5 α /pCU1. A total of 162 white transformants (on Xgal Km plates) from the four ligations were tested by spot mating with recipient HB101Rif. No isolates showed significantly reduced transfer frequencies. One isolate with a two fold reduction in mobilization was sequenced and found to contain a point mutation in the first IR.

4.3.2 Site-specific mutation of *nic* by PCR cloning with degenerate oligonucleotides

The 10 bp *nic* sequence represented the best candidate for site-specific mutagenesis using degenerate oligonucleotide primers and PCR. The sequence targeted was a 20 bp sequence encompassing the 10 bp known to contain *nic* (see Figure 4.9). This sequence contains six A residues on one strand and seven on the other. Two oligonucleotides, ORIT4 and ORIT4 ϕ , 29 and 30 bp in length respectively, complementary to opposite strands of DNA encompassing the 20 bp target sequence were synthesized. The A phosphoramidite used in the synthesis was mixed with low levels of each of the other three phosphoramidites during the synthesis at each of the 13 A sites in the targeted area. Such an arrangement would result in mutations at either the T or A residues of each strand during complementary strand synthesis via PCR. The level of contamination was set at 3% for each base. This was predicted to produce a pool of oligos in which approximately 35% of the molecules contain one mutation and 10%

Figure 4.8 PCR cloning of *nic* mutants. In the first round of PCR, two reactions were conducted, one containing the pUC reverse (solid arrow) and ORIT4op (hatched arrow) primers and the other containing the pUC forward (open arrow) and ORIT4 (shaded arrow) primers. Each reaction synthesized one half of the pCU1 *oriT* region, some containing mutations (x) in the target region. The PCR products from the first reaction overlapped in the region of *nic*. The PCR products from both first round reactions were added to a second round of PCR together with both forward and reverse pUC primers. In the second round, the 3' ends of the PCR products from the first reaction anneal and are extended to produce full length *oriT* fragments which can be amplified by the forward and reverse primers. The products of the second round of PCR were digested with *EcoRI* and *SalI* and cloned into pSP6Km cut with the same enzymes

Round 1

Round 2

contain two (Sambrook *et al.*, 1989).

These two oligos were used as outlined in Figure 4.8 to produce a full length *oriT* fragment containing mutations in the target region. Each primer was used together with either the reverse or forward primers to produce separate halves of the *oriT* region with a 30 bp overlap spanning *nic*. These PCR products were combined in a second reaction where the 3' complementary ends would result in the synthesis of full length *oriT* sequences. The PCR generated *oriT* fragments were cloned into pSPKm6. One half of the ligation was transformed into DH5 α /pCU1 and the 60 white transformants isolated on Xgal Km Sp plates were spot mated with HB101Rif. Of the four isolates which showed no growth on RifKm, digest of miniprep DNA demonstrated that only one contained an insert. One additional isolate showing sparse growth on RifKm was also shown to contain an insert. Sequencing of these two isolates, *nic3* and *nic16* identified mutations within *nic* (see Figure 4.9).

On the possibility that not all point mutations within the target area would result in a reduction in mobilization that could be detected by spot mating, mini-filter matings were conducted on the 60 DH5 α /pCU1/putative *nic* mutants to increase the sensitivity of the mobilization assay. This identified 10 additional isolates with a 2 to 5 fold reduction in mobility. Of these, four were shown by sequencing to contain mutations (*nic11*, 59, 15 and 45).

The second half of the original PCR-pSPKm6 ligation was transformed into DH5 α producing 272 white transformants on Xgal Km plates. Triparental spot mating with HB101Rif and DH5 α /pCU1 identified 6 isolates with significantly reduced mobilization

Figure 4.9 Point and deletion mutations within the *nic* region of the pCU1 *oriT*.

The middle sequence represents the upper strand of the *nic* region of the pCU1 *oriT*. The 10 bp junction in pMUR/pCU1 hybrids is underlined. The open arrow above the sequence is the second direct repeat, the closed arrow is the left arm of IR₂. The / above the sequence indicates the location of *nic* in R388 as identified by *in vitro* assay with *trwC* (de la Cruz, personal communication).

The top two sequences show the extent of the complementary degenerate ORIT primers used in the PCR reaction. * indicates those bases which were targeted for mutation during synthesis of the primers.

Base substitutions in each mutant derivative are indicated by the letters below the sequence; -- indicates deleted bases. ^ indicates a duplication of the region indicated at the bottom.

The mobilization frequency of each derivative is presented to the right. The frequency is the number of transconjugants receiving the Km marker of the *nic* mutant divided by the number of transconjugants receiving the Sp marker of conjugative helper plasmid pCU1. Values are the average of at least two trials. Most replica trial values varied less than 5 fold from each other, with a maximum variation of 20 fold observed with *nic*238. The mobilization frequency for *nic*224 is not listed due to the presence of a wildtype *nic* sequence in the duplicated region.

of the Km marker (none to several isolated colonies on the spot) and 8 isolates with slightly reduced mobilization (less confluent than controls but no isolated colonies). These 14 isolates as well as 1 isolate from each of 12 spot mating plates showing the greatest degree of reduction in mobility (representing the isolate most likely to contain a mutation from each plate), were sequenced using the pUC reverse primer. Of the 6 isolates showing significantly reduced transfer, two different deletions (nic136 and nic144) and one point mutation were identified (nic238). Of the 8 isolates with slightly reduced transfer, three unique point mutations within the target sequence were identified (nic109, nic111, and nic188) and four unique point mutations outside the target sequence (nic 120, nic121, pSP137 and pSP139) were identified. Of the 12 isolates chosen for their possible reduction in transfer frequency, three containing *oriT* mutations were identified, two with duplicated target sequences (nic224 and nic185) and one with a mutation outside the target sequence which had already been isolated.

A total of 38 isolates was sequenced. Of these, 26 contained some form of mutation, representing a total of 18 unique isolates. Of these 18 unique mutations, 8 were within the target sequence while 10 lay outside and were presumably caused by incorporation errors during PCR. 10 of the 26 isolates carried more than one mutation, but no isolates contained two point mutations within the target region. The 8 mutations within the target region resulted in mutations at 6 of the possible 13 A/T sites in the targeted sequence (5 within the 10 bp nic sequence). Two of these sites had two different mutations. Of the 10 mutations which lay outside the target region, 5 were found in or next to one of the four inverted repeat sequences and are discussed further below (see

Figure 4.10). The other 5 were located in a 7 bp sequence, AGGCGCA, just to the right of the targeted region which appears to be a hotspot for mutation. Two different deletions were generated (nic136 and nic144) and three isolates contained a duplication of the *nic* region as if the two halves of the *oriT* generated in the first round of PCR were joined end to end instead of overlapping (nic120, nic224 and nic45).

Large filter matings of each of the 18 unique mutants were conducted using DH5 α /pCU1/*nic* constructs as the donors in biparental matings with HB101Rif. The results for those within the *nic* region are shown in Figure 4.9. Of these only the two deletion mutants nic136 and nic144 and point mutants nic3, nic16 and nic238 showed a marked reduction in transfer (10^3 - 10^4 fold). Mobilization results of the deletion derivatives and the point mutations lying outside the *nic* region are discussed below (see Section 4.3.3).

Isolates nic16 and nic238 contain transition and transversion mutations of the same A residue. Surprisingly, nic16 is mobilized at a frequency 100 fold lower than nic238, indicating that the transition had greater effect on transfer than the transversion. This suggests that the A/T content of sequence surrounding *nic* may be important. nic3 is a transition two bases to the left of nic16 and nic238. The mutation in nic3 which occurs at the base identified as the 3' base of the *nic* site of R388 (de la Cruz, personal communication) has less effect on mobility than mutations in nic16 which lie two bases to the right suggesting that either pCU1 and R388 have different *nic* sites or that bases on either side of *nic* are less important than those slightly upstream or downstream.

The other five point mutations within the targeted region, resulted in either no

reduction or very little reduction (less than 10 fold) in mobility which may explain why more mutations within the targeted sequence were not identified. It is possible other point mutations were overlooked due to the lack of a detectable phenotype. The mutation in *nic109* resulted in the introduction of an *RsaI* site within the 10 bp *nic* sequence which was used to construct additional deletion mutants pSP6 and pSP7 described in Section 4.3.4.

It should be noted that 6 of the 8 point mutations within the targeted sequence were switches to G residues. This is probably the result of having unequal concentrations of contaminating phosphoamidites during synthesis of the oligonucleotide primers as very small quantities (0.05 ml) were measured with a 1 ml syringe. Another possible explanation is that the phosphoamidites were not all equally active as they have a very short life (2 weeks) on the synthesis machine and were not all reconstituted at the same time. The percentage of conversions of A residues verses T residues is also unequal (6 of 8) suggesting that the concentration of the two PCR products from the first round of PCR added to the second round was not equal.

None of the mutations within the hot spot sequence AGGCGCA located just to the right of the 10 bp *nic* region, resulted in more than a 10 fold reduction in mobility. The presence of a potential 7 bp inverted repeat located directly to the right of this sequence may explain its increased propensity for mutation. Increased incidence in PCR misincorporation directly next to the 3' end of the primer and just upstream from a hairpin loop structure has been described by Loewen and Switala (1995).

The other five PCR generated mutations located outside the targeted sequence

(including pSP8, see below) show only slight if any reduction in mobility (see Figure 4.10). The localization of four of five of these mutations within or near inverted repeat sequences suggests that these structures may also cause the Taq polymerase to be more error prone.

The lack of significant reductions in mobility with all but three of the point mutations within both the targeted region and within inverted repeats outside the *nic* region, argued against further attempts to generate point mutations within the *oriT* region either randomly by PCR or by site-directed PCR mutagenesis. While this approach was useful for detecting the essential *nic* sequence, the probability of point mutations at other sites within *oriT* having significant effects is low. It was at this point that deletion mutagenesis was considered and this is described below.

4.3.3 Deletion mutagenesis of *oriT*

In addition to the two deletion mutations generated during PCR cloning with degenerate oligonucleotides, six other deletion derivatives were constructed.

From the sequence alignment, it is apparent that the structural features common to pCU1, pMUR274 and R388 in the right portion of the *oriT* end after the last IR. Furthermore, the sequence of the *mob* genes downstream from *oriT* in this direction indicates that the *traK* coding region begins three bases to right of the base defining the right most limit of *oriT* in pBP1 and pSP28 (see Section 4.5.4). Between the last IR and the previously defined end of *oriT* there is a *BstYI* site which was used to delete 37 bases from the right end of *oriT*. pSP24 was constructed by cloning the *EcoRI-BstYI* fragment

from pSP25 back into pSPKm6 cut with *EcoRI-BamHI*. When this clone demonstrated no loss in transfer (see Figure 4.10), two additional deletions were made using PCR directed deletion mutagenesis.

Oligonucleotide primers complementary to the right arm of IR₃ (ORIT2) and to IR₄ (ORIT3) were synthesized with a *BamHI* site at the 5' end for cloning (see Figure 4.10). These oligonucleotides were used separately with the pUC' reverse primer to synthesize truncated *oriT* regions by PCR. The PCR generated fragments were digested with *EcoRI-BamHI* and ligated into pSPKm6. The resultant clones were screened by digestion of miniprep DNA and possible clones sequenced to ensure no mutations were introduced during the PCR process. The desired derivatives were named pSP9 (from oligo ORIT3) and pSP10 (from oligo ORIT2). In the case of the cloning involving oligo ORIT3, an isolate carrying a PCR generated mutation was identified. This derivative, pSP8, carries an A->T mutation in IR₃ (see Figure 4.10).

pCU1 was introduced into DH5 α /pSP9 and pSP10 and mobilization frequencies were compared to those of pSP24 (see Figure 4.10). Mobilization of pSP9 at a frequency comparable to that of pSP24 indicates that the minimal pCU1 *oriT* can be reduced to 266 bp. As evidenced by a 10⁴ fold drop in the mobilization frequency of pSP10, further deletions into the right end of the *oriT* have a definite negative effect on transfer.

The second region targeted for deletion analysis is the AT rich region of the pCU1 *oriT*. As seen in the alignment in Figure 4.5, both pCU1 and pMUR274 have a large AT rich region in their *oriT* which is not present in the *oriT* of the IncW plasmid

Figure 4.10 Deletion and point mutations lying outside the *nic* region.

Upper figure. Nucleotide sequence of the 325 bp pCU1 *oriT*. The 10 bp junction sequence containing *nic* is shaded. Point mutations are indicated below the sequence. Primers ORIT1-3 are indicated as lines below the sequence. Restriction sites are in bold letters. Inverted and direct repeats are indicated by arrows above the sequence.

Lower figure. Schematic representation of deletions and point mutations. Central heavy line represents the 325 bp *oriT* region in pSP28. Inverted and direct repeats are shown above the line. Thin lines below the central line represent the extent of the *oriT* sequence present in each derivative. The numbers at the ends of the lines indicate the last nucleotide present on the derivative. Derivatives represented by a single letter are point mutations of pSP24.

Mobilization frequency of each mutant by pCU101 in biparental matings with HB101Rif is shown to the right. Frequencies are given as the number of transconjugants receiving the *oriT* mutant per transconjugant receiving pCU101 and are the average of at least two trials. Replica trial values varied less than 5 fold from each other. *, results reported previously (Paterson and Iyer, 1992).

RsaI ————— *SspI*
 ACTGGTATTCATGCAAGCATGAATACGGAATATTGAACGCTTTT TAGTCA 50
 G G
 nic59 *nic139*
 ————— *SnaBI*
 GTTAGTCCTTAACCTACGTAGATCGCTCCTACCAAAAAAGATGATTTTTT 100

 ATCCCTTAACCTGCTATACACCTAACGCAGCGGCATAAATCACGCTATCA 150
 G C A G
 DraI <---*nic136*--->
 SspI
 CACTGATTTAAAAAATAAAAAATAAATAGCGTCAATATTCGGTAATAAT 200
 C
 nic137
 TAGGTATTAAAAACACCTAATAACTCATTTGTTTTATAATGAGTTTGCTTT 250
 TC ORIT2 A
 pSP8 *nic144* *nic14*
 BstYI
 ATTACTTTGGTAATACTAAGATTTTCAGCTTGTAGATCCCAATCTTTTGT 300
 ORIT3

 TTAAGAAATGCACAATGTGAGGACT 325
 ORIT1

—————
RsaI *SspI* *SnaBI* *nic* *DraI* *SspI* *BstYI* *BamHI*
 100 200 300 pSP28 1.7x10⁻⁶

32	—————	125	pSP12	5.1x10 ⁻⁶
60	—————	125	pSP19	1.6x10 ⁻⁶
100	GCAG	100	<i>nic136</i>	3.3x10 ⁻⁶
112	—————	100	<i>nic144</i>	2.0x10 ⁻⁶
118	—————	100	pSP6	3.6x10 ⁻⁶
117	—————		pSP7	6.0x10 ⁻⁶
183	—————		pBP11	3.5x10 ⁻⁶
143	—————		pSP10	1.5x10 ⁻⁶
166	—————		pSP9	8.3x10 ⁻⁶
159 187	—————	100	pSP29	2.8x10 ⁻⁶
100	—————		pSP24	1.1x10 ⁻⁶
100	—————		<i>nic59</i>	3.1x10 ⁻⁶
100	—————		<i>nic139</i>	2.5x10 ⁻⁶
100	—————		<i>nic137</i>	2.1x10 ⁻⁶
100	—————		pSP8	1.3x10 ⁻⁶
100	—————		<i>nic15</i>	3.1x10 ⁻⁶
100	—————		<i>nic14</i>	2.1x10 ⁻⁶

R388. On the belief that only non-essential sequences are lost or altered significantly, part of the AT rich region of the pCU1 *oriT* was deleted. pSP24 DNA was cut in separate reactions with *DraI-BglI* and *SspI-BglI*. The 500 bp *SspI-BglI* carrying the right end of *oriT* and the 2.6 Kb *DraI-BglI* fragment carrying the left end of *oriT* were isolated by Gene Clean and ligated together (see Figure 4.10). The *oriT* sequence of the resultant clone, pSP29, was confirmed by sequencing ds DNA using the ORIT1 primer (see Figure 4.10).

pCU1 was introduced into DH5 α /pSP29 and mated with HB101Kif. The 10⁴ fold drop in mobilization frequency of pSP29 with respect to pSP24 indicates that the AT rich region plays a significant role in conjugal transfer.

The two deletions mutants generated during PCR cloning with degenerate oligonucleotides nic144 and nic136, show significant reductions in transfer to 10⁻⁴ however this frequency is still 100 fold greater than background (10⁻⁶) supporting the placement of *nic* within the 10 bp pMUR-pCU1 hybrid junction still present in both deletion mutants. The mobilization frequency of nic136 and nic144 is 10 fold lower than the frequency obtained with derivative pSP19 which carries a deletion to the *SnaBI* site. Mobilization of both derivatives at comparable frequencies would suggest that deletion of the direct repeat from nic144 but not nic136 does not adversely affect transfer. However it should be noted that both nic136 and nic144 carry additional PCR generated mutations. nic136 carries the A->C point mutation found in nic11 as well as a three mutations within the AGGCGCA hot spot. By comparison with nic11 and nic121 which carry each of these mutations alone, it can be seen that these mutations alone could

account for the additional 10 fold reduction in mobility observed with *nic136* over *pSP19*. The *nic144* deletion mutant also carries a T->C point mutation in *IR₄*. This point mutation is directly next to the mutation carried by *pSP8* which can be seen to have no effect on transfer. This may suggest that while deletion of the first direct repeat results in a 10 fold reduction in mobilization (*pSP12* vs *pSP19*), deletion of the second direct repeat from *nic144* may result in a further 10 fold reduction.

As indicated above, the mutation in *nic109* converted the ATAC to GTAC which is recognized by the restriction enzyme *RsaI*. *pSP6* was constructed by cloning the right hand portion of the *oriT* from *nic109* on a *RsaI-HindIII* fragment into *pSPKm6* cut with *HindIII-HincII* while *pSP7* contains the left hand portion of *oriT* on a *EcoRI-RsaI* fragment cloned into the *EcoRI-HincII* sites of *pSPKm*. As can be seen from the results in Figure 4.10, neither of these two constructs are mobilized efficiently. *pSP6* is mobilized at a frequency 100 fold less than deletion derivative *nic144* indicating that the loss of the five bases to the right of the *nic144* deletion, the first five bases of the 10 bp *nic* region, reduces transfer to background levels. This could indicate that *nic* lies within these five bases. The frequency of *pSP7*, while as low as *pSP6* is no lower than observed with deletions to the right of *SspI₂* in *pBP11*, or with *pSP29* which carries the deletion in the AT rich region between the *DraI* and *SspI₂* sites and is only 10 fold lower than the small deletion in *pSP10* removing the last inverted repeat. These results suggest that while *nic* plays a significant role in transfer, the right hand portion of *oriT* is equally important.

4.4 SITE-SPECIFIC RECOMBINATION AT *oriT*

4.4.1 Locating *nic* by intramolecular site-specific recombination

As described in Section 4.2, it was possible to localize *nic* to the 10 bp junction in the hybrid pCU1-pMUR *oriT* due to sequence differences between the two parental sequences directly to the left and right of the GCTATACACC junction sequence. This approach was used to further localize *nic* within the 10 bp sequence by examining the sequence of junctions created by site-specific recombination between one wild type pCU1 *oriT* and a second pCU1 *oriT* carrying one of the mutations introduced into the 10 bp *nic* sequence by oligonucleotide PCR mutagenesis (see Figure 4.9).

