GENETIC AND TRANSCRIPTIONAL ORGANIZATION OF A REGION ESSENTIAL FOR DNA TRANSPORT IN Tn*5252*

By

FRANCISCO J. ALARCON-CHAIDEZ

Bachelor of Science
Instituto Tecnologico de Culiacan
Sinaloa, Mexico
1982

Master of Science
Texas A&M University
College Station, Texas
1984

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of the
requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 1997

GENETIC AND TRANSCRIPTIONAL ORGANIZATION

OF A REGION ESSENTIAL FOR DNA

TRANSPORT IN Tn*5252*

Thesis Approved:

M. N. Vijayakumar

Thesis Advisor

David A. Wegas

Jeffrey Hadwiger

Robert L. Burnap

Richard C. Essenberg

Thomas C. Collins

Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express sincere gratitude to my major advisor, Dr. Moses Vijayakumar for his encouragement and advice through my graduate program. My sincere appreciation extends to my other committee members, Dr. David Demezas, Dr. Jeffrey Hadwiger, Dr. Richard Essenberg and Dr. Robert Burnap whose suggestions and support were very helpful throughout the study.

Special thanks are due to my colleagues and friends, Salhu Ayalew, Ali Kiliç, Ursula Muñoz-Najar, Janardhan Sampath, and Pravina Srinivas for their suggestions, constructive comments, and friendship.

I would also like to give my special appreciation to my wife Judith and my son Eric for their love and understanding throughout this whole process.

My deepest appreciation is extended to my mother who always believed in my abilities and constantly provided me with support, moral encouragement, and understanding.

Finally, I would like to thank Dr. Moses Vijayakumar and the Department of Microbiology and Molecular Genetics for providing me with this research opportunity and their generous financial support.

TABLE OF CONTENTS

Chapter Page

1. INTRODUCTION 1

2. LITERATURE REVIEW 5

DNA Translocation 5
Bacterial Transformation 6
Bacterial Conjugation 8
Conjugal Transfer of Plasmids in Gram-negative Organisms 8
Conjugative Plasmid Transfer in Gram-positive Organisms 10
Conjugative Transposition 12
Mechanism of Transposition 13

3. MATERIALS AND METHODS 24

Bacterial Strains and Plasmids 24
Growth Conditions 25
Chemicals, Restriction Endonucleases, and Media 30
Molecular Cloning Techniques 30
Preparation of *S. pneumoniae* competent cells 31
Transformation of *E. coli* 32
Transformation of *S. pneumoniae* 32
Plasmid Isolation from *E. coli* 32
Chromosomal DNA Isolation from *S. pneumoniae* 34
DNA Sequence Analysis 35
Generation of Nested Deletions 35
Preparation of Double-Stranded DNA Template 37
Denaturation of Template 38

Chapter Page

Annealing of Primers to Template 38
Labeling of DNA 38
Termination 39
Denaturing Polyacrylamide Sequencing Gel 39
Gel Electrophoresis 40
Analysis of DNA sequences 41
Protein Expression 41
Cell Fractionation 42
Periplasmic Fraction 42
Membrane and Cytoplasmic Fractions 43
SDS-PAGE Analysis 44
Conjugation Analysis 45
Southern Hybridization 46
Transfer of DNA 46
Nick Translation 47
Prehybridization and Hybridization 48

4. GENETIC AND TRANSCRIPTIONAL ORGANIZATION OF A REGION ESSENTIAL FOR DNA TRANSPORT IN Tn5252 49

RESULTS AND DISCUSSION 49

Identification and Nucleotide Sequence Analysis 49
Sequence Analysis 50
Nucleotide Sequence Organization 51
Analysis of the ORFs Predicted Amino Acid Sequences 68
ORF20 Predicted Nucleotide Sequence 68
ORF21 Predicted Product 70
Deduced Products From ORF22 to ORF25b 80
ORF26 Predicted Product 85
Analysis of the ORF27 Predicted Product 90
Homologies of the ORF28 Predicted Product 91
Expression and Identification of the ORF26 and ORF27 Products 104

Chapter Page

5. CONJUGATIVE TRANSFER OF TRANSFER-DEFICIENT Tn*5252* BY COMPLEMENTATION 110

RESULTS AND DISCUSSION 110

Intraspecific Mobilization of Transfer-Deficient Tn*5252* 110
Transfer Properties of Transconjugants 115
Interspecific Mobilization of Transfer-Deficient Tn*5252* 116
Analysis of Transconjugants by Southern Hybridization and Transformation 120

6. CONCLUSIONS 126

DNA Sequence Analysis and Protein Expression 126
Complementation Analysis 127

BIBLIOGRAPHY 130

LIST OF TABLES

Table Page

1. *E. coli* Strains and Cloning Vectors 26
2. Streptococcal Strains 28
3. Antibiotics and Concentrations 29
4. Summary of the DNA Analysis of the 9.5-kb *BamHI* DNA Fragment 54
5. Characteristics of the Predicted Products in the 9.5-Kb DNA Fragment 74
6. Intraspecific Transfer Frequencies of Tn*5251* and Tn*5252* 117
7. Interspecific Transfer Frequencies of Tn*5251* and Tn*5252* 119
8. Transformation Properties of Selected Transconjugants 125

LIST OF FIGURES

Figure Page

1. Composite structure of Tn*5253* 19
2. Predicted rearrangements of Tn*5253* upon insertion of heterologous DNA within the *Sau3AI* and *Hinc*IIII sites localized downstream from Tn*5251* 21
3. Identification of regions involved in transfer in Tn5252 by using insertion and deletion mutagenesis 23
4. Schematic representation of the restriction map of the 9.5-kb DNA fragment from Tn*5252* and the predicted ORFs 56
5. Complete nucleotide sequence of the 9.5-kb *BamHI* fragment from Tn*5252* 58
6. Hydropathy profile and secondary structure prediction of the deduced amino acid sequence from ORF20 69
7. Multiple alignment of the predicted product from ORF21 and its homologs 77
8. Hydropathy profile and secondary structure prediction of the deduced amino acid sequences from ORF22 and ORF23 82
9. Hydropathy profile and secondary structure prediction of the deduced amino acid sequences from ORF24 and ORF25a 83

Figure | Page

10. Hydropathy profile and secondary structure prediction of the deduced amino acid sequence from ORF25b 84

11. Multiple sequence alignment of the predicted amino acid sequence from ORF26 and its homologs 88

12. Multiple sequence alignment of the predicted amino acid sequence from ORF27 and its homologs 93

13. Hydropathy profile and secondary structure prediction of the deduced amino acid sequences from ORF27 and ORF28 95

14. Multiple sequence alignment of the predicted amino acid sequence from the amino-terminal domain from ORF28 and its bacteriophage homologs 98

15. Multiple sequence alignment between the carboxyl-terminal domain of ORF28 and its homologs 102

16. Protein profile from the 4.2-kb *BamHI* fragment from Tn*5252* 107

17. Protein profile from the 3.2-kb *BamHI* fragment from Tn*5252* 108

18. Protein profile from the 3.2-kb *BamHI* fragment from Tn*5252* using 17% Polyacrylamide slab gels 109

19. Strategy For Inter- and Intraspecific Transfer of Tn*5251* and Tn*5252* 113

20. Identification of Tn*5251* and Tn*5252* by DNA hybridization 122

CHAPTER 1

Introduction

In recent years, the appearance of multiple antibiotic resistance in *Streptococcus pneumoniae* has been of particular concern with some strains showing resistance to more than eight different antibiotics (9,10). Previous reports have demonstrated that resistance traits in *S. pneumoniae* were capable of transfer by a deoxyribonuclease resistant process, not only into *S. pneumoniae* but also into other groups of streptococci (10). Endogenous extrachromosomal elements are rarely observed in this species, and the transfer of resistance determinants has been associated with a novel class of genetic elements called conjugative transposons (10,18). Conjugative transposons differ from classical transposons in that no duplication of the target sequences at which they insert takes place and insertion of the element into the host genome confers on it the ability to act as a conjugational donor. Additionally, donor and recipient strains need not belong to the same species or even the same genus (60). Due to their involvement in the spread of antibiotic resistance among clinical isolates of Gram-positive bacteria, conjugative transposons are of

considerable medical importance. Some of the best studied conjugative transposons include Tn*916* (18), Tn*1545* (11), Tn*3701* (5), and Tn*5253* (78). Genetic characterization of these elements has revealed that all of them carry *tet*(M), a gene encoding resistance to tetracycline that is expressed in both Gram-positive and Gram-negative bacteria. The presence of the *tet* determinant and the significant homology between regions of DNA surrounding this gene in most of the conjugative transposons, suggested that a smaller conjugative element such as *tet* carrying Tn*916* could have served as a progenitor in the evolution of the larger elements (10,60). In addition, the transposons may contain other antibiotic resistance determinants such as kanamycin, erythromycin, and chloramphenicol.

Tn*916*, the smallest of the conjugative transposons, is an 18-kb element that was first detected in *Streptococcus faecalis* DS16 (18). This transposon was considered to be the prototype of a growing family of conjugative transposons that were commonly found in streptococci and enterococci (10,61). Tn*916* has been shown to insert at different sites in the bacterial chromosome of a large number of streptococci but also appeared to have preference for a unique target site in some strains (10). Transposition of these elements occurs via an excision-insertion mechanism that resembles that of lambdoid phages (62).

Studies of the larger conjugative transposable elements such as Tn*3701* (5) and Tn*5253* (2), formerly called Ω *cat-tet*, revealed that they

were composite elements containing a Tn*916*-like element within a larger transposable element. Interestingly, the composite element can transpose as a large unit or the smaller element can transpose independently.

Tn*3701* is a 50-kb composite transposon carried by *Streptococcus agalactiae* A454, that encodes resistance to erythromycin, tetracycline and minocycline (37,38). This element has been shown to preferentially insert at a unique site in the chromosome of *E. faecalis*. The antibiotic resistance determinants are located in a internal 19.7-kb DNA element designated Tn*3703* that resembles the structure of Tn*916* (37). This smaller element has the ability to transpose to different sites in the chromosome of *E. faecalis* (38). The remaining sequences in Tn*3701* outside Tn*3703* showed significant similarity to other composite structures such as Tn*5253* (78) and Tn*3951* (30), and it has been suggested that Tn*3701*-like structure is the prototype of composite elements.

Tn*5253*, a 65-kb DNA segment, originally identified in the chromosome of *Streptococcus pneumoniae* BM6001 (78), is a self-transmissible element that encodes resistance to chloramphenicol and tetracycline. This element has been shown to insert at a specific target site in the pneumococcal genome (78). Physical analysis of Tn*5253* resulted in a detailed restriction map of this element that made possible the localization of drug resistance determinants and the identification of its junction and target regions in the pneumococcal chromosome (79). Tn*5253* also has been shown to be a composite structure of two

conjugative transposons, Tn*5251* and Tn*5252.* Tn*5251* is a 18-kb segment of DNA carrying the *tet* determinant which is capable of independent conjugative transposition when removed from the larger element (2). Deletion of this fragment from within Tn*5253* does not affect mobility of the remaining sequences (Tn*5252*). This element was able to insert randomly at many sites and it showed structural and functional similarity to Tn*916* and Tn*1545*.

The process of mobilization of Tn*5252* is still not completely understood. The objectives of this study were to perform genetic and functional analyses of an 8.6-kb *BamH*I DNA region from Tn*5252* that appears to encode genes potentially involved in the process, and to determine whether this region plays a role in regulation of the transfer of the element.

CHAPTER 2

LITERATURE REVIEW

DNA Translocation

The transfer of molecules across bacterial cell membranes has been studied for some time. Information gathered from these studies has shown that transport of many small molecular weight compounds into a cell is usually accomplished by the driving force of an ion-coupled transport system (54). On the other hand, the cell membrane transport of larger molecules, including DNA and RNA, has proved to be more complex and detailed information regarding the process is still unavailable. Two of the best studied mechanisms of DNA transfer through bacterial membranes include transformation and conjugation. In recent years, interest in the study of these mechanisms has increased due to the important contribution of DNA transfer to bacterial genetic diversity and the potential risks of releasing genetically engineered microorganisms into the environment.

Bacterial Transformation

During bacterial transformation, DNA released into the environment is taken up by induced or naturally transformable (competent) bacterial cells via a DNase sensitive process. Some of the most well studied systems of natural genetic transformation include the Gram-positive *S. pneumoniae*, *Bacillus subtilis* and *Streptococcus sanguis*, and the Gram-negative *Haemophilus influenzae* and *Neisseria gonorrhoeae.* Competence in *S. pneumoniae* and *S. sanguis* develops during early to late log phase of growth in response to a secreted protease-sensitive signal that has been identified as a 17-residue peptide (27). Export of this factor into the surrounding environment appears to be mediated by an ATP-dependent transport protein encoded by the gene *comA* in *S. pneumoniae* (27). On the other hand, competence of *B. subtilis* arises during early stationary phase and, in contrast with *S. pneumoniae*, only a small fraction of a given population becomes competent. Structural changes in the cells leading to the ultimate transfer of DNA has been suggested to involve the formation of a membrane-bound receptor complex consisting of an endonuclease and a putative DNA-binding protein. In *S. pneumoniae*, the membrane-bound endonuclease, EndA, degrades one of the strands of the double-stranded DNA resulting in the transport of the other DNA strand into the cell in a 3' -to- 5' polarity in a Ca^{2+} dependent process (41). Translocation of the single-stranded DNA into the cytoplasm involves formation of a

complex with another protein that appears to provide protection against endonuclease attack or to facilitate recombination. A similar mechanism has also been proposed for *B. subtilis* with the only difference being that transport of the single strand of DNA appears not to be polar since both 5' and 3' ends can be transferred (13).

The mechanism of transformation in Gram-negative resembles that of Gram-positive. Competence in *H. influenzae* appears during the early stationary phase and occurs in practically all cells in a population. However, no competence factor has been identified. The DNA binding/uptake system in Gram-negative bacteria consists of membrane vesicles composed mainly of phospholipids and polypeptides that are released during the state of competence. These vesicles termed "transformasomes" recognize specific sequences in the incoming DNA that are only found in similar or closely related species, and transport them to the recipient cell cytoplasm. Integration of donor DNA into the chromosome of the recipient appears to be similar in both Gram-positive and Gram-negative transformation systems (41). A heteroduplex is formed between homologous regions, which results in assimilation of the donor strand and displacement of the recipient strand. This process seems to be catalyzed by a RecA-like mechanism.

Bacterial Conjugation

Bacterial conjugation consists of the transfer of DNA from donor to recipient cells by a process that requires cell to cell contact (36). This process has been extensively studied in *Escherichia coli* K-12 strains containing the F plasmid, a 100-kb conjugative plasmid that encodes about 40 transfer (*tra*) genes that actively participate in the conjugative process (19).

Conjugal Transfer of Plasmids in Gram-negative Organisms

In general, bacterial conjugation in F plasmids begins by nicking at a specific origin of transfer (*oriT*) on the DNA strand that is to be transferred. Cleavage takes place in the presence of the TraI and TraY gene products and Mg^{2+} ions (36). TraI appears to function as the DNA relaxase that initiates transfer of F and is thought to interact with the cell membrane through the TraD membrane protein (19). DNA unwinding by the helicase function of TraI results in displacement of the nicked strand into the recipient cell. In IncQ plasmids, mobilization of DNA is controlled by the MobA, MobB and MobC gene products that have shown in vitro strand-specific cleavage of the *oriT* region in plasmid RSF1010 (25). In vitro relaxation in IncP conjugative plasmids is carried out by the nicking activity of the TraI and TraJ gene products and it is enhanced by the

presence of the TraK protein. T-DNA transfer from the plant pathogen *Agrobacterium tumefaciens* into some plant species also requires nicking of specific sequences in Ti plasmids before transfer can take place. These specific sequences are known as border sequences and consist of 25-bp imperfect direct repeats that are recognized and nicked by the VirD2 protein, a homolog of the TraI endonuclease of broad-host-range plasmid RP4 (49). In recent studies, the transfer regions of most Ti plasmids have been shown to encode transfer genes that are related to those in F and IncP plasmids (1,14). The TraA protein from nopaline-type Ti plasmid pTiC58 was shown to share segments of homology with the relaxase MobA from the mobilizable plasmid RSF1010. Similarly, the Ti TraF and TraG proteins were homologous to TraF and TraG of IncP plasmid RP4 which form part of the mating bridge that is associated with the cell membrane (14). It has been suggested that arrangement of the transfer genes in Ti plasmids may have evolved by acquisition of genes from different conjugal transfer systems (1).

The mechanism of transport of the DNA strand into the recipient cell is still under study. The ability of RP4 TraI to bind to the 5'-end of the mobilized strand suggested that this protein might act as the "pilot" protein that leads it through the membrane (48). Some genes from F and I plasmids have been shown to encode single-stranded DNA binding (SSB) proteins that bind DNA without sequence specificity. However, the fact that these proteins are not essential for conjugation, suggests that SSB

proteins function to protect single-stranded DNA after it has been transferred to the recipient, rather than to promote transport. Furthermore, F-like plasmids transfer DNA without any detectable association with plasmid-encoded proteins (53). Transfer of T-DNA in Ti plasmids appears to be promoted by the VirE2-encoded protein, which binds DNA nonspecifically to protect it during transfer through bacterial membranes (36).

Conjugative Plasmid Transfer in Gram-positive Organisms

Most of the information concerning the mechanisms of bacterial conjugation has come from studies on Gram-negative systems. However, understanding of the genetics of gene exchange in Gram-positive is of major importance specially because conjugal transfer of DNA has proved to be critical in the spread of antimicrobial resistance among Gram-positive pathogens. As in Gram-negative systems, conjugal transfer in Gram-positive requires cell-to-cell contact which may or may not be induced by pheromones (43). Plasmids that have no need for pheromones are usually broad host range plasmids that show a wide range of transfer to a variety of other Gram-positive bacteria (43). A number of conjugative plasmids have been identified in clinical isolates of staphylococci and streptococci. Streptococcal plasmids range from 26 to 33 kb and are known to encode resistance to macrolide, lincosamide, and streptogramin

B (MLS) group of antibiotics and to chloramphenicol (*cat*) (10,43). Among this type of plasmids, *Streptococcus agalactiae* pIP501 (81), and *Enterococcus faecalis* pAMβ1 (9) have been studied as the conjugal transfer models. Plasmid pIP501 is 30.2 kb in size and has been shown to transfer to a wide variety of streptococcal species (43). The transfer functions of this plasmid reside in an operon that consists of six ORFs preceded by a putative *oriT* locus similar to that of Gram-negative plasmids. All ORFs were essential for conjugal transfer and the *oriT*-like sequence underwent nicking when used for in vivo experiments in *E. coli* (81,82). These results suggested similarity in the mechanisms of transfer between Gram-positive and Gram-negative conjugative plasmids.

Staphylococcal conjugative plasmids have been shown to transfer resistance to erythromycin, penicillin, neomycin, and gentamycin among clinical isolates of *Staphylococcus aureus* (43). Examples of staphylococcal conjugative plasmids include pGO1 (46,72) and pSK41 (16). Genetic analysis of the conjugative transfer regions of pGO1 (*trs*) and pSK41 (*tra*) has revealed remarkable similarity between both plasmids at the amino acid sequence level. The size of these regions ranges from 13 kb in pGO1 to 14.4 kb in pSK41 and sequence analysis revealed the presence of 14 and 15 putative open reading frames, respectively, that were mostly transcribed in the same direction. This arrangement was similar to those of plasmids of Gram-negative species such as *E. coli* and *A. tumefaciens*

(16,46). However, the functions of each of the predicted products remain to be clarified.

Conjugative Transposition

Conjugative transposons transfer from one bacterial cell to another by a process requiring cell to cell contact (65). These elements show a broad host range since they are able to conjugatively transpose to Gram-negative bacteria and can conjugatively transfer among many species of Gram-positive bacteria with frequencies ranging from 10^{-4} to 10^{-9} (9,62). Similar to conjugative plasmids, they are important in the spread of antibiotic resistance among pathogenic bacteria that belong to different species and even different genera. Representatives of conjugative transposons detected in Gram-positive bacteria include Tn*916*, isolated from *E. faecalis* (18), and Tn*5253*, formerly called the Ω *cat/tet* element, that was found in *S. pneumoniae* BM6001(2,65,77,78). A family of transfer-deficient but mobilizable insertion elements called NBU (Nonreplicating *Bacteroides* units) have been found in human colonic *Bacteroides* species (55,66,67). These elements are unable to promote their own excision, however, they can be transferred by the excision and mobilization functions provided by *Bacteroides* conjugative transposons. The circular intermediate formed by the NBUs after excision contains an internal origin of transfer (*oriT*) that allows it to be mobilized (55). It has been suggested that transfer

proceeds by transport of a single-stranded copy of the NBU thorough the mating pore formed by a conjugative transposon (40).

Mechanism of Transposition

The 18-kb element Tn*916*, one of the smallest transposons reported, and the closely related Tn*1545* (25 kb), have been the best-studied conjugative transposons in Gram-positive bacteria (51,55,59,62). These elements usually carry a *tetM* determinant, a gene that encodes resistance to tetracycline. In addition, Tn*1545* also carries the *apha-3* and *ermAM* genes for kanamycin and erythromycin resistance respectively (62).

A transfer mechanism suggested for Tn*916* involves a covalently closed circular intermediate produced by excision of the transposon from the donor molecule, which is then transferred for reinsertion into the recipient (61). However, failure to locate an internal origin of transfer (*oriT*) in the element has raised questions about this process (55). The model proposed for excision involves staggered nicks made on each strand of the DNA about five bases 3' of the transposon ends that generate a 5' overhang. The bases excised with the transposon called "coupling sequences", are ligated together to produce an excisant and the mismatch between these bases is resolved by replication (59,60). Based on this information, conjugative transposition has been suggested to be similar to that of excision and integration of phage lambda. However, unlike lambda

integration, branch migration would not apply for Tn*916* because the coupling sequences are not homologous. Insertion of the transposon in the recipient chromosome is a reciprocal of excision. Staggered nicks are made in the target DNA and the ends are ligated to the ends of the nicked transposon. The mismatches generated at insertion are resolved during replication resulting in the introduction of coupling sequences derived from a previous target (60). This has been demonstrated by analysis of the sequence adjacent to the transposon Tn*916* after being used to transform *Bacillus subtilis* protoplasts (59). In each case, the sequence on one side of the element corresponded to that of the coupling sequences found in the circular transposition molecule. Recent studies on Tn*916* and Tn*1545* have revealed that the functions needed for conjugative transposition reside in the element itself (55). The *int* and *xis* genes of these transposons share similarity in their sequence. The *int* gene is required for excision and for integration as well, while the *xis* gene is not essential for excision but appears to stimulate the frequency of this process (51).

Larger conjugative transposons such as Tn*3701* (38), Tn*3951* (30), and Tn*5253* (2) are composite elements in which a Tn*916* -like element appears to be inserted into the larger element. Tn*3701* is a 50-kb conjugative element carried by *Streptococcus pyogenes* A454 that contains the Tn*916*-like element Tn*3703* (37). Comparison of the sequences from Tn*3701*, Tn*5253*, and Tn*3951* revealed significant similarities not only in

their DNA-DNA hybridization profiles but also in their restriction maps in regions outside the internal transposon (38). Additionally, Tn*3701* also shows a preferred target site in the chromosome of *E. faecalis*. However, the sequence of the junctions between the transposon and the host chromosome remain to be identified (37). The internal 19.7-kb element in Tn*3701* carries resistance to erythromycin (*erm*), tetracycline, and minocyclin [*tet(M)*]. This element, designated Tn*3703*, was capable of independent transposition from hemolysin plasmids to many sites in the host chromosome but it could not transfer by conjugation from one *E. faecalis* chromosome to another (38).

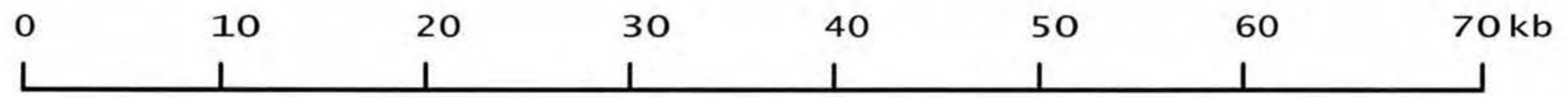
Tn*5253*, a 65.5-kb conjugative transposon originally identified in *Streptococcus pneumoniae* BM001 (78), is also a composite element that carries a smaller Tn*916*-like element (Fig. 1) (2). Tn*5252* is a 47-kb conjugative element that was generated upon removal of Tn*5251* from Tn*5253* (2). Transposition of Tn*5252*, similarly to parental Tn*5253*, is site-specific and involves a 72-bp target region in the chromosome of *S. pneumoniae* that appears to serve as a signal for its integration and excision. It has been suggested that the crossover point of integration is nonrandom with respect to this region and that the mechanism of insertion may resemble that of site-specific temperate phages (79).

Of particular interest in Tn*5253* has been the nucleotide sequence spanning the tetracycline determinant *tet*(M) which exhibits a high degree of homology to other *tet*(M) determinants. Initial studies in this specific

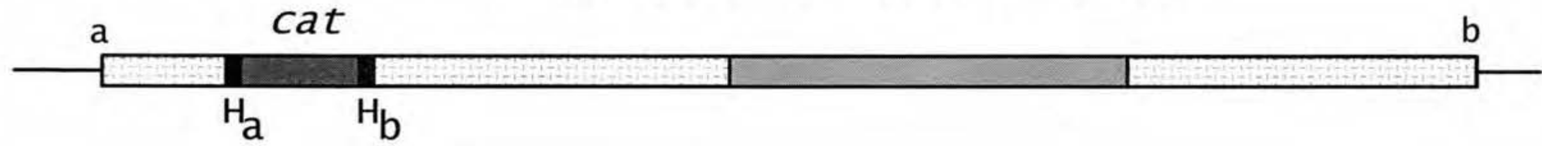
region of Tn*5253* suggest the presence of potential regulatory sequences that may be involved in transposition itself (77). By use of directed insertion of a non-replicating plasmid vector, Vijayakumar et al. (77) were able to locate a mutation in Tn*5253* which abolished all drug resistances and transfer functions. Plasmid vector pVA891, a derivative of pACYC184 carrying a streptococcal *erm* determinant (42), was used to generate a series of clones carrying fragments from different sites in Tn*5253*. These vector molecules were expected to insert into the host chromosome only when directed by the ligated passenger fragments in an homology-dependent recombination reaction (52). After transformation of pneumococcal strain DP132 competent cells carrying the element, six transformants resistant to all three drugs, erythromycin, tetracycline, and chloramphenicol, were isolated, demonstrating that the vector had inserted into the transposon. All strains were able to transfer the Ω element to wild type cells by conjugation on filters except for an isolate designated GP45 that was conjugation deficient (Tra^-). Although this strain was sensitive to chloramphenicol and tetracycline, its DNA was able to transform DP1322 cells giving rise to Em^r transformants that were Cm^s Tc^s Tra^- (40%), Cm^s Tc^s Tra^- (40%) or Cm^r Tc^r Tra^+ (20%). Vijayakumar et al. (77) suggested that insertion of the plasmid vector in a regulatory sequence resulted in a variety of genomic rearrangements that produced the results observed. The insertion of the plasmid vector in strain GP45 was located at a *Sau3AI* site upstream from the *tet* determinant (Fig. 2). Recent studies

on Tn*5252* using insertion and deletion mutagenesis have resulted in the location of a number of regions in this element that have been shown to be involved in the process of conjugal transfer of Tn*5252* (Fig. 3). Some of these regions have been mapped in an 8.6-kb *BamH*I fragment located at coordinates 33.0 to 41.6 in the restriction map of Tn*5252* (Fig. 3). Insertion of heterologous DNA into an internal *Xba*I site in this region has been shown to impair the transfer abilities of Tn*5252* (34). This site is located approximately 200 bp upstream from a 0.3-kb *Hind*III-*Sau3A*I region that was previously suggested to play an essential role in regulation of transposition (Fig. 2) (77). Conjugal transfer of T*n5252* was also abolished when an internal 1.9-kb *Bgl*II fragment was removed from the 8.6-kb *BamH*I DNA segment. This deletion was mapped at coordinates 37.9 to 39.8, approximately 3.2 kb downstream from the *Xba*I site (Fig. 3).

Figure 1. Composite structure of Tn*5253*. Ha and Hb represent insertion sequences that are flanking the *cat* (chloramphenicol acetyl transferase) gene. Removal of Tn*5251* from the larger transposon results in the formation of Tn*5252*. Occasionally, the *cat* gene cassette will also be lost giving rise to Tn*5252*Δ*cat* that is still transfer-proficient.

