MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 3, No. 3, pp. 174-181, 1990

Analysis of the C_4-Dicarboxylate Transport Genes of *Rhizobium meliloti*: Nucleotide Sequence and Deduced Products of *dctA*, *dctB*, and *dctD*

Robert J. Watson

Plant Research Centre, Agriculture Canada, Ottawa, Ontario K1A OC6, Canada.
Received 24 August 1989. Accepted 2 January 1990.

Rhizobium meliloti **transports succinate, fumarate, malate, and aspartate by means of the dicarboxylate transport system, which is encoded by *dct* genes located on the *exo* megaplasmid. Analysis of these genes using Tn*5* insertion mutagenesis revealed three complementation groups within a 5.9-kb *Hin*dIII fragment. The sequence of this fragment and the sites of Tn*5* insertion were determined. Three genes, *dctA*, *dctB*, and *dctD*, were identified as the only three open-reading frames in locations consistent with the complementation data. The *dctA* gene is preceded by the sequence CTGGCACG-N_4-TTGCT, which is characteristic of promoters requiring the *ntrA*-encoded protein for activation. The *dctA*-encoded protein is highly hydrophobic and contains eight potential transmembrane helices, indicating that it is probably the structural component of the transport system responsible for movement of dicarboxylates from the periplasm across the inner membrane. The *dctB* and *dctD* genes are transcribed in the opposite direction to *dctA*. They encode proteins with homology to the *R. leguminosarum* bv. *viceae* dicarboxylate transport proteins regulating expression of *dctA* and to other proteins comprising two-component regulatory systems. The *dctB*-encoded protein includes a putative periplasmic N-terminal domain that senses the presence of dicarboxylates and a C-terminal cytoplasmic domain that activates the *dctD*-encoded protein. The C-terminus of the *dctD*-encoded protein shows homology to several DNA-binding proteins, indicating that it is probably the domain which binds DNA in the *dctA* promoter region to regulate *dctA* transcription. All the *R. meliloti* mutants altered in *dctA*, *dctB*, and *dctD* were complemented by the *dct* region from *R. l.* bv. *viceae*.**

Additional keywords: *Medicago sativa*, membrane-associated proteins, symbiotic nitrogen fixation.

Rhizobium meliloti Dangeard fixes nitrogen in a symbiotic interaction within nodules formed on the roots of the legume *Medicago sativa* L. (alfalfa). Bacteria within the nodule, known as bacteroids, are dependent upon the plant for all their metabolites, and in exchange they provide nitrogen to the plant. The source of the energy for nitrogen fixation is plant photosynthate. It is probably provided to the bacteroids in the form of one of the C_4-dicarboxylates. The dicarboxylates succinate, fumarate, and malate have been implicated in the nitrogen fixation process by studies of carbon source utilization by purified bacteroids (Bergersen and Turner 1967; Miller *et al.* 1988) and by analysis of the symbiotic effectiveness of mutants lacking the ability to utilize these compounds (Ronson *et al.* 1981).

C_4-Dicarboxylates are transported into rhizobia by the dicarboxylate transport system, which is induced in the presence of succinate, fumarate, or malate. In *R. meliloti* the dicarboxylate transport (*dct*) genes have been cloned by cosmid complementation of mutants unable to grow on these compounds due to an inability to transport them into the cell (Bolton *et al.* 1986; Engelke *et al.* 1987; Watson *et al.* 1988; Yarosh *et al.* 1989). The *dct* genes have been localized to a 5.9-kilobase (kb) *Hin*dIII fragment located on the *exo* megaplasmid. Mutations within this region result in an inability to transport succinate, fumarate, malate, or aspartate, together with either of two symbiotic phenotypes: one type of mutant produces nodules in which the bacteria differentiate into bacteroids that appear normal, but senesce rapidly and do not fix nitrogen; the second type, resulting from mutations within regulatory genes, produces nodules that are effective, but less than the wild type ($Fix^{+/-}$). The requirement for the dicarboxylate transport regulatory genes appears to be bypassed in the nodule, possibly by other regulatory genes that are symbiosis-specific.

In *R. leguminosarum* bv. *viceae* Jordan, a cluster of three genes has been identified encoding structural and regulatory proteins necessary for the uptake of dicarboxylates in free-living cells, and a model for the functions of these genes has been proposed (Ronson *et al.* 1987a). The structural gene, which encodes a membrane protein, is *dctA* (Ronson and Astwood 1985). Expression of *dctA* is regulated by the products of the *dctB* and *dctD* genes, (DctB and DctD proteins, respectively), which are produced constitutively at a low level. DctB contains a periplasmic component that senses the presence of dicarboxylate and then activates DctD on the cytoplasmic side of the inner membrane. Transcription of *dctA* is stimulated by activated DctD. In addition to the *dct* regulatory genes, *dctA* expression requires the product of the *ntrA* (also called *rpoN*) gene, NtrA (Ronson *et al.* 1987b). The NtrA protein is a sigma factor that binds to the core RNA polymerase and permits

Nucleotide and/or amino acid sequence data is to be submitted to GenBank as accession number JO3683.

transcription from promoters of a type having the consensus sequence −26 CYGGYAYR-N_4-TTGCA −10 (Gussin *et al.* 1986).

The *dctB* and *dctD* genes belong to a class of two-component regulatory systems in which a sensor component obtains information from the environment, such as the presence of metabolites, and transmits the information to a regulatory component (Ronson *et al.* 1987c). The signal is transmitted through an interaction between a C-terminal domain of the sensor protein and the N-terminal domain of the regulator protein. This triggers a change in the regulator protein, perhaps through phosphorylation or a change in conformation, that switches it to an activated state. In this state, the protein enhances the expression (transcription) of genes able to utilize, or respond to, the original environmental stimulus. Two-component regulatory systems other than the *dctB/dctD* genes are *Escherichia coli* Migula (Castellani and Chalmers) genes responding to nitrogen limitation (*ntrB/ntrC*), phosphate limitation (*phoR/phoB*), and osmolarity changes (*envZ/ompR*), and *Agrobacterium tumefaciens* (Smith and Townsend) Conn genes responsive to plant exudate (*virA/virG*).

In this study, the *dct* region of *R. meliloti* has been analyzed to determine the number and location of its genes. The nucleotide sequence of the three dicarboxylate transport genes, *dctA, dctB*, and *dctD*, was determined, and the amino acid sequences of their gene products were deduced. The *dctB* and *dctD* genes were found to be very similar to the corresponding genes in *R. l.* bv. *viceae*, as expected from previous studies of *dct* mutants, and to contain protein domains homologous to those characteristic of two-component regulatory systems and DNA-binding proteins.