Double *oriT* derivatives containing one wildtype (wt) pCU1 *oriT* and one mutated (mt) pCU1 *oriT* separated by the Km marker were constructed as outlined in Figure 4.11 and designated pNKN plus the number of the mutant derivative used. Derivatives with the mutated *oriT* in the E position (in the *EcoRI* site of pSP28) which would be nicked first were designated E derivatives while those with the mutated *oriT* in the B position (in the *BamHI* site of pSP28) which would be nicked second are designated the B derivatives. Km^r hybrid *oriT*s formed from E derivatives but not B derivatives should carry the mutation if it lies to the left of *nic*. The opposite would be true of mutations which lie to the right of *nic*, Km^r *oriT* hybrids from the B derivatives would carry the mutation while those from E derivatives would not.

pNKN3E and B and pNKN109E derivatives were constructed from the *nic*3 and *nic*109 *oriT* mutants as outlined in Figure 4.11. pNKN109B derivatives could not be

Figure 4.11 Construction of pNKNE and pNKNB double *oriT* derivatives. pNKNB derivatives were constructed by simultaneously cloning the Tn5 Kanamycin gene (Km, solid line) from plasmid pKan/*oriT* on a *Bam*HI-*Sma*I fragment together with the mutated pCU1 *oriT* (hatched arrow) on a *Pvu*II-*Hind*III fragment into pSP28 cut with *Hind*III and *Bam*HI. pNKNE derivatives were constructed by simultaneously cloning the Tn5 Km gene (solid line) from plasmid pNKN9B on an *Eco*RI-*Hind*III fragment together with the mutated pCU1 *oriT* on an *Eco*RI-*Hind*III fragment into the *Eco*RI site of pDP28. Relevant restriction sites are shown. Ap, ampicillin; Km, kanamycin; Rep, pUC18 replicon.

constructed as discussed further below. Recombinant molecules containing hybrid wt/mt *oriT*s were isolated by triparental mini-filter matings with HB101/pCU101 and HB101Rif followed by screening for Km^r isolates. In order to determine whether the resultant recombinant *oriT* contained the mutated or wildtype sequence, the Km^r isolates were screened in one of two ways. Since the *nic109* mutation introduces an *RsaI* site, the pNKN109Km^r isolates were screened by digesting plasmid DNA with *RsaI*. In the case of the pNKN3Km^r isolates, those carrying mutated *oriT*s were identified by subsequent spot mating with C600Nal since the *nic3* mutation results in a 1000 fold decrease in mobility.

Spot matings with C600Nal demonstrated that all of the 16 pNKN3BKm^r hybrids tested from two separate experiments were unable to transfer efficiently, indicating that they contained the *nic3* mutation. On the other hand, the 16 pNKN3EKm^r hybrids (from two separate experiments) were all able to transfer efficiently confirming that *nic* lies to the left of the mutation in *nic3*. The presence and absence of mutation was confirmed in one pNKN3E and two pNKN3BKm^r isolates by sequencing.

Restriction digestion of pNKN109EKm^r miniprep DNA indicated that all of the 10 hybrids tested from one mating contained an *RsaI* site within the *oriT* region, indicating that they contained the *nic109* mutation and hence that *nic* must lie to the right of the *nic109* mutation. Again, sequencing confirmed the presence of the mutation in two pNKN109EKm^r isolates from two separate tests.

In all cases, sequencing of the Km^r hybrids indicated no alterations in sequence other than the expected mutation. The absence of small (1 to several bases) insertions

or deletions indicates that the mutated *oriT*s are nicked at the same site as the wildtype sequence.

That the *nic109* and *nic3* mutations are present in the *Km^r* recombinants from opposite constructs (E vs B) indicates that *nic* must lie between them. Since there is only one nucleotide between the two mutations, *nic* must lie directly to the right or left of this intervening T residue as illustrated below:

The *nic* site identified for the IncW plasmid R388 matches the right possibility.

In the process of trying to construct pNKN double-*oriT* derivatives it was observed that some derivatives were more difficult to construct than others. As indicated above, I was unable to construct pNKN109B. Four separate attempts were made. In each attempt, derivatives containing two *oriT*s separated by the *Km* gene were obtained, however upon sequencing and/or digestion with *RsaI*, both *oriT*s were found to be wildtype - that is to say that neither *oriT* contained the *nic109* mutation which resulted in the *RsaI* site. A total of 35 transformants were screened from the four cloning attempts. This apparent 'correction' of the second cloned sequence was also observed in the pNKN109E constructs, however, on the second cloning attempt, two out of six clones were obtained which contained one wt and one mutant *oriT*. In the first attempt, all twelve clones tested contained two *oriT* regions but no extra *RsaI* site. It should be noted that pNKN109E and pNKN109B constructs were attempted at the same time with the same DNA and therefore it is unlikely that these results stem from technical errors

during the cloning process.

Similar irregularities were observed in the construction of pNKN238B. Initial sequencing of the pNKN238B and pNKN238E constructs suggested that each contained the mutation in the proper *oriT*. However, sequencing two Km^r hybrids from each construct demonstrated that while the E hybrids both contained the wild type *oriT* sequence as expected from the location of the mutation in *nic238* to the right of *nic*, only one of the two pNKN238B Km^r hybrids contained the mutation as expected. The other contained the wild type sequence. A second set of pNKN238BKm^r isolates were spot mated with C600Nal to identify hybrids containing the mutation from those with wild type sequence (the *nic238* mutation reduces transfer 100 fold, Figure 4.9). Of the 24 Km^r hybrids tested, all except one transferred inefficiently and therefore contained the mutation. The exception, along with two others showing reduced transfer, were sequenced. As expected the transfer proficient isolate contained the wild type *oriT* sequence while the two with reduced transfer contained the *nic238* mutation. These results lead me to conclude that the original pNKN238B construct was either contaminated with low levels of wt/wt construct or that the second mutated *oriT* was being corrected at a low frequency. Again the lack of small deletions or insertions within the sequence indicate that both *oriT*s were nicked at the same site.

Further evidence of the same irregularity was observed in attempts to construct other pNKN derivatives. The first attempt to construct pNKN29B (from deletion derivative pSP29, see Figure 4.10) produced a construct with two wild type *oriT*s, however in a second attempt, all three of the clones obtained contained one mutated and

one wild type *oriT* sequence. Similarly, both pNKN16E and pNKN16B contained two wild type *oriT*s in the single construction attempt. While the reason for this phenomenon is unknown, it would appear from the results presented to be more prevalent in attempts to clone the mutated *oriT* into the position which represents the termination site as evidenced by the difficulties in cloning pNKN109B, 29B, 238B and 16B. It should be stressed that alteration of the mutated *oriT* occurred in the absence of *tra* functions and hence this phenomenon is not the result of inaccurate nicking during transfer.

4.4.2 Identification of initiation and termination functions in the pCU1 *oriT*

The frequency of site-specific recombination between mutant and wild type *oriT* sequences in the double-*oriT* constructs described above can also be used as an indication of whether the mutation affects initiation of transfer or termination (Brasch and Meyer, 1987). Mutations which reduce recombination frequency when located in the *oriT* nicked first (as in E constructs) represent sites important for initiation, while those which reduce recombination when located in the *oriT* nicked second (as in B constructs) represent sites which affect termination.

The frequency of recombination observed with the various pNKN derivatives, as measured by the percentage of Km^r transconjugants, is presented in Table 4.4. As evidenced by the pNKN9B and E constructs in which both *oriT*s contain wild type sequences, termination at the second *oriT* occurs in 91 % of the transferred DNA strands. This is almost 4 time greater than the recombination frequency observed between the pCU1 and pMUR *oriT*s during mobilization by pCU101 (see pMKN2B, Table 2). The

Table 4.4 Mobilization and recombination frequencies of pNKN derivatives.

Double <i>oriT</i> Derivative	Mobilization frequency ^a	Km ^r /Ap ^r	
		Actual	Percentage
pNKN9B ^b	1.4x10 ⁰	183/200	91.5
pNKN9E ^b	9.4x10 ⁻¹	182/200	91.0
pNKN10B	1.5x10 ⁰	190/200	95.0
pNKN10E	1.4x10 ⁻²	7/200	3.5
pNKN29B	6.6x10 ⁻¹	0/200	<0.5
pNKN29E	3.4x10 ⁻²	3/200	1.5
pNKN3B	8.9x10 ⁻²	23/345	6.7
pNKN3E	9.3x10 ⁻²	75/288	26.0
pNKN238B	1.3x10 ⁻¹	181/396	45.7
pNKN238E	1.0x10 ⁻⁰	159/292	40.6
pNKN109B	ND	ND	ND
pNKN109E	1.0x10 ⁻²	99/200	49.5

^a mobilization frequency is given as the number of transconjugants receiving pNKN derivative per donor (not per transconjugant receiving pCU101 helper plasmid since this is a triparental rather than biparental mating). Frequencies are the average of at least two trials. Replica trial values varied less than 5 fold from each other with the exception of pNKN10B in which one of three values was 100 fold higher than the other two.

^b these are positive controls where both *oriT* are wildtype.

difference indicates the effect of non-conserved sequences on termination.

The only mutation that does not affect recombination is the deletion in pSP10 (loss of IR₄ at the right most side of *oriT*) when located in the *oriT* nicked second as in pNKN10B. By comparison, recombination in pNKN10E is reduced significantly suggesting that the far right end of *oriT* is required for initiation but not for termination of transfer. This is consistent with a role in binding of the relaxosome proteins to the *oriT* which would occur in the *oriT* nicked first but not second.

All of the other mutations tested affected both initiation and termination. Of the mutations located close to *nic*, the *nic3* mutation, which affects transfer most significantly, also has the greatest affect on recombination. The results suggest that this site is slightly more important in initiation of transfer than in termination. Surprisingly, *nic238* which reduces transfer 100 fold more than *nic109* shows no greater reduction in recombination. This would suggest that sequences on both sides of *nic* are important to recombination. In the case of the *nic238* mutation, initiation and termination are affected equally. The effect of the *nic109* mutation on termination can not be determined due to the lack of a pNKN109E construct.

The most interesting results are those of the pNKN29 constructs, wherein both initiation and termination appear to be greatly affected by deletion of the AT rich sequence over 40 nt to the right of *nic*. While an explanation can be put forward for the AT rich region to act as part of the relaxosome assembly site or to promote localized denaturation to allow nicking during initiation, neither of these functions would explain its role in termination wherein the nicking enzyme is already bound to the 5' end of the

oriT nicked first, and denaturation would be concurrent with unwinding of the DNA by helicase. These results suggest that the AT sequence may be playing a previously unsuspected role in transfer.

4.5 TRANSPOSON MUTAGENESIS AND SEQUENCING OF THE *mob* REGION

Early Tn5 mutagenesis of pCU1 identified eight complementation regions within the *tra* region (Thatte and Iyer, 1985). Six of these regions (*traA* to *traF*) were shown to be involved in pili formation. It was therefore concluded that the genes involved in conjugal DNA metabolism or mobilization (*mob* genes) must lie within the two remaining complementation regions, *traG* and *traH* which are located in the region extending from the *SmaI* site at approximately 2.8 Kb on the pCU1 map to the end of *tra*. This same region on the closely related plasmid pKM101 has been shown to contain five complementation regions: *fip*, *traH*, *tral*, *traJ* and *traK* (Winans and Walker, 1985₁). All but *fip* (fertility inhibition of P plasmids) are required for transfer (Winans and Walker, 1985₁). Initial sequence information (V. Thatte, unpublished data) from this region suggested that there were sequence similarities between the pCU1 region corresponding to *traJ* of pKM101 and that of *traD* of the F plasmid as well as between the pCU1 regions corresponding to pKM101 *traH* and *tral* and the *tral* of F. The fact that there were no open reading frames of significant size and that the sequence of several small pCU1 ORFs corresponded to the sequence of a single *mob* gene suggested that there were probably four *mob* complementation regions corresponding to those seen in pKM101 but that they were not detected earlier due to errors in the

sequence. To confirm these suspicions, this region was mutagenized with *TnphoA*. These mutants were then used in complementation tests with *Tn5* mutants of pKM101 and the DNA sequence determined by sequencing from the ends of the transposon inserts. It was also hoped that these mutants could be used to demonstrate nicking *in vivo* through the use of site-specific recombination between two *oriT* regions. These experiments are described below.

4.5.1 Construction of pSP34

Rather than attempt to mutagenize a large plasmid such as pCU1 or even the *Tra*⁺ derivative pCU101, a smaller derivative carrying only the proposed *mob* region was used. A previously described *Bal31* deletion derivative of pCU57, pCU57 Δ 14 (Paterson and Iyer, 1992), believed at the time to carry the entire *mob* region (excluding *fip*), was used to construct pSP34 (see Figure 4.12). pSP34 was constructed to provide both a target for transposon mutagenesis of the pCU1 *mob* region as well for use in an *in vivo* nicking assay (see Section 4.6). The pCU1 *mob* region carried by pSP34 extends an estimated 500 bp from the *Bgl*III site at 3.3 Kb on the pCU1 map (see Figure 1.1) to a *Tn5* insert approximately 350 bp past *oriT*. In addition to the N *mob* region, pSP34 carries the IncP *oriT* from RP4, cloned from pKan/*oriT* into the *Bam*HI site of pCU57 Δ 14. The function of the IncP *oriT* is discussed below in the intermolecular recombination assay (see section 4.6).

Figure 4.12 Restriction and functional map of pSP34. pCU1 *mob* region is indicated by the open region with *oriT* represented by an open arrow; direction of the arrow corresponds to the orientation presented in previous figures. The solid region is part of Tn5 cloned with the pCU1 *mob* from a Tn5 derivative of pCU1. The shaded region is the P *oriT* cloned on a *Bam*HI fragment from plasmid pKan/*oriT*. The single line portion of the plasmid is the vector pACYC184 containing the chloramphenicol (Cm) gene and the plasmid replicon (Rep). Relevant restriction sites are indicated.

4.5.2 Mutagenesis of pCU1 *mob* by *TnphoA*

TnphoA was introduced into BW11397/pSP34 by transfection with λ ::*TnphoA*. Isolates carrying pSP34::*TnphoA* were identified by their ability to grow in the presence of high concentrations of kanamycin. Miniprep DNA was isolated from Cm^r Km^r colonies and analyzed by *Hind*III and *Eco*RI digests. 18 unique insertions were isolated, 15 below the *Bgl*II site and three above.

TnphoA was introduced into pET-traHI, a construct carrying the pCU1 region above the *Bgl*II site cloned into the pET23d expression vector (see Section 4.5.5) by passing the plasmid through *E. coli* strain PN1 which carries the transposon in the chromosome. Three unique pET-traHI::*TnphoA* insertions were obtained.

An additional mutant, pSP34 Δ Kpn was constructed by cutting pSP34 at the unique *Kpn*I site, removing the 4 base overhang with T4 DNA polymerase and religating.

4.5.3 Complementation tests between pSP34::*TnphoA* and pKM101::Tn5

Each of the pSP34::*TnphoA* derivatives were introduced into separate strains carrying different pKM101 mutants with Tn5 in each of the four complementation regions identified by Winans and Walker (1985₂). Matings were conducted to determine which of the pSP34 mutants could complement the pKM101 mutants.

The results of the complementation tests are presented in Table 4.5. Complementation of mutations in all four of the pKM101 complementation regions by pSP34 supports the previously held view that pCU1 and pKM101 are closely related. Typically mutations lying in different complementation groups allow transfer of both

Table 4.5 Complementation of pSP34::Tn*phoA* mutants by pKM101::Tn5 mutants.

pSP34 insert	pSP34 transfer freq ^a	pKM101 <i>traH1143</i> transfer freq ^a	pSP34 transfer freq ^a	pKM101 <i>traI1136</i> transfer freq ^a	pSP34 transfer freq ^a	pKM101 <i>traJ35</i> transfer freq ^a	pSP34 transfer freq ^a	pKM101 <i>traK1217</i> transfer freq ^a
58	1.3x10 ⁻¹	1.6x10 ⁻¹	9.6x10 ⁻²	5.5x10 ⁻²				
51	5.0x10 ⁻³	7.8x10 ⁻³	2.7x10 ⁻²	2.2x10 ⁻²				
52	5.5x10 ⁻³	5.0x10 ⁻³	1.1x10 ⁻²	1.4x10 ⁻²	1.0x10 ⁰	7.9x10 ⁻¹	6.2x10 ⁰	1.3x10 ⁰
ΔKpn	6.3x10 ⁻⁴	2.4x10 ⁻³	5.0x10 ⁻⁴	2.1x10 ⁻³	1.2x10 ⁰	1.1x10 ⁰	1.6x10 ⁰	2.0x10 ⁰
16	1.2x10 ⁻¹	1.7x10 ⁻¹	1.2x10 ⁻¹	2.9x10 ⁻¹	2.8x10 ⁻⁴	2.7x10 ⁻⁴	1.6x10 ⁰	2.2x10 ⁰
48					1.0x10 ⁻⁶	1.6x10 ⁻⁵		
11			2.0x10 ⁻¹	4.0x10 ⁻¹	1.1x10 ⁻⁵	1.2x10 ⁻⁵		
47					1.1x10 ⁻³	1.2x10 ⁻³		
12					4.8x10 ⁻⁵	5.8x10 ⁻⁵		
13					3.9x10 ⁻⁵	6.8x10 ⁻⁵		
53					1.8x10 ⁻⁶	6.9x10 ⁻⁷		
56					1.7x10 ⁻⁵	9.0x10 ⁻⁶		
54					3.1x10 ⁻³	4.5x10 ⁻³		
61	1.1x10 ⁻¹	1.6x10 ⁻¹	1.0x10 ⁻¹	3.0x10 ⁻¹	1.5x10 ⁻³	2.3x10 ⁻³	2.5x10 ⁰	2.3x10 ⁰
60					1.0x10 ⁻³	1.0x10 ⁻³	1.8x10 ⁰	1.5x10 ⁰
31							1.9x10 ⁰	1.8x10 ⁰
43							2.8x10 ⁰	2.2x10 ⁰
70					1.3x10 ⁻²	1.4x10 ⁻²	2.0x10 ⁰	1.8x10 ⁰
19	5.1x10 ⁻³	1.8x10 ⁻¹	5.6x10 ⁻²	1.3x10 ⁻¹	2.0x10 ⁻²	1.9x10 ⁻²	3.0x10 ⁰	2.8x10 ⁰
30	2.3x10 ⁻¹	2.7x10 ⁻¹	3.6x10 ⁻¹	6.3x10 ⁻¹	1.1x10 ⁰	2.0x10 ⁰	4.1x10 ⁻⁴	2.3x10 ⁻⁴
pSP34	9.6x10 ⁻²	4.0x10 ⁻¹	8.6x10 ⁻²	3.8x10 ⁻¹	1.6x10 ⁰	1.7x10 ⁰	1.1x10 ¹	2.5x10 ⁰
none	N/A	1.4x10 ⁻⁶	N/A	1.9x10 ⁻⁶	N/A	2.8x10 ⁻⁸	N/A	<5.5x10 ⁻⁸

shaded areas indicated test not done.

^a complementation of pKM101::Tn5 mutants as measured by the transfer frequency of each plasmid. Calculated as the number of recipients receiving the plasmid antibiotic marker (Ap or Cm) divided by the number of donors. Results are the average of at least two trials. Most replica values varied less than 10 fold from each other with a maximum variation of 100 fold in some instances.

plasmids at a frequency of 10^{-1} or better, comparable to the frequency obtained with pSP34. Mutations within the same complementation group resulted in transfer frequencies at least 1000 fold lower than observed with pSP34.

The results clearly identify fourteen pSP34 mutants which are unable to complement the pKM101*traJ* mutant. Selected putative pSP34*traJ* mutants were able to complement the other three pKM101 mutants. One interesting observation is the ability of the *traJ* mutants with inserts closest to the *oriT* to complement the pKM101 *traJ* mutant to a greater extent than those with inserts closer to the *Bgl*II site. Once the sequence had been determined (see Section 4.5.4), it was found that the amino terminus is proximal to the *oriT*. Considering the supposedly polar effect of transposon mutations the variation in complementation results would seem backwards as one would expect amino terminal inserts to exert at least as great an effect as carboxyl terminal inserts.

Only one pSP34::Tn*phoA* mutant (#30) was unable to complement the pKM101*traK* mutant. This putative pSP34*traK* mutant was able to complement each of the other three pKM101 mutants.

The results obtained with the pKM101*traH* and *traI* mutants were not as clear. Unfortunately only three Tn*phoA* inserts in the *mob* region above the *Bgl*II site were isolated and they were all located within 600 bp of the *Bgl*II site, placing them all in the region corresponding to the pKM101*traI* region. The pSP34ΔKpn mutant should lie in the pCU1 region corresponding to the pKM101*traH* region. Despite this, all of these four derivatives produced similar complementation results with both the pKM101*traH* and pKM101*traI* derivatives. In all cases, the transfer frequencies were only reduced 10-

1000 fold unlike the *traJ* and *traK* results in which transfer frequencies could be reduced 10^4 and 10^5 fold. These results do not provide conclusive evidence as to whether there are one or two complementation regions above the *Bgl*II site. However, both the sequence of this region in pKM101 (M. More and S. Winans, personal communication) discussed in Section 4.5.4, and protein expression experiments presented in Section 4.5.5, suggest that this region encodes one large protein.

4.5.4 Analysis of the nucleotide sequence of the *mob* region of pCU1

DNA was isolated from each of the pSP34::Tn*phoA* isolates and sequenced using the Tn*phoA* and/or Tn5 primer (in the latter case, a DNA fragment containing only the right end of the Tn*phoA*, and hence only one Tn5 primer site, was extracted from an agarose gel via Gene Clean). The sequence of both strands of DNA below the *Bgl*II site were determined using these transposon inserts together with four internal primers as outlined in Figure 4.13. The nucleotide sequence of this region together with the amino acid sequence of potential protein coding regions is presented in Figure 4.14.