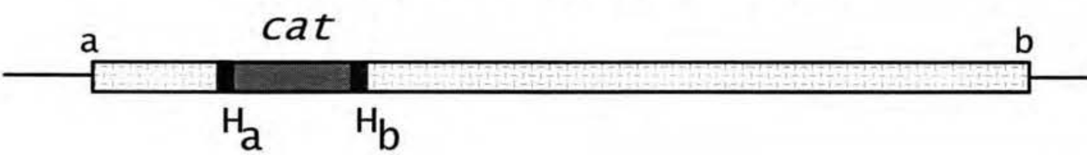


Tn*5253* (65.5 kb, Cm^r, Tc^r Tra^+)



Tn*5251*(18 kb, Tc^r Tra^+)

Tn*5252*(47.5 kb, Cm^r, Tra^+)



Tn*5252* Δ*cat*(39.2 kb, Tra^+)

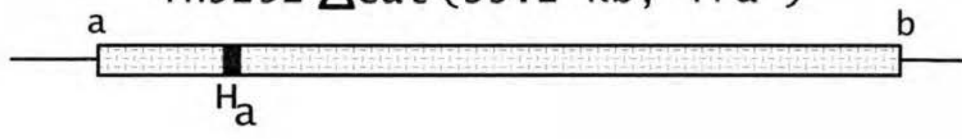
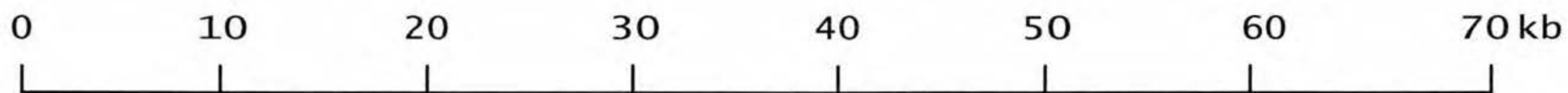
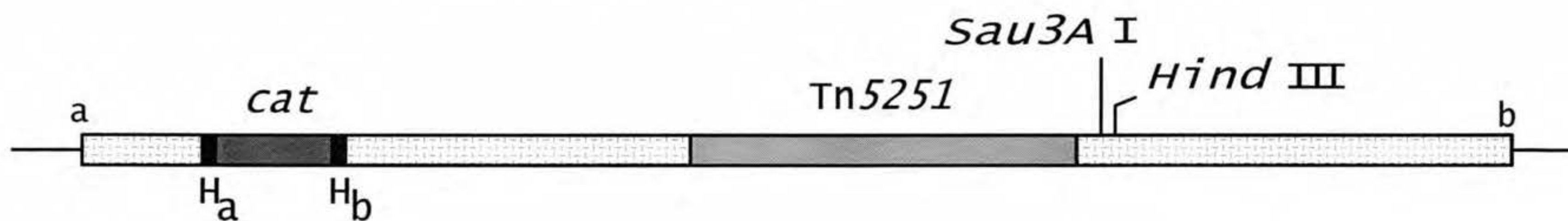
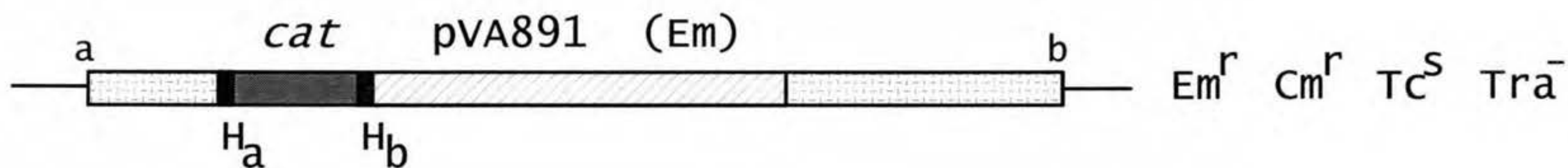
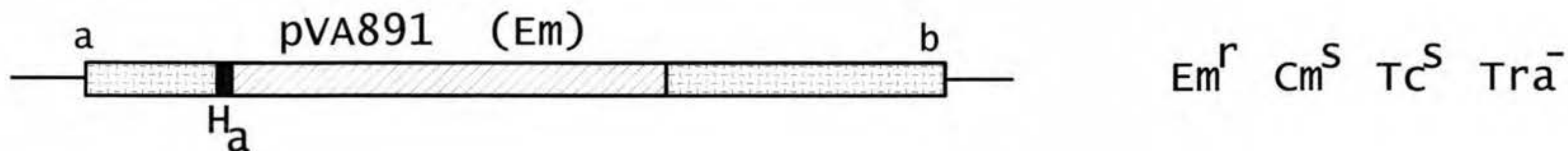


Figure 2. Predicted rearrangements of Tn*5253* upon insertion of heterologous DNA within the *Sau3A*I and *Hind*III sites localized downstream from Tn*5251*. Ha and Hb represent insertion sequences that are flanking the *cat* (chloramphenicol acetyl transferase) gene. Hr represents an insertion sequence located at an unknown position in the chromosome.


0
10
20
30
40
50
60
70 kb
Tn*5253* (65.5 kb, Cm^r, Tc^r Tra^+)


Sau3A I
Hind III
a
cat
Tn*5251*
b
Ha
Hb


a
cat
pVA891 (Em)
b
Ha
Hb
Em^r Cm^r Tc^s Tra^-


a
pVA891 (Em)
b
Ha
Em^r Cm^s Tc^s Tra^-

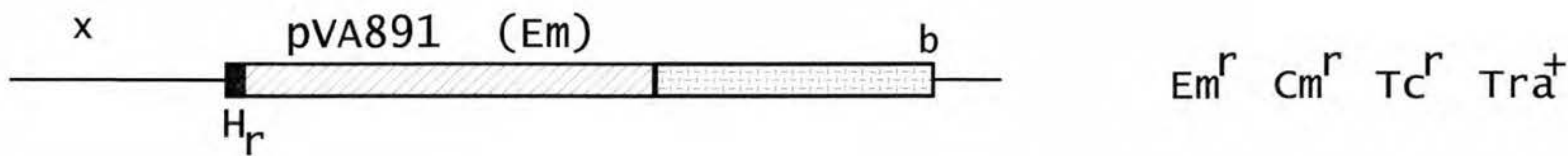
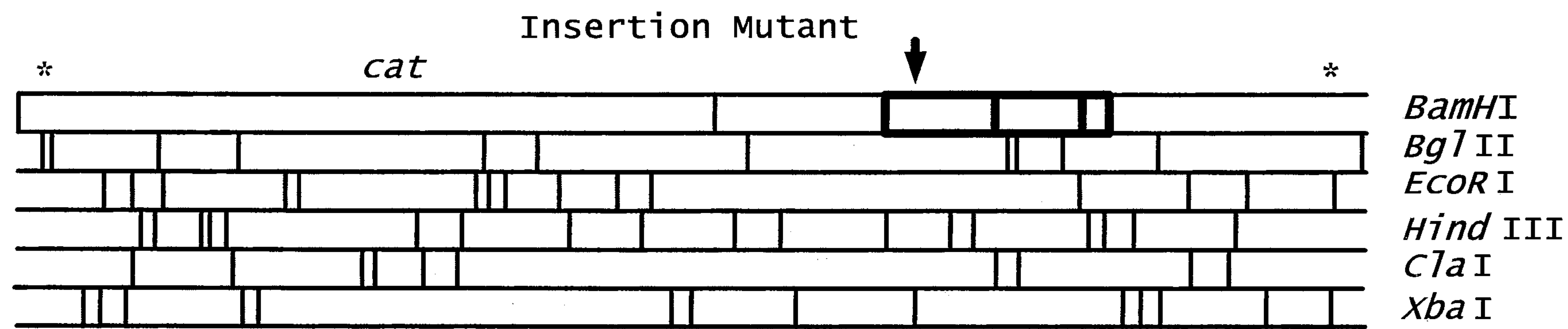

x
pVA891 (Em)
b
Hr
Em^r Cm^r Tc^r Tra^+

Figure 3. Identification of regions involved in transfer in Tn*5252* by using insertion and deletion mutagenesis. The arrow indicates an insertion at a *Xba*I site internal to the 4.2-kb *BamH*I DNA fragment. Hatched boxes below the restriction map indicate fragments deleted from Tn*5252* that resulted in impairment of transfer functions.



8.6-kb *BamH* I fragment

Deletion Mutants

1.9-kb *Bgl* II fragment

CHAPTER 3

MATERIALS AND METHODS

Bacterial Strains and Plasmids

Bacterial strains and plasmids used in this study are described in Tables 1 and 2. *Escherichia coli* strains used for amplification and purification of recombinant plasmids were recombination deficient with the exception of strain C600. *Streptococcus pneumoniae* strains were derived from the non encapsulated wild type strain Rx1 (63). Strains DP1322 (63) and SP1000 (2) are Rx1 derivatives that carry Tn*5253* and Tn*5252* respectively. Strains SP1254 and SP1256 were created by directed insertion of *E. coli* plasmid pVA891 (42) into the Tn*5252* element of SP1000 (34). SP1256 carries an insertion within a *BamH*I 4.2-kb region in Tn*5252* that renders the strain transfer deficient (Tra^-). Insertion of pVA891 in SP1254 does not affect the transfer abilities of Tn*5252* (Tra^+) and this strain was used as a control in conjugation experiments.

Growth Conditions

Strains of S. *pneumoniae* were routinely maintained in CAT broth containing 1.0 % (w/v) casein hydrolysate, 5.0 % (w/v) tryptone, 0.1 % (w/v) yeast extract, and 0.5 % (w/v) NaCl. Sterile CAT broth was supplemented with 0.5 % (w/v) glucose and 15 mM K_2HPO_4 for buffering before use. Pneumococcal strains were grown at 37°C without aeration to an OD_{550nm} of 0.2 (ca. 2×10^8 CFU/ml) to prevent autolysis and the cells were used immediately or stored at -80°C in 10% (v/v) glycerol. Strains of *E. coli* were propagated at 37°C in Luria Bertani (LB) broth supplemented with the appropriate antibiotics to maintain recombinant plasmids. Cultures were stored in 10% (v/v) glycerol at -80°C. For solid medium, 1.5 - 2.0 % (w/v) agar was included in the broth when desired. *Streptococcus pneumoniae* was plated on CAT agar supplemented with 2% (v/v) bovine blood using agar overlay (65) to provide anaerobic conditions while *E. coli* strains were plated on LB agar surfaces. Antibiotic supplements used for selection of bacteria are listed in Table 3.

TABLE 1

E. coli STRAINS AND CLONING VECTORS

Strain or plasmid	Relevant characteristics	Reference or Source
E. coli		
DH5α	Φ80*lacZ*ΔM15, *recA*1, *endA*1, *gyrA*96, *thi-1*, *hsdR*17 (r_K^-, m_K^+), *supE*44, *relA*1, *deoR*, Δ*lacYU*169	Bethesda Research Laboratories
C600	*supE*44, *hsdR*, *thi-1*, *thr-1*, *leuB*6, *lacY*1, *tonA*21	
JM109	*recA*1 *supE*44 *endA*1, *hsdR*17 *gyrA*96 *relA*1 *thiA*, (*lac-proAB*) [F' *traD*36 *proAB* + *lacI*^{q}*lacZ*ΔM15]	
BL21(DE3)	F$^-$ *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal* *dcm* (DE3)	Novagen
Plasmids		
pBluescript SK(+)	2.9-kb E. coli cloning vector; *lacZ*, Amp^r	Stratagene
pLG130	4.2-kb *BamH*I fragment from Tn*5252* cloned into pUC18	(34)
pET30a(+)	5422-bp T7 promoter expression vector; Kan^r	Novagen
pET30b(+)	5423-bp T7 promoter expression vector; Kan^r	Novagen

TABLE I (Continued)

pET30c(+)	5421-bp T7 promoter expression vector; Kan[r]	Novagen
pAF102	4.2-kb *BamHI* fragment from Tn*5252* cloned into pET30b(+)	This study
pAF105	3.3-kb *BamHI* fragment from Tn*5252* cloned into pET30a(+)	This study
pAF106	3.3-kb *BamHI* fragment from Tn*5252* cloned into pET30b(+)	This study
pAF107	3.3-kb *BamHI* fragment from Tn*5252* cloned into pET30c(+)	This study

TABLE 2

STREPTOCOCCAL STRAINS

Strain	Relevant characteristics	Reference
S. pneumoniae		
Rx1	*hex*	(65)
DP1002	*nov*-1	(24)
DP1004	*str-1*	(24)
DP1322	Tn*5253* (*cat tet*)	(69)
SP1000	Tn*5252* (*cat*)	(2)
SP1254	*str-1 fus* Tn*5252* (*cat*ΩEm) Tra$^+$	(34)
SP1256	*str-1 fus* Tn*5252* (*cat*ΩEm) Tra$^-$	(34)
SP1704	*nov-1* Tn*5251* (*tet*) Tra$^+$	(33)
SP1705	*nov-1* Tn*5251* (*tet*) Tra$^+$	(33)
AF101F	*str-1* Tn*5251* (*tet*) Tra$^+$ Tn*5252* (*cat*ΩEm) Tra$^-$	This study
AF101R	*nov-1* Tn*5251* (*tet*) Tra$^+$ Tn*5252* (*cat*ΩEm) Tra$^-$	This study
AF301R	*nov-1* Tn*5251* (*tet*) Tra$^+$ Tn*5252* (*cat*ΩEm) Tra$^+$	This study
AF601	*nov-1* Tn*5251* (*tet*)	This study
AF602	*nov-1* Tn*5252* (*cat*ΩEm)	This study
S. pyogenes		
21547	*opt*	Stillwater Med.Cent.
AF801	*opt* Tn*5252* (*cat*ΩEm)	This study
AF902	*opt* Tn*5251* (*tet*)	This study
AF904	*opt* Tn*5252* (*cat*ΩEm)	This study

TABLE 3

ANTIBIOTICS AND CONCENTRATIONS

Phenotype	Antibiotic	Concentration (μg/ml)	
		Overlay	Stab Plate
S. pneumoniae			
cat (Tn*5253*)	chloramphenicol	15	5
em^r	erythromycin	5	3
fus	fusidic acid	50	10
nov	novobiocin	10	10
rif	rifampicin	10	10
str	streptomycin	200	200
tet	tetracycline	5	2
spc	spectinomycin	500	200
S. pyogenes			
tet (Tn*5251*)	tetracycline	5	2
Em^r	erythromycin	5	1
nov	novobiocin	20	20
str	streptomycin	1000	600
opt	optochin	20	20
E. coli			
Cm^r	chloramphenicol		10
Tc^r	tetracycline		10
Em^r	erythromycin		200
Km^r	kanamycin		50
Ap^r	ampicillin		50
Sm^r	spectinomycin		200

Chemicals, Restriction Endonucleases, and Media

Bacteriological agar and media were obtained from Difco. Proteinase K and all chemicals and antibiotics were obtained from Sigma Chemical Co. Restriction endonucleases, T4 ligase, labeling kits and modifying enzymes were obtained from United States Biochemical (USB), New England Biolabs (NEB), Promega Corp., Bethesda Research Laboratories (BRL), or Boehringer-Mannheim Corp., and used as described by the manufacturer. Agarose and acrylamide reagents were purchased from Fisher and Bio Rad Laboratories. Radionuclides were obtained from New England Nuclear Co. (NEN) and Amersham Co.

Molecular Cloning Techniques

DNA digestions, ligations and agarose gel electrophoresis were performed as outlined by Sambrook *et al.* (56). DNA samples were prepared with an appropriate amount of tracking dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 20% (v/v) glycerol, 0.1 M EDTA, pH 8.0) and loaded onto the gel along with a molecular weight standard marker. Electrophoresis was routinely carried out using 0.8 to 1.2% (w/v) agarose gels and TBE buffer (89 mM Tris-base, 89 mM boric acid, and 2 mM EDTA, pH 8.0). After electrophoresis, gels were stained with 1.0 μg/ml ethidium bromide for 10 min and destained in deionized

water for 10 to 20 min. DNA was visualized on the gels using an UV transilluminator and photographed using a Kodak Polaroid camera. For resolution and recovery of DNA fragments, samples were run in Ultra Pure DNA Grade agarose (Bio Rad) gels using the above conditions. Fragments were extracted and purified by electroelution essentially as described by Sambrook *et al.* (56).

Preparation of *S. pneumoniae* competent cells

Preparation of pneumococcal competent cells was performed as described by Guild and Shoemaker (24). Bacterial cells were grown at 37°C in CAT medium to an OD_{550} of 0.2 (c.a. $2x10^8$ CFU/ml). Cells were then diluted 100-fold in competence medium (CTM) (CAT broth supplemented with 10 mM $CaCl_2$ and 0.2% (w/v) bovine serum albumin [Fraction V]). After 70 min of growth, 1-ml aliquots of cells were collected at 10-min intervals, mixed with 10% glycerol, frozen in acetone-dry ice for 5 min, and stored at -80°C. Stored cells were tested for level of competence by plotting the number of transformants as a function of time. After the optimum competence time was determined, a stock of competent cells was obtained by inoculating a 250 ml-volume of CTM and collecting 20 ml-aliquots of cells at the competence peak time observed in the previous experiment. Cells were mixed with 10% (v/v) glycerol and stored at -80°C.

Transformation of *E. coli*

Recombinant plasmids containing the relevant fragments from Tn*5252* were introduced in *E. coil* competent cells by transformation using the technique described by Hanahan (26). Transformant cells were isolated by plating appropriate dilutions of the transformation mixture on LB agar supplemented with the appropriate antibiotics. IPTG (isopropyl-β-D thiogalactopyranoside) and X-Gal (5-bromo-4-chloro-3-indolyl-β-D galactoside) were also included in the plates when needed.

Transformation of *S. pneumoniae*

Competent cells were thawed on ice, mixed with donor DNA and incubated at 37°C for 30 min. DNase I was then added to a final concentration of 10 μg/ml followed by incubation at 37°C for 5 min. Transformant cells were transferred to ice and appropriate dilutions were plated on CAT agar using the overlay method.

Plasmid Isolation from *E. coli*

Plasmid DNA screening in *E. coli* was performed by using the mini-lysis technique described by Sambrook et al. (56). A 1-3 ml volume of an overnight culture of *E. coli* was centrifuged at 11,750 x g at room

temperature for 3 min. Supernatant was decanted and pellet was resuspended in 200 μl of a solution containing 50 mM glucose, 25 mM Tris-HCl and 10 mM EDTA, pH 8.0. The cell suspension was lysed by addition of 200 μl of a fresh solution of 0.2N NaOH and 1% (w/v) SDS and then neutralized with 200 μl of 3M potassium acetate, pH 4.5. Lysate was kept on ice for 5 min and centrifuged for 5 min at 11,750 x g. Supernatant was transferred to a fresh tube and DNA was precipitated by adding 2 volumes of 95% ethanol and incubating at -20°C for 20 min. The DNA solution was centrifuged as above and the pellet was dried under vacuum and resuspended in 100 μl of TE buffer containing 100 μg/ml DNase-free pancreatic RNase. Further purification of plasmid DNA isolated in this way was carried out by a second extraction with phenol-chloroform, followed by ethanol precipitation when required.

Large-scale isolation of plasmid DNA from *E. coli* was performed using cesium chloride-ethidium bromide density gradient ultracentrifugation. A crude lysate was obtained by following a scaled-up version of the protocol used for plasmid DNA screening. An overnight 250-ml culture of *E. coli* was centrifuged at 5,800 x g for 10 min at 4°C. The bacterial pellet was resuspended in 5ml of resuspension solution (50 mM glucose, 25 mM Tris-HCl and 10 mM EDTA, pH 8.0)and transferred to 25-ml Oakridge tubes. A 5-ml volume of lysis solution (0.2N NaOH and 1% SDS) was added to the cell suspension followed by addition of 5 ml of

neutralization solution (3M potassium acetate, pH 4.5) and incubation on ice for 10-20 min. The lysate was then centrifuged at 28,000 x g for 20 min at 4°C and DNA was precipitated by adding two volumes of 95% ethanol and incubating at -20°C for 20 min. Precipitated DNA was pelleted by centrifugation at 28,000 x g for 20 min at 4°C, dried under vacuum and resuspended in 4 ml of TE (10mM Tris-HCl, 1 mM EDTA, pH 8.0) buffer. For density gradient centrifugation, a XL-70 Beckman Ultracentrifuge and a VTi65 rotor were used. The DNA solution was mixed with 4.0 grams of cesium chloride and 0.4 ml of 10 mg/ml ethidium bromide and then transferred to 5-ml quick-seal tubes. Centrifugation was carried out at 227,640 x g, 17°C for 18 h, or 383,700 x g for 4 h. Plasmid DNA was recovered with a 16-G needle attached to a 3-ml syringe and purified by extracting several times with water-saturated butanol followed by dialysis against TE buffer for 18 h. Quality of the DNA was spectrophotometrically assessed by determining the $A_{260/280}$ ratio and by gel electrophoresis.

Chromosomal DNA Isolation from *S. pneumoniae*

Broth cultures (200 ml) of pneumococcal strains were grown in CAT broth supplemented with 0.02% (w/v) choline chloride. After reaching an OD_{550} of 0.3-0.4, cultures were mixed with 10 ml 10 mM EDTA and kept on ice for 10 min. Cells were washed twice with 50 mM Tris-HCl, 20 mM EDTA, pH 7.5 by centrifugation at 5,000 x g in a Sorvall RC-5B centrifuge

(DuPont Instruments). Cell pellets were resuspended in 5 ml of the same buffer followed by addition of 1 ml of a solution containing 0.6% (w/v) triton X-100, 0.06% (w/v) sarkosyl, 0.6% (w/v) sodium deoxycholate (DOC), and 300 μg/ml RNase . The cell suspension was incubated at 37°C until lysis was evident. Lysis was completed by adding 1 ml of 1% (w/v) SDS and 350 μg/ml Proteinase K and overnight incubation at 65°C. DNA was purified by phenol-chloroform extraction and ethanol precipitation followed by drying and resuspension in 400 μl of TE buffer.

DNA Sequence Analysis

Sequence analysis was performed by the Sanger dideoxy-chain termination method with denatured double-stranded DNA as the template using a Sequenase Version 2.0 kit (USBC) and [α-^{35}S]-ATP (NEN), as well as by automated sequencing at the Recombinant DNA/Protein Resource Core Facility, Oklahoma State University. Manual sequence reactions were carried out as directed by the manufacturer specifications.

Generation of Nested Deletions

To determine the sequence of the segments of interest, overlapping DNA fragments were obtained by generating nested deletions from fragments cloned in pUC and pBluescript SK+ vectors. For this purpose the

double strand specific 3' to 5' exonuclease activity of the modifying enzyme Exonuclease III (USB) was used. The recombinant plasmids were digested with two different restriction enzymes in such a way as to generate a protected 4-base 3' end and a 5' overhang or blunt end adjacent to the insert from which deletions were to proceed. Digested DNA (ca. 5 μg) was mixed with 10 μl of 10X exonuclease III buffer (0.6 M Tris-HCl pH 8.0, 6 mM $MgCl_2$) and 450 units of exonuclease III enzyme. The reaction mixture was incubated on ice and 10 μl aliquots were removed every minute for 10 min. These samples were transferred to tubes containing 30 μl of S1 nuclease (BRL) (30 units in 10% (v/v) glycerol, 60 mM NaCl, 1.3 mM $ZnSO_4$, and 8 mM potassium acetate pH 4.6) and held on ice until all samples were taken. Tubes were removed from ice and the reaction was allowed to proceed for 20 min at room temperature. The reaction was terminated by adding 4 μl of stop buffer (0.7 M Tris-Base pH 8.0, 50 mM EDTA) and by heat inactivating at 70°C for 10 min. Samples were extracted with phenol-chloroform and chloroform-isoamyl alcohol followed by ethanol precipitation, self-ligated and used to transform *E. coli* competent cells. Transformants were screened for recombinant plasmids bearing inserts of smaller size and suitable for sequencing. When the partial sequences obtained from these clones were not sufficient to generate overlaps, synthetic oligomeric DNA primers were synthesized to complete the entire sequence on both strands.

Preparation of Double-Stranded DNA Template

Double-stranded DNA was prepared from overnight cultures of *E. coli* with the Magic mini-prep kit (Promega Corp.) by using a modification of the supplier's methodology. Cells from a 10 ml culture were pelleted at full speed in a centrifuge and resuspended in 200 µl of 50 mM Tris-HCl, pH 7.5, 10 mM EDTA and 100 µg/ml RNase A. Cells were the lysed by addition of 200 µl of a solution containing 1% (w/v) SDS and 0.2N NaOH at room temperature. After mixing, 200 µl of a neutralization solution (2.55M potassium acetate) was added to the tube, mixed again and centrifuged at full speed. Supernatants were transferred to a new tube to which 2 volumes of DNA purification resin (Promega) were added. The lysate-resin solution was transferred to a syringe and the contents were passed through a purification column. The column was washed two more times by passing 2 ml of a wash solution (0.2M NaCl, 20 mM Tris-HCl, pH 7.5, 5mM EDTA and 50% (v/v) ethanol). Excess wash was removed by centrifugation for 20 sec at full speed. DNA was eluted from the column by adding 100 µl of TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) at 65°C and centrifugation for 20 sec.

Denaturation of Template

Template DNA was denatured by adding 0.2N NaOH and 0.2mM EDTA at room temperature to a solution containing at least 1 μg of double-stranded DNA. Neutralization was carried out by addition of 0.3M Tris-HCl (pH 4.5) and 0.8M sodium acetate. DNA was then precipitated with two volumes of 95% ice-cold ethanol, washed with 70% ethanol and vacuum-dried. Denatured DNA was stored at -20C.

Annealing of Primers to Template

Primers for DNA sequencing were obtained from the Recombinant DNA/Protein Resource Core Facility, Oklahoma State University. A reaction containing the template, primer (30 ng per assay) and 1X reaction buffer (200 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl, 0.1 M dithiothreitol (DTT)) was incubated at 37C for 30 min for annealing.

Labeling of DNA

Template-primer mixtures were labeled by adding 1 μl of 0.1M DTT, 2 μl of labeling nucleotide mix (1.5 μM dGTP, 1.5 μM dCTP, and 1.5 μM ^{35}S-dATP) and 2 μl of Sequenase Version 2.0 enzyme. Extension was carried out on ice for 10 min.

Termination

Extension of DNA was terminated by transferring 3.5 µl aliquots of the mixture to each of four pre-warmed (49°C) microcentrifuge tubes containing 2.5 µl of the appropriate dideoxynucleotides (ddNTP). Incubation at 49°C for 4 min was followed by termination of the reaction by addition of 4µl of stop solution (95% (v/v) formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% (w/v) xylene cyanol FF). Samples were heated at 90°C before loading onto a sequencing gel or stored at -20°C for later use.

Denaturing Polyacrylamide Sequencing Gel

Polyacrylamide gels [7% (w/v)] used for sequencing were prepared by mixing two solutions. Acrylamide solution A contained 17.5 % (v/v) of a 40% polyacrylamide stock solution (38% (w/v) acrylamide and 2% (w/v) bis-acrylamide), 25% (w/v) urea, and 10% 10X TBE buffer. Acrylamide solution B had the same composition except that it was supplemented with 25% (w/v) sucrose and 0.005 g (w/v) of bromophenol blue. Gels containing a concentration gradient were prepared by using a ratio of solution A to B of 1:5 when filling the chamber of a gel apparatus (Sequi-Gen Nucleic Acid Sequencing Cell, BIORAD). The acrylamide solutions were degassed separately for 10 min and then 0.3% (w/v) ammonium sulfate and 0.05%

(v/v) N, N, N',N'-tetramethylethylene-diamine (TEMED) were added. After pouring, the gel was allowed to polymerize at room temperature for at least 18h.

Gel Electrophoresis

For electrophoresis, the top reservoir of the gel apparatus was filled with 0.5X TBE buffer and 1X TBE buffer was added to the bottom reservoir. The gel was prerun at 1500 V and 35W until the temperature indicator reached 45C. Samples were loaded in a staggered manner to avoid migration by diffusion into the gel and electrophoresis continued at the same settings. Electrophoresis was usually allowed to proceed for 5-10 h depending on the length of the sequence desired. At the end of the run, the glass plate to which the gel was attached was carefully separated from the integral plate and fixed in a solution containing 10% glacial acetic acid for 1 h with occasional agitation. The gel was rinsed with distilled water and dried overnight at 65°C in a dry air oven. Autoradiograms were prepared by exposing the gels to X-ray films (Kodak XAR-5) for an appropriate length of time.

Analysis of DNA sequences

Sequence data was assembled and analyzed by using MacVector v 4.5 (IBI), AssemblyLign (IBI) and DNASIS (Hitachi) software programs. Prediction of transmembrane segments was done by using TMpred and TMbase as described by Hofmann and Stoffel (29). Search for homologies was performed by comparing the deduced amino acid sequences to protein databases using the BLAST Network Service at NCBI via University of Oklahoma, and the Genetics Computer Group (GCG, Madison, Wisconsin) package for sequence analysis.

Protein Expression

For detection of products expressed from the cloned fragments, the *E. coli* pET expression system (Novagen) was employed. Recombinant clones containing the fragments of interest in the correct orientation were used to transform the expression host *E. coli* BL21 (DE3). After a target plasmid was established in this strain, a single colony from a fresh plate was inoculated into 2 ml of LB medium containing 50 μg/ml of kanamycin. Cultures were incubated with shaking at 37°C until the OD_{600} reached 0.6-1.0 and then stored at 4°C overnight. The cells from these cultures were pelleted by centrifugation for 30 sec and resuspended in an equal volume of fresh LB medium. These cells were used as inoculum for 50-ml cultures.

Expression of the target DNA was performed by incubating at 37°C to a recommended OD_{600} of 0.6 (approximately 3 h). Cells were divided in two 25 ml cultures for use as control and for titering. Induction was initiated by adding 1 mM IPTG to the growing culture and incubation was continued for another three hours. After induction, cells were kept on ice for 5 min and 500 µl samples were pelleted by centrifugation at 5000 x g for 5 min at 4°C. Supernatant was removed and pellet was washed twice in 1/4th culture volume of cold 2mM EDTA, 50mM Tris-HCl pH 8.0. The cells were finally resuspended in 50 µl of SDS sample buffer (1.0% (w/v) SDS, 1.0% (v/v) mercaptoethanol, 10% (v/v) glycerol, 60 mM Tris-HCl, pH 6.8) and stored as a frozen pellet at -70°C.

Cell Fractionation

Periplasmic Fraction

A 500-µl sample from an induced culture was centrifuged at 7000 x g for 10 minutes at 4°C. Supernatant was discarded and cells were resuspended in 200 µl (80ml/g cells) of 30mM Tris HCl pH 8.0, 20% (w/v) sucrose and 1 mM EDTA, followed by incubation for 5-10 minutes at room temperature with shaking or stirring. Cells were then centrifuged at 10,000 x g for 10 minutes at 4°C and the pellet resuspended in an equal volume of ice-cold 5mM $MgSO_4$. After 10 min of incubation on ice, cells

were centrifuged as above. The supernatant, which contained the periplasmic fraction, was transferred to a new centrifuge tube and stored at -20°C. The pellet was kept for membrane proteins isolation.

Membrane and cytoplasmic fractions

Pellets resulting from periplasmic proteins isolation were gently resuspended in 100 μl of 30 mM Tris-HCl pH 8.0, 20% sucrose. Cells were then disrupted by using a Sonic Dismembrator (Fisher Scientific) with a microtip sonicator. Cells were sonicated ten times for 20 sec each. The solution was centrifuged at 10,000 x g for 15 min to separate membrane (pellet) and cytoplasmic fractions (supernatant). Proteins in the cytoplasmic fraction were precipitated by adding 400 μg/ml of yeast tRNA and trichloroacetic acid (TCA) to a final concentration of 10% (v/v). The protein mixture was incubated at 4°C for 30 min and then centrifuged at 11,750 x g for 5 min at 4°C. The supernatant was discarded and pellets containing the protein were washed twice by using an equal volume of ice-cold acetone and centrifugation as above. Protein pellet was dried at 65°C and resuspended in an appropriate volume of 10 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS and 1% (v/v) mercaptoethanol. For electrophoresis, an equal volume of SDS-sample buffer was added to the samples prior to loading.

SDS-PAGE Analysis

SDS-PAGE was carried out according to the discontinuous buffer system of Laemmli (35) in acrylamide (30:0.8 acrylamide:bis-acrylamide w/w) vertical slab gels (16 cm x 18 cm x 1 mm). A 25-ml volume of 12% (w/v) acrylamide solution containing 0.1% (w/v) SDS and 0.375 M Tris-HCl, pH 8.8 was used for the separating gel. The solution was degassed for 10 min at room temperature, 250 μg/ml ammonium persulfate and 5 μl/ml TEMED were added and then it was poured between glass plates. The acrylamide mixture was overlaid with 2-5 ml of butanol and allowed to polymerize for 30 min at room temperature. After polymerization, butanol was removed from the gel surface and rinsed with a solution containing 0.1 % (w/v) SDS and 0.375 M Tris-HCl, pH 8.8. For the stacking gel, 10 ml of a 3% acrylamide mixture containing 0.1% (w/v) SDS and 0.125 M Tris-HCl, pH 6.8 was used. After degassing for 10 min, 150 μg/ml ammonium persulfate and 1.5 μl/ml TEMED were added to the solution which was poured on top of the separating gel and allowed to polymerize overnight with a comb in place. Tris-glycine buffer (25 mM Tris-Base, 0.192 M glycine, 0.1% (w/v) SDS) was used for the running buffer. Samples containing approximately 10 μg of protein dissolved in SDS sample buffer were heated to 90°C for 5 min and loaded onto the stacking gel. Electrophoresis was carried out at 20 mA constant current until the bromophenol blue marker dye had reached the bottom of the stacking gel. The power supply was then adjusted to 45 mA constant current and the gel

run until the dye front was about 0.5 cm from the bottom of the separating gel. The gel was removed from the glass plates, stained overnight in a solution containing 0.2% (w/v) coomassie blue, 50% (v/v) methanol, and 7% (v/v) glacial acetic acid, and destained in several changes of 7% (v/v) glacial acetic acid.

