CONSTRUCTION OF A GENE BANK AND USE OF THE CHROMOSOME WALKING TECHNIQUE FOR THE DETECTION OF NEW PUTATIVE AGROCIN GENES IN *Agrobacterium tumefaciens* STRAIN D286.

Gerardo Herrera

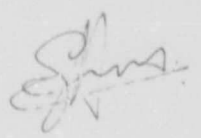
Submitted in partial fulfilment of the requirements for the degree of MSc. in the Faculty of Science, University of the Witwatersrand.

February 1987.

CONTENTS

DECLARATION

I declare that this dissertation is my own, unaided work, except where otherwise stated in the text. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any Degree or examination at any other University.



Gerardo Herrera

19th day of February, 1987

Acknowledgements

I wish to thank to my supervisors Professor J.A. Thomson and Ms. C.L.Botha for their enthusiastic support and encouragement of my efforts during this work and for suggesting the topic of this project as well as for making critically important suggestions for the approach to the problem.

I thank my wife Veronica for her moral and financial support and for her help during the typing of this project report.

ABSTRACT

A gene bank of *Agrobacterium tumefaciens* D286 wt has been constructed by cloning D286 wt DNA partially digested with EcoRI in the cosmid vector pLAFRI. The library; composed of 1750 members with a 27.7 kb average insert size was probed with pCDTn5-3, a cosmid vector carrying a D286:: Tn5 insert from the strain D286:: Tn5 Ag-. In this manner one recombinant cosmid of the library, pCD0932, was detected. The insert DNA of pCD0932 has sequences homologous to the D286:: Tn5 insert of pCDTn5-3, therefore it carries putative wt agrocin D286 genes (or sequences affecting agrocin D286 production). The insert DNA of pCD0932 was isolated and used to probe the D286 wt gene library. This first step in chromosome walking resulted in the detection of pCD2375. EcoRI restriction digestions and DNA homology studies of pCD0932 and pCD2375 showed that their D286 wt inserts are both composed of 4 EcoRI DNA sub-fragments totalling 21.8 and 24.8 kb respectively, with an overlapping sequence extending 3.5 kb.

In order to overcome the failure to detect *A. tumefaciens* cells transformed with pCD0932, attributed to Tc^r (conferred by pCD0932) not being an efficient selectable marker in most *Agrobacterium* strains, we constructed vector pSUP204-1. Such vector has been derived from pSUP204 which we slightly altered by cloning into it a 700 bp λ DNA SalI fragmet. This resulted in insertion inactivation of the Tc^r gene, a feature that, in addition to its Amp^r marker, allows the use of pSUP204-1 as a subcloning vector in conjugations and transformations involving pCD0932 or pCD2375 and strains D286:: tn5 Ag- and C58 C1G.

Preliminary work on the characterisation of pCD0932 and pCD2375 through restriction endonuclease mapping indicate that in addition to EcoRI, SalI and BglII will be useful enzymes since they produce a reasonable spread of not too numerous DNA fragments.

As a result of this work two recombinant cosmids bearing D286 wt DNA inserts, at least one of which (pCDO932) contains DNA sequences putatively affecting agrocin D286 production, are now available for further genetic manipulations. pSUP204-1 should prove useful as a subcloning vector for transformations and conjugations involving recombinant cosmids from the D286 wt gene bank and *Agrobacterium* strains. Future work on the molecular biology of agrocin D286 production is discussed.

Abbreviations

Amp	ampicillin
Cm	chloramphenicol
cpm	counts per minute
CSIR	Council for Scientific and Industrial Research
dCTP	deoxycytosine triphosphate
dpm	desintegrations per minute
hs	hours
kb	kilobase
λ	lambda phage
λHE or λIII	lambda Hind III/EcoRI digest
λH orλII	lambda Hind III digest
LB	Luria broth
LMCB	Laboratory for Molecular and Cell Biology (Johannesburg)
NA	nutrient agar
Nal	nalidixic acid
nm	nanometres
Nm	[illegible]omycin
rpm	revolutions per minute
TBE	Tris borate EDTA buffer
Tc	Tetracycline
TE	Tris EDTA buffer
U	unit
UV	ultraviloet
V	volt
wt	wild type