Two open reading frames (ORF) were identified within the sequenced region; both reading in the same direction away from the *oriT*. The ORF closest to *oriT* corresponds to the *traK* complementation region and would encode a 139 aa protein with an estimated molecular weight of 15.3 kDa. The ORF begins 61 bp from the end of *oriT*. The only ATG start codon is located six bases downstream from a putative Shine Dalgarno ribosomal binding sequence (RBS). The upstream region contains several possible promoter sequences, some of which overlap the *oriT* region.

Figure 4.13 Strategy for sequencing the *mob* region of pCU1. Upper line composed of four arrows represents the 3 Kb region extending from the *oriT* in pSP34. The open arrows correspond to the *oriT*, *traK*, *traJ* and the 5' end of the *traHI* open reading frames (ORF). *TnphoA* inserts are indicated by numbers above the line. Location of the four *mob* primers are indicated by solid arrowheads above the line. The extent and direction (strand) of sequence obtained from each *TnphoA* derivative or primer are indicated by the thin arrows below the main line. *TnphoA* derivatives were sequenced using the *TnphoA* primer unless indicated by Tn5 following the derivative designation.

pBE42
pho30Tn5
mob4
pho19
pho43
pho70
pho31
mob6
pho60Tn5
pho61Tn5
pho13
pho16
pho58
pho52
pho51
pho30
mob3
pho19Tn5
pho31Tn5
mob5
pho60
pho61
pho47
pho48
mob2

Figure 4.14. DNA sequence and putative ORFs of pCU1 *traK* and *traJ* regions. Sequence extends from the end of *oriT* (bold letters) to the *Bgl*II site to include the entire *traK* and *traJ* coding regions and the start of the *traHI* coding region. Putative promoters upstream of the *traK* ORF are indicated. The location of Tn*phoA* insertions are indicated ▼ ; the transposon is oriented such that the *phoA* gene (IS50L) is on the side of the arrowhead with the insert number. RBS, putative ribosomal binding sites are indicated. Relevant restriction sites are indicated.

RsaI *SspI*
 GTACTGGTATTTCATGCAAGCATGAATACGGAATATTGAACGCTTTTTAGTCAGTTAGTCC 60

 TTAACCTACGTAGATCGCTCCTACCAAAAAAGATGATTTTTTATCCCTTAACCTGCTATA 120

 CACCTAACGCAGCGGCATAAATCACGCTATCACACTGATTTAAAAAATAAAAAATAAAT 180

 AGCGTCAATATTTCGGTAATAATTAGGTATTAAAAACACCTAATAACTCATTGTTTTATAA 240

 TGAGTTTGCTTTATTACTTTGGTAATACTAAGATTTTTCAGCTTGTTAGATCCCAATCTTTT 300

 GTTTAAGAAATGCACAATGTGAGGACTGAATGCCAATAATAACGCAAAAGTATCAGATG 360
 M P I I T A K V S D
 AACTTTTGGCCTATATAGACCTGGTTTCAGGAGGTAATCGGTCAGATTATCTGCGGCGCT 420
 E L L A Y I D L V S G G N R S D Y L R R
 GCCTTGAAGCTGGCCCAGGTGATCGGGAGTCCGGGTTGAAAATAGTTGCCGACCGGCTGA 480
 C L E A G P G D R E S G L K I V A D R L
 GCGACGTAAATCGCAAACCTGGATTATCTTTTGACCGTGCTTCAGATGCGGATTTTGGCC 540
 S D V N R K L D Y L F D R A S D A D F G
 CACTGCGCGATGAGCTGAAGGCAATAACGAAACGCTATCTGGCGTAAATTTCCACCGG 600
 P L R D E L K A I T D T L S G V K F P P
 CAGGACAGATGATGTTGCATGAATCACTCGCTATTGAAACGCTAATCCTCCTGCGCTCGA 660
 A G Q M M L H E S L A I E T L I L L R S
 TAGCAGAACCGGGCAAACCAAAGCGGCAAAGCAGAGGTTGAACGTAACGGCTACAAAG 720
 I A E P G K T K A A K A E V E R N G Y K
 TCTGGGAACCAAAAAGGAGCGCTGAAATGGACGATAGAGAAAGAGGCTTAGCATTTTTA 780
 V W E P K K E R * M D D R E R G L A F L
 TTTGCAATTACTTTGCCTCCAGTGATGGTATGGTTTCTAGTTGCAAATTTACCTACG 840
 F A I T L P P V M V W F L V A K F T Y G
 ATTGATCCATCCACGGCTAAATACCTGATTCCGTATCTGGTTAAGAATACTTTTACCTA 900
 I D P S T A K Y L I P Y L V K N T F S L
 TGGCCTTTATGGTCAGCTTTAATTGCAGGCTGGTTTATTGGTGTGGCGGTCTGATCGCT 960
 W P L W S A L I A G W F I G V G G L I A
 TTTATCATTTATGATAAATCACGCGTGTTTAAAGGCGAAAGATTCAAAAAGATTTATCGT 1020
 F I I Y D K S R V F K G E R F K K I Y R
 GGTACAGAGCTTGTTTCGCGCCAGAACAACCTCGCTGATAAAACACGCGAAAGAGGTGTCAAC 1080
 G T E L V R A R T L A D K T R E R G V N
 CAGTTAACCGTGGCTAATATCCCCATACCTACATACGCTGAGAACTTGCAATTTTTCGATT 1140
 Q L T V A N I P I P T Y A E N L H F S I

*Rsa*I *Rsa*I
GCCGGTACAAACCGGTACTGGTAAAACCACAATTTTCAATGAACTGTTATTTAAGAGCATC 1200
A G T T G T G K T T I F N E L L F K S I
*Ssp*I
ATTAGAGGCGGCAAAAATATTGCTTTAGATCCAAATGGGGGGTTCTTAAAGAATTTCTAT 1260
I R G G K N I A L D P N G G F L K N F Y
*Dra*I
CGTCCCGGCGATGTTATTTTAAACGCCTATGATAAACCECACTGAAGGCTGGGTGTTTTTC 1320
R P G D V I L N A Y D K R T E G W V F F
AATGAAATTCGCCGTTTCATATGATTACGAGCGCTTAGTGAACTCTATTGTTTCAGGAAAGC 1380
N E I R R S Y D Y E R L V N S I V Q E S
CCTGATATGGCTACTGAAGAATGGTTGGCCTATGGCCGTCTTATTTTTAGTGAAGTTTCG 1440
P D M A T E E W L A Y G R L I F S E V S
60 ↓
AAAAAAGCTTACAGCCTATACAGCACAGTAACTATGGAAGAAGTTATTTCACTGGGCCTGT 1500
K K L H S L Y S T V T M E E V I H W A C
AACGTTGACCAGAAAAAATTAAAGAATTTTAAATGGGGACGCCTGCCGAAGCTATTTTTT 1560
N V D Q K K L K E F L N G T P A E A I F
61 ↓
TCCGGGTCTGAAAAAGCAGTTGGAAGCGCGCGATTGTTCTCAGTAAGAATCTTGCCCCA 1620
S G S E K A V G S A R F V L S K N L A P
CATTGAAAATGCCGGAAGGTAATTTTTCCCTGCGTGACTGGCTTGATGATGGAAAGCCG 1680
H L K M P E G N F S L R D W L D D G K P
54
GGAACCCTGTTTATCACCTGGCAGGAAGAATGAAAAGGTCACCTTAATCCGCTAATTTCC 1740
G T L F I T W O E E M K R S L N P L I S
TGCTGGCTGGATTTCGATTTTTTCTATCGTGCTGGGTATGGGTGAAAAAGAAAGCCGCATT 1800
C W L D S I F S I V L G M G E K G S R I
↓ 13 ↓ 56
AATGTATTTATTGACGAGCTGGAATCACTCCAGTTTCTGCCAAACCTCAACGATGCACTG 1860
N V F I D E L E S L Q F L P N L N D A L
47 ↓
ACCAAAGGGCGTAAAAGCGGTCTGTGTGTTTATGCTGGCTATCAAACCTATTCTCAGCTG 1920
T K G R K S G L C V Y A G Y Q T Y S Q L
↓ 12 11 ↓ 48 ↓ ↓ 16
GTTAAGGTTTATGGTCGGGATATGGCTCAGACCATCTGGCTAACATGCGTTCTAACATC 1980
V K V Y G R D M A Q T I L A N M R S N I
GTGCTGGGCGGCACGGCTCTCGGTGATGAAACGTTGGATCAAATGTCGCGCTCACTCGGT 2040
V L G G T A L G D E T L D Q M S R S L G
GAGATAGAAGGCGAAGTTGAGCGTAAAGAATCCGATCCTCAGAAGCCCTGGATTGTCCGT 2100
E I E G E V E R K E S D P Q K P W I V R
AAACGCCGCGACGTTAAAGTTGTTTCGTGCCGTAAACGCCTACCGAAATATCAATGTTGCCA 2160
K R R D V K V V R A V T P T B I S M L P
AACCTCACCGGCTATCTGCGTTGCCTGGTGATATGCCCGTCGCTAAGTTCAAGGCTAAAC 2220
N L T G Y L R C L V I C P S L S S R L N
RBS_ *Bgl*II
ACGTTAAATACCATCGCAAAAACCGTTCCTGGCATTGAACTGAGGGAGATCTGATGCT 2280
T L N T I A K T L P L A L N . M L

However, pET23-traH*phoA2* also shows a protein larger than predicted by the nucleotide sequence. While this could reflect inaccuracies in measuring large proteins by SDS-PAGE, the fact that both proteins run significantly slower than the protein size standards of comparable size could indicate that the carboxyl terminus of the TraH protein which is present in the pET23-traH and pET23-traH::*phoA2* isolates may be modified post translationally.

In the SDS-PAGE in Figure 4.16, both pET-traH::*pho3* and *pho7* show less intense proteins which are 10-15 kDa smaller than the main protein band. Similar bands have been observed on other gels. The presence of these smaller bands could represent a restart protein in which translation is reinitiated approximately 300 bases downstream of the predicted start codon 1 base downstream of the *Bgl*III site. There is a potential ATG start codons 281 bp downstream of the putative start codon of the *traH* sequence which could act as a secondary start site.

4.5.6 Amino acid sequence analysis of pCU1 *mob* proteins

An alignment of the TraK amino acid sequence with that of the IncW R388 TrwA protein and the F TraY protein demonstrates 26% and 14% sequence homology respectively (Figure 4.17). Amino acid sequence identity between TraK and TrwA is greatest in the carboxyl terminus. While TraK shows remnants of the leucine zipper motif seen in TrwA, there is little evidence of such a structure in TraY.

An alignment of the TraJ amino acid sequence with that of the F plasmid TraD and the R388 plasmid TrwB proteins is seen in Figure 4.18 and demonstrates 40%

4.5.5 Expression of *mob* proteins

The *mob* region of pCU1 was cloned into the expression vector pET23d in three pieces. The 5' end of each pET23d-*tra* junction as determined by DNA sequencing, is shown in Figure 4.15. In the first construct, pET-*traK*, the entire putative *traK* coding region was cloned on a *BsrYI-HpaI* fragment from pSP34 into the *BamHI-HincII* sites of pET23d. The pET-*traK* construct contains its native RBS and hence should produce a native TraK protein. The second construct, pET-*traJ*, contains the carboxyl terminus of the *traJ* region cloned on a *HpaI-HindIII* fragment from pSP34 into the *HincII-HindIII* sites of pET23d. As depicted in Figure 4.15, this portion of the *traJ* coding region is in frame with the pET23d coding region which should produce a fusion protein truncated at the N terminal. To construct pET-*traHI*, the entire *traHI* coding region on a *BglII-HindIII* fragment was cloned from pSP34 into the *BamHI-HindIII* sites of pET23d. This construct lacks the putative *traHI* RBS, however, the sequence of the junction clearly indicates that the *traHI* coding region is not in frame with the pET23d coding region and hence the weak RBS (gga) five bases upstream of the putative ATG start codon of the *traHI* coding region within the *BglII* site must function adequately.

SDS-PAGE of the proteins expressed from these three constructs upon induction with IPTG is presented in Figure 4.16. A 14.3 kDa protein estimated from the gel for pET-*traK* corresponds well to the 15.3 kDa predicted from the nucleotide sequence. It should be noted that it is possible that this band represents the amino terminus of the TraJ protein which is also present in this construct and would result in a peptide with a molecular weight of 16.6 kDa. Despite the overall faintness of the protein bands in the

Figure 4.15 5' end of pCU1 *ira* genes cloned into the pET23d expression vector. Nucleotide sequences are shown above with corresponding amino acid sequence below. The pCU1 nucleotide sequence is shown in lower case letter, the pET nucleotide sequence is given in upper case letters. The amino acid sequences derived from pET23d are shown in normal font, while the amino acid sequence derived from pCU1 is italicized. Restriction sites and ribosomal binding sites (RBS) are indicate in bold. Putative ATG start codons are underlined.

Figure 4.16 SDS-PAGE of pCU1 Mob proteins. Molecular weights of protein standards (lanes 1) are indicated to the left. Molecular weights of unique protein bands corresponding to expressed Mob proteins are indicated to the right (see below). Bands labelled with * indicate secondary products.

lane	Plasmid construct	Level of IPTG induction (mM)	MW of Mob protein from nucleotide sequence (kDa)	MW of Mob protein(s) from SDS-PAGE (kDa)	
2	pET23d	0.4	none	none seen	
3	pET23-traK	0.4	15.3	none seen	
4	pET23-traK	0.05	15.3	14.3	
5	pET23-traJ	0.4	47	46.5	44.5*
6	pET23-traHI	0.4	123	150	
7	pET23-traHI::Tn ϕ hoA2	0.4	104	128	
8	pET23-traHI::Tn ϕ hoA3	0.4	58.7	56	45*
9	pET23-traHI::Tn ϕ hoA7	0.4	41.7	45	29*

pET-TraK sample induced with 0.4 mM IPTG (Figure 4.16, lane 3), repeated attempts to induce the *traK* gene with 0.4 mM IPTG produced no visible 14.3 kDa protein band. However, when low levels of IPTG (0.05 mM) were used (Figure 14.6 lane 4) the 14.3 kDa band was visible. While the simplest explanation for these results is that the over expression of the *traK* gene results in cell death and selection of plasmidless cells over time, a second possibility is that the *traK* gene negatively autoregulates itself.

The pET-traJ construct is seen to express a 46.5 kDa protein which corresponds well to the molecular weight of 47 kDa predicted for the TraJ carboxyl fusion protein. The abundance of this protein relative to that of the TraK or TraHI proteins most likely reflects the elevated efficiency of the pET23d ribosomal binding site over the native ribosomal binding sites of the *tra* genes. There appears to be a smaller, less intense protein band below the heavier one. This band corresponds to a 44.5 kDa protein. Expression of a protein this size would require a start codon just downstream of the *HpaI* site. As seen in Figure 4.15, there is a TTG rare start codon 39 bp from the *HpaI* site.

The final construct, pET-traHI, expresses a large protein with a MW of 150 kDa. While this clearly indicates that the *traHI* region encodes one protein as suggested by the nucleotide sequence rather than two small proteins as suggested by the complementation experiments of Winans and Walker (1985₂), the observed size is significantly larger than the 123 kDa predicated from the nucleotide sequence of the *traHI* region carried by pSP34 together with the 42 amino acids from the pET23d vector. Truncated TraHI proteins produced by pET-traHI::Tn*phoA* derivatives pET23-traHI*phoA*3 and 7 show good agreement with the size predicted from the nucleotide sequence (see Figure 4.16).

However, pET23-traHI*phoA2* also shows a protein larger than predicted by the nucleotide sequence. While this could reflect inaccuracies in measuring large proteins by SDS-PAGE, the fact that both proteins run significantly slower than the protein size standards of comparable size could indicate that the carboxyl terminus of the TraHI protein which is present in the pET23-traHI and pET23-traHI::*phoA2* isolates may be modified post-translationally.

In the SDS-PAGE in Figure 4.16, both pET-traHI::*pho3* and *pho7* show less intense proteins which are 10-15 kDa smaller than the main protein band. Similar bands have been observed on other gels. The presence of these smaller bands could represent a restart protein in which translation is reinitiated approximately 300 bases downstream of the predicted start codon 1 base downstream of the *Bgl*II site. There is a potential ATG start codons 281 bp downstream of the putative start codon of the *traHI* sequence which could act as a secondary start site.

4.5.6 Amino acid sequence analysis of pCU1 *mob* proteins

An alignment of the TraK amino acid sequence with that of the IncW R388 TrwA protein and the F TraY protein demonstrates 26% and 14% sequence homology respectively (Figure 4.17). Amino acid sequence identity between TraK and TrwA is greatest in the carboxyl terminus. While TraK shows remnants of the leucine zipper motif seen in TrwA, there is little evidence of such a structure in TraY.

An alignment of the TraJ amino acid sequence with that of the F plasmid TraD and the R388 plasmid TrwB proteins is seen in Figure 4.18 and demonstrates 40%

Figure 4.17 Amino acid sequence alignment of pCUI TraK (middle), R388 TrwA (upper) and F TraY (lower). Matches between TraK and TrwA or TraY are indicated by |; Leucine residues in putative leucine zipper motif in TrwA and corresponding residues in pCUI are underlined. Alignment prepared by Clustal V, version 3.52 (Higgins and Sharp, 1989) using default parameters.

TrwA	M-----AL	GDPIQVRLSP	EKQALLEDEA	ARKGKRLATY	LRELLES---	40
TraK	M-----PI	---ITAKVSD	E---LLAYID	LVSGGNRSDY	LRRCLEAGPG	37
TraY	MKRFGTRSAT	GKMVKLKLPV	DVESLLIEAS	NRSGRSRS--	----FEAV--	42
TrwA	--ENDL-----	-----	-----QG	ELAALRREV	SLHHVIEDLA	66
TraK	DRESGLKIVA	DRLSDVNRKL	DYLFDRASDA	DFGPLRDE--	-LKALTETLS	84
TraY	---IRLKDHL	HRYPKFNRA-	GNIYGK-SLV	KYLTMRLLDDE	TNQLLIAAKN	87
TrwA	DTGLRSDQSG	PGQNAVQIET	LLLLRAIAGP	ERMKPVKGEL	KRLGIEVWTP	116
TraK	GVKFPPADRM	MLHESLAIET	LILLRSIAEP	GKTKAAKAEV	ERNGYKVWEP	134
TraY	RSGWCKTDEA	ADR---VIDH	LIKFPDFYNS	EIFREADKE-	EDITFNTL--	131
TrwA	EGKED					121
TraK	K-KER					138
TraY	-----					131

sequence identity with TrwB and 23% identity with TraD. Both the type A and type B nucleotide binding sites seen in TraD and TrwB are conserved in TraJ (aa 126-138 and 350-355 respectively). The hydrophobicity plot seen in Figure 4.19 demonstrates that there are two prominent hydrophobic regions spanning the first 80 aa of the amino terminus of the TraJ protein which correspond with similar domains in both the R388 and F analogues. Although the hydrophobic F sequences are slightly larger and spaced further apart the sequences align with the hydrophobic regions of TraJ and TrwB.

The amino acid sequence alignment between TraHI and TrwC of R388 and TraI of the F plasmid is shown in Figure 4.20. TrwC and TraHI show 41% sequence identity while TraI and TraHI show only 25% identity. TraI is much larger than either TraHI or TrwC and the alignment shows that the two smaller proteins align with two separate regions of the larger. The first region covers the first 342 amino acids of TraI while the second spans from amino acid residue 969 to approximately 1469. This leaves a central portion of approximately 600 amino acids as well as 300 amino acids in the carboxyl portion of TraI with no comparable sequences in the smaller two proteins. The type A nucleotide binding site observed in the F TraI and R388 TrwC is conserved in TraHI. TraHI contains two sets of two tyrosine residues (amino acids 18/19 and 26/27) reminiscent of the IncP-type motif I which align with identical residues in both TraI and TrwC. There is a histidine rich sequence (aa 49-67) similar to motif III of the IncP TraI but no evidence of a motif II like sequence.

Figure 4.18 Amino acid sequence alignment of pCU1 TraJ (middle), R388 TrwB (upper) and F TraD (lower). Matches between TraJ and TrwB or TraD by |: Type A and type B nucleotide binding sites conserved in all three sequences are indicated by lines over the sequences. RGD sequence found in TrwB and TraD is indicated by a line under the sequences. Alignment prepared by Clustal V, version 3.52 using default parameters.