Conjugation Analysis

Broth cultures of donor and recipient strains were grown at 37°C without aeration to a density of 2 X 10^8 cfu/ml in CAT medium supplemented with 0.001% (w/v) choline chloride and no selection. Both donor and recipient cells were mixed at a ratio of 1:1 in the presence of 10 mM $MgSO_4$, 2 mg/ml bovine serum albumin (BSA), 1 mM $CaCl_2$, 0.2% yeast extract, and 1 mg/ml DNase I. A 5 ml volume of the cell suspension (ca. 2 X 10^9 cells) was passed through nitrocellulose filters (pore size: 45µm, Millipore). Filters were then placed cell-side-down onto 2% CAT agar containing 10 mM $MgSO_4$, 2 mg/ml BSA, 1 mM $CaCl_2$, 0.2% (w/v) yeast extract, and 100 µg/ml DNase I. Filters were overlaid with 6 ml of the same agar and the plates were incubated for 4 h at 37°C. After incubation, the filters and the top and bottom agar portions were removed from the plates and transferred to 3 ml of CAT broth containing 10 mM $MgSO_4$, 2 mg/ml BSA, 150 µg/ml DNase I, and 10% (v/v) glycerol. The cell suspension was vortexed thoroughly to remove cells from filters and serial dilutions were plated on selective media using the agar overlay method.

Prior to the addition of medium containing antibiotics, plates were incubated at 37°C for 90 min to allow for phenotypic expression. The frequency of conjugation was calculated by dividing the number of transconjugants per ml by the number of donor cells present after the mating.

Southern Hybridization

Transfer of DNA

Southern transfer of DNA on 0.8% agarose gels to GeneScreen Plus membranes (Du Pont, NEN Research Products) was done by a protocol of Southern (71) and according to the manufacturer's directions. Nylon membranes were pre-wetted in deionized water for 10 min and in 10X SSC buffer (1.5 M NaCl, 0.15 M Sodium citrate) for another 10 min. DNA was denatured by soaking the gel in 0.5 N NaOH for 30 minutes with gentle shaking. The gel was placed on top of the equilibrated membrane inside a vacuum blotter (Model 785 Vacuum Blotter, Bio Rad). Using 10X SSC as the transfer solution, DNA transfer was carried out for 90 min at 5 inches of Hg. After transfer was completed, the membrane was removed from the gel, soaked in 2X SCC for 10 min and air-dried at room temperature.

Nick Translation

Probes were labeled with [α-^{32}P]dCTP (Amersham Corporation) by following the protocol described by Sambrook *et al.* (56). The labeling reaction was made by mixing 200 ng of probe DNA, 10 µl 5X nucleotide mix (10 mM dATP, dGTP, and dTTP in 50 mM Tris-HCl, pH 7.5), 0.5µl DNase I (1 U/µl), 3 µl [α-^{32}P]dCTP (10 mCi/ml), 0.5 µl *E. coli* DNA polymerase I (900 U/µl) and sterile water to a final volume of 50 µl. The mixture was incubated at 16°C for 2 h and then stopped by adding 25 µl of 0.5 M EDTA, pH 8.0, 25 µl of salmon sperm DNA (2 mg/ml), 50 µl of 7.5M ammonium acetate, and 50 µl TE buffer. Labeled probe was precipitated by adding 2 volumes of 95% ethanol and by incubating at -20°C for 15 min. DNA was pelleted, washed with 70% ethanol, dried under vacuum and resuspended in 600 µl of TE buffer. To determine the total radioactivity of the labeled probe, 1 µl of the reaction mixture was transferred to 5 ml of Ready-Solv liquid scintillation cocktail (Beckman) and the activity counted in a Beckman LS6000C scintillation counter. An appropriate volume of probe was mixed with 100 µl of 20X SSC buffer, 400 µl of 2 mg/ml salmon sperm DNA and the total volume was brought to 2 ml by adding sterile deionized water. The labeled probe was used for hybridization immediately or stored at -20°C for later use.

Prehybridization and Hybridization

Membranes were placed in heat sealable bags and prehybridized with 10 ml of a solution containing 1 M NaCl, 1% (w/v) SDS and 10% (w/v) dextran sulfate. Prehybridization was carried out overnight with gentle agitation at 65°C. The probe solution was denatured by heating for 10 min at 95-100°C and placed on ice for 5 min before adding to the prehybridization buffer. The probe concentration in the bag was adjusted to 1.0 to 4.0 X 10^6 dpm. After addition of the probe, bags were resealed and agitated overnight at 65°C for a minimum of 16 h. Membranes were removed from the bags and washed twice in excess 2X SSPE buffer (3 M NaCl, 200 mM NaH_2PO_4, 20 mM EDTA, pH 7.4) for 15 min at room temperature. Washings were repeated using 2X SSPE, 2% SDS for 45 min at 65°C, 0.1X SSPE for 30 min at room temperature, and 3 mM Tris-Base (unbuffered) for 15 min at room temperature. Excess buffer was blotted from membranes by drying at 37°C for 10 min. Membranes were placed in plastic wrap and exposed to X-ray film (Kodak BioMax MS) with an intensifying screen and stored at -80°C for an appropriate period of time.

CHAPTER 4

GENETIC AND TRANSCRIPTIONAL ORGANIZATION OF A REGION ESSENTIAL FOR DNA TRANSPORT IN Tn*5252*

RESULTS AND DISCUSSION

Identification and Nucleotide Sequence Analysis

Earlier studies dealing with the composite transposon Tn*5253* suggested the presence of a potential regulatory sequence around a 0.3-kb *Hind*III/*Sau3A*I DNA fragment located downstream from the right end of Tn*5251* (Fig. 2) (77). Deletion of an 8.6-kb *BamH*I DNA segment spanning this region was recently shown to render Tn*5252* incapable of conjugal transfer (34). This segment is located between coordinates 33.0 and 41.6 in the restriction map of Tn*5252* shown in Fig 3. Directed insertion and deletion mutagenesis in this DNA section revealed the presence of two sites that appeared to be potentially associated with conjugal transfer. An inactivating insertion was mapped to an internal *Xba*I site (coordinate 34.4) while an inactivating deletion was found after removing a 1.9-kb *Bgl*II fragment (coordinates 37.9 to 39.8). In order to further characterize the

entire 8.6-kb area, smaller DNA fragments derived from this segment were subcloned into pUC8 and pBluescript SK(+). Recombinant plasmids generated were pLG130 containing a 4.2-kb *BamHI* DNA segment (coordinates 33.0 to 37.2); pDR8, containing a 3.2-kb *BamHI* fragment (coordinates 37.2 to 40.5); and pDR9 containing a 1.1-kb *BamHI* fragment (coordinates 40.5 to 41.6) (34). Additional sequence was obtained from pDR14, a recombinant plasmid carrying a 0.9-kb *Hind*III DNA fragment (coordinates 41.4 to 42.3) contiguous to the right end of the 8.6-kb segment.

Sequence Analysis

To determine the nucleotide sequence, a series of overlapping DNA fragments was obtained by Exonuclease III and S1 nuclease treatment of the recombinant plasmids. A nested set of 20 deletion derivatives was sequenced using the M13 universal and reverse primers. Additional synthetic primers were designed when necessary to complete the sequence in both strands (Fig. 4). After assembly and analysis of the sequences from the different DNA fragments, the entire region was found to be 9,522 nt in length (coordinates 33.0 to 42.5 in Fig. 3) with a G+C content of approximately 38.0% as expected for streptococcal DNA. As shown in Fig. 4, transcription of the potential ORFs proceeds from right to left. The complete sequence presented in the orientation of transcription is shown

in Fig. 5. Nucleotide 1 starts at coordinate 42.5 and proceeds to the end of the segment to coordinate 33.0 in the Tn*5252* restriction map.

Nucleotide Sequence Organization

Nucleotide sequence analysis identified 10 open reading frames (ORFs) that were all predicted to be transcribed in the same orientation (Fig. 4). A summary of their characteristics is shown in Table 5. The sequence of the passenger DNA fragments in plasmids pDR9 and pDR14 were 1,099 and 989 nt respectively. Since these two fragments overlap 217 nt, the combined size of both sequences was found to be 1,871 nt (Fig. 3, coordinates 42.5 to 40.6). Examination of the sequence revealed two ORFs, designated ORF20 and ORF21, that were 339 and 1,410 nt in length respectively (Table 5). ORF20 is predicted to utilize an ATG codon located at nt 468 and a TAA stop codon at nt 805. ORF21 extended 228 nt into the contiguous 3.2-kb *BamH*I with a GTG start codon at nt 909 and a TGA stop codon located at nt 2,315 (Fig. 5). Consensus ribosomal site (RBS) were located within 10 nt from their corresponding start sites (Table 4). Sequence of the contiguous *BamH*I DNA fragment, contained in pDR8, was found to be 3,243 nt in length (Fig. 3, coordinates 40.6 to 37.3). Analysis of the nucleotide sequence showed 5 small ORFs that were designated ORF22, ORF23, ORF24, ORF25a, and ORF25b, and the amino-terminal of a longer ORF (ORF26) that extended 1,171 nt in the same

direction of transcription. The small ORFs ranged in sizes from 195 to 342 nt and all of them showed consensus RBS (Table 4). With the exception of ORF25a, no transcriptional start sites commonly reported for prokaryotic genes were detected for the rest of the ORFs. ORFs 22 and 23 utilized TGA as the stop codon (Fig. 5, nt 2581 and 3075 respectively) while ORFs 24, 25a, and 25b used a TAA stop codon (Fig. 5, nt 3508, 3771, and 3920 respectively). ORF25b overlapped both the carboxyl-terminal of ORF25a and the amino-terminal of ORF26 (Fig. 5). The latter started at a TTG codon located at nt 3,867 which was preceded by consensus RBS (Table 4). The TAA stop codon of ORF26 was found to reside in the contiguous 4.2-kb *BamHI* fragment and it was localized at nt 6,228.

The passenger DNA in pLG130 was found to be 4193 nt (coordinates 33.0 to 37.2) and its analysis revealed the presence of 2 more intact ORFs which were also transcribed in the same orientation (Fig. 4). These ORFs designated ORF27 and ORF28 were 1,560 and 1,116 nt in size respectively. Predicted ORF27 contained a TTG translational start codon at nt 6,313 and a TAA stop codon at nt 7,870. ORF28 extended from a GTG codon at nt 7,956 to a TGA codon at nt 9,069. Both ORFs showed near-consensus ribosomal binding sites within 8 to 10 nt from the start site (Table 4). Downstream from ORF28, a 53-nt DNA sequence directly repeated twice was located that was followed by a putative rho factor-independent terminator having a stem length of 10 nt and a loop size of 4 nucleotides (6) (Fig. 4). The direct repeats showed a consistent sequence

pattern in which some A's and T's in the first repeat are replaced with G's and C's respectively in the second repeat. In addition, the first 20 bases showed a sequencing artifact where an extra base was added during replication in one orientation while a base was skipped when the sequence was replicated in the opposite orientation.

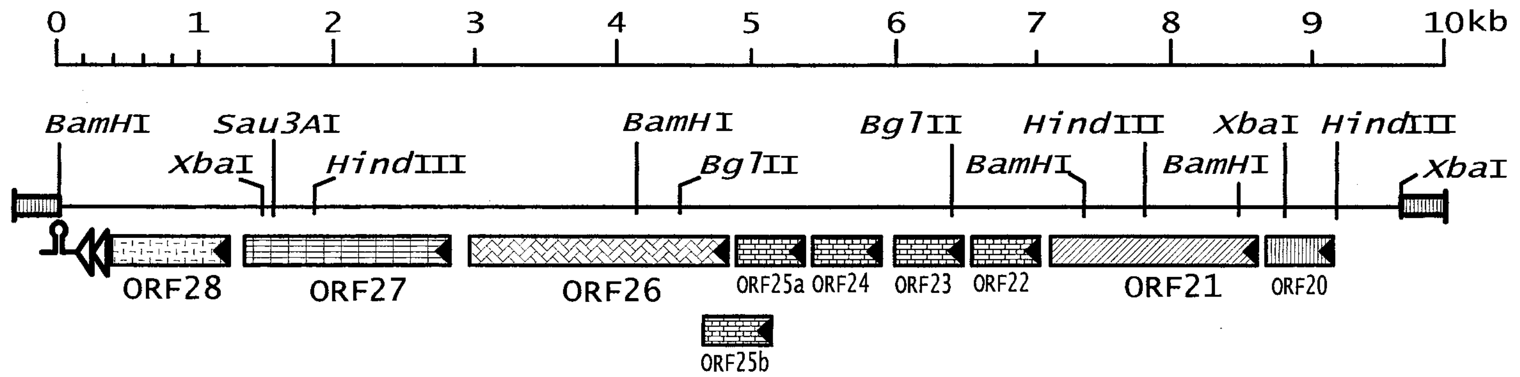
TABLE 4

SUMMARY OF THE DNA ANALYSIS OF THE 9.5-Kb *BamHI* DNA FRAGMENT

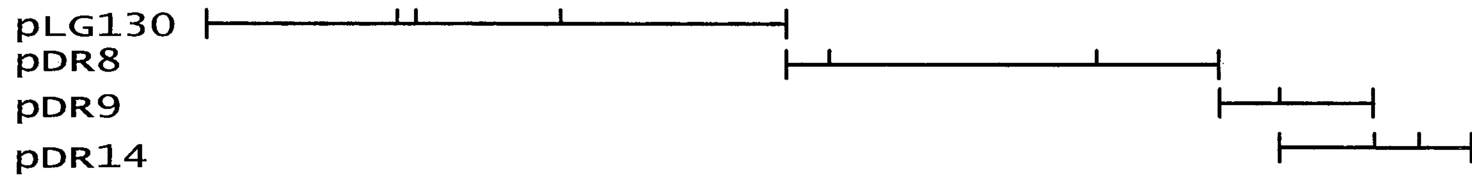
ORF	Size (bp)	Start Codon	Stop Codon	Putative RBS[a]
ORF20	339	ATG	TAA	CAAACAGAAAGAGAGGTGAG
ORF21	1,410	GTG	TGA	ATAGAAATAAGAATATTGCG
ORF22	195	ATA	TGA	CCATAGAATAAGAGGAAATA
ORF23	342	TTT	TGA	CTGAAAATTTATTTGAAGGA
ORF24	237	ATT	TAA	TTTAAGTTGATGAATGGAAG
ORF25a	210	ATG	TAA	CCTTTTTAGAAAGGAAAGTC
ORF25b	207	TTT	TAA	TCTGGCAATTTGGAGATTGG
ORF26	2,364	TTG	TAA	AATTTCAGGAAAGAAAGGAC
ORF27	1,560	TTG	TAA	TAAAAGATGAGAAGAAATTC
ORF28	1,116	GTG	TGA	AAAACAAGAAAGAAAAATTT

[a]RBS, Putative Ribosomal Binding Site indicated by underlining of the corresponding sequence

Figure 4. Schematic representation of the restriction map of the 9.5-kb DNA fragment from Tn*5252* and the predicted ORFs. Relevant restriction sites are diagramed above the map. Hatched boxes below the restriction map represent ORFs 20 to 28. The filled arrows inside the boxes indicate the direction of transcription. Open arrows at the left end of ORF28 represent a 53-bp direct repeat that is followed by a putative transcription terminator (Ω). (A) Thick horizontal lines represent the DNA segments contained in the subclones used for sequencing and their designations are shown on the left of the diagram. (B) Strategy used for sequencing both DNA strands which is indicated by thin lines with filled arrows going in both directions.



A



B

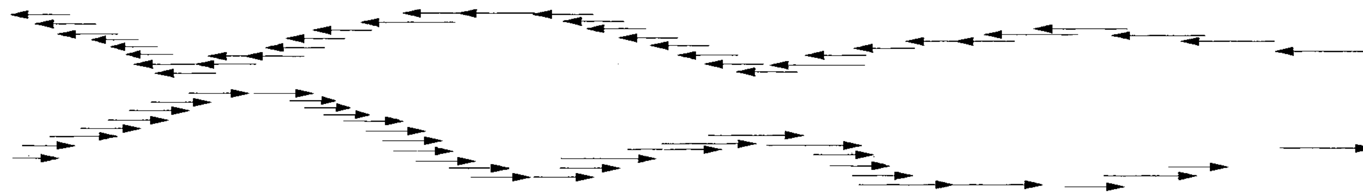


Figure 5. Complete nucleotide sequence of the 9.5 *BamHI* fragment from Tn*5252.* Nucleotide 1 starts at coordinate 42.5 in Tn*5252* (Fig. 3). Translated sequences from open reading frames are indicated above the DNA sequence. Start and stop codons are shown in bold. Potential ribosomal binding sites (RBS) are underlined. Relevant restriction sites in the DNA fragment are shown below the nucleotide sequence. Long arrows indicate a 53-bp direct repeat from nt 9128 to 9235.

```
   1  TGAACAACAGGAACGAATGGCAGTTCAATATGCTGAGCGTAGTCTTTTATTCACTGTCAA

  61  AAGTCTTTTAAAGATTCTAGACGTCAGGCTGTAGCACAGGATTCCGCCTATAAGATAGGG
                        XbaI
 123  GCGCAGAAATTAGAAGAGTTGCTACAATCTCCTCATTCGATTGATACGATTAATCTGTAA

 181  AAAGATTTCTTAGACCAACCAATTGATATAGAGAAATTTAAAGCTTTTTTAGAAAAAGAA
                                                      HindIII
 241  GAGATTCCTTTAGCCATCGCTTGGCAAGGAGATTCTCTGCATTTCTACACGAAAGTACCG

 301  TTCGATTCTAGACAATCATTTAGACCATCTGTTAGAAAAAATGGTTAATGATCCGGAGTA

 361  AATTAGCTGATTTTACCATGGATAAGTCATTAGACGATGCAATTGATGAGGCTAAATCCC
                                              ORF20 ->
   1                                                  M  M  Y  S
 421  AAATTACCTTTAGACAAGAAGGGGCCGTCAAACAGAAAGAGAGGTGAGATGATGTACAGT

   5  G  K  K  F  L  L  F  S  L  L  G  I  L  L  G  Y  L  F  H  R
 481  GGAAAGAAATTCCTACTATTCTCACTGTTAGGCATCTTACTAGGCTATCTTTTTCATCGT

  25  L  T  L  L  Y  D  S  Y  T  G  N  T  L  D  K  W  T  R  L  L
 541  TTGACGCTTTTGTATGATTCCTATACTGGAAATACTTTAGATAAATGGACTCGTCTCCTG

  45  M  E  G  Q  E  E  V  L  Q  S  P  W  N  I  S  F  T  G  K  S
 601  ATGGAAGGTCAAGAGGAAGTTCTTCAGTCGCCATGGAATATTTCTTTCACTGGAAAATCA

  65  S  A  F  F  L  L  G  F  V  M  M  L  L  V  Y  L  Y  L  E  T
 661  AGTGCTTTTTTTCTACTAGGCTTTGTGATGATGTTGCTGGTTTATCTCTATCTAGAGACC
                                                            XbaI
  85  G  K  K  Q  Y  R  E  G  V  R  I  R  G  A  P  V  L  E  L  L
 721  GGTAAGAAACAATATCGAGAAGGAGTTAGAATACGGGGAGCGCCCGTTTTGGAACTCCTA

 105  K  E  K  N  L  L  L  R  *
 781  AAAGAAAAGAATCTCCTTTTACGGTAAGGAATTTTCTCCATGATACGATCTTTAGCTCCA

 841  AGATGTTCCGTTTGACATTTATTAGATAGAGAGGCCCCACCCAATATGATAGAAATAAGA
      ORF21 ->
   1          V  I  G  G  S  G  S  G  K  T  F  R  F  V  K  P  N  L
 901  ATATTGCGGTGATCGGAGGTTCAGGAAGTGGGAAGACATTTCGCTTTGTGAAACCCAATC

  19   I  Q  M  N  S  S  N  I  V  V  D  P  K  D  H  L  A  E  K  T
 961  TGATTCAGATGAATAGTTCTAATATTGTAGTGGATCCTAAAGATCACTTGGCCGAAAAAA
                                          BamHI
  39   G  K  L  F  L  E  H  G  Y  Q  V  K  V  L  D  L  V  N  M  K
1021  CAGGCAAACTCTTTTTAGAACATGGCTACCAAGTAAAGGTGTTAGATTTAGTTAATATGA

  59   N  S  D  G  F  N  P  F  R  Y  I  E  T  E  N  D  L  N  R  M
1081  AGAACTCAGATGGCTTCAATCCTTTTCGCTATATAGAGACAGAAAATGATTTGAATCGCA

  79   L  A  V  Y  F  N  N  T  K  G  S  G  S  R  S  D  P  F  W  D
1141  TGCTGGCGGTTTATTTCAATAACACCAAAGGCTCTGGCTCCCGTAGTGATCCATTTTGGG
```

```
  99   E  A  S  M  T  L  V  R  A  L  A  S  Y  L  V  D  F  Y  N  P
1201  ATGAAGCTTCTATGACTTTGGTTCGAGCTTTAGCCTCCTACTTGGTCGATTTCTATAATC
        HindIII
 119   P  K  T  R  E  Q  L  I  E  E  S  R  L  S  Q  K  E  Y  Q  N
1261  CACCTAAAACAAGAGAACAGCTCATAGAAGAAAGTCGTTTAAGTCAAAAAGAATACCAAA

 139   L  L  K  R  Q  K  K  E  V  E  E  R  K  K  R  G  R  L  S  K
1321  ACTTGTTGAAACGTCAAAAAAAAGAAGTGGAAGAGCGAAAAAAACGAGGGCGGTTATCCA

 159   F  C  E  S  Q  N  S  L  N  T  Y  P  R  V  K  T  R  K  S  V
1381  AGTTTTGTGAATCTCAAAACTCATTAAACACTTATCCAAGGGTGAAAACCAGAAAAAGTG

 179   L  E  I  L  F  E  N  Y  A  K  K  Y  G  T  E  N  F  T  M  R
1441  TCTTAGAAATTCTATTTGAAAATTATGCTAAAAAGTATGGAACTGAAAATTTTACCATGC

 199   N  W  A  D  F  Q  N  Y  K  D  K  T  L  D  S  V  I  A  V  T
1501  GAAATTGGGCAGATTTTCAAAATTATAAGGATAAGACTCTGGATTCTGTTATAGCTGTAA

 219   T  A  K  F  A  L  F  N  I  Q  S  V  M  D  L  T  K  R  D  T
1561  CCACTGCTAAATTTGCCCTCTTCAATATTCAAAGTGTCATGGATTTGACCAAAAGAGATA

 239   L  D  M  K  T  W  G  Q  E  K  S  M  V  Y  L  V  I  P  D  N
1621  CCCTTGATATGAAGACATGGGGCCAGGAAAAATCAATGGTTTACTTAGTTATCCCAGATA

 259   D  S  T  F  R  F  L  S  A  L  L  F  F  N  P  Y  F  Q  T  P
1681  ACGATAGTACCTTTCGCTTTCTTTCAGCCCTCCTTTTTTTCAACCCGTATTTCCAAACCC

 279   N  K  T  S  Q  I  L  M  L  R  V  R  L  P  L  H  V  R  V  Y
1741  CTAACAAGACAAGCCAGATATTGATGTTAAGGGTCAGATTGCCTCTTCATGTGAGAGTTT

 299   L  D  E  F  A  N  I  G  E  I  P  D  F  A  E  Q  T  S  T  V
1801  ACTTAGATGAGTTCGCAAATATTGGAGAAATCCCAGATTTTGCTGAACAAACCTCAACAG

 319   R  S  R  N  M  S  L  V  P  I  L  Q  N  I  A  Q  L  Q  G  L
1861  TCCGTTCTCGTAATATGAGTCTCGTTCCCATTCTTCAAAATATTGCCCAACTTCAAGGAC

 339   Y  K  E  K  E  A  W  K  T  I  L  G  N  C  D  S  L  V  Y  L
1921  TCTATAAAGAAAAAGAAGCTTGGAAAACCATTCTTGGGAACTGTGATAGCTTAGTCTACT
                        HindIII
 359   G  G  N  D  E  D  T  F  K  F  M  S  G  L  L  G  K  Q  T  I
1981  TAGGTGGGAATGATGAAGATACCTTTAAATTTATGAGTGGGTTACTCGGTAAACAAACCA

 379   D  V  R  N  T  S  R  S  F  G  Q  T  G  S  G  S  L  S  H  Q
2041  TTGATGTTCGAAATACTAGTCGTTCCTTTGGCCAGACAGGTTCAGGATCCCTTTCTCATC
                                                       BamHI
 399   K  I  A  R  D  L  M  T  P  D  E  V  G  N  M  K  R  H  E  C
2101  AAAAGATTGCTCGTGATTTAATGACACCTGATGAAGTCGGAAATATGAAACGGCATGAAT

 419   L  V  R  I  A  N  M  P  V  F  K  S  K  K  Y  N  S  T  K  H
2161  GCCTGGTTCGAATTGCCAATATGCCTGTCTTTAAAAGCAAAAAATACAATTCAACTAAGC
```

```
 439   P  N  W  K  Y  L  A  N  Q  E  T  D  E  R  R  W  E  L  S  N
2221  ATCCAAACTGGAAGTACCTAGCCAATCAAGAAACCGATGAACGGCGGTGGGAACTATCAA

 459   Q  S  F  K  S  K  T  R  K  S  S  *
2281  ATCAATCCTTTAAATCAAAGACAAGAAAATCATCTTGAAGGCCTTAGAATTCCGACGATT
                                                ORF22 ->
   1                                                     I  D  D  T  H
2341  TAACTGTTTGAATCCTAGTTTAAAATAACCATAGAATAAGAGGAAATAGATGATACGCAT

   5  F  K  V  L  G  G  L  W  S  Q  G  T  H  S  A  I  V  L  A  T
2401  TTTAAGGTTTTGGGGGGTCTATGGAGCCAGGGGACGCACAGTGCTATTGTTCTGGCAACA

  25  K  L  I  V  V  G  L  Q  E  G  I  D  C  G  W  C  L  S  Q  C
2461  AAGCTTATTGTTGTCGGGTTACAAGAAGGGATTGATTGCGGCTGGTGCCTTTCTCAGTGT

  45  V  D  A  T  L  Q  P  T  L  R  W  W  T  R  C  S  E  D  Q  S
2521  GTGGATGCAACCTTGCAACCAACATTAAGATGGTGGACCAGGTGTTCGGAATGCCAATCT

  65  *
2581  TGAAATTGTCGGTGGAGTTATGGTATGGGCTGCTGGAGCCTTTGTAACCCAGATTTCAAT

2641  TTAGGAGGATAAACAATGACATGAATCTTACTTTAGTCTCACCCTGGTGTATACCTTGCA
                                  ORF23 ->
   1                                          F  N  V  H  L  Q  S  P  V
2701  TCTGAAAATATCAGACTGAAAATTTATTTGAAGGATTTAATGTACATTTACAATCACCGG

  10   D  L  I  K  S  L  S  S  Y  N  P  T  V  W  T  Y  M  S  S  I
2761  TAGATCTGATTAAATCTCTATCTAGCTACAATCCAACTGTTTGGACTTATATGTCTAGTA
       BglII
  30   T  K  S  V  M  Q  P  L  G  V  A  I  L  S  V  V  L  I  L  E
2821  TTACTAAAAGTGTCATGCAGCCTCTTGGAGTTGCGATTTTATCAGTTGTTCTCATCTTAG

  50   F  S  K  M  A  K  K  I  A  N  S  G  G  A  M  T  F  E  A  L
2881  AATTTTCGAAGATGGCAAAGAAAATTGCTAACTCAGGTGGAGCGATGACTTTTGAAGCAT

  70   A  P  M  L  I  S  Y  I  M  V  A  V  V  I  T  N  T  T  V  I
2941  TAGCGCCGATGTTGATTAGTTATATTATGGTCGCAGTTGTAATTACCAATACTACCGTTA

  90   V  E  A  I  I  G  I  A  I  T  P  L  N  N  A  S  I  G  S  R
3001  TTGTAGAAGCTATCATCGGGATTGCGATCACGCCATTGAACAATGCTTCGATTGGCTCAC

 110   W  G  K  V  *
3061  GGTGGGGCAAAGTATGATACACTCTCTGGATTAAAAGGTTCAGGATTTATTGGCCGGATG

3121  ATTGTGGGCTTTTTCGCCCTCCTCATTTGGCTTGTTCGGATAGTAAGTGCAGCCATGGTT

3181  AATCTTTTGGTATCTATTCGATTTATTCAACTCTACCTTATGATTCCATTTGCCCCTCTT
                                  ORF24 ->
   1                                            I  G  I  G  Y  L  K  N
3241  ACGATTCCAACATTTTTAAGTTGATGAATGGAAGTCTATTGGTATTGGCTATTTAAAAAA
```

```
  10  I  M  V  Y  A  V  Q  G  V  L  I  F  L  I  V  S  L  V  P  L
3301  TATTATGGTTTATGCGGTACAAGGGGTTCTCATTTTTCTGATTGTTTCTCTTGTTCCTTT

  30  F  E  S  A  G  K  I  A  V  S  N  G  A  G  V  L  Q  S  L  A
3361  GTTTGAATCTGCTGGGAAAATAGCTGTTTCAAATGGTGCAGGAGTCTTGCAATCACTTGC

  50  I  M  F  G  S  L  V  Q  A  I  L  L  I  I  A  L  V  G  S  Q
3421  GATTATGTTTGGTAGTTTGGTACAAGCTATCTTACTGATTATTGCCCTCGTTGGTTCTCA

  70  R  T  A  R  S  I  L  G  M  *
3481  ACGTACGGCTCGCTCAATTTTAGGTATGTAATTAGATAAAGCTAGGAAGTGATTGCTTC
                ORF25a ->
   1                             M  N  T  R  V  F  K  D  I  S  K  Y
3541  TTAGCCTTTTAGAAAGGAAAGTCATGAATACACGTGTCTTTAAAGACATCTCAAAATAC

  13  Q  H  R  A  W  L  G  F  T  T  R  Q  I  I  F  V  L  P  A  F
3601  CAACACAGGGCTTGGTTAGGCTTCACTACAAGGCAAATCATCTTTGTCCTACCAGCCTTT
                                                 ORF25b ->
   1                                                          F  V
  33  I  V  T  Y  C  F  G  L  E  S  L  F  L  A  I  W  R  L  V  C
3661  ATTGTCACTTATTGTTTTGGGCTTGAATCTCTTTTTCTGGCAATTTGGAGATTGGTTTGT

   3  Y  G  F  V  F  A  F  T  N  P  L  M  L  F  G  V  Y  N  P  M
  53  L  R  F  C  V  C  L  Y  Q  S  P  H  A  F  W  S  L  *
3721  TTACGGTTTTGTGTTTGCCTTTACCAATCCCCTCATGCTTTTTGGAGTCTATAACCCAAT

  23  I  Y  D  F  E  H  Y  L  K  Y  R  L  H  F  E  L  T  I  P  L
3781  GATTTATGATTTTGAACATTATTTGAAATACCGTCTTCATTTTGAACTAACGATACCCCT

  43  R  T  I  S  G  K  K  G  L  E  H  E  K  K  I  K  Y  I  K  E
   1                ORF26 ->   L  N  M  K  R  K  S  N  T  L  K  K
3841  ACGCACAATTTCAGGAAAGAAAGGACTTGAACATGAAAAGAAAATCAAATACATTAAAGA

  63  T  T  N  F  N  D  *
  13   Q  Q  T  S  T  T  N  K  K  E  E  V  K  D  K  K  E  E  V  L
3901  AACAACAAACTTCAACGACTAATAAAAAGGAAGAAGTTAAAGATAAAAAAGAGGAAGTGT

  33   P  S  T  A  N  T  L  S  Y  Q  A  L  Y  Q  N  G  L  M  Q  V
3961  TACCATCAACGGCTAATACTCTTTCCTATCAAGCCTTGTATCAAAATGGTCTGATGCAGG

  53   K  E  D  Y  F  S  Q  S  Y  L  L  G  D  V  N  Y  Q  T  V  G
4021  TAAAGGAAGATTATTTTTCACAAAGCTATTTACTTGGTGATGTCAATTACCAGACCGTTG

  73   L  E  D  K  G  A  I  I  E  K  Y  S  D  L  I  N  P  L  D  D
4081  GTTTAGAAGATAAGGGCGCAATCATTGAGAAGTATTCTGATTTGATTAATCCTTTAGATG

  93  Q  T  N  F  Q  L  T  I  F  N  K  R  L  N  L  E  K  F  R  H
4141  ACCAAACCAACTTCCAATTGACCATCTTTAATAAAAGATTGAATTTAGAAAAATTCAGAC
```

```
113   S  V  L  Y  E  E  K  E  D  G  Y  D  S  Y  R  K  E  L  N  R
4201  ACAGTGTTTTGTATGAGGAAAAAGAAGATGGGTACGATAGCTATCGTAAAGAATTGAATC

133   M  M  N  Q  N  L  D  S  G  E  N  N  F  S  A  V  K  L  I  S
4261  GGATGATGAATCAAAATTTAGACAGTGGGGAAAATAACTTTTCAGCTGTGAAACTGATTA

153   F  G  R  K  D  S  N  P  K  Q  A  Y  R  S  L  S  Q  I  G  E
4321  GCTTTGGTAGGAAGGATTCTAATCCCAAACAAGCCTATCGTTCCTTGTCTCAAATTGGAG

173   Y  F  K  S  G  F  S  E  I  D  A  R  F  E  S  L  A  G  E  E
4381  AATATTTCAAGAGTGGTTTCTCAGAAATTGATGCTCGATTTGAATCCTTGGCTGGAGAAG

193   R  V  N  L  L  A  D  M  L  R  G  E  H  H  L  P  F  S  Y  R
4441  AACGGGTGAACTTGTTGGCAGATATGCTTAGAGGAGAACATCATCTTCCTTTTTCTTACC

213   D  L  T  R  S  G  Q  T  T  R  H  F  I  A  P  N  L  L  D  F
4501  GTGATTTAACGAGATCTGGTCAGACAACTCGTCACTTCATAGCACCTAATCTCTTGGATT
                BglII
233   K  N  K  N  Y  L  Q  I  N  D  R  L  L  Q  I  V  Y  V  R  D
4561  TTAAAAACAAGAATTACCTACAAATCAATGACCGCTTATTACAGATTGTCTATGTGAGAG

253   Y  G  M  E  L  G  D  Q  F  I  R  D  L  M  Q  G  D  L  E  L
4621  ACTACGGTATGGAATTAGGGGATCAGTTTATCCGAGACCTCATGCAAGGAGATCTGGAAT

273   I  V  S  L  H  A  Q  S  S  T  K  S  D  A  M  K  K  L  R  T
4681  TGATTGTAAGCCTTCATGCTCAAAGTTCGACCAAGTCAGATGCCATGAAGAAACTACGAA

293   K  K  T  L  M  E  S  Q  K  I  G  E  Q  Q  K  L  A  R  T  G
4741  CAAAGAAAACCTTAATGGAATCCCAAAAGATTGGGGAACAACAAAAACTAGCTCGTACAG

313   I  Y  L  E  K  V  G  H  V  L  E  S  N  I  D  E  A  E  E  L
4801  GTATCTATTTGGAAAAAGTAGGTCATGTATTAGAAAGCAATATCGATGAAGCCGAGGAAC

333   L  K  T  M  T  E  T  G  D  K  L  F  Q  T  V  F  L  I  G  V
4861  TCTTAAAAACCATGACCGAGACAGGAGATAAACTATTTCAAACGGTCTTCTTGATTGGGG

353   F  G  Q  D  E  E  E  L  K  Q  A  L  D  T  V  Q  Q  V  A  G
4921  TCTTTGGTCAGGATGAAGAAGAACTCAAACAAGCCCTAGACACTGTCCAACAAGTGGCCG

373   S  N  D  L  M  I  D  K  L  P  Y  M  Q  E  A  A  F  N  S  L
4981  GCTCAAATGACCTAATGATTGATAAACTTCCATATATGCAAGAAGCAGCTTTTAATAGTT

393   L  P  F  G  C  D  F  L  E  G  V  S  R  S  L  L  T  S  N  I
5041  TGCTGCCATTTGGTTGTGATTTTTTAGAGGGAGTATCACGGAGTTTATTAACGTCCAATA

413   A  V  N  S  P  W  T  S  V  D  L  Q  D  R  S  G  K  Y  Y  G
5101  TAGCAGTGAACTCACCTTGGACTTCAGTAGACCTACAAGACCGTAGTGGGAAATATTACG

433   I  N  Q  I  S  S  N  I  I  T  I  D  R  S  L  L  N  T  P  S
5161  GTATCAATCAAATATCAAGTAATATTATTACCATTGATCGCAGCCTATTAAATACACCGT
```

```
453   G  L  I  L  G  T  S  G  A  G  K  G  M  A  T  K  H  E  I  I
5221  CTGGTCTGATTTTAGGAACATCTGGAGCAGGGAAAGGGATGGCAACCAAGCATGAAATTA

473   T  T  K  I  K  E  S  G  E  N  T  E  I  I  I  V  D  P  E  A
5281  TCACGACCAAAATCAAGGAATCTGGTGAAAATACTGAAATTATCATCGTGGATCCAGAAG
                                                        BamHI

493   E  Y  S  V  I  G  R  T  F  G  G  E  M  I  D  I  A  P  D  S
5341  CAGAGTACAGTGTCATTGGACGGACTTTTGGGGGAGAAATGATTGATATTGCGCCTGATT

513   E  T  Y  L  N  V  L  D  L  S  E  E  N  M  D  E  D  P  V  K
5401  CCGAAACCTATCTCAATGTCCTTGACTTGTCTGAGGAAAATATGGATGAAGATCCTGTAA

533   V  K  S  E  F  L  L  S  F  I  G  K  L  L  D  R  K  M  D  G
5461  AGGTAAAATCAGAATTTCTTTTATCCTTTATCGGCAAGTTATTGGATAGAAAAATGGATG

553   R  E  K  S  I  I  D  R  V  T  R  L  T  Y  Q  S  F  K  E  P
5521  GAAGAGAAAAAATCGATTATCGACCGAGTTACGAGGCTCACCTATCAGTCATTTAAAGAGC

573   S  L  E  E  W  V  F  V  L  S  Q  Q  P  E  E  E  A  Q  N  L
5581  CTTCTTTGGAAGAATGGGTCTTTGTATTGAGTCAACAACCAGAAGAAGAAGCGCAGAATT

593   A  L  D  M  E  L  Y  V  E  G  S  L  D  I  F  S  H  K  T  N
5641  TGGCACTTGATATGGAACTGTATGTCGAAGGTTCTCTTGATATTTTTTCTCATAAGACCA

613   I  Q  T  G  S  N  F  L  I  Y  N  V  K  K  L  G  D  E  L  K
5701  ATATTCAGACAGGATCTAATTTCTTGATTTATAACGTTAAGAAGTTAGGAGATGAGCTGA

633   Q  I  A  L  M  V  V  F  D  Q  I  W  N  R  V  V  R  N  Q  K
5761  AACAAATCGCCCTTATGGTTGTTTTTGATCAGATATGGAATCGTGTCGTTCGGAACCAAA

653   L  G  K  K  T  W  I  Y  F  D  E  I  E  L  L  L  L  D  K  Y
5821  AATTAGGGAAGAAGACCTGGATTTATTTTGATGAAATCGAGCTTCTCTTATTAGATAAAT

673   P  S  D  F  F  F  K  L  W  S  R  V  R  K  Y  G  A  S  P  T
5881  ATCCGAGTGATTTCTTCTTTAAATTGTGGAGTCGTGTCAGAAAATATGGAGCCAGTCCGA

693   G  I  T  Q  N  V  E  T  L  L  L  D  P  N  G  R  R  I  I  A
5941  CTGGAATAACTCAAAACGTCGAAACCTTATTGTTAGATCCAAATGGTAGACGGATTATTG

713   N  S  E  F  M  I  L  L  K  Q  A  K  N  D  R  E  E  L  V  Q
6001  CAAATAGTGAATTTATGATTCTCCTCAAGCAAGCAAAAAATGATAGAGAAGAACTGGTTC

733   L  L  G  L  S  K  E  L  E  K  Y  L  V  N  P  E  K  G  A  G
6061  AACTCTTAGGCTTGTCAAAAGAACTTGAAAAATACCTTGTCAATCCAGAAAAAGGGGCAG

753   L  I  K  A  G  S  V  V  V  P  F  K  N  K  I  P  Q  G  S  Q
6121  GACTGATAAAAGCTGGTTCAGTTGTTGTTCCCTTTAAAAATAAGATTCCTCAAGGATCTC

773   L  F  D  I  M  R  S  D  P  D  K  M  A  S  N  *
6181  AATTGTTTGATATCATGAGATCAGATCCTGATAAAATGGCTTCTAATTAAGGGGAAGGTA
```

