MOLECULAR GENETICS OF BACTERIAL PLASMIDS AND TRANSPOSONS

COLLECTED REPRINTS 1973 - 1989

RONALD ANTHONY SKURRAY

A.U.A. (Pharm), B.Sc., Ph.D. (Adelaide)

Thesis submitted for the degree of Doctor of Science

in the Faculty of Science, The University of Adelaide,

1991.

CONTENTS

Preface 2

Declaration 3

Acknowledgements 4

List of Publications 5

Summary of Work 12

Statement on Authorship and Contribution 24

Published Works

SUMMARY OF WORK

The work presented in this thesis is summarized under three broad headings.

(i) Molecular genetics of the conjugative F plasmid

The studies undertaken during my Ph.D. candidature addressed, at a biochemical and genetic level, some physiological changes in *Escherichia coli* cells during conjugation. Previous reports in the literature had shown that *E.coli* recipient (F^-) cells were sensitive to an excess of donor (Hfr) cells. This phenomenon of "lethal zygosis" was characterized in detail as it was believed to represent an amplification of normal physiological changes in a recipient cell during conjugation (papers 1,2). Studies showed that conjugation was accompanied by the following perturbations in the recipient cell: (i) inhibition of DNA synthesis; (ii) inhibition of beta-galactosidase induction; (iii) altered transport and accumulation of galactosides; and (iv) leakage of beta-galactosidase into the supernatant fluid. These results suggested that there are alterations induced in the cytoplasmic membrane of a cell by cell-cell communication associated with conjugation (2).

In order to further characterize primary events accompanying conjugation, conjugation-defective (Con) mutants were isolated on the basis of a recipient cell's resistance to lethal zygosis. Examination of the cell envelope of Con^- mutants revealed an alteration in a major outer membrane protein, 3a (now designated OmpA) (4). It was suggested that OmpA serves either for the attachment of the F pilus or is involved in the transmission of conjugal DNA through the outer membrane. These studies and the ideas they fostered provided a basis for the development, in collaboration with Mark Achtman, of a model for the mating cycle (70).

Donor cells carrying the F plasmid seemed largely insensitive to killing by "lethal zygosis" and from studies with F^+ and Hfr donor strains it was inferred that there exists an immunity to lethal zygosis (Ilz^+) associated with the F factor (3). Donor strains with deletions or point mutations in their F factor were examined in

order to locate *ilz*$^{+}$ cistrons on the F factor genetic map. Results indicated *ilzA* is located near the transfer (*tra*) gene *traS* whereas a second locus, *ilzB*, was postulated to map to the right of *traI* (5). In hindsight it is now clear that *ilzA* is equivalent to *traT* and since *traT* is able to be expressed independently of *traJ* (40), the need for a second locus is in doubt. It is most likely that one of the prime reasons for the carriage of the surface exclusion genes, *traS* and *traT*, by the F plasmid is to reduce lethal zygosis among donor populations.

The F plasmid provides an ideal system for the genetic and molecular analysis of conjugal transfer of antibiotic resistance. Since F is the archetypal plasmid of the F-like group, which includes a number of antibiotic resistance (R factors) and virulence plasmids, its study is also of significance in the analysis of the evolution of plasmid-encoded functions among this group. My postdoctoral work at Berkeley and Edinburgh concentrated on a molecular genetic analysis of the F plasmid itself rather than the process of conjugation.

The major study undertaken at Berkeley was the molecular cloning of various DNA segments of the F sex factor in order to provide a series of overlapping fragments (6,8). The clones demonstrated one or more of the following F-specified functions: inhibition of female-specific phage (T7), formation of F pili, surface exclusion, immunity to lethal zygosis, incompatibility and self-replication (7,12,13,14). In collaboration with Mark Achtman's group in Berlin, these chimeric plasmids (pRS series) were employed to assign *Eco*RI fragment locations for all of the known *tra* genes (11) and also provided templates for the identification of the proteins specified by these *tra* cistrons (9,10). This work represented one of the first reported clonings and dissections of a series of complex operons, in this case, the transfer (*tra*) operons.

