

---

**Deduced products of C<sub>4</sub>-dicarboxylate transport regulatory genes of *Rhizobium leguminosarum* are homologous to nitrogen regulatory gene products**

---

Clive W. Ronson<sup>1,2,3</sup>, Patricia M. Astwood<sup>1</sup>, B. Tracy Nixon<sup>2,4</sup> and Frederick M. Ausubel<sup>2</sup>

---

<sup>1</sup>Grasslands Division, Department of Scientific and Industrial Research, Private Bag, Palmerston North, New Zealand and <sup>2</sup>Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA

---

Received August 10, 1987; Accepted September 9, 1987

---

**ABSTRACT**

We have sequenced two genes *dctB* and *dctD* required for the activation of the C<sub>4</sub>-dicarboxylate transport structural gene *dctA* in free-living *Rhizobium leguminosarum*. The hydrophobic profile of the *dctB* gene product (DctB) suggested that its N-terminal region may be located in the periplasm and its C-terminal region in the cytoplasm. The C-terminal region of DctB was strongly conserved with similar regions of the products of several regulatory genes that may act as environmental sensors, including *ntrB*, *envZ*, *virA*, *phoR*, *cpxA*, and *phoM*. The N-terminal domain of the *dctD* gene product (DctD) was strongly conserved with N-terminal domains of the products of several regulatory genes thought to be transcriptional activators, including *ntrC*, *ompR*, *virG*, *phoB* and *sfra*. In addition, the central and C-terminal regions of DctD were strongly conserved with the products of *ntrC* and *nifA*, transcriptional activators that require the alternate sigma factor *rpoN* (*ntrA*) as co-activator. The central region of DctD also contained a potential ATP-binding domain. These results are consistent with recent results that show that *rpoN* product is required for *dctA* activation, and suggest that DctB plus DctD-mediated transcriptional activation of *dctA* may be mechanistically similar to NtrB plus NtrC-mediated activation of *glnA* in *E. coli*.

**INTRODUCTION**

The C<sub>4</sub>-dicarboxylate transport structural gene *dctA* is expressed in both free-living and symbiotic forms of *Rhizobium leguminosarum*. In free-living cells, expression of *dctA* is induced by succinate, fumarate or malate, and *dctA* product (DctA) is required for growth on these C<sub>4</sub>-dicarboxylates as carbon sources. DctA is also essential for symbiotic nitrogen fixation, as *dctA* mutants form nodules that contain bacteroids but do not fix nitrogen (1-4). The symbiotic role of DctA has not been defined, but it is likely that DctA is required for transport of C<sub>4</sub>-dicarboxylates into bacteroids for use as energy substrates (reviewed in 5,6).

The *dctA* gene encodes a hydrophobic polypeptide of 444 amino acids and is the only structural gene required for C<sub>4</sub>-dicarboxylate transport in both symbiotic and free-living rhizobia (2,5). The *dctA* promoter (5) is strongly homologous to the consensus promoter -26 CTGGYAYR-N4-TTCA -10 recognized by the *rpoN* (*ntrA*) sigma factor (7), and recent results have shown that *rpoN*

product (RpoN) is required for dctA expression (8). Other promoters known to be recognised by RpoN such as glnAp2 and nif promoters also require the product of another regulatory gene, either the nitrogen regulator ntrC or the nif-specific regulator nifA, for activation (reviewed in 7). The ntrC and nifA products (NtrC and NifA respectively) are structurally similar, sharing extensive regions of homology outside of their N-terminal domains that may include a domain that interacts with RpoN (9,10). These observations suggest that the dctA promoter may require an activator analogous to NtrC or NifA.

The regulation of dctA is complex. The dctA gene may be under negative autogenous control, since a dctA-lacZ fusion is expressed constitutively in dctA mutants (5,11). In addition, two regulatory loci dctB and dctD that are closely linked to dctA but probably transcribed divergently to dctA (1,2,5, 11), are required in addition to inducer and RpoN for dctA activation in free-living rhizobia. The regulatory components of the dct regulon may also be required for symbiotic nitrogen fixation, since nitrosoguanidine-induced dctB mutants form nodules that fail to fix nitrogen (1,2). However only one out of 12 Tn5-induced dctB mutants formed nodules that fixed nitrogen at significantly less than wild-type rates; the other Tn5-induced dctB mutants and three Tn5-induced dctD mutants formed nodules that fixed nitrogen as well as nodules formed by wild-type rhizobia (11). The reason for the disparity in symbiotic phenotypes between the chemically-induced and transposon-induced dctB mutants is not clear.

In this paper, we report the DNA sequence of dctB and dctD, and the precise location and orientation of the Tn5 insertions. We report that dctB product (DctB) is homologous over its C-terminal region to a variety of proteins including ntrB product (NtrB) that may act as environmental sensors, and that dctD product (DctD) is homologous over its N-terminal region with a variety of proteins including NtrC that may act in conjunction with a sensor component to activate transcription in response to environmental stimuli. Furthermore DctD is strongly homologous in other regions to NtrC and NifA. Based on these results, we propose that DctB is a receptor for C4-dicarboxylates and transduces a signal across the membrane to DctD which then acts in conjunction with RpoN to activate transcription of dctA. We also propose a model to account for the symbiotic phenotypes of the regulatory mutants.