6241 AATGAAGGATAAAAGAGAAATCATACGTGCCCGAAAGCATTTAGAAGAAGTCTAAAAGAT

ORF27 ->

1 L K Q G K K E V R K Q K K D S A
6301 GAGAAGAAATTC**TTG**AAACAAGGAAAGAAGGAGGTGAGGAAACAGAAAAAAGATTCCGCT

17 G L D E K A W K K E I K E K L E E M R E
6361 GGACTGGATGAAAAAGCATGGAAAAAAGAGATAAAAGAAAAGCTAGAGGAGATGAGAGAA

37 A S K A R V K Q A N E D Y N H I L Q N S
6421 GCTTCAAAGGCTAGAGTAAAACAAGCAAATGAAGACTACAATCATATTCTTCAAAATAGT

57 P P S L L N R K E L R D R R L P H A R K
6481 CCTCCATCTCTTTTAAATCGCAAAGAATTAAGAGACAGACGGTTGCCTCATGCTAGGAAA

77 R L K I A K K Q F K E G S K G R S K R R
6541 CGATTGAAAATAGCCAAGAAGCAATTTAAGGAAGGAAGCAAAGGTAGAAGCAAAAGAAGA

97 K K E S R K E R K P I K N F S T G Q E S
6601 AAGAAAGAGAGTCGTAAAGAAAGAAAACCAATCAAAAATTTTTCTACGGGTCAGGAATCG

117 K H K S N F F F Q G K S L E E L K A K K
6661 AAACATAAATCTAATTTTTTCTTTCAAGGGAAGAGTTTAGAAGAATTAAAAGCTAAGAAA

137 E V K A A K E N L K S T K Q V Y K S K K
6721 GAAGTCAAGGCCGCAAAAGAGAATCTAAAATCTACTAAACAAGTCTATAAGTCCAAAAAA

157 V S R K A K T F L Y V L G R E G G E L A
6781 GTCAGTAGGAAAGCCAAAACTTTTCTTTATGTCCTTGGACGTGAAGGTGGAGAGTTAGCT

177 S E N E D L E G Y H T L Q E T I R K G K
6841 TCAGAAAATGAAGATTTAGAAGGTTATCACACACTTCAAGAGACAATTAGAAAAGGAAAA

197 R Y S R L S Y N L G K A S V K T G Q A T
6901 CGCTACAGTCGCCTTTCTTATAACCTTGGAAAAGCTAGTGTCAAAACAGGACAAGCAACA

217 G R F T K K R L T N T K E R Y H H F K D
6961 GGTCGTTTTACCAAGAAAAGACTGACCAACACAAAAGAGCGATACCATCATTTTAAGGAT

237 G K G W K L A K D N P S S F K N R F R K
7021 GGAAAAGGATGGAAACTAGCGAAAGATAACCCAAGTTCTTTTAAAAATCGGTTTCGAAAA

257 L K K Q G L T S V R N I Y Q K L K A A F
7081 TTAAAGAAACAAGGTCTTACAAGTGTCCGAAATATCTATCAAAAACTAAAAGCAGCCTTT

277 S F F T F A A G N L V T W I V G G I V F
7141 TCCTTCTTTACATTTGCGGCTGGAAATCTTGTAACCTGGATAGTTGGAGGAATAGTCTTT

297 L L L L I M S F F L G F S S A S L I Q Q
7201 CTTCTTTTACTTATAATGAGCTTCTTTTTAGGATTTTCATCTGCTAGTTTGATTCAACAA

317 D E F E L T K A Y T H L T W E D A E H T
7261 GATGAATTTGAATTAACAAAAGCTTATACCCACCTAACTTGGGAAGATGCAGAACATACT

337 R T N D K G I T Y Y T K V D D V M G Y M
7321 CGCACAAATGACAAAGGAATTACTTATTACACAAAAGTTGATGATGTGATGGGCTATATG

357 N F K F H D Y E L H K P V H L F S S E S
7381 AACTTTAAATTCCATGACTATGAGTTACACAAACCAGTTCACTTATTTAGTTCAGAATCT

377 Y K D Y L S T L W H D L N D G D D L K S
7441 TACAAGGATTATCTGTCTACTTTGTGGCATGATTTAAACGATGGGGATGATTTGAAATCC

397 M Q D L Y E T P K Y K L S K D D Q E E I
7501 ATGCAAGACCTCTATGAAACTCCTAAGTATAAACTATCGAAAGACGATCAAGAGGAAATA

417 K E L K E E G V Y A S M Q X L D N P F E
7561 AAGGAACTAAAAGAAGAGGGTGTCTATGCTTCCATGCAGGRATTGGACAATCCATTTGAG

437 G K S V P D S L T M T Y R Y G Y Y D L D
7621 GGGAAAAGCGTGCCAGATAGTCTAACCATGACTTATCGTTATGGATACTATGATTTAGAC

457 G K P T L Q E Y I L L E A K A H Q T I V
7681 GGAAAACCTACTCTTCAGGAGTATATTCTACTAGAAGCGAAGGCTCACCAAACAATTGTC

477 A P M D G V V S L D G D D V I L T N G K
7741 GCACCAATGGATGGTGTTGTATCTCTAGACGGAGATGATGTTATTCTCACTAACGGAAAA

XbaI

497 G E N E S R L T L Y S I H N G R A I E G
7801 GGAGAGAATGAGAGTCGATTGACCTTGTATTCTATTCATAATGGCCGTGCGATTGAGGGG

517 T R V *
7861 ACAAGAGTCTAACGGGTGATATTATTGGTGAAACACCAGACGATACACCTTTGAAAGTTT

ORF28 ->

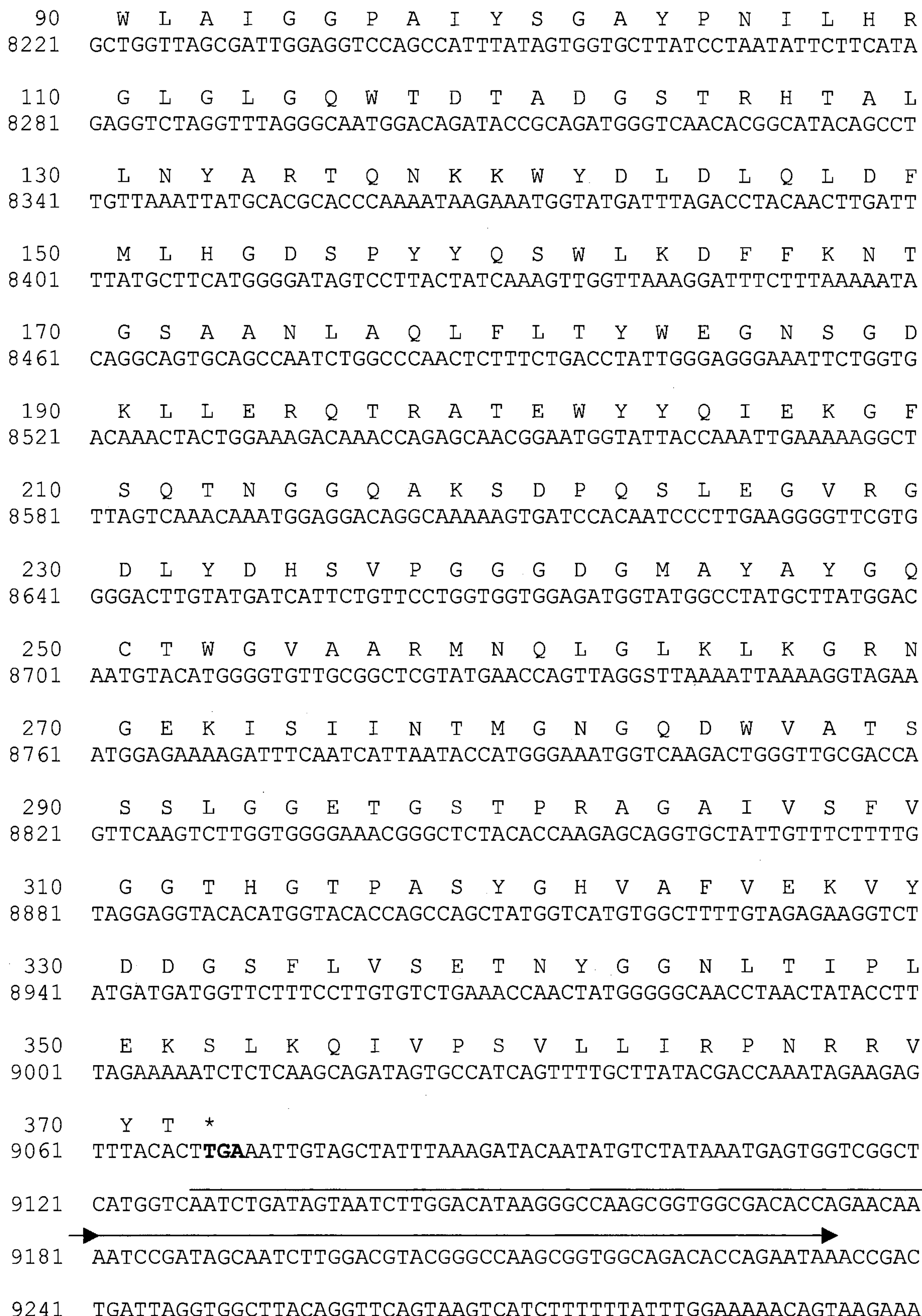
1 V Y V N P Q F Y F
7921 CCTATCAAAAGTATAAAAACAAGAAAGAAAAATTT**GTG**TATGTCAATCCGCAATTTTATT

10 P K V I Q L Q T T I L P A I G Q F G G D
7981 TTCCAAAAGTCATTCAACTTCAGACCACTATCTTACCAGCCATTGGTCAGTTTGGTGGGG

30 E F E R A K H I Y E F L K S Q G A S P Q
8041 ATGAGTTTGAACGAGCAAAACATATTTATGAGTTTTTGAAATCTCAAGGGGCAAGTCCCC

50 A I A A I L G N W S V E S S I N P K R A
8101 AAGCCATTGCGGCAATTTTAGGAAATTGGTCGGTAGAGTCTTCTATTAATCCTAAACGAG

70 E G D Y L T P P V G V P I P P W D D E S
8161 CTGAAGGAGATTATTTAACTCCTCCTGTTGGCGTACCGATTCCTCCATGGGATGATGAAA



90 W L A I G G P A I Y S G A Y P N I L H R
8221 GCTGGTTAGCGATTGGAGGTCCAGCCATTTATAGTGGTGCTTATCCTAATATTCTTCATA

110 G L G L G Q W T D T A D G S T R H T A L
8281 GAGGTCTAGGTTTAGGGCAATGGACAGATACCGCAGATGGGTCAACACGGCATACAGCCT

130 L N Y A R T Q N K K W Y D L D L Q L D F
8341 TGTTAAATTATGCACGCACCCAAAATAAGAAATGGTATGATTTAGACCTACAACTTGATT

150 M L H G D S P Y Y Q S W L K D F F K N T
8401 TTATGCTTCATGGGGATAGTCCTTACTATCAAAGTTGGTTAAAGGATTTCTTTAAAAATA

170 G S A A N L A Q L F L T Y W E G N S G D
8461 CAGGCAGTGCAGCCAATCTGGCCCAACTCTTTCTGACCTATTGGGAGGGAAATTCTGGTG

190 K L L E R Q T R A T E W Y Y Q I E K G F
8521 ACAAACTACTGGAAAGACAAACCAGAGCAACGGAATGGTATTACCAAATTGAAAAAGGCT

210 S Q T N G G Q A K S D P Q S L E G V R G
8581 TTAGTCAAACAAATGGAGGACAGGCAAAAAGTGATCCACAATCCCTTGAAGGGGTTCGTG

230 D L Y D H S V P G G G D G M A Y A Y G Q
8641 GGGACTTGTATGATCATTCTGTTCCTGGTGGTGGAGATGGTATGGCCTATGCTTATGGAC

250 C T W G V A A R M N Q L G L K L K G R N
8701 AATGTACATGGGGTGTTGCGGCTCGTATGAACCAGTTAGGSTTAAAATTAAAAGGTAGAA

270 G E K I S I I N T M G N G Q D W V A T S
8761 ATGGAGAAAAGATTTCAATCATTAATACCATGGGAAATGGTCAAGACTGGGTTGCGACCA

290 S S L G G E T G S T P R A G A I V S F V
8821 GTTCAAGTCTTGGTGGGGAAACGGGCTCTACACCAAGAGCAGGTGCTATTGTTTCTTTTG

310 G G T H G T P A S Y G H V A F V E K V Y
8881 TAGGAGGTACACATGGTACACCAGCCAGCTATGGTCATGTGGCTTTTGTAGAGAAGGTCT

330 D D G S F L V S E T N Y G G N L T I P L
8941 ATGATGATGGTTCTTTCCTTGTGTCTGAAACCAACTATGGGGGCAACCTAACTATACCTT

350 E K S L K Q I V P S V L L I R P N R R V
9001 TAGAAAAATCTCTCAAGCAGATAGTGCCATCAGTTTTGCTTATACGACCAAATAGAAGAG

370 Y T *
9061 TTTACACT**TGA**AATTGTAGCTATTTAAAGATACAATATGTCTATAAATGAGTGGTCGGCT

9121 CATGGTCAATCTGATAGTAATCTTGGACATAAGGGCCAAGCGGTGGCGACACCAGAACAA

9181 AATCCGATAGCAATCTTGGACGTACGGGCCAAGCGGTGGCAGACACCAGAATAAACCGAC

9241 TGATTAGGTGGCTTACAGGTTCAGTAAGTCATCTTTTTATTTGGAAAAACAGTAAGAAA

9301 GCGTAAAAAATCAATTCCTGTACTAGCGTGTAATTGAAATTTTTTAGTTTTGGCGTTAGA

9361 ACTGTTTCCTCAGTCCTAGCCCCTTCTCTTTTTGATAAGTATGCTTGACGGAAAGTAGTT

9421 TGTAAATGATTCTGAGAATCTTATGAGCATAAGCAATGACCGGCCTTCATCTTGCTTACT

9481 CTTTGGGAAATTCGATTATAAAAAAAAGAAAAAGCTGGATCC

BamHI

Analysis of the ORFs Predicted Amino Acid Sequences

Nucleotide and derived amino acid sequences from each of the open reading frames were compared against DNA and protein sequence databases using the BLAST Network Service at NCBI via University of Oklahoma and the results are shown in Table 5.

ORF20 Predicted Nucleotide Sequence

The predicted amino acid sequence from ORF20 was 13 kDa in size with a calculated isoelectric point (pI) of 9.49 (Table 5). Database searches did not reveal any significant homology to known proteins. However, weak similarity was shared with the dipeptide transporter protein dppC from alkaliphilic bacteria *Bacillus firmus*, a membrane protein that appears to be involved in complementation of K^+-uptake deficiency in *Escherichia coli* (accession number 1813497). ORF20 also showed homology to the PTB gene product, a polypyrimidine tract-binding protein involved in binding of pre-mRNA from the protein alpha-tropomyosin (50). Hydropathy profiles of the ORF20 amino acid sequence indicated a strong tendency to form transmembrane segments (TMS) between residues 8 to 29 and 64 to 82 (Fig. 6) (29).

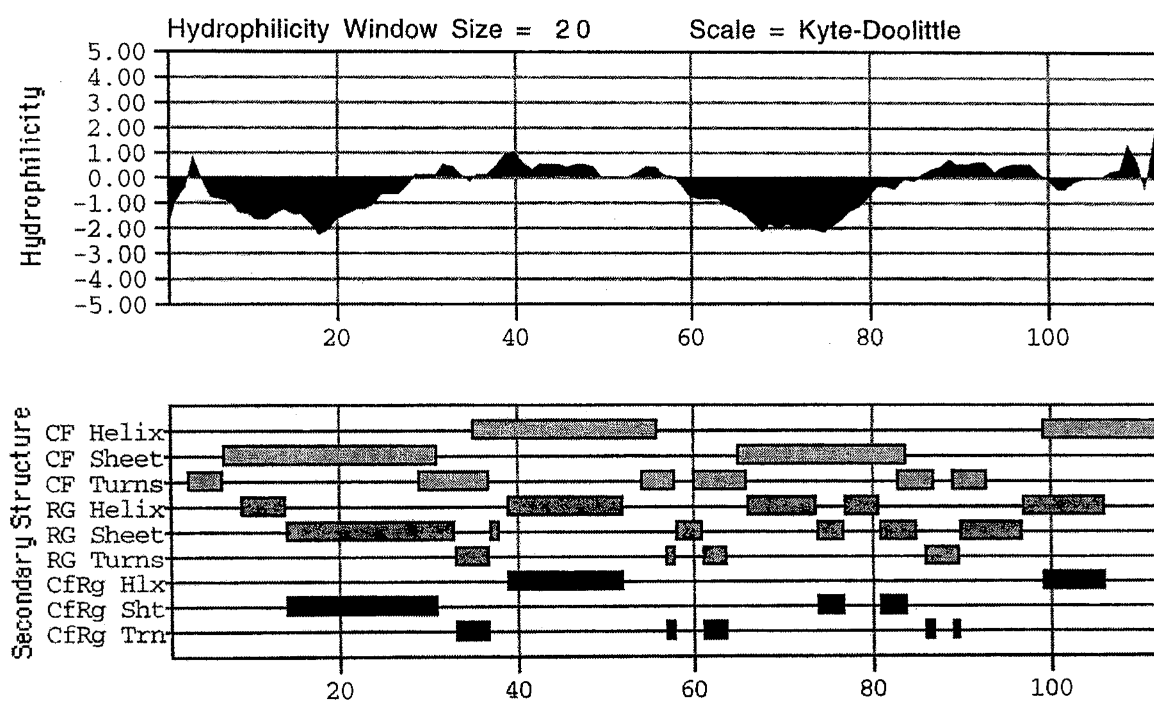


Figure 6. Hydropathy profile and secondary structure prediction of the deduced amino acid sequence from ORF20.

ORF21 Predicted Product

The ORF21 deduced amino acid sequence was found to be 54 kDa in size with a calculated isoelectric point of 9.48 (Table 5). Analysis of this protein revealed significant similarity (7.7×10^{-22}) to the TraK and TrsK gene products from the transfer regions of conjugative staphylococcal plasmids pSK41 (16) and pGO1 (46) respectively. The *tra* region in pSK41 is 14,399 nt in length and analysis of the translated sequence showed 15 genes potentially involved in the conjugative process. All the putative genes appeared to be under control of a common regulatory factor. The transfer region in pGO1 (*trs*) is 13,612 nt with 14 identified ORFS, 13 of which are transcribed in the same direction. Deletion and complementation analyses on these genes revealed that they were essential for conjugative mobilization of their respective plasmids. However, attempts to further characterize the predicted products were unsuccessful due to the inability of these genes to be transcribed and translated in *E. coli* expression systems. ORF21 also shared homology to the TraG proteins found in the conjugal transfer system of octopine- (2.0×10^{-09}) and nopaline-type (3.9×10^{-08}) Ti plasmids from *Agrobacterium tumefaciens*. The transfer regions of these plasmids have been shown to share lineage with the transfer systems of IncP (RP4) and F plasmids (1,14). In addition, the ORF21 product showed a significant level of similarity (1.1×10^{-08}) to the ORF10 protein of unknown function from *Helicobacter pylori* that is

located in the *cag* gene. This gene has been reported to be a pathogenicity island that encodes disease-associated virulence factors (7). Other proteins from transfer-related systems that presented weak homology ($3.2x10^{-03}$) were the traG products from IncP plasmids RP4 and RP751 (83). The TraG from RP4 plasmid and TraG and VirD4 proteins from Ti plasmids have been suggested to play a role in initiation of DNA transfer by interacting with both the relaxosome and the mating bridge complex (14). A similar role has been suggested for TraD, a protein found in the inner membrane of F^+ cells and an analog of RP4 TraG and Ti plasmids TraG. TraD appears to function either by transporting nucleic acids through the membrane or by promoting interaction between the relaxosome and other membrane proteins to form a membrane pore (39). This protein has also proved to be essential for the infection of RNA phages such as R17 and MS2 (19) and for efficient DNA replacement synthesis in donor cells (31).