LIST OF TABLES

LIST OF FIGURES

1 LITERATURE REVIEW OF CROWN GALL DISEASE

1.1 SUMMARY

A. tumefaciens causes crown gall disease in a wide range of plants (De Cleene and De Ley, 1976). The disease is characterised by the formation of tumors called crown galls (Brown and White, 1943). Crown gall cells are able to grow in the absence of phytohormones (Brown, 1958) and produce opines (Tempe and Goldman, 1982) which can be catabolized by virulent bacteria (De Greeve *et al.*,1984; Holsters *et al*,1980) and may induce conjugal transfer of Ti-plasmids (Ellis *et al.*,1982; Petit *et al.*,1978). The events leading to plant tumor formation are, chronologically, as follow : (i) Agrobacteria penetrate and multiply at wounds sites, attachment seems to be an essential step in tumor induction (Douglas *et al.*,1985).(ii) The T-region of the Ti plasmid then becomes integrated in the plant nuclear genome (Chilton *et al.*,1977). (iii) In the plant cells the T-DNA is transcribed in a characteristic set of different mRNAs (Willimitzer *et al.*,1983) some of which are translated into enzymes that mediate production of opines while others determine enzymes that are involved in the production of the phytohormones indole acetic acid (an auxin) (Garfinkel *et al.*,1981; Ooms *et al.*,1981) and isopentyl AMP (a cytokinin) (Barry *et al.*,1984; Akiyoshi *et al.*,1984).

Besides the T-region the only other part of the Ti-plasmid necessary for tumor induction is the virulence region (Yanofsky *et al.*,1985; Stachel and Zambrisky, 1986) but during crown gall induction the virulence region is not integrated in the plant nuclear genome. The Ti plasmid must be, therefore, processed so that only the T-region is inserted. The *vir* genes, which are expressed in the bacterium only after induction factors present in plant exudates (Stachel *et al.*,1985), play a role in T-region

processing and transfer (Bolton *et al.*,1986). A physical linkage between T-region and virulence region is not necessary for T-region transfer (Schell and Montagu, 1983). The border sequences surrounding the T-region contain 24 bp direct repeats that are recognition signals for T-DNA transfer, the right border being essential for transfer and integration (Shaw *et al.*,1984; Wang *et al.*,1984; Joos *et al.*,1983). Recently described sequences called *overdrive* act as transfer enhancers in octopine Ti-plasmids (Peralta *et al.*, 1986). Since the inner region of the T-region is not necessary for T-DNA integration , it can be deleted eliminating oncogene sequences encoding opine production. This has revolutionised the development of Ti-plasmid derived plant gene vectors some of which have been successfully used for expression of foreign genes (using the NOS promoter sequences) in transformed plant cells (Herrera-Estrella *et al*, 1983).

The host range of crown gall is perhaps the widest among bacterial plant diseases, dicotyledoneous plants are most susceptible (De Cleene and De Ley, 1976) but other families are also genetically colonised by *A. tumefaciens*. The host range of crown gall is determined , at least partly, by the Ti-plasmid (Loper and Kado, 1979; Thomashow, 1980) and oncogene 4 (*cyt*:cytokinin) is clearly involved (Buccholz and Thomashow, 1984; Hoekema *et al.*,1984) but the determinants of host range have not yet been entirely defined.

Ti-plasmids contain the necessary information for the catabolism of opines which seemingly enter the bacterium's cell via a high affinity active transport system involving a periplasmic permease (Murphy and Roberts, 1979). *A. radiobacter* 84, in turn produces agrocin 84, a bacteriocin-like antibiotic which acts as an illegitimate substrate for the high affinity uptake system of nopaline type plasmids (Murphy *et al*, 1979) therefore inhibiting virulent strains carrying such Ti-plasmids. *A. radiobacter*, in addition, competes for attachment sites which appar-

ently produces some degree of biological control as well. *A. radiobacter* 84 has proven effective for the biological control of crown gall in several situations around the world. However, agrocin 84 does not inhibit virulent *Agrobacterium* strains harboring octopine type plasmids and not all nopaline producing strains are controlled either. Crown gall from grapevines can not be controlled by *A. radiobacter* 84 (Thomson, 1986). In addition, breakdown of biological control can occur (Kerr and Tate, 1983), therefore in some laboratories efforts have been directed into discovering strains producing agrocins with broader and/or different host ranges than that of *A. radiobacter* 84. In this regard, progress has been made in the LMCB (Johannesburg). Studies on *A. tumefaciens* strains J73 and D286 have been the most successful. Agrocin D286 has a broader host range , at least *in vitro*, than agrocin 84 (Hendson *et al.*,1983). *A. tumefaciens* J73 can inhibit crown gall *in vitro* in grapevines.

1.2 HOST RANGE OF CROWN GALL AND BIOTYPES OF A. TUMEFACIENS

A review published by De Cleene and De Ley (1976), which is to date the most extensive source of information on crown gall susceptibility of plants, indicate that no lower plant (*Fungi*, *Briophyta*, *Pteridophyta*) is known to be a host for crown gall but 60% of the gymnosperms and the dicotyledonous angiosperms examined were sensitive to crown gall. The gymnosperms contain at least two classes which are susceptible to crown gall; the class *Coniferopsida* and the class *Taxopsida*. The angiosperms, in particular dicotyledons, are good hosts for crown gall; 84 families are susceptible. Monocotyledons are largely nonsusceptible; only three families, belonging to the orders *Liliales* and *Arales* were found to be susceptible. The list of plants tested, it should be noted, is a minimal fraction of the tracheophytes (vascular plants), which amount to more than 200.000 species.

The distribution of crown gall is world-wide. Before the advent of biological control of crown gall the disease caused considerable economic losses in most countries exporting fruits, parts of America, Australia, Greece and Europe were particularly affected (Schroth *et al.*,1971; Alconero, 1980; Lehoczky, 1978; Panagopoulos *et al.*,1983; Kerr, 1969).

The genus *Agrobacterium* in Bergey's Manual (Allen and Holding, 1974) is speciated primarily according to phytopathogenic effects. Strains causing crown gall are placed in *Agrobacterium tumefaciens*, those causing hairy-root in *A. rhizogenes*, those causing cane gali on *Rubus spp.* in *A. rubi* and non pathogens in *A. radiobacter* and *A. tumefaciens* are not distinguished from each other except for their phytopathogenic effect. Although it has been proposed that the species name *A. radiobacter* be rejected (Holmes and Roberts, 1981) it is still used, seemingly for the sake of simplicity.

What is more important for the workers who d[illegible]*robacterium* strains is the discrimination of the bacterial group [illegible]he different existing biotypes. Biotyping is based on a number of biochemical diagnostic characters put forward by Kerr and Panagopoulos (1977) (Table 1.1). There are three biotypes currently recognised and although the biochemical characteristics that define them are not determined by the Ti-plasmid, there is a correlation between the biotype of virulent strains and their host range. biotype 1 has a broad host range , biotype 2 is commonly found in stonefruit and biotype 3 has a host range limited almost exclusively to grapevines (Kerr and Panagopoulos, 1977; Panagopoulos *et al*, 1978).

1.3 TI-PLASMID OF A. TUMEFACIENS

1.3.1 SUMMARY

Virulent strains of *A. tumefaciens* carry large plasmids (Zaenen *et al.*, 1974) essential for oncogenicity (Van Larebeke *et al.*, 1974; Watson *et*

	3-keto-lactose	2% NaCl	Max. growth temp.	Litmus milk	Acid from		Alkali from			Selective media of	
					erythritol	melezitose	malonate	L-tartrate	propionate	Schroth et al.	New and Kerr
Biotype 1	–	+	37	Alk.→Rdn.	–	+	–	–	–	–	–
Biotype 2	–	–	29	Acid	+	–	+	+	–	–	+
Biotype 3	–	+	35	Alk.	–	–	+	+	–	–	–

TABLE 1.1 Diagnostic characters for determination of biotypes
From: Kerr and Panagopoulos.1977. Phytopath. Z., 90: 172-179.

al., 1975). These large plasmids (generally more than 200 kb in size) are called tumor inducing (Ti) plasmids. Gene transfer from *A. tumefaciens* to the plant tumor cell was demonstrated by Chilton *et al.* (1977) who reported the presence of a fragment of the Ti-plasmid in crown gall cells. Further studies (Chilton *et al.*, 1980) showed that the transferred Ti-DNA fragment became integrated in the plant nuclear DNA. The fragment is called T-region as part of the Ti-plasmid and T-DNA as part of the plant tumor cell. The information necessary for tumor induction is carried in the T-region, which in addition bears genes that direct the synthesis of unique compounds known as opines (Petit *et al.*, 1978). Opines are produced in the tumor cells and since different Ti-plasmids determine the synthesis of different opines (of which nopaline, octopine and agropine are the most common), Ti-plasmids are classified into groups depending on the type of opines synthesis that their corresponding T-DNA directs in the transformed plant cells. In addition, Ti plasmids have information coding for the catabolism of the opines whose synthesis their T-region DNA induce. The Ti-plasmid can ,therefore, be considered an unusual type of catablic plasmid which induces the synthesis of its own substrate in transformed plant cells. The production of opines in plentiful supply is ensured by oncogenes, also borne in the T-region of the Ti-plasmid, which once integrated in the plant cells nuclear genome direct the synthesis of plant growth regulator substances (Garfinkel *et al.*, 1981; Ooms *et al.*, 1981), resulting in the neoplastic growth of opine producing cells. Sensitivity to bacteriocin-like antibiotics known as agrocins is also (with one reported exception,; Webster *et al.*, 1986) determined by the Ti-plasmid.

1.3.2 TRANSFER OF TI-PLASMIDS AND INTEGRATION OF T-REGION

As early as 1956 (Stonier) it was suggested that *A. tumefaciens* enters into a very close association with the host cells which can not be dem-

onstrated by standard bacteriological or cytological techniques. More than twenty years later (Chilton *et al.*, 1977) it was shown that the molecular basis of crown gall tumorigenesis was the stable incorporation of plasmid DNA into the plant cells. Restriction endonuclease digestions and DNA hybridization techniques were then used to conclude that highly conserved DNA of the Ti-plasmid overlapped the T-DNA maintained in plant tumors (Chilton *et al.*, 1978). The T-DNA was shown to be located in the nucleus of crown gall plant cells (Willimitzer, 1980; Chilton *et al.*, 1980). Yadav *et al.* (1980) isolated by molecular cloning a border fragment T-DNA and flanking plant DNA from a crown gall and showed that the T-DNA was covalently joined to DNA of the host plant nuclear DNA. In a more complete study Thomashow *et al.* (1980b) isolated the Ti-plasmid sequences from an octopine producing crown gall tumor by molecular cloning using the bacteriophage λ vector Charon 4A. Analysis of the DNA fragments indicated that the plasmid sequences were covalently joined to plant nuclear DNA. Thus, they demonstrated that genetic recombination of unusual nature, since it involved an eukaryotic and a prokaryotic organisms, had taken place. It was later shown by Thomashow *et al.* (1980a)that several tumor lines contained a "core" T-DNA which was co-linear with the Ti-plasmid.

With the underlying hypothesis that the T-DNA is a discrete physical and genetic element, several authors have contributed to the identification and physical and functional characterization of the DNA sequences at its edges. The T-DNA in several independent nopaline tumor lines is a 23 bp DNA segment co-linear with the corresponding T-region in the T-i plasmid (Lemers *et al.*, 1980). In octopine Ti-plasmids at either end of the T-region there is a 25 bp repeated element (Yadav *et al.*, 1982).These repeats signal the T-DNA borders since T-DNA ends occur in or near these repeats in several different tumors (Holsters *et al.*, 1983; Zambrisky *et al.*, 1982). Deletions removing the right border of the T-region attenuate virulence drastically in most plants even though the tumor maintenance

genes remain intact (Shaw *et al.*, 1984; Wang *et al.*, 1984; Joos *et al.*, 1983). Wang *et al.* (1984) showed that T-DNA transmission requires only a right border repeat in *cis* but Ti-plasmid sequences that lie to the right of the right border repeat stimulate its function (Jen, 1986). Wang *et al* . (1984) observed that efficiency of tumor formation is highly influenced by the orientation of the 25 bp right border repeat, indicating that the transfer and /or integration induced by the sequence is strongly polar. They suggested that this would explain why the right border is essential for T-DNA transfer and the left border is not. in the absence of the right border, the left border acting in a polar fashion would only direct transfer leftward, away from the T-DNA.

In the work described so far on T-DNA transmission it was not possible to distinguish between transfer and integration since the assay system involved the formation of a tumor which requires both processes to occur. This problem was overcome by Horsh and Klee (1986) who used a binary vector transformation system developed by Fraley *et al.* (1985), based on T-DNA borders and a selectable marker. Horsh and Klee (1986) using this system were able to categorically demonstrate that the 25 bp border right border sequence is necessary and sufficient to direct the transfer to plant cells and subsequent integration of the T-DNA. In addition these authors observed that a binary vector such as pMON508 (11kb) containing the left border would be transferred and lead to stable cell transformation, in contrast to what occurred with the large (250 kb) co-integrate vector pTiB6S35E (also containing only a left border). They suggested that size was the critical factor determining efficient transfer of pMON508. A single border is perhaps sufficient to permit Ti-plasmid transfer but the transfer would be inefficient if the plasmid was large. Thus, Horsh and Klee (1986) suggest that the primary role of a circular T-DNA intermediate described by Kaukolikova-Nicola *et al.* (1985) may be to reduce the size of the Ti-plasmid to a more manageable size than 250 kb.

The octopine plasmid pTiA6NC contains two adjacent but non-contiguous T-regions designated TL and TR (for left and right T-DNA). The left T-DNA contains all the genes required for tumor maintenance (Thomashow *et al.*, 1980). In octopine-type tumors a 13.5 kb TL-DNA is always present, and an additional 6-7 kb TR-DNA can some times be found (Thomashow *et al.*, 1980). In contrast to the situation in nopaline plasmids where the right border alone promotes efficient T-DNA transfer and integration, in the octopine-type Ti-plasmid pTiA6NC efficient T-DNA transmission requires 2 discrete sequences, a 23-bp T-DNA right border repeat and a second sequence (*overdrive*) lying to the right of the right border repeats (Peralta *et al.*, 1986). The 23-bp T-DNA border repeat promotes transmission only weakly, *overdrive* by itself does not promote transfer at all. Peralta *et al.* (1986) have identified a region to the right of the nopaline- type plasmid pTi37, a right border repeat which acted similarly to *overdrive* when an octopine-type plasmid supplied the *vir* functions but acted only weakly when the functions were provided by a nopaline plasmid. Since there was homology of this region with the core sequences of *overdrive* the authors concluded that *overdrive* is necessary when an octopine-type Ti-plasmid provides the *vir* genes but not when they are provided by the nopaline Ti-plamids. Based on the suggested function of some of the Ti-plasmid-carried *vir* genes, namely to encode proteins, in response to induction from plant exudates, which act at the right border repeat to initiate T-DNA transmission (Stachel *et al.*, 1985), it has been proposed that *overdrive* enhances interaction between the right border and the appropriate *vir* proteins. Therefore, most T-DNA transmission events will initiate at the right border and move leftward through the DNA in a similar fashion as has been suggested for nopaline-type Ti-plasmids.

1.3.3. ONCOGENES OF THE TI PLASMID

The T-DNA genes of Ti-plasmids can induce tumorous growth on their host plants. Studies on nopaline and octopine Ti-plasmids (the most studied