### MATERIALS AND METHODS

The region of DNA containing dctB and dctD was sequenced using specific restriction fragments and nested sets of deletions generated using Bal 31

nuclease (Fig. 1). Appropriate restriction fragments were subcloned from the *dct* plasmids pPN103 and pPN104 (2) into M13 mp8 and mp9 (12). The 1.2 kb BamHI/EcoRI fragment and the 2.1 SalI fragment (Fig. 1) were subcloned into pUC8 (13) and the resultant plasmids used as the sources of DNA for the construction of Bal 31-generated deletions as described (8). Sequencing was done by the dideoxy method using a 17-base universal primer and [<sup>35</sup>S]dATP as label (14).

Sequence was also determined from the ends of several Tn5 insertions in *dctB* and *dctD*. Recombinant plasmids that contained Tn5 insertions previously used to generate *dctB* and *dctD* mutants by homogenotization (11), were digested with EcoRI plus BamHI or BamHI alone, and the fragments cloned into pUC8. Appropriate clones were screened for kanamycin resistance to determine the orientation of the Tn5 insertion. The fragments were then cloned into M13 mp8 for sequence determination using a Tn5-specific primer as described (8). The sequence generated also served to sequence across the EcoRI site in *dctB* (Fig. 1).

DNA sequences were compiled using the DBAUTO and DBUTIL programs (15), and were analysed using programs from the University of Wisconsin (16). The program FASTP (17) was used to screen the NBRF Protein Identification Resource data base for sequences homologous to DctB or DctD. Pair-wise comparisons between proteins were made with the ALIGN program, using the mutation data matrix and a gap penalty of 16 (18). The alignment scores given represent the number of standard deviations separating the maximum score of the real sequences from the mean of the maximum scores of 100 random permutations of the two sequences (18).

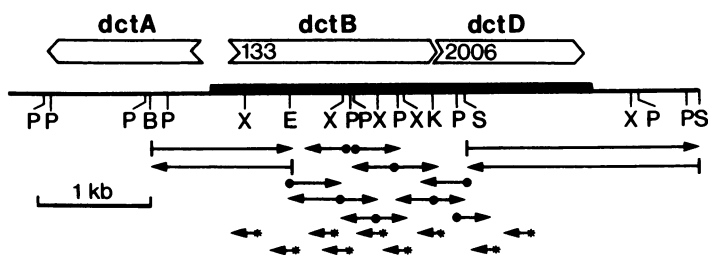


Figure 1. Restriction map of the *dct* regulon from *R. leguminosarum* (2,11), and strategy used to sequence the *dctB* and *dctD* genes. The heavy line indicates the region, the sequence of which is presented in Fig. 2. The sequence was determined from restriction fragments (●→), deletions generated using Bal 31 nuclease (▲→), and from the ends of Tn5 insertions (\*→). Restriction sites: B, BamHI; E, EcoRI; K, KpnI; P, PstI; S, SalI; X, XhoI.

**RESULTS AND DISCUSSION**

**Nucleotide sequences of *dctB* and *dctD***

In previous work, we defined the region of DNA containing *dctB* and *dctD* by genetic complementation studies and transposon mutagenesis (2,5,11). We have now sequenced this region (Fig. 1), and also determined the orientation and precise sites of several Tn5 insertions. The sequence data (Fig. 2) revealed two open-reading frames (ORFs). The first ORF of 1866 nucleotides was designated *dctB* since it corresponded precisely in location and orientation to the region previously genetically defined to be *dctB* (2). Furthermore sequence analysis showed that all Tn5 insertions that gave a *DctB*<sup>-</sup> phenotype (leaky growth on succinate) were located in this ORF. The translation initiation site of *dctB* was arbitrarily assigned at the first of two possible ATG codons which were separated by 12 bp. Neither ATG codon was preceded by a good match to the consensus ribosome binding site defined for *E. coli* genes (19). The first ATG codon of *dctB* was separated from that of *dctA* by 223 bp, and the genes were transcribed divergently (data not shown). The second ORF was designated *dctD* since the three Tn5 insertions that gave a *DctD*<sup>-</sup> phenotype (no growth on succinate) mapped within the ORF. The translation initiation site of *dctD* was also assigned at the first of two potential ATG codons which were separated by 3 codons, giving an ORF of 1344 nucleotides. The *dctD* sequence was preceded by two potential ribosome binding sites, the first of which (AAGG) overlapped *dctB*, and *dctB* and *dctD* were separated by 4 nucleotides (Fig. 2).

Eight out of the nine Tn5 insertions in *dctB* and *dctD* that were analyzed were in the same orientation, with transcription of the *nptII* gene from left to right with reference to Fig. 1. The insertion in strain 659, which is the only Tn5-induced *dctB* or *dctD* mutant strain that formed *Fix*<sup>-</sup> nodules (11), was the only insertion in the opposite orientation.

**Structure of the *dctB* product**

A search of the Protein Identification Resource protein data base using the FASTP program (17) revealed that *DctB* shared homology over its C-terminal region to the product of *envZ*, an *E. coli* gene involved in osmoregulation (20, 21). It has recently been reported that *envZ* product shares extensive homology over its C-terminal 250 amino acids to analogous regions of the products of the regulatory genes *ntrB* (22), *cpxA* (23), *phoR* (24), and *virA* (25). Comparison of the C-terminal region of *DctB* with the other gene products using the ALIGN program showed that several residues grouped in 5 regions were either invariant or substituted by a conservative change in all six proteins. The product of a further gene *phoM* (26) was also found to contain these homologous regions (Fig. 3).