Multiple alignment between the ORF21 predicted protein and the sequences with the highest level of homology is shown in Figure 7. The translated sequence of ORF21 was also analyzed by using the MOTIFS program from the GCG package in order to search for conserved patterns. Results from this analysis revealed a type-A ATP/GTP-binding site motif also known as the P-loop (57). The consensus sequence for the P-loop is reported in the literature as [GA]X{4}GK[ST] and appears to be part of an ATP-binding structure that is usually found in a family of exporter systems

known as bacterial ABC transporters (15). Comparisons using the pattern construction algorithm PIMA (pattern-induced multiple alignment), revealed that all sequences present at least one highly conserved nucleotide triphosphate binding motif. Figure 7 shows that the type A NTP-binding motif (P-loop) is shared by ORF21 and the Ti TraG protein (1,14). In agreement with reports from Firth *et al.* (16) and Morton *et.al.* (46), this motif appears to be less conserved in TraK and TrsK from the staphylococcal plasmids due to the presence of a tyrosine residue in place of the last conserved glycine. IncP TraG (3) also contained a less conserved pattern while the motif was absent from ORF10 of *H. pylori*. On the other hand, a type B NTP-binding site was highly conserved in all sequences considered but less conserved in ORF21 which showed 5 identical or conserved amino acids out of 8. In ABC transporters or traffic ATPases (23), binding of ATP takes place at a central loop formed by the consensus sequence that involves an interaction with a conserved lysine residue and the phosphate groups in Mg^{2+}-ATP. The Mg^{2+} molecule is then chelated by an aspartic acid residue on the second binding site known as the B site (D[ED]) (23,80). The importance of the presence of NTP-binding sites and their role in DNA transfer was assessed by Balzer *et al.* (3) by performing site-directed mutagenesis on the type A and type B motifs present in RP4 TraG. Single amino acid exchanges in each of the motifs completely inactivated DNA transmission and a mutation only in motif A resulted in a 2000-fold reduction in the transmission frequency. The

presence of conserved nucleotide binding motifs in ORF21 may be an indication of a similar function for this protein that could provide the energy for the export of DNA or other proteins involved in DNA transfer.

TABLE 5

CHARACTERISTICS OF THE PREDICTED PRODUCTS IN THE 9.5-kb DNA FRAGMENT

ORF	Predicted Product			Homologues
	Amino Acid Residues	Molecular Weight (Da)	pI	
ORF20	113	13,078	9.49	Dipeptide transporter protein from *B. firmis*
ORF21	470	54,268	9.48	TraK of pSK41, TrsK of pGO1, ORF10 of *H. pylori*, TraG of pTiC58
ORF22	65	7,062	5.28	Plasminogen activator receptor
ORF23	114	12,097	9.36	Calcium channel α-subunit, skeletal muscle
ORF24	79	8,089	9.95	Sodium channel α-subunit, skeletal muscle
ORF25a	70	8,266	10.57	Membrane bound sugar transport protein of *Synechocystis sp.*
ORF25b	69	8,227	8.89	Histidine kinase of *Dictyostelium discoideum*
ORF26	787	89,743	5.06	ORF5 of pIP501, TraE of pSK41, TrsE of pGO1

TABLE 5 (Continued)

ORF27	520	60,143	9.55	GARP product of *P. falciparum*, nuclear protein of *E. minor*, C11G6.3 product of *C. elegans*
ORF28	372	40,635	5.98	TraG of pSK41, TrsG of pGO1, Isp of *S. pyogenes*, ORF18 of pneumococcal phage Cp-1, ORF1 of *S. aureus* Newman, gene 13 protein of Bacillus phages ϕ-29, PZA, and B103

Figure 7. Multiple alignment of the predicted product from ORF21 and its homologs. Conserved amino acids are shaded in black while conserved substitutions are shaded in gray. The position of putative type-A and type-B NTP-binding sites are marked and their consensus sequence is shown above the motifs. The sequences aligned are ORF21, TraK and TrsK from staphylococcal conjugative plasmids pSK41 and pGO1 respectively, TraG from pTiC58, and ORF10 from *H. pylori.*

```
ORF21          1 ------------------------------------------------------------
TraK           1 ------------------------------------------------------------
TrsK           1 ------------------------------------------------------------
TraG pTiC58    1 ------------------------------------------------------------
ORF10          1 MEDFLYNTLYFIEDYKLVVIFSFIGLIALFFLYKFIKAQKKAFKDKANQPQKKKSFKEII

ORF21          1 ------------------------------------------------------------
TraK           1 ------------------------------------------------------------
TrsK           1 ------------------------------------------------------------
TraG pTiC58    1 ---------------MTMNRLLLLILPAIIMLAAMFATSGMEQRLAAFGTSPQAKLMLGR
ORF10         61 IDGLKERVKTFGFWLQAILLLSYSFITSGLFFLILLGNFYDDNRSPESDDDLFDIWVYAI
                                                                        . .

ORF21          1 ------------------------------------------------------------
TraK           1 ------------------------------------------------------------
TrsK           1 ------------------------------------------------------------
TraG pTiC58   46 AGLALPYIAAVAIGIIGLFAANGSANIKAAGL-SVLAGSGVVI---------TIATLREV
ORF10        121 QDFPNYYFKALGFSSLKIYCFNISLVVYSSILCSYIFITFFVWFLKYLTRTRDLCANLKV
                       .  . .    . ... . .  .     . . .      .            ..  . .

ORF21          1 ------------------------------------------------------------
TraK           1 ------------------------------------------------------------
TrsK           1 ------------------------------------------------------------
TraG pTiC58   96 IRLNSIASSVPAEQSVLA-YADPVTMIGA--SIAFISGMFAL--RVAIKGNAAFATTAPK
ORF10        181 DDLFGSASWETEEKMIKAKLITPNNKKRAFDKREVIVGRRGLGDFIAYAGLAFIGLIAPT
                   .    .          . .      .     .          .  ..    ..  ...  .  ..

ORF21          1 ------------------------------------------------------------
TraK           1 --------------------------------------------------------MTTL
TrsK           1 --------------------------------------------------------MTTL
TraG pTiC58  151 RIGGKRAVHGEADWMKIQEAAKLF-P-----ERGGIIIGERYRVDRDSVAAMPFRADDRG
ORF10        241 R-SGKGVGFIMPNMINYPQNIVVFDPKADTMETCGKIREKRENQKVFIYEPFSLKTHRFN
                 .  ..          .         .. .     .  .  .    ..              .    .

                     Type-A NTP-binding Site     [GA]XXGXGK[TS]

ORF21          1 -------------------------------VIGGSGSGKTFRFVKPNLIQMNSSNIVV-
TraK           5 LGELESKATSINKKNLVIQDFDTKCPNRNIFVVGGPGSYKSAGYVIPNVIVKNQQSIVVT
TrsK           5 LGELESKATSINKKNLVIQDFDTKCPNRNIFVVGGPGSYKSAGYVIPNVIVKNQQSIVVT
TraG pTiC58  205 SWG---AG-----GKSPLLCFDGSFGSSHGIVFAGSGGFKTTSVTIPTA-LKWGGGLVV-
ORF10        300 PEAYVDFGNDVVLTEDILSQIDTRL-KGHGMV-ASGGDFSTQIFGLAK---------IVF-
                  ..    .  .     ...  ....     ...*....*....  ..........    .*.

ORF21         29 DPKDHLAEKTG--KLFLEHGY-QVNVLDLV----NMKNSD--GFNPFRYIETENDINRML
TraK          65 DPSGEVYEKTS--NIKRMQGF-DVRVVNFK----NFLASD--RQNFFDYIDKDSDC-SIV
TrsK          65 DPSGEVYEKTS--NIKRMQGF-DVRVVNFK----NFLASD--RQNFFDYIDKDSDC-SIV
Trag pTiC58  255 --LDPSSEVAPMVCEHRRQAGRKVIVLDPTAGGVGFNALDWIGRHGNT---KEEDIVAVA
ORF10        349 --PERPNEKDPF---FSNQA-RNLFVINCN----IYRDLMW-TKKGLEF-VKRKKIIMPE
                 .. . . *...   ... .... ..*.      .. . .. ........... ... ..
```

```
ORF21          80 AVYFNNTKGSGSRSDPFWDEASMTLVRALASYLVDFYNPPKTREQLIEESRLSQKEYQNL
TraK          115 AELIIKSAGDSKINKDVWYKASVGLLNSL--LLYAKYEFEPEKRTIGNIIKFIQNNKPN-
TrsK          115 AELIIKSAGDSKINKDVWYKASVGLLNSL--LLYAKYEFEPEKRTIGNIIKFIQNNKPN-
TraG pTiC58   310 TWIMTDNPRTASARDDFFRASAMQLLTAL---IA---------LVCLSGHTETEDQTLR-
ORF10         397 TPTMFFIGSMASGINLIDEDTNMEKVVSL---M-----------EFFGGE---EDKSGD-

ORF21         140 LKRQK-KEVEERKKRGRLSKFCESQNSLN---TYPRVKTRKSVLEILFENYAKKYGTENF
TraK          172 -QDEE-GSVELDKRFNELSKDHPARESYEFGFAVSEGRTRASII----------------
TrsK          172 -QDEE-GSVELDKRFNELSKDHPARESYEFGFAVSEGRTRASII----------------
TraG pTiC58   357 -RVRANLSEPEPKLRARLTKIYEESES-------DFVKENVSVF----------------
ORF10         439 -NLRA-LSPATRNMWNNFKTMGGAKET---------YSSVQGVY----------------

ORF21         196 TM----RNWAD-FQNYKDKTI-DSVIAVTTAKFALFNIQSVMDLT--KRDTLDMKTWGQE
TraK          214 -L----TFVAD-LRNYIDNEI-RS------------------YT--SDSDFDLRDVGLR
TrsK          214 -L----TFVAD-LRNYIDNEI-RS------------------YT--SDSDFDLRDVGLR
TraG pTiC58   393 -VNMTPETFSGVYANAVKETHWLSY-------------PNYAGL--VSGDSFSTDDLADG
ORF10         472 -----TSAFAP-YNNAMIRNF--TS-------------ANDFDFRRLRIDEVSIGVIANP

                                     Type-B NTP-binding Site        D[ED]

ORF21         248 KSMVYLVIPDNDS-TFRFLSALIF---FNPYFQTPNKTSQILMLRVRLPLHVRVYLDEFA
TraK          246 KTIIYVMLPVLGN-TVQSLSSLFFSQMFCQLYRLGDENG------ARLPVPVDFLLDEWP
TrsK          246 KTIIYVMLPVLGN-TVQSLSSLFFSQMFCQLYRLGDENG------ARLPVPVDFLLDEWP
TraG pTiC58   437 GTDIFIALDLKVLEAHPGLARVVIGSLLNAIYN-RN---------GNVKGKTLFLLDEVA
ORF10         511 KESTIVGPILELF-----FNVMIYSNLILPIH---D---------PQCKRSCLMLMDEFT

ORF21         304 NIGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQG--LYKE-KEAWKTILGNCDSLVYLGG
TraK          299 NIGAIPDYEETLATCRKYGISITTIVQSISQAVI--KYNK-DKA-NAIIGNHAVILCLGN
TrsK          299 NIGAIPDYEETLATCRKYGISITTIVQSISQAVI--KYNK-DKA-NAIIGNHAVILCLGN
TraG pTiC58   487 RLGYLRILETARDAGRKYGIMLTMIFQSLGQMRE--AYG-GRDATSKWFESASWISFAAI
ORF10         554 LCGYIETEVKAVGIMAEYNMRPAFVFQSKAQLENDPPLGYGRNSAKTIIDNLSINMYYGI

ORF21         361 NDEDTFKF---MSGLLGK-QTIDVRNTSR--SFGQTGSGSLSHQKI---ARDLMTPDEVG
TraK          355 VNNDTAKY---IKEELGN-ATVEFETSSEGSSSGKDSRSKSSNKNVSYTGRALINEDEIS
TrsK          355 VNNDTAKY---IKEELGN-ATVEFETSSEGSSSGKDSRSKSSNKNVSYTGRALINEDEIS
TraG pTiC58   544 NDPDTADY---ISKRCGD-TTVEVDQTNR--STGMKGSSRSRSKQL--SRRPLILPHEVL
ORF10         614 NNDNYYEHFEKISKVLGKYTRQDVSRSID--DNTGKTNTSISNKE-----RFLMTPDEIM

ORF21         412 NMKRHE-CLVRIENMPVFKSKKYNS--TK-HPNW--KYLANQEIDERFWELSNQSFKSKT
TraK          411 NMQQDKAILITKNRNPKIIKKLAYF--KM-FPNITETYIAPQNEYKRK---KNQSMIDKH
TrsK          411 NMQQDKAILITKNRNPKIIKKLAYF--KM-FPNITETYIAPQNEYKRK---KNQSMIDKH
TraG pTiC58   596 HMRSDEQIVFTSSNPPLRCGRAIVFRRDDMKACVGENRFHRTGTGTDTHTEHAPPWQKEG
ORF10         667 TM-GDELIIIENTLKPIKCHKALYY--DD--PFFTIELIKVSPSLSKKYKLGKVPDQTTF
```

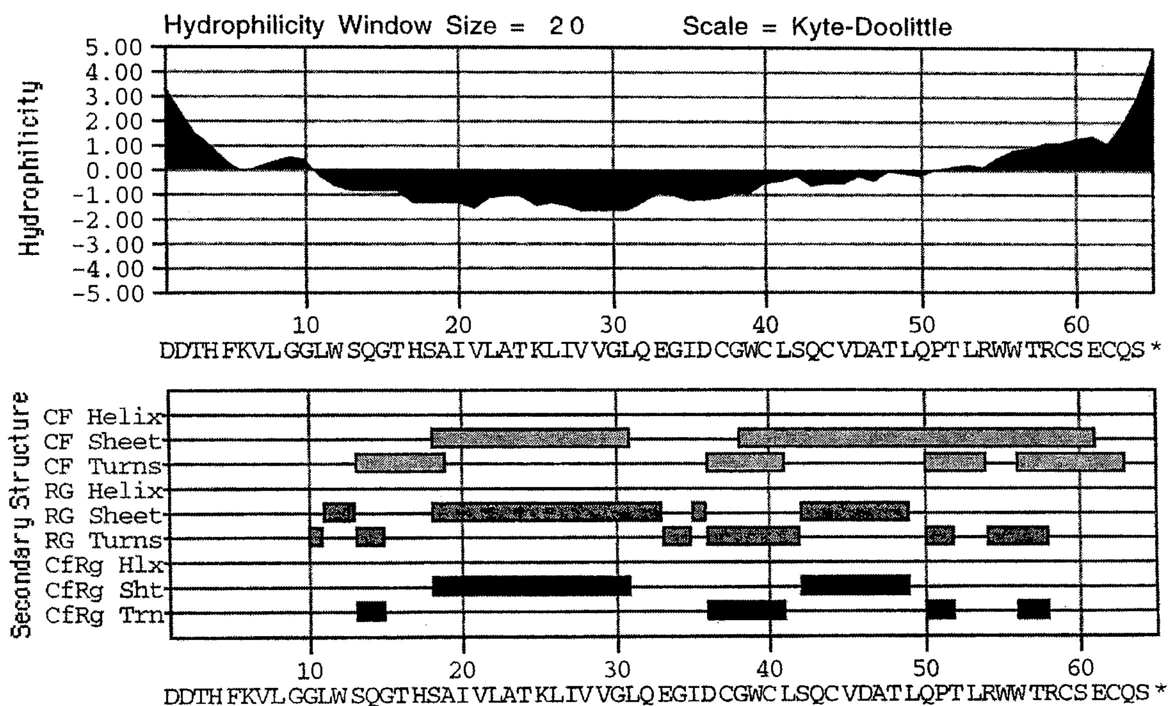
```
ORF21         466 RKSS--------------------------------------------------------
TraK          465 NRLREEWDKKQEKNKQQKLEEYKEEQRQKELEKEQQPQEEQQNEKVKESKDKNDQEDKKD
TrsK          465 NRLREEWDKKQEKNKQQKLEEYKEEQRQKELEKEQQPQEEQQNEKVKESKDKNDQEDKKD
TraG pTiC58   656 TRP---------------------------------------------------------
ORF10         722 YDDLQAAKTKGELSYDKSLVPVGSSEL---------------------------------
                    .       . .      .

ORF21             ----------------------
TraK          525 EQQKINDSKKAFYEKLKQKQAK
TrsK          525 EQQKINDSKKAFYEKLKQKQAK
TraG pTiC58       ----------------------
ORF10             ----------------------
```

Deduced Products from ORF22 to ORF25b

Predicted products from ORF22 to ORF25b ranged in size from 7.0 to 12 kDa and most of them were calculated to have a basic isoelectric point, the only exception being ORF22 which had an isoelectric point of 5.28 (Table 5). The hydropathy profiles of their amino acid sequences are shown in Figures 8, 9, and 10. The high content of hydrophobic residues observed in all of them indicated membrane association. Comparisons of their deduced sequences with protein databases failed to detect any significant homology to other known proteins. However, all ORFs shared weak similarity with a number of membrane-related proteins (Table 5). The 7.0 kDa protein from ORF22 showed similarity to a kinase receptor found in rats (accession number X71899). The amino acid sequence from ORF23 (12.0 kDa) presented some degree of relatedness to the æ-1 subunit protein characteristic of L-type Ca^{2+} channel transcripts in human fibroblasts (70). The product from ORF24 also revealed some homology to the -subunit protein of the sodium channel from mammalian skeletal muscle (75). Similarly, the sequence from ORF25a shared significant homology ($2.1x10^{-4}$) with a membrane-bound protein from *Synechocystis sp.* that is involved in sugar transport (32). Finally, ORF25b presented some level of similarity to a histidine kinase from *Dictyostelium discoideum* (58).

This particular region has proved to be essential for conjugal transfer of Tn*5252* (34), and deletions in this region could result in a defective transport mechanism. The membrane nature of these homologs resembles that of the ORFs predicted sequences as judged by their hydropathy profiles (Figures 8, 9, and 10), and suggests an important membrane interaction for this region. Furthermore, this segment has been previously shown to be essential for transposition when a deletion of a 1,749 nt *Bgl*II fragment (coordinates 2762 to 4511 in Figure 3) impaired conjugal transfer of Tn*5252* (34).

ORF22



ORF23

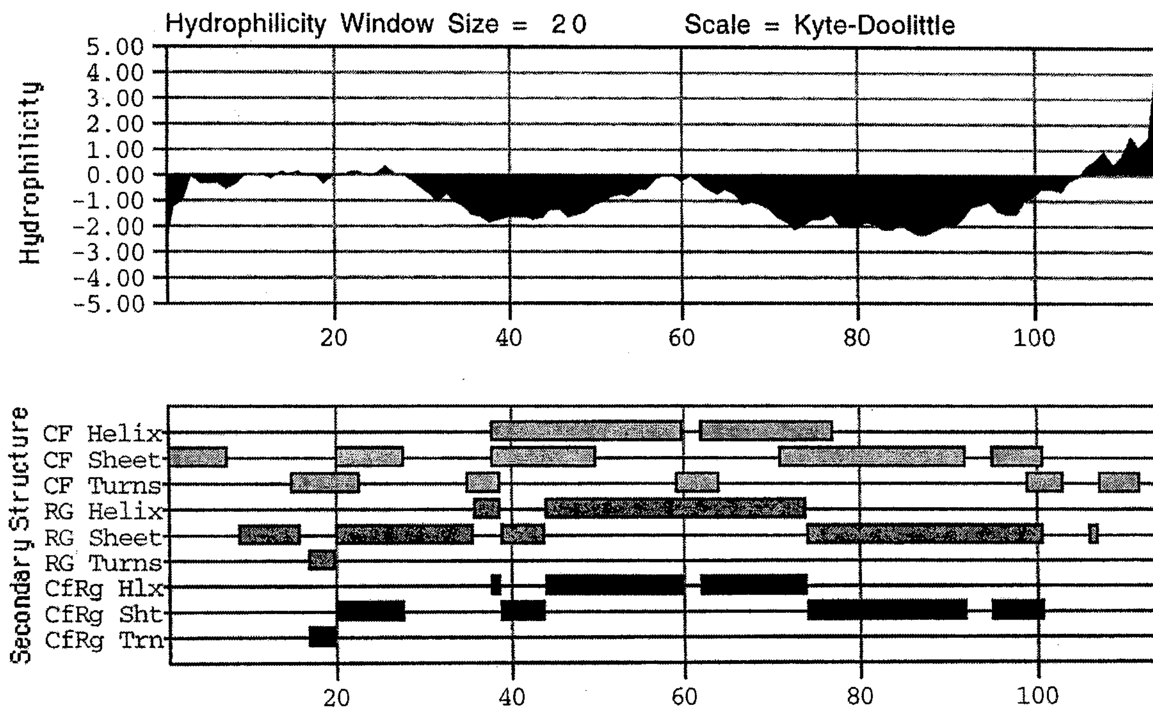
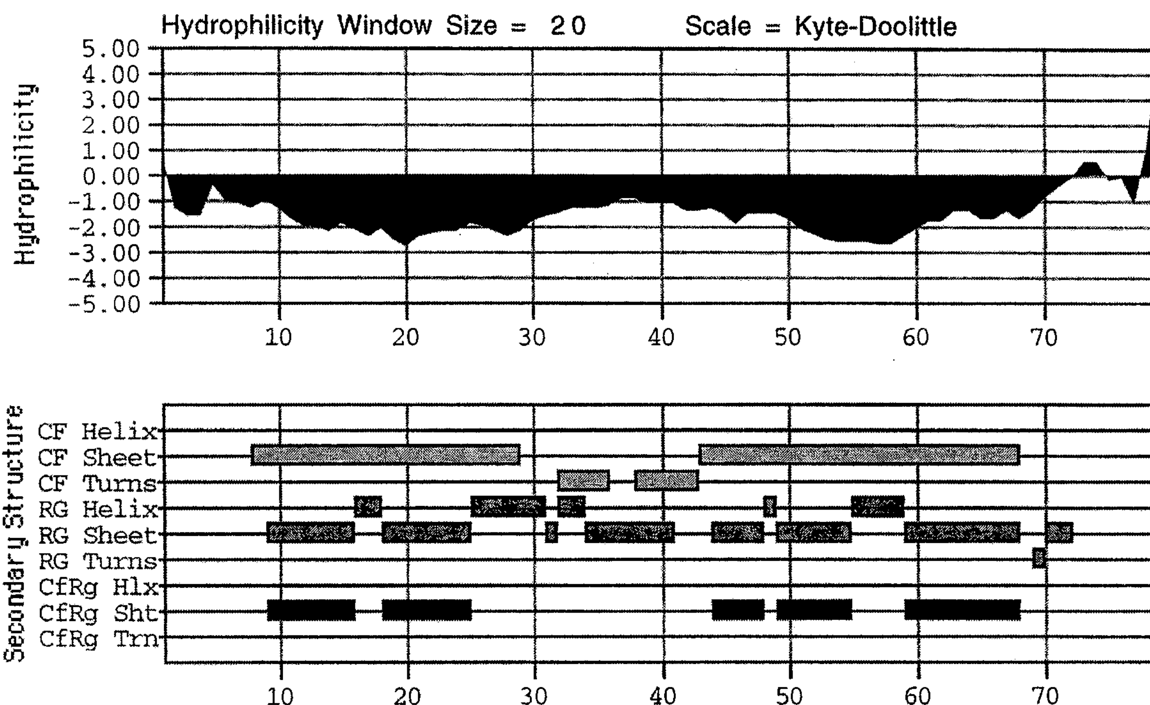


Figure 8. Hydropathy profile and secondary structure prediction of the deduced amino acid sequences from ORF22 and ORF23.

ORF24



ORF25a

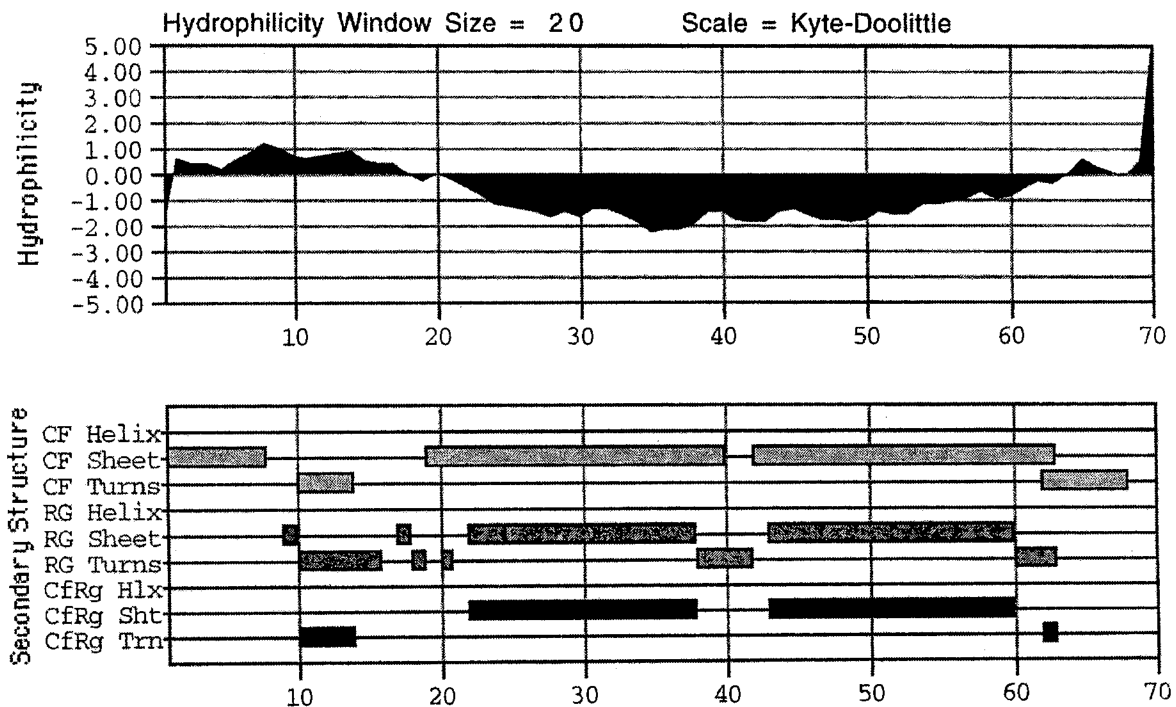


Figure 9. Hydropathy profile and secondary structure prediction of the deduced amino acid sequences from ORF24 and ORF25a.

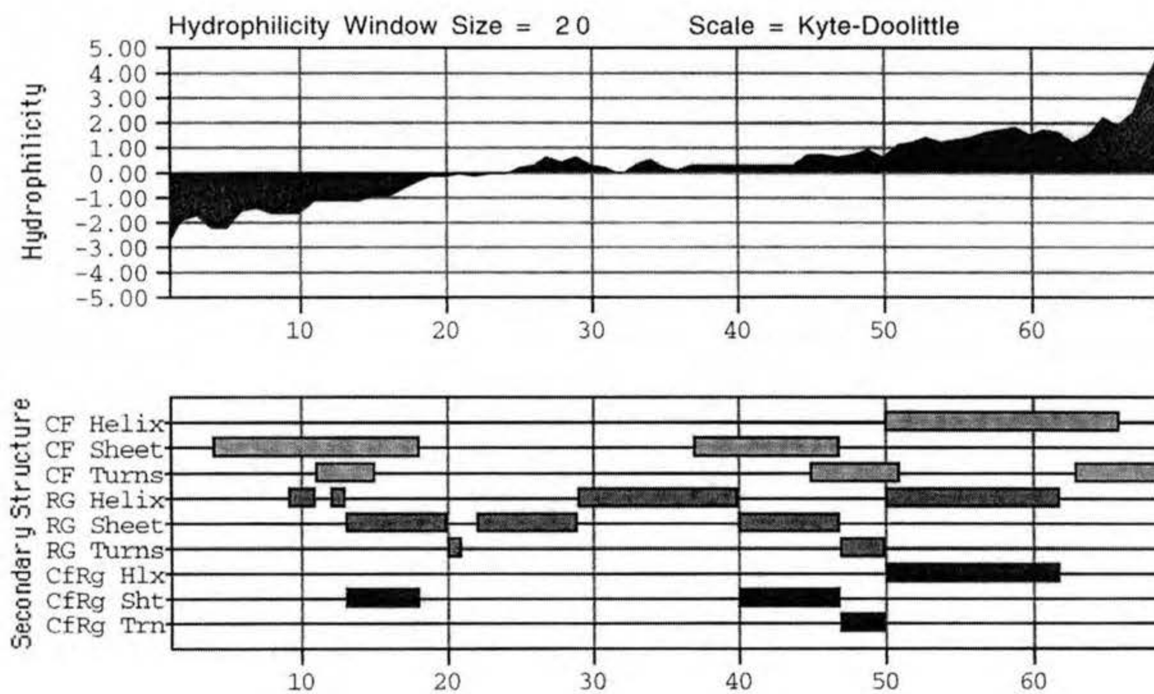


Figure 10. Hydropathy profile and secondary structure prediction of the deduced amino acid sequence from ORF25b.

ORF26 Predicted Product

The ORF26 translated sequence also showed an acidic isoelectric point and a molecular weight of 89 kDa (Table 5). Computer search of protein databases revealed similarity ($5.8x10^{-10}$) to the ORF5 predicted product from the transfer region A of pIP501, a conjugative plasmid from *Streptococcus agalactiae* (81). ORF5 was shown to be essential for pIP501 mobilization but its specific function in the transfer process is still under study. The ORF26 amino acid sequence was also similar to the transfer-related genes *traE* ($5.8x10^{-10}$) and *trsE* ($1.9x10^{-09}$) from staphylococcal conjugative plasmids pSK41 and pGO1 respectively (16,46). As with TraK and TrsK proteins, attempts to identify the TraE and TrsE products were unsuccessful due to their hydrophobic nature. Database searches on their predicted products detected a type-A ATP/GTP binding motif and an ATPase activity was suggested. Multiple alignment between ORF26 and the homologue proteins (Fig. 11) shows that a less conserved NTP-binding motif is also present in ORF26 with 5 identical or conserved residues out of 8. Another important feature found in this sequence was a motif that is characteristic of a family of prokaryotic transcriptional regulators known as LysR (28). The consensus pattern for this motif is [LIVMFYT]-x{2}-[STGLAV]-[STA]-x{5}-[PSTA]-[PNQHKR]-x{2}-[LIVMA]-[STA]-x{2}-[LIVMFW]-x{2}-[LIVMFW]-[RKEQA]-x{2}-[LIVMFYNT] and the location is shown in Figure 8.

Members of the LysR family bind DNA through a helix-turn-helix motif located in their N-terminal domain and they also appear to negatively regulate their own expression (28). An example of these proteins is the *iciA* gene that controls the initiation of chromosomal replication in *E. coli* (73). The IciA product blocks initiation by binding to the 13-mer sequences that are opened by the DnaA protein during replication. Other activator proteins of the same family are known to regulate biosynthetic pathways (28).