TrwB	MH-----	-----PDDQR	KVS-----	---AGIVIVL	P-LIFWITAV	QKTEVLG---	33
TraJ	MD-----	-----DRER	GLA-----	---FLFAITL	PPVMVWFLVA	KFTYGDID---	33
TraD	MSFNAKDMTQ	GGQIASMRIR	MFSQIANIML	YCLFIFFWIL	VGLVLWIKIS	WQTFVNGCIY	60
TrwB	-----S	P--KLLALWE	LM---KLTPQ	KPILLLSALG	G-----	---L--AVGVL	66
TraJ	-----P	STAKYLIPYL	VK---NTFSL	WP--LWSALI	AGW-----	-FI--GVGGL	67
TraD	WWCTTLEGMR	DLIKSQPVYE	IQYYGKTFRM	NAAQVLHDKY	MIWCSEQLWS	AFVLAAVVAL	120
TrwB	FV-----	-WLLNSVGQG	EPGGAPFK-R	FLRGTRIVSG	GKLKRMTRER	A-KQVTVAGV	115
TraJ	IA-----	-FIIYDKSRV	-FKGERFK-K	IYRGTELVR	RTLADKTRER	GVNQLTVANI	116
TraD	VICLITFFV	SWILGRQGKQ	QSENEVTGGR	QLTDNPKDVA	RMLKKDGKDS	DIR---IGDL	177
TrwB	PMPRDAEPRH	LLVNGATGTG	KSVLLRELAY	TGLLRGDRMV	IVDPNGDMLS	KFGRDK-DII	174
TraJ	PIPTYAENLH	FSIAGTTGTG	KTTFNELLF	KSIIRGGKNI	ALDPNGGFLK	NFYRPG-DVI	175
TraD	PIIRDSEIQN	FCRHGTVGAG	KSEVIRRLAN	YARQRGDMVV	IYDRSGEFVK	SYDPSIDKI	237
TrwB	LNPDYDQRTG	WSFFNEIRND	YDQRYALSV	VPRGKTDEAE	EWASYGRLLL	RETAKKL-AL	233
TraJ	LNAYDKRTEG	WVFFNEIRRS	YDYERLVNSI	VQESPDMAE	EWLAYGRLIF	SEVSKKLHSL	235
TraD	LNPLDARCAA	NDLWKECLTQ	PDFDNTANTL	IRDGKTEDPF	-WQSGSRTIF	AEAAYLMRN-	295
TrwB	IGTPSMRELF	HWTTIATFDD	LRGFL----	E GTLAESLFAG	SNEASKALTS	ARFVLSDKLP	289
TraJ	YSTVTMEEVI	HWACNVQKK	LKEFL----	N GTPARAI FSG	SE---KVGGS	ARFVLSXNLA	288
TraD	DPNRSYSKLV	DTLLSIKIEK	LRTYLVIHRO	ANLVEEKIEK	TAISIRAVLT	N-YVKAIRYL	354
TrwB	EHVTMPDGD	SIRSWL----	EDPNGGNLFI	TWREDMGPAL	RPLISAWVDV	VCTSILSLPE	345
TraJ	PHLKMPGNGF	SLRDWL----	DDGKPGTLFI	TWQEEMKRSL	NPLISCWLDS	IFSIVLGMGE	344
TraD	QGIEHNGEPP	TIRDWMRGVR	EDQKNGWIFI	SSNADTHASL	KPVISMWLSI	AIRGLLAME	414
TrwB	EPKRRLWLF	DELASLEKLA	SLADALTKGR	KAGLRVVAGL	OSTSOLDDVY	GVKEAOTLRA	405
TraJ	KGSR-INVFI	DELESLOFLP	NLNDALTKGR	KSGLCVYAGY	QTYSQLVKVY	GRDMAQTILA	403
TraD	NRNRRVWFFC	DELPTLHKLP	DLVEILPEAR	KFGGCVVFGI	QSYAQLEDIY	GEKAAASLFD	474
TrwB	SFRSLVVVGG	SRTDPKTNE	MSLSLGEHEV	ERDRYSKNTG	KHHS-TGRAL	ERVRRERVMP	464
TraJ	NMRSNIVVGG	SRLGDETLDQ	MSRSLGEIEG	EVERKESDPQ	KPW1-VRKRR	DVKVVRVATP	462
TraD	VMTTRAFFRS	PSHKIAEFAA	GEIGFKEHLK	ASEQYSYGAD	PVRDGVSTGK	DMERQTLVSY	534
TrwB	AEIANLPDLT	AYVGFAGNRP	IAKVPLEIKQ	FANRQPAFVE	GTI-----	-----	507
TraJ	TEISMLENLT	GYLRLVICP	SLSSRLNT--	LNTIAKTLFL	ALN-----	-----	503
TraD	SDIQSLPDLT	CYVTLPGPYP	AVKLSLKYQT	RPKVAPEFIP	RDINPEMENR	LSAVLAAREA	594
TrwB	-----	-----	-----	-----	-----	-----	
TraJ	-----	-----	-----	-----	-----	-----	
TraD	EGRQMASLFE	PDVPEVVSSE	DVTQAEQPPQ	PVSPAINDKK	SDS-VNVPAG	GIEQELMKMP	654
TrwB	-----	-----	-----	-----	-----	-----	
TraJ	-----	-----	-----	-----	-----	-----	
TraD	EEEMEQQLPP	GISESGEVVD	MAAYEAWQOE	NHPDIQQQMQ	RREEVNINVH	RRERGEDVEPG	714
TraJ	---	---	---	---	---	---	507
TrwB	---	---	---	---	---	---	503
TraD	DDF	---	---	---	---	---	717

Figure 4.19 Comparison of hydrophobicity plots of pCU1 TraJ (lower), R388 TrwB (middle) and F TraD (upper). Area above the dotted line is hydrophobic; area below the dotted line is hydrophilic. Prepared using Grease program with a window size of 7 (Kyte and Doolittle, 1982).

Figure 4.20 Amino acid sequence alignment of pCU1 TraHI (middle), R388 TrwC (upper) and F Tral (lower). Matches between TraHI and TrwC or Tral by |: the tyr residues of the P type are indicated as are the P type motif III and the type A nucleotide binding site. Location of the pSP34::Tn*phoA* inserts (▼) and pET23d-traHI::*phoA* inserts (▽) are indicated above the three sequences. Putative restart codons in the pCU1 sequence (see Discussion) are indicated by ✱ above the three sequences. Alignment prepared by Clustal V, version 3.52 using default parameters.

				motif I					
T1WC	MLSHMVLTRQ	DIGRAASYYE	DGADDDYYAKD	GDASEWQGGK	AEELGLSGEV	DSKRFRELLA		60	
T14HI	MLLITTTITRQ	NVTSVVGYYYS	DAKDDYYYSKD	SSFTSWQGGTG	AEALGLSGDV	ESARFKELLV		60	
T14I	MMSIAQV -R	SAGSAGNYTT	D-KDNYVVLG	SMGERWAGRG	AEQLGLQGSV	DKDVFTRLLE		57	
T1WC	GNIGEGHRIM	RSATRODSKE	RIGLDLTFSA	PKSVSLQALV	AGDAEIIKAH	DRAVARTLEQ		120	
T14HI	GEIITFTTHMC	RHVG-DAKKE	RLGYDLTFSA	PKGVSMQALI	HGDKTIIIEAH	EKAVAAAVRE		119	
T14I	GKLPDGDALS	R-MQDGSNRH	RFGYDLTFSA	PKSVSMAMML	GGDKRLIDAH	NQAVDFAVRQ		116	
	▼58		▼52	motif III					
T1WC	AEARQAQAKK	IQGKTRIEET	GNLVIGKFRH	ETSREDDPOL	HTHAVILNMT	KRSDGQWRAL		180	
T14HI	AEKLAQAATT	ROGKSVTQNT	NNLVVATFRH	ETSRALEDDL	HTHAFVMNMT	QREDGQWRAL		179	
T14I	VEALASTAVM	TUGQSETVLT	GNLVMALFNH	QTSRDQEPOL	HTHAVVANVT	QH NGEWKT		175	
				▼51					
T1WC	KNDEIVK -	- - - -RTRY	LGAVYNAELA	HELQKLGYYL	RY-GKDGNFQ	LAHIDRQQIE		210	
T14HI	KNDELMR -	- - - - -NKM	LGDVYQELA	LELAKAGYGL	RYNSKNNTFD	MAHFSDEQIR		210	
T14I	SEDKVGKTGF	LENVYANQIA	FGRLYEKLLK	EJVEALGYET	EVVGKHHMWE	MPGVF - - -VE		207	
T1WC	GFSKRTEGTA	EWYAARGLDP	NSVLELKQA	AKVLSRAKKT	SVDREALRAE	WQATAYFSGI		290	
T14HI	AFSRRSEQIE	KGLAAMGLTK	ETADAQTKSK	VSMATREKKT	EHSREEIHQQ	WASRAKTLGI		290	
T14I	AFSGRSQTIR	EAVG - - -	EDASLKRSDV	AALDTRKSKQ	HVDPEIKMAE	WMQTLKETGF		286	
T1WC	DF - - -SRRE	WSGREK - -	GSSEKQAHSE	MPSDEAA -	KRAVRYAI	NHLTER - - -		333	
T14HI	DF - - -DNRE	WQGHGKPLEA	DIARNMADDF	TSFEVKA - -	-DRAIQFAV	KSLSER - - -		337	
T14I	DIRAYRDAAD	QRADLFLLT	GPASQDGFEN	QCAVTQAIAG	LSERKVQFTY	TDVLARTVGI		346	
T1WC			ERELVDATMK		HQ VQARLEDIC	KELLR - - -		366	
T14HI			REKLIQIANF		- - - - -V	LGHATIADVE	KAYLKA - - -	370	
T14I	LPPENGVIER	ARASIDEAIS	RELLIFLDRE	KSLFTSGIHV	LDELSVRALS	RDIMKQNPVT		406	
T14I	VHFEKSVPR	ACYSQAVSVL	AQIFPSLAIV	SGQGBAAGSK	ERVAVELVMA	REGGREVCII		466	
T1WC					TETG - - -	YLIREAI		577	
T14HI					V - - - -	ATIESEA		581	
T14I	AAWRKQMNM	KQTERLGGEL	ITDRQLLES	MAFTPGSTVI	WVGEKLSLK	ETLTILDGAA		616	
T1WC	K							578	
T14HI	E							581	
T14I	ABNYVLTII	SGKATDTGSA	LMAMKIASIN	TYRWGGSEF	PATIISEPIA	NRYAKLASL		586	

				motif I							
TrwC	MLSHMVLTRQ	DIGRAASYYS	DGADDYAKD	GDASEWQGGK	AEELGLSGEV	DSKRFRELLA					60
TraHI	MLDITITITRQ	NVTSVVGYS	DAKDDYYSKD	SSFTSWQGTG	AEALGLSGDV	ESARFKELLV					60
TraI	NMSIAQV--R	SAGSAGNYIT	D-KDNYVLG	SMGERWAGRG	AEQLGLQGSV	DKDVFTRLLE					57
				*							
TrwC	GNIGEGHRIM	RSATRODSKE	RIGLDLTFSA	PKSVSLQALV	AGDAEIIKAH	DRAVARTLEQ					120
TraHI	GEIDTFTHMQ	RHVG-DAKKE	RLGYDLTFSA	PKGVSMAALI	HGDKTIIKAA	EKAVAAAVRE					119
TraI	GRLPDGADLS	R-MQDGSNRH	RPGYDLTFSA	PKSVSMAMML	GGDKRLIDAH	NQAVDFAVRQ					116
		▼58		▼52	motif III						
TrwC	AEARQAARQK	IQGKTRIEET	GNLVIGKFRH	ETSRERDPOL	HTHAVILAMT	KRSDGQWRAL					180
TraHI	AUKLAQARTT	EQGKSVTONT	NNLVVATFRH	ETSRALDPDL	HTHAFVMMNT	QREDGQWRAL					179
TraI	VEALASTKVM	TDGQSETVLT	GNLVMALFNH	ETSRDOEPOL	HTHAVVANVT	QH-NGEWKTL					175
				▼51							
TrwC	KNDEIVK---	-----RTRY	LGAVYNAELA	HELQKLGYYL	RY-GKDGDFD	LAHIDRQQIE					210
TraHI	KNDELMR---	-----NKKH	LGDVYKQELA	LELAKAGYGL	RYNSKNNTFD	MAHFSDEQIR					210
TraI	SSDKVGKTGF	IENVYANQIA	FGRLYREKLK	EQVEALGYET	EVVGKHGMWE	MPGVP---VE					207
TrwC	GFSKRTEQIA	EWYAARGLDP	NSVSLEQKQA	AKVLSRAKKT	SVDREALRAE	WQATAI LGI					290
TraHI	AFSRRSEQIE	KGLAAMGLTR	ETADAQTKSR	VSMATREKKT	EHSREEIHQQ	WASRAKTLGI					290
TraI	AFSGRSQTIR	EAVG-----	EDASLKSRDV	AALDTKSKQ	HVDPEIKMAE	WMQTLKETGF					286
TrwC	DF-----SRRE	WSGREK----	GGSEKQAHSF	MPSDEAA---	--KRAVRYAI	NHLTER----					333
TraHI	DF-----DNRE	WQSHGKPLEA	DIARNMAPDF	TSPEVKA---	--DRAIQFAV	KSLSER----					337
TraI	DIRAYRDAAD	QRADLRILT.	GPASQDGPDV	QCAVTQAIAG	LSEKRVQFTY	TDVLARTVGI					346
TrwC	-----	-----QSVMD	ERELVDTAMK	-----HA	VGAARLEDIQ	KELLRQ----					366
TraHI	-----	-----DASFE	REKLIQIANK	-----QV	LGHATIADVE	KAYLKA----					370
TraI	LPPENGVIER	ARAGIDEAIS	REQLIPLDRE	KGLFTSGIHV	LDELSVRALS	RDIMKQNRVT					406
TrwC	-----	-----	-----	-----	-----	-----					366
TraHI	-----	-----	-----	-----	-----	-----					370
TraI	VHPEKSVPRT	AGYSDAVSVL	AQDRPSLAIV	SGQGGGAAGQR	ERVAELVMAA	REQGREVQII					466
TrwC	-----	-----	-----	-----	TETG-----	---YLIREAP					377
TraHI	-----	-----	-----	-----	VQKG-----	---ALIEGEA					381
TraI	AADRRSQMMN	KQDERLSGEL	ITGRRQLLEG	MAFTPGSTVI	VDQGEKLSLK	ETLTLLDGAA					526
TrwC	R-----	-----	-----	-----	-----	-----					378
TraHI	R-----	-----	-----	-----	-----	-----					382
TraI	RHNVQVLITD	SGQRTGTGSA	LMAMKDAGVN	TYRWQGGEGE	PATIISEFDR	NVRYARLAGD					586

TrwC	YRPGCQTGPT	DEPGK	-----	-----	-----	-----	-----	393
TrwHI	YQSTLKVGAS	VMAET	-----	-----	-----	-----	-----	397
TrwI	FAASVKAGEE	SVAQVSGVRE	QAILTQAIRS	ELKTQGVLSL	PEVTMTALSP	VWLDKSKRYL	-----	646
TrwC	-----	TR AEW	-----	-----	-----	VA ELAAK	-----	405
TrwHI	-----	LTR KEW	-----	-----	-----	ID SLTNS	-----	410
TrwI	RDMYRPGTVM	EQWNPETRSH	DRYVIDRVTA	QSHSLTLRD	AQGETQVVRIS	SLDSSWSLPR	-----	706
TrwC	-----	-----	G MK-QGAAREP	VDNAIKTGGI	VPIEPR	-----	-----	431
TrwHI	-----	-----	G ME ADKARFA	VDDGIKNGRL	KLTSHR	-----	-----	436
TrwI	PEKMPVADGE	RLRVTKIPG	LRVSGGDRLO	VASVSEDAMT	VVVPGRAEPA	TLPVSDSFT	-----	766
TrwC	-----	-----	YTTQTALE	REKRILQ-IE	RDGR	-----	-----	452
TrwHI	-----	-----	VTTVEGIR	LERSILT-IE	SRGR	-----	-----	457
TrwI	ALKLENGWVE	TPGHSVSDSA	TVFASVTQMA	MDNATLNGLA	RSGRDVRLYS	SLDETRTAEK	-----	826
TrwC	-----	-----	-----	-----	-----	-----	-----	452
TrwHI	-----	-----	-----	-----	-----	-----	-----	457
TrwI	LARHPSFTVV	SEQIKTRAGE	TSLETAISHQ	KSALHTP	ALHLA PVVE	SKKLAFSMVT	-----	856
TrwC	-----	-----	G AVAPVIA	-----	-----	AEAAAPF	-----	466
TrwHI	-----	-----	G QNPROLT	-----	-----	AEIAAPF	-----	471
TrwI	LLTEAKSFAA	EGTGPTLGG	EINAQIKRGD	LLYVDVARGY	G'GLLVSKAS	YEAEKSLPR	-----	946
TrwC	RLASTN	-----	-----	-----	nucleotide binding	-----	-----	506
TrwHI	LLSGKT	-----	-----	-----	-----	-----	-----	511
TrwI	ILEGKEAVMP	LMERVPGELM	EKLTSQGRRA	TRMILETSR	PTVVGGYAGV	SKTT-PRAM	-----	1006
TrwC	DTAKOMIEGE	GYHVRALAAY	CSQVKALREL	NVEANTLAGE	LPKADYNDS	PT	-----	566
TrwHI	MAAFELLESQ	GLYVTALAPY	CTQKFALEDD	GLPARTVAAP	LPKADKFLRE	KI	-----	567
TrwI	SAVINMLPESE	RPRVVGLSPT	HRAVGEMRSA	GVDAITLAGE	SH	ITLGL	PGSTHPRON	1064
TrwC	VLVIDEAGV	VPTRLMQTL	KLAEKAGAKV	VLMGDIAGTY	AVEAGKFFP	VLQAA MLTA	-----	616
TrwHI	VVFIDEAGV	IPAROMKQLM	EVIEKHNAAR	VFGSDTSQTY	AVEAGKFFP	LLIYA MLTA	-----	617
TrwI	TLFLDESSM	VGNTOMARAY	ALLAAGGGKA	VASGDTLGL	ALAI DDPFL	LLQKAA MLTA	-----	1024
TrwC	HMREIQROKN	PELKIAVELA	AAG-KASSS	LEPIV	-----	ITVETVGNHE	-----	618
TrwHI	YMKDIGROKN	EVLLEAVKYA	AEG-NAARA	LYNIT	-----	ITVETVGNHE	-----	619
TrwI	IMKCIVRQTP	ELREAVYSLI	NRLVERALSG	LESVYPSQVP	PLQGANAPEN	ITVETVGNHE	-----	1146

OF / DE

3

PRECISIONSM RESOLUTION TARGETS

TrwC	RR-----	-----	-----	AAVAE	AYIALKPDER	DRTLIVSGTN	EARRBINQIV	696
TraHI	RL-----	-----	-----	AQLAD	RYLSLSSEQQ	DATLIISGTN	ASRKTLDYI	701
TraI	AKLAAQOKA	MLKGEAFPDV	PMTLYEAIVR	DYTGRTPEAR	EQTLIVTHLN	EDRRVLNSMI	1244	
TrwC	---REGLGTA	GKG-IEFDTL	VRVDTTQAE	RH--SKNYQV	GHVIQPERDY	AKTGLQRGFL	750	
TraHI	---RGNLGLA	GTG-ETPTLL	DRVDSTQAE	RD--SRYFSK	GQIIIEQDY	-KNGMKRGES	754	
TraI	HDVREKAGEL	GKEQVMVPVL	NTANIRDGEL	RRLSTWETHR	DALVLVDNVY	-----	1294	
TrwC	YRVVETGPGN	RLTVIGENDG	QRIQFSPMTH	TK--ISVYQP	ERAELAVGDT	IRITRNDKHL	808	
TraHI	YQVLDTGPGN	KLTV-ESISG	EQIAFSPRTH	TK--LSVYQA	VSAELAPGAK	VMVTRNDKTL	811	
TraI	HRIAGISKDD	GLITLQDAEG	NTRLISPREA	VAEGVTLYTP	DTIRVGTGDR	MRFTKSDRER	1354	
TrwC	DLANGDRMKV	VAVEDRKVTV	TDGK--RNVE	LPTDKPI-HV	DHAYATTVHS	SQGLTSDRVL	865	
TraHI	DVANGDRFTV	KTVEGEKLT	EDKK-GRTVE	LDKKQAS-YL	SYAYATTVHK	SQGLTCDRVL	869	
TraI	GYVANSVWTV	TAVSGDSVTL	SDGQOTKEIR	PGQEQAQCHI	DLAYAITAHG	AQGASETFAI	1414	
TrwC	IDAHAESRTT	RKDVIY--YV	AISRARFEAR	VFTNDRGKLP	AAIARENIXS	AAHDL-----	917	
TraHI	FNIDTKSLTT	SKDVF--YV	GISRARHEVE	IFTDDKSLA	SSVSRDSPKT	TAAEI-----	921	
TraI	ALEGTEGNRK	LMAGFESAYV	ALSRMEQHVO	VYTDNRQGT	DAINNAVQKG	TAHDVFEPKP	1474	
TrwC	-----AR-	-----	-----	-----	-----	DRGORS	AAAER-----Q	931
TraHI	-----DRF	FGLEARFKDI	G-----RDT	SLET-----	-----	RSAEKGLPE	ATGESMAFNQ	961
TraI	DREVMNAERL	FSTARELRDV	AAGRAVLROA	GLAGGDSPAR	FIAPGRKYPO	PYVALPAFDR	1534	
TrwC	RE-----	-----	-----	QQRER E-----	-----	-----	939	
TraHI	KP-----	-----	-----	DEHNM TTG-TDYQPV	-----	SNAEDAFHL	986	
TraI	NGKSAGIWLW	PLTTDDGNGL	RGFSGEGRVK	GSGDAQFVAL	QGSRNESLL	ADNMQDGVRI	1694	
TrwC	-----R-	-----	-----	-----	-----	-----	940	
TraHI	KQNPMDDSVG	LR-----	-----	-----	-----	-----	998	
TraI	ARDNPDGQVV	VRIAGEGRPW	NPGAITGGRV	WGDIPDNSVQ	PGAGNGEPVT	AEVLAQRQAE	1754	
TrwC	---NRQTOQ-	-----	PAHDR	QKAARE-----	-----	-----	957	
TraHI	---RHEAQON	DAELANDYA-	---AADDQ	QSAQYADY	EHYAEA---S	DYDFDSSIYD	1046	
TraI	EAIRRETERE	ADIVRKMAE	NKPDLPDQKT	EQAVREIAGQ	ERDRAAITER	EAALPEGVLR	1814	
TrwC	-----	-----	---AERGMEAG	R-----	-----	-----	966	
TraHI	DYAMPQTS--	-----QA	EQSHTQKHT	HEHEHEEGGH	EI	-----	1078	
TraI	EPQVRREAVR	EIARENLLQE	RLQOMERDMV	RDLOKEKTLG	GD	-----	1756	

4.6 IDENTIFICATION OF *mob* FUNCTIONS INVOLVED IN NICKING AT *oriT*

4.6.1 Intermolecular site-specific recombination assay with pSP34::TnphoA

As previously described in Section 4.2 and 4.4, intramolecular site-specific recombination between two *oriT*s on the same plasmid has been observed and used to locate *nic* within *oriT*. As described below, an intermolecular site-specific recombination between *oriT*s carried on two separate plasmids was used to determine which of the *mob* functions are required for nicking and religation at *oriT* during transfer.