1  
 CCACATCCCGTCCAGTTTAAAGTCCCAACCAACCGATTGAATTTTATTAATGAAAATTTACGAGACTGCTAACACACTAACTAAGCAATGCGAAATCCGCACAAATCCATTTCGGCTT  
 121  
 TCGTGCAGAAAATGACAAAATCCCACTGTCCGTTGTGCGAAAATTTGGCTTCCCTGCCCTGCAACACCCOCATGCCCGGATGTGGTGCACATGCCCGCTCCCTCTCCGCG  
 M H K S A M S V S R K L W P S L P L Q H R I R R M W W T Y A A L A F L A  
 241  
 DctB  
 GTCGTCCCAAGCCTTGGACACCGCCGAGTGCGCCAGCACCGGCGGAGCTGCCCTCGAGGACGAGCCGCCATGDTGTGACCTGATCGGGCTTCTGCCACGCTCTTGA  
 V V A S L W T S G E I G Q H R A E A L E Q A R M D V A L L E A R L T V L E  
 361  
 AAATACCGGCGCTCCCTCTGCTGTGCGAGACACCCGCTGCCCGGCGCTGTGGGAAACGATGCCCGACGTTGACAGCGCTCAGTCAGAACTGCGAACTGCGCGCGGCG  
 K Y R A L P F V L S Q D T A L A A A L V G N D A G T F E R L S Q K L E I L A G A  
 481  
 ACGAAGCCCGCTCATCTATGTCTGACGAAGGACGATCCCGCTTTCGCCAGCACTGCCGCGACCGACGCTTGTCCGCAATGACTACCGCTTCCCGGATATTTACAGGG  
 T K A A V I Y V I D K D G I A V S A S N W R E P T S F V G N D Y R F R E Y F R G  
 601  
 CGGTCCAGAGCGCCAGCCGACCTTCCCTCGCCAGCTTACGCAAGAACCGGCTTTATATCTCGACGCGAATTCGCGCAGCAATGCCCTGCTGGTGTGCTGTGGTCAAG  
 A V E R G Q A E H F A L G T V S K K P G L Y I S Q R I S G S N G L L G V V V V K  
 721  
 V533  
 GTCGAAATCGACACGTCGAGCGGATGGAACCGCTCCGCGACGCTCCGCTACGTCGTTGACGAGCGTGGCATGTGCTCATCACCGATCTTCATCATGCGCGTTCATGACGATCGG  
 V E F D D V E A D W N A S G T P S Y V V D E R G I V L I T S L P S W R F M T I G  
 841  
 CGGATGCCGAGATGCCCTGACAGCGATCCGCAAAAGCTCCAATTCGCGCTCGCCACTTCAACCCCTGCCCTCGACATGTCGGAACCTCGCGCAAGGCGTCAATGCTGCTCGAG  
 R I A E D R L T A I R E S L Q F G A A P L Q P L P L D M V R N L G E G L D V E  
 961  
 ATGTCATGCCCGCAGTCCCGGAAAACAGTTTCTGATGTCGCAACGCTCGGCGGACCGGATGCGATTCGACGATCGTGGCGTGGCGCGCTCCGTCGATCGGGGAT  
 I V M P G D A G K E N F A L G T V S K K P G L Y I S Q R I S G S N G L L G V V V V K  
 1081  
 V535  
 CGCGCAACCGCATCGCATCTCATCTCTGCACTGCTCGCGGACGAGCTTCTCTCGCGCGCGCGCTTACGATCGCGCTCGGATCGCGTCCGACACGACAGCAGCAGAG  
 R E A R M L L L L L L L L L L L L L L L L L L R R R R H T A I L R I S S E Q Q A R E  
 1201  
 V543  
 GAATCGAAGCGCGTGTGCGAGCGAGCTGATCTGCAAGCGCGAGACCGCTGACGCTGAAATCATCGCCCAAGAGCAGCGAAGAACTCGAGCGGTCGACGAGAT  
 E L E R R V V E R T L D L S Q A R D R L Q A E I I G H K S T E R K L Q A V Q Q D  
 1321  
 CTGGTGCAGCCATCGCTGCCATGCTGCCAGGTAGCCCGCGGCTCCCTCATGATCAACCGCGTGGCGACATCGCTGCTATCGGATATGCCCGACCTCTCCGAT  
 L V Q A N R L A I L G Q V A A G V A H E I N Q P V A T I R A Y A D M A R T F L D  
 1441  
 CGCGCCGAGCTGCGCTCCCGGAAAATCGCGGCTCGCGGACGACGATGATGATCGACCGGACGCTGCGCGCTTCCGCGCAAGCGCCGCGCGACGCGCGA  
 R G Q T A P A G E N L E S I A A L T E R I G S I T E E L K T F A R K R G S A  
 1561  
 CCACAGGATGAAGAGCTCATCGAGGCGCGGATGCTGTCGCGAGCGGTTTCCCGCGCATGATACGCTCGACATCGACCTCCCGCGCGAGACTCGAGTGTGGAAAC  
 P T Q L K D V I E G A V M L L R S R F A G R M D T L D I D L P P D E L Q V M G N  
 1681  
 V659  