Figure 11. Multiple sequence alignment of the predicted amino acid sequence from ORF26 and its homologs. Conserved amino acids are shaded in black while conserved substitutions are shaded in gray. The position of a putative type-A NTP-binding site and a bacterial regulator motif are marked and their consensus sequence is shown above the motifs. Sequences shown are ORF26, TraE and TrsE from staphylococcal conjugative plasmids pSK41 and pGO1 respectively, and ORF5 from conjugative streptococcal plasmid pIP501.

```
ORF26     1 LTIFNKRLNLEKFRHSVLYEEKEDGYDSYRKELNRMMNQNLDSGENNFSAVKLISFGRKD
TraE      1 ------------------------------------------------------------
TrsE      1 ------------------------------------------------------------
ORF5      1 ------------------------------------------------------------

              Family of Bacterial Regulator Proteins  lysR      YXXLTX

ORF26    61 SNPKQAYRSLSQIGEYFKSGFSEIDARFESLAGEERVNLLADMLRGEHHLPFSYRDLTR-
TraE      1 -------------------------------MIMAFLKKKKQEQVTNKIHNESFQKLTEV
TrsE      1 -------------------------------MIMAFLKKKKQEQVTNKIHNESFQKLTEV
ORF5      1 ------------------------------------------MEKIPKEKIVLIPEV
                                            ............... ............

             XXXXTRXXIAXXLXXFKXXN

ORF26   120 -SGQITRHFIAPNLLDFKNKNYLCINDRLLQIVYVRDYGMELGDQFIRDLMQGDLELIVS
TraE     30 LDTDSVDSIYPFSW--IEKKSHIETGENYIKNLLVVDYPQSVKGAYLSNLLKKNGNIQIT
TrsE     30 LDTDSVDSIYPFSW--IEKKSHIETGENYIKNLLVVDYPQSVKGAYLSNLLKKNGNIQIT
ORF5     16 -DTDVVSDLAPFNE--TVERDKLLIDESYAVPYVITKYNNKPRGNWFNRIRKMSGDITIS
            ............ .  ........ ............*......................

ORF26   179 LHAQSSTKSDAMKKLRTKKTLMESQKIGEQQKLARTGIYLEKVGHVLESNIDEAEELLKT
TraE     88 KFIRPANIERMIDHLNNS----IKNKTAEQMR-TTDPKRNAT----IKREIESSKKQLDK
TrsE     88 KFIRPANIERMIDHLNNS----IKNKTAEQMR-TTDPKRNAT----IKREIESSKKQLDK
ORF5     73 HYYTKANGNSLNDYYNRT----IKNKQAEIDR-SHDPLTIIR----LEREMKIAQTQLEQ
            ..................    ...*..*... .........    . ...... ...*..

ORF26   239 MTETGDKLFQTVFLIGVFGQDEEELKQALDTVQQVAGSNDLMIDKLPYMQEAAFNSLLPF
TraE    139 FLDEKTGFMYMYMYITLNGDSYEKIQALEKDVKRTLTRLRLKTHTPTNAMRESFHTVLPL
TrsE    139 FLDEKTGFMYMYMYITLNGDSYEKIQALEKDVKRTLTRLRLKTHTPTNAMRESFHTVLPL
ORF5    124 AVDENTSYLYEYTYVLIKSKSEDKLKKLCEDFETRCIASGVKALIPYTMIDKAYWSSLPL
            ...................... ... .......................... ... ....**.

ORF26   299 GCDFLEGVSRSLITSNIAVN-SPWTSVDLQ--DRSGKYYGINQISSNIITID---RSLLN
TraE    199 NRNFLSAFTQQNMDIATAGHFFMFDDSEIIDLTPNTSVFGINKNTDSLVAVDFNNKEKTL
TrsE    199 NRNFLSAFTQQNMDIATAGHFFMFDDSEIIDLTPNTSVFGINKNTDSLVAVDFNNKEKTL
ORF5    184 ESNEVPEYTYTIANSISASSIFPFDDNELSVFTKNMIIEGINKDTENIVSIDYTNRKLVV
            ...................*.... ................***.... ....*..... ..

               GXXXXXGK[ST]   Type-A NTP-binding Site

ORF26   353 TPSGLILGTSGAGKGMATKHEIITT---KIKESGENTEIIIVDPEAEYSVIGRTFGGEMI
TraE    259 NKNMTIIGTSGVGKSTLNMRMILDNI-----KKSIRQFII--DPEDEFSYITKYYGGTVV
TrsE    259 NKNMTIIGTSGVGKSTLNMRMILDNI-----KKSIRQFII--DPEDEFSYITKYYGGTVV
ORF5    244 NRNKFVFGLSGGGKITYLTSDYLKKYAFSDNSTELRHRIVLFDPEDEQVERVRSLGGEII
            .......*.**.**............    .......*.. ***.*.........** ..

ORF26   410 DIAPDSETYLNVLDL-SEENMD----EDPVKVKSEFL-----------------------
TraE    312 NISTSSNIKINPFEIFSEEVFEKE-DETDNETTSDVLTENNE--------HESQIDTLVR
TrsE    312 NISTSSNIKINPFEIFSEEVFEKE-DETDNETTSDVLTENNE--------HESQIDTLVR
ORF5    304 NLSSMSDVRINPFQIYSRNTLEVELKESLSDFEEDELVENIEIKHKDYEMTENDIDKEIS
            .....*....*.....*........ .*.........*.....        ..........
```

```
ORF26   442 -----LSFIGKLLDRKMDGREKSIIDRVTRLTYQSFKE--------------PSLEEWVF
TraE    363 SKIGKLKTFFRVLKDEISQTEISVLSSTLRQLYQDKGFKGNAKLSDFKSEDYPTLTELYN
TrsE    363 SKIGKLKTFFRVLKDEISQTEISVLSSTLRQLYQDKGFKGNAKLSDFKSEDYPTLTELYN
ORF5    364 KRMNILTPYFLMVDHSLTDSQLSIIKIEAKKCYTTLYEKKN--LSKMENTDFPTFSDLEN
            .....*........ .........*.........*..... ...............*.......

ORF26   483 VLS----QQPE--EEAQNLALDMELYV--EGSLDIFSHKTNIQTGSNFLIYNVKKLGDE-
TraE    423 KLKDL---DPEKYEVLKDLTLIIEDYTMEHGTTTIFDGYTNIKLDNEIVTFNLKPLQTEK
TrsE    423 KLKDL---DPEKYEVLKDLTLIIEDYTMEHGTTTIFDGYTNIKLDNEIVTFNLKPLQTEK
ORF5    422 RLKALEETDPKRYKRIEDFIYSLEDFTI--GSRTIFNGHTNIDLNNPLICFSLRDLQTEE
            .*...   .*..............*.......*...**...***..............*..*.

ORF26   534 -LKQIALMVVFDQIWNRRVVRNQKLGKKTWIYFDEIELLLLDKYPSDFFFKLWSRVRKYGA
TraE    480 DVQSAAYLNIFSFLWDEITKDRT--TETVLMTDELHFLATNEYSLDFYYQAYKRIRKYGG
TrsE    480 DVQSAAYLNIFSFLWDEITKDRT--TETVLMTDELHFLATNEYSLDFYYQAYKRIRKYGG
ORF5    480 GIRDLAYLNSFSYLFEEITNNPQ--IVTSVYADEFHFLLKNKISADFFFQAYKRFRKYNA
            .....*....*......... ..  ..*.. .**...* .. ...**.......*.***..

ORF26   593 SPTGITQNVETLLLDPN--GRRIIANS--EFMILLKQAKNDREELVQLLGLSKELEKYLV
TraE    538 GAIASTQQIKDILRTSQEIGSAIIENSHTKFFFGMDNVGVDDVVDKLGLKFSDQEISHLT
TrsE    538 GAIASTQQIKDILRTSQEIGSAIIENSHTKFFFGMDNVGVDDVVDKLGLKFSDQEISHLT
ORF5    538 DCTVSTQQIDDVLKAPDNIGKAIIGNSFTKVFFGLDETEAQGISNELKLKLIKKELSFIT
            .. ..**.....*.. ...*..**.**.....................*. . .......

ORF26   649 NPEKGAGLLKAGSVVVPFKNKL-PKGSQLEEIMRSDPDKMASN-----------------
TraE    598 KKKKGEALLLYGTQRAFIRIDLDREETRLWNKELYETIYEEPADVEPNYVEQLGLTDIDL
TrsE    598 KKKKGEALLLYGTQRAFIRIDLDREETRLWNKELYETIYEEPADVEPNYVEQLGLTDIDL
ORF5    598 SKRQGEALLFHGTKRAKIKVDLTQEEMRLLNPGEYEDIYGVSPKKEINWILRSKIQ----
            ....*..*...*...............*........... .................

ORF26       ---------------
TraE    658 AELEEELRQAEMIYE
TrsE    658 AELEEELRQAEMIYE
ORF5        ---------------
            ...............
```

Analysis of the ORF27 Predicted Product

The ORF27 amino acid sequence was 60 kDa in size with an isoelectric point of 9.55 (Table 5). This predicted product showed significant homology (1.9×10^{-5}) to a *Plasmodium falciparum* antigen known as glutamic acid-rich protein (GARP) (74), and weak homology (0.00045) to nuclear proteins from the bivalve mollusc *Ensis minor* (4). The GARP protein belongs to a group of antigens that are known to be associated with the host membrane. Nuclear proteins from *E. minor* are, on the other hand, small basic proteins found in the mollusc sperm. Multiple alignment in Figure 12 shows that the region of ORF27 that shares the highest homology has a high lysine content. Hydropathy analysis of the deduced polypeptide revealed a potential transmembrane segment (TMS) between residues 286 to 305 (Fig. 13) (29). Furthermore, a secondary structure analysis performed on the amino acid sequence according to Chou and Fasman (8), showed a predicted tendency to form α-helices (Fig. 13). These results indicate that the homology observed is possibly structural rather than functional since the GARP and nuclear proteins are not related to any DNA-transfer systems. Characterization of this ORF was one of the main goals of the present study since an insertional mutation directed at an *Xba*I site at its carboxyl-terminal (Fig. 4) was shown to abolish transfer of Tn*5252* (34).

Homologies of the ORF28 Predicted Product

The deduced amino acid sequence of ORF28 had a molecular weight of 40 kDa and an isoelectric point of 5.98 (Table 5). Hydropathy profiles of the predicted product indicated the presence of a stretch of 13 hydrophobic amino acids at its amino-terminus that resembles a potential signal peptide (Fig. 13). In addition, the sequence also showed predicted tendency to form TMS between residues 92 to 107 and 303 to 319 (29).

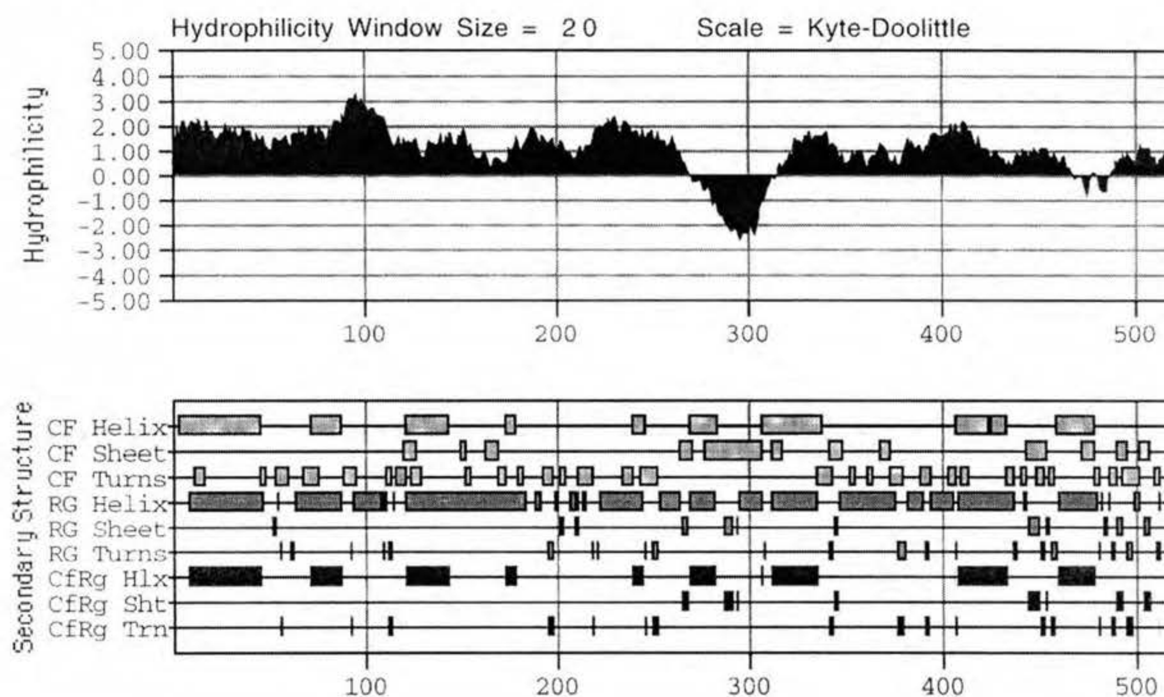
The N-terminal domain of ORF28 showed a high level of homology with several morphogenetic proteins from bacteriophages (Fig. 14). High similarity (9.1×10^{-10}) was shared with the ORF18 product from pneumococcal lytic bacteriophage Cp-1 (44) and the gene 13 protein from the closely related *Bacillus* phages ϕ-29 (2.7×10^{-08}) (76), PZA (1.2×10^{-05}) (47), and B103 (8.0×10^{-06}) (Accession number X99260).

Figure 12. Multiple sequence alignment of the predicted amino acid sequence from ORF27 and its homologs. Conserved amino acids are shaded in black while conserved substitutions are shaded in gray. Sequences shown are ORF27, GARP protein from *P. falciparum*, nuclear protein from *E. minor*, and C11G6.3 protein from *C. elegans*.

```
ORF27      1 ------------------------------------------------------------
GARP       1 MNVLFLSYNICILFFVVCTLNFSTKCFSNGLLKNQNILNKSFDSITGRLINEIELEKNKD
NucProt    1 -------------------------------------------------MAGSKSRSRSR
C11G6.3    1 ------------------------------------------------------------
                                                              .  .

ORF27      1 ------------------------------------------------------------
GARP      61 DNSKSETLLKEEKDEKDDVPTTSNDNLKNAHNNNEISSSTDPTNIINVNDKDNENSVDKK
NucProt   12 SRSKSKSPAKSASPKKAASPKAS----KSPKRSKS-PKAGKSRKRSRSRSKSSKRSASKK
C11G6.3    1 ------------------------------------------------------------
                ... .  .     .   .   .    .  .                 .    . ..

ORF27      1 --------------------------------------------------LKQGKKEVRK
GARP     121 KDKKE-KKHKKDKKEKKEKKDKEKKDKK--EKKHKKEKKHKKDKKKKENSEVMSLYKTGQ
NucProt   67 RRSRSPKKKSKSKKRSASKKRKSKPRKKSASKKRSASKKRKSRKPSASKKRKSASRKRST
C11G6.3    1 --MFLFKLEFFCQYSTPVLKLKIKFNPNLQTPKAEEPEPMRPDTTRSSGSRPASSKSNYH
             .  .  ... . .. .  .. . .  ..  ....  .... .... . ... ....  .

ORF27     11 QKKDSAGLDEKAWKKEIKEKLEEMREASKA-RVKQANEDYNHILQNSPPSLLNRKELRDR
GARP     178 HKPKNAT--EHGEENLDEEMVSEINNNAAGGLLLSSPYQYREQGGCGIISSVHETSNDTK
NucProt  127 SKKKSASK-KRSKSPKRSASRNKSNNNTNNSAKKRS--RSRKRSASKKRSRSRKPSASKK
C11G6.3   59 NEPPPPP----ANPPPLKFPFKNLFNLGDQ--------DVKKEEPSSMTPESSRPGSSLE
             .. ....  ...  . ..... ..... .   .. .  ....  .... . ..... . .

ORF27     70 RLPHARKRLKIAKKQFKEG-----SKGRSKRPKKESPKERKPIKNFSTGQESKHKSNFFF
GARP     236 DNDKENISEDKKEDHQQEEMLKTLDKKEKKQKEKEMKEQEKIEKKKKKQEEKEKKKQEKE
NucProt  184 RSHSRKRSASKKRSHSPKR--SASKKRSHSRKRSASKKRSKSRKRSASKKRSKSRKRSAS
C11G6.3  107 TPSSSSSSKHHHHHHKKE------RKDKEHKKHKKDKEHRDREKEKERDERKEREKQQKE
             .  . ... ...... ..      .*........... ... .*..  .   .... ..

ORF27    125 QGKS--LEELKAKKEVKAAKENLKSTKQVYKSKKVSKKAKTFLYVLGREGGELASENE--
GARP     296 RKKQEKKERKQKEKEMKKQKKIEKERKKKEEKEKKKKKHDKENEETMQQPEQESEETNNE
NucProt  242 KKRSKSRKRSASKKRSKSRKRSASKKRSKSRKRSASKKRSKSRKRSASKKRSKSRKRS--
C11G6.3  161 KEREEAARREIEEKAEMDAKRVAEEEEEERRKEKEKRREEKKKQKELLKEKERSERKEK--
             ..... ...    *... .*....... . .......... .... .... ...

ORF27    181 ----------DLEGYHTLQETIPKGKRYSRLSYNLGKASVKTGQATGRFTKKRLTNTKER
GARP     356 IMVPLPSPLTDVTTPEEHKEGEHKEEEHKEG--EHKEGEHKEEEHKEEEHKKEEHKSKEH
NucProt  300 ---------ASKKRSKSRKRSASKKRSHSR---KRRPSKKRSDSRKRRASKKRSHSRKRS
C11G6.3  219 ----------ERELEREKEKSREKEREKERE--KEREKEREKEREKQKEREKEREKEREK
                       . .  ...... .*......   .......... .. .. .*  ......

ORF27    231 YHHF-------KDGKGWK-LAKDN--PSSFKNR-FRKLKKQG--------LTSVRNIYQK
GARP     414 KSKGKKDKGKKEKGKHKK--AKKEKVKKHVVKNVIEDEDKDGVEIINIEDKEACEEQHIT
NucProt  348 ASK------KRSHSPKRR--ASRSGVVTKVMNAIAHCKSSKGCSAQAIRKYLAAHSKLTG
C11G6.3  267 EKK------REEEARRKKEEASTPVIIRPLLS--QEDDDSNG-------SESSEEIWICP
              ..      ... .....   * .  .. .... . .... .*      .   .  .  .
```

```
ORF27     272 LK-AAFSFFTFAAG---NLVTWIVGGIVFLLLLIMSFFLGFSSASLIQQDEFELTKAYTH
GARP      472 V-----ESRPLSQP-QCKLIDEPEQLTLMDKSKVEEKNLSIQEQLIGTIGRVNVVPRRDN
NucProt   400 VFLNFHVRRALAAGMKNHLLAHPKGSNNFILAK---KKAPRRRRRVAK--KVKKAPKRAK
C11G6.3   312 V-----CSVAYTVGA--NMIGCDQCQDWFHWHCVGLTAEPTDSKWFCT--RCTKGNKSKK
              .     .......   .... . .  .... ..  . ..  .. ...  .. ..... .

ORF27     328 LTWEDAEHTRTNDKGITYYTKVDDVMGYMNFKFHDYELHKPVHLFSSESYKDYLSTLWHD
GARP      526 HKKKMAKIEEAELQKQKHVDKEEDKKEESKEVQEESKEVQEDEEEVEEDEEEEEEEEEEE
NucProt   455 KTKKVAKRKSKKAKKSKKARKARSHKKAAAPKRRATRRRASSRRRRATKAH---------
C11G6.3   363 HGKRSATGPHDRDPSAKKKKKSH-------------------------------------
              .....*.. . .... .. .*  ...      . .... .   .    .  ..     .

ORF27     388 LNEGDDLKSMQDLYETPKYKLSKDDQEEIKELKEEGVYASMQXLDNPFEGKSVPDSLTMT
GARP      586 EEEEEEEEEEEEEEEEEEEEEEEDEDEEDEDDAEEDEDDAEEDEDDAEEDDDEEDDDEEDDDED
NucProt       ------------------------------------------------------------
C11G6.3       ------------------------------------------------------------
                . ..    .  .        .. ..  .  ...        .    .     .

ORF27     448 YRYGYYDLLGKPTLQEYILLEAKAHQTIVAPMDGVVSLDGDDVILTNGKGENESRLTLYS
GARP      646 EDEDEEDEEEEEEEEEESEKKIKRNLRKNAKI----------------------------
NucProt       ------------------------------------------------------------
C11G6.3       ------------------------------------------------------------
                    . .      .      .      . .

ORF27     508 IHNGRAIEGTRV
GARP          ------------
NucProt       ------------
C11G6.3       ------------
```

ORF27



ORF28

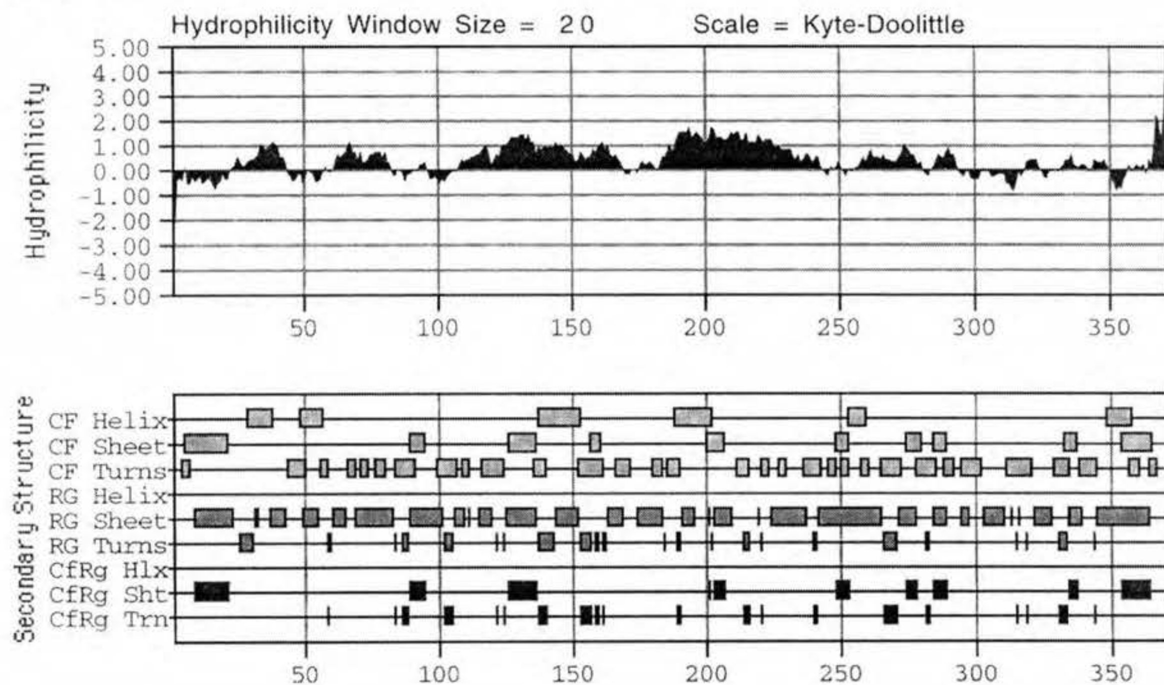


Figure 13. Hydropathy profile and secondary structure prediction of the deduced amino acid sequences from ORF27 and ORF28.

These bacteriophages belong to a class of DNA viruses that use a protein-priming mechanism for the initiation of replication. The exact function of the morphogenetic proteins is not known. The gene 13 protein from *Bacillus* phages ϕ-29 and PZA are known to be contiguous to other genes (genes 14 and 15) that appear to be involved in disruption of the cell membrane (21,47). On the other hand, details regarding the function of the gene 13 product from phage B103 are not available. Interestingly, the ORF28 product also shared weak similarity with a major pneumococcal autolysin (20). The role of the ORF18 product from pneumococcal phage Cp-1 in viral morphogenesis is not known. However, this protein was shown to share significant similarity to other tail proteins from *E. coli* phages that bind to bacterial receptors (adhesins) (44).

Figure 14. Multiple sequence alignment of the predicted amino acid sequence from the amino-terminal domain from ORF28 and its bacteriophage homologs. Conserved amino acids are shaded in black while conserved substitutions are shaded in gray. Sequences included are ORF28, ORF18 product from pneumococcal bacteriophage Cp-1, gene 13 protein from *Bacillus* phages PZA, ϕ-29, and B103.

```
ORF28     1 -------------------------------------VYVNPQFYFPKVIQLQTTILPAI
PZA       1 ------------------------------------------------------MVYVSN
ϕ-29      1 ------------------------------------------------------MVYVSN
B103      1 ------------------------------------------------------MFYSKN
ORF18     1 MYLSRDIGVKYKETTEGDKTYIEILSNITGSVAEIGRKNSRNYRTSGSSSGGGSGSTNES
                                                                  ...  .

ORF28    24 SQFGGDEF-ERAKHIYEFLKS--QGASPQAIAAILGNWSVESSINPKRAEGDYLTPPVGV
PZA       7 KYLTMSEMKVNAQYILNYLSN--NGWTKQAICGMLGNMQSESTINP------------GL
ϕ-29      7 KYLTMSEMKVNAQYILNYLSS--NGWTKQAICGMLGNMQSESTINP------------GL
B103      7 AYLSMKEMTVNAQYILNYLLP--RGWTKNAICGMLGNMQTESTINP------------GI
ORF18    61 GDISTEE--SRVRLVTRIIKKLVPDATAEGIAGIIGNFSAESNVTAKKYEADYAT---GY
            ..... *. ..........    ........*....**...**....           *.

ORF28    81 PIPPWDDE--------SWLAIGG-------PAIYSGAYPNILHRGLGLGQWTDTADGSTR
PZA      53 ----------------WQN--------------LDEGNTSL---GFGLVQWTP-------
ϕ-29     53 ----------------WQN--------------LDEGNTSL---GFGLVQWTP-------
B103     53 ----------------WQN--------------LDEGNTSL---GFGLVQWTP-------
ORF18   116 EYEKMESEPTAENLMGSWGAFASLYSISLNEAGYRGSDGN-HWIGIGTGQWT-----GPR
                            *..              ........   *.*..***.

ORF28   126 HTALLNYARTQ--NKKWYDLDLQLDFMLHGESPY---YQSWL-KDFFKNTGSAANLAQLF
PZA      73 ASNYINWANNQGIPYKNMDSELKRIIWEVNNNAQWNNLRDMTFKEYIKSTKTPRELAMIF
ϕ-29     73 ASNYINWANSQGLPYKDMDSELKRIIWEVNNNAQWINLRDMTFKEYIKSTKTPRELAMIF
B103     73 ATKYLNWADRNGLKRDHMDSQLKRILWEVDNNEQWINLRNMTFKEFTKSTKSANELAMIF
ORF18   170 AEELLNFARSQ--GKSLWDFNLQFQFMNQESRA-----DTF--RRVASSTASASTNASDF
            .. ..*.* .... .. .*..*....... ..... ... ...... ..*... ..*..*

ORF28   180 LTYREG---NSGDKLLERQTRATEWYYQIEKGFSQTNGGQAKSDPQSLEGVRGDLYDHSV
PZA     133 LASYE--RPANPNQPVRGDQAEYWYKNLSGGGGGGLQLAQFPMDIINITQGENGSFSHKG
ϕ-29    133 LASYE--RPANPNQPERGDQAEYWYKNLSGGGGGGLQLAQFPMDIINISQGENGSFSHKG
B103    133 LASYE--RPANPNQPERGTQAEYWFKTLTGKGSTGLQLAQFPMDIINITQGENGSFSHKG
ORF18   221 MNNWEGVAYKEAERIEH-------------------------------------------
            ....*  ........... ....... ... .  .........................

ORF28   237 PGGGDGMAYAYGQCTWGVAARMNQLGLKLKGRNGEKISIINTMGNGQDWVATSSSLGGET
PZA     191 TLCIDFVGKTEKYPYYAPCDCTCVWRGDASAYLAWTSDKEVMCADGSVRYITWVNVHESP
ϕ-29    191 TLCIDFVGKTEKYPYYAPCDCTCVWRGDASAYLAWTSDKEVMCADGSVRYITWVNVHESP
B103    191 TLCIDFVGKHEKYPYYAPCDCTCVWRGDESAYLAWTSDKEVMCADGIVRYITWVCVHDEN
ORF18       ------------------------------------------------------------
            ......... ................. ......................... ...

ORF28   297 GSTPRAGAIVSFVGGTHGTPASYGHVAFVEKVYDDGSFLVSETNYGGNLTIPLEKSLKQI
PZA     251 LPEDVGKKLKKGDLMGHTGIGGNVTGDHWHFNVIDGKEYQGWTKKPDSCLAGTELHIYDV
ϕ-29    251 LPEDVGKKLKKGDLMGHTGIGGNVTGDHWHFNVIDGKEYQGWTKKPDSCLAGTELHIYDV
B103    251 LLYNVGKKLKKGELMGHSGKGGRATGDHLHLNVIEGNKYQGWVKKPDSALAGTELHIYDV
ORF18       ------------------------------------------------------------
            . . ............... ..  .... . .....  .......... ...........
```

```
ORF28     357 VPSVLLIRPNRRVYT------------------------------------------
PZA       311 FAVNNVEIINGNGYDWKTSDWQDG-DGGDGGD-DNENNKTKDLITLLLSDALHGWKA
φ-29      311 FAVNNVEIINGNGYDWKTSDWQDG-DGGDGDD-DNDNNKTKDLITLLLSDALHGWKA
B103      311 FAVNGVEIVNGLGYDWKTSDWVDGSDENNGDDKDKDKDETKNIVNLLLCGALNGW--
ORF18         ---------------------------------------------------------
              .... ....... .......... .. .    . . . .   .. .. ...  .. ..
```

The carboxy-terminal portion of ORF28 showed a high level of homology to the TraG ($1.5x10^{-17}$) and TrsG ($7.5x10^{-17}$) products from the transfer regions of staphylococcal conjugative plasmids pSK41 and pGO1 respectively (16,46) (Fig. 8). The TraG predicted product was found to be a hydrophobic protein localized to the membrane. However, and similar to TraE and TrsE, its function remains to be identified after unsuccessful attempts at expression in an *E. coli* system. The TrsG protein from pGO1 has been identified as a 38-kDa protein that belongs to a complementation group involving other *trs* genes (TrsD through TrsK). This protein has proved to be essential for conjugation but its specific function during the transfer process has not been investigated further. Other significant homologies at the C-terminal domain of ORF28 included relatedness to the Isp conserved locus in Group A streptococci (GAS) that encodes an immunogenic secreted protein ($2.6x10^{-13}$) (45). A distinctive feature of the Isp protein was a bacterial signal peptide at its amino-terminus that is recognized by the Sec pathway. Homology to the TraG and TrsG proteins was also observed but the significance of the similarity was not discussed. ORF28 also shared similarity ($3.7x10^{-08}$) to ORF1 from *Staphylococcus aureus* Newman (Accession number X97985). This product appears to be involved in regulation of biosynthesis of staphyloxanthin, a major pigment produced by these microorganisms. Information regarding this study is not yet available.

