JOURNAL OF BACTERIOLOGY, Apr. 1989, p. 1952–1959
0021-9193/89/041952-08$02.00/0

Vol. 171, No. 4

Evidence that the Transcription Activator Encoded by the *Pseudomonas putida nahR* Gene Is Evolutionarily Related to the Transcription Activators Encoded by the *Rhizobium nodD* Genes

MARK A. SCHELL* AND MANDHANA SUKORDHAMAN

Department of Microbiology, University of Georgia, Athens, Georgia 30602

Received 26 October 1988/Accepted 6 January 1989

The *nahR* gene of the 83-kilobase naphthalene degradation plasmid NAH7 of *Pseudomonas putida* encodes a 34-kilodalton polypeptide which binds to the *nah* and *sal* promoters to activate transcription of the degradation genes in response to the inducer salicylate. The DNA sequence of the *nahR* gene was determined, and a derived amino acid sequence of the NahR protein was obtained. A computer search for homologous proteins showed that within the first 124 amino-terminal residues, NahR has approximately 35% identity with the transcriptional activator proteins encoded by the *nodD* genes of *Rhizobium* species. Allowing for ultraconservative amino acid substitutions, greater than 47% overall similarity was found between NahR and NodD, while 32% similarity was found between NahR and another transcription activator, LysR of *Escherichia coli*. The region of greatest similarity among all three proteins contained a probable helix-turn-helix DNA-binding motif as suggested by homology with the proposed consensus sequence for Cro-like DNA-binding domains. The high level of amino acid identity between NahR and NodD, in conjunction with the observations that *nahR* and *nodD* are 45% homologous in DNA sequence, are divergently transcribed from homologous promoters near the structural genes they control, and have similar DNA-binding sites, strongly suggests that these two genes evolved from a common ancestor.

The 83-kilobase (kb) NAH7 plasmid from the soil bacterium *Pseudomonas putida* encodes enzymes for the metabolism of naphthalene or salicylate as the sole carbon and energy source (8). The 14 genes encoding the enzymes for this metabolism are organized in two operons: *nah* (*nahA-F*), encoding six enzymes required for metabolism of naphthalene to salicylate and pyruvate, and *sal* (*nahG-M*), encoding eight enzymes which metabolize salicylate to pyruvate and acetaldehyde (43). Expression of these enzymes is increased over 20-fold by growth in the presence of the inducer salicylate or the nonmetabolizable inducer 2-aminobenzoate (2). Induction requires the product of only one regulatory gene, *nahR*, encoding a 36-kilodalton (kDa) polypeptide (33, 36, 44); it mediates induction by activating transcription from both the *nah* and *sal* promoters (33). NahR activates transcription from the *sal* promoter in *trans*, only in the presence of salicylate. Deletion experiments have shown that the *sal* promoter sequences between −83 and −45 are required for both DNA binding and transcription activation by NahR (35, 36). Other experiments have shown that the NahR protein protects a highly conserved region (−82 to −47) of both the *nah* and *sal* promoters from DNase I digestion in the presence or absence of salicylate and that this binding is necessary for transcription activation (35). It has been suggested that the promoter-bound NahR protein, upon binding salicylate, undergoes a conformational change which results in increased transcription from that promoter (35).

An analogous system is found in another genus of soil bacteria, *Rhizobium*. Members of this genus harbor large megaplasmids encoding gene products which allow these bacteria to colonize plant roots by formation of symbiotic root nodules containing nitrogen-fixing bacteroids (21). Initiation of nodule formation is induced by transcriptional activation of at least three plasmid-encoded operons (*nodABC*, *nodFE*, and *nodH*) in response to various signal molecules in root exudates (12, 24, 28). This induction requires the product of a regulatory gene, *nodD*, and specific flavonoid inducers (24, 37). Recent experiments suggest that NodD binds specifically to conserved sequences (nod box; 30) located upstream (−80 to −27) of the *nodABC*, *nodH*, and *nodFE* operons (12, 17); binding apparently does not require the presence of flavonoid inducers.

Present knowledge of the structure and mechanism of action of procaryotic transcriptional activator proteins is derived from a few *Escherichia coli* proteins which have been intensely studied by genetic, biochemical, and biophysical methods. Many other positive regulatory gene products are more poorly understood at the molecular level because they are less amenable to such analysis. However, derivation of amino acid sequences from DNA sequence analysis followed by computer searches for homologies with previously characterized proteins has provided insight into structure-function relationships and possible mechanisms of action of regulatory proteins. Thus, to gain insight into the evolution and mechanism of action of NahR, we determined its DNA sequence and derived from it the amino acid sequence of its activator product. A computer search for proteins homologous to NahR revealed a striking similarity between NahR and an identically sized transcriptional activator encoded by the *nodD* genes of *Rhizobium* species. The search also found that NahR is partially homologous to the LysR activator of *E. coli* and that NahR is a member of the LysR family of procaryotic transcriptional activator proteins recently described by Henikoff et al. (16).

MATERIALS AND METHODS

Bacterial strains and plasmids. Bacterial strains used in this study were *E. coli* JM101 and *E. coli* JM83 (42).

* Corresponding author.

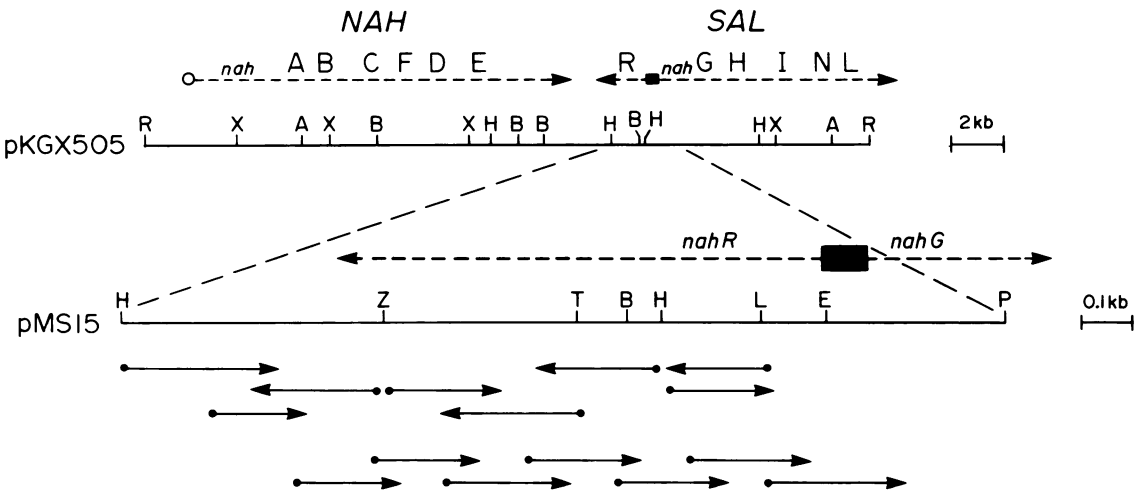

FIG. 1. Physical map and location of *nahR*. Upper portion shows the approximate location of the transcripts (– – – →) and genes (from data in references 15 and 43) of the *nah* operon (*nahABCFDE*), the *sal* operon (*nahGHINL*), and *nahR* on the 23-kb *Eco*RI fragment of plasmid NAH7 cloned on plasmid pKGX505 (32). The *nah* promoter is indicated by an open circle and the *sal-nahR* promoters are indicated by a solid box. The middle section shows the restriction endonuclease cleavage map of pMS15, the location of the *nahR* transcript (←– – –), and the divergent *sal* transcript (*nahG* – – – →). Strategy for sequencing *nahR* (solid arrows) is shown under the pMS15 map. Restriction endonuclease cleavage sites are designated as follows: R, *Eco*RI; X, *Xho*I; A, *Sma*I; B, *Bgl*II; H, *Hin*dIII; Z, *Sph*I; T, *Stu*I; L, *Sal*I; P, *Pst*I; E, *Eag*I.

Plasmids used were pMS10 (35), pMS15 (36), and pTZ18U/19U (5, 23).

Construction of plasmids. All plasmids used for DNA sequence determination of *nahR* were derived by subcloning restriction fragments from the NAH7-derived inserts on pMS15 and pMS10 into pTZ18U or pTZ19U vectors (5, 23) (United States Biochemical Corp., Cleveland, Ohio) and by subsequently making restriction fragment deletion derivatives of the *nahR* subclones. Methods used for subcloning and deletion were standard and have been described previously (22, 32, 33).

DNA-sequencing procedures. DNA sequence determination was by the dideoxy-chain termination method of Sanger et al. (31) utilizing a kit (no. 410) from New England BioLabs, Inc. (Beverly, Mass.) with [α-^{35}S]dATP (Amersham Corp., Arlington Heights, Ill.) and pTZ18U/19U *nahR* subclones by the strategy shown in Fig. 1. Single-stranded template DNA was prepared from *E. coli* JM101 containing the pTZ subclones after infection with the helper bacteriophage M13K07 by the method recommended by the United States Biochemical Corp. In some cases, deaza-GTP (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) was used in sequencing reactions to reduce G+C compression. The entire *nahR* DNA sequence derived from pTZ18/19U subclones utilizing the M13 sequencing primer (no. 1212; New England BioLabs) was rechecked utilizing *nahR*-specific 17-mer oligonucleotide primers derived from preliminary sequence data and synthesized on a Applied Biosystems 380A DNA synthesizer. The DNA sequence of the 1.5-kb DNA fragment between the *Eag*I and left terminal *Hin*dIII sites of the NAH7-derived insert on pMS15 was sequenced on both strands except for the 70 base pairs (bp) upstream of the *Sph*I site and the last 300 bp upstream of terminal *Hin*dIII site, 150 bp downstream of the end of *nahR* coding region. The 280-bp DNA sequence of the region upstream of the *Sal*I site of the NAH7 insert on pMS15 (into *nahG*) was reported previously (34).

Computer analysis of nucleotide and protein sequences. Analysis of the *nahR* DNA sequence and NahR protein sequence was performed with the Pustell sequence analysis programs (Version 5.1) from International Biotechnologies, Inc. (New Haven, Conn.). The initial computer search for homologous proteins was performed by the FASTP method of Lipman and Pearson (20) on the Protein Identification Resource of the National Biomedical Research Foundation (4,253 sequences; 1,029,056 residues). Later searches were performed against procaryotic and viral DNA sequences in the GenBank (1987 update; 4,993,541 residues) with the Cyborg database manager (Version 1.8) of International Biotechnologies, Inc.

RESULTS

DNA sequence analysis and derivation of NahR amino acid sequence. Previous experiments with plasmid pMS15 (Fig. 1) showed that *nahR* was contained within a 1.5-kb *Pst*I-partial *Hin*dIII fragment derived from plasmid NAH7 (36). The *nahR* promoter, transcription start site, and putative amino-terminal (N-terminal) amino acid sequence had been previously localized on this fragment near the *Sal*I site (34). Thus, we subcloned the 1.45-kb region downstream of the *nahR* transcription start site and determined its DNA sequence by the strategy shown in Fig. 1. The resultant DNA sequence (Fig. 2; GenBank accession no. J04233) has only one open reading frame (ORF) that is longer than 90 residues and begins within 350 bp of the *nahR* transcription start site. This ORF begins at nucleotide 55, 5 bp downstream from a possible ribosome-binding site (i.e., a 7-bp sequence with significant homology to *Pseudomonas aeruginosa* 16s rRNA) (34). The ORF terminates at nucleotide 955, giving the encoded polypeptide (300 amino acids) a calculated molecular mass of 34 kDa, a size within 3% of the observed size of the NahR polypeptide detected in *E. coli* maxicells (36). Insertion of termination codons into the *Sph*I or *Ssp*I sites near the carboxy terminus (C terminus) of the NahR ORF (Fig. 2) resulted in synthesis of nearly full length NahR proteins which were dramatically reduced (>15-fold) in their ability to specifically bind to the NahR-binding site of the *sal*

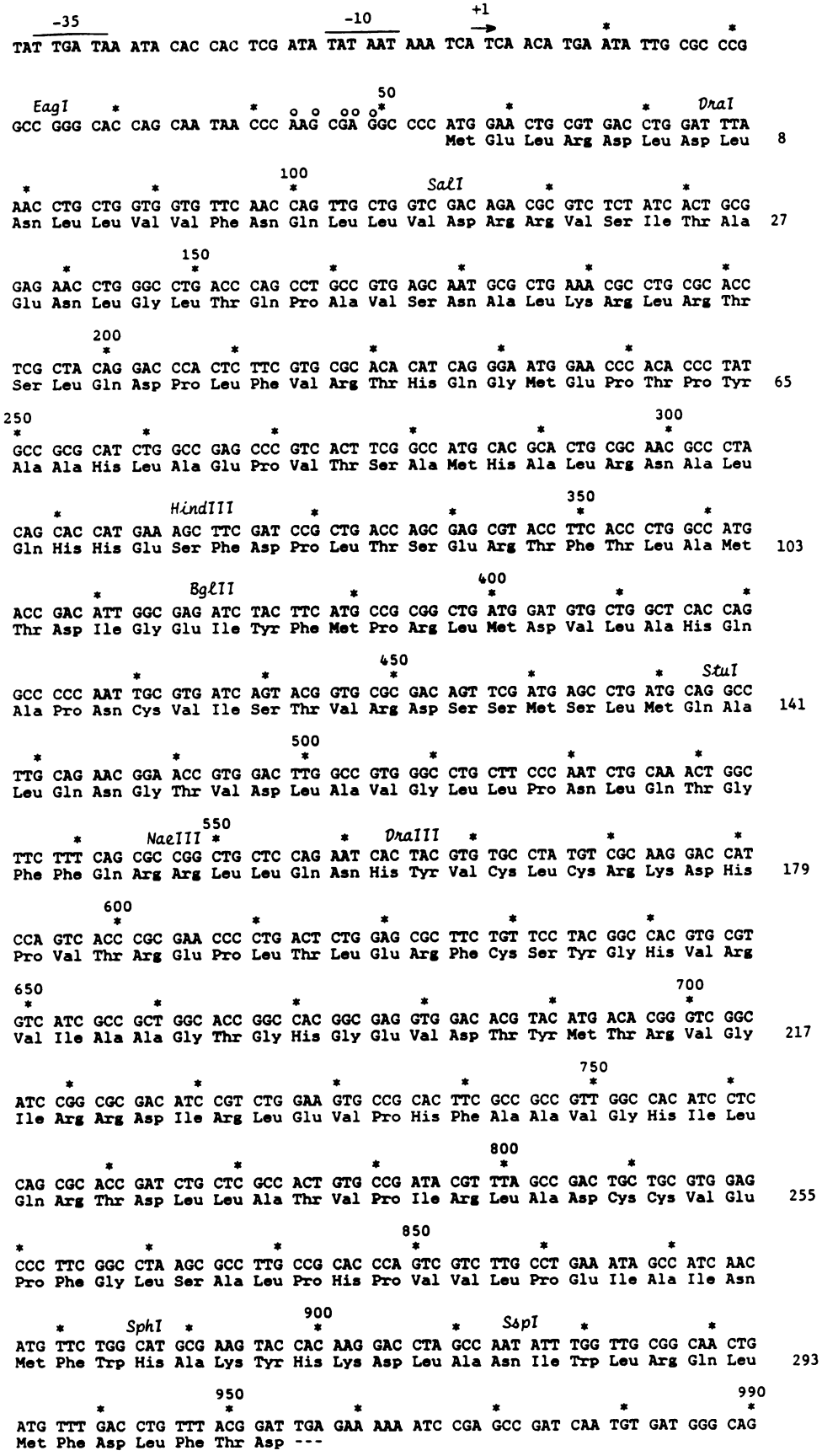

FIG. 2. Nucleotide sequence of *nahR* and derived amino acid sequence of NahR protein. Nucleotide sequence (5′→3′; noncoding strand) starts at the −35 sequence of the *nahR* promoter (28 bp upstream of the divergent transcription start site of the *sal* operon [34]). Promoter elements (−35 and −10 sequences) are overlined. Nucleotide numbering (on top of sequence) begins with transcription start site of *nahR* (+1, →); asterisks are placed every 10 bp; possible ribosome-binding site of *nahR* is indicated by open circles near nucleotide 50. Amino acid residue numbers of the NahR protein for last residue on that line are on the right. The nucleotide sequence of the NAH7 DNA following the *nahR* coding region (nucleotides 991 to 1446) up to the terminal *Hin*dIII site of pMS15 was determined but is not shown.

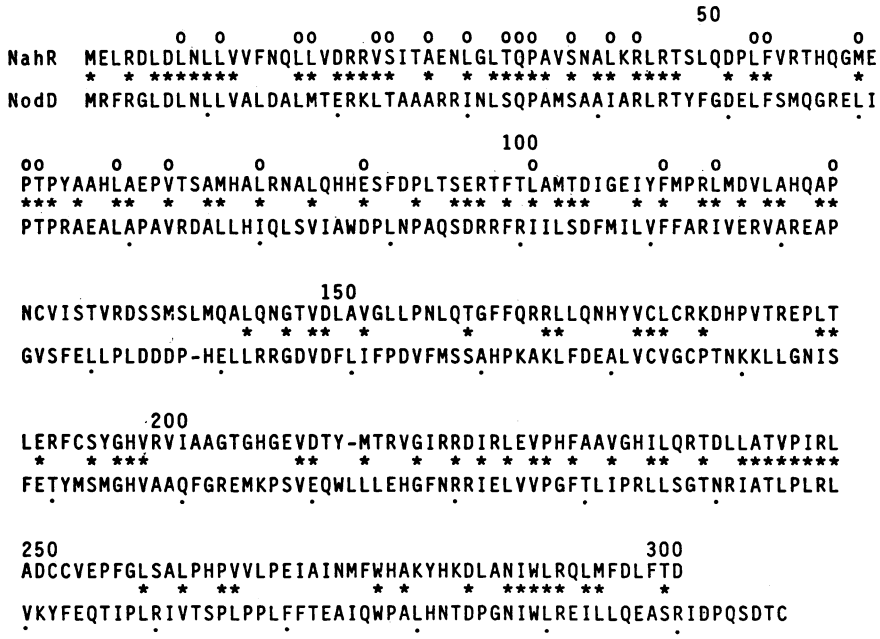

FIG. 3. Homology between NahR and NodD proteins. The derived amino acid sequences of NahR (upper lines) and NodD from *R. meliloti* (9) (lower lines) are presented in single-letter code aligned by the N termini. Asterisks indicate identical-similar residues (see text). Similar residues found in NodD, NahR, and LysR are indicated (○). Dots are placed every 10 residues.

promoter; this provides further evidence that the 34-kDa ORF is NahR (M. Schell and P. Brown, unpublished data).

The amino acid composition of the polypeptide encoded by the NahR ORF is similar to the composition expected for a procaryotic protein (7) except that it contains 60% more leucine but 40% less isoleucine than average. NahR contains 80% less lysine but 60% more arginine than the average protein (7) and has a calculated pI of 6.8. The G+C content of the NahR-coding region (60%) is similar to that found for the *P. putida* chromosome (26) and the NAH7 plasmid (R. Farrell, Ph.D. thesis, University of Illinois, Urbana, 1979). Codon utilization is typical for a *P. putida* gene in that U and A, especially at the third position, are uncommon (25). For example, GUA (Val), GGU (Gly), and UCA (Ser) are not used in NahR, whereas they are fairly common in *E. coli* (25); GUU is used once in 23 valine codons and AAA is used once in 4 lysine codons, whereas in *E. coli* these are the preferred codons. All data strongly suggest that the 34-kDa ORF is the NahR polypeptide.

Homology between NahR and NodD. The Protein Identification Resource of the National Biomedical Research Foundation was searched for proteins with homology to the derived amino acid sequence of NahR. Initially, two proteins with significant amino acid homology to NahR were detected: NodD (9) and LysR (39). A more extensive search of the Genbank for coding sequences homologous to NahR identified 15 putative proteins with homology scores greater than 122 (Hatch level 2; modified PAM250 matrix). The top seven scores were for NodD proteins from different *Rhizobium* species. Although the three highest scores (430, 425, and 418) were for NodDs from the broad-host-range *Rhizobium* strain MPIK3030 (18), *Bradyrhizobium* strain ANU289 (1), and *Rhizobium leguminosarum* (37), respectively, the greatest homology and best alignment with the least gaps was for NodD from *Rhizobium meliloti* (9). Significant homology with LysR, LeuO (16), and IlvY (16) was also detected. Homology between NodDs and LysR has been described previously (1, 16). All these proteins (except LeuO) are nearly identical in size (305 ± 10 amino acids) and have been reported to be transcriptional activators of divergently transcribed structural genes (16, 33).

The aligned amino acid sequences of NodD and NahR are shown in Fig. 3 with identities and similarities marked. Amino acid similarities refer to nonidentical pairs of amino acids which are interchangeable in many proteins since they represent amino acid substitutions which occur frequently in evolution (4). The only pairs of similarities used here (I = L; S = T; V = L; M = L; V = I; D = E; K = R) represent seven of the most frequent substitutions occurring in nature; their scores in the PAM250 matrix of Dayhoff (4, 7) are only 50% less than the score for identity. Henceforth, the term similarity will be used to refer to the sum of both amino acid similarities and identities.

The homology between NodD and NahR is striking (Fig. 3). The two transcriptional activators differ in size only by the 8 additional residues at the C terminus of NodD, and within the first 125 residues the two proteins are 36% identical and 52% similar. In comparison, the identity between different NodD proteins encoded by genes from different *Rhizobium* species is 70% within the first 125 residues (1). Homology between the last 175 residues of NahR and NodD is less (32% similar, 23% identical) and is concentrated in four regions: (i) residues 186 to 199 with 54% similarity, (ii) residues 220 to 249 with 62% similarity, (iii) residues 277 to 295 with 55% similarity, and (iv) between residues 142 and 173 with 30% similarity. Although between residues 250 and 276 there is almost no similarity, this region of both proteins is very hydrophobic as measured by the method of Kyte and Doolittle (19).

Comparison of the amino acid sequence of NahR with that of LysR shows much less homology (22% identity, 36% similarity) within the first 125 N-terminal residues; similarity between the second halves of the proteins is low and diffuse (data not shown). A three-way comparison between the first

```
                    +++   +     +
Cro       QTKTAKDLGVYQSAINKAIHAGR
           * **  *** * **  **   *
NahR      VSITAENLGLTQPAVSNALKRLR
           *  *  * * * **  **   *
NodD      LTAAARRINLSQPAMSAAIARLR

CONSENSUS xoxxA±xhxxoQxAhxxAhxxxR

           *  ** *  *** **  * *
LysR      LTEAAHLLHTSQPTVSRELARFE
```

FIG. 4. Comparison of common domains from NahR, NodD, and LysR with DNA-binding domain of Cro. Amino acid sequences (single-letter code) of the common domain of NodD and NahR (residues 23 to 45) are aligned with the HTH DNA-binding domain of Cro (residues 16 to 38). Residues implicated in DNA binding by Cro are marked (+); residues similar to those of Cro are indicated (*). Below is a suggested consensus sequence for the common domain in Cro, NahR, and NodD; o, residue with hydroxyl R group; ±, charged residue; h, hydrophobic residue; x, no preference. The common domain of LysR (residues 21 to 43) is shown below; residues similar to the suggested consensus sequence or present in both NahR and NodD are indicated (*).

125 residues of NahR, NodD, and LysR utilizing our stringent similarity rules (Fig. 3) showed that although the overall similarity among all three proteins in this region was only 21%, one domain (residues 23 to 45; "common domain") showed a high degree of similarity (40%) in all three transcriptional activators.

Identification of possible DNA-binding domain. The region which was 40% similar in NodD, NahR, and LysR (common domain; Fig. 4) was examined in more detail. Since all three proteins activate transcription, they are probably DNA-binding proteins, and it is possible that the common domain is involved in DNA binding. Prediction of the probable secondary structure of each of the common domains by the method of Chou and Fasman (3) indicated possible α-helical structure. Furthermore, all three regions show evidence for existence of a helix-turn-helix (HTH) motif when analyzed by the rules of Pabo and Sauer (27). The HTH motif is often found in DNA-binding proteins, and in some cases, it has been shown that it contains the amino acid residues which physically interact with the nucleotides of the DNA-binding sites (27). Comparison of the common domains of NodD and NahR with the DNA-binding domain (i.e., HTH) of the λ Cro protein showed 45 and 55% similarity, respectively (Fig. 4). Analysis of these regions for an HTH structure by the numerical method of Dodd and Egan (6) also suggested a reasonable probability that the common domain is a Cro-like DNA-binding motif (Dodd-Egan scores: NahR, 1,171; NodD, 749; LysR, 1,475; Cro, 1,699). Both NahR and NodD have Gln (Q) at position 34, Arg (R) at position 45, and a residue containing a hydroxyl R group at position 33 in the common domain; these residues are also found at the same positions in the Cro HTH and have been implicated in its interactions with its DNA-binding site (27). The DNA sequence of the Cro consensus binding site and the central portion of the NahR consensus binding site are also very similar (Fig. 5), providing further evidence that the common domain could be involved in DNA binding. Recently, Henikoff et al. (16) have independently come to a similar conclusion about the common domain of LysR and NodD. Our data and those of Henikoff suggest that these proteins utilize a Cro-like mechanism for DNA binding. However, most of the evidence is circumstantial, and it will be necessary to use

```
NahR   TATTCA-CGCTGGTGATA
       *** ** *** *******
Cro    TAT-CACCGCCGGTGATA
```

FIG. 5. Comparison of DNA-binding sites of NahR and Cro. Shown are the DNA sequences (5'→3') of the consensus binding site of λ Cro protein (27) and the central portion (nucleotides −75 to −59) of the NahR binding site (35) of the *nah* operon promoter. Homologous nucleotides are marked (*).

site-directed mutagenesis to alter the amino acid sequence of these domains to define the role of this domain in DNA binding.

DNA homology between *nahR* and *nodD* and their target sites. The NahR-NodD amino acid sequence homology suggests parallel evolution or a common precursor. Comparison of the DNA sequences of the two genes suggests common ancestry since the coding regions of *nodD* and *nahR* are 45% homologous in DNA sequence with few gaps (data not shown). In addition, the 23-bp sequences (−18 to +5) surrounding their transcription start sites are nearly 70% homologous (Fig. 6A). The observation that both *nodD* and *nahR* are divergently transcribed from start sites near the promoters of one set of structural genes they control provides further suggestive evidence for a common ancestor. In fact, the transcription start sites of the regulatory and structural genes are so close (50 bp for *nahR* and 20 bp for *nodD*) that the promoters for the regulatory and structural genes partially overlap (Fig. 6A), suggesting a potential for simultaneous regulation. The LysR system apparently has a similar divergent promoter organization (39, 40).

Comparison of the upstream sequences of a NahR-regulated promoter (*nah*) and a nodD-regulated promoter (*nodH*) (13) aligned by their transcription start sites showed extensive homology between −80 and −70 (Fig. 6B). The 7-bp sequence TATTCAC is found at the same location in both promoters. This 7-bp sequence is part of the *nah* promoter region that is protected from DNase I by the NahR protein (35) and part of the conserved nod box sequences of the *nodH* promoter which are probably involved in NodD binding (12, 13, 17). In addition, mutation of nucleotide 2 or 3 in the TATTCAC sequence eliminates NahR binding and activation of a NahR-regulated promoter (35). Other promoter sequence homologies are evident (GAT at −60; TNCAA at −49; TTNACNAAT at −35) and are separated by approximately one helical turn, placing them in adjacent grooves on one face of the helix. These results suggest that NodD and NahR have similar DNA-binding sites and utilize similar mechanisms to recognize, bind to, and activate the promoters they control. A consensus binding site sequence for NahR-regulated promoters is very similar to one derived for an analogous location in three NodD-regulated promoters (Fig. 6C) and the Cro binding site (Fig. 5). However, under conditions in which we observed binding of NahR to one of its regulated promoters, we were unable to detect binding of NahR to a NodD-regulated promoter (*nodA*) (data not shown). Further experimentation is required to clarify these apparent binding site homologies.

DISCUSSION

DNA sequence analysis of the *nahR* gene detected a 34-kDa ORF which probably represents the amino acid sequence of the NahR protein. This amino acid sequence has extensive homology with the derived sequences of the NodD proteins of *Rhizobium* species and lower but significant

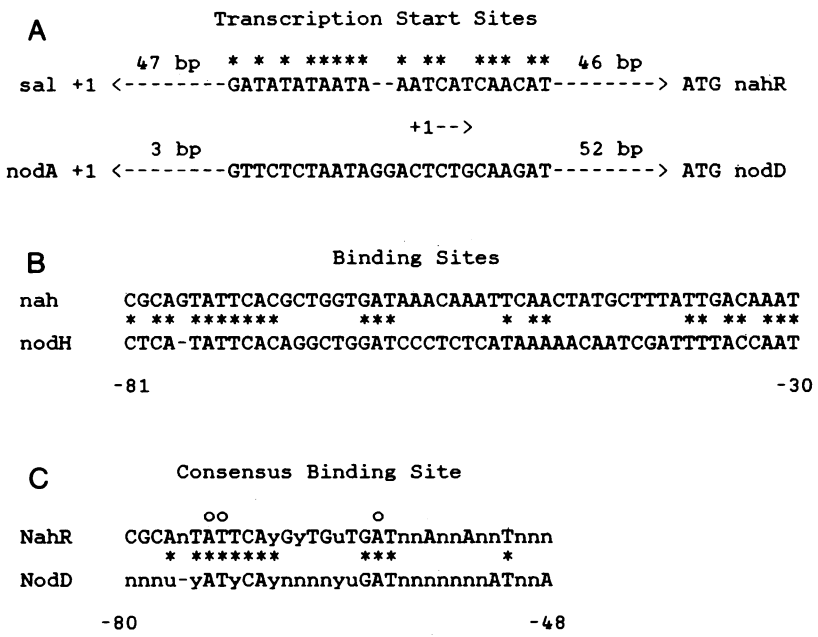

FIG. 6. DNA sequence homology between the transcription start sites and DNA-binding sites of NahR and NodD. (A) The 23-bp DNA sequence surrounding the transcription start sites (+1 - - →) of *nahR* (34) and *nodD* (11) aligned by the start sites. Homologous nucleotides for panels A, B, and C are indicated (*). Distances to the ATG start codons of each gene and distances to the transcription start site of the divergently transcribed structural genes (+1) are also indicated. (B) Upstream promoter sequences (ca. −80 to −30; aligned by the transcription start sites) of the *nah* operon and *nodH* gene. (C) Consensus sequences for upstream promoter regions (−80 to −48) for NahR-regulated promoters (*nah* and *sal*) and NodD-regulated promoters (*nodFE*, *nodH*, *nodABC*) (13). Nucleotides required for NahR activation of NahR-regulated promoters (35) are indicated (o). n, No nucleotide preference; y, pyrimidine; u, purine.

homology with the LysR gene product of *E. coli*. All three of these proteins are nearly identically sized transcriptional activators, and it is likely that the regions of greatest similarity between them represent protein domains involved in common functions (e.g., DNA binding or transcriptional activation or both). The amino acid sequence of the domain of NahR containing the region of greatest similarity between the three proteins (common domain; residues 23 to 45) had a protein sequence which was very similar to the amino acid sequence of the HTH DNA-binding motif of the lambda Cro protein. In addition, the common domains from all three proteins showed good fit to the consensus sequence proposed for HTH DNA-binding motifs (6, 27).

Just prior to submission of this manuscript, Henikoff et al. (16) reported discovery of a family of procaryotic transcriptional activators evolutionarily related to LysR. This family included the NodD proteins and seven other proteins or ORFs (LysR, IlvY, LeuO, CysB, MetR, AmpR, AntO), all of which are approximately 300 residues in size and most, if not all, of which are transcriptional activators of divergently transcribed structural genes. Comparison of the NahR protein sequence with the Henikoff consensus sequence for this protein family clearly showed that NahR is a member of the family (42% similarity in the first 150 residues; 32% similarity overall). However, NahR has much greater similarity to the NodD proteins and LeuO ORF than to any other family members, suggesting the possible existence of a subfamily within the LysR family of transcriptional activators. In addition, by analysis of the regions of greatest sequence similarity, Henikoff et al. (16) independently concluded that what we refer to as the common domain of NahR, LysR, and NodD contains an HTH motif and is probably involved in DNA binding. In support of our HTH hypothesis, we have isolated activation-deficient mutants of *nahR* and shown that several of these NahR proteins which are defective in binding to the *sal* promoter are also altered at amino acid positions in the proposed HTH region (M. A. Schell, P. Brown, and S. Raju, manuscript in preparation). Burn and Johnson (Proceedings of the 4th International Symposium on Plant-Microbe Interactions, May 1988) have obtained similar preliminary results with the HTH of NodD protein.

Comparison of various features of the *nodD* and *nahR* regulons provides a significant amount of circumstantial evidence suggesting that the *nodD* gene of *Rhizobium* species is evolutionarily related to the *nahR* gene of *Pseudomonas* species. The similarities in DNA and amino acid sequence, function, promoter structure and organization, target site, autoregulation (17, 35), and plasmid location strongly suggest that *nahR* and *nodD* share a common ancestor. The NodD protein has apparently evolved into a variety of different NodDs present in various *Rhizobium* species. Each NodD protein responds to different flavonoids in root exudates of specific host plants and probably influences host range by its recognition specificity for molecules in root exudates (29, 38). While all NodD proteins have very similar N-terminal amino acid sequences, the C-terminal halves diverge substantially. Horvath et al. (18) have shown that inducer recognition specificity of NodD probably lies in the C-terminal domain; therefore, the C-terminal divergence of NodD is probably the result of the evolution of different recognition specificities. Since NahR shows the same pattern of homology and divergence with NodD, it is possible that NahR is a divergent form of NodD with recognition specificity for salicylate. In support of this idea, the NahR inducer molecule salicylate is vaguely similar in structure to portions of some flavonoid inducers of NodD proteins in that both contain hydroxylated aromatic rings. All these observations are consistent with the proposal that NahR and

NodD evolved from a common ancestral transcriptional activator gene. Evolutionary relatedness of *nahR* and *nodD* is not unexpected since genetic exchange between *Rhizobium* species, *Pseudomonas* species, and other bacteria in the rhizosphere is probably extensive, and the fact that both these genes are plasmid borne may have facilitated evolution of the genes.

Other NAH7 plasmid genes have been shown to have extensive DNA and protein sequence homology with genes of similar function on another degradative plasmid (TOL pWW0) found in *Pseudomonas* species. Harayama et al. (15) have shown that the *nahH* gene-gene product (which is part of a *nahR*-regulated transcript) has 80% DNA and 84% amino acid sequence homology with the *xylE* gene-gene product of the TOL plasmid. Additional sequence homology between *nahH* and *xylE* was detected downstream of the coding regions, and it was suggested that the *xylEGFJ* and *nahHINL* genes evolved from a common ancestor (15). Hybridization experiments suggest homology between portions of NAH7 and other aromatic degradation plasmids (10), but the nature and extent of the homology has not been clarified. We observed that the putative N-terminal amino acid sequence of the *clcR* gene product of *Pseudomonas* species (14) shows significant homology with the N-terminal sequence of NahR (40% similarity in the first 75 N-terminal residues), as does the N-terminal sequence of the *tfdO* ORF (16, 41) of *Alcaligenes eutrophus* (35% similarity in the first 75 residues). These genes are found on two different plasmids, each encoding for the degradation of the aromatic hydrocarbon 3-chlorocatechol. *clcR* is thought to encode a transcriptional activator of divergently transcribed structural genes involved in degradation of 3-chlorocatechol (14; A. Chakrabarty, personal communication). All these results suggest that genetic interaction and cross-talk between different soil bacteria and their resident plasmids is quite prevalent. If so, the LysR-NodD-NahR type of regulatory system may be very widespread in the procaryotic community.

While this manuscript was in review, You et al. (45) reported a DNA sequence for *nahR*. However, their derived NahR ORF was 42 kDa, which they admitted was inexplicably larger (>20%) than the expected value of 35 kDa. Their *nahR* sequence differs from ours in three important ways; You et al. (45) found an extra G after our position 894 (AAGgTAC), an additional A after our position 925 (TTGaCGG), and a C, not T, at position 932. These anomalies cause the increased size of their NahR ORF. In several determinations of the sequence in this region on both strands, we observed no evidence of these extra nucleotides. Moreover, in sequencing other G+C-rich genes, we have often found that unusual numbers of arginine residues in a small region of a *Pseudomonas* ORF, like that found at the C terminus of the You et al. ORF, are caused by an incorrect reading frame in G+C-rich regions. Based on this and the facts that our NahR ORF is 34 kDa (within 3% of the expected size reported by us and You et al. [45]) and shows such extensive homology with the LysR family of activators, it is likely that our sequence is the correct *nahR* sequence and ORF.

ACKNOWLEDGMENTS

We thank R. Nagao, R. Geever, and P. Brown for assistance with the computer homology searches, E. Poser for technical assistance, and P. Bates for typing the manuscript.

This research was supported in part by Public Health Service grant GM-3322-04 from the National Institutes of Health.

LITERATURE CITED

1. **Applebaum, E. R., D. V. Thompson, K. Idler, and N. Chartrain.** 1988. *Rhizobium japonicum* USDA 191 has two *nodD* genes that differ in primary structure and function. J. Bacteriol. **170**:12–20.
2. **Barnsley, E. A.** 1975. Induction of the enzymes of naphthalene metabolism in *Pseudomonas* by salicylate and 2-aminobenzoate. J. Gen. Microbiol. **88**:193–196.
3. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:45–148.
4. **Dayhoff, M.** 1978. Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Silver Springs, Md.
5. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**: 1645–1655.
6. **Dodd, I. B., and J. B. Egan.** 1987. Systematic method for the detection of potential λ Cro-like DNA-binding regions in proteins. J. Mol. Biol. **194**:557–564.
7. **Doolittle, R. F.** 1986. Of URFs and ORFs, p. 55. University Science Books, Mill Valley, Calif.
8. **Dunn, N. W., and I. C. Gunsalus.** 1973. Transmissible plasmid encoding early enzymes of naphthalene oxidation in *Pseudomonas putida*. J. Bacteriol. **144**:974–979.
9. **Egelhoff, T. T., R. F. Fisher, T. W. Jacobs, J. T. Mulligan, and S. R. Long.** 1985. Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. DNA **4**:241–248.
10. **Farrell, R., and A. Chakrabarty.** 1979. Degradative plasmids: molecular nature and mode of evolution, p. 97–109, *In* K. N. Timmis (ed.), Plasmids of medical, environmental, and commercial importance. Elsevier Science Publishing, Inc., New York.
11. **Fisher, R. F., H. L. Brierley, J. T. Mulligan, and S. R. Long.** 1987. Transcription of *Rhizobium meliloti* nodulation genes: identification of a *nodD* transcription initiation site *in vitro* and *in vivo*. J. Biol. Chem. **262**:6849–6855.
12. **Fisher, R. F., T. T. Egelhoff, J. T. Mulligan, and S. R. Long.** 1988. Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. Genes Dev. **2**:282–293.
13. **Fisher, R. F., J. Swanson, J. T. Mulligan, and S. R. Long.** 1987. Extended nodulation gene region in *Rhizobium meliloti* 1021. II. Sequence, protein products, and transcription start site for *nodFEGH*. Genetics **117**:191–201.
14. **Frantz, B., and A. Chakrabarty.** 1987. Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation. Proc. Natl. Acad. Sci. USA **84**: 4460–4464.
15. **Harayama, S., M. Rekik, A. Wasserfallen, and A. Bairoch.** 1987. Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. Mol. Gen. Genet. **210**:241–247.
16. **Henikoff, S., G. Haughn, J. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85**:6602–6606.
17. **Hong, G.-F., J. E. Burn, and A. W. B. Johnston.** 1988. Evidence that DNA involved in the expression of nodulation (*nod*) genes in *Rhizobium* binds to the product of the regulatory gene *nodD*. Nucleic Acids Res. **15**:9677–9689.
18. **Horvath, B., C. W. B. Bachem, J. Schell, and A. Kondorosi.** 1987. Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the *nodD* gene product. EMBO J. **6**:841–848.
19. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
20. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
21. **Long, S. R.** 1986. Genetic advances in the study of *Rhizobium* nodulation, p. 135–150. *In* Genetic engineering—principles and methods. J. K. Setlow and A. Hollaender (ed.), Plenum Publishing Corp., New York.