 CGGATGCTGTCGAGCGCTCTCATCACTGCTCGAAGCCGCTGGAGCGGTCGCGCGAAGCGCGGAGGCTCGCGTGGATCAGAACATCAACTGACCGCGGATGTAGG  
 R I R L E L I M L I M L E A V A P K A G E R V E I R A T S T D A G W T  
 1801  
 GTGACGTCGCGCACAGCGGATCCACGGAATCCCGAAAGCTTGTTCAGCGGCTTCAATACCTCGAAGGAAAGCGCTGCGCTGCGCTGCGTATTCGATCGAGATATC  
 V T V A D N G P G I P T E I R K G L F T P F N T S K E S Q L G L Q L V I S K D I  
 1921  
 TCGCGACTACCGCGCGGATGAGCTTCCAGCGACCGGTTGACCGGTTTATCGTTCAGCTGAGGAAGCTTGAAGTTCATGACACACTGATCCCGCTTCCGCTGATCGAGA  
 V G D Y G G R M D V A S D S G G T R F I V Q L R K A M D T L M P V A L I D D  
 2041  
 V538  
 DctD  
 CGCAAGATCTGCGCGTCCACCGCGAGCTCGAATCCCGGATTTCCGTTTCCGCTATGAGCTCGAAGAACCGCGGACTGCGCGCGACTTTCGCGCGGCTTTCGCGCGGCT  
 D K D L R A T K E N R L P L I G Q T P V M E N L R N I L R H I A D T D V D V L V A  
 2161  
 CGTACCGATTCGCGCGTCCGCGGATCGAGCGACTGACCTTCCGACCGCTCGAGGATGATGTGACCTCGCGGCTGATGATGACCGCGCGCGGATTTCCGATCGCGCT  
 V T D I R M P E I D Q L Q L F A T L L Q M D V D L P V I L M T G H G D I P M A V  
 2281  
 TCGAGGATTCAGGAGCGGCTATGATTTCAACCAAGCCCTTCCGAGCGGATCGCTGCTCGAGCGCTCGCTCGCGCAAGCGAGAGCGCGGCTTGTTCGAGAGCCGATGCT  
 Q A I Q D G A Y D F I A K P F A A D R L V Q S V R R A S E K R R L V L E N R M L  
 2401  
 V663  
 CGCGAAGCAGCGGAGATGCGCGAGAAATTTCCGCTGATCGCCAGACCGCTGTGATGAAAACCTCAGAACATCTTCCGACATCGCGATCCGATGTGACGCTACTGCTGC  
 R K A E A D G E N R L P L I G Q T P V M E N L R N I L R H I A D T D V D V L V A  
 2521  
 CGCGAAGCAGCGGAGAGAGTGTGCGCGAGCTGCTGATGAGCGACCGCGGAAAGGCAATTCGTCGCGCTGAGAGAACCGCGGAGATTTCCGCGCTTCCGCAAGCGCT  
 G E T G S G K E V V A G I L H Q W S H R R K G N F V A L N C G A L P E T V I E S  
 2641  
 CGAAGCTTTCGCGCACAGCCGCGGCTTTCACCGCGCCGAGAGCCGCGACCGCGGATCGAACATCGAAGCGCGGACCGCTTGTCTCGACGATCGAAGCATCGCGCGC  
 E L F G H E R G A F T G A Q K R R T G R I E H A S G Q T L F L D E I E S M P A A  
 2761  
 V537  
 CAGCGAGTCAAGTCTGAGGCTGCTGAGAGTGCAGGATCACCGGCTCGACCAATGAGTGGCGCGCTCAATCTCGGCTGCTCGCGCTGCGAGATCGACCTCGCGGCGC  
 T R V K M L R V L E M R E I T P L G T N E V R P V N L R V V A A K I D L G D P  
 2881  
 CGCGGTTCCGCGGATTTCCGCGGATCTTATTCAGCGCTGATGTCGAGATCTCCATTCGCGCGCTGAGAGAACCGCGGAGATTTCCGCTGCTGTTTCCGATCTCCGCG  
 A Y R G D F R E D L Y L R L N V V T I S I P P L R E R R D I L P L F S H F A A  
 3001  
 TCGCGCGGCGGCTTCCGCGGATTCGCGCGCTTTCACCTGATGTCGCGCGGATCTCGCTCGCACATGCGCAAGCAATGTCGCGGAGCTTCGCAATATCGCGAAGCGT  
 R A A E R F R R D V P L S P D V R R H L A S H T W P G N V R E L S H Y A E R V  
 3121  
 GTCCTTTCGCGGATGAGGCGCGGAGCGGCTCCCTCTCGCGGAGCGGTCGCGGCTCCCGCAAGCGCTGAGAGCTACGAGGCGGAGATCATCGCGACCGCTGTCAGCCAA  
 V L G V E G G A A A V P P T G A T L P E R L E R I E A I R D T L S A A  
 3241  
 TGCAGCGGCTGCGCGGATTCGCGGATTCGCGGAGCGGCTTTCAGGAAAGCGTTTATGCAAACTTCAGCGCGCACGCGATCAACCGCGCGCTATAGCTCGCGCGAGTAACTCGGTCG  
 D Q D V R R T I E A L G I P R K T F Y D K L Q R H G I N R G G Y S S R K

Figure 2. Nucleotide sequence of *dctB* and *dctD* genes, and deduced amino-acid sequence of their products. Possible ribosome binding sites and ATG initiation codons are overlined. The potential membrane-spanning regions of *dctB* product are underlined. Sites of Tn5 insertions are indicated by arrowheads together with the number of the induced mutant strain. Strains 660, 533, 535, 543, 536 and 659 are *dctB* mutants, and strains 538, 663 and 537 *dctD* mutants (11).



Figure 3. Alignment of C-terminal regions of products homologous to *dctB* product (RLDCTB), including the products of *nrB* from *Klebsiella pneumoniae* (KPNTB; 27) and *Bradyrhizobium* spp. [*Parasponia*] (BPNTB; 22), *envZ* (ECENVZ; 20), *cpxA* (ECCPXA; 23), *phoR* (ECPHOR; 24) and *phoM* (ECPHOM; 26) from *Escherichia coli*, and *virA* (ATVIRA; 25) from *Agrobacterium tumefaciens*. The multiple alignment was prepared from the pair-wise alignments determined using ALIGN by minimizing the number and location of gaps manually. This procedure slightly reduced the alignment scores between any given pair of sequences, but highlighted strongly conserved regions. Boxed residues are conserved, defined as more than 70% of amino acids belonging to one of the groups P A G S T; H K R; Q N E D; I L V M F Y W C.

All the gene products homologous to DctB except NtrB contain two regions of sufficient hydrophobicity to span the membrane, and *virA* and *envZ* products have been shown to reside in the inner membrane (25,28). The hydropathy profile of DctB (Fig. 4) showed that it also had two uncharged strongly hydrophobic regions, one at residues 25 to 42 and one at residues 321 to 338. The first potential membrane-spanning region was preceded by three arginine residues, while the second region was preceded by two arginine residues and

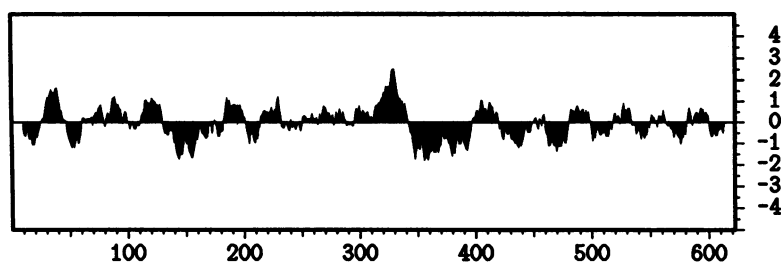


Figure 4. Hydropathy profile of *dctB* product at window-length 19, calculated according to Kyte and Doolittle (29). Hydrophobic regions have positive value and hydrophilic regions negative value.

followed by three arginine residues. Similar hydrophobic regions of 17 or more residues surrounded by charged residues are known to be able to span the hydrophobic interior of the membrane (30,31). Hence the first hydrophobic region of DctB may act as an uncleaved signal sequence, and the second region as a "stop-transfer" or anchor sequence, as demonstrated for other transmembrane proteins (30,31). Most of the N-terminal 320 amino acids of DctB would then be periplasmically-located, while its conserved C-terminal region would be located in the cytoplasm. The chemosensory receptor proteins of *E. coli* have a similar structure (32) and the periplasmic domain has been shown to bind the ligands (33,34). The binding of ligand is thought to cause a signal to be transduced across the membrane to the cytoplasmic portion of the molecule. Similarly, binding of ligands to the periplasmic faces of the DctB homologues may cause activation of their conserved cytoplasmic regions.

#### Structure of the *dctD* product

Like DctB, the regulatory proteins shown in Fig. 3 act in concert with a second regulatory protein to regulate gene expression (reviewed in 35). Members of this second family which include *ntrC* (10,22), *ompR* (20), *sfrA* (36), *phoB* (37), *virG* (38,39) and "ORF2" (the gene downstream of *phoM* (26)) products share extensive homology to each other and to the products of the *cheY* (40), *cheB* (40), *spo0A* (42) and *spo0F* (43) genes in their N termini of about 120 amino acids. DctD also shared this homology (Fig. 5). The N-terminal region is known to perform a regulatory function in *cheB* product (44). In the two-component systems, it may act as the receptor for signals transmitted through the DctB homologues (22,35,38).

*ompR*, *sfrA*, *phoB*, *virG* and "ORF2" products share homology over their entire length of about 240 amino acids, whereas NtrC diverges sharply after the N-terminal domain. However, the remainder of NtrC is highly homologous to analogous portions of NifA, the positive regulator of *nif* gene expression



Figure 5. Alignment of homologous regions of products conserved with *cdtD* product (RLDCTD), including *ntnC* products from *K. pneumoniae* (KPNTRC; 9,10), *Bradyrhizobium* spp. [Parasponia] (BPNTRC; 22) and *R. meliloti* (RMNTRC; 45); N-terminal regions of the products of *ompR* (ECOMPR; 20), *sfiA* (ECSFRA; 36), *phoB* (ECPHOB; 37), *cheY* (ECHEY, 41) and *cheB* (ECHEB; 41) from *Escherichia coli* (*cheB* and *cheY* from *Salmonella typhimurium* are similar (40)); and central and C-terminal regions of *nifA* products from *K. pneumoniae* (KNIFA; 9,10), *R. meliloti* (RMNIFA; 9) and *R. leguminosarum* (RLNIFA; 46). Alignments were prepared as for Fig. 3. The potential nucleotide-binding site is underlined.



Table 1. Pair-wise alignment scores of the central regions of products homologous to dctD product .

	KPNTRC	BPNTTC	RMNTRC	KPNIFA	RMNIFA	RLNIFA
RLDCTD (145-389)	43.9	46.9	45.8	40.5	38.3	43.6
KPNTRC (140-384)		53.6	58.6	46.2	47.3	44.3
BPNTTC (140-383)			83.4	44.9	45.8	48.4
RMNTRC (139-382)				54.3	42.6	49.5
KPNIFA (212-455)					54.8	59.5
RMNIFA (200-443)						66.2

\* Abbreviations are as in Figure 5.