MATERIALS AND METHODS

Strains and plasmids. Bacterial strains and plasmids are listed in Table 1. *R. meliloti* strains were grown at 30°

Table 1. Bacterial strains and plasmids used in this study

Designation	Relevant characteristics[a]	Source or reference
Rhizobium meliloti		
JJ1c10	Wild-type parent, derivative of IZ450; Dct^+ Fix^+ Rif^r	Selvaraj *et al.* 1987
4F6	*dctA* mutant derived from JJ1c10 by Tn*5* mutagenesis; Dct^- Fix^- Rif^r Km^r	Watson *et al.* 1988
R646, R647	*dctA*::Tn*5-51*, *dctA*::Tn*5-35*, JJ1c10 derivatives containing genomic Tn*5* inserts introduced by recombination; Dct^- Fix^- Rif^r Km^r	Watson *et al.* 1988; this study
R639, R645, R648	*dctB*::Tn*5-33*, *dctB*::Tn*5-15*, *dctB*::Tn*5-20*, JJ1c10 derivatives containing genomic Tn*5* inserts introduced by recombination; Dct^- $Fix^{+/-}$ Rif^r Km^r	Watson *et al.* 1988; this study
R649, R650, R651	*dctD*::Tn*5-39*, *dctD*::Tn*5-22*, *dctD*::Tn*5-52*, JJ1c10 derivatives containing genomic Tn*5* inserts introduced by recombination; Dct^- $Fix^{+/-}$ Rif^r Km^r	Watson *et al.* 1988; this study
R652	Tn*5-11* in ORF distal to *dctD* introduced into JJ1c10 by recombination; Dct^+ Fix^+ Rif^r Km^r	Watson *et al.* 1988; this study
4R21	*ntrA* mutant derived from JJ1c10 by Tn*5* mutagenesis; Dct^- Fix^- Rif^r Km^r	Selvaraj *et al.* 1987; this study
Escherichia coli		
HB101, DH5α	*E. coli* host strains	Boyer and Roulland-Dussoix 1969; J. Jessee[b]
Plasmids		
pRK2013	ColE1 replicon containing RK2 transfer genes; Km^r	D. R. Helinski; Univ. of Calif., San Diego, La Jolla; Fugurski and Helinski 1979
pLAFR1	Broad host range cosmid cloning vehicle; Tc^r	F. M. Ausubel, Harvard Medical School, Boston; Friedman *et al.* 1982
pRK310	Broad host range cloning vehicle; Tc^r	G. Ditta, Univ. of Calif., San Diego, La Jolla; Ditta *et al.* 1985
pUC8, pUC18, PUC19	*E. coli* cloning vehicles; Ap^r	Yanisch-Perron *et al.* 1985
pBR322	*E. coli* cloning vehicle; Tc^r Ap^r	Bolivar *et al.* 1977
pBB107	pLAFR1 cosmid carrying 23 kb of the *exo* megaplasmid pRmeJJ1c10b, including the *dct* region; Tc^r	Watson *et al.* 1988
pBB122	pRK310 carrying the 5.9-kb *Hin*dIII fragment containing the *dct* region; Tc^r	Watson *et al.* 1988
pBB127	5.9-kb *Hin*dIII fragment containing the *dct* region cloned into pUC8; Ap^r	This study
pBB128	5.9-kb *Eco*RI fragment of the *dct* region cloned in *Eco*RI site of pBR322; Tc^r Ap^r	This study
pBB129	2.9-kb *Eco*RI fragment of the *dct* region cloned in the *Eco*RI site of pBR322; Tc^r Ap^r	This study
pBB124	0.6-kb *Xho*I fragment of the *dct* region cloned in the *Sal*I site of pUC18; Ap^r	This study
pBB125	1.6-kb *Pst*I fragment of the *dct* region cloned in pUC18; Ap^r	This study
pBB126	1.0-kb *Pst*I fragment of the *dct* region cloned in pUC18; Ap^r	This study
pBB123	pLAFR1 cosmid carrying the *ntrA* gene of *R. meliloti*, complements 4R21; Tc^r	This study
pPN108	pLAFR1 cosmid carrying the *R. leguminosarum* bv. *viceae dctA*, *dctB*, and *dctD* genes in adjacent 4.4- and 4.8-kb *Eco*RI fragments; Tc^r	C. W. Ronson, Biotechnica Internatl., Cambridge, MA; Ronson *et al.* 1984

[a] Rif, rifampicin; Km, kanamycin; Tc, tetracycline; Ap, ampicillin; ^{r}, resistant; ORF, open-reading frame; and kb, kilobase.

[b] J. Jessee. New subcloning efficiency competent cells: $> 1 \times 10^8$ transformants/μg. Pages 8-10 in: Focus, Vol 8, No. 4, 1986. Bethesda Research Laboratories, Gaithersburg, MD.

C on tryptone-yeast extract medium containing $CaCl_2$ (Beringer 1974) as a rich medium or on M9 minimal media (Maniatis *et al.* 1982) supplemented with 20 mM mannitol or glucose as the carbon source. Carbon sources used to assay the Dct phenotype were 20 mM succinate, fumarate, malate, and aspartate. Yeast extract was included at 50 μg/ml in liquid media. Antibiotics used with *R. meliloti* were rifampicin (100 μg/ml), tetracycline (5 μg/ml), gentamycin (25 μg/ml), and kanamycin (40 μg/ml). *E. coli* strains were grown at 37° C on Luria-Bertani medium (Maniatis *et al.* 1982); antibiotics used were ampicillin (20 μg/ml), tetracycline (10 μg/ml), kanamycin (20 μg/ml), and streptomycin (100 μg/ml).

Bacterial matings and DNA manipulations. Triparental matings including the helper plasmid pRK2013 were used to transfer plasmids into *R. meliloti* (Ditta *et al.* 1980). DNA digests, ligations, and transformations were done as described by Maniatis *et al.* (1982). Localization of the *ntrA* region to the chromosome was done by hybridization using pBB123 as described previously for pBB107. Other genetic manipulations, plant growth, and acetylene reduction assays have also been described (Watson *et al.* 1988).