The *tra* region of the F plasmid is some 33 kb in length and encodes at least 23 genes involved in pilus biosynthesis and assembly, conjugal DNA metabolism and surface exclusion (72). The *tra* region contains one of the longest and most complex of prokaryotic operons, the *traYI* operon, which is under the positive control of the *traJ* gene product; *traJ* is in a separate operon. Analysis of the organization of the structural genes of these operons and their control sequences formed the basis of a

long term project which I began in Edinburgh, under the direction of Neil Willetts, and continued at Monash. The work undertaken in Edinburgh was largely concerned with obtaining a detailed understanding of the biochemical mechanisms involved in the control of F transfer by fertility inhibition. Unlike the F sex factor, the majority of F-like plasmids are unable to express their full potential for transfer due to a plasmid-determined system that inhibits the expression of the *tra* operons. Such Fin$^+$ (fertility inhibition$^+$) plasmids not only control the expression of their own transfer but also that of F. In addition to this FinOP system, five other systems of fertility control of F have been recognized (71). The studies reported in paper 15 were directed towards defining the mechanisms of control of *tra* gene expression by the FinQ (R62), FinU (JR66a) and FinV (R485) fertility inhibition systems.

At Monash, work concentrated on a molecular genetic analysis of the following aspects of F plasmid biology:

Leading region in DNA transfer. One of the major efforts was an analysis of the 13kb leading region of the F plasmid which is believed to be the first DNA to enter the recipient cell during conjugation (16). The significance of this region was examined by detailed molecular analysis using cloned, overlapping fragments. Analysis of ^{35}S-labelled products synthesized by maxicells carrying these chimeric plasmids identified polypeptides encoded within the region, one of which was subsequently identified by other workers as a single-strand DNA binding protein (18,25). *In vitro* deletions of cloned fragments, as well as transposon-insertion mutagenesis, aided in locating genes encoding these polypeptides. Promoters were identified by promoter-probe cloning, nucleotide sequencing and S1 transcript mapping (25,55,69). Studies also identified a plasmid maintenance function expressed by sequences within the F plasmid leading region (26,42,55). Our combined nucleotide sequencing, S1 transcript mapping, *lacZ* fusion analysis and mRNA stability studies demonstrated that the plasmid maintenance function, Flm, is dependent upon expression of the *flmA* polypeptide which is lethal for the host cell; expression of *flmA* is inhibited by antisense *flmB* mRNA and the differential stability of *flmA* (long-lived) and *flmB* (short-lived) mRNAs provides a mechanism which ensures the predominance

of plasmid-carrying cells in the population (for model see 55). We have also postulated that production of *flmA* is dependent upon expression of the product of ORF70, a gene mapping within the *flm* region which we designated *flmC; flmC* provides translation-dependent post-transcriptional control of *flmA*. Sequence analysis and functional studies have provided insights into the evolutionary significance of the leading region which is retained among F-like and non-F-like conjugative plasmids (69).

Genes involved in F-plasmid-mediated inhibition of T7 virus infection. T7 phage infection is aborted at an early stage if an F plasmid is present in the host. The genes (*pif*) required for this phage inhibition mediated by F were cloned (28). Potential *pif* gene products were examined in the *E.coli* maxicell system and results demonstrated that the *pif* region specifies at least two polypeptides and that the genes for these polypeptides lie within a single operon. The direction of transcription across the *pif* region was established by fusing potential *pif* promoters to sequences of the *E.coli* beta-galactosidase (*lacZ*) structural gene (29,32).