Ti-plamids) have shown that although they are distantly related their T-regions share a highly homologous 9 kb sequence involved in oncogenicity (*onc*) (Engler *et al.*, 1981). This common DNA includes the T-DNA genes 1,2,4,6a and 6b (Willimitzer *et al.*, 1983). At least three genes of the common DNA code for specific enzymes that are involved in phytohormone synthesis. Tryptophan 2-monooxygenase and indoleacetamide hydrolase, two enzymes that convert tryptophan to indole acetic acid (the major naturally occurring auxin), are encoded by the "auxin" genes 1 and 2. Their combined effect is auxin-like since mutants in genes 1 or 2 (*Shi*$^-$: *Sh*oot *i*nhibition) produce tumors that sprout shoots (Garfinkel *et al.*, 1981; Ooms *et al.*, 1981). Gene 4 (*cyt*) encodes isopentenyltransferase, an enzyme that catalyzes the first step in cytokinin biosynthesis (Barry *et al.*, 1984; Akiyoshi *et al.*, 1984). Mutants in gene 4 (*Roi*$^-$: *R*oot *i*nhibition) produce tumors that sprout shoots (Joos *et al.*, 1983; Garfinkel *et al.*, 1981). In analogy to the effect of plant hormones on non-transformed plant tissue (Skoog and Miller, 1957). The effect of gene 4 can be described as "cytokinin-like". From the foregoing it follows that the undifferentiated appearance of a crown gall tumor and its independence of exogenous hormones for continuous growth is a reflection of the combined activities of genes 1,2 and 4.

It is well known that the T-region is transferred to plants and becomes stably incorporated in the nuclear DNA where at least some of the several loci defined so far are transcribed (Willimitzer et al., 1983). What is not clear is the origin of the loci in the T-region. White et al. (1982) in homology studies using T-region DNA probes covering the entire T-region of *A. rhizogenes* showed that there exists at least one copy of homologous DNA sequence in the nuclear DNA of the uninfected, normal tobacco plant cells. It was later shown, in addition, that Ti-plasmids of *A. tumefaciens* and *A. rhizogenes* had homologous T-region sequences for these probes. It has been suggested that cellular DNA sequences homologous to T-DNA oncogenes appear to have been maintained as endogenous cellular plant

genes during evolution (Yang, 1985). An interesting prospect, if this is true, is that cDNA probes of specific T-DNA phytohormone genes could be used in selecting and identifying the cellular auxin and cytokinin gene sequences from plants. Morover, Yang (1985) suggests that these primary gene sequences could then be used as probes for isolating additional cellular DNA still coding for enzymes or proteins involved in auxin and cytokinin activity which could have diverged through evolution from original sequences.

1.3.4 PRODUCTION AND ROLE OF OPINES

Opine is the generic term for compounds produced by plant cells which have been transformed by T-DNA. They serve as nutrients for agrobacteria responsible for inciting their production and their catabolism requires enzymes determined by sequences in the Ti-plasmid outside the T-region. Messens *et al*, (1985) defines an opine as a compound ratisfying the following criteria. It is (i) synthesised by transformed but not normal plant cells, (ii) made available to the colonising agrobacteria and (iii) metabolized by the inciting bacteria.

Genes for opine biosynthesis have been found to map in the T-region DNA (Garfinkel *et al.*, 1981; Hernalsteens *et al.*,1980; Murai *et al.*, 1982) whereas genes for opine catabolism map outside the T-region DNA (De Greeve *et al.*, 1981; Holsters *et al.*, 1980). The T-region DNA encoding nopaline and octopine has been sequenced and found to possess a eukaryotic gene structure, despite their prokaryotic origin (Bevan *et al.*, 1983; De Greeve *et al.*, 1983).

Octopine Ti-plasmids carry genes that determine the biosynthesis of octopine, octopinic acid (Menage and Morel, 1964,1965), lysopine (Lioret 1957, as quoted by Ellis *et al.*, 1982) and histopine (Kemp, 1977). This opines result from reductive condensation of piruvic acid with arginine, ornithine lysine and histidine respectively. Octopine Ti-plasmids can also determine the synthesis of agropine (Firmin and Fenwick, 1978) which is a derivative of mannose and glutamine. Nopaline Ti-plasmids determine the synthesis of nopaline (Goldman *et al*, 1969) and nopalinic acid (Firmin and Fenwick, 1977). These opines result from the reductive condensation of α-ketoglutaric acid with arginine and ornithine respectively. Some Ti-plasmids formerly known as "unusual nopaline" plasmids (because strains carrying them were able to utilise nopaline but did not produce any opines) have recently been shown to determine the synthesis of