DNA sequencing. The nucleotide sequence of the 5.9-kb *Hin*dIII fragment was determined by the dideoxynucleotide sequencing method (Sanger *et al.* 1977). Modified T7 DNA polymerase (Sequenase, U.S. Biochemical Corporation, Cleveland, OH) was used to incorporate [α-^{35}S]dATP and [α-^{35}S]dGTP in the sequencing reactions. Covalently closed circular double-stranded DNA was prepared for use as the sequencing template as described by Chen and Seeburg (1985). Sequencing was done primarily using pBB127, containing the entire 5.9-kb *Hin*dIII fragment, as the template. A sequence was first obtained using M13 universal and reverse sequencing primers (New England Biolabs, Beverly, MA). This sequence was then used to choose primers, which were synthesized using a Biosearch 8750 DNA Synthesizer, to further extend the known sequence. This process was repeated in a stepwise manner until the entire fragment was sequenced on both strands. Plasmids pBB125, pBB126, pBB129, pBB124, and pBB128, containing cloned internal *Pst*I, *Xho*I, and *Eco*RI fragments of the 5.9-kb *Hin*dIII fragment, were used to obtain internal sequences for primer synthesis such that sequencing could be conducted simultaneously from multiple sites.

The insertion sites of Tn*5* were determined after subcloning *Eco*RI fragments from pBB107::Tn*5* derivatives into pUC18 or pUC19 plasmids and selecting ampicillin- and kanamycin-resistant transformants to obtain those with the transposon. These were used directly for DNA sequencing of insertion sites for Tn*5-15*, Tn*5-22*, and Tn*5-39* using primers to adjacent *R. meliloti* DNA. The insertion sites for the other derivatives were determined after removing one Tn*5* arm from each derivative by excision of a *Bam*HI fragment, and then using an internal Tn*5* sequence 60 nucleotides within its arms, TTACCATGTTAGGAGG (16mer), as a primer. The orientations of the Tn*5* inserts were determined by restriction mapping using the same plasmids.

Analysis of the DNA sequence was done using programs from PCGene software (Intelligenetics Corporation, Mountain View, CA). Searches for sequence similarity were done using Pesearch to search the Swiss-Prot Protein Sequence Data Bank, Release 10. Analysis for open-reading frame (ORF) identification was done using the Transl and Cod-Rny programs. Membrane-associated alpha helices were found using the program RaoArgos (base line value, 1.05; peak minimal value, 1.22; and minimal helix length, 16). Inverted repeat DNA structures were found using the Hairpin program. DNA and protein sequence comparisons were made using the Nalign and Palign programs, respectively.

RESULTS AND DISCUSSION

Complementation analysis of the *dct* region. A series of Tn*5* insertions into the *R. meliloti dct* region have been used to define its location on the *exo* megaplasmid (Watson *et al.* 1988). These insertions, within cosmid pBB107 and integrated into the megaplasmid, were used to determine the number of complementation groups in the *dct* region. Pairwise testing of each cosmid Tn*5* insertion with the genomic *dct* mutants revealed three complementation groups, designated *dctA, dctB*, and *dctD* (Table 2, Fig. 1). The *dctA* gene is within a region previously found to be essential for symbiotic nitrogen fixation. The *dctB* and *dctD*

Table 2. Complementation analysis of the *dct* region[a]

	Recipient									
Plasmid	R646 (51)	R647 (35)	4F6	R639 (33)	R645 (15)	R648 (20)	R649 (39)	R650 (22)	R651 (52)	JJ1c10
pBB122, pBB107	+	+	+	+	+	+	+	+	+	+
pBB107::										
Tn*5-51*	−	−	−	+	+	+	+	+	+	+
Tn*5-35*	−	−	−	+	+	+	+	+	+	+
Tn*5-33*	+	+	+	−	−	−	+	+	+	+
Tn*5-15*	+	+	+	−	−	−	+	+	+	+
Tn*5-20*	+	+	+	−	−	−	+	+	+	+
Tn*5-39*	+	+	+	+	+	+	−	−	−	+
Tn*5-22*	+	+	+	+	+	+	−	−	−	+
Tn*5-52*	+	+	+	+	+	+	−	−	−	+
pPN108	+	+	+	+	+	+	+	+	+	+

[a] Each plasmid was transferred into the *Rhizobium meliloti* recipient strains by conjugation and tested for the ability to grow on minimal media plates containing succinate, fumarate, malate, or aspartate as the sole carbon source. The numbers in brackets below some of the strain numbers indicate the Tn*5* insertion used for their derivation. + indicates growth for all four compounds; − indicates no growth.

genes are adjacent within a region found to be nonessential for nitrogen fixation in the nodule, although necessary for dicarboxylate transport in the free-living cells (Table 3).

To establish the relationship of the genes described here to other genes involved in dicarboxylate utilization, *R. meliloti* mutants unable to grow on dicarboxylates were compared by phenotypic and complementation analysis. A collection of symbiotically defective mutants derived by Tn*5* mutagenesis (Selvaraj *et al.* 1987) was screened for growth on succinate, fumarate, malate, and aspartate. Mutants were identified that were cleanly blocked in the ability to utilize these substrates, but grew well on mannitol or glucose. These were divided into two classes by selecting cosmids, pBB107 and pBB123, that were able to restore the ability to utilize these compounds, then using these to group the mutants by complementation analysis. One class, represented initially by *dct* mutant 4F6 and cosmid pBB107, has been described previously (Watson *et al.* 1988) and is analyzed in detail here. The second class, represented by mutant 4R21 and cosmid pBB123, was identified as

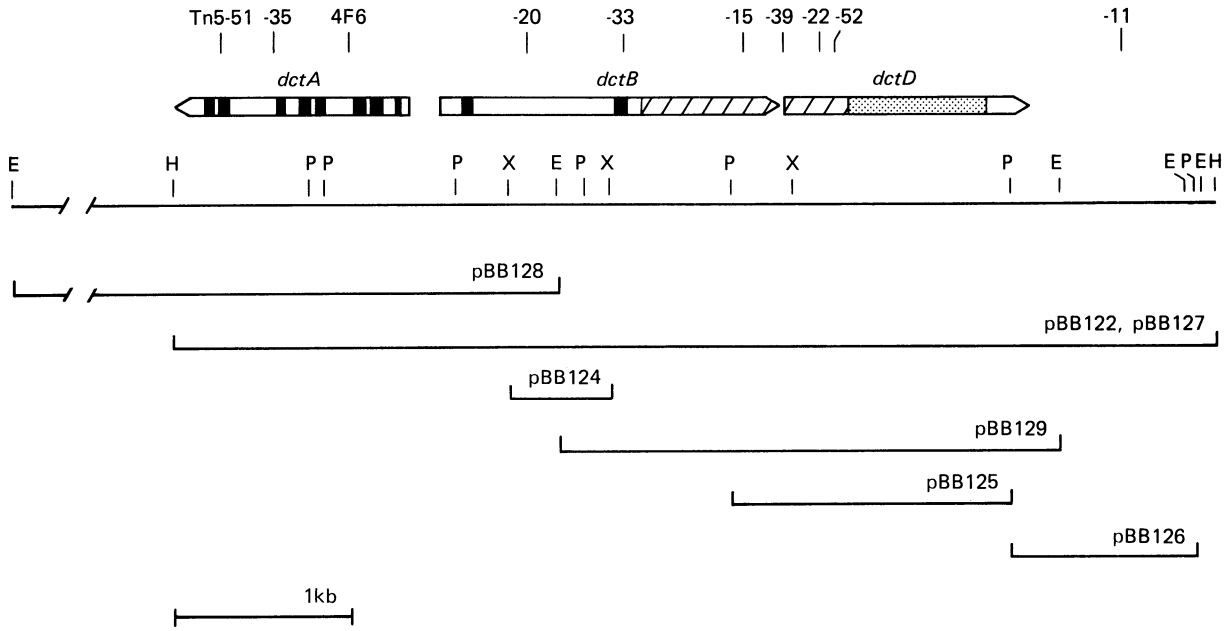

Fig. 1. Organization of the *dct* region in *Rhizobium meliloti*. The locations and orientations of the *dctA*, *dctB*, and *dctD* genes are shown relative to the restriction map of the region. Transmembrane segments in the proteins encoded by the *dctA* and *dctB* genes are shown as darkened areas. The crosshatched region in *dctB* encodes the C-terminal domain that interacts with the N-terminal domain of the *dctD*-encoded protein. In *dctD*, the crosshatched, stippled, and open portions represent segments encoding the N-terminal, central, and C-terminal (DNA-binding) domains of the *dctD*-encoded protein, respectively. The sites of Tn*5* insertions in the original *dctA* mutant, 4F6, and those derived subsequently by localized mutagenesis are shown above the genes. Cloned segments used for DNA sequencing are shown below. Restriction enzymes: E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; and X, *Xho*I.

Table 3. *Rhizobium meliloti* genes involved in transport of C_4-dicarboxylates[a]

Strain	Genotype or characteristic	Cosmid	Phenotypes				
			Growth				
			Suc	Asp(C)	Asp(N)	Nitrate	Fix
4F6	*dctA*::Tn*5-f6*	pBB122	−	−	+	+	−
R646	*dctA*::Tn*5-51*	pBB122	−	−	+	+	−
R647	*dctA*::Tn*5-35*	pBB122	−	−	+	+	−
R639	*dctB*::Tn*5-33*	pBB122	−	−	+	+	+/−
R645	*dctB*::Tn*5-15*	pBB122	−	−	+	+	+/−
R648	*dctB*::Tn*5-20*	pBB122	−	−	+	+	+/−
R649	*dctD*::Tn*5-39*	pBB122	−	−	+	+	+/−
R650	*dctD*::Tn*5-22*	pBB122	−	−	+	+	+/−
R651	*dctD*::Tn*5-52*	pBB122	−	−	+	+	+/−
4R21	*ntrA-r21*	pBB123	−	−	+	−	−
JJ1c10	Wild type		+	+	+	+	+

[a] The altered gene in each of the mutant strains is shown, together with the cosmid that carries the wild-type region. Growth was assayed on M9 media with 20 mM succinate (Suc) or aspartate (Asp) as the carbon source. Tests of the ability to use aspartate or nitrate as the nitrogen source were done on M9 media without NH_4Cl and with 20 mM glucose as the carbon source, using 20 mM aspartate or 0.5 or 5 mM potassium nitrate. Growth is indicated by +, no growth by −. Nitrogen fixation (Fix) was assayed by acetylene reduction. Wild-type levels (95 nmol per hour per plant) are indicated by +, partial nitrogen fixation ability (>50% of wild-type levels) by +/−, and no nitrogen fixation (<5% of wild-type levels) by −.

mutated in the *ntrA* gene by testing for nitrate assimilation and by comparison of the restriction map of pBB123 with that of the *ntrA* region described by Ronson *et al.* (1987b). The *ntrA* region in pBB123 was found to be chromosomal. Except for nitrate utilization, the *ntrA* mutant, 4R21, showed the same growth and nitrogen fixation phenotypes as the *dctA* mutant, 4F6, including the ability to use aspartate as a nitrogen source, but not as a carbon source (Table 3).

Sequence of the *dct* region. The 5.9-kb *Hin*dIII fragment was subcloned and its nucleotide sequence determined. Analysis of the sequence showed that the complementation groups corresponding to *dctA, dctB*, and *dctD* could be unambiguously assigned to three ORFs (Fig. 1), although there are several possible candidates for the initiation sites within each. The *dctA* gene is positioned such that its termination codon is part of one of the terminal *Hin*dIII sites (Fig. 2). The *dctB* gene is separated from *dctA* by about 230 nucleotides and is orientated such that the two genes diverge. The *dctD* gene is three nucleotides distal to *dctB* in the same orientation (Fig. 3). An inverted repeat, which may be a transcription terminator, is located about 40 nucleotides downstream of *dctD*. The remaining 979 nucleotides of the 5.9-kb *Hin*dIII fragment (sequence not shown) contain the 3′ terminus of an ORF extending into the fragment through the terminal *Hin*dIII site and ending near the *Eco*RI site distal to *dctD*. A Tn*5* insertion, Tn*5-11*, within this ORF has been described previously (Watson *et al.* 1988). It had no effect on dicarboxylate transport or nitrogen fixation.

The organization of the *R. meliloti dct* genes is the same as that reported for the *dct* genes of *R. l.* bv. *viceae* (Ronson *et al.* 1987a), and comparison of the sequences of DNA segments containing the *dctB* and *dctD* genes from the two species showed them to be 69% homologous. These similarities are consistent with the observed phenotypes of *dct* mutants in the two species. In particular, *dctA* mutants are Fix^- since this gene encodes the structural membrane protein responsible for dicarboxylate transport in both free-living cells and in bacteroids. Mutants in the *dctB* and *dctD* genes are Fix^+ or $Fix^{+/-}$. These loci are regulatory in free-living cells, but apparently are not required for activation of *dctA* in bacteroids (Table 3).

The *R. meliloti dctA* gene is preceded by the sequence CTGGCACG-N_4-TTGCT, which belongs to the consensus sequence characteristic of promoters activated by NtrA (Fig. 2). This agrees with the finding of Ronson *et al.* (1987b) that transcription of the *R. l.* bv. *viceae dctA* gene is dependent on NtrA, and this is consistent with the observation that *R. meliloti ntrA* mutants are defective in dicarboxylate transport and phenotypically similar to *dctA* mutants (Table 2).

Although candidates for the promoters of the *dctB* and *dctD* genes are present in the sequence, their identification will require techniques such as determination of the transcription start points. However, it is notable that the *dctD* gene must have its own promoter if Tn*5* insertions into *dctB* are polar, since complementation was observed between *dctB* and *dctD* mutants. If so, the close proximity of the *dctB* and *dctD* genes requires that the *dctD* promoter be located within the 3′ end of the *dctB* gene.

Ronson *et al.* (1987a) suggested that production of the *dctD*-encoded protein (DctD) by *dctB* mutants would result in a Fix^- phenotype if inactivated DctD inhibited *dctA* expression in the nodule. This would account for their observation that *dctB* mutations considered to be nonpolar, namely an *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine-derived *dctB* mutation and one Tn*5*-induced *dctB* mutation, resulted in a Fix^- phenotype, while other Tn*5*-induced *dctB*

```
TTCCGCGGCAAACTGGCACGCATGTTGCTGACCAGCTCCACAAGGCAGCCACTGCTGTCG   60
        NtrA-activated promoter

ATCTTCGGAAAGCGGCCCGGGAGGCCCGGCATGTTGCCGGACTGGGCCTGCCACGTGGAG  120
                  (rbs)     (M)                          rbs

GATATCATGATCATCGAACATTCCGCGGAGGTCCGCGGCAAGACACCCCTTTACCGCCAT  180
      M  I  I  E  H  S  A  E  V  R  G  K  T  P  L  Y  R  H      18

CTCTATGTTCAGGTGCTGGCGGCGATCGCCGCGGGCATCCTGCTCGGGCATTTCTATCCT  240
 L  Y  V  Q  V  L  A  A  I  A  A  G  I  L  L  G  H  F  Y  P     38

GATATCGGCACGGAGCTCAAACCGCTCGGCGACGCCTTCATCAGGCTCGTGAAGATGATC  300
 D  I  G  T  E  L  K  P  L  G  D  A  F  I  R  L  V  K  M  I     58

ATCGCGCCGGTGATCTTCCTGACGGTCGCGACCGGGATTGCCGGCATGACCGATCTCGCC  360
 I  A  P  V  I  F  L  T  V  A  T  G  I  A  G  M  T  D  L  A     78

AAGGTCGGCCGCGTCGCCGGCAAGGCGATGATCTACTTTCTCGCCTTCTCCACCCTCGCG  420
 K  V  G  R  V  A  G  K  A  M  I  Y  F  L  A  F  S  T  L  A     98

CTTGTCGTCGGCCTCGTCGTCGCAAACGTGGTGCAGCCGGGCGCGGGCATGCATATCGAT  480
 L  V  V  G  L  V  V  A  N  V  V  Q  P  G  A  G  M  H  I  D    118

CCGGCCTCGCTGGACGCCAAGGCAGTCGCGACCTATGCCGAGAAGGCGCATGAGCAGTCG  540
 P  A  S  L  D  A  K  A  V  A  T  Y  A  E  K  A  H  E  Q  S    138

ATCACCGGCTTCCTGATGAACATCATCCCGACGACGCTCGTCGGCGCCTTCGCCGAAGGC  600
 I  T  G  F  L  M  N  I  I  P  T  T  L  V  G  A  F  A  E  G    158

GACATCCTGCAGGTCCTGTTCATCTCGGTGCTCTTCGGTATCTCGCTGGCGATCGTCGGC  660
 D  I  L  Q  V  L  F  I  S  V  L  F  G  I  S  L  A  I  V  G    178

AAGAAAGCCGAGCCCGTGGTCGATTTCCTGCAGGCGCTGACGCTGCCGATCTTCCGGCTC  720
 K  K  A  E  P  V  V  D  F  L  Q  A  L  T  L  P  I  F  R  L    198

GTTGCGATCCTGATGAAGGCCGCCCCGATCGGCGCCTTCGGCGCCATGGCCTTCACCATC  780
 V  A  I  L  M  K  A  A  P  I  G  A  F  G  A  M  A  F  T  I    218

GGCAAGTACGGTATCGCCTCGATCGCCAATCTCGCCATGCTGATCGGCACCTTCTATCTG  840
 G  K  Y  G  I  A  S  I  A  N  L  A  M  L  I  G  T  F  Y  L    238
                                    <Tn5-35
ACGTCGTTTCTCTTCGTCTTCATCGTGCTCGGCGCGGTCGCACGCTATAACGGCTTCTCG  900
 T  S  F  L  F  V  F  I  V  L  G  A  V  A  R  Y  N  G  F  S    258

ATCCTCTCGCTCATCCGCTACATCAAGGAGGAGCTGCTGCTGGTGCTCGGGACGTCTTCT  960
 I  L  S  L  I  R  Y  I  K  E  E  L  L  L  V  L  G  T  S  S    278

TCGGAGGCGGCTCTTCCGGGTCTCATGAACAAGATGGAGAAGGCCGGCTGCAAGCGCTCG 1020
 S  E  A  A  L  P  G  L  M  N  K  M  E  K  A  G  C  K  R  S    298

GTCGTCGGTCTCGTCATTCCGACCGGTTATTCCTTCAACCTGGACGGCACCAATATCTAC 1080
 V  V  G  L  V  I  P  T  G  Y  S  F  N  L  D  G  T  N  I  Y    318

ATGACGCTTGCGGCCCTGTTCATCGCCCAGGCGACCGATACGCCGCTCTCATACGGCGAC 1140
 M  T  L  A  A  L  F  I  A  Q  A  T  D  T  P  L  S  Y  G  D    338
                                    Tn5-51>
CAGATCCTGCTGCTCCTCGTCGCAATGCTGAGCTCGAAGGGTGCGGCCGGCATTACCGGC 1200
 Q  I  L  L  L  L  V  A  M  L  S  S  K  G  A  A  G  I  T  G    358

GCCGGCTTCATCACGCTTGCCGCAACGCTCTCGGTCGTTCCCTCCGTGCCGGTCGCCGGC 1260
 A  G  F  I  T  L  A  A  T  L  S  V  V  P  S  V  P  V  A  G    378

ATGGCGCTGATCCTCGGCATCGACCGCTTCATGTCGGAATGCCGCGCCCTGACCAATTTC 1320
 M  A  L  I  L  G  I  D  R  F  M  S  E  C  R  A  L  T  N  F    398

GTCGGCAACGCGGTTGCGACGATCGTGGTGGCGAAGTGGGAAGGCGAGCTCGATCAGGCG 1380
 V  G  N  A  V  A  T  I  V  V  A  K  W  E  G  E  L  D  Q  A    418

CAGCTTTCCGCAGCTCTCGGCGGCGAGGCGTCCGTCGAGGCCATCCCGGCGGTCGTCCAG 1440
 Q  L  S  A  A  L  G  G  E  A  S  V  E  A  I  P  A  V  V  Q    438

CCCGCCGAATAAGCTT                                              1456
 P  A  E  -                                                    441
```