Figure 15. Multiple sequence alignment between the carboxyl-terminal domain of ORF28 and its homologs. Sequences compared are TraG and TrsG from staphylococcal conjugative plasmids pSK41 and pGO1 respectively, Isp protein from *S. pyogenes*, and ORF1 product from *S. aureus* Newman.

```
ORF28      1 ------------------------------------------------------------
TraG       1 ------------------------------------------------------------
TrsG       1 ------------------------------------------------------------
Isp        1 MKKRKLLAVTLLSTILLNSAAPLVVADTSLRNNTSSTDQPTTADTDTDDESETAKKDKKS
ORF1       1 ------------------------------------------------------------

ORF28      1 ------------------------------------------------------------
TraG       1 ------------------------------------------------------------
TrsG       1 ------------------------------------------------------------
Isp       61 KETASQHDTQKDHKPSYNHPTPPSNDTKQTDQASSEATDKPNKDKNDTKQPDSSDQSTPS
ORF1       1 ------------------------------------------------------------

ORF28      1 -------------------------------------------VYVNPQFYFPKVIQEQT
TraG       1 ----------------------MWSAIAKGITMLMKKKALSKILKPVIAIVIGFFILLFI
TrsG       1 ----------------------MWSAIAKGITMLMKKKALSKILKPVIAIVIGFFILLFI
Isp      121 PKDQSSQKESQNKDGRPTPSPDQQKDQTPDKTPEKGPEKAAEKTPEPNRDAPKPIQPPLG
ORF1       1 ------------------------------------------------------------
                                                 .         .         . .

ORF28     18 TILPAIGQFGGDEFERAKHIYEFLKS--QGASPQAIAAILGNWSVES--SINPKRAECDY
TraG      39 MIAYLVGSFTKEINESGEQKMKEAES-------TATGA-CTGGSVEGS-GISAFEKNAKG
TrsG      39 MIAYLVGSFTKEINESGEQKMKEAES-------TATGA-CTGGSVEGS-GISAFEKNAKG
Isp      181 AAAPVFAPWRESDKDLSKLKPSSRSSAAYVRHWTCDSA-YTHNLLSRRYGITAEQLDGFL
ORF1       1 ------------------------------------------------------------
              .. ... .  .  . . .    .         .. ..  .  ...   ....   .

ORF28     74 LT--P------PVGVPIPPWDDESWLAIGGPAIYSGAYPNILHRGLGL--GQ---WIDTA
TraG      90 GAL-------EGKGKDMVKIAKKNKIPPKLFIAIVASESEWGRGANAT--KQKNPLSIMG
TrsG      90 GAL-------EGKGKDMVKIAKKNKIPPKLFIAIVASESEWGRGANAT--KQKNPLSIMG
Isp      240 NSLGIHYDKERLNGKRLLEWEKLTGLDVRAIVAIAMAESSLGTQGVA---KEK-GSNVFG
ORF1       1 -------------------------MKKIVTATIATAGLATIAFAGHDAQAAEQNNNG
              .          .. ..   .    . ..  ... .... ... . .    ...    .. .

ORF28    121 DGSTRHTALLNYARTQNKKWYDLDLQLDFMLHGDSPYYQSWLKD--FFKNTGSAANLAQL
TraG     141 AGPLQVYP--SIEEGLDKGA--KNLY-DLYI---SEGLTTPEK--IGPKYAPVG------
TrsG     141 AGPLQVYP--SIEEGLDKGA--KNLY-DLYI---SEGLTTPEK--IGPKYAPVG------
Isp      296 YGAFDFNPN-NAKKYSDEVAI-FHMVEDTII---ANKNQTFERQDLKAKKWSLGQLDTLI
ORF1      34 YNSNDAQS-YSYTYTTDAQG-NYHYTWTGNW--NPSQLTQNNT-YYYNNYNTYSYNNASY
              .     .  .    ... .  .  .   .   .   .  ... ..  .   ..   ..

ORF28    179 FLTYWEGNSGDKLLERQTRATEWYYQIEKGFSQTN-GGQAK-------SDPQSLEGVRGD
TraG     185 ---------ASNDPDDL---NSNWVPTVKKIMKSFG-GKEAKC-STE---SSGGSSDGKGF
TrsG     185 ---------ASNDPDDL---NSNWVPTVKKIMKSFG-GKEAKC-STE---SSGGSSDGKGF
Isp      351 DGGVYFTDTSGSGQR----RADIMTKLDQWIDDH--GNTPDI-PEHLKITSGTQFSEVPV
ORF1      89 NNYYNH---SYQYNN----YTNNSQTATNNYYTGGSGASYSTTSNNVHVTTTAAPSSNGR
                     .     .     .... ..  .  .  .  .  *   ..   .       .....    ..

ORF28    231 LYDHSVP----GGGDGMAYAYGQCTWGVAARMNQLGLKLKGRNGEKISIINTMGNGQDWV
TraG     229 DFKGEFPKPDKSKYNGQSYPWGQCTWYVHQRRKEIG----------KPVPLTWGNGGDWG
TrsG     229 DFKGEFPKPDKSKYNGQSYPWGQCTWYVHQRRKEIG----------KPVPLTWGNGGDWG
Isp      404 GYKRSQPQ-NVLTYKSETYSFGQCTWYAYNRVKELG----------YQVDRYMGNGGDWQ
ORF1     142 SISNGYA------SGSNLYTSGQCTYYVFDR---VG----------GKIGSTWGNASNWA
              ..  ..       . ...* .****... .* ...*          .   ..**...*.
```

```
ORF28     287 ATSSSLGGETGSTPRAGALVSFVGGTHGTPASYGHVAFVEKVYDDGSFLVSETNY-GGNL
TraG      279 DNAKAQGWEVGSKPKAGAGASVKPGNFGAPPPYGHIMFVEKVKKDGGIVVSEANV-KGLG
TrsG      279 DNAKAQGWEVGSKPKAGRGASVKPGNFGAPPPYGHIMFVEKVKKDGGIVVSEANV-KGLG
Isp       453 RKP---GFVTTHKPKVGYVVSFAPGCAGADATYGHVAVVEQIKEDGSILISESNV-MGLG
ORF1      183 NAAASSGYTVNNTPKVGAIMQ------TTQGYYGHVAYVEGVNSNGSVRVSEMNYGHGAG
                .   *.......*..*....  ... ....  ***...**... .*....** *. .*..

ORF28     346 TIPLEKSLKQIVPSVLLIRPNRRVYT
TraG      338 VISSREFSKAETQRMQFIY-DK----
TrsG      338 VISSREFSKAETQRMQFIY-DK----
Isp       509 TISYRTFTAEQASLLTYVVGDKLPRP
ORF1      237 VVTSRTISANQAGSYNFIH-------
              ..... ...     ....   ..
```

Expression and Identification of the ORF26 and ORF27 products

The 4.2-kb *BamHI* fragment at coordinates 33.0 to 37.2 in Tn*5252* (Fig. 3) has been of particular importance due to its potential role in the regulation of transfer (34,77). In order to verify the predicted ORFs in this region, the fragment was cloned in an appropriate orientation into the *BamHI* site of the *E. coli* expression vector pET30b(+)which is under control of a T7 promoter (Novagen). The resultant recombinant plasmid, designated pAF102, was transformed into *E. coli* BL21(DE3), a strain bearing a chromosomal copy of the T7 RNA polymerase gene under the control of *lacUV5* promoter. After 3 h of induction in the presence of 3 mM IPTG, cells were harvested and fractionated into cytoplasmic, periplasmic and membrane components and analyzed by SDS-PAGE (Fig. 16). The protein profile of the plasmid pAF102 showed three unique protein bands that fractionated with the cell membrane (Lane 7). Electrophoretic mobilities of the overexpressed proteins revealed products with apparent molecular masses of 63 kDa and 44 kDa. These observed values agreed with those calculated from the predicted amino acid sequences of 60.1 and 40.6 kDa for ORF27 and ORF28 respectively. A third band with a molecular weight of approximately 30 kDa was also detected. This protein was presumed to be a truncated form of the product encoded by ORF26. The size of this truncated protein was found to be in agreement with the one obtained from the predicted sequence

(34.3 kDa) of the 4.2-kb fragment. Analysis of the deduced amino acid sequence of ORF27 and ORF28 revealed extensive homology to membrane proteins found in a number of conjugative plasmids. Although the hydropathy profiles of these sequences did not show a significant content of hydrophobic residues, the presence of predicted transmembrane-spanning regions suggested association with membranes or with membrane proteins. Results from this experiment confirmed the location of these protein products. Interestingly, ORF28 appears to be expressed in greater concentrations than either ORF27 or the partial form of ORF26 that was present in the 4.2-kb *BamHI* fragment. The fact that a mutation that impairs conjugal transfer of the element lies upstream from ORF28 and it only affects the very end of ORF27 (Fig. 4) could imply that interruption of ORF28 expression rather than disruption of ORF27 is the cause for the inability of Tn*5252* to mobilize. Further studies utilizing ORF28 in mutagenesis and complementation analysis would be needed to confirm this assumption.

Another region that also proved to be essential in the transfer process of Tn*5252* was the 3.2-kb *BamHI* DNA segment localized at coordinates 40.6 to 37.3 in Tn*5252* (Fig. 3) (34). The deduced sequence revealed part of the end of ORF21 and 5 smaller ORFs preceding the start of ORF26 (Fig. 4). This fragment was also cloned in the appropriate orientation for transcription into the *BamHI* site of the expression vectors pET30-a(+), b(+) and c(+) and the resulting recombinant plasmids

designated pAF105, pAF106, and PAF107 respectively, were used to transform *E. coli* BL21(DE) cells. Transformants were then induced and processed as before. Analysis of total crude protein from induced cells is shown in Fig. 17. The total protein profile from pAF105 showed a protein band of approximately 60.0 kDa (Lane 5). This size is in reasonable agreement with the expected size of 55.5 kDa calculated from the deduced sequence from the section of ORF26 that resides in the 3.2-kb *BamHI* fragment. Additionally, pAF107 which carries the passenger DNA in a different reading frame, also showed a cluster of bands with molecular weights ranging from 15.0 to 16.0 kDa (Lane 13). These bands seemed to belong to the small ORFs 22 to 25b; however, the sizes from the deduced sequences from most of them were shown to be below 10.0 kDa (Table 5). In order to verify the size and number of bands present, a second SDS-PAGE analysis was performed with the same samples but using a 17% polyacrylamide gel. Results are shown in Fig. 18. In this gel, the cluster of bands appeared to be composed of at least 4 protein bands of sizes ranging from 13.0 to 15.0 kDa. The fact that analysis of the deduced sequence from some of these ORFs did not show a defined start site, suggests that some of them could be part of bigger ORFs which would account for the larger sizes of the protein bands observed in Figure 18. Therefore, it is possible that some of these ORFs may be part of larger ones and the predicted sizes could be a result of artifacts in the DNA sequence. No other proteins of smaller size were detected.

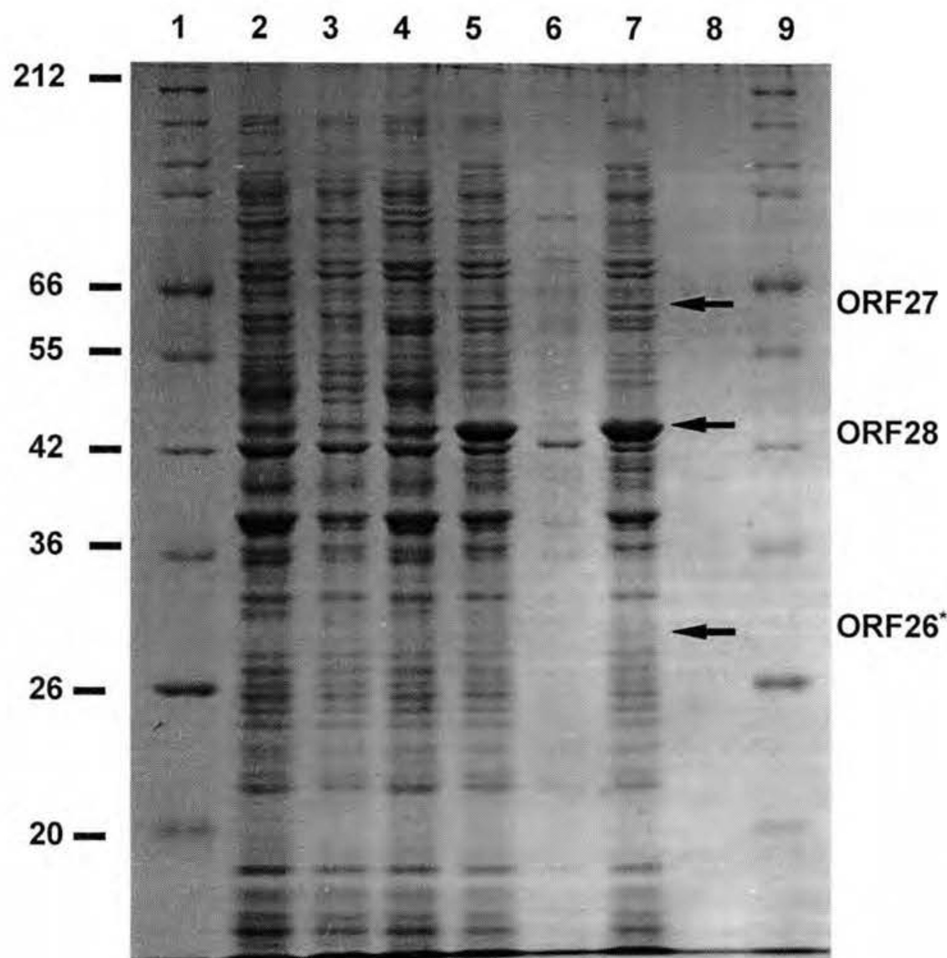


Figure 16. Protein profile from the 4.2-kb *BamH*I fragment from Tn*5252*. *E. coli* BL21(DE3) cells harboring recombinant plasmid pAF102 were induced for expression. Total protein from induced cells (lane 5) was fractionated into cytoplasmic (lane 6), membrane (lane 7) and periplasmic (lane 8) fractions. Controls used were total protein from uninduced (lane 2) and induced (lane 3) vector pET30b(+) and uninduced pAF102 (lane 4). Lanes 1 and 9 show protein marker (NEB) and the molecular weights (in kDa) are shown on the left of the gel. Positions of ORF27, ORF28 and a truncated form of ORF26 (ORF26*) are indicated by arrows.

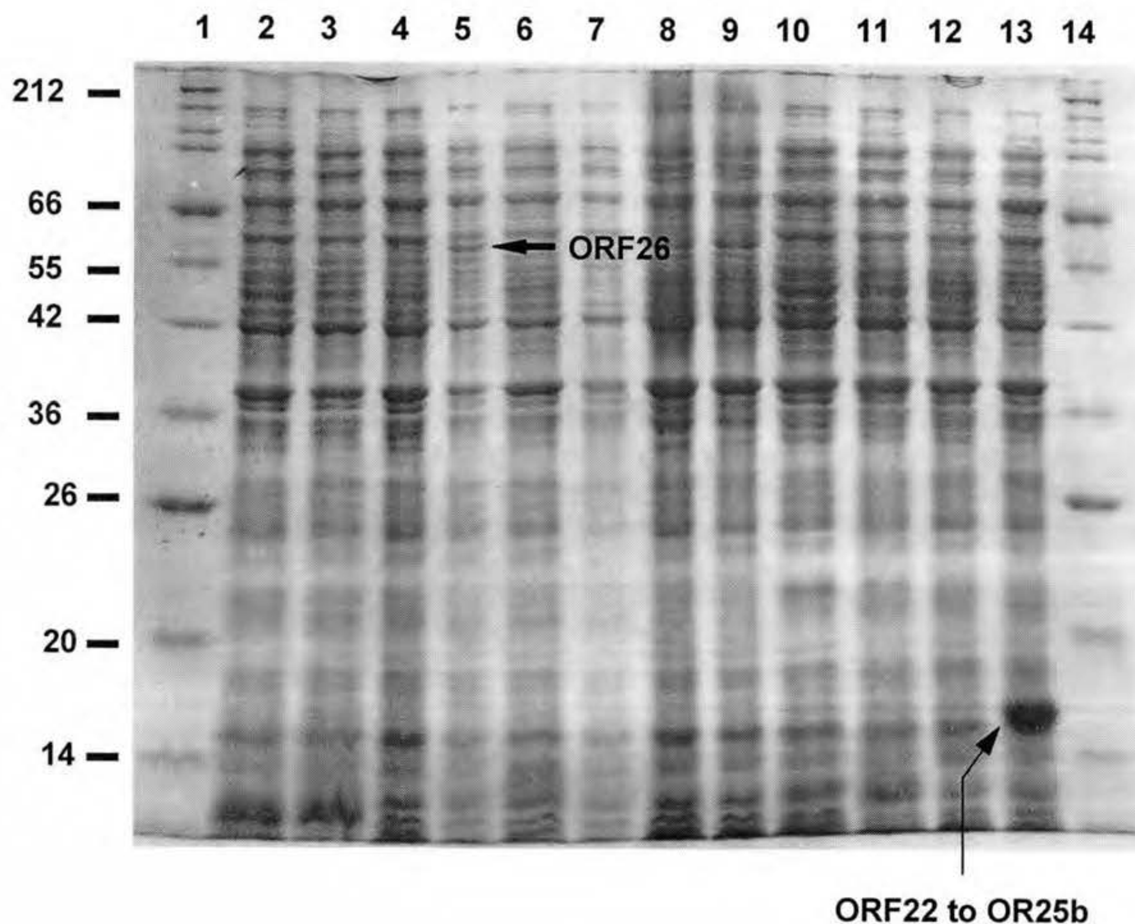


Figure 17. Protein profile from the 3.2-kb *BamH*I fragment from Tn*5252*. Profiles shown correspond to uninduced and induced vectors pET30a(+) (lanes 2 and 3), pET30b(+) (lanes 6 and 7), and pET30c(+) (lanes 10 and 11), and uninduced and induced pAF105 (lanes 4 and 5), pAF106 (lanes 8 and 9), and pAF107 (lanes 12 and 13), respectively. Lanes 1 and 14 show the protein marker (NEB) and the molecular weights (in kDa) are shown on the left of the gel. Positions representing partial ORF26 and ORF22 to ORF25b are indicated by arrows.

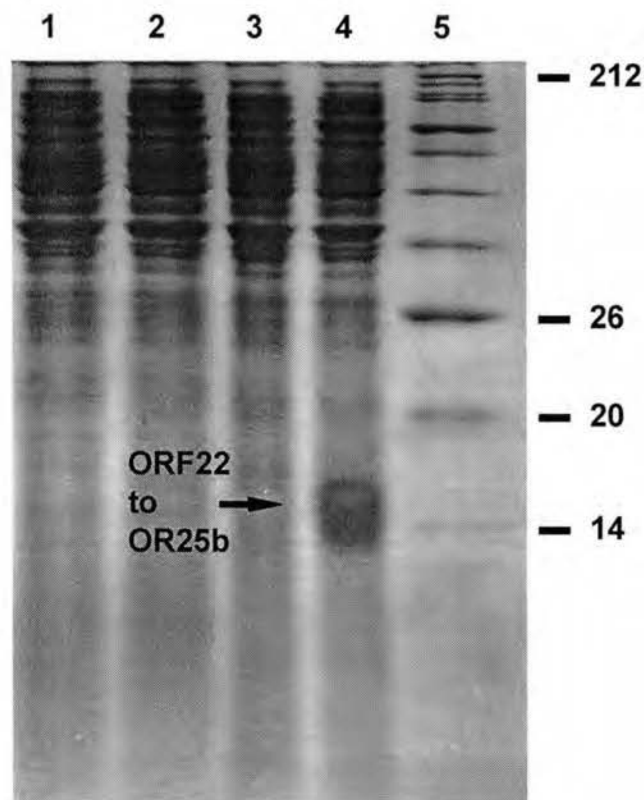


Figure 18. Protein profile from the 3.2-kb *BamH*I fragment from Tn*5252* using 17% Polyacrylamide slab gels. Profiles shown correspond to uninduced and induced pET30c(+) (lanes 1 and 2) and uninduced and induced pAF107 (lanes 3 and 4) respectively. Lane 5 shows the protein marker (NEB) and the molecular weights (in kDa) are shown on the right of the gel. Positions representing ORF22 to ORF25b are indicated by an arrow.

CHAPTER 5

CONJUGATIVE TRANSFER OF TRANSFER-DEFICIENT Tn*5252* BY COMPLEMENTATION

RESULTS AND DISCUSSION

Intraspecific Mobilization of Transfer Deficient Tn*5252*

In a recent study (34), transfer deficient pneumococcal strain SP1256 was constructed by insertion of heterologous DNA at an internal *Xba*I site of the 4.2-kb *BamH*I region in Tn*5252* (Fig. 3). Sequence analysis of this region has located this insertion in the carboxy-terminal region of ORF27 (Fig. 4). The next step to confirm the requirement of this protein in the process of conjugation was to perform complementation analysis to provide ORF27 in *trans* on a separate plasmid to the mutant strain SP1256. However, attempts to generate replicating shuttle vectors bearing the intact 4.2-kb fragment were unsuccessful, possibly because the fragment was very unstable and/or the product expressed in high copy plasmids was toxic to the bacterial cells. Due to these problems, a different approach was undertaken. Transposon Tn*5251* has been shown

to encode its own transfer properties when separated from the context of the larger element Tn*5253* (2). It was suspected that functions provided by Tn*5251* could complement those inactivated in Tn*5252*. In order to investigate this possibility, strains carrying both transfer-proficient Tn*5251* and transfer-deficient Tn*5252* were constructed to study their transfer properties. Pneumococcal strain SP1704 carrying Tn*5251* (33) was used as donor in conjugation experiments with the recipient strain SP1256 that carries transfer-deficient Tn*5252*. Parallel to this, SP1705, a strain carrying Tn*5251* on a different location in the chromosome (33) was also crossed with SP1256. A third mating between SP1704 and transfer-proficient SP1254 bearing Tn*5252* (34) was used as a control. Conjugation was performed on filters and using a 1:1 ratio of donor to recipient cells as previously described. Transconjugants were selected on CAT agar containing tetracycline (Tn*5251*) and streptomycin (recipient). Additionally, transconjugants arising from the reciprocal matings were also selected by using erythromycin (Tn*5252*) and novobiocin (donor). Screening of unselected markers was performed by replica plating. Results from the conjugation experiments are shown in Table 6 and Fig. 19 summarizes the strategy used for inter- and intraspecific mobilization of Tn*5252* and Tn*5251* and the parental strains that were used along with the transconjugants obtained.

As expected, Tn*5251* was mobilized from SP1704 into SP1256 at a frequency of 2.0×10^{-8} Tc^r Str^r transconjugants per donor cell. Similarly, the

reciprocal mating showed back transfer of Tn*5252* from SP1256 into the donor SP1704 at a frequency of $3.4x10^{-8}$ Em^r Novr transconjugants per donor cell, indicating that Tn*5251* was able to mobilize transfer-deficient Tn*5252* into the donor by complementation.

Conjugation between SP1705 and SP1256 showed similar results (Table 6). In the second mating, Tn*5251* was transferred from SP1705 at a frequency of $6.6x10^{-8}$ Tc^r Str^r transconjugants per donor cell into SP1256. Again, complementation was observed in the reciprocal cross with Tn*5252* being transferred into SP1705 at a frequency of $1.7x10^{-7}$ per donor cell. Overall, frequencies in both directions fell within the range of those obtained for the control cross between transfer-proficient strains SP1704 and SP1254 (Table 6). Additional crosses using SP1254 and SP1256 as donors and transposon-free DP1002 as the recipient were also performed to confirm their transfer properties. As expected, Tn*5252* was mobilized from SP1254 into DP1002 at a frequency of $1.0x10^{-6}$ Em^r Novr transconjugants per donor cell while no transconjugants were detected ($<1.0x10^{-9}$) when SP1256 carrying transfer-deficient Tn*5252* was used as the donor.

Figure 19. STRATEGY FOR INTER- AND INTRASPECIFIC TRANSFER OF Tn*5251* AND Tn*5252*[1].

Forward:

SP1704 Xg SP1256 => AF101F-105F => AF101F Xg DP1004 => Tc => No transconjugants
Tc^r,Tra^+ Em^r,Tra^- Tc^r,Em^r Tc^r,Em^r Str^r Em => No transconjugants

=> AF104F Xg 21547=> Tc => No transconjugants
Tc^r,Em^r Opt^r Em => No transconjugants

SP1705 Xg SP1256 => AF201F => AF201F Xg DP1004 => Tc => No transconjugants
Tc^r,Tra^+ Em^r,Tra^- Tc^r,Em^r Tc^r,Em^r Str^r Em => No transconjugants

Reciprocal:

SP1256 Xg SP1704 => AF101R => AF101R Xg 21547 => Tc => No transconjugants
Em^r,Tra^- Tc^r,Tra^+ Tc^r,Em^r Tc^r,Em^r Opt^r Em => AF904
Em^r,Opt^r

SP1256 Xg SP1705 => AF201R
Em^r,Tra^- Tc^r,Tra^+ Tc^r,Em^r

Figure 19. (Continued)

Controls:

SP1254 Xg DP1002
Em^r,Tra^+ Nov^r

SP1256 Xg DP1002 => No transconjugants
Em^r,Tra^- Nov^r

SP1704 Xg SP1254 => AF301
Em^r,Tra^+ Tc^r,Tra^+ Tc^r,Em^r

SP1254 Xg SP1704 => AF301R, AF302R => AF301R Xg DP1004 => Tc => AF601 (Tc^r,Str^r); Em => AF602 (Em^r,Str^r)
Em^r,Tra^+ Tc^r,Tra^+ Tc^r,Em^r — AF301R: Tc^r,Em^r; DP1004: Str^r

=> AF301R Xg 21547 => Tc => AF902 (Tc^r,Opt^r); Em => AF801 (Em^r,Opt^r)
AF301R: Tc^r,Em^r; 21547: Opt^r

[1] Xg: conjugation, Em: erythromycin, Tc: tetracycline, Str: streptomycin, Nov: novobiocin, Opt: optochin.

Transfer Properties of Transconjugants

The results obtained during the reciprocal conjugations strongly indicated that inactivated functions in mutant Tn*5252* could be complemented by those in Tn*5251*. To confirm these results, transconjugants AF101F and AF201F carrying both transposons (Fig. 19) were used in matings with pneumococcal recipient DP1002. In theory, the presence of both transposons in the chromosome would produce the same effect of complementation as the one observed during reciprocal mating. However, as seen in Table 6, no transconjugants were obtained from these crosses. To further investigate this, transconjugant AF301R, obtained from the reciprocal control mating (Fig. 19) was also tested for its transfer functions. This strain also carried both transposons but unlike AF101F or AF201F, the elements were transfer-proficient. Transconjugants were selected on agar containing tetracycline (Tn*5251*) or erythromycin (Tn*5252*). As expected, AF301R was able to mobilize Tn*5251* into the recipient DP1004 (Str^r), at a frequency of 5.7×10^{-7} per donor cell respectively (Table 6). However, colonies resistant to tetracycline were sensitive to erythromycin indicating that only Tn*5251* had been transferred. Similarly, colonies resistant to erythromycin were also tetracycline-sensitive suggesting that only Tn*5252* was mobilized into the recipient. The frequency of Tn*5252* transfer from AF301R into DP1004

was 1.5×10^{-7} per donor cell. Two transconjugants from these matings, AF601 (Tc^r) and AF602 (Em^r) (Fig. 19), were selected for further studies.

Interspecific Mobilization of Transfer-Deficient Tn*5252*

Results from conjugal transfer of Tn*5252* between different strains of *S. pneumoniae* indicated that transfer-deficient Tn*5252* could be mobilized by Tn*5251*. However, this raised the question of transformation being responsible for the presence of transfer-deficient Tn*5252* in the reciprocal transconjugants rather than complementation. *S. pneumoniae* strains become highly competent during the late log phase of growth (41) and undergo a high frequency of transformation in the presence of exogenous DNA. Although mating mixtures from the conjugation experiments were kept on ice and DNase was used at all times, the possibility existed that transformation had taken place at some point during incubation at 37°C after plating. To rule out genetic transformation between *S. pneumoniae* strains, selected transconjugants that were previously shown to carry both transposons in their chromosome were used in mating experiments with *Streptococcus pyogenes* 21547. *S. pyogenes* is not naturally transformable and mobilization of the transposons into these strains can only take place by conjugal transfer. Matings were performed as described before and the transfer frequencies are described in Table 7.

TABLE 6

INTRASPECIFIC TRANSFER FREQUENCIES OF Tn*5251* AND Tn*5252*

Donor		Recipient		Transconjugants per donor		
Strain	Relevant phenotype	Strain	Relevant phenotype	Em Nov	Tc Str	Em Str
SP1704	Tn*5251* Tra^+	SP1254 SP1256	Tn*5252* Tra^+ Tn*5252* Tra^-	$1.0x10^{-5}$ $3.4x10^{-8}$	$5.2x10^{-8}$ $2.0x10^{-8}$	
SP1705	Tn*5251* Tra^+	SP1256	Tn*5252* Tra^-	$1.7x10^{-7}$	$6.6x10^{-8}$	
SP1254	Tn*5252* Tra^+	DP1002	Nov^r	$1.0x10^{-6}$		
SP1256	Tn*5252* Tra^-	DP1002	Nov^r	$<1.0x10^{-9}$		
AF101F	Tn*5252* Tra^-, Tn*5251* Tra+	DP1002	Nov^r	$<1.0x10^{-9}$	$<1.0x10^{-9}$	
AF201F	Tn*5252* Tra^-, Tn*5251* Tra^+	DP1002	Nov^r	$<1.0x10^{-9}$	$<1.0x10^{-9}$	
AF301R	Tn*5251* Tra^+, Tn*5252* Tra^-	DP1004	Str^r		$5.7x10^{-7}$	$1.5x10^{-7}$

Transconjugants AF104F and AF105F (Fig. 19), containing both transfer-deficient Tn*5252* and transfer-proficient Tn*5251* were used in matings with *S. pyogenes* 21547. As seen in Table 7, no transconjugants were detected. These results were in agreement with those observed for AF101F and AF201F during intraspecific matings (Table 6). On the other hand, when AF101R from the reciprocal mating was crossed with *S. pyogenes* 21547, only Em^r transconjugants were detected (4.4×10^{-7} per donor cell), indicating that only Tn*5252* was transferred. Apparently, Tn*5251* in AF101R was able to complement Tn*5252* transfer functions, however, Tn*5251* could not transfer for an unknown reason. A transconjugant from this cross (AF904) was selected to test for transformation properties. Furthermore, conjugation between AF301R from the control mating and *S. pyogenes* 21547 gave transfer frequencies similar to those observed in intraspecific matings (Table 5). Tn*5252* and Tn*5251* were mobilized independently from AF301R into 21547 at frequencies of 6.6×10^{-7} Em^r and 2.8×10^{-5} Tc^r transconjugants per donor cell, respectively. Replica plating of a number of colonies from this cross failed to detect isolates resistant to both erythromycin and tetracycline. Two of these transconjugants designated AF801 (Em^r) and AF902 (Tc^r) were selected to study their transformation properties.