In this *in vivo* nicking assay, different TnphoA mutants of pSP34 were tested for their ability to mobilize pSP27 from *E. coli* donor S17-1. S17-1 carries the P *tra* system on its chromosome and hence is able to donate pSP34 efficiently via the P *oriT* it carries. pSP27 is a derivative of pMK2004 carrying the 325 bp pCU1 *oriT*. It cannot be donated by the P *tra* system carried by S17-1, however it can be conducted by pSP34 via site-specific recombination at the N *oriT*s carried by each plasmid. In these experiments, pSP34 was shown to mobilize pSP27 from S17-1 at a frequency 100 fold greater than pMK2004 (see Table 4.6). Restriction digests of plasmid DNA from the S17-1/pSP34/pSP27 transconjugants demonstrated a co-integrate plasmid consistent with integration occurring at the two *oriT* regions. The *oriT* of pSP34 is contained on a 2.3 Kb *HpaI* fragment (see Figure 4.12) while pSP27 has no *HpaI* sites. Therefore intermolecular site-specific recombination at *oriT* can be followed by the replacement of the pSP34 2.3 Kb *HpaI* fragment with a 7.7 Kb fragment (see Figure 4.21). DNA from S17-1/ pSP34/pMK2004 transconjugants demonstrates either the same digestion pattern as the parental plasmids in the donor cells or a pattern consistent with cointegration of

Table 4.6 Mobilization of pSP27 by pSP34 *tra* mutants from S17-1.

pSP34 derivative	Mobilization frequency ^a	Plasmid structure in transconjugant			
		cointegration between <i>oriT</i> s	cointegration in vector	parental plasmid(s)	other
pSP34:: <i>phoA52</i> (<i>traHI</i>)	2.3x10 ⁻⁶	3	7	6	0
pSP34ΔKpn (<i>traHI</i>)	2.7x10 ⁻⁶	2	0	2	1
pSP34:: <i>phoA16</i> (<i>traJ</i>)	2.4x10 ⁻⁴	1	0	9	0
pSP34:: <i>phoA11</i> (<i>traJ</i>)	7.4x10 ⁻⁵	6	0	2	0
pSP34:: <i>phoA54</i> (<i>traJ</i>)	5.1x10 ⁻⁴	3	0	2	0
pSP34:: <i>phoA31</i> (<i>traJ</i>)	2.9x10 ⁻⁴	5	0	0	0
pSP34:: <i>phoA19</i> (<i>traJ</i>)	1.5x10 ⁻⁴	7	0	2	1
pSP34:: <i>phoA30</i> (<i>traK</i>)	2.7x10 ⁻⁶	3	4	4	3
pSP34/pSP27 ^b	9.1x10 ⁻⁴	21	0	4	0
pSP34/pMK2004 ^b	7.1x10 ⁻⁶	0	5	7	0

^a mobilization frequencies are given as the number of transconjugants receiving pSP27 (or pMK2004) per transconjugant receiving pSP34 derivative. Results are the average of at least two trials except for the *traJ*::*phoA* derivatives 11, 54 and 31. Replica result values varied from 2-20 fold from each other with a maximum variation of 100 fold in the pSP34ΔKpn results.

^b positive and negative controls. pMK2004 is the parental plasmid of pSP27.

Figure 4.21 Schematic diagram comparing the results of mobilization of pSP27 by pSP34 from S17-1 by conduction. Upper cell represents the S17-1 donor cell with the *P-tra* system (box with diagonal lines) on its chromosome. The donor cell contains plasmid pSP27 with the *N oriT* (solid arrow) and the Tc gene (hatched box) and plasmid pSP34 with the *N oriT* (solid arrow) contained on a 2.3 Kb *HpaI* (Hp) fragment, the *P oriT* (box with vertical lines), the Cm gene (dotted box), and the *N mob* (open box).

Lower cells represent the transconjugants after mating S17-1/pSP27/pSP34 with HB101Rif (Rif gene, box with large hatch, carried on the chromosome). Lower left, site-specific recombination between *oriT* sequences (site of recombination represented by ^) results in the integration of pSP27 into the 2.3 Kb *HpaI* fragment of pSP34. In the cointegrate, the pSP27 DNA is bordered by two Inc *N oriTs* (solid arrows) and contained on a 7.7 bp *HpaI* fragment. Lower right, site-specific recombination not involving the *oriT* sequences results in integration of pSP27 into pSP34 at a site (^) outside the 2.3 Kb *HpaI* fragment. In this case the 2.3 Kb *HpaI* fragment of pSP34 is not altered.

pMK2004 into the vector portion of pSP34. Integration into the vector portion of the plasmid could occur via recombination at the *oriV* sites on either plasmid (Kilbane and Malamy, 1980; O'Connor and Malamy, 1984) or by other RecA independent mechanisms (Reinmann and Hass, 1993). The presence of two unaltered parental plasmids would presumably result from the resolution of such a cointegrate.

As can be seen from Table 4.6, the mobilization frequency of pSP27 varied with the *mob* gene affected. Insertions in *traHI* reduced mobilization of pSP27 by 1000 fold, the *traK* insertion resulted in a 100 fold reduction and *traJ* mutations demonstrated only a 10 fold reduction in the mobilization frequency of pSP27 relative to that observed with pSP34.

While the transfer results clearly suggest that *traHI* and *traK* are more important for nicking than *traJ*, it is important to note that pCU57 Δ 14 from which pSP34 was constructed (contains pCU1 *mob* but not the P *oriT*), is donated from S17-1 at a frequency comparable to the frequency with which pSP27 is conducted by pSP34 from S17-1. This suggests that the N *mob* functions can interact to some extent with the P pilus bridge to allow transfer at a low frequency. This would indicate that pSP27 may be donated from S17-1/pSP34 as well as conducted by site-specific recombination and therefore that the mobilization frequencies of pSP27 by S17-1/pSP34::Tn*phoA* may not reflect an accurate nicking frequency per se.

To clearly demonstrate whether the pSP27 plasmid was donated or conducted from S17-1/pSP34::Tn*phoA* derivatives, plasmid DNA was extracted from transconjugants and digested with *HpaI*. As seen from the results presented in Table 4.6, while most (84%)

of the transconjugants from the pSP34 pSP27 mating show co-integration within *oriT*, none of those from the pSP34 pMK2004 mating do. The plasmid DNA from the latter transconjugants demonstrate either co-integration in the vector portion of pSP34 (42%) or no co-integration at all (58%).

Transconjugant DNA from 58% of *traJ* transconjugants (representing 5 different pSP34::Tn*phoA* derivatives) demonstrated digests consistent with integration of pSP27 at *oriT*. The remaining 39% had patterns identical to the parental plasmids. While a greater percentage of the transconjugants have *oriT* cointegrates than parental size plasmids, it is clear that the percentage is reduced in comparison to the results obtained with unmutated pSP34.

The *traK* and *traHI* mutants demonstrate a more definitive effect on site specific recombination at *oriT*. Only 21% and 25% of *traK* and *traHI* transconjugants carry cointegrates formed by site-specific recombination at *oriT*.

These results support the mobilization frequencies and the conclusion that both *traK* and *traHI* play a more important role in site-specific recombination at *oriT* than does *traJ*.

5. DISCUSSION

The N type *oriT* of plasmid pMUR274 was cloned on a 474 bp *RsaI*-*SspI* fragment and sequenced. Alignment of this sequence with that of the pCU1 *oriT* demonstrated over 50% sequence identity. Site-specific recombination between directly repeated copies of the pCU1 and pMUR274 *oriTs* resulted in deletion of the intervening DNA and the formation of a hybrid *oriT* containing opposite halves of each parental *oriT*. The junction between the two parental sequences, presumably containing *nic*, occurred within the 10 bp sequence GCTATACACC common to both parental sequences. Similar recombination events between directly repeated copies of wildtype and mutated pCU1 *oriT* sequences further located *nic* within the 10 bp junction to either side of the second T residue (nt 119 from the *RsaI* site which defines the left end of *oriT*). Lack of any variation in the sequence of the junctions in any of the hybrid *oriTs* indicates that a unique site is nicked, that this site is the same in both parental plasmids and that mutations within the sequence surrounding *nic* do not alter its location. Use of a RecA background, combined with the fact that no other junction sequences were observed despite extended sequences of homology between the two *oriT* sequences as well as other sequences outside *oriT*, supports the conclusion that this is a *recA* independent, site-specific recombination event.

As seen in Figure 5.1, the 10 bp junction sequence surrounding *nic* shares sequence identity at eight and seven bases of the *nic* sequences of R388 and the F plasmid, respectively. These sequences align such that the *nic* sites of F and R388

Figure 5.1 Comparison of *nic* sequence in pCU1, R388 and F plasmids. The *nic* sites are indicated by ^.

G	C	T	A	T	A	G	A	C	A
				^					
G	C	T	A	T	A	C	A	C	C
				^	^				
G	C	A	C	C	A	C	A	C	C
				^					

R388

pCU1

F

correspond to the putative pCU1 *nic* site identified to the right of the T residue in the 10 bp junction sequence. Based on the direction of transfer observed for IncN plasmid pKM101, which shares 100% homology with the *oriT* region of pCU1 (in Lanka, 1985), and assuming that transfer occurs in the 5' 3' direction as predicted for all other transfer systems studied to date (Lanka and Wilkins, 1995), it would be the strand complementary to that depicted in all figures in this study that is nicked. This agrees with findings for both R388 and F plasmids (Thompson *et al.*, 1989; de la Cruz, personal communication).

Of the point mutations generated within the sequence surrounding *nic*, only three (*nic3*, *nic16* and *nic238*) to the right of *nic*, which represents the 3' side of the nicked strand, reduce transfer significantly. Two mutations on the 5' end of the nicked strand (*nic188* and *nic109*) as well as several more distant mutations on the 3' side (*nic11*, *nic211* and the five unique mutations in the 7 bp hot spot) had little or no effect. These results are consistent with mutations in the *oriT* regions of other plasmids. In the IncP and IncQ plasmids, mutations within the 8 bases of the 3' end of the nicked strand between *nic* and the inverted repeat, result in the greatest reduction in transfer. This region has been implicated in recognition by the nicking enzyme (Waters *et al.*, 1991; Zhang and Meyer, 1995; Pansegrau and Lanka, 1996). Similarly, two point mutations 4 and 8 bp from the 3' side of the F *nic* result in significant reductions while several derivatives carrying six bp substitutions to the 5' side of *nic* showed little or no reduction in transfer (Luo *et al.*, 1995). Deletion of the five bases to the left of the T residue which forms one side of the pCU1 *nic* results in a 100 fold reduction in transfer

(compare *nic144* and *pSP6*). These results could suggest that the immediate 5' end of the nicked strand is important in the transfer of *pCU1* or that *nic* lies to the left of the T residue which is deleted in *nic144*. If this is the case, then the *nic* site in *pCU1* may not align with the *nic* sites of F and R388 as suggested in Figure 5.1.

That the two bases in the R388 sequence that do not match the *pCU1* and *pMUR* sequence lie to the right of *nic* may explain the lack of success in attempts to obtain a R388-*pCU1* hybrid *oriT* by site-specific recombination (personal observation, Llosa *et al.* 1994₁). Considering the sequence conservation to the right of *nic* in *pCU1* and the F plasmid, it may be possible to create a hybrid between these two *oriTs* (unless *nic* proves to lie to the left of the T residue in which case the base at the 3' end of the two *nics* would be different).

It should be noted that in each of the *IncP*, *IncQ*, *Inc11*, *IncF* and *IncW* *oriTs* there is an inverted repeat sequence less than 10 bp to the 3' side of the nicked strand. In all cases, deletion or alteration of the arm of the IR closest to *nic* results in loss of transfer (Luo *et al.* 1995; Waters *et al.* 1991; Furuya and Komano, 1991; Brasch and Meyer, 1986). While the *pMUR* plasmid contains an 11 bp imperfect IR 10 bases to the right of the T residue, the closest IR to this residue in *pCU1* is a 7 bp perfect IR 18 bp to the right.

Of the point mutations within the *nic* region, mutation of what may be the 3' end of *nic* (*nic3*) affected both initiation and termination to a greater extent than either the mutation on the 5' side (*nic109*) or further from *nic* on the 3' (*nic238*). In particular, the *nic3* mutation affected termination more severely than initiation. Similar results were

observed with an F point mutation 4 bases on the 3' side of *nic* (Gao *et al.*, 1992).

The deletion in pSP9 reduced the minimal pCU1 sequence required *in cis* for transfer to 266 bases, extending from the *RsaI* site on the left side to the end of *IR_L* on the right. While deletions into either end of this 266 bp fragment affect transfer, deletions into the right end have a much greater effect than deletions into the left. A deletion (pSP12) into the left end of *oriT*, removing the *IR_L* sequence, results in only a 10 fold reduction in transfer (Figure 4.10; Paterson and Iyer, 1992). Even a derivative carrying a deletion to within 5 bases of the T residue representing one side of *nic* (*nic144*) is transferred at a low frequency (10^4). By contrast, a deletion into the right side of *oriT* removing only the last inverted repeat as in pSP10 or a 23 bp deletion in the AT rich sequence to the right of *nic* found in pSP29 results in a significant decrease in transfer frequency (10^6). Evidently the 3' trailing region of the *oriT* is more critical to transfer than the 5' leading region.

The ability of the hybrid pMUR-pCU1 *oriT* plasmids to be mobilized only by the parental plasmid corresponding to the right portion of the *oriT* suggests that it is this portion of the *oriT* that dictates plasmid specificity. In other conjugative systems plasmid specificity is determined by those functions that bind to the *oriT*. This would suggest that *mob* proteins of pCU1 may interact specifically with the *oriT* sequences to the right of *nic* namely the AT rich sequence and/or the inverted repeat sequences. Since deletion of a portion of either of these two sequences results in elimination of transfer either is possible. Evidence from the recombination experiments involving the AT deletion in pNKN29 and the *IR* deletion in pNKN10 suggests that these two sequences

play different roles since the former affects both termination and initiation while the latter affects only initiation.

The F plasmids are the only other group of plasmids which contain a significant AT rich sequence in their *oriT*. The AT rich region in the F plasmids begins 11 bases from the 3' end of the nicked strand and extends approximately 80 bp. The *nic* proximal end of the AT stretch contains two IR while the distal end encompasses the TraY and IHF binding sites. The TraM binding sites are located beyond the AT rich sequence in the F like *oriTs*. By comparison, the AT rich sequence of the N *oriT* starts almost 40 bp from *nic* and the inverted repeat sequences span the distal portion. In the F plasmid, deletion of the distal portion of the AT rich sequence which contains TraY binding sites affects initiation of transfer but not termination while deletions up to 30 bases on the 3' side of *nic* affected both functions (Gao *et al.*, 1994). These results are somewhat similar to the situation in pCU1 where deletion of the last IR affects initiation only, while deletion of part of the AT sequence closer to *nic* affects both initiation and termination. This would suggest that TraK, the putative TraY analogue, binds to the inverted repeats in the right portion of the AT rich sequence and that *nic* proximal portion of the AT rich sequence plays another role in transfer.

One possible role which could be played by the proximal portion of the AT rich sequence of pCU1 is recognition by the nicking enzyme. In the IncQ system, recognition of the *oriT* by the nicking enzyme is attributed to localized melting of the sequence between *nic* and the IR on the 3' side of *nic* by the *mobC* function (Zang and Meyer, 1994). In the F plasmids, similar permanganate sensitivity studies suggested that the AT

rich sequence was not significantly melted during the initial stages of transfer, however, it was suggested that the AT tracks could induce bending providing access to *nic* by Tral (Luo *et al*, 1995). The AT rich region in pCU1 could play a comparable role. Despite the high degree of sequence homology between the *oriT* regions of the IncW plasmid R388 and pCU1, the former does not contain an AT rich sequence. This would suggest that there is an intrinsic difference in the transfer of these two plasmids which is discussed further below.

The modest effects of deletions into the left end of *oriT* would suggest a less critical role in transfer. Similarly modest reductions in transfers have been observed with the deletion of the 5' leading end of *oriT* in the IncP plasmids, however this region of the RP4 *oriT* binds the TraK protein for which there is no pCU1 analogue (Furste *et al*, 1992). Conservation of IR₁ at the left end of both N *oriT*s as well as the W *oriT* would suggest that it is important to the function of this region. That the sequence of IR₁ is not conserved between the IncW and IncN plasmids might suggest that the structure is more important than the sequence. Since this sequence represents the leading end of the transferred strand, one possible role might be in complementary strand synthesis in the recipient. The lower strand, to the left of *nic* which is conserved in the two IncN plasmids, contains a sequence with five of six bases matching the consensus sequence of the Pri binding site of ϕ X174 (AAGCGG). This site may be used as a loading site for host encoded primosome proteins which could initiate strand synthesis. The proximity of IR₁ is consistent with this possibility. Deletion of IR₁ in pSP12 causes a 10 fold reduction in transfer while deletion of the putative Pri binding site in pSP19 results in

an additional 10 fold reduction (Paterson and Iyer, 1992). Recognition of the left portion of *oriT* by host encoded functions is consistent with the lack of plasmid specificity of this region.

It should be noted that pSP19 also lacks one of the 8 bp direct repeats which could also be the reason for the reduction in transfer observed with this derivative. Comparable mobilization results for *nic136* and *nic144* suggest that deletion of the second direct repeat (present in *nic136* but not *nic144*) has no additional affect on transfer. However, while both show a 10 fold reduction in transfer relative to pSP19, *nic136* carries several point mutations to the right of *nic* that might account for this reduction while the single point mutation in *nic146* within the fourth inverted repeat most likely would not result in a 10 fold reduction in transfer as indicated by the transfer frequency of a similar mutation in pSP8. These results could suggest that deletion of each of the direct repeats results in a 10 fold reduction in transfer.