Fig. 2. Nucleotide sequence of the *dctA* gene and deduced amino acid sequence of DctA. The sequence is given in the orientation of the *dctA* gene, as shown in Figure 1, ending at the *Hin*dIII site. The sequence characteristic of NtrA-activated promoters is underlined. Possible ribosome binding sites (rbs) and their ATG initiation codons are also indicated. Sites of Tn*5* insertions are indicated by overlining the ends of the nine-nucleotide segments duplicated during transposition. The orientations of the Tn*5* inserts are indicated by < or > to show the direction of the *nptII* gene. The deduced amino acid sequence of DctA is shown below the nucleotide sequence and numbered on the right. The underlined protein segments are those found to contain membrane-spanning helices by the method of Rao and Argos (1986).

mutations with Tn*5* in the opposite orientation produced a Fix$^+$ phenotype. Thus, the Fix$^-$ phenotype could be due to *dctB* mutations of a type that permitted continued transcription of *dctD* and production of the inactivated DctD protein.

To examine this question more closely, the insertion sites and orientations of the Tn*5* inserts in the *R. meliloti dct* region were determined (Figs. 2 and 3). Three of the Tn*5* insertion sites were determined by sequencing from both ends with primers complementary to *R. meliloti* DNA and converging on the transposon. It was confirmed that the transposon generated a 9 base pair repeat at each target site, as is characteristic. The other Tn*5* insertion sites were determined using a primer complementary to the Tn*5* internal sequence. Their orientations were determined by mappings using *Sma*I and *Bam*HI. Six of the eight Tn*5* insertions, including the three insertions in the *dct*B gene, were found to be orientated such that the *nptII* gene would be transcribed in the opposite direction to the gene in which it was inserted (Figs. 2 and 3). This orientation corresponds

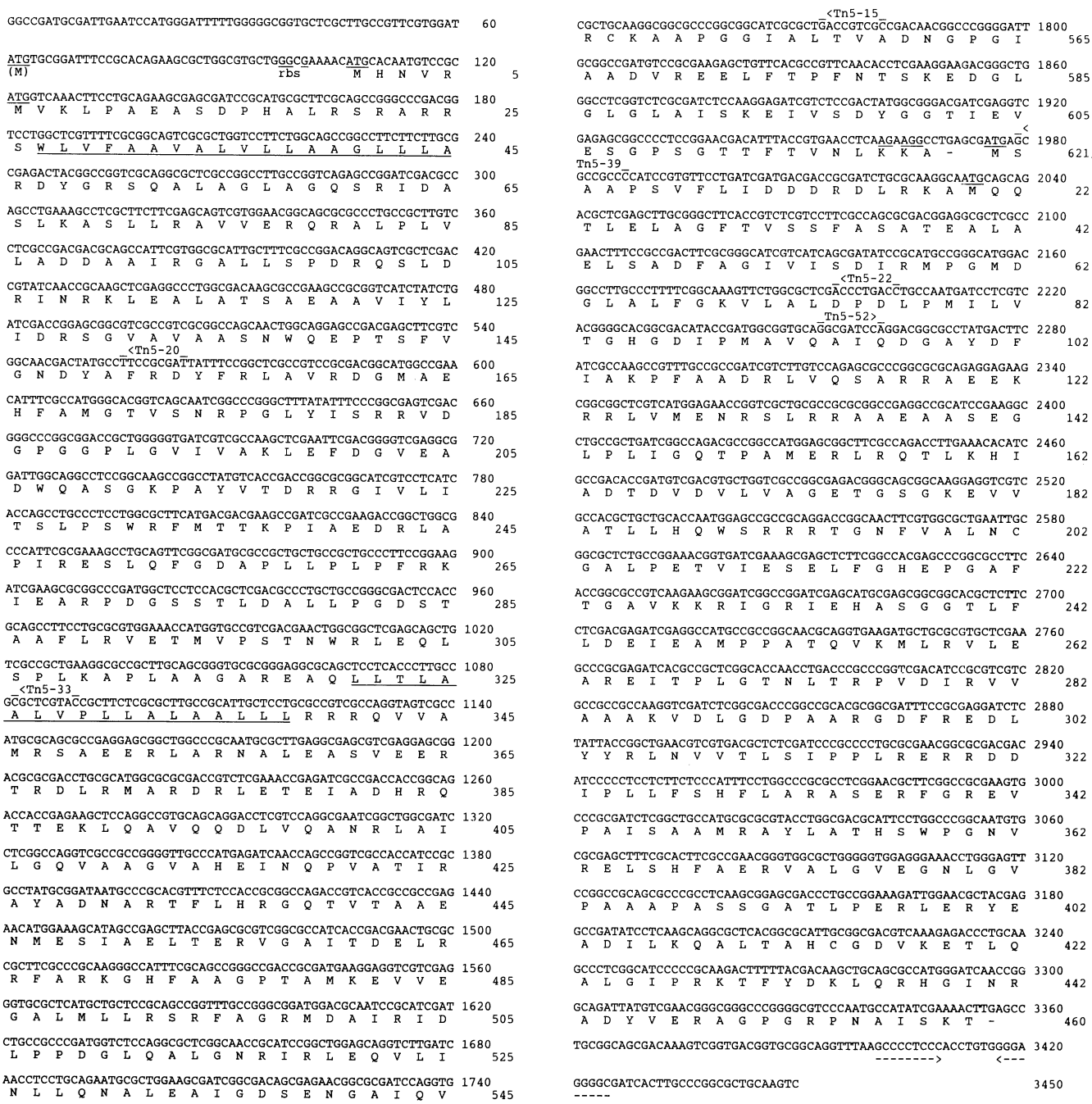

Fig. 3. Nucleotide sequence of the *dctB* and *dctD* genes and deduced amino acid sequences of DctB and DctD. The sequence shown is continuous with that in Figure 2, but in the opposite direction. Possible initiation codons, ribosome binding sites, Tn*5* insertions, and membrane-spanning protein segments are indicated as in Figure 2. An inverted repeat in the nucleotide sequence following *dctD* is indicated beneath the sequence by dotted lines with arrows at the ends.

to that of the Fix⁻ *R. l.* bv. *viceae* mutant with Tn*5* in *dctB* as described above. Since these *R. meliloti dctB* mutations do not result in a Fix⁻ phenotype, the *R. l.* bv. *viceae* model does not appear to apply to the *dct* genes in *R. meliloti.*

The DctA protein. The *dctA* gene shown in Figure 2 is the 3′ end of a large ORF, 4.5 kb, which starts within the *dctD* region of the sequence. There are two possible start sites for the coding portion of the gene, both of which are preceded by good candidates for a ribosome binding site (underlined in Fig. 2). The second of these, used here, corresponds to the start site predicted in *R. l.* bv. *viceae.* DctA is a 441 amino acid protein that is highly hydrophobic. Analysis of the amino acid sequence to predict transmembrane helices by the method of Rao and Argos (1986) showed eight segments that may span the cytoplasmic membrane (underlined in Fig. 2). These characteristics support predictions that DctA is an integral membrane protein responsible for the transport of dicarboxylates from the periplasm to the cytoplasm. Direct evidence that DctA is membrane-associated has been obtained by Yarosh *et al.* (1989) using Tn*phoA* fusions to this protein.

