

MINIREVIEW

Domain Architectures of σ^{54} -Dependent Transcriptional Activators

David J. Studholme^{1*} and Ray Dixon²

Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA,¹ and Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH,² United Kingdom

Transcription is the key control point for regulation of numerous cellular activities. Bacteria regulate levels of gene expression by using transcription factors that modulate the recruitment of RNA polymerase (RNAP) to promoter elements in the DNA. Many bacteria also control gene expression by using a second class of transcription factor that uses energy from nucleotide hydrolysis to actively promote transcription initiation.

The sigma (σ) subunit is required for promoter recognition and initiation of transcription by the bacterial RNAP. Typically a bacterial cell may contain several alternative σ subunits with differing sequence specificities that direct the RNAP holoenzyme to different sets of promoters. The σ^{54} subunit (also known as RpoN, NtrA, and σ^N) is unique (42) in that it shares no detectable homology with any of the other known sigma factors (e.g., σ^{70} and σ^{28}). σ^{54} -RNAP binds to specific promoter sites, with the consensus DNA sequence YTGGCACG rNNTTGCW (6), to form a transcriptionally inactive closed complex consisting of holoenzyme bound to double-stranded DNA. In contrast to σ^{70} -RNAP bound at its cognate promoter sites, σ^{54} -RNAP is unable to spontaneously isomerize from a closed complex to a transcriptionally competent open complex (11, 72). To proceed with initiation of transcription, the closed complex must participate in an interaction with a transcriptional activator, involving nucleotide hydrolysis. This transcriptional activator is usually bound at least 100 bp upstream of the promoter site, and DNA looping is required for the activator to contact the closed complex and catalyze formation of the open promoter complex. In this respect the activator resembles the enhancer-binding proteins (EBPs) found in eukaryotic systems, and for this reason such activators are known as bacterial EBPs.

From a protein structural point of view, EBPs share in common a σ^{54} interaction module (Pfam accession number PF00158) but typically have at least one additional domain (Fig. 1). In nearly all of those investigated so far, there is a DNA-binding domain containing a helix-turn-helix sequence motif, enabling the protein to bind to specific DNA enhancer elements upstream of σ^{54} -dependent promoters (44, 47, 52, 56, 72). One exception to this scenario was recently reported (30), where *Pseudomonas aeruginosa* FleQ can activate transcription while bound in the downstream vicinity of the promoter.

Although the physiological advantages conferred by the σ^{54} -

EBP mode of transcription are not yet clear, activation of σ^{54} -dependent transcription is highly regulated by environmental cues through regulatory modules in the EBPs and in some cases by interactions with other regulatory proteins. Sensory modules in EBPs include CheY-like response regulator domains, PAS domains, GAF domains, PRD modules, and V4R domains, often represented within an N-terminal region (Fig. 1). These sequence features are described in more detail later in this article. Intriguingly, recent complete genome sequences have revealed some unusual EBPs containing regions of homology to other signal transduction domains and enzymes. With the large number of complete bacterial genomes now available, and with the importance of accurate annotation of future sequence data, we feel that it is timely to survey the variety of domain architectures found in these important proteins.

σ^{54} -DEPENDENT EBPs ARE AAA+ PROTEINS

The AAA+ proteins are a large family of chaperone-like ATPases that often act as molecular machines involved in the formation and remodeling of protein and protein-nucleic acid complexes (45, 51). The sequence feature common to all bacterial EBPs is the σ^{54} -RNAP interaction module (Pfam accession number PF00158). This stretch of about 300 amino acids is similar in sequence to the regions that define the AAA+ family (45, 73) and comprises several structural domains.

The EBPs share seven highly conserved sequence motifs, several of which are shared with other AAA+ proteins (44, 56, 72). Open reading frames with homology to EBPs can be identified in genome sequences by using Blast searches and from protein domain databases such as Pfam (8). However, an additional sequence feature that is diagnostic for genuine EBPs is the GAFTGA motif (73), which is found at residues 215 to 220 in *Escherichia coli* NtrC. This sequence motif, which is almost invariant in most EBPs, represents a structural feature that probably interacts with the σ factor and is unmasked or stabilized by nucleotide binding or hydrolysis (73). Related proteins that lack this GAFTGA motif (e.g., TyrR and *Rhodobacter capsulatus* NtrC) are unable to activate transcription at σ^{54} -dependent promoters (10, 48).

PspF-LIKE EBPs LACK A SENSORY INPUT DOMAIN

A number of EBPs comprise just the σ^{54} interaction module plus a DNA-binding domain. The best characterized of these is the phage shock protein, PspF, in enteric bacteria. The

* Corresponding author. Mailing address: Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom. Phone: 44 (0)1223 834244. Fax: 44 (0)1223 494919. E-mail: ds2@sanger.ac.uk.