22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Mead, D. A., E. S. Skorupa, and B. Kemper.** 1985. Single stranded DNA SP6 promoter plasmids for engineering mutant RNA's and proteins: synthesis of a 'stretched' parathyroid hormone. Nucleic Acids Res. **13**:1103–1118.
24. **Mulligan, J. T., and S. R. Long.** 1985. Induction of *Rhizobium meliloti nodC* expression by plant exudate required nodD. Proc. Natl. Acad. Sci. USA **82**:6609–6613.
25. **Nakai, C., H. Kagmiyama, M. Nozaki, T. Nakazawa, S. Inouye, Y. Ebina, and A. Nakazawa.** 1983. Complete nucleotide sequence of the metapyrocatechase gene on the TOL plasmid of *Pseudomonas putida* mt-2. J. Biol. Chem. **258**:2923–2928.
26. **Normore, M. W.** 1976. Guanosine-plus-cytosine (GC) composition of the DNA of bacteria, fungi, algae and protozoa, p. 65–240. *In* G. D. Fasman (ed.), Handbook of biochemistry and molecular biology. CRC Press, Inc., Boca Raton, Fla.
27. **Pabo, C. O., and R. J. Sauer.** 1984. Protein DNA recognition. Annu. Rev. Biochem. **53**:293–321.
28. **Peters, N. K., J. W. Frost, and S. R. Long.** 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science **233**:971–1008.
29. **Rossen, L., E. O. Davis, and A. W. B. Johnston.** 1987. Plant induced expression of *Rhizobium* genes involved in host specificity and early stages of nodulation. Trends Biochem. Sci. **12**:430–433.
30. **Rostas, K., E. Kondorosi, B. Horvath, A. Simoncsits, and A. Kondorosi.** 1986. Conservation of extended promoter regions of nodulation genes in *Rhizobium*. Proc. Natl. Acad. Sci. USA **83**:1757–1761.
31. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **Schell, M. A.** 1983. Cloning and expression in *E. coli* of the naphthalene degradation genes from plasmid NAH7. J. Bacteriol. **153**:822–828.
33. **Schell, M. A.** 1985. Transcriptional control of the *nah* and *sal* hydrocarbon-degradation operons by the *nahR* gene product. Gene **36**:301–309.
34. **Schell, M. A.** 1986. Homology between nucleotide sequences of promoter regions of *nah* and *sal* operons of NAH7 plasmid of *Pseudomonas putida*. Proc. Natl. Acad. Sci. USA **83**:369–373.
35. **Schell, M. A., and E. Poser.** 1989. Demonstration, characterization, and mutational analysis of NahR protein binding to *nah* and *sal* promoters. J. Bacteriol. **171**:837–846.
36. **Schell, M. A., and P. Wender.** 1986. Identification of the *nahR* gene product and nucleotide sequences required for its activation of the *sal* operon. J. Bacteriol. **166**:9–14.
37. **Shearman, C. A., L. Rossen, A. W. B. Johnston, and J. A. Downie.** 1986. The *Rhizobium leguminosarum* nodulation gene *nodF* encodes a polypeptide similar to acyl-carrier protein and is regulated by *nodD* plus a factor in pea root exudate. EMBO J. **5**:647–652.
38. **Spaink, H. P., C. A. Wijffelman, E. Pees, R. J. H. Okker, and B. J. J. Lugtenberg.** 1987. Rhizobium nodulation gene *nodD* as a determinant of host specificity. Nature (London) **328**:337–340.
39. **Stragier, P., O. Danas, and J.-C. Patte.** 1983. Regulation of diaminopimelate decarboxylase synthesis in *Escherichia coli*. III. Nucleotide sequence and regulation of the *lysR* gene. J. Mol. Biol. **168**:333–350.
40. **Stragier, R., F. Richard, F. Borne, and J.-C. Patte.** 1983. Regulation of diaminopimelate decarboxylase in *Escherichia coli*. I. J. Mol. Biol. **168**:297–320.
41. **Streber, W. R., K. N. Timmis, and M. H. Zenk.** 1987. Analysis, cloning, and high-level expression of 2,4-dichlorophenoxyacetate monooxygenase gene *tfdA* of *Alcaligenes eutrophus* JMP134. J. Bacteriol. **169**:2950–2955.
42. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
43. **Yen, K.-M., and I. C. Gunsalus.** 1982. Plasmid gene organization: naphthalene/salicylate oxidation. Proc. Natl. Acad. Sci. USA **79**:874–878.
44. **Yen, K.-M., and I. C. Gunsalus.** 1985. Regulation of naphthalene catabolic plasmid NAH7. J. Bacteriol. **162**:1008–1013.
45. **You, I. S., D. Ghosal, and I. C. Gunsalus.** 1988. Nucleotide sequence of plasmid NAH7 gene *nahR* and DNA binding of the *nahR* product. J. Bacteriol. **170**:5409–5415.