(9,10). Analysis of DctD revealed that it was highly conserved with NtrC and NifA proteins from several species in the regions where the NtrC and NifA proteins are homologous to each other (Fig. 5). The regions of homology of DctD with NtrC and NifA are consistent with the domain structure previously proposed for the latter proteins (10). Hence DctD and NtrC share a common N-terminal domain linked by a hydrophilic segment (residues 120 to 142 of DctD) to a conserved central region also found in NifA. The central region may consist of more than one functional domain, since it contains segments of strong homology separated by regions of lesser conservation. In particular, the segment corresponding to residues 332 to 345 of DctD is hydrophilic but poorly conserved in all proteins, suggesting it may be an inter-domain linker. Following the conserved central region, and separated from it by a region of variable length, all three products contain a helix-turn-helix motif characteristic of a DNA-binding domain (ref. 10; asterisked in Fig. 5). The linker between the conserved central and C-terminal domains is shortest in DctD, accounting for most of the difference in size between DctD (448 amino acids) and NtrC (469-480 amino acids) products.

The conserved central region and DNA-binding domain are likely to be involved in functions common to NtrC, NifA and DctD. The central conserved region of NtrC, NifA and DctD may be involved in interaction with RpoN or RNA polymerase as all three proteins require RpoN as co-activator (8; reviewed in 7). Interestingly, the homology in the central region between NtrC proteins from different species (mean score 65.2) and between NifA proteins from different species (mean score 60.1) was greater than the homology between NtrC and NifA proteins of the same species (mean score 44.4), or between NifA and DctD from *R. leguminosarum* (score 43.6) (Table 1). This suggests that the products diverged functionally before the species diverged evolutionarily. Hence it seems likely that a progenitor gene encoding the conserved central region has been adapted during evolution for use in several regulatory systems that will require RpoN as co-activator. The consensus sequence characteristic

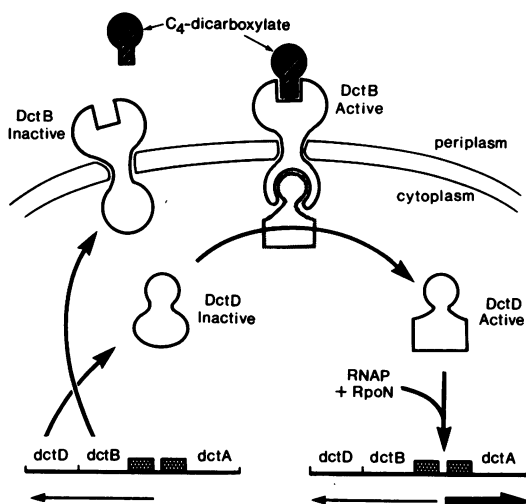


Figure 6. Model for dctA activation. See text for details.

of RpoN-activated promoters has been observed in several structural genes including xylA (47), carboxypeptidase G2 (48) and pilin genes (49) of Pseudomonas species, and flagellin genes of Caulobacter crescentus (50). It seems likely that each of these operons will require as co-activator a regulatory gene product with homology to the central domain found in DctD, NtrC and NifA.

The proposed DNA-binding domain may confer activator specificity on DctD, NtrC and NifA, as NtrC and NifA each appear to recognize a specific nucleotide sequence with dyad symmetry located at least 100 bp upstream of the promoters they activate (51,52,53). The residues within the second helix of the helix-turn-helix motif of the proteins, the helix that is likely to recognise the specific binding site (54), are strongly conserved within the NtrC or NifA proteins, but diverge between NtrC and NifA and with DctD. The dctA promoter also contains regions of dyad symmetry upstream of the transcriptional initiation site (unpublished data) and we are investigating whether these regions are required for transcriptional activation by DctD.

#### DctD contains a potential nucleotide-binding site in its central region

It has recently been shown that NtrB activates NtrC by phosphorylating it (55). Examination of the sequence of NtrB or homologous regions of DctB did not reveal homology to known protein kinases or ATP-binding sites. However a potential ATP-binding site was observed in strongly conserved regions of the central portions of DctD, NtrC and NifA (underlined in Fig. 5). Walker et al

(56) first observed two regions of homology present in several nucleotide-binding proteins, and these regions have been shown to form part of the ATP-binding site of adenylate kinase (57). The first region in the central portion of DctD etc., [G E S/T G S/T G K E] shows one mismatch to the consensus [G X X G X G K S/T], while the second region [A X G G T L F L D] contained two mismatches to the consensus [G X X X L Hy Hy Hy D], where Hy represents an hydrophobic amino acid.

The presence of the potential ATP-binding site suggests that ATP may play a role in the process of transcriptional activation mediated by DctD, NtrC and NifA. In other proteins where it has been investigated, part of the consensus ATP-binding site [G X X G X G K] is thought to form a flexible loop structure that undergoes conformational change in response to either substrate binding, as shown for adenylate kinase, or interaction with another protein or domain (57). It has also been shown that mutations in a similar sequence of the ras oncogene are responsible for the transforming ability of its product (58,59). Mutagenesis of this region of DctD and its homologues should reveal whether this sequence is involved in their activation in response to stimuli.

#### Model for dctA activation

We have shown that DctB and DctD are members of a conserved family of two-component regulatory systems that respond to environmental stimuli (Figs 3 and 5; 22,35,38). These systems include E. coli genes responding to osmolarity (envZ/ompR), nitrogen limitation (ntrB/ntrC), phosphate limitation (phoR/phoB) and toxic compounds (changes in membrane proteins controlled by cpxA/sfra), and genes controlling the virulence of Agrobacterium tumefaciens in response to plant exudate (virA/virG). One component of each system is thought to act as a sensor that transduces a signal to the second component which then activates the response. The conservation in the C-terminal regions of the sensors including DctB and the N-terminal regions of the activators including DctD may reflect a common mechanism of signal transduction.

Given the homology between NtrB and DctB over their C-terminal regions and the homology of NtrC with DctD over their entire length, it is likely that the Ntr and Dct systems will show common mechanisms in their mode of action. On this basis, we propose that Dct regulation involves the following features (Fig. 6): DctB and DctD are constitutively produced at low levels (11). Binding of a C4-dicarboxylate to the periplasmic face of DctB causes an allosteric change in the cytoplasmic portion of DctB, causing activation of DctB. Activated DctB then interacts with the N-terminal region of DctD, the interaction resulting in an allosteric or covalent modification of the remaining portion of DctD. Activated DctD in conjunction with RpoN then activates transcription of dctA. This activation may involve ATP, since DctD

contains a possible ATP-binding site in its central region. Activated DctB may be a protein kinase that phosphorylates DctD, since NtrB is a protein kinase that activates NtrC by phosphorylating it (55). Alternatively, activated DctB may induce a conformational change in DctD rendering it capable of autophosphorylation.

One feature of dctA regulation not accounted for in the model is the observation that a dctA-lacZ fusion is constitutively expressed in a dctA mutant background (5,11). A similar phenomenon has been observed in another two-component system that regulates a transport operon. Strains mutated within the pst operon that encodes the high-affinity phosphate transport system and is regulated by phoR/phoB, express other phosphate-regulated operons constitutively (60). The three-prime gene of the pst operon phoU is not required for phosphate transport, and it has been proposed that it functions directly or indirectly with a co-repressor transported by the pst system to repress pho gene expression under conditions of phosphate excess (61,62,63). However an alternative explanation also applicable to dctA is that the structural gene products are in contact with the sensor component and hold it in a particular mode in the absence of changes in stimuli. We are isolating dctA dctB double mutants to test this idea.

The finding that most Tn5-induced dctB and dctD mutants form nodules that fix nitrogen at wild-type rates suggested that dctA must be regulated differently in symbiotic rhizobia (11). However dctB mutants induced by a chemical mutagen, and one Tn5-induced dctB mutant, strain 659, formed nodules severely affected in their ability to fix nitrogen (1,11). The Tn5 insertion in strain 659 was in the opposite orientation to those in the other dctB and dctD mutants. We suggest that most Tn5-induced dctB mutants are polar on dctD whereas the chemically-induced mutants and the Tn5 insertion in strain 659 are apolar. Hence in these latter strains DctD will be present in inactive form. Inactive NtrC is able to bind to the glnAp2 promoter but does not activate transcription (51,52). Similarly, inactive DctD may bind to the dctA promoter region and interfere with dctA activation by another system, possibly NifA, in the symbiotic state.

### ACKNOWLEDGEMENT

Part of this work was supported by a grant from Hoechst AG to Massachusetts General Hospital.

<sup>3</sup>Present address and address for correspondence: Biotechnica International Inc., 85 Bolton Street, Cambridge, MA 02140, USA

<sup>4</sup>Present address: Department of Molecular and Cellular Biology, Pennsylvania State University, University Park, PA 16802, USA

---

**REFERENCES**

1. Ronson, C.W., Lyttleton, P. and Robertson, J.G. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4284-4288.
2. Ronson, C.W., Astwood, P.M. and Downie, J.A. (1984) *J. Bacteriol.* 160, 903-909.
3. Finan, T.M., Wood, J.M. and Jordan, D.C. (1981) *J. Bacteriol.* 148, 193-202.
4. Finan, T.M., Wood, J.M. and Jordan, D.C. (1983) *J. Bacteriol.* 154, 1403-1413.
5. Ronson, C.W. and Astwood, P.M. (1985) in *Nitrogen Fixation Research Progress*, Evans, H.J., Bottomley, P.J. and Newton, W.E., Eds., pp. 201-207, Martinus Nijhoff, Dordrecht.
6. Guerinot, M.L. and Chelm, B.K. (1987) in *Plant-Microbe Interactions: Molecular and Genetic Perspectives*, Vol. II, Kosuge, T. and Nester, E.W., Eds., pp. 103-146, MacMillan, New York.
7. Gussin, G.N., Ronson, C.W. and Ausubel, F.M. (1986) *Ann. Rev. Genet.* 20, 567-591.
8. Ronson, C.W., Nixon, B.T., Albright, L.M. and Ausubel, F.M. (1987) *J. Bacteriol.* 169, 2424-2431.
9. Buikema, W.J., Szeto, W.W., Lemley, P.V., Orme-Johnson, W.H. and Ausubel, F.M. (1985) *Nuc. Acids Res.* 13, 4539-4555.
10. Drummond, M., Whitty, P. and Wooten, J. (1986) *EMBO Journal* 5, 441-447.
11. Ronson, C.W. and Astwood, P.M., manuscript in preparation.
12. Yanisch-Perron, C., Viera, J. and Messing, J. (1985) *Gene* 33, 103-119.
13. Viera, J. and Messing, J. (1982) *Gene* 19, 259-268.
14. Bankier, A.T. and Barrell, B.G. (1983) in *Techniques in Life Sciences, B5: Nucleic Acids Biochemistry*, Flavell, R.A., Ed., pp. 1-34, Elsevier/North Holland, Limerick.
15. Staden, R. (1982) *Nuc. Acids Res.* 10, 4731-4751.
16. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nuc. Acids Res.* 12, 387-395.
17. Lipman, D.J. and Pearson, W.R. (1985) *Science* 227, 1435-1441.
18. Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) *Meth. Enzymol.* 91, 524-545.
19. Shine, J. and Delgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1346.
20. Comeau, D.E., Ikenaka, K., Tsung, K. and Inouye, M. (1985) *J. Bacteriol.* 164: 578-584.
21. Hall, M.N. and Silhavy, T.S. (1981) *Ann. Rev. Genet.* 15, 91-142.
22. Nixon, B.T., Ronson, C.W. and Ausubel, F.M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7850-7854.
23. Albin, R., Weber, R. and Silverman, P.M. (1986) *J. Biol. Chem.* 261, 4698-4705.
24. Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) *J. Mol. Biol.* 192, 549-556.
25. Leroux, B., Yanofsky, M.F., Winans, S.C., Ward, J.E., Ziegler, S.F. and Nester, E.W. (1987) *EMBO J.* 6, 849-856.
26. Amemura, M., Makino, K., Shinagawa, H. and Nakata, A. (1986) *J. Bacteriol.* 168, 294-302.
27. MacFarlane, S.A. and Merrick, M. (1985) *Nuc. Acids Res.* 13, 7591-7606.
28. Liljestrom, P. (1986) *FEMS Microbiol. Lett.* 36, 145-150.
29. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
30. Davis, N.G., Boeke, J.D. and Model, P. (1985) *J. Mol. Biol.* 181, 111-121.
31. Davis, N.G. and Model, P. (1985) *Cell* 41, 607-614.
32. Krikos, A., Mutoh, N., Boyd, A. and Simon, M.I. (1983) *Cell* 33, 615-622.
33. Krikos, A., Conley, M.P., Boyd, A., Berg, H.C. and Simon, M.I. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1326-1330.
34. Park, C. and Hazelbauer, G.L. (1986) *J. Bacteriol.* 167, 101-109.
35. Ronson, C.W., Nixon, B.T. and Ausubel, F.M. (1987) *Cell* 49, 579-581.
36. Drury, L.S. and Buxton, R.S. (1985) *J. Biol. Chem.* 260, 4236-4242.

37. Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) *J. Mol. Biol.* 190, 37-44.
38. Winans, S.C., Ebert, P.R., Stachel, S.E., Gordon, M.P. and Nester, E.W. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8278-8282.
39. Melchers, L.S., Thompson, D.V., Idler, K.G., Schilperoort, R.A. and Hooymaas, P.J.J. (1986) *Nuc. Acids Res.* 14, 9933-9942.
40. Stock, A., Koshland, D.E. and Stock, J. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7989-7993.
41. Mutoh, N. and Simon, M.I. (1986) *J. Bacteriol.* 165, 161-166.
42. Ferrari, F.A., Trach, K., LeCoq, D., Spence, J., Ferrari, E. and Hoch, J.A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2647-2651.
43. Trach, K.A., Chapman, J.W., Pigot, P.J. and Hoch, J.A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7260-7264.
44. Simms, S.A., Keane, M.G. and Stock, J. (1985) *J. Biol. Chem.* 260, 10161-10168.
45. Szeto, W.W., Nixon, B.T., Ronson, C.W. and Ausubel, F.M. (1987) *J. Bacteriol.*, 169, 1423-1432.
46. Gronger, P., Manian, S.S., Reilander, H., O'Connell, M., Priefer, U.B. and Puhler, A. (1987) *Nuc. Acids Res.* 15, 31-49.
47. Inouye, S., Ebina, Y., Nakazawa, A. and Nakasawa, T. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1688-1691.
48. Minton, N.P. and Clarke, L.E. (1985) *J. Mol. Appl. Genet.* 3, 26-35.
49. Johnson, K., Parker, M.L. and Lory, S. (1986) *J. Biol. Chem.* 261, 15703-15708.
50. Minnich, S.A. and Newton, A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 1142-1146.
51. Hirschmann, J., Wong, P.-K., Sei, K., Keener, J. and Kustu, S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7525-7529.
52. Reitzer, L.J. and Magasanik, B. (1986) *Cell* 45, 785-792.
53. Buck, M., Miller, S., Drummond, M. and Dixon, R. (1986) *Nature* 320, 374-378.
54. Pabo, C.O. and Sauer, R.T. (1984) *Ann. Rev. Biochem.* 53, 293-321.
55. Ninfa, A.J. and Magasanik, B. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5909-5913.
56. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO Journal* 1, 945-952.
57. Fry, D.C., Kuby, S.A. and Mildvin, A.S. (1986) *Proc. Natl. Acad. Sci. USA* 83, 907-911.
58. Tabin, C.J., Bradley, E., Bargmann, C., Weinberg, R., Papageorge, A., Scolnick, E.M., Dhar, R., Lowy, D. and Chang, E. (1982) *Nature* 300, 143-148.
59. Reddy, E.P., Reynolds, R.K., Santos, E. and Barbacid, M. (1982) *Nature* 300, 149-152.
60. Zuckier, G. and Torriani, A. (1981) *J. Bacteriol.* 145, 1249-1256.
61. Surin, B.P., Rosenberg, H. and Cox, G.B. (1985) *J. Bacteriol.* 161, 189-198.
62. Amemura, M., Makino, K., Shinagawa, H., Koboyashi, A. and Nakata, A. (1985) *J. Mol. Biol.* 184, 241-250.
63. Shinagawa, H., Makino, K. and Nakata, A. (1983) *J. Mol. Biol.* 168, 477-488.