Control of conjugational transfer by fertility inhibition. The major *tra* operon of F is subject to indirect negative control by the FinOP fertility inhibition system which acts to inhibit the expression of the positive control gene, *traJ* (72). The *finO* gene from the F-like plasmid R6-5 was cloned and mapped more precisely by transposon mutagenesis (31). Using a *finO*- specific probe, the *finO* gene was mapped on F and shown to have suffered an IS*3* insertion, thereby providing an explanation for the derepressed state of the F plasmid transfer system (36,52). Subsequently, the nucleotide sequences of *finO* on F and R6-5 were completed, as were the adjacent sequences at the distal end of the *tra* region on both plasmids (manuscript in preparation). Studies at Edinburgh showed that the FinQ system specified by the I-like plasmids R62 and R820a inhibits *tra* operon transcription (15). At Monash the *finQ* gene was cloned and sequenced; the sequence predicted a predominantly hydrophilic protein with features of a DNA binding protein (62). The *finV* region of the plasmid R485 has also been cloned and sequenced (manuscript in preparation) but,

as with the FinQ system, the exact site and mechanism of FinV inhibition remains to be defined.

Organization and expression of the *tra* region. Analysis of the regulation of *tra* gene expression has often relied upon the use of small cloned fragments of the *tra* region of the F plasmid. In order to carry out studies with a more natural system, a lambda-derived cosmid, pRS2405, which is conjugative but can infect F^- cells from a virus head, was constructed. Such a system was used to provide information regarding the transcriptional and translational controls exerted on the conjugational transfer machinery (39,40). These studies also provided support for the notion that some genes within the major *traYI* operon could be expressed independently of *traJ* control and the *traYI* promoter (40). Using promoter and terminator fusions and S1 transcript mapping, a number of active promoters and a strong terminator (at the end of *traT*) were detected within the *traYI* operon (44,59,64). These findings have raised the possibility that the *traYI* operon is not uniquely transcribed as a single transcriptional unit (64). Also, the presence of these promoters and the terminator has provided some indication of the component parts that may have been assembled during the evolution of the *tra* region.

With the entire sequence of the 33 kb F plasmid *tra* region now completed in a collaborative programme with laboratories in Canada and the USA, my laboratory has begun fundamental biochemical and molecular studies to establish the exact roles of the *tra* gene products in pilus biosynthesis and conjugal DNA transfer. Initially, we have concentrated on the bifunctional 110 kilodalton (Kd) protein produced by *traG* which is involved in pilus synthesis and mating aggregate formation. Sequence and polypeptide analyses have identified the domain of the protein involved in pilus biosynthesis; immunolabelling studies indicate that *traG* is processed, possibly into two polypeptides each with a distinct function (manuscript submitted).

(ii) Molecular genetics of plasmids and transposons associated with antibiotic resistance in *Staphylococcus aureus* and *S. epidermidis*

Multiply antibiotic resistant strains of *Staphylococcus aureus*, often termed methicillin-resistant *S. aureus* (MRSA), became a world-wide problem as a cause of infections among hospitalized patients in the late 1970's. By 1980 these strains, which in some cases were resistant to twenty or more antimicrobial compounds, were epidemic in many hospitals in Australia, Europe and North America. Moreover, many strains of *S. epidermidis*, a common skin bacterium, were also multiresistant. It became apparent to me, after an examination of the literature, that little work had been undertaken on the molecular and genetic basis of resistance in these staphylococcal strains. I therefore initiated studies on this problem in my laboratory at Monash University in 1980. The work concentrated on three main themes.

Genetic basis of resistance. Our molecular and genetic studies showed that most MRSA from Australian hospitals possessed chromosomally-encoded determinants for resistance to erythromycin, fusidic acid, methicillin, minocyline, penicillin, rifampicin, streptomycin, sulphonamides, tetracycline, cadmium and mercury (17,19,20). Chloramphenicol resistance was exclusively encoded on small plasmids (19,41,56). Resistance to antiseptics and disinfectants was carried by members of the pSK1 family of multiresistance plasmids (20 - 42 kb in size) which additionally encode resistance to trimethoprim (Tp^r) via transposon Tn*4003*, penicillin (Pc^r) via Tn*4002* and linked resistance to the aminoglycosides gentamicin, tobramycin and kanamycin ($Gm^rTm^rKm^r$) via Tn*4001*. The structure and evolution of pSK1 plasmids has been extensively analysed in my laboratory, as have the transposons and insertion sequences carried by these plasmids (43,53,73). The most important findings relating to these structures are as follows:

(a) Gentamicin resistance: role of Tn*4001*. Linked $Gm^rTm^rKm^r$ is most frequently mediated by pSK1 family plasmids but has also been detected on the chromosome of some isolates (45,49). Our physical and genetic analyses established that $Gm^rTm^rKm^r$ is encoded by the 4.5 kb transposon Tn*4001* (22,45,66), the first composite transposon

from a Gram-positive organism to be identified and fully characterised. Molecular analysis of Tn*4001*, including cloning, mutagenesis and DNA sequencing (48) has clarified the nature of Gm^{r}Tm^{r}Km^r in *S. aureus*. Tn*4001* encodes the gene *aacA-aphD* which specifies a 57 Kd bifunctional protein with aminoglycoside acetyltransferase [AAC(6')] and aminoglycoside phosphotransferase [APH (2")] activities. Our molecular analyses allowed us to tentatively associate the N-terminal 170aa with AAC(6') activity and Tm^{r}Km^r, and the C-terminal region with APH(2") activity and Gm^r. Analysis of the primary aa structure and comparison with other resistance proteins indicated homology between the predicted AAC(6') region and the streptothricin acetyltransferase from *Streptomyces*; a 40aa sequence in the predicted APH(2") domain is highly conserved with other APH's and may represent active site sequences. Interestingly, secondary structure predictions suggested that the structure of the C-terminal portion of the protein is dependent on N-terminal sequences (48).

The central 1.9 kb region of Tn*4001* that encodes the AAC-APH protein and a small ORF for a 132 aa polypeptide is flanked by identical inverted repeats of the 1.3 kb insertion sequence IS*256* (47). Nucleotide sequence analysis of IS*256* has revealed structural features including 26 bp inverted terminal repeats and a single ORF for a 45.6 Kd putative transposase protein; sequence information identified potential promoters on IS*256* which are likely to be responsible for expression of Gm^{r}Tm^{r}Km^r (68). IS*256* sequences are present independent of Tn*4001* on plasmids and the chromosomes of MRSA and also occur in direct tandem duplication adjacent to the sequences for the AAC-APH protein (47). In the later case, we have proposed that promoter sequences on the duplicated IS*256* provide enhanced expression and result in increased levels of Gm^{r}Tm^{r}Km^r in the clinical isolates carrying such structures. Tn*4001* and Tn*4001*-like elements are also present on plasmids other than those of the pSK1 family. Large (ca 50 kb) conjugative plasmids, such as pSK41 and pUW3626 from North American isolates, carry the *aacA-aphD* determinant (45), however, the repeats bounding this determinant consist of a truncated IS*256* fused to a copy of IS*257* (53; Byrne, Gillespie & Skurray, Antimicrobial Agents and Chemotherapy (1990) *34*: 2106-2113). These and other studies from my laboratory gave support to our hypothesis that Tn*4001* and Tn*4001*-like elements have been

responsible for the worldwide dissemination of gentamicin resistance among Gram-positive bacteria.

(b) Penicillin resistance /ß-lactamase production; role of Tn*4002*. pSK1 family plasmids which encode Pc^r carry the 6.7 kb transposon Tn*4002* (57). We have determined the structure of Tn*4002* and shown it to be largely equivalent to the ß-lactamase (*bla*) region of ß-lactamase/heavy-metal resistance plasmids, such as pSK74, and to a previously described transposon, Tn*552* (57). The nucleotide sequences of the *blaZ* gene and the 121 bp inverted repeat sequence at the ends of Tn*4002* have been determined (67). The results supported our hypothesis that Tn*4002* and Tn*552* are closely related or even identical ß-lactamase transposons. DNA-DNA hybridization demonstrated that sequences homologous with Tn*4002* are also present on the chromosome of most MRSA isolates. We have proposed that Tn*4002* transposed from the chromosome to a pSK1-family plasmid on a number of occasions to give rise to the various insertions we have detected (53,57).