TABLE 7

INTERSPECIFIC TRANSFER FREQUENCIES OF Tn*5251* AND Tn*5252*

Donor		Recipient		Transconjugants per donor	
Strain	Relevant phenotype	Strain	Relevant phenotype	Em Opt	Tc Opt
S. pneumoniae		*S. pyogenes*			
AF103F	Tn*5252* Tra^-, Tn*5251* Tra^+	21547	Opt^r	$<1x10^{-9}$	$<1x10^{-9}$
AF104F	Tn*5252* Tra^-, Tn*5251* Tra^+	21547	Opt^r	$<1x10^{-9}$	$<1x10^{-9}$
AF105F	Tn*5252* Tra^-, Tn*5251* Tra^+	21547	Opt^r	$<1x10^{-9}$	$<1x10^{-9}$
AF101R	Tn*5252* Tra^-, Tn*5251* Tra^+	21547	Opt^r	$4.4x10^{-7}$	$<1x10^{-9}$
AF301R	Tn*5252* Tra^+, Tn*5251* Tra^+	21547	Opt^r	$6.6x10^{-7}$	$2.8x10^{-5}$

Analysis of Transconjugants by Southern Hybridization and Transformation

The presence of both the transposons in the transconjugants considered was confirmed by Southern blots and by transformation. Chromosomal DNAs from the parental strains and selected transconjugants were digested with *EcoR*I, electrophoresed, blotted, and hybridized with the plasmid probe pAM118. The plasmid pAM118 is pVA838 carrying transposon Tn*916* (22). The Tn*916* portion of the probe was expected to hybridize to Tn*5251* while the vector portion (pVA838) should hybridize to the pVA891 insertion in the transfer-deficient Tn*5252*.

The hybridization patterns of the digested DNAs probed with ^{32}P-labeled pAM118 are shown in Fig. 20. The probe hybridized with control DNAs as expected. Rx1 (lane 2), SP1000 (Lane 4), and DP1004 (Lane 8) did not hybridize to the probe because they did not contain Tn*5251*. DP1322 carrying the parental transposon, Tn*5253* (Lane 3), showed two hybridizing bands of approximately 22.9 and 12.2 kb. SP1256 (lane 5) does not contain *Tn5251*, but it carries pVA891 inserted in Tn*5252*, which reacted with the vector portion of the probe. SP1704 on the other hand, carries Tn*5251* and showed two fragments of 20.0 kb and 3.5 kb hybridizing with the probe (Lane 6). Transconjugant AF101F showed the same fragment patterns found in the parental strains SP1704 and SP1256 demonstrating the presence of both elements in its chromosome (lane 7). The size of the hybridizing fragments in AF101F indicated that the Tn*5251*

element was inserted at the same location as in the donor strain. Previous studies have shown that Tn*5251* inserts at multiple sites in the chromosome of *S. pneumoniae* (78). In this case, this element appeared to have inserted at the same site as that in the donor strain.

Transconjugant strains AF301R and AF302R (lanes 10 and 11), derived from the control reciprocal mating (Fig. 19) showed a hybridization profile identical to their parental strains SP1704 (lane 6) and SP1254 (lane 9). Two fragments of approximately 3.5 kb and 20.0 kb hybridized to the Tn*916* probe while another two fragments of 6.4 and 4.4 kb reacted with the pVA838 portion of the probe, indicating the presence of Tn*5252*. Conjugation experiments using AF301R as a donor in intraspecific matings, revealed that when both transfer-proficient transposons Tn*5251* and Tn*5252* were in the same host but in different locations, they could transfer independently to a transposon-free recipient (Table 6). Transconjugant strains AF601 and AF602, which arose from these matings, were also analyzed by Southern blotting. Digested DNA from tetracycline resistant AF601 (lane 12) revealed the same profile as that of parental strain SP1704 confirming the presence of Tn*5251*. Similarly, erythromycin resistant transconjugant AF602 showed only two bands of 4.4 kb and 6.4 kb that hybridized to the pVA891 insertion in Tn*5252* (Lane 13).

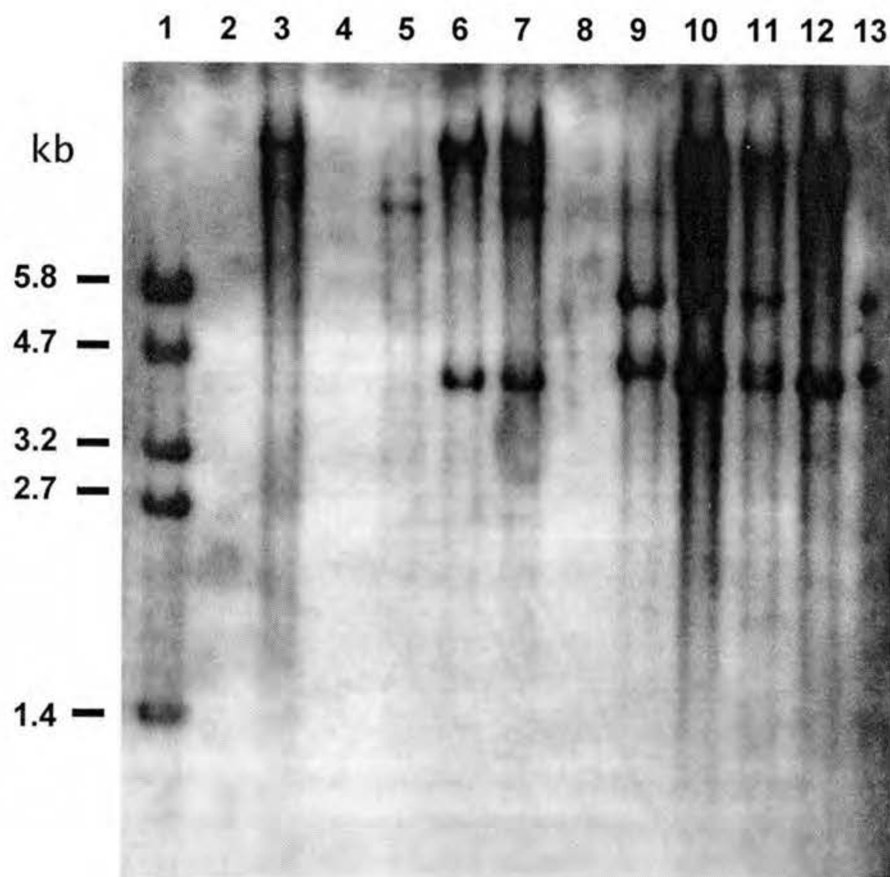


Figure 20. Identification of Tn*5251* and Tn*5252* by DNA hybridization. Chromosomal DNA from all the strains were digested with *EcoR*I and hybridized with the ^{32}P-labeled probe pAM118, a pVA838-based plasmid carrying Tn*916*. Lanes 5, 6, 8, and 9 show the parental strains SP1256, SP1704, DP1004, and SP1254 respectively. Lanes 7 and 10 to 13 show transconjugants AF101F, AF301R, AF302R, AF601, and AF602 respectively. Controls for this experiment were Rx1 (lane 2), DP1322 (lane 3), and SP1000 (lane 4). The molecular weight marker is in lane 1 with the molecular masses shown on the left of the figure.

Selected isolates from the interspecific conjugation that showed resistance to either tetracycline or erythromycin were used in transformation experiments to confirm the presence of the elements in their chromosome. The transconjugants selected were *S. pyogenes* transconjugant strains AF801 (Em^r) and AF902 (Tc^r), derived from pneumococcal donor strain AF301R (Tc^r, Em^r), and *S. pyogenes* AF904 (Em^r), derived from pneumococcal donor strain AF101R (Tc^r, Em^r) (Fig. 19). Additionally, parental *S. pneumoniae* strains AF104F and AF101R were also tested for their transformation properties. Partially purified DNAs from these strains were used to transform the wild-type strain Rx1, SP1000 carrying Tn*5252*, and Tn*5253*-bearing strain DP1333 that carries a *tet-3* point mutation. If resistance to tetracycline and erythromycin came from Tn*5251* and Tn*5252* respectively, transformation of a recipient strain carrying these elements should be more efficient than transformation of the wild-type strain Rx1. Transformation of the latter would require insertion of a larger segment from the element, an event that has been demonstrated to be very inefficient (64). Transformation frequencies are shown in Table 8. As evident from the results, DNAs from the donor strains were able to transform the element-bearing strains at much higher frequencies than the wild-type Rx1. This confirmed that resistance to the antibiotics was due to the presence of the elements in the chromosome and not due to spontaneous mutations. Exceptions to this were the DNAs from pneumococcal strains AF104F and AF101R that transformed DP1333

to tetracycline resistance but were unable to transform SP1000 to erythromycin resistance. This represents a puzzling finding since AF101R was able to transfer Tn*5252* (Em^r) but not Tn*5251* (Tc^r) into *S. pyogenes* 21547 by conjugation (Table 7). Furthermore, although AF104F was also able to transform DP1333 to tetracycline resistance, previous conjugation experiments proved this strain incapable of transfer of either element in intra- or interspecific matings.

TABLE 8

TRANSFORMATION PROPERTIES OF SELECTED TRANSCONJUGANTS

Donor	Selection[1]	Transformants per ml		
		RX1	SP1000	DP1333
AF801	Em	5	850	
AF902	Tc	0	0	$1.0x10^5$
AF904	Em	50	$1.6x10^3$	
AF104F	Em Tc		ND	$7.7x10^3$
AF101R	Em Tc		ND	$1.0x10^4$
DP1617	*ery*	$6.4X10^5$	$1.9x10^4$	

[1] Em: erythromycin, Tc: tetracycline; *ery*: chromosomal point mutation conferring low level of resistance to erythromycin; ND: Not detected

CHAPTER 6

CONCLUSIONS

DNA Sequence Analysis and Protein Expression

During the course of characterizing mutations that impaired conjugal transfer of Tn*5252*, the sequence of a 9.5-kb *BamH*I fragment (coordinates 33.0 to 42.5) spanning these sites was obtained. This DNA segment could potentially encode several proteins showing strong sequence similarity to transfer genes of conjugative plasmids from Gram-positive and from Gram-negative microorganisms as well. This region is composed of 10 potential ORFs that are transcribed in the same orientation suggesting that these are part of an operon. All of the ORFs were preceded by a consensus ribosomal binding sequence located within appropriate distance from the translational start site. Predicted products encoded by these ORFs varied in size from 8.0 to 89.7 kDa and with the exception of the ORF26 amino acid sequence, all predicted proteins showed a strong tendency to form transmembrane segments, suggesting membrane association. No consensus motifs characteristic of bacterial

promoters were detected. However, is likely that the promoter may be located still upstream of the start-site of ORF20 and outside of the context of the sequence considered here.

Complementation Analysis

Conjugative transposons are highly promiscuous and they can be transferred to a broad range of different species and even different genera (10,55,60,62). Furthermore, these elements are not subject to host restriction during conjugation. Earlier reports have shown that the presence of two copies of a transposon in a cell appears to stimulate transposition in a process called transactivation. This process seems to be mediated by the *trans* activity of the Int and Xis gene products from one transposon on the other but the exact mechanism is unknown (17). A similar process has been suggested for *Bacteroides* conjugative transposons. However, in this case, the stimulatory process appears to come from the *trans* activity of regulatory proteins and not from the action of excision proteins (55). These reports strongly suggest that transfer functions from homologous or otherwise related transposons could be used interchangeably. In the present study, a mutation that impaired conjugal transfer of Tn*5252* was complemented in *trans* by the Tn*916*-like transposon Tn*5251*. These results suggested a mechanism of retrotransfer in which the presence of the Tn*5251* element in the recipient

allowed transfer-deficient Tn*5252* to be mobilized into the donor by providing the missing functions. Similar findings were reported by Sia *et al* (68) who demonstrated that retrotransfer required the transfer of conjugation genes to the recipient. They suggested that at least one function involved in formation and maintenance of the conjugation bridge is needed for retrotransfer to occur. Interestingly, most of the products encoded by the ORFs characterized in this study were shown to be significantly related to proteins participating in mating bridge formation in Gram-negative conjugative systems. In this context, the ATP-binding domains detected in ORF21 could provide the energy required for DNA translocation across the cell membrane. ABC transporters usually require assistance from accessory factors that function to facilitate the release of the transported products into the extracellular medium (12). These factors are usually linked to genes encoding the ABC transporters. In the present study, the smaller hydrophobic proteins encoded by ORFs 22 to 25b could play the role of the accessory factors. Similarly, the high pI (9.55) of the predicted product from ORF27 and its predicted tendency to form α-helical structures could suggest a role of this protein in binding and transport of DNA. Finally, the signal to initiate formation of the mating bridge could be regulated by the product encoded by ORF26 which showed a highly conserved motif typical of bacterial regulatory proteins (28).

Evidence to support the participation of these putative genes in the transport channel formation may come from the experiments of reciprocal

conjugations. Transfer-deficient Tn*5252* could have been able to transfer back into the donor by using a bridge formed by Tn*5251* already present in the donor chromosome. On the other hand, once both transposons were in the same cell, the resultant transconjugant was only able to mobilize Tn*5252* but not Tn*5251*. This could have been due to Tn*5251* being repressed by some unknown *trans*-acting factor produced by Tn*5252*. Earlier studies on the characteristics of the composite element Tn*5253* demonstrated that Tn*5251* was not able to mobilize independently from the context of the larger element (2,78). Therefore, it could be speculated that a similar kind of repression is taking place even though they are now separated. In contrast, when Tn*5251* is mobilized into a strain that already contains transfer-deficient Tn*5252*, the resulting transconjugant is now no longer capable of mobilizing any of the elements. It could be argued that when Tn*5252* is mobilized, it inserts at a site from which it is able to transfer when complemented by Tn*5251*. However, this is unlikely since it has been demonstrated that this element has a preferred site of insertion (2,78). The possibility of transformation was ruled out since the same transfer patterns were observed during intraspecific conjugations involving *S. pyogenes* 21547 as the recipient. Studies on the transfer properties of the transconjugants bearing only one type of element and localization of these elements in their chromosome would provide information leading to characterization of the mechanism of transfer of this novel class of mobile elements.

BIBLIOGRAPHY

1. Alt-Mörbe, 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 plasmids is closely related to the transfer system of an IncP plasmid and distantly related to Ti plasmid *vir* genes. J. Bacteriol. 178:4248-4257.

2. Ayoubi, P., A. O. Kilic., and M. N. Vijayakumar. 1991. Tn*5253*, the pneumococcal Ω(*cat tet*) BM6001 element, is a composite structure of two conjugative transposons, Tn*5251* and Tn*5252*. J. Bacteriol. 173:1617-1622.

3. Balzer, D., W. Pansegrau, and E. Lanka. 1994. Essential motifs of relaxase (TraI) and TraG proteins involved in conjugative transfer of RP4. J. Bacteriol. 176:4285-4295.

4. Bandiera, A., U. A. Patel, G. Manfioletti, A. Rustighi, V. Giancotti, and C. Crane-Robinson. 1995. A precursor-product relationship in molluscan sperm proteins from *Ensis minor*. Eur. J. Biochem. 233:744-799.

5. Boccard, F., T. Smokvina, J. L. Pernodet, E. Friedman, and M. Guerineau. 1989. Molecular analysis of a composite chromosomal conjugative element (Tn*3701*) of *Streptococcus pyogenes*. Plasmid. 21:59-70.

6. Brendel, V., and E. N. Trifonov. 1984. A computer algorithm for testing potential prokaryotic terminators. Nucl. Acids Res. 12:4411-4427.

7. Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type-1 specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. USA. 93:14648-14653.

8. Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid structure. Adv. Enzymology. 47:45-148.

9. Clewell, D. B. 1981. Plasmids, drug resistance, and gene transfer in the genus *Streptococcus*. Microbiol. Rev. 45:409-436.

10. Clewell, D. B. and C. Gawron-Burke. 1986. Conjugative transposons and the dissemination of antibiotic resistance in streptococci. Ann. Rev Microbiol. 40:635-659.

11. Courvalin, P., C. Carlier, and F. Caillaud. 1987. Tn*1545*:A conjugative shuttle transposon. Mol. Gen. Genet. 206:259-264.

12. Dreiseikelmann, B. 1994. Translocation of DNA across bacterial membranes. Microbiol. Rev. 58:293-316.

13. Dubnau, D. 1991. Genetic competence in *Bacillus subtilis*. Microbiol. Rev. 55:395-424.

14. Farrand, S. K., I. Hwang, and D. M. Cook. 1996. The *tra* region of tghe nopaline-type Ti plasmid is a chimera with elements related to the transfer systems of RSF1010, RP4, and F. J. Bacteriol. 178:4233-4247.

15. Fath, M. J., and R. Kolter. 1993. ABC transporters: bacterial exporters. Microbiol. Rev. 57:995-1017.

16. Firth, N., K. P. Ridgway, M. E. Byrne, P. D. Fink, L. Johnson, I. T. Paulsen, and R. A. Skurray. 1993. Analysis of a transfer region from the staphylococcal conjugative plasmid pSK41. Gene. 136:13-25.

17. Flannagan, S. E. and D. B. Clewell. 1991. Conjugative transfer of Tn*916* in *Enterococcus faecalis*: *trans* activation of homologous transposons. J. Bacteriol. 173:7136-7141.

18. Franke, A. E., and D. B. Clewell. 1981. Evidence for a chromosome-borne resistance transposon (T*n916*) in *Streptococcus faecalis* that is capable of "conjugal transfer" in the absence of a conjugative plasmid. J. Bacteriol. 145:494-502.

19. 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.

20. Garcia, P., J.L. Garcia, E. Garcia, and R. Lopez. 1986. Nucleotide sequence and expression of the pneumococcal autolysin gene from its own promoter in Escherichia coli. Gene. 43:265-272.

21. Garvey, K. J., M.S. Saedi, and J. Ito. 1986. Nucleotide sequence of *Bacillus* phage ø29 genes 14 and 15: Homology of gene 15 with other phage lysozymes. Nucl. Acids Res. 14:10001-10008.

22. Gawron-Burke, C. and D. B. Clewell. 1984. Regeneration and insertionally inactivated streptococcal DNA fragments after excision of transposon Tn*916* in *Escherichia coli*: Strategy for targeting and cloning genes from Gram-positive bacteria. J. Bacteriol. 159:214-221.

23. 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.

24. Guild, W. R., and N. B. Shoemaker. 1976. Mismatch correction in pneumococcal transformation: donor length and hex-dependent marker efficiency. J. Bacteriol. 125:125-135.

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

26. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557-580.

27. Håvarstein, L. S., G. Coomaraswamy, and D. A. Morrison. 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA. 92:11140-11144.

28. Henikoff, S., G. W. Haughn, J. M. Calvo, and J. C. Wallace. 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA. 85:6602-6606.

29. Hofmann, K., and W. Stoffel. 1993. TMbase - A database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler. 347:166.

30. Inamine, J. M. and V. Burdet. 1985. Structural organization of a 67-kilobase streptococcal conjugative element mediating multiple antibiotic resistance. J. Bacteriol. 161:620-626.

31. Ippen-Ihler, K., and R. A. Skurray 1993. Genetic organization of transfer-related determinants on the sex factor F and related plasmids. pp. 23-52. *In* D. B. Clewell (ed.), Bacterial conjugation. Plenum Press, New York.

32. Kaneko, T., A. Tanaka, S. Sato, H. Kotani, T. Sazuka, N. Miyajima, M. Sugiura, and S. Tabata. 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. DNA Res. 2:153-166.

33. Kiliç, A. O. 1994. Genetic characterization of a composite streptococcal conjugative transposon Tn*5253*. Ph. D. Thesis. Oklahoma State University.

34. Kiliç, A. O., M. V. Vijayakumar, and S. F. Al-khaldi. 1994. Identification and nucleotide sequence analysis of a transfer-related region in the streptococcal conjugative transposon Tn*5252*. J. Bacteriol. 176:5145-5150.

35. Laemmli, U. K. 1970. Clevage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227:680-685.

36. Lanka, E. and B. M. Wilkins. 1995. DNA processing reactions in bacterial conjugation. Ann. Rev. Biochem. 64:141-169.

37. Le Bougenéc, C., G. De Cespédès, and T. Horaud. 1988. Molecular analysis of a composite chromosomal conjugative element (Tn*3701*) of *Streptococcus pyogenes*. J. Bacteriol. 170:3930-3936.

38. Le Bougenéc, C., G. De Cespédès, and T. Horaud. 1990. Presence of chromosomal elements resembling the composite structure Tn*3701* in streptococci. J. Bacteriol. 172:727-734.

39. Lessl, M., D. Balzer, W. Pansegrau and E. Lanka. 1992. Sequence similarities between the RP4 Tra2 and the Ti VirB region strongly support the conjugation model for T-DNA transfer. J. Biol. Chem. 267:20471-20480.

40. Li, L. Y., N. B. Shoemaker, and A. A. Salyers. 1993. Characterization of the mobilization region of a *Bacteroides* insertion element (NBU1) that is excised and transferred by *Bacteroides* conjugative transposons. J. Bacteriol. 177:4992-4999.

41. Lorenz, M. G. and W. Wackernagel. 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. 58:563-602.

42. Macrina, F. L., R. P. Evans, J. A. Tobian, D. L. Hartley, D. B. Clewell, and K. R. Jones. 1983. Novel shuttle plasmid vehicles for *Escherichia-Streptococcus* transgenic cloning. Gene. 25:145-150.

43. Macrina, F. L., and Archer, G. L. 1993. Conjugation and broad host range plasmids in streptococci and staphylococci. pp. 313-329. *In* D. B. Clewell (ed.), Bacterial Conjugation. Plenum Press, New York.

44. Martin, A. C., R. Lopez, and P. Garcia. 1996. Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1. J. Virol. 70:3678-3687.

45. McIver, K. S., S. Subarao, E. M. Kellner, A. S. Heath, and J. R. Scott. 1996. Identification of *isp*, a locus encoding an immunogenic secreted protein conserved among group A streptococci. J. Bacteriol. 64:2548-2555.

46. Morton, T. M., D. M. Eaton, J. L. Johnston, and G. L. Archer. 1993. DNA sequence and units of transcription of the conjugative transfer gene complex (*trs*) of *Staphylococcus aureus* plasmid pGO1. J. Bacteriol. 175:4436-4447.

47. Paces, V., C. Vlcek, and P. Urbánek. 1986. Nucleotide sequence of the late region of *Bacillus subtilis* phage PZA, a close relative of ø29. Gene. 44:107-114.

48. Pansegrau, W., G. Ziegelin, and E. Lanka. 1990. Covalent association of the *traI* gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation nick site. J. Biol. Chem. 265:10637-10644.

49. Pansegrau, W., D. Balzer, V. Kruft, R. Lurz, and E. Lanka. 1991. Common sequence motifs in DNA relaxases and nick regions from a variety of DNA transfer systems. Nucl. Acids Res. 19:3455.

50. Patton, J. G., S. A. Mayer, P. Tempst, and B. Nadal-Ginard. 1991. Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. Gens Dev. 5:1237-1251.

51. Poyart-Salmeron, C., P. Trieu-Cot, C. Carlier and P. Courvalin. 1989. Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn*1545*: homologies with other site-specific recombinases. EMBO J. 8:2425-2433.

52. Pozzi, G. and W. R. Guild. 1985. Modes of integration of heterologous plasmid DNA into the chromosome of *Streptococcus pneumoniae*. J. Bacteriol. 161:909-912.

53. Rees, C. E. D., D.E. Bradley, and B. M. Wilkins. 1990. Protein transfer into the recipient cell during bacterial conjugation: studies with F and RP4. Mol. Microbiol. 4:1199-1205.

54. Sabelnikov, A. G. 1994. Nucleic acid transfer through cell membranes: Towards the underlying mechanisms. Prog. Biophys. Molec. Biol. 62:119-152.

55. Salyers, A. A., N. B. Shoemaker, A. M. Stevens, and L.-Y. Li. 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol. Rev. 59:579-590.

56. Sambrook, J. E., E. F. Fritsch, and N. T. Maniatis. 1989. *Molecular Cloning. A Laboratory Manual*, 2nd edn: Cold Spring Harbour Laboratory Press.

57. Saraste, M., P. R. Sibbald, and A. Wittinghofere. 1990. The P-loop - a common motif in ATP- and GTP- binding proteins. Trends Biochem. Sci. 15:430-434.

58. Schuster, S. S., A. A. Noegel, F. Oehme, G. Gerisch, and M. I. Simon. 1996. The hybrid histidine kinase DokA is part of the osmotic response system of Dictyostelium. EMBO J. 15:3880-3889.

59. Scott, J. R., P. A. Kirchman, and M. G. Caparon. 1988. An intermediate in transposition of the conjugative transposon Tn*916*. Proc. Natl. Acad. Sci. USA. 85:4809-4813.

60. Scott, J. R. 1992. Sex and the single circle: Conjugative transposition. J. Bacteriol. 174:6005-6010.

61. Scott, J. R., F. Bringel, D. Marra, G. Van Alstine, and C. K. Rudy. 1994. Conjugative transposition of Tn*916*; preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. Mol. Microbiol. 11:1099-0008.

62. Scott, J. R. and G. G. Churchward. 1995. Conjugative transposition. Ann. Rev Microbiol. 49:367-397.

63. Shoemaker, N. B., and W. R. Guild. 1974. Destruction of low efficiency markers is a slow process occurring at a heteroduplex stage of transformation. J. Bacteriol. 128:283-290.

64. Shoemaker, N. B., M. D. Smith, and W. R. Guild. 1979. Organization and transfer of heterologous chloramphenicol and tetracycline resistance of genes in pneumococcus. J. Bacteriol. 139:432-441.

65. Shoemaker, N. B., M. D. Smith, and W. R. Guild. 1980. DNAse-resistant transfer of chromosomal *cat* and *tet* insertions by filter mating in pneumococcus. Plasmid. 3:80-87.

66. Shoemaker, N. B., G.-R. Wang, and A. A. Salyers. 1996. The *Bacteroides* mobilizable insertion element, NBU1, integrates into the 3' end of Leu-tRNA gene and has an integrase that is a member of the lambda integrase family. J. Bacteriol. 178:3594-3600.

67. Shoemaker, N. B., G.-R. Wang, and A. A. Salyers. 1996. NBU1, a mobilizable site-specific integrated element from *Bacteroides* spp., can integrate nonspecifically in *Escherichia coli*. J. Bacteriol. 178:3601-3607.

68. Sia, E. A., D. M. Kuehner, and D. H. Figurski. 1996. Mechanism of retrotransfer in conjugation: Prior transfer of the conjugative plasmid is required. J. Bacteriol. 178:1457-1464.

69. Smith, M. D. 1981. The genetics of multiple drug resistance in *Streptococcus pneumoniae.* Ph. D. Thesis. Duke University.

70. Soldatov, N. M. 1992. Molecular diversity of L-type Ca2+ channel transcripts in human fibroblasts. Proc. Natl. Acad. Sci. USA. 89:4628-4632.

71. Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

72. Thomas, W. D. and G. L. Archer. 1989. Identification and cloning of the conjugative transfer region of *Staphylococcus aureus* plasmid pGO1. J. Bacteriol. 171:684-691.

73. Thöny, B., D. S. Hwang, L. Fradkin, and A. Kornberg. 1991. *iciA*, an *Escherichia coli* gene encoding a specific inhibitor of chromosomal initiation of replication *in vitro.* Proc. Natl. Acad. Sci. USA. 88:4066-4070.

74. Triglia, T., H-D. Stahl, P. E. Crewther, A. Silva, F. R. Anders, and D. J. Kemp. 1988. Structure of a *Plasmodium falciparum* gene that encodes a glutamic acid-rich protein (GARP). Molecular and Biochemical Parasitology. 31:199-202.

75. Trimmer, J. S., S. S. Cooperman, S. A. Tomiko, J. Zhou, S. M. Rean, M. B. Boyle, R. G. Kallen, Z. Sheng, R. L. Barchi, F. J. Sigworth, R. H. Goodman, W. S. Agnew, and G. Mandel. 1989. Primary structure and functional expression of a mammalian skeletal muscle sodium channel. Neuron. 3:33-49.

76. Vicek, C., and V. Paces. 1986. Nucleotide sequence of the late region of *Bacillus phage* ø29 completes the 19,285-bp sequence of ø29 genome. Comparison with the homologous sequence of phage PZA. Gene. 46:215-225.

77. Vijayakumar, M. N., S. D. Priebe, G. Pozzi, J. M. Hageman, and W. R. Guild. 1986. Cloning and physical chracterization of chromosomal conjugative elements in Streptococci. J. Bacteriol. 166:972-977.

78. Vijayakumar, M. N., S. D. Priebe, and W. R. Guild. 1986. Structure of a conjugative element in *Streptococcus pneumoniae.* J. Bacteriol. 166:978-984.

79. Vijayakumar, M. N. and S. Ayalew. 1993. Nucleotide sequence analysis of the termini and chromosomal locus involved in site-specific integration of the streptococcal conjugative transposon Tn*5252*. J. Bacteriol. 175:2713-2719.

80. Walker, J. E., E. M. Sartse, M.J. Runswick, and N. J. Gay. 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, mysoin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1:945-951.

81. Wang, A., and F. L. Macrina. 1995. Characterization of six linked open reading frames necessary for pIP501-mediated conjugation. Plasmid. 34:206-210.

82. Wang, A. and F. L. Macrina. 1995. Streptococcal plasmid pIP501 has a functional *oriT* site. J. Bacteriol. 177:4199-4206.

83. Ziegelin, G., W. Pansegrau, B. Strack, D. Balzer, M. Kroger, V. Kruft, V. and E. Lanka. 1991. Nucleotide sequence and organization of genes flanking the transfer origin of promiscuous plasmid RP4. DNA Sequence. 1:303-327.

VITA

Francisco J. Alarcon-Chaidez

Candidate for the Degree of

Doctor of Philosophy

Thesis: GENETIC AND TRANSCRIPTIONAL ORGANIZATION OF A REGION ESSENTIAL FOR DNA TRANSPORT IN Tn*5252*

Major Field: Microbiology, Cell, and Molecular Biology

Biographical:

Personal Data: Born in Culiacan, Sinaloa, Mexico, October 9, 1960, son of Jose N. Alarcon and Beatriz Chaidez.

Education: Graduated from Instituto Tecnologico de Culiacan High School, Sinaloa, Mexico, in August 1978; received Bachelor of Science Degree in Biochemical Engineering from Instituto Tecnologico de Culiacan, Sinaloa, Mexico, in August 1982; received Master of Science Degree in Food Science and Technology from Texas A&M University, College Station, Texas, in August 1984; completed requirements for the Doctor of Philosophy Degree at Oklahoma State University in July 1997.

Professional Experience: Assistant Professor, Instituto Tecnologico de Veracruz, Veracruz, Mexico, September 1984 to December 1991. Research Assistant, Department of Animal Science, Oklahoma State University, January 1992 to July 1993. Research and Teaching Assistant, Department of Microbiology and Molecular Genetics, Oklahoma State University, August 1993 to May 1997.

Professional Memberships: American Society for Microbiology