The DctB and DctD proteins. The *dctB* and *dctD* genes have several possible translation start sites, as shown in Figure 3; those chosen here have the best candidate ribosomal binding sites and correspond to the start sites chosen by Ronson *et al.* (1987a) for these genes in *R. l.* bv. *viceae.* In *dctD*, the choice of the earlier initiation codon is favored by the observation that it immediately precedes a Tn*5* insertion found to be within the *dctD* complementation group. DctB and DctD are 621 and 460 amino acids long, respectively. The *R. meliloti* DctB protein is 60% homologous to that of *R. l.* bv. *viceae*; the DctD protein is 77% homologous. *R. meliloti* DctD has 12 additional amino acids near the C-terminus.

The DctB protein was analyzed to search for transmembrane helices using the same program parameters as described above for DctA. Two membrane-spanning segments were predicted, one near the N-terminus and a second near the middle of the protein. These correspond to the two membrane-spanning segments known in *R. l.* bv. *viceae* DctB (Fig. 1, underlined in Fig. 3). The significance of these segments is indicated by the observation that, although they are uniquely and identically predicted on the basis of structure, they are less than 50% homologous between the two species, and in the first segment only 6 of 19 amino acids match. The domain of DctB between these transmembrane segments is expected to be periplasmic, where it has been proposed that it functions in the detection of dicarboxylates, prior to induction of the transport system.

Homology searches through the Swiss-Prot Protein Sequence Data Bank were made using amino acid sequences from *R. meliloti* DctB and DctD comprising domains corresponding to those in proteins belonging to other two-component regulatory systems. A search with the C-terminal portion of DctB, amino acids 372–621, the putative sensor component domain, showed homology to the *ntrB*-encoded protein from *Klebsiella pneumoniae* Schroeter (Trevisan) and *E. coli* and to the *phoR*-encoded protein from *E. coli.* The N-terminal portion of DctD, amino acids 1–117, the putative regulator component domain that receives the signal from the DctB sensor, was found to have homology to proteins encoded by *ntrC, phoB*, and *ompR* from *E. coli*; *ntrC* from *K. pneumoniae*; and *virG* from *A. tumefaciens.* The central portion of DctD, amino acids 118–388, the domain thought to change conformation between active and inactive forms in conjunction with the C-terminus, showed homology to the *nifA*-encoded protein from *R. meliloti, R. l.* bv. *viceae, Bradyrhizobium japonicum* (Buchanan) Jordan, *Azorhizobium caulinodans* Dreyfus *et al.*, *K. pneumoniae*, and *Azotobacter chroococcum* Beijerinck, and to the *ntrC*-encoded protein from *K. pneumoniae* and *E. coli.* These homologies are to proteins

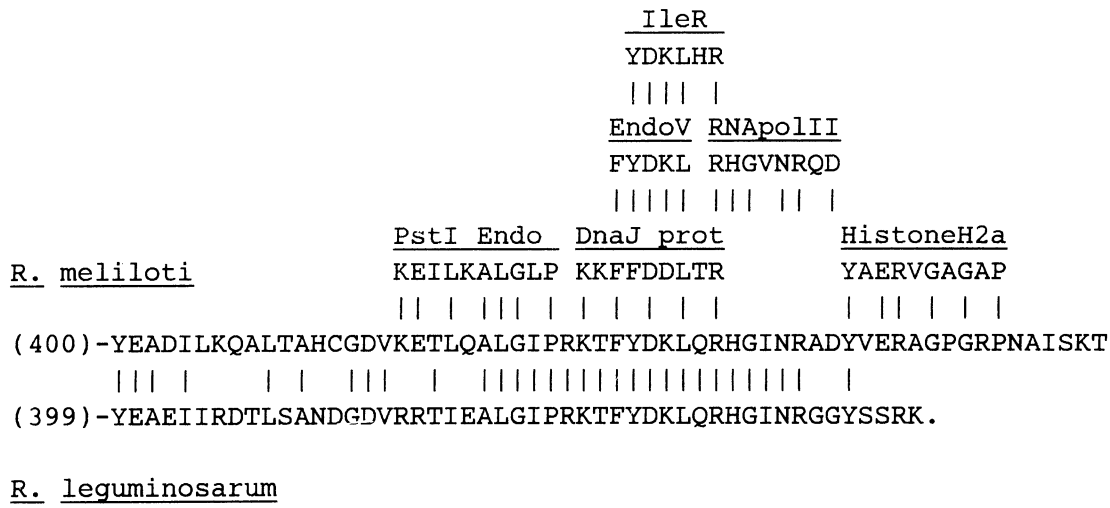

Fig. 4. Homologies with the C-terminal amino acid sequence of the *Rhizobium meliloti* DctD protein. The 60 amino acids of the C-terminal of *R. meliloti* DctD are shown aligned with the C-terminus of *R. leguminosarum* bv. *viceae* DctD. Homologous segments of other proteins thought to bind DNA are indicated above. Abbreviations shown for the protein segments refer to the following protein segments retrieved from the Swiss-Prot Protein Sequence Data Bank: IleR, amino acids 61–67 of the isoleucine repressor of *Escherichia coli* (ILER$ECOLI); EndoV, amino acids 60–65 of endonuclease V of bacteriophage T4 (END5$BPT4); RNApolII, amino acids 1385–1393 of RNA polymerase II, 215,000-dalton polypeptide from mouse (RPO2$MOUSE); PstI Endo, amino acids 10–20 of restriction endonuclease *Pst*I from *Providencia stuartii* (NUPS$PROST); DnaJ prot, amino acids 368–377 of the *dnaJ*-encoded protein of *E. coli* (DNAJ$ECOLI); and HistoneH2a, amino acids 39–49 of histone H2A (H2A$BOVIN).

known to belong to two-component regulatory systems as described by Ronson *et al.* (1987c). The similarity of the *R. meliloti* DctB and DctD protein sequences to others proposed to belong to two-component systems lends further support to this model.

The C-terminal portion of DctD, amino acids 389–460, the proposed DNA-binding domain, showed homology to proteins that interact with DNA, including the *Pst*I restriction endonuclease, the *dnaJ*-encoded protein of *E. coli*, and histone H2A from numerous eucaryotic species (Fig. 4). This DctD domain differs between *R. meliloti* and *R. l.* bv. *viceae* in that *R. meliloti* DctD contains a unique 18 amino acid sequence at its C-terminus. This extension may be an integral part of the *R. meliloti* DctD DNA-binding site since it contains the region homologous to histone H2A.

Complementation of *R. meliloti dctA* mutants by *R. l.* bv. *viceae dct* genes was tested to determine whether the *R. l.* bv. *viceae* transport system could functionally replace that of *R. meliloti* (Table 3). A plasmid containing all three *R. l.* bv. *viceae dct* genes, pPN108, was found to restore the ability of all the *R. meliloti dct* mutants to grow on C_4-dicarboxylates.

The *R. meliloti* dicarboxylate transport system has been found to be the product of three *dct* genes. Two of the *dct*-encoded proteins sense the presence of dicarboxylates outside the cell and activate the production of the third component, an integral membrane protein that affects the movement of dicarboxylates across the inner membrane. This system is broadly related to a series of others comprising two-component regulatory systems in bacteria and is specifically related to the dicarboxylate transport system in *R. l.* bv. *viceae*. The sequences and organization of the *dct* genes and their protein products are well-conserved between *R. meliloti* and *R. l.* bv. *viceae*, although differences are apparent. The DctB proteins of the two species vary in sequence to a greater extent than do the DctD proteins, though both proteins have retained their essential similarity to those of other two-component systems. Since neither of these regulatory proteins is essential for symbiotic nitrogen fixation, it is of interest to identify an alternate regulatory system that is able to activate the *dctA* gene in nodules.

ACKNOWLEDGMENTS

The dedicated technical assistance of Roselyn Heys is greatly appreciated.

This work was supported in part by Strategic Grant 40182 from the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED

Bergersen, F. J., and Turner, G. L. 1967. Nitrogen fixation by the bacteroid fraction of breis of soybean root nodules. Biochim. Biophys. Acta 141:507-515.

Beringer, J. E. 1974. R factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. 84:188-198.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95-113.

Bolton, E., Higgisson, B., Harrington, A., and O'Gara, F. 1986. Dicarboxylic acid transport in *Rhizobium meliloti*: Isolation of mutants and cloning of dicarboxylic acid transport genes. Arch. Microbiol. 144:142-146.

Boyer, H. W., and Roulland-Dussoix, D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. 41:459-472.

Chen, E. Y., and Seeburg, P. H. 1985. Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4:165-170.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 1980. Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA 77:7347-7351.

Ditta, G., Schmidhauser, T., Yakobson, E., Lu, P., Liang, X.-W., Finlay, D. R., Guiney, D., and Helinski, D. R. 1985. Plasmids related to the broad host range vector pRK290 useful for gene cloning and for monitoring gene expression. Plasmid 13:149-153.

Engelke, T., Jagadish, M. N., and Pühler, A. 1987. Biochemical and genetical analysis of *Rhizobium meliloti* mutants defective in C_4-dicarboxylate transport. J. Gen. Microbiol. 133:3019-3029.

Figurski, D. H., and Helinski, D. R. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided *in trans*. Proc. Natl. Acad. Sci. USA 76:1648-1652.

Friedman, A. M., Long, S. R., Brown, S. E., Buikema, W. J., and Ausubel, F. M. 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene 18:289-296.

Gussin, G. N., Ronson, C. W., and Ausubel, F. M. 1986. Regulation of nitrogen fixation genes. Annu. Rev. Genet. 20:567-591.

Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Miller, R. W., McRae, D. G., Al-Jobore, A., and Berndt, W. B. 1988. Respiration supported nitrogenase activity of isolated *Rhizobium meliloti* bacteroids. J. Cell. Biochem. 38:35-49.

Rao, J. K. M., and Argos, P. 1986. A conformational preference parameter to predict helices in integral membrane proteins. Biochim. Biophys. Acta 869:197-214.

Ronson, C. W., and Astwood, P. M. 1985. Genes involved in the carbon metabolism of bacteroids. Pages 201-207 in: Nitrogen Fixation Research Progress. H. J. Evans, P. J. Bottomley, and W. E. Newton, eds. Martinus Nijhoff, Dordrecht, The Netherlands.

Ronson, C. W., Astwood, P. M., and Downie, J. A. 1984. Molecular cloning and genetic organization of C_4-dicarboxylate transport genes from *Rhizobium leguminosarum*. J. Bacteriol. 160:903-909.

Ronson, C. W., Astwood, P. M., Nixon, B. T., and Ausubel, F. M. 1987a. Deduced products of C_4-dicarboxylate transport regulatory genes of *Rhizobium leguminosarum* are homologous to nitrogen regulatory gene products. Nucleic Acids Res. 15:7921-7934.

Ronson, C. W., Lyttleton, P., and Robertson, J. G. 1981. C_4-Dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. Proc. Natl. Acad. Sci. USA 78:4284-4288.

Ronson, C. W., Nixon, B. T., Albright, L. M., and Ausubel, F. M. 1987b. *Rhizobium meliloti ntrA* (*rpoN*) gene is required for diverse metabolic functions. J. Bacteriol. 169:2424-2431.

Ronson, C. W., Nixon, B. T., and Ausubel, F. M. 1987c. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell 49:579-581.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Selvaraj, G., Hooper, I., Shantharam, S., Iyer, V. N., Barran, L., Wheatcroft, R., and Watson, R. J. 1987. Derivation and molecular characterization of symbiotically deficient mutants of *Rhizobium meliloti*. Can. J. Microbiol. 33:739-747.

Watson, R. J., Chan, Y.-K., Wheatcroft, R., Yang, A.-F., and Han, S. 1988. *Rhizobium meliloti* genes required for C_4-dicarboxylate transport and symbiotic nitrogen fixation are located on a megaplasmid. J. Bacteriol. 170:927-934.

Yanisch-Perron, C., Vieira, J., and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

Yarosh, O. K., Charles, T. C., and Finan, T. M. 1989. Analysis of C_4-dicarboxylate transport genes in *Rhizobium meliloti*. Mol. Microbiol. 3:813-823.