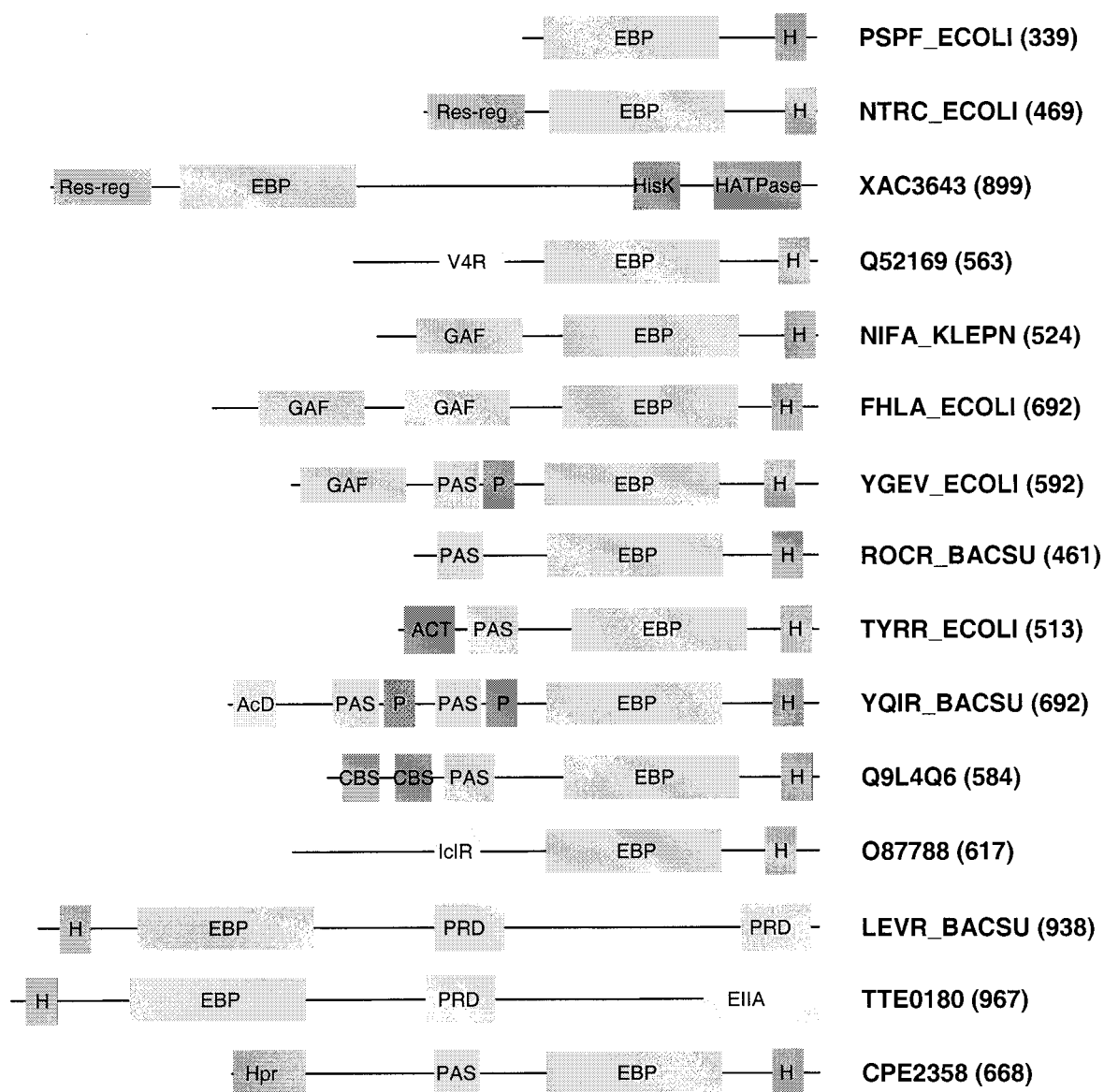


FIG. 1. Major domain architectures of bacterial EBPs. Examples of each of the known domain organizations found in bacterial EBPs are given. Sequences are identified by SwissProt/trEMBL accession numbers, except for XAC3643, TTE0180, and CPE23358, which are identified by their genome-specific open reading frame identifiers. The domain boundaries are shown to scale as predicted by Pfam. Each domain type is indicated by a different shade of gray. Matches to known protein families and domains are indicated as follows: EBP, enhancer-binding protein; H, helix-turn-helix domain; Res-reg, response regulator domain; HisK, histidine kinase domain; HATPase, ATPase domain; V4R, V4R domain; GAF, GAF domain; PAS, PAS domain; P, PAC domain; AcD, acetaldehyde dehydrogenase domain; CBS, CBS domain; IclR, IclR domain; PRD, PRD domain; EIIA, PTS enzyme IIA domain; Hpr, PTS Hpr domain. See text for a full description of these domains. Note that although no helix-turn-helix domain was detected in protein sequence XAC3643 by using the current release of Pfam (7.7), this does not imply that such a DNA-binding domain is absent. The lengths of the protein sequences are indicated in parentheses. The source organisms for each sequence are *B. subtilis* (ROCR_BACSU, YQIR_BACSU, and LEVR_BACSU), *C. sticklandii* (Q9L4Q6), *C. perfringens* (CPE2358), *E. coli* (PSPF_ECOLI, NTRC_ECOLI, FHLLA_ECOLI, YGEV_ECOLI, and TYRR_ECOLI), *P. putida* (Q52169 and O87788), *T. tencongensis* (TTE0180), and *Xanthomonas campestris* pv. Citri (XAC3643).

pspABCDE operon encodes several proteins that appear to help the cell adapt to changes in membrane integrity (19). It is induced by filamentous phage infection and by the presence of secretin proteins, which assist export of large protein complexes from the cell, including virulence factors and type II and type III secretion systems. PspA appears to have important functions in helping *E. coli* to cope with insertions of secretins

in the membrane in addition to its function as a regulator of PspF, and it may be these additional functions that have selected for its physical uncoupling from PspF. The *pspC* gene of *Yersinia enterocolitica* was found to be important for normal growth when the Ysc type III secretion system was expressed in the laboratory (14).

Strikingly, the PspF protein sequence does not contain any

known regulatory input domains; its activity is controlled by formation of a repressive complex with the PspA protein. Elderkin et al. (19) suggest that PspA may be an escaped regulatory domain of PspF. However, our phylogenetic analysis suggests that PspF is closely related to AlgB and belongs to a clade of response regulator EBPs (see below) (Fig. 2). Another well-studied example of EBPs lacking regulatory input domains is the HrpR/HrpS regulators of plant pathogenicity genes in *Pseudomonas syringae*. These EBPs are also negatively regulated by another protein, HrpV, although HrpV displays no detectable homology with PspA.

EBPs THAT ARE TWO-COMPONENT RESPONSE REGULATORS

Perhaps the best-known class of EBPs are those that contain an archetypal two-component response regulator receiver CheY-like domain (Pfam accession number PF00072) at their N termini. Two-component systems have evolved to allow bacteria to sense and respond to a wide range of stresses and environmental cues (25, 62). Examples include the nitrogen assimilation regulatory protein NtrC and the dicarboxylic acid transport regulator DctD. These proteins are activated by phosphorylation of an aspartate residue commonly located at position 54 in the N-terminal domain. In the case of NtrC, phosphorylation causes oligomerization and hence activation of transcription by σ^{54} -RNAP, while in DctD, phosphorylation relieves the inhibitory effect that the N-terminal domain has on the σ^{54} interaction module. In both cases, phosphorylation is carried out by a separate histidine kinase. Unlike other members of the NtrC family, *Rhodobacter capsulatus* NtrC lacks the GAFTGA motif and activates transcription by σ^{70} -RNAP rather than the σ^{54} holoenzyme. Transcriptional activation by this EBP requires ATP binding but not ATP hydrolysis (10).

Intriguingly, the *Xanthomonas anoxopodis* XAC3643 response regulator EBP also has a histidine kinase domain (Pfam accession number PF00512) fused to its C terminus (Fig. 1). The gene encoding this novel protein lies close to *ybjY*, which is preceded by a strong candidate σ^{54} -dependent promoter. Therefore, XAC3643 probably activates transcription of *ybjY*, encoding a protein that resembles an ABC transporter permease.

EBPs THAT CONTAIN PTS COMPONENTS

In addition to the two-component system, the other major bacterial signal transduction apparatus that relies on transient phosphorylation of regulatory proteins is the phosphoenolpyruvate-dependent phosphotransferase system (PTS). The main function of the PTS in *E. coli* is to carry out the fundamental process of substrate-level phosphorylation, whereby the transport and activation of sugar substrates are thermodynamically coupled to dephosphorylation of the glycolytic intermediate phosphoenolpyruvate. The phosphoryl relay proceeds sequentially from phosphoenolpyruvate to enzyme I (EI), HPr, enzyme II (EII), and finally the incoming sugar that is transported across the membrane and concomitantly phosphorylated by EII. Additionally, PTS components participate in signal transduction, chemotaxis, and the regulation of some key physiological processes, e.g., carbohydrate transport, catabolite

repression, carbon storage, and coordination of carbon and nitrogen metabolism (50, 53, 54, 58, 67). The molecular basis of the phosphoenolpyruvate-dependent phosphorylation cascades is well characterized thanks to many recent structural and functional studies, including the solution of three-dimensional structures of several of the protein components (69).

Those proteins that are subject to direct regulation by the PTS share a common domain called the PTS regulation domain (PRD) (69). PRDs (Pfam accession number PF00874) are found, usually in pairs, in the BglG/SacY antiterminators as well as in a variety of transcriptional activators such as *Bacillus stearothermophilus* MtlR and *Bacillus subtilis* LicR, and their transcription-regulatory activities are modulated by the phosphorylation states of conