JOURNAL OF BACTERIOLOGY, Oct. 1993, p. 6067–6074
0021-9193/93/196067-08$02.00/0

Vol. 175, No. 19

MINIREVIEW

The σ^{54} Bacterial Enhancer-Binding Protein Family: Mechanism of Action and Phylogenetic Relationship of Their Functional Domains

ENRIQUE MORETT[1]* AND LORENZO SEGOVIA[2]†

Departments of Molecular Ecology[1] and Molecular Genetics,[2] Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México, Apdo. Postal 565-A, Cuernavaca, Morelos, Mexico

Because of their ability to stimulate gene expression from distant sites and regardless of their orientation, transcriptional enhancer elements are a puzzling class of regulatory DNA sequences (42). Although enhancers are a common feature of eukaryotic and viral genes, the mechanism by which they communicate with promoters and stimulate gene expression is still not very well understood. However, our understanding of the mechanism of gene activation at a distance has been boosted by the study of naturally occurring enhancer-like elements (ELE) in prokaryotes. The ELE are palindromic or nearly palindromic nucleotide sequences normally found between 100 and 200 bp upstream of a certain class of promoters, and, like the eukaryotic enhancers, they still stimulate transcription when moved thousands of nucleotides upstream or downstream from the transcription start and regardless of their orientation (7, 47). The extensive genetic characterization of many bacterial ELE and the development of defined in vitro transcription systems has greatly facilitated the study of activation by enhancers, and now some of the molecular details of distal activation are well established (34).

It has been demonstrated that the ELE are binding sites for regulatory proteins named bacterial enhancer-binding proteins (EBPs). The best-studied EBPs are NtrC and NifA, which stimulate the expression of genes required for nitrogen assimilation and nitrogen fixation, respectively, in a number of organisms (37). These proteins activate genes transcribed by the RNA polymerase holoenzyme containing the alternative sigma factor σ^{54} ($E\sigma^{54}$). The σ^{54} system is widely represented in the alpha and gamma proteobacteria (previously called purple bacteria), and it has also been reported to be present in *Bacillus subtilis* (13, 14). Among other physiological processes controlled by σ^{54} are nitrogen assimilation and fixation, dicarboxylic acid transport, hydrogen oxidation, alginate utilization, degradation of aromatic compounds, pilus formation and formate utilization (for reviews, see references 35 and 56). A common feature of all these processes is that they are not absolutely required for cell survival, and each is activated by a different EBP interacting with a distinctive ELE nucleotide sequence.

In addition to its ability to be activated at a distance, the $E\sigma^{54}$ system has several other unique characteristics that are not common among bacterial holoenzymes. $E\sigma^{54}$ recognizes promoter DNA with unusually close boxes, centered at around −12 and −24 from the transcription start, instead of the more common −10 and −35 boxes (39). Upon binding, $E\sigma^{54}$ forms a stable, closed promoter complex, but isomerization to the transcriptionally active open promoter complex is absolutely dependent on an activator protein bound at the ELE. This step requires energy, which is obtained by ATP hydrolysis catalyzed by the EBP itself (34). In this sense, $E\sigma^{54}$ can be regarded as a "defective" holozyme that only initiates transcription in concert with a second transcriptional factor.

The unique properties of σ^{54} are also reflected in its structure. This protein is the only sigma factor that does not have sequence similarity to any other known prokaryotic sigma factor, indicating that σ^{54} has a different origin than the rest of the known sigma factors (for a review, see reference 36). Interestingly, several functional motifs normally found in eukaryotic transcriptional factors, such as a leucine zipper, an acidic region, and a glutamine-rich region, have been identified in σ^{54} (51). Another remarkable particularity of σ^{54} is its ability to interact with the promoter DNA in the absence of core RNA polymerase (5). The distinctive form in which $E\sigma^{54}$ interacts with the promoter DNA, the strict dependency on regulatory proteins bound far upstream, and the requirement for nucleotide hydrolysis for activation of transcription is reminiscent of the activation process by eukaryotic RNA polymerase II (57).

Early experiments devised to address how EBPs bound far from the promoter can activate transcription showed that the relative position of the ELE and the promoter is critical for activation (3, 6, 38, 48). It was observed that activation of transcription only occurs efficiently when the ELE is located over a certain face of the DNA helix. To explain these results, Buck and collaborators (6) proposed that the EBP may simultaneously contact the ELE and the $E\sigma^{54}$ holoenzyme bound at the promoter, causing the looping of the intervening DNA. This hypothesis was elegantly demonstrated by direct observation of such looped structures by electron microscopy with the *glnA*p2 promoter (55). For several genes, the loop formation that brings together the EBP and $E\sigma^{54}$ is facilitated by the integration host factor, a heterodimeric protein that binds between the ELE and the promoter and causes a sharp bending of the DNA (10, 29).

Notwithstanding the central role of the interaction of the EBP with the ELE in gene expression, this is only one step in a complex mechanism of transcription activation. It has been observed that certain strong promoters can be partially activated even if they lack an ELE. These observations revealed that DNA binding and activation of transcription

* Corresponding author.
† Present address: National Eye Institute, National Institutes of Health, Bethesda, MD 20892.

are two different functions of the same regulatory protein, as has been shown for several eukaryotic transcriptional factors (26). What, then, is the role of the ELE? It has been demonstrated that binding to the ELE increases the local concentration of the EBP in the vicinity of the promoter (58) and tethers it in the right position for activation. Moreover, since each EBP recognizes a specific ELE sequence, binding to these elements confers specificity to the initiation of transcription (39).

Sequence analysis, deletion experiments, and chimeric constructions indicate that the EBPs, like the eukaryotic transcriptional activators (for a review, see reference 26), are modular in structure (19, 20, 30, 31, 40). Three functionally different domains have been described: a COOH-terminal domain containing a helix-turn-helix DNA-binding motif involved in binding to the ELE, a conserved central domain responsible for the activation of transcription, and a nonconserved NH_2-terminal domain presumed to be the target for certain regulatory signals (8, 19, 49). Experimental data from several laboratories has shown that these domains are structurally independent. Deletion analysis of the NifA and DctD proteins showed that the isolated central domain retains the ability to activate transcription even in the absence of the DNA-binding domain (30, 31, 40), while the isolated COOH-terminal domain interacts specifically with the ELE, although it is unable to stimulate transcription (9, 20, 40).

The observation that the COOH-terminal and the central domains retain their respective functions when expressed separately and that the NH_2-terminal domain is highly divergent indicates that these are true domains that can be correctly folded separately from each other. These observations suggest that the different EBP domains could have evolved independently. In this article, we display an updated amino acid sequence alignment of the known EBPs and present the evolutionary relationship of these proteins and their different functional domains.

AMINO ACID SEQUENCE ALIGNMENT

To date, more than 30 sequences of 13 different EBPs from 20 species have been reported (Fig. 1). We generated a multiple sequence alignment of the EBPs shown in Fig. 1 by using the GCG sequence analysis program PILEUP. This analysis shows that the central domain is common to all members of the EBP family and is highly conserved with a pairwise similarity of more than 40% and that the COOH-terminal domain also has some conserved regions at its carboxy end, whereas the NH_2-terminal domains do not share obvious similarities between the different protein classes and greatly differ in length. This alignment sustains the previously proposed multidomain model of Drummond et al. (19).

Central domain. The central domain is about 240 amino acids long (Fig. 1). It is highly conserved along its entire length, and very few gaps were needed for the alignment. The alignment shown in Fig. 1 revealed that the conserved residues are arranged in seven regions (C1 to C7).

Region C1 is glycine rich and displays remarkable similarity to the glycine-rich flexible loop motif found at the nucleotide-binding site of adenylate kinase and several other ATP-binding proteins (consensus G--G-GK-) (see reference 27 and references therein). It has been well established by both genetic and biochemical criteria that NtrC catalyzes ATP hydrolysis as an essential step for open promoter complex formation and that mutations that disrupt the integrity of this region affect ATP hydrolysis (1, 46, 53, 59). Structural studies of the *Escherichia coli* adenylate kinase have shown that the alpha phosphoryl group of ATP is in close contact with the lysine of this motif (27). Adjacent to it there are two other lysines separated by three residues, two of them hydrophobic. The hydrophobic residues form part of a pocket in which the adenine-ribose moiety of ATP is located (27). Region C1 of the EBPs also has hydrophobic residues at similar positions adjacent to a positively charged amino acid (histidine instead of lysine). There is a second motif in adenylate kinase involved in ATP hydrolysis; this region, called segment 3, is found 90 amino acids away from region C1 and is structured as a hydrophobic strand of parallel beta-pleated sheet terminated by an aspartate (QPTLLLYVD) (27). The carboxylate residue has been proposed to be at the reaction center. It is interesting that there is a region, C5 (QakLLRVLqe) in the EBPs, found about 80 amino acids away from region C1, which has similarity to segment 3 of adenylate kinase. Thus, although no mutants have been isolated in this region, it is possible that the EBPs bind ATP by means of their C1 and C5 regions in a manner very similar to that of adenylate kinase. Interestingly, the HydG protein of *E. coli* lacks the conserved glycines at C1, hence it is unlikely that this protein binds ATP.

Region C2 is hydrophobic and, except for the NtrC proteins that have methionine residues, has the only conserved cysteine in the EBP family. The secondary structure of region C3 is predicted to fold into two alpha helices separated by a turn, according to the Garnier-Osguthorpe-Robson method. The mutagenesis of *Salmonella typhimurium ntrC* has shown that the change of the conserved glycine 219 to lysine at the end of the second alpha helix specifically impairs the activation function of NtrC with no effect on ATP hydrolysis (59). The NtrC protein of *Rhodobacter capsulatus* lacks this highly conserved region. This protein has been show to activate gene expression in an *R. capsulatus* mutant strain devoid of σ^{54} and, therefore, it might function with a different sigma factor (25, 32, 33). *E. coli* TyrR is a regulatory protein that controls the expression of a number of genes involved in tyrosine biosynthesis (45).

FIG. 1. Amino acid sequence alignment of the σ^{54} bacterial enhancer-binding protein family. The sequences of 31 EBPs were aligned by using the Genetics Computer Group program PILEUP (15). The alignments were manually edited with the program LINEUP from the same package. The code for each sequence is the name of the protein followed by a three-letter code for the species (described below). The very long NH_2 end of FhlAEco and the COOH end of LevRBsu are not shown. Due to their high degrees of similarity, only one sequence of each of the enterobacterial NifA, NtrC and HydG proteins is shown. The consensus amino acids are presented with a black background. The conserved cysteines of the oxygen-regulated NifA proteins are also presented with a black background. The boxes indicate the limits of each domain. Each of the seven central regions is also indicated. The codes for the species are as follows: *Abr*, *Azospirillum brasilense*; *Aca*, *Azorhizobium caulinodans*; *Aeu*, *Alcaligenes eutrophus*; *Atu*, *Agrobacterium tumefaciens*; *Avi*, *Azotobacter vinelandii*; *Bja*, *B. japonicum*; *Bsp*, *Bradhyrhizobium* spp.; *Bsu*, *Bacillus subtilis*; *Ccr*, *Caulobacter crescentus*; *Eco*, *Escherichia coli*; *Hse*, *Herbaspirillum seropeidae*; *Kpn*, *Klebsiella pneumoniae*; *Rca*, *Rhodobacter capsulatus*; *Rle*, *Rhizobium leguminosarum*; *Rlp*, *R. leguminosarum PRE*; *Rme*, *R. meliloti*; *Rtr*, *R. trifoli*; *Ppu*, *Pseudomonas putida*; *Psy*, *P. syringae*.


NH2
CENTRAL
C1
C2
C3
C4
C5
C6
C7
COOH
H
T
H

Although the currently available data indicates that TyrR activates the expression of genes transcribed by the RNA polymerase holoenzyme with the housekeeping sigma factor σ^{70}, it has a region homologous to the EBP central domain (more than 45% similarity [Fig. 1]) (12). TyrR also lacks the second alpha helix of region C3, as found for *R. capsulatus* NtrC. It thus seems that these proteins were σ^{54} activators that suffered an internal deletion that impaired their ability to activate σ^{54} promoters. Therefore, it seems reasonable to believe that this region could be involved in a specific interaction between the EBP and the $E\sigma^{54}$ required for open promoter complex formation.

Region C4 is rich in glycine and is negatively charged; by contrast, C6 is positively charged and is rich in aromatic residues and proline. Finally, region C7 has a core of eight highly conserved amino acids. No function has yet been ascribed to any of these regions.

We looked for sequences with similarity to the seven regions of the central domain in the SwissProt, GenBank, and EMBL data banks. The only protein found with significant similarity was the PgtA regulatory protein from *S. typhimurium*. This protein showed scattered sequence similarity along the entire central domain, and the similarity to region C7 is very high. We also detected two partially sequenced ORFs in the *pspABCD* and *purHD* operons of *E. coli* that had similarity to the central domain. These findings indicate that the complexity of the σ^{54} system extends to at least six regulons in the enterobacteria.

COOH-terminal domain. The COOH-terminal domain is the shortest domain of the EBP family. This domain is between 65 and 130 amino acids long. The only exception is the *B. subtilis* LevR protein, which extends for 600 amino acids after the central domain. The alignment shown in Fig. 1 revealed a conserved motif at the carboxy end of this domain. It has been proposed that this motif (AL-X9-AA-X2-LG) forms a helix-turn-helix DNA-binding motif (19). Its consensus sequence generally agrees with the criteria established by Brennan and Mathews (4) and Dodd and Egan (17) for the prediction of helix-turn-helix motifs. Model building and mutagenesis analyses of the *Klebsiella pneumoniae nifA* and *ntrC* genes support this proposition; only the mutations that lie on the surface of the second helix, the putative recognition helix, affected specific DNA binding (11, 40).

The amino acids in this motif that were conserved between the different classes of EBPs correspond to the first helix and to the turn, with no significant homology at the recognition helix. This is consistent with the observation that each EBP recognizes a different ELE nucleotide sequence.

The *Bradyrhizobium japonicum* and *Rhizobium meliloti* NifA proteins are directly sensitive to the oxygen tension of the cell (16, 24). Activation of transcription by these proteins only occurs in microaerobic conditions, and a shift from microaerobic to aerobic growth rapidly impairs the ability of *B. japonicum* NifA to bind at the ELE and to activate *nifD* gene expression (41). Fisher et al. (23) have proposed that NifA senses oxygen by the redox state of a metal cofactor bound between its central and COOH-terminal domains. Around this region are four cysteine residues that are essential for NifA activity (23). The alignment of Fig. 1 shows that the cysteine residues are conserved in the NifA proteins of *Rhizobium*, *Bradyrhizobium*, *Azospirillum*, *Azorhizobium*, *Herbaspirillum*, and *Rhodobacter* species. Thus, as observed for *B. japonicum* and *R. meliloti*, it seems likely that the activity of all these NifA proteins could be directly controlled by the oxygen tension of the cell. Conversely, the *K. pneumoniae* and *Azotobacter vinelandii* NifA proteins and the AnfA and VnfA proteins of the latter organism, which are insensitive to the oxygen tension, lack this region.

The putative binding site for the metal cofactor lies between the central and the COOH domains. Thus, any change in the redox state of the cofactor could influence the activity of both the central and the COOH-terminal domain (41).

NH_2-terminal domain. The NH_2-terminal domain is the least-conserved domain in the entire EBP family. There are no regions of similarity encompassing all the members of this family. Two proteins, the *Rhizobium leguminosarum* bv. trifolii NifA and the *Pseudomonas syringae* HrpS proteins, have very short NH_2 domains, consisting of 12 and 74 amino acids, respectively, while the *E. coli* FhlA NH_2 sequence is almost 400 amino acids long. However, a subgroup formed by the NtrC, DctD, HydG, AlgB, HoxA, and FlbD proteins displayed a high degree of similarity at this domain (more than 38% amino acid similarity). These proteins belong to the signal-transducing family of "two-component" sensor-activator regulatory proteins (for a review, see reference 54). It has been shown that an aspartic acid residue on the NtrC NH_2-terminal domain is phosphorylated in response to the nitrogen status of the cell by the histidine kinase protein NtrB (60). The phosphorylated form of NtrC is the transcriptionally active form of this protein. This residue, as well as a second aspartic acid and a lysine, is invariant in all these proteins except FlbD (54).

The activity of NifA is regulated by the oxygen and fixed-nitrogen status of the cell through the action of the NifL protein in *K. pneumoniae*. Immunological studies have shown that NifA and NifL form a complex (28). Also, NifA proteins lacking the NH_2-terminal domain are inactivated by NifL regardless of the oxygen and nitrogen conditions of the cell (20). This has led to the proposition that the NH_2 domain prevents inactivation by NifL at low oxygen and fixed-nitrogen levels (20). It is interesting that although NifL proteins have only been reported for *K. pneumoniae* and for *A. vinelandii*, all of the NifA proteins have some similarity in this region (Fig. 1). Although the NifA protein of *R. leguminosarum* bv. trifolii lacks this domain and is proficient in activation, it is still not known whether this domain has any function in the other NifA proteins.

ESTIMATION OF PHYLOGENETIC RELATIONSHIPS

Several proteins from both prokaryotic and eukaryotic organisms consist of structurally independent domains that evolved separately (2). As discussed above, the EBPs display a modular structure; however, it is not well understood how independent the different functional domains are and how they have evolved. In order to estimate the phylogenetic relationships of the EBPs and of each of their functional domains, we first created a multiple sequence alignment by using the GCG program PILEUP (15). This program uses a simplification of the progressive alignment method of Feng and Doolittle (22). This procedure begins with the determination of all possible pairwise similarity scores. The two most similar sequences are aligned by using the Needleman and Wunsch algorithm (43), forming the first cluster, and then the next most related sequences are progressively aligned to this cluster. We calculated the genetic distances by using the GCG program DISTANCES (15) with the Dayhoff PAM-250 matrix (52). This program does not take into account multiple replacements at a given site. The clustering was made with the Neighbor-Joining method of

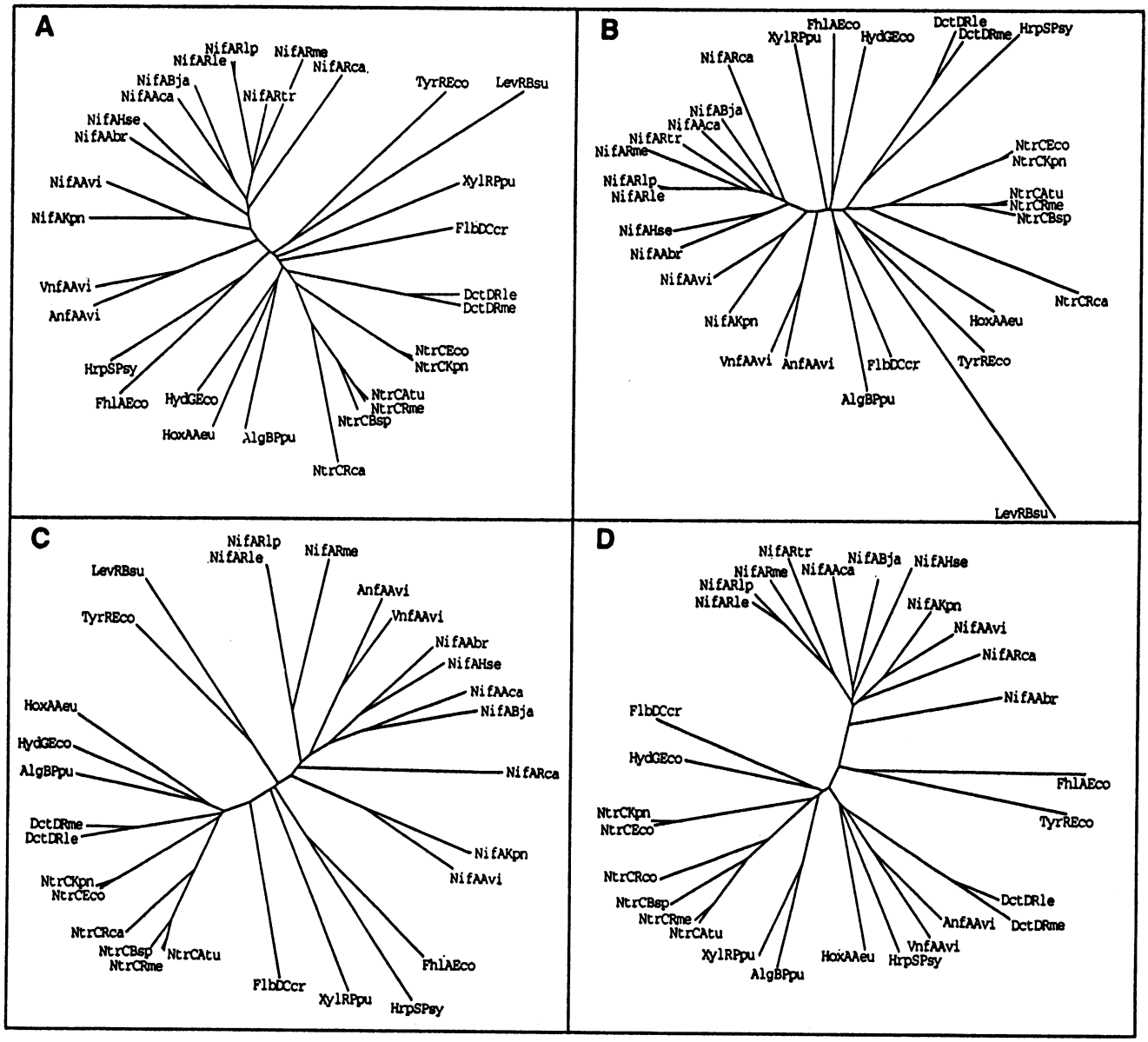

FIG. 2. Unrooted phylogenetic trees of the entire protein sequence (A), the central domain (B), the NH_2-terminal domain (C), and the COOH-terminal domain (D) of the σ^{54} bacterial enhancer-binding protein family. We performed a phylogenetic analysis of the 31 EBP sequences. Initially, a multiple sequence alignment of all the members of this protein family was generated by using the GCG sequence analysis package program PILEUP (15). The alignment obtained was further refined with the manual multiple alignment program LINEUP. For each domain the alignments were generated independently of the alignment of the entire sequences. The genetic distances were calculated by using the Dayhoff PAM-250 matrix (52) with the GCG sequence analysis package program DISTANCES. Subsequently, we determined the phylogenetic relationships among these sequences by using the Neighbor-Joining method of Saitou and Nei (50) with the NEIGHBOR program of J. Felsenstein's PHYLIP phylogeny inference package (21).

Saitou and Nei (50), using the program NEIGHBOR of J. Felsenstein's PHYLIP phylogeny inference package (21). This algorithm sequentially identifies neighbor pairs that minimize the total length of the tree without making any statistical analysis to prove the topologies of the resulting trees. We took the alignment of Fig. 1 to establish the limits of each domain. Then, to optimize the individual alignments, each domain was realigned with the GCG PILEUP program and further improved manually.

The phylogeny of the sequences of the entire proteins shown in Fig. 2A demonstrates that each class of EBP forms a cluster independently of the others. NtrC and NifA, the EBPs for which the most sequences have been determined, are arranged in two main groups. In the NifA cluster, there is one group comprising the proteins from the organisms belonging to the alpha subdivision of the proteobacteria and a second group encompassing the proteins from the gamma subgroup organisms *Klebsiella* and *Azotobacter* spp. Likewise, the NtrC proteins are arranged in two main groups, with the enterobacterial proteins on one side and the rhizobial and *R. capsulatus* NtrC proteins on the other, also clearly reflecting the separation of the alpha and gamma

subdivisions of the proteobacteria. Thus, the NtrC and NifA sequences are clearly paralogous; they are related by an ancestral duplication, and their subsequent evolutionary histories are in agreement with one another and with the ribosomal 16S RNA genes of the same organisms (61, 62).

In the current model for activation of transcription by the EBPs, their central domain interacts with the $E\sigma^{54}$-promoter complex, presumably directly with σ^{54}, to catalyze open promoter complex formation. It has been shown that several EBPs from many organisms can activate transcription in *E. coli* with the host σ^{54} factor. This implies the functional conservation of the amino acid residues involved in protein interaction. One possibility was that within a single species, since all the activators interact with the same $E\sigma^{54}$ holoenzyme, the central domains or a part of them would be more similar among them than they would be to the central domains of the same class of protein from other species. Our alignment of the central domain did not reveal any species-specific trait and, consistent with this, the phylogeny of the central domain generally coincided with that of the whole protein, showing the clustering of NifA and NtrC proteins accordingly to the phylogeny of their host organism (Fig. 2B). The only exception is the *R. capsulatus* NtrC protein, which fell out of the NtrC group. As discussed above, this protein has some internal deletions within this domain and activates transcription in the absence of σ^{54}. In addition, it has several changes in positions that are normally well conserved. We believe that these changes are the result of genetic drift of this perhaps nonfunctional *R. capsulatus* NtrC domain.

The topology of the central domain tree showed a very long branch with the LevR protein of *B. subtilis*, the only firmibacterial example. This is due to the high degree of conservation of the central domain, which makes the distance between the protein of this gram-positive organism and the rest of the EBPs more apparent.

The topology of the phylogenetic tree of the NH_2-terminal domain showed two clearly recognizable groups. One encompassed the EBPs that belong to the two-component family of regulatory proteins, which are clearly related, and a second group was formed by the NifA proteins (Fig. 2C). The very low degree similarity between them and the rest of the EBPs is suggestive of different origins. In this domain, the alternative NifA homologs of *A. vinelandii* are more similar to the *Azospirillum brasilense* and *Herbaspirillum seropediae* NifA proteins than they are to NifA of their own species, also indicating that their duplication predates the speciation events that gave rise to the genus *Azotobacter*.

We also determined the phylogenetical relationship of the COOH-terminal domain. Despite the short length of the amino acid sequences and the low degree of similarity, which makes the phylogenetic relationship of the COOH-terminal domain less reliable, some relationships could be established unambiguously. The NifA and the NtrC proteins clustered independently and were arranged generally according to the 16S phylogeny (Fig. 2D). The *A. vinelandii* AnfA and VnfA proteins fell out of the NifA cluster. This is consistent with the observation that these proteins recognize different ELE sequences than those recognized by NifA. A remarkable observation is the clustering of XylR and AlgB from *Pseudomonas putida*. Interestingly, these proteins are more homologous at this domain than at the conserved central domain. This observation indicates that the COOH domain of these proteins has the same and perhaps a recent origin within the genus *Pseudomonas*.

There is cumulative evidence that the EBPs are multimeric proteins. To date, we do not know where the dimerization function resides. However, Kustu and collaborators have proposed that the NtrC proteins dimerize through their COOH-terminal domain (44). This proposition is based on their findings that the *E. coli* Fis protein, a small protein for which the crystal structure has been solved and the dimerization determinants established, is homologous to the COOH domain and that the dimerization motif is only present in NtrC proteins. For the rest of the EBPs, the dimerization determinants must be located in a different domain (44). The Fis protein has more similarity to the alpha proteobacterial NtrC than do the NtrC proteins themselves (data not shown). This suggests the lateral transfer of the Fis gene from an alpha proteobacterium to a gamma proteobacterium. Nevertheless, more Fis sequences from other lineages must be analyzed to validate this hypothesis.

CONCLUSIONS

The modularity in the organization of proteins has recently been recognized as a key element in the understanding of their differences in function. The many combinations of a finite number of independently evolved functional domains is a factor that greatly contributes to the versatility of protein function (2, 18). The EBPs are a family of transcriptional factors that activate σ^{54} promoters from distant sites. Deletion analysis and amino acid sequence comparison have demonstrated that the members of this family of proteins are modular in structure, consisting of three different functional domains. Our phylogenetic analysis, aimed at understanding the evolution of these different functional domains, showed that the central and COOH-terminal domains have a common origin for the whole family, while the NH_2-terminal domain seems to be common only for a subgroup of proteins that belong to the two-component family of activators. Although some proteins showed mixed clustering of their domains, the overall conservation of the topologies observed with the entire proteins and with each domain indicates that there has not been extensive shuffling of the functional domains but rather that they have coevolved. The most intriguing proteins are the *A. vinelandii* AnfA and VnfA. Although they have been only found in *Azotobacter* spp., their origin seems to predate the speciation events that gave rise to this genus and might even predate the split between alpha and gamma proteobacteria. The absence of these alternative systems in the other lineages indicates that they must have been lost.

The fact that the EBPs have been found in both firmibacteria and proteobacteria indicates that σ^{54} is part of a very ancient regulatory system. Clearly, the origin of this sigma factor and of the EBP family as a whole will be understood by acquiring more EBP sequences from several other bacterial groups.

ACKNOWLEDGMENTS

We are very grateful to Jaime Mora for his advice and generous support. We thank Sydney Kustu, Gary Olsen, Peter Frederikse, and Julio Collado-Vides for their critical reading of the manuscript. We are also grateful to David Link for his continuous support with every matter concerning computers.

REFERENCES

1. **Austin, S., and R. Dixon.** 1992. The prokaryotic enhancer binding protein NTRC has an ATPase activity which is phosphorylation and DNA dependent. EMBO J. **11**:2219–2228.
2. **Baron, M., D. G. Norman, and L. D. Campbell.** 1991. Protein

modules. Trends Biochem. Sci. **16:**13–17.
3. **Birkmann, A., and A. Bock.** 1989. Characterization of a *cis* regulatory DNA element necessary for formate induction of the formate dehydrogenase gene (*fdhA*) of *Escherichia coli*. Mol. Microbiol. **3:**187–195.
4. **Brennan, R. G., and B. W. Matthews.** 1989. The helix-turn-helix DNA binding motif. J. Biol. Chem. **264:**1903–1906.
5. **Buck, M., and W. Cannon.** 1992. Specific binding of the transcription factor sigma-54 to promoter DNA. Nature (London) **358:**422–424.
6. **Buck, M., W. Cannon, and J. Woodcock.** 1987. Transcriptional activation of the *Klebsiella pneumoniae* nitrogenase promoter may involve DNA loop formation. Mol. Microbiol. **1:**243–249.
7. **Buck, M., S. Miller, M. Drummond, and R. Dixon.** 1986. Upstream activator sequences are present in the promoters of nitrogen fixation genes. Nature (London) **320:**374–378.
8. **Buikema, W. J., W. W. Szeto, P. V. Lemley, W. H. Orme-Johnson, and F. M. Ausubel.** 1985. Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. Nucleic Acids Res. **13:**4539–4555.
9. **Cannon, W., W. Charlton, and M. Buck.** 1991. Organization and function of binding sites for the transcriptional activator NifA in the *Klebsiella pneumoniae nifE* and *nifU* promoters. J. Mol. Biol. **220:**915–931.
10. **Cannon, W. V., R. Kreutzer, H. M. Kent, E. Morett, and M. Buck.** 1990. Activation of the *Klebsiella pneumoniae nifU* promoter: multiple and overlapping upstream NifA binding sites and involvement of integration host factor. Nucleic Acids Res. **18:**1693–1701.
11. **Contreras, A., and M. Drummond.** 1988. The effect on the function of the transcriptional activator NtrC from *Klebsiella pneumoniae* of mutations in the DNA-recognition helix. Nucleic Acids Res. **16:**4025–4039.
12. **Cornish, E. C., V. P. Argyropoulos, J. Pittard, and B. Davidson.** 1986. Structure of the *Escherichia coli* K-12 regulatory gene *tyrR*. J. Biol. Chem. **261:**403–410.
13. **Debarbuille, M., I. Martin-Verstraete, A. Klier, and G. Rapoport.** 1991. The transcriptional regulator LevR of *Bacillus subtilis* has domains homologous to both sigma 54 and phosphotransferase system-dependent regulators. Proc. Natl. Acad. Sci. USA **88:**2212–2216.
14. **Debarbuille, M., I. Martin-Verstraete, F. Kunst, and G. Rapaport.** 1991. The *Bacillus subtilis sigL* gene encodes an equivalent of σ^{54} from Gram-negative bacteria. Proc. Natl. Acad. Sci. USA **88:**9092–9096.
15. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
16. **Ditta, G., E. Virts, A. Palomares, and K. Choong-Hyun.** 1987. The *nifA* gene of *Rhizobium meliloti* is oxygen regulated. J. Bacteriol. **169:**3217–3223.
17. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019–5026.
18. **Doolittle, R. F.** 1992. Reconstructing history with amino acid sequences. Protein Sci. **1:**191–200.
19. **Drummond, M., P. Whitty, and J. Wootton.** 1986. Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae*: homologies to other regulatory proteins. EMBO J. **5:**441–447.
20. **Drummond, M. H., A. Contreras, and L. A. Mitchenal.** 1990. The function of isolated domains and chimaeric proteins constructed from the transcriptional activators NifA and NtrC of *Klebsiella pneumoniae*. Mol. Microbiol. **4:**29–37.
21. **Felsenstein, J.** 1993. Phylip-Phylogeny inference package. Version 3.5c. (Distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle.)
22. **Feng, D. F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25:**351–360.
23. **Fischer, H. M., T. Bruderer, and H. Hennecke.** 1988. Essential and non-essential domains in the *Bradyrhizobium japonicum* NifA protein: identification of indispensable cysteine residues potentially involved in redox activity and/or metal binding. Nucleic Acids Res. **16:**2207–2224.
24. **Fischer, H. M., and H. Hennecke.** 1987. Direct response of *Bradyrhizobium japonicum nifA*-mediated *nif* gene regulation to cellular oxygen status. Mol. Gen. Genet. **209:**621–626.
25. **Foster-Hartnett, D., and R. G. Kranz.** 1993. Analysis of the promoter and upstream sequence of *nifA1* and *nifA2* in *Rhodobacter capsulatus*: activation requires *ntrC* but not *rpoN*. Mol. Microbiol. **6:**1049–1060.
26. **Frankel, A. F., and P. S. Kim.** 1991. Modular structure of transcriptional factors: implications for gene regulation. Cell **65:**717–719.
27. **Fry, D. C., S. A. Kuby, and A. S. Mildvan.** 1986. ATP-binding site of adenylate kinase: mechanistic implications of its homology with ras-encoded p21, F1-ATPase, and other nucleotide-binding proteins. Proc. Natl. Acad. Sci. USA **83:**907–911.
28. **Henderson, N., S. A. Austin, and R. A. Dixon.** 1989. Role of metal ions in negative regulation of nitrogen fixation by the *nifL* gene product from *Klebsiella pneumoniae*. Mol. Gen. Genet. **216:**484–491.
29. **Hoover, T. R., E. Santero, S. Porter, and S. Kustu.** 1990. The integration host factor stimulates interaction of RNA polymerase with NifA, the transcriptional activator of nitrogen fixation operons. Cell **63:**11–22.
30. **Huala, E., and F. M. Ausubel.** 1989. The central domain of *Rhizobium meliloti* NifA is sufficient to activate transcription from the *R. meliloti nifH* promoter. J. Bacteriol. **171:**3354–3365.
31. **Huala, E., J. Stigter, and F. M. Ausubel.** 1992. The central domain of *Rhizobium leguminosarum* DctD functions independently to activate transcription. J. Bacteriol. **174:**1428–1431.
32. **Jones, R., and R. Haselkorn.** 1989. The DNA sequence of the *Rhodobacter capsulatus ntrA*, *ntrB* and *ntrC* gene analogues required for nitrogen fixation. Mol. Gen. Genet. **215:**507–516.
33. **Kranz, R. G., V. M. Pace, and I. M. Caldicott.** 1990. Inactivation, sequence, and *lacZ* fusion analysis of a regulatory locus required for repression of nitrogen fixation genes in *Rhodobacter capsulatus*. J. Bacteriol. **172:**53–62.
34. **Kustu, S., A. K. North, and D. S. Weiss.** 1991. Prokaryotic transcriptional enhancers and enhancer-binding proteins. Trends Biochem. Sci. **16:**397–402.
35. **Kustu, S., E. Santero, D. Popham, and J. Keener.** 1989. Expression of σ^{54} (*ntrA*)-dependent genes is probably united by a common mechanism. Microbiol. Rev. **53:**367–376.
36. **Lonetto, M., M. Gribskov, and C. A. Gross.** 1992. The σ^{70} family: sequence conservation and evolutionary relationships. J. Bacteriol. **174:**3843–3849.
37. **Merrick, M.** 1992. Regulation of nitrogen fixation genes in free-living and symbiotic bacteria, p. 835–847. *In* G. Stacy, R. Burris, and H. Evans (ed.), Biological nitrogen fixation. Chapman Hall, New York.
38. **Minchin, S. D., S. Austin, and R. A. Dixon.** 1989. Transcriptional activation of the *Klebsiella pneumoniae nifLA* promoter by NTRC is face-of-the-helix dependent and the activator stabilizes the interaction of sigma 54-RNA polymerase with the promoter. EMBO J. **8:**3491–3499.
39. **Morett, E., and M. Buck.** 1989. In vivo studies on the interaction of RNA polymerase-σ^{54} with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters: the role of NifA in the formation of an open promoter complex. J. Mol. Biol. **210:**65–77.
40. **Morett, E., W. Cannon, and M. Buck.** 1988. The DNA-binding domain of the transcriptional activator protein NifA resides in its carboxy terminus, recognises the upstream activator sequences of *nif* promoters and can be separated from the positive control function of NifA. Nucleic Acids Res. **16:**11469–11488.
41. **Morett, E., H. M. Fischer, and H. Hennecke.** 1991. Influence of oxygen on DNA binding, positive control, and stability of the *Bradyrhizobium japonicum* NifA regulatory protein. J. Bacteriol. **173:**3478–3487.
42. **Muller, M. M., T. Gerster, and W. Schaffner.** 1988. Enhancer sequences and the regulation of gene transcription. Eur. J. Biochem. **176:**485–495.

43. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol. **48:**443–453.
44. **North, A. K., K. E. Klose, K. M. Stedman, and S. Kustu.** 1993. Prokaryotic enhancer-binding proteins reflect eukaryotic-like modularity: the puzzle of nitrogen regulatory protein C. J. Bacteriol. **175:**4267–4273.
45. **Pittard, J., and B. E. Davidson.** 1991. TyrR protein of *Escherichia coli* and its role as repressor and activator. Mol. Microbiol. **5:**1585–1592.
46. **Popham, D. L., D. Szeto, J. Keener, and S. Kustu.** 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. Science **243:**629–635.
47. **Reitzer, L. J., and B. Magasanik.** 1986. Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter. Cell **45:**785–792.
48. **Reitzer, L. J., B. Movsas, and B. Magasanik.** 1989. Activation of *glnA* transcription by nitrogen regulator I (NRI)-phosphate in *Escherichia coli*: evidence for a long-range physical interaction between NRI-phosphate and RNA polymerase. J. Bacteriol. **171:**5512–5522.
49. **Ronson, C. W., P. M. Astwood, B. T. Nixon, and F. M. Ausubel.** 1987. Deduced products of C4-dicarboxylate transport regulatory genes of *Rhizobium leguminosarium* are homologous to nitrogen regulatory gene products. Nucleic Acids Res. **15:**7921–7934.
50. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
51. **Sasse-Dwight, S., and J. D. Gralla.** 1990. Role of eukaryotic-type domains found in the prokaryotic enhancer receptor factor σ^{54}. Cell **62:**945–950.
52. **Schwartz, R. M., and M. O. Dayhoff.** 1979. Dayhoff table, p. 353–358. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, D.C.
53. **Shiau, S.-P., P. Chen, and L. J. Reitzer.** 1993. Effect of insertions and deletions in *glnG* (*ntrC*) of *Escherichia coli* on nitrogen regulator I-dependent DNA-binding and transcriptional activation. J. Bacteriol. **175:**190–199.
54. **Stock, J. B., A. M. Stock, and J. M. Mottonen.** 1990. Signal transduction in bacteria. Nature (London) **344:**395–400.
55. **Su, W., S. Porter, S. Kustu, and H. Echols.** 1990. DNA looping and enhancer activity: association between DNA-bound NTRC activator and RNA polymerase at the bacterial *glnA* promoter. Proc. Natl. Acad. Sci. USA **87:**5504–5508.
56. **Thöny, B., and H. Hennecke.** 1989. The −24/−12 promoter comes of age. FEMS Microbiol. Rev. **63:**341–358.
57. **Wang, W., M. Carey, and J. D. Gralla.** 1992. Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. Science **255:**450–453.
58. **Wedel, A., D. S. Weiss, D. Popham, P. Droge, and S. Kustu.** 1991. A bacterial enhancer functions to tether a transcriptional activator near a promoter. Science **248:**486–490.
59. **Weiss, D. S., J. Batut, K. E. Klose, J. Keener, and S. Kustu.** 1991. The phosphorylated form of the enhancer-binding protein NTRC has an ATPase activity that is essential for activation of transcription. Cell **67:**155–167.
60. **Weiss, V., and B. Magasanik.** 1988. Phosphorylation of nitrogen regulator I (NRI) of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85:**8919–8923.
61. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
62. **Young, J. P. W.** 1992. Phylogenetic classification of nitrogen-fixing organisms, p. 43–86. *In* G. Stacy, R. Burris, and H. Evans (ed.), Biological nitrogen fixation. Chapman Hall, New York.