DNA sequencing of the pCU1 *mob* region has identified three putative coding regions corresponding to *traK*, *traJ* and *traHI* genes (the sequence of the latter gene was determined by M. More and S. Winans). While the results presented in this study are do not clearly support or contradict earlier complementation experiments suggesting the presence of two complementation regions above the *Bgl*III site corresponding to the *traH* and *traI* genes (Winans and Walker, 1985), sequence and SDS-PAGE results indicate the existence of one large coding region above the *Bgl*III site. To prevent confusion with the earlier pKM101 results, this region has been designated *traHI*. Complementation results coupled with the nearly 100% sequence identity over corresponding regions of pKM101

and pCU1 that have been sequenced to date, suggest that the *mob* regions of these two plasmids are almost identical, similar to the situation with the IncP plasmids RP4 and RK2 or the IncQ plasmids R1162 and RSF1010.

Alignments between the N, F and W *tra* proteins would suggest that the pCU1 TraHI protein is an analogue of the F Tral and the R388 TrwC proteins, both of which encode an endonuclease function necessary for nicking. This is consistent with the finding that the *traHI* gene is required for intermolecular recombination between pCU1 *oriTs*. Each of the TraHI, TrwC and Tral proteins contain two sets of two tyrosine residues within the first 30 amino acids (aa 18/19 and 26/27 in TraHI and TrwC and 16/17 and 23/24 in Tral). The location of these residues coincides with tyr22 and tyr29 in motif I of the IncP Tral protein and the Ti VirD2 protein, respectively. In the IncP Tral and the VirD2 proteins, tyr22 is believed to be the catalytic centre of the nickase and has been shown to be involved in covalent binding of the protein to the 5' end of the nicked strand (Vogel and Das, 1992₂; Pansegrau *et al.*, 1993; Pansegrau and Lanka, 1996). The presence of two closely spaced sets of two adjacent tyr residues in TraHI is reminiscent of the two tyr residues separated by three amino acids in the A protein of phage ϕ X174 (Hanai and Wang, 1993). In the ϕ X174 system, it was postulated that the two tyrosine residues alternate in nicking within the replication origin. This situation could assist in the intermolecular recombination event observed in this study. It is conceivable that each tyr residue (or one from each pair) nicks and binds to separate pSP27 and pSP34 *oriTs*, bringing them into close proximity of each other to allow the transesterification step between the free 3' OH of one plasmid and the 5'-tyr of the other

resulting in the formation of a cointegrate plasmid.

Of the other two motifs observed in the IncP-type nickases, only motif III is partially conserved in the IncN, IncW and IncF enzymes. Conservation of the his residues within this motif is consistent with the belief that one of the his residues of motif III interacts with the tyr residues of motif I as part of the catalytic centre of the nickase. On the assumption that motif II is involved in recognition and non-covalent binding of the nickase to the 3' end of the *nic* sequence, the lack of a conserved IncP-like motif II sequence in the F, N and W nickase is consistent with the lack of DNA conservation between the *nic* region of these plasmids and the IncP like plasmids (Balzer *et al.*, 1994; Pansegrau *et al.*, 1996).

While SDS-PAGE clearly demonstrates that the *traHI* region encodes a large protein, the complementation results reflect the possibility that this protein contains more than one functional domain as seen with functional analogues in each of the IncF, IncW, IncP, IncQ and Ti systems. As indicated in Figure 4.20, the amino and carboxyl termini of the pCU1 TraHI and R388 TrwC proteins align separately with two regions of the F TraI protein. The putative nickase motifs lie within the first region of alignment while the nucleotide binding site which could be associated with the ATPase activity of a helicase function lies at the start of the second region of alignment. The presence of a restart protein within the *traHI* coding region would help to explain the inconsistent complementation results. In both the F and P systems, a restart protein has been identified however no function has been assigned to it (Traxler and Minkley, 1988; Pansegrau *et al.*, 1990).

There is a putative ATG start codon at aa 518, just upstream of the nucleotide binding site sequence which is preceded by a strong ribosomal binding site. Translation initiation from such a site would result in a 50-55 kDa protein which should be visible on the SDS-PAGE of the pET-traHI::Tn*phoA2* isolate. Such a band is not observed.

The only possible translational restart proteins visible on the SDS-PAGE are seen in pET-traHI::Tn*phoA3* and 7. Both isolates show a less intense protein band which is 10-15 kDa smaller than the main protein. A reduction of this size would require a restart site approximately 300 bp downstream of the main *traHI* start codon. There is an ATG codon 285 bp from the putative start codon which is preceded by a weak RBS.

Unfortunately, such a restart protein would not completely explain the complementation results obtained in this study. Each of the three pSP34::Tn*phoA* inserts used in the complementation tests lie within the first region of alignment with F TraI, downstream from the putative restart codon at aa 95. While the exact location of the pKM101*traI*::Tn5 insert is not known, it was mapped close to the *Bgl*III site. If it lies upstream of the putative restart codon, it is possible that transcriptional readthrough from the Tn5 could produce sufficient transcript to allow expression of the restart protein. This restart protein could provide the helicase functions of the carboxyl terminus of TraHI while the pSP34::Tn*phoA* inserts provide the nicking function of the amino terminus. It should be noted that while each of the pSP34 mutants would produce truncated proteins containing the potentially active tyr residues, only insert 51 would produce a protein containing the motif III like sequence. A translational restart site at aa 95 also does not explain the results obtained with the pSP34ΔKpn derivative. By the

reasoning presented above, the pKM101*traI* insert should be complemented by pSP34ΔKpn if the former is able to supply the helicase activity on a restart protein while the latter supplies the nicking functions on a truncated amino terminal protein. However, neither pKM101*traI* or *traH* mutants appear to be complemented by pSP34ΔKpn.

The possibility of a truncated protein comprising the amino terminus of TraHI carrying full nickase activity is not supported by the intermolecular recombination results. The Kpn deletion derivative should produce a protein containing the first 523 amino acids of TraHI. This is significantly larger than a deletion derivative carrying the first 272 amino acid of TrwC which has been shown to retain full nicking activity (Llosa *et al.*, 1994₁). Based on this information, one would expect the Kpn deletion derivative to demonstrate nicking activity by mobilizing pSP27 efficiently from S17-1. However such is not the case. Further investigation into the possible functional domains of the TraHI protein is required.

The TraK protein of pCU1 shows 26% sequence identity to the TrwA protein of R388. By contrast, TraK shows very little sequence identity with the TraY protein of the F plasmid. However, the finding that *traK* is required for intermolecular site-specific recombination at *oriT* suggests that it may represent a functional analogue of the F plasmid *traY* gene. Sequence conservation between TraK and TrwA is most pronounced in the carboxyl terminus which is consistent with the higher degree of sequence identity between the carboxyl terminus of the different F like TraY proteins. In the F like plasmids, TraY demonstrates plasmid specificity which is believed to reflect variations in the nucleotide sequence of the binding site. As a result, it has been suggested that it

is the less conserved amino terminus that probably specifies binding (Bowie and Sauer, 1990). The TrwC protein contains a leucine zipper motif which suggests that it forms a dimer in solution. While TraK shows remnants of such a motif, there is no indication that the TraY protein contains such a structure which is consistent with findings that TraY exists as a monomer in solution (Nelson *et al.*, 1993). These results taken together, would suggest that TraK and TraY are functional analogues despite little or no sequence identity while TraK and TrwA contain conserved sequences despite a lack of apparent functional similarity. This lack of functional similarity is demonstrated by the observation that TrwC is not required for intramolecular site-specific recombination at the R388 *oriT* (Llosa *et al.*, 1994₁) while TraK is required for intermolecular site-specific recombination. Plasmid specificity of the pCU1-pMUR hybrid *oriT* for transfer functions is dependant upon the right side of the *oriT*, suggesting that the pCU1 *tra* proteins interact with the AT rich sequence and the inverted repeat sequences located in this region. The presence of the AT rich sequence in both the F and pCU1 *oriTs* but not that of R388, may correlate with the variation in function observed for the TraK, TraY and TrwC proteins.

As discussed above, the pNKN10 recombination results suggest that if TraK does bind to the pCU1 *oriT* in a manner analogous to the binding of TraY to the F *oriT*, the binding site most likely corresponds to the inverted repeats that lie to the right of *nic*. Presence of the *traK* gene directly to the right of *oriT*, and the possibility that the *traK* promoter overlaps the *oriT*, suggests that *traK* may regulate its own expression in the same manner observed with the *traY* and *traM* genes of the F plasmid (although *traY*

does not lie next to *oriT* it contains TraY binding sites) (Inamoto and Ohstubo, 1990; Schwabb *et al.*, 1993; Luo *et al.*, 1994).

Several attempts to visualize the *traK* product expressed by the pET-*traK* derivative following induction with 0.4 mM IPTG were unsuccessful. It was finally determined that the TraK protein band could only be seen after induction with low concentrations of IPTG (0.05 mM). Since high levels of IPTG also reduce the level of host protein expression it is difficult to compare protein samples from cells induced with different concentrations of IPTG. With this limitation in mind, the SDS-PAGE results in this thesis (as well as other gels not shown) could suggest that high induction levels reduce the amount of TraK protein produced. Low yields of TraK following induction with high levels of IPTG could indicate that overproduction of TraK causes cell death and subsequent proliferation of plasmidless cells. Low TraK yields could also suggest that TraK negatively regulates the expression of *traK*.

The putative start codon of the *traK* ORF is located 61 bp downstream from the end of *oriT*. The *BstYI* site used to construct pET23-*traK* lies 17 bp downstream from the end of *oriT*. The lack of any portion of the *oriT* within the pCU1 sequence cloned into pET-*traK* argues against the possibility that the *traK* is regulated by binding of the TraK protein to the *oriT* region. However, as indicated in Figure 4.14, there appears to be more than one possible promoter sequence upstream of *traK* and hence the sequence to the right of the *BstYI* site used to construct pET-*traK* may represent only one of several TraK binding sites. Primer extension studies have demonstrated that transcription of the *traY* gene of F is initiated from more than one promoter (Silverman

et al. 1991). Interaction of the TraK protein with *oriT* and the promoter sequence of *traK* requires further study.

The TraJ protein shows amino acid sequence identity to the TraD protein of the F plasmid. Like the TraD protein, it contains two pronounced hydrophobic stretches at the amino terminus, which may indicate that it too is associated with the inner membrane. These hydrophobic domains are retained in the IncW protein TrwB. Both the A type and B type ATP binding sites observed in the TraD and TrwB proteins are conserved in TraJ however the RGD motif identified in both the TraD and TrwB protein and proposed as a mechanism of protein-protein interaction is not fully conserved in the TraJ protein (Yoshioka *et al.* 1990).

Overexpression of the carboxyl terminus of the TraJ protein also appears to produce a truncated protein. In this case, the second protein is only 3 kDa smaller than the main product. A protein of this size would require a start codon close to the *HpaI* site used in the construction of pET23-traJ. While there are no ATG codons within this region, there is a rare TTG start codon 40 bp down stream of the *HpaI* site which is preceded by a weak ribosomal binding site which could act as a restart codon to produce a protein of the expected size. Assuming that there is at least a low level of transcriptional readthrough from the transposon, the possibility of a translational restart site downstream of the *HpaI* site would help to explain the higher complementation results obtained with the promoter proximal inserts in *traJ* (#70 and 19) upstream of the *HpaI* site.

The apparent, albeit slight, involvement of the *traJ* function in the site-specific

recombination event has not been observed for the TraJ analogue of any other transfer system studied to date. While this may reflect a difference in the sensitivity of the *in vivo* assay used in this study compared to the relaxation assay used in other studies, it may also reflect a basic difference in the relaxation process in pCU1 which has been resistant to the demonstration of the nicking complex via the SDS-relaxation assay. One could speculate that anchorage of pSP27 and pSP34 DNA to the membrane via TraJ prior to nicking could enhance intermolecular recombination by bringing the two plasmids into closer proximity. This could also prevent isolation of nicked molecules from the cleared lysate and hence prevent the demonstration of nicking by SDS-relaxation.

The sequence data presented in this study confirms earlier reports of the close relationship between the IncN and IncW plasmid (Llosa *et al.*, 1991) and clearly demonstrates that like the IncW system, the IncN conjugal DNA processing system is more closely related to the IncF system than to the IncP (Llosa *et al.*, 1994¹; Llosa *et al.*, 1994₂). This is contrary to early predictions based on similarities between conjugative pili and host range studies (Bradley, 1980). Furthermore, that the genes involved in the assembly of the N type conjugative pili appear to be more closely related to the *vir* genes of *Agrobacterium* (Kado, 1993; Polhman *et al.*, 1994), despite very little similarity between the N *mob* genes and the *vir* genes, is further evidence of the chimeric nature of plasmids (Farrand *et al.*, 1996; Alt-Morbe, 1996). The possibility that the chimeric nature extends beyond conjugal transfer functions is supported by the finding that the pMUR plasmid do not belong to the same incompatibility group as pCU1 despite obvious similarities in conjugative functions (Krishnan *et al.*, 1991).

6. REFERENCES

- Abo, T., S. Inamoto and E. Ohtsubo. 1991. Specific binding of the TraM protein to the *oriT* region of plasmid R100. *J. Bacteriol.* **173**:6347-6354.
- Abo, T. and E. Ohtsubo. 1995. Characterization of the functional sites in the *oriT* region involved in DNA transfer promoted by sex factor plasmid R100. *J. Bacteriol.* **177**:4350-4355.
- Abdel-Monem, M. and H. Hoffmann-Berling. 1976. Enzymatic unwinding of DNA. 1. Purification and characterization of a DNA-dependant ATPase from *Escherichia coli*. *Eur. J. Biochem.* **65**:431-440.
- Abdel-Monem, M., H. Durwald and H. Hoffmann-Berling. 1976. Enzymatic unwinding of DNA. 2. Chain separation of a ATP-dependant DNA unwinding enzyme. *Eur. J. Biochem.* **65**:441-449.
- Abdel-Monem, M., G. Taucher-Scholz and M. Klinkert. 1983. Identification of *Escherichia coli* DNA helicase I as the *traI* product of the F sex factor. *Proc. Natl. Acad. Sci. USA.* **80**:4659-4663.
- Achtman, M., P.A. Manning, C. Edelbluth and P. Herrlich. 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor *tra* cistrons in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **76**:4837-4841.
- Albright, L.M., M.F. Yanofsky, B. Leroux, D. Ma and E.W. Nester. 1987. Processing of the T-DNA of *Agrobacterium tumefaciens* generates border nicks and linear, single-stranded T-DNA. *J. Bacteriol.* **169**:1046-1055.
- Al-Doori, Z., M. Watson and J. Scaife. 1982. The orientation of transfer of the plasmid RP4. *Genet. Res. Camb.* **39**:99-103.
- Alt-Morbe, J., J.L. Stryker, C. Fuqua, P-L. Li, S.K. Farrand and S.C. Winans. 1996. The conjugal transfer system of *Agrobacterium tumefaciens* octopine-type Ti plasmids is closely related to the transfer system of an IncP plasmid and distantly related to the Ti plasmid *vir* genes. *J. Bacteriol.* **178**:4248-4257.
- Ando, T. and T. Arai. 1981. Genetic structure of the IncN plasmid N3. *Plasmid* **6**:293-301.
- Balzer, B., W. Pansegrau and E. Lanka. 1994. Essential motifs of relaxase (TraI) and TraG proteins involved in conjugative transfer of plasmid RP4. *J. Bacteriol.* **176**:4285-4295.

- Barlett, M.M., M.J. Erickson and R.J. Meyer. 1990. Recombination between directly repeated origins of conjugative transfer cloned in M13 bacteriophage DNA models ligation of the transferred plasmid strand. *Nuc. Acids Res.* **18**:3579-3586.
- Bastia, D. 1978. Determination of restriction sites and the nucleotide sequence surrounding the relaxation site of *ColE1*. *J. Mol. Biol.* **124**:601-639.
- Beijersbergen, A., A.D. Dulk-Ras, R.A. Schilperoort and P.J.J. Hooykaas. 1992. Conjugative transfer by the virulence system of *Agrobacterium tumefaciens*. *Sci.* **256**:1324-1327.
- Benz, I. and H. Muller. 1990. *Escherichia coli* DNA helicase I. Characterization of the protein and of its DNA-binding properties. *Eur. J. Biochem.* **189**:267-276.
- Bhattacharjee, M.K. and R.J. Meyer. 1991. A segment of a plasmid gene required for conjugal transfer encodes a site-specific, single-strand DNA endonuclease and ligase. *Nuc. Acids Res.* **19**:1129-1137.
- Bhattacharjee, M., X-M. Rao and R.J. Meyer. 1992. Role of the origin of transfer in termination of strand transfer during bacterial conjugation. *J. Bacteriol.* **174**:6659-6665.
- Bhattacharjee, M.K. and R.J. Meyer. 1993. Specific binding of MobA, a plasmid-encoded protein involved in the initiation and termination of conjugal DNA transfer, to single-stranded *oriT* DNA. *Nuc. Acids Res.* **21**:4563-4568.
- Bolland, S., M. Llosa, P. Avila and F. de la Cruz. 1990. General organization of the conjugal transfer genes of the *IncW* plasmid R388 and interactions between R388 and *IncN* and *IncP* plasmids. *J. Bacteriol.* **172**:5795-5802.
- Bowie, J.U. and R.T. Sauer. 1990. TraY proteins of F and related episomes are members of the *arc* and *mnt* repressor family. *J. Mol. Biol.* **211**:5-6.
- Boyd, A.C. and D.J. Sheratt. 1986. Polar mobilization of the *Escherichia coli* chromosome by the *ColE1* transfer origin. *Mol. Gen. Genet.* **203**:496-504.
- Bradley, D.E. 1979. Morphology of pili determined by the N incompatibility group plasmid N3 and interaction with bacteriophages PR4 and I_{ke}. *Plasmid* **2**:632-636.
- Bradley, D.E. 1980. Morphological and serological relationships of conjugative pili. *Plasmid* **4**:155-169.
- Bradshaw, H.D. Jr., B.A. Traxler, E.G. Minkely, Jr., E.W. Nester and M.P. Gordon. 1990. Nucleotide sequence of the *traI* (helicase I) gene from the sex factor F. *J. Bacteriol.* **172**:4127-4131.

- Brasch, M.A. and R.J. Meyer. 1986. Genetic organization of plasmid R1162 DNA involved in conjugative mobilization. *J. Bacteriol.* **167**:703-710.
- Brasch, M.A. and R.J. Meyer. 1987. A 38 base-pair segment of DNA is required in *cis* for conjugative mobilization of broad host-range plasmid R1162. *J. Mol. Biol.* **198**:361-369.
- Buchanan-Wollaston, V., J.E. Passiatore and F. Cannon. 1987. The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants. *Nat.* **328**:172-175.
- Cabezón, E., E. Lanka and F. de la Cruz. 1994. Requirements for mobilization of plasmids RSF1010 and ColE1 by the IncW plasmid R388: *tnrB* and RP4 *traG* are interchangeable. *J. Bacteriol.* **176**:4455-4458.
- Carter, J.R. and R.D. Porter. 1991. *traY* and *tral* are required for *oriT*-dependant enhanced recombination between *lac*-containing plasmid and λ plac5. *J. Bacteriol.* **173**:1027-1034.
- Chatfield, L.K., E. Orr, G.J. Boulnois and B.M. Wilkins. 1982. DNA primase of plasmid ColIB is involved in conjugal DNA synthesis in donor and recipient. *J. Bacteriol.* **152**:1188-1195.
- Chatfield, L.K. and B.M. Wilkins. 1984. Conjugative transfer of IncII plasmid DNA primase. *Mol. Gen. Genet.* **197**:461-466.
- Clark, A.J. and G.J. Warren. 1979. Conjugal transmission of plasmids. *Ann. Rev. Genet.* **13**:99-125.
- Clewell, D.B. and D.R. Helinski. 1969. Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. *Proc. Natl. Acad. Sci. USA.* **62**:1159-1166.
- Clewell, D.B. and D.R. Helinski. 1970. Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. *Biochem.* **9**:4428-4440.
- Cole, S.P., E. Lanka and D.G. Guiney. 1993. Site-directed mutations in the relaxase operon of RP4. *J. Bacteriol.* **175**:4911-4916.
- Cook, D.M. and S.K. Farrand. 1992. The *oriT* region of the *Agrobacterium tumefaciens* Ti plasmid pTiC58 shares DNA sequence identity with the transfer origins of RSF1010 and RK2/RP4 and with the T-region borders. *J. Bacteriol.* **174**:6238-6246.

Coupland, G.M., A.M.C. Brown and N.S. Willetts. 1987. The origin of transfer (*oriT*) of the conjugative plasmid R46: Characterization by deletion analysis and DNA sequencing. *Mol. Gen. Genet.* 208:219-225.

Dash, P.K., B.A. Traxler, M.M. Panicker, D.D. Hackney and E.G. Minkley Jr. 1992. Biochemical characterization of *Escherichia coli* DNA helicase I. *Mol. Microbiol.* 6:1163-1172.

Datta, N. 1974. Epidemiology and classification of plasmids. In *Microbiology- 1974*. D. Schlessinger (ed). Am. Soc. Microbiol. Washington. pp. 9-15

Datta, N. and V.M. Hughes. 1983. Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics. *Nature* 306:616-617.

Derbyshire, K.M., G. Hatful and N.S. Willetts. 1987. Mobilization of the non-conjugative plasmid RSF1010: A genetic analysis of its origin of transfer. *Mol. Gen. Genet.* 206:154-160.

Di Laurenzio, L., L.S. Frost, B.B. Finlay and W. Paranchych. 1991. Characterization of the *oriT* region of the IncFV plasmid pED208. *Mol. Microbiol.* 5:1779-1790.

Di Laurenzio, L., L.S. Frost, W. Paranchych. 1992. The TraM protein of the conjugative plasmid F binds to the origin of transfer of the F and ColE1 plasmids. *Mol. Microbiol.* 6:2952-2959.

Dobritsa, A.P., T.G. Mikhailova and V.I. Dubovaya. 1985. Physical and genetic structure of the IncN plasmid R15. *Plasmid* 14:99-105.

Dotto, G.P., K. Horiuchi and N.D. Zinder. 1982. Initiation and termination of phage ϕ 1 plus-strand synthesis. *Proc. Natl. Acad. Sci. USA.* 79:7122-7126.

Drolet, M. and P.C.K. Lau. 1992. Mobilization protein-DNA binding and divergent transcription at the transfer origin of the *Thiobacillus ferrooxidans* pTF1 plasmid. *Mol. Microbiol.* 6:1061-1071.

Durrenberger, F., A. Cramer, B. Hohn and Z. Koukolikova-Nicola. 1989. Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. *Proc. Natl. Acad. Sci. USA.* 86:9154-9158.

Eckert, K.A. and T.A. Kunkel. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nuc. Acids Res.* 18:3739-3743.

Erickson, M.J. and R.J. Meyer. 1993. The origin of greater-than-unit-length plasmid generated during bacterial conjugation. *Mol. Microbiol.* 7:289-298.

- Everett, R. and N. Willetts. 1980. Characterisation of an *in vivo* system for nicking at the origin of conjugal DNA transfer of the sex factor F. *J. Mol. Biol.* **136**:129-150.
- Farrand, S.K., I. Hwang and D. Cook. 1996. The *tra* region of the nopaline-type Ti plasmid is a chimera with elements related to the transfer systems of RSF1010, RP4 and F. *J. Bacteriol.* **178**:4233-4247.
- Filichkin, S.A. and S.B. Gelvin. 1993. Formation of a putative relaxation intermediate during T-DNA processing directed by the *Agrobacterium tumefaciens* VirD1, D2 endonuclease. *Mol. Microbiol.* **8**:915-926.
- Finlay, B.B., L.S. Frost and W. Paranchych. 1986₁. Nucleotide sequences of the R1-19 plasmid transfer genes *traM*, *finP*, *traJ*, *traY*, and the *traYZ* promoter. *J. Bacteriol.* **166**:368-374.
- Finlay, B.B., L.S. Frost and W. Paranchych. 1986₂. Origin of transfer of IncF plasmids and nucleotide sequences of the type II *oriT*, *traM*, and *traY* alleles from ColB4-K98 and the type IV *traY* allele from R100-1. *J. Bacteriol.* **168**:132-139.
- Fowler, T., L. Taylor and R. Thompson. 1983. The control region of the F plasmid transfer operon: DNA sequence of the *traJ* and *traY* genes and characterisation of the *traY*->*traZ* promoter. *Gene* **26**:79-89.
- Frost, L. 1993. Conjugative pili and pilus-specific phages. *in*, Bacterial conjugation. ed. D.B. Clewell. Plenum Press. New York. pp 189-222.
- Frost, L.S., K. Ippen-Ihler and R.A. Skurray. 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58**:162-210.
- Fu, Y-H.F., M-M. Tsai, Y. Luo and R.C. Deonier. 1991. Deletion analysis of the F plasmid *oriT* locus. *J. Bacteriol.* **173**:1012-1020.
- Fukuda, H and E. Ohtsubo. 1995. Large scale purification and characterization of TraI endonuclease encoded by sex factor plasmid R100. *J. Biol. Chem.* **270**:21319-21325.
- Furste, J.P., W. Pansegrau, G. Ziegelin, M. Kroger and E. Lanka. 1989. Conjugative transfer of promiscuous IncP plasmids: Interaction of plasmid-encoded products with the transfer origin. *Proc. Natl. Acad. Sci. USA* **86**:1771-1775.
- Furuya, N. and T. Komano. 1991. Determination of the nick site at *oriT* of IncI1 plasmid R64: Global similarity of *oriT* structures of IncI1 and IncP plasmids. *J. Bacteriol.* **173**:6612-6617.
- Furuya, N., T. Nisioka and T. Komano. 1991. Nucleotide sequences and functions of

the *oriT* operon of IncII plasmid R64. *J. Bacteriol.* **173**:2231-2237.

Furuya, N. and T. Komano. 1995. Specific binding of the NikA protein to one arm of 17-base-pair inverted repeat sequences within the *oriT* region of plasmid R64. *J. Bacteriol.* **177**:46-51.

Gao, Q., Y. Luo and R.C. Deonier. 1994. Initiation and termination of DNA transfer at F plasmid *oriT*. *Mol. Microbiol.* **11**:449-458.

Ghai, J. and A. Das. 1989. the *virD* operon of *Agrobacterium tumefaciens* Ti plasmid encodes a DNA-relaxing enzyme. *Proc. Natl. Acad. Sci. USA* **86**:3109-3113.

Gielen, B., M. De Beuckeleer, J. Seurinck, F. Deboeck, H. De Greve, M. Lemmers, M. Van Montagu and J. Schell. 1984. The complete nucleotide sequence of the TL-DNA of the plasmid pTiAch5. *EMBO J.* **3**:835-846.

Glocker, B. and I. Rasched. 1990. Genetic analysis of the transfer region of the IncN plasmid N3. *Res. Microbiol.* **141**:621-631.

Goldner, A., H. Graus and G. Hogenauer. 1987. The origin of transfer of p307. *Plasmid* **18**:76-83.

Gorbalenya, A.E. and E.V. Koonin. 1990. Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. *J. Mol. Biol.* **213**:583-591.

Grandoso, G., M. Llosa, J.C. Zabala and F. de la Cruz. 1994. Purification and biochemical characterization of TrwC, the helicase involved in plasmid R388 conjugal DNA transfer. *Eur. J. Biochem.* **226**:403-412.

Gruss, A. and S.D. Ehrlich. 1989. The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. *Microbiol. Rev.* **53**:231-241.

Guiney, D.G. and D.R. Helinski. 1975. Relaxation complexes of plasmid DNA and protein. III. Association of protein with the 5' terminus of the broken DNA strand in the relaxed complex of plasmid ColE1. *J. Biol. Chem.* **250**:8796-8803.

Guiney, D.G. and D.R. Helinski. 1979. The DNA-Protein relaxation complex of the plasmid RK2: Location of the site-specific nick in the region of the proposed origin of transfer. *Mol. Gen. Genet.* **176**:183-189.

Guiney, D.G. and E. Yakobson. 1983. Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. *Proc. Natl. Acad. Sci. USA.* **80**:3595-3598.

Guiney, D.G., C. Deiss and V. Simnad. 1988. Location of the relaxation complex nick site with the minimal origin of transfer region of RK2. *Plasmid* 20:259-265.

Guiney, D.G. 1989. Conjugative transfer of IncP plasmids. *in*, Promiscuous plasmids of Gram-negative bacteria. *ed.* C.M. Thomas. Academic Press. Toronto. pp 27-56.

Guiney, D.G. and E. Lanka. 1989. Conjugative transfer of IncP plasmids. *in*, Promiscuous plasmids of Gram-negative bacteria. *ed.* C.M. Thomas. Academic Press. Toronto. pp 27-56.

Guiney, D.G. 1993. Broad host range conjugative and mobilizable plasmids in Gram-negative bacteria. *in*, Bacterial conjugation. *ed.* D.B. Clewell. Plenum Press. New York. pp 75-104.

Guyer, M.S., N. Davidson and A.J. Clark. 1977. Heterduplex analysis of *tra* Δ F' plasmids and the mechanism of their formation. *J. Bacteriol.* 131:970-980.

Hames, B.D. 1981. An introduction to polyacrylamide gel electrophoresis. *in* Gel electrophoresis of proteins. A practical approach. *eds* B.D. Hames and D. Rickwood. IRL Press Ltd. Washington DC.

Hanai, R. and J.C. Wang. 1993. The mechanism of sequence specific DNA cleavage and strand transfer by ϕ X174 gene A' protein. *J. Biol. Chem.* 268:23830-23838.

Harrington, L.C. and A.C. Rogerson. 1990. The F pilus of *Escherichia coli* appears to support stable DNA transfer in the absence of wall-to-wall contact between cells. *J. Bacteriol.* 172:7263-7264.

Hengen, P.N. and V.N. Iyer. 1992. DNA cassettes containing the origin of transfer (*oriT*) of two broad-host-range transfer systems. *Biotechniques* 13:58-62

Herrera-Estrella, A., Z-M. Chen, M. Van Montagu and K. Wang. 1988. VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent DNA-protein complex at the 5' terminus of T-strand molecules. *EMBO J.* 7:4055-4062.

Higgins, D.G. and P.M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS* 5:151-153.

Hirayama, T., T. Muranaka, H. Ohkawa and A. Oka. 1988. Organization and characterization of the *virCD* genes from *Agrobacterium rhizogenes*. *Mol. Gen. Genet.* 213:229-237.

Holmes, D.S. and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114:193-197.

Horowitz, B. and R.C. Deonier. 1985. Formation of Δtra F' plasmids: specific recombination at *oriT*. *J. Mol. Biol.* **186**:267-274.

Howard, E.A., B.A. Winsor, G. De Vos and P. Zambryski. 1989. Activation of the T-DNA transfer process in *Agrobacterium* results in the generation of a T-strand protein complex: tight association of VirD2 with the 5' ends of T-strands. *Proc. Natl. Acad. Sci. USA.* **86**:4017-4021.

Howard, M.T., W.C. Nelson and S.W. Matson. 1995. Stepwise assembly of a relaxosome at the F plasmid origin of transfer. *J. Biol. Chem.* **270**:28381-28386.

Howland, C.J. and B.M. Wilkins. 1988. Direction of conjugative transfer of IncII plasmid ColIb-P9. *J. Bacteriol.* **170**:4958-4959.

Hughes, V.M. and N. Datta. 1983. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. *Nature* **302**: 725-726.

Inamoto, S., Y. Yoshioka and E. Ohtsubo. 1988. Identification and characterization of the products from the *traJ* and *traY* genes of plasmid R100. *J. Bacteriol.* **170**:2749-2757.

Inamoto, S. and E. Ohtsubo. 1990. Specific binding of the TraY protein to *oriT* and the promoter region for the *traY* gene of plasmid R100. *J. Biol. Chem.* **265**:6461-6466.

Inamoto, S., T. Abo and E. Ohtsubo. 1990. Binding sites of integration host factor in *oriT* of plasmid R100. *J. Gen. Appl. Microbiol.* **36**:287-293.

Inamoto, S., Y. Yoshioka and E. Ohtsubo. 1991. Site- and strand-specific nicking *in vitro* at *oriT* by the TraY-Tral endonuclease of plasmid R100. *J. Biol. Chem.* **266**:10086-10092.

Inamoto, S., H. Fukuda, T. Abo and E. Ohtsubo. 1994. Site- and strand-specific nicking at *oriT* of plasmid R100 in a purified system: enhancement of the nicking activity of Tral (helicase I) with TraY and IHF. *J. Biochem.* **116**:838-844.

Ippen-Ihler, K.A. and R.A. Skurray. 1993. Genetic origination of transfer-related determinants on the sex factor and related plasmid. *in*, *Bacterial conjugation*. ed. D.B. Clewell. Plenum Press. New York. pp 23-52.

Ippen, K.A. and R.C. Valentine. 1967. The sex hair of *E. coli* as a sensory fibre, conjugation tube or mating arm. *Biochem. Biophys. Res. Comm.* **27**:674-680.

Iyer, V.N. 1989. IncN group plasmids and their genetic systems. *In* *Promiscuous plasmids of gram-negative bacteria*. Academic Press. London. pp. 165-183.

- Jacobson, A. 1972. Role of F pili in the penetration of bacteriophage f1. *J. Virol.* **10**:835-843.
- Jalajakumari, M.B. and P.A. Manning. 1989. Nucleotide sequence of the *traD* region in *Escherichia coli* F sex factor. *Gene* **81**:195-202.
- Jasper, F., C. Koncz, J. Schell and H-H. Steinbiss. 1994. *Agrobacterium* T-strand production *in vitro*: sequence-specific cleavage and 5' protection of single-stranded DNA templates by purified VirD2 protein. *Proc. Natl. Acad. Sci. USA.* **91**:694-698.
- Jayaswal, R.K., K. Veluthambi, K., S.B. Gelvin and J.L. Slightom. 1987. Double-stranded cleavage of T-DNA and generation of single-stranded T-DNA molecules in *Escherichia coli* by a *virD*-encoded border-specific endonuclease from *Agrobacterium tumefaciens*. *J. Bacteriol.* **169**:5035-5045.
- Jessee, J. 1986. New subcloning efficiency competent cells: $> 10^6$ transformants/ μ g. Bethesda Research Laboratories Focus **8**:4
- Jouanin, L., D. Bouchez, R.F. Drong, D. Tepfer and J.L. Slightom. 1989. Analysis of TR-DNA/plant junctions in the genome of a *Convolvulus arvensis* clone transformed by *Agrobacterium rhizogenes* strain A4. *Plant Mol. Biol.* **12**:75-85.
- Kado, C.I. 1993. *Agrobacterium*-mediated transfer and stable incorporation of foreign genes in plants. *in*, Bacterial conjugation. *ed* D.B. Clewell. New York. pp 243-254.
- Kahn, M., R. Kolter, C. Thomas, D. Figurski, R. Meyer, E. Remaut and D.R. Heninski. 1979. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K and RK2. *Methods Enzymol.* **68**:268-280.
- Keen, N.T., S. Tamaki, D. Kobayashi and D. Tollinger. 1988. Improved broad-host-range plasmids for DNA cloning in Gram-negative bacteria. *Gene* **70**:191-197.
- Khatoon, H., R.V. Iyer and V.N. Iyer. 1972. A new filamentous bacteriophage with sex-factor specificity. *Virol.* **48**:145-155.
- Kilbane, J.J. and M.H. Malamy. 1980. F factor mobilization of non-conjugative chimeric plasmids in *Escherichia coli*: general mechanisms and a role for site-specific *recA*-independent recombination at *oriV1*. *J. Mol. Biol.* **143**:73-93.
- Kim, K. and R.J. Meyer. 1989. Unidirectional transfer of broad host-range plasmid R1162 during conjugative mobilization. Evidence for genetically distinct events at *oriT*. *J. Mol. Biol.* **208**:501-505.
- Kline, B.C. and D.R. Helinski. 1971. F₁ sex factor of *Escherichia coli*. Size and

purification in the form of a strand-specific relaxation complex of supercoiled deoxyribonucleic acid and protein. *Biochem.* **10**:4975-4980.

Klinkert, M-Q., A. Klein and M. Abdel-Monem. 1980. Studies on the functions of DNA helicase I and DNA helicase II of *Escherichia coli*. *J. Biol. Chem.* **255**:9746-9752.

Komano, T., A. Toyoshima, K. Morita and T. Nisioka. 1988. Cloning and nucleotide sequence of the *oriT* region of the IncII plasmid R64. *J. Bacteriol.* **170**:4385-4387.

Konarska-Kozłowska, M. and V.N. Iyer. 1981. The genetic and physical basis of variability in *Escherichia coli* strains carrying a reference IncN group plasmid. *Can. J. Microbiol.* **27**:616-626.

Koukolikova-Nicola, Z., R.D. Shillito, B. Hohn, K. Wang, M. Van Montagu and P. Zambryski. 1985. Involvement of circular intermediates in the transfer of T-DNA from *Agrobacterium tumefaciens* to plant cells. *Nat.* **313**:191-196.

Krishnan, B.R., Kim, H-Y. and V.N. Iyer. 1991. pMUR274 and pMUR545 are not IncN group plasmids. *Plasmid* **26**:78-81.

Kuhn, B., M. Abdel-Monem and H. Hoffmann-Berling. 1978. DNA Helicases. Cold Spring Harbour Symp. Quant. Biol. **43**:63-67.

Kuhn, B., M. Abdel-Monem, H. Krell and H. Hoffmann-Berling. 1979. Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. *J. Biol. Chem.* **254**:11343-11350.

Kupersztoch-Portnoy, Y.M., M.A. Lovett and D.R. Helinski. 1974₁. Strand and site specificity of the relaxation event for the relaxation complex of the antibiotic resistance plasmid R6K. *Biochem.* **13**:5484-5490.

Kupersztoch-Portnoy, Y.M., G.L.G. Miklos and D.R. Helinski. 1974₂. Properties of the relaxation complexes of supercoiled deoxyribonucleic acid and protein of the R plasmids R64, R28K and R6K. *J. Bacteriol.* **120**:545-548.

Kyte, J. and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105-132.

Lahue, E.E. and S.W. Matson. 1988. *Escherichia coli* DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. *J. Biol. Chem.* **263**:3208-3215.

Lahue, E.E. and S.W. Matson. 1990. Purified *Escherichia coli* F-Factor TraY protein binds at *oriT*. *J. Bacteriol.* **172**:1385-1391.

- Lanka, E., E. Scherzinger, E. Gunther and H. Schuster. 1979. A DNA primase specified by I-like plasmids. *Proc. Natl. Acad. Sci. USA* **76**:3632-3636.
- Lanka, E. and P.T. Barth. 1981. Plasmid RP4 specifies a deoxyribonucleic acid primase involved in its conjugal transfer and maintenance. *J. Bacteriol.* **148**:769-781.
- Lanka, E., R. Lurz, M. Kroger and J.P. Furste. 1984. Plasmid RP4 encodes two forms of a DNA primase. *Mol. Gen. Genet.* **194**:65-72.
- Lanka, E., J.P. Furste, E. Yakobson and D.G. Guiney. 1985. Conserved regions at the DNA primase locus of the IncP α and IncP β plasmids. *Plasmid* **14**:217-223.
- Lanka, E. and B.M. Wilkins. 1995. DNA processing reactions in bacterial conjugation. *Ann. Rev. Biochem.* **64**:141-169.
- Lessl, M., W. Pansegrau and E. Lanka. 1992. Relationship of DNA-transfer-systems: essential transfer factors of plasmids RP4, Ti and F share common sequences. *Nuc. Acids Res.* **20**:6099-6100.
- Lessl, M., D. Balzer, K. Weyrauch and E. Lanka. 1993. The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor specific phage propagation. *J. Bacteriol.* **175**:6415-6425.
- Lessl, M. and E. Lanka. 1994. Common mechanisms in bacterial conjugation and Ti-mediated T-DNA transfer to plant cells. *Cell* **77**:321-324.
- Liss, L.R. 1987. New M13 host: DH5 α F' competent cell. *Focus* **9**:13.
- Llosa, M., S. Bolland and F. de la Cruz. 1991. Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. *Mol. Gen. Genet.* **226**:467-472.
- Llosa, M., S. Bolland and F. de la Cruz. 1994₁. Conjugation-independent, site-specific recombination at the *oriT* of the IncW plasmid R388 mediated by TrwC. *J. Bacteriol.* **176**:3210-3217.
- Llosa, M., S. Bolland and F. de la Cruz. 1994₂. Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J. Mol. Biol.* **235**:448-464.
- Loewen, P.C. and J. Switala. 1995. Template secondary structure can increase the error frequency of the DNA polymerase from *Thermus aquaticus*. *Gene* **164**:59-63.
- Lopez, J., L. Salazar, I. Andres, J.M. Ortiz and J.C. Rodriguez. 1991. Nucleotide sequence of the *oriT-traM-finP* region of the haemolytic plasmid pSU316: comparison

to F. *Nuc. Acids Res.* **19**:3451-3451.

Lovett, M.A., D.G. Guiney and D.R. Helinski. 1974. Relaxation complexes of plasmids ColE1 and ColE2: unique site of the nick in the open circular DNA of the relaxed complexes. *Proc. Natl. Acad. Sci. USA.* **71**:3854-3857

Lovett, M.A and D.R. Helinski. 1975. Relaxation complexes of plasmid DNA and protein. II. Characterization of the proteins associated with the unrelaxed and relaxed complexes of plasmid ColE1. *J. Biol. Chem.* **250**:8790-8795.

Luo, Y., Q. Gao and R.C. Deonier. 1994. Mutational and physical analysis of F plasmid *traY* protein binding to *oriT*. *Mol. Microbiol.* **11**:459-469.

Luo, Y., Q. Gao and R.C. Deonier. 1995. Boundaries of the nicking region for the F plasmid transfer origin, *oriT*. *Mol. Microbiol.* **15**:829-837.

Machida, Y., S. Usami, A. Yamamoto, Y. Niwa and I. Takebe. 1986. Plant-inducible recombination between the 25 bp border sequences of T-DNA in *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **204**:374-382.

Manoil, C. and J. Beckwith. 1985. *TnphoA*: A transposon probe for protein export signals. *Proc. Natl. Acad. Sci. USA* **82**:8129-8133.

Matson, S.W. and B.S. Morton. 1991. *Escherichia coli* DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid *oriT*. *J. Biol. Chem.* **266**:16232-16237.

Matson, S.W., W.C. Nelson and B.S. Morton. 1993. Characterization of the reaction product of the *oriT* nicking reaction catalyzed by *Escherichia coli* DNA helicase I. *J. Bacteriol.* **175**:2599-2606.

McIntire, S.A. and W.B. Dempsey. 1987. *oriT* sequence of the antibiotic resistance plasmid R100. *J. Bacteriol.* **169**:3829-3832.

Merryweather, A., P.T. Barth and B.M. Wilkins. 1986₁. Role and specificity of plasmid RP4-encoded DNA primase in bacterial conjugation. *J. Bacteriol.* **167**:12-17.

Merryweather, A., C.E.D. Rees, N.M. Smith and B.M. Wilkins. 1986₂. Role of *sog* polypeptides specified by plasmid Collb-P9 and their transfer between conjugative bacteria. *EMBO J.* **5**:3007-3012.

Meyer, R. 1989. Site-specific recombination at *oriT* of plasmid R1162 in the absence of conjugative transfer. *J. Bacteriol.* **171**:799-806.

- Nelson, W.C., B.S. Morton, E.E. Lahue and S.W. Matson. 1993. Characterization of the *Escherichia coli* F factor *traY* gene product and its binding sites. *J. Bacteriol.* **175**:2221-2228.
- Nelson, W.C., M.T. Howard, J.A. Sherman and S.W. Matson. 1995. The *traY* gene product and integration host factor stimulate *Escherichia coli* DNA helicase I-catalyzed nicking at the F plasmid *oriT*. *J. Biol. Chem.* **270**:28374-28380.
- Nijkamp, H.J.J., R. DeLang, A.R. Stuitj, P.J.M. Van Den Elzen, E. Veltkamp and A.J. Van Putten. 1986. The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13. *Plasmid* **16**:135-160.
- Nomura, N., H. Masai, M. Inuzuka, C. Miyazaki, E. Ohtsubo, T. Itoh, S. Sasamoto, M. Matsui, R. Ishizaki and K-I. Arai. 1991. Identification of eleven single-strand initiation sequences (*ssr*) for priming of DNA replication in the F, R6K, R100 and ColE2 plasmids. *Gene* **108**:15-22.
- Norrande, J., T. Kempe and J. Messing. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* **26**:101-106.
- O'Connor, M.B. and M.H. Malamy. 1984. Site-specific recombination in the *oriV* region of the F factor. *Cold Spring Harbor Symp. Quant. Biol.* **49**:421-434.
- Ohki, M. and J-I. Tomizawa. 1970. Asymmetric transfer of DNA strands in bacterial conjugation. *Cold Spring Harbour Symp. Quant. Biol.* **33**:651-658.
- Ostermann, E., F. Kricek and G. Hogenauer. 1984. Cloning the origin of transfer region of the resistance plasmid R1. *EMBO J.* **3**:1731-1735.
- Panicker, M.M. and E.G. Minkley, Jr. 1985. DNA transfer occurs during a cell surface contact stage of F sex factor-mediated bacterial conjugation. *J. Bacteriol.* **162**:584-590.
- Panicker, M.M. and E.G. Minkley. 1992. Purification and properties of the F sex factor TraD protein, an inner membrane conjugal transfer protein. *J. Biol. Chem.* **267**:12761-12766.
- Pansegau, W., G. Ziegelin and E. Lanka. 1988. The origin of conjugative IncP plasmid transfer. Interaction with plasmid-encoded products and the nucleotide sequence at the relaxation site. *Biochim. Biophys. Acta.* **951**:365-374.
- Pansegau, W., D. Balzer, V. Kruft, R. Lurz and E. Lanka. 1990. *In vitro* assembly of relaxosomes at the transfer origin of plasmid RP4. *Proc. Natl. Acad. Sci. USA* **87**:6555-6559.

- Pansegrau, W., G. Ziegelin and E. Lanka. 1990₂. Covalent association of the *tral* gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation nick site. *J. Biol. Chem.* **265**: 10637-10644.
- Pansegrau, W. and E. Lanka. 1991. Common sequence motifs in DNA relaxases and nick regions from a variety of DNA transfer systems. *Nuc. Acid Res.* **19**:3455-3455.
- Pansegrau, W., F. Shoumacher, B. Hohn and E. Lanka. 1993₁. Site-specific cleavage and joining of single-stranded DNA by VirD2 protein of *Agrobacterium tumefaciens* Ti plasmids: Analogy to bacterial conjugation. *Proc. Natl. Acad. Sci. USA.* **90**:11538-11542.
- Pansegrau, W., W. Schroder and E. Lanka. 1993₂. Relaxase (Tral) of IncP α plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. *Proc. Natl. Acad. Sci. USA.* **90**:2925-2929.
- Pansegrau, W., W. Schroder and E. Lanka. 1994. Concerted action of three distinct domains in the DNA cleaving-joining reaction catalyzed by relaxase (Tral) of conjugative plasmid RP4. *J. Biol. Chem.* **269**:2782-2789.
- Pansegrau, W. and E. Lanka. 1996. Mechanisms of initiation and termination reactions in conjugative DNA processing. *J. Biol. Chem.* **271**:13068-13076.
- Paterson, E.S. and V.N. Iyer. 1992. The *oriT* region of the conjugative transfer system of plasmid pCU1 and specificity between it and the *mob* region of other *N tra* plasmids. *J. Bacteriol.* **174**:499-507.
- Pohlman, R.F., H.D. Genetti and S.C. Winans. 1994. Common ancestry between IncN conjugal transfer genes and macromolecular export systems of plant and animal pathogens. *Mol. Microbiol.* **14**:655-668.
- Porter, S.G., M.F. Yanofsky and E.W. Nester. 1987. Molecular characterization of the *virD* operon from *Agrobacterium tumefaciens*. *Nuc. Acids Res.* **15**:7503-7517.
- Reinmunn, C. and D. Haas. 1993. Mobilization of chromosomes and non-conjugative plasmids by cointegrative mechanisms. *In*, Bacterial conjugation. *ed.* D.B. Clewell. Plenum Press. New York. pp 137-188.
- Rees, C.E.D. and B.M. Wilkins. 1989. Transfer of *tra* proteins into the recipient cell during bacterial conjugation mediated by plasmid ColIb-P9. *J. Bacteriol.* **171**:3152-3157.
- Rees, C.E.D. and B.M. Wilkins. 1990. Protein transfer into the recipient cell during bacterial conjugation: Studies with F and RP4. *Mol. Microbiol.* **4**:1199-1205.

- Reygers, U., R. Wessel, H. Muller and H. Hoffmann-Berling. 1991. Endonuclease activity of *Escherichia coli* DNA helicase I directed against the transfer origin of the F factor. *EMBO* 10:2689-2694.
- Rohrer, J. and D.E. Rawlings. 1992. Sequence analysis and characterization of the mobilization region of a broad-host range plasmid, pTF-FC2, isolated from *Thiobacillus ferrooxidans*. *J. Bacteriol.* 174:6230-6237.
- Rohrer, J. and D.E. Rawlings. 1993. Regulation of mobilization of the broad-host-range plasmid pTF-FC2. *Mol. Microbiol.* 9:1051-1059.
- Rupp, W.D. and G. Ihler. 1970. Strand separation during bacterial mating. *Cold Spring Harbour Symp. Quant. Biol.* 33:647-650.
- Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular cloning. A laboratory manual. 2nd ed. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press.
- Schaffer, H.E. and R.R. Sederoff. 1981. Improved estimation of DNA fragment lengths from agarose gels. *Anal. Biochem.* 115:113-122.
- Scheiffele, P., W. Pansegrau and E. Lanka. 1995. Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of superhelical T-border DNA *in vitro*. *J. Biol. Chem.* 270:1269-1276.
- Scherzinger, E. R., Lurz, S. Otto and D. Dobrinski. 1992. *In vitro* cleavage of double and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. *Nuc. Acids Res.* 20:41-48.
- Scholz, P., V. Haring, B. Whittmann-Liebold, K. Ashman, M. Bagdasarian and E. Scherzinger. 1989. Complete nucleotide sequence and gene organization of broad host-range plasmid RSF1010. *Gene* 75:271-288.
- Schwabb, M., H. Gruber and G. Hogenauer. 1991. The TraM protein of plasmid R1 is a DNA binding protein. *Mol. Microbiol.* 5:439-446.
- Schwabb, M., H. Reisenzein and G. Hogenauer. 1993. TraM of plasmid R1 plasmid regulates its own expression. *Mol. Microbiol.* 7:975-803.
- Silverman, P.M., E. Wichersham and R. Harris. 1991. Regulation of the F plasmid *traY* promoter in *Escherichia coli* by host and plasmid factors. *J. Mol. Biol.* 218:119-128.

- Simon, R., U. Priefer and A. Puhler. 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Biotechnol.* 1:784-791.
- Sims, J. and E.W. Benz, Jr. 1980. Initiation of DNA replication by the *Escherichia coli dnaG* protein: evidence that tertiary structure is involved. *Proc. Natl. Acad. Sci. USA.* 77:900-904
- Smith, C.A. and C.M. Thomas. 1989. Relationships and evolution of IncP plasmids. *in*, Promiscuous plasmids of Gram-negative bacteria. *ed.* C.M. Thomas. Academic Press. Toronto. pp 57-78.
- Snijder, A., A.J. van Putten, E. Veltkamp and H.J.J. Nijkamp. 1983. Localization and nucleotide sequence of the *bom* region of Cl₆DF13. *Mol. Gen. Genet.* 192:444-451.
- Stachel, S.E., B. Timmermann and P. Zambryski. 1986. Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens* to plant cells. *Nat.* 322:706-712.
- Stachel, S.E., B. Timmermann and P. Zambryski. 1987. Activation of *Agrobacterium tumefaciens vir* gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region: requirement for 5' *virD* gene products. *EMBO J.* 6:857-863.
- Strack, B., M. Lessl, R. Calendar and E. Lanka. 1992. A common sequence motif, E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P- type DNA primases and the α protein of the *Escherichia coli* satellite phage P4. *J. Biol. Chem.* 267:13062-13072.
- Studier, F.W., A.H. Rosenberg, J.J. Dunn and J.W. Dunbendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* 185:60-89.
- Thatte, V., D.E. Bradley and V.N. Iyer. 1985. N conjugative transfer system of plasmid pCU1. *J. Bacteriol.* 163:1229-1236.
- Thompson, R., L. Taylor, K. Kelly, R. Everett and N. Willetts. 1984. The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin specific nicks. *EMBO* 3:1175-1180.
- Thompson, T.L., M.B. Centola and R.C. Deonier. 1989. Location of the nick at *oriT* of the F plasmid. *J. Mol. Biol.* 207:505-512.
- Traxler, B.A. and E.G. Minkley, Jr. 1987. Revised genetic map of the distal end of the F transfer operon: implications for DNA Helicase I, nicking at *oriT*, and conjugal DNA transport. *J. Bacteriol.* 169:3251-3259.

- Traxler, B.A. and E.G. Minkley, Jr. 1988. Evidence that DNA helicase I and *oriT* site-specific nicking are both function of the F Tral protein. *J. Mol. Biol.* **204**:205-209.
- Tsai, M.-M., Y.-H. F. Fu and R.C. Deonier. 1990. Intrinsic bends and integration host factor binding at F plasmid *oriT*. *J. Bacteriol.* **172**:4603-4609.
- Valentine, C.R.I and C.I Kado. 1989. Molecular genetics of IncW plasmids. *in*. Promiscuous plasmids of Gram negative plasmids. *ed.* C.M. Thomas. Academic Press. Toronto. pp 124-164.
- van de Pol, H., E. Velkamp and H.J.J. Nijkamp. 1978. Genetic analysis of the mobilization of the non-conjugative plasmid CloDF13. *Mol. Gen. Genet.* **160**:134-149.
- van Mansfeld, A.D.M., H.A.A.M. van Teeffelen, P.D. Baas and H.S. Jansz. 1986. Two juxtaposed tyrosyl-OH groups participate in ϕ X174 gene A protein catalysed cleavage and ligation of DNA. *Nuc. Acids Res.* **14**:4229-4239.
- Veluthambi, K., R.K. Jayaswal and S.B. Gelvin. 1987. Virulence genes *A*, *G*, and *D* mediate the double-stranded border cleavage of T-DNA from the *Agrobacterium* Ti plasmid. *Proc. Natl. Acad. Sci. USA.* **84**:1881-1885.
- Vieira, J. and J. Messing. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**:3-11.
- Vielmetter, W., F. Bonhoeffer and A. Schutte. 1968. Genetic evidence for transfer of a single DNA strand during bacterial conjugation. *J. Mol. Biol.* **37**:81-86.
- Vogel, A.M. and A. Das. 1992₁. Mutational analysis of *Agrobacterium tumefaciens virD2*: tyrosine 29 is essential for endonuclease activity. *J. Bacteriol.* **174**:303-308.
- Vogel, A.M. and A. Das. 1992₂. The *Agrobacterium tumefaciens* VirD3 gene is not essential for tumorigenicity on plants. *J. Bacteriol.* **174**:5161-5164.
- Wang, K., A. Herrera-Estrella and M. Van Montagu. 1990. Overexpression of *virD1* and *virD2* genes in *Agrobacterium tumefaciens* enhances T-complex formation and plant transformation. *J. Bacteriol.* **172**:4432-4440.
- Wang, K., S.E. Stachel, B. Timmerman, M. Van Montagu and P. Zambryski. 1987. Site-specific nick in the T-DNA border sequence as a result of *Agrobacterium vir* gene expression. *Sci.* **235**:587-591.
- Ward, E.R. and W.M. Barnes. 1988. VirD2 protein of *Agrobacterium tumefaciens* very tightly linked to the 5' end of T-strand DNA. *Sci.* **242**:927-930.

- Warren, G.J., A.J. Twigg and D.J. Sherratt. 1978. ColE1 plasmid mobility and relaxation complex. *Nat.* **274**:259-261.
- Warren, G.J. and A.J. Clark. 1979. Conjugal transmission of plasmids. *Ann. Rev. Genet.* **13**:99-125.
- Warren, G.J., M.W. Saul and D. J. Sherratt. 1979. ColE1 plasmid mobility: essential and conditional functions. *Mol. Gen. Genet.* **170**:103-107.
- Warren, G.J. and A.J. Clark. 1980. Sequence-specific recombination of plasmid ColE1. *Proc. Natl. Acad. Sci. USA.* **77**:6724-6728.
- Waters, V.L., K.H. Hirata, W. Pansegrau, E. Lanka and D.G. Guiney. 1991. Sequence identity in the nick regions of IncP plasmid transfer origins and T-DNA borders of *Agrobacterium* Ti plasmids. *Proc. Natl. Acad. Sci. USA* **88**:1456-1460.
- Waters, V.L. and D.G. Guiney. 1993. Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. *Mol. Microbiol.* **9**:1123-1130.
- Wessel, R., H. Muller and H. Hoffmann-Berling. 1990. Electron microscopy of DNA-helicase-I complexes in the act of strand separation. *Eur. J. Biochem.* **189**:277-285.
- Wilkins, B.M. 1975. Partial suppression of the phenotype of *Escherichia coli* K-12 *dnaG* mutants by some I-like conjugative plasmids. *J. Bacteriol.* **122**:899-904.
- Wilkins, B.M., G.J. Boulnois and E. Lanka. 1981. A plasmid DNA primase activity in discontinuous bacterial DNA replication. *Nat.* **290**:217-221.
- Wilkins, B.M. and E. Lanka. 1993. DNA processing and replication during plasmid transfer between Gram-negative bacteria. *in*, Bacterial conjugation. *ed.* D.B. Clewell. Plenum Press. New York. pp 105-136.
- Willetts, N. 1980. Interactions between the F conjugal transfer system and CloDF13::Tn4 plasmids. *Mol. Gen. Genet.* **180**:213-217.
- Willetts, N. and R. Skurray. 1980. The conjugation system of F-like plasmids. *Ann. Rev. Genet.* **14**:41-76.
- Willetts, N. and C. Crowther. 1981. Mobilization of the non-conjugative IncQ plasmid RSF1010. *Genet. Res.* **37**:311-316.
- Willetts, N. and B. Wilkins. 1984. Processing of plasmid DNA during bacterial conjugation. *Microbiol. Rev.* **48**:24-41.

- Willetts, N. and J. Maule. 1985. Specificities of IncF plasmid conjugation genes. *Genet. Res.* 47:1-11.
- Winans, S.C. and G.C. Walker. 1985₁. Conjugal transfer system of the IncN plasmid pKM101. *J. Bacteriol.* 161:402-410.
- Winans, S.C. and G.C. Walker. 1985₂. Fertility inhibition of RP1 by IncN plasmid pKM101. *J. Bacteriol.* 161:425-427.
- Yakobson, E. and G. Guiney. 1983. Homology in the transfer origins of broad host range IncP plasmids: definition of two subgroups of P plasmids. *Mol. Gen. Genet.* 192:436-438.
- Yakobson, E., C. Deiss, K. Hirata and D.G. Guiney. 1990. Initiation of DNA synthesis in the transfer origin region of RK2 by the plasmid-encoded primase: detection using defective M13. *Plasmid* 23:80-84.
- Yanofsky, M.F., S.G. Porter, C. Young, L.M. Albright, M.P. Gordon and E.W. Nester. 1986. The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease. *Cell* 47:471-477.
- Yoshioka, Y., Y. Fujita and E. Ohtsubo. 1990. Nucleotide sequence of the promoter-distal region of the *tra* operon of plasmid R100, including *traI* (DNA helicase I) and *traD* genes. *J. Mol. Biol.* 214:39-53.
- Young, C. and E.W. Nester. 1988. Association of the VirD2 protein with the 5' end of T strands in *Agrobacterium tumefaciens*. *J. Bacteriol.* 170:367-3374.
- Zavitz, K.H. and K.J. Marians. 1991. Dissecting the functional role of PriA protein-catalysed primosome assembly in *Escherichia coli* DNA replication. *Mol. Biol.* 5:2869-2873.
- Zhang, S. and R.J. Meyer. 1995. Localized denaturation of *oriT* DNA within relaxosomes of broad host range plasmid R1162. *Mol. Micro.* 17:727-735.
- Ziegelin, G., J.P. Furste and E. Lanka. 1989. TraJ protein of plasmid RP4 binds to a 19-base pair invert sequence repetition with the transfer origin. *J. Biol. Chem.* 264:11989-11994.
- Ziegelin, G., W. Pansegrau, B. Strack, D. Balzer, M. Kroger, V. Kruft and E. Lanka. 1991. Nucleotide sequence and organization of genes flanking the transfer origin of promiscuous plasmid RP4. *DNA Seq. Map* 1:303-327.
- Ziegelin, G., W. Pansegrau, R. Lurz and E. Lanka. 1992. TraK protein of conjugative

plasmid RP4 forms a specialized nucleoprotein complex with the transfer origin. J. Biol. Chem. **267**:17279-17286.

END

1 6-0 6-9 7

FIN