(c) Trimethoprim resistance; role of Tn*4003*. The mechanism of Tp^r encoded by pSK1 is the synthesis of a dihydrofolate reductase (DHFR) with reduced affinity for Tp. This DHFR was designated type S1 and the Tp^r gene, *dfrA*, was cloned from pSK1 and localized by mutagenesis (37,51); this region is flanked by three directly repeated copies of IS*257* which together make up the 4.7 kb Tp^r transposon Tn*4003*. Our nucleotide sequence and biochemical analyses of Tn*4003* showed that the central region encodes not only *dfrA* but also a thymidylate synthase (*thyE*); these genes were predicted to form a coordinately controlled operon, *thyE-dfrA*, similar to that on the chromosomes of *Bacillus* and *Lactobacillus* (61). Amino acid homology relationships allowed us to make deductions regarding the active sites of these enzymes and the likely origin of the *thyE-dfrA* region on Tn*4003* (61).

The insertion sequence IS*257* (790bp) encodes a putative transposase of 224 aa with two helix-turn-helix DNA recognition motifs (61). IS*257* shares sequence similarities with members of the IS*15* family of sequences from Gram-negative

organisms and with ISS*1* from *Streptococcus lactis*. We have determined the evolutionary distances between members of this superfamily of insertion sequences and explored the effect of G+C content of these elements and their hosts on transposase polypeptide evolution (63).

(d) Antiseptic and disinfectant resistance. The *qacA* (for quaternary ammonium compound) gene is located on pSK1 family plasmids and mediates resistance to acriflavine (Ac), ethidium bromide (Eb) and to quaternary ammoniums (Qa) such as cetrimide (38). *qacA* has been cloned and expressed in *E. coli* (33). Tn*5* mutagenesis of *qacA* combined with analysis of proteins expressed in minicells and maxicells identified the *qacA* product as a 50 Kd protein. Fluorimetric and isotopic assays demonstrated that *qacA* specifies an energy-dependent export system coupled to H^+ uptake, a mechanism which presumably limits the concentration of antiseptic within the bacterial cell (60). Nucleotide sequence analysis has given some exciting insights into the nature of the *qacA* polypeptide; viz., that this cytoplasmic membrane protein with 14 predicted transmembrane segments shares significant homology with other proteins responsible for export-mediated resistance to antibiotics and the uptake of sugars (Rouch, Cram, DiBerardino, Littlejohn & Skurray, Molecular Microbiology (1990) *4*: 2051-2062).

The staphylococcal gene pool and transfer of resistance. Analysis of plasmids and resistance determinants in strains of coagulase-negative staphylococci, such as *S.epidermidis,* was undertaken to explore the hypothesis that such organisms could contribute to the gene pool of the more invasive *S. aureus* (27,41). Detailed molecular analysis, using cloned fragments as specific resistance probes, demonstrated the presence of identical resistance genes in both species (54; unpublished). It is also clear that other bacterial species can contribute to or exchange genes with staphylococci; e.g., Tn*4001*-like structures are present on plasmids from *Enterococcus faecalis* (48). Studies also demonstrated that the majority of MRSA isolates in Australia do not carry conjugative plasmids, as have been described in North

American isolates, but that the resistance determinants may be transferred by a procedure designated mixed-culture transfer (22).

Origin and evolution of resistance determinants in MRSA. Molecular techniques were also employed to provide an epidemiological and evolutionary analysis of resistance determinants in staphylococci. There have been periodic episodes with resistant *S. aureus*, both in hospitals and the community, over the past forty years and our retrospective studies have demonstrated that contemporary MRSA do possess some of the determinants present in strains from earlier episodes (21,34). Significantly, some of these determinants, e.g., those for penicillin, mercury and tetracycline resistance, are chromosomally encoded in current MRSA yet plasmid borne in earlier strains suggesting translocation events have been involved in the evolution of contemporary MRSA. Our studies confirmed the importance of modular evolution in MRSA; transposon and plasmid integration were documented thoroughly (35,53) giving direct evidence that transposition and genetic rearrangement of pre-evolved determinants have been the predominant forces in the evolution of both plasmids and the multiresistant chromosome in MRSA. These mechanisms also appear to have played a prominent role in the development of multiresistant coagulase-negative staphylococci, now recognized as important pathogens in their own right.

Of some significance was our recognition that the IS*257* family of insertion sequences appears to have played a major role in the acquisition and spread of a number of antimicrobial resistance determinants; e.g., Tp^r (Tn*4003*), Hg^r (Tn*4004*) and the methicillin-mercury-tetracycline resistance cluster on the chromosome (50,61,63, unpublished). Moreover, large conjugative plasmids, such as pSK41, each carry at least 6 copies of IS*257*. We have mapped the precise locations of IS*257* on these plasmids and established what determinants these sequences flank (Byrne, Gillespie & Skurray, Antimicrobial Agents and Chemotherapy (1990) *34*: 2106-2113). Finally, significant advances in our understanding of the evolutionary origins of staphylococcal resistance determinants have been obtained through DNA and protein homology studies and from analysis of the G+C content of staphylococcal resistance determinants (48,61).

(iii) Molecular genetics of degradative plasmids in *Pseudomonas* and related genera

The biochemistry and genetics of the degradation of aromatic compounds by bacteria is an important field of study because of the environmental impact of such compounds as pollutants and the likely use of these compounds as precursors for commercial synthetic processes. Analysis of the biochemical pathways for the degradation of cresols and xylenols and the genes involved had been undertaken in the laboratory of Dr. R.C. Bayly for some years prior to my joining Monash University. It was suspected that some of the genes involved were plasmid encoded and a collaboration was developed to exploit my expertise in the molecular and genetic analysis of plasmids.

The degradation of *p*-cresol in *Acaligenes eutrophus* strain 345 was shown to be chromosomally encoded. However, this strain contains a plasmid, pRA1000, that encodes for the degradation of *m*- and *p*-toluate. This 85 kb Tol-like plasmid was characterized by restriction analysis and found to be closely related to the archetypal Tol (toluene degrading) plasmid pWWO. It differs from pWWO in that pRA1000, when present in *A. eutrophus* 345, regulates the expression of the chromosomally-encoded *p*-cresol *meta* degradative pathway (23). Biochemical studies with wild type *A. eutrophus* 345 and a series of derivatives demonstrated that a number of steps in the catechol *meta*-cleavage pathways employed in the degradation of phenol, *p*-cresol and *m*- and *p*- toluate are mediated by isofunctional enzymes (24). In the case of the two 2-hydroxy-6-oxohepta-2,4-dienoate hydrolases, one is encoded by pRA1000 and the other is chromosomally encoded. Both enzymes were purified and characterized and shown to be immunologically distinct (46).

The physical and biochemical analysis of a strain of *Pseudomonas putida* showed that the ability to degrade 3,5-xylenol is encoded on a 500 kb plasmid, pRA500. Restriction analysis of deleted derivates of this plasmid limited the genes for 3,5-xylenol degradation within a 130 kb region; the majority of genetic information for this pathway is encoded on a 50-70 kb segment (30). Studies with cloned

fragments of this region have provided a more detailed understanding of the organization and control of this degradative pathway (unpublished). A preliminary analysis of a strain of *P. putida* able to utilize *p*-cresol and 2,4-xylenol demonstrated that the plasmid carried, pRA4000, encodes genes for the degradation of *p*-cresol and resistance to inorganic mercury. The genes for 2,4-xylenol degradation are chromosomally encoded in this strain but their expression appears to require a positive regulatory gene encoded on pRA4000 (65). Much of the work from these studies on degradative pathways has been summarized in a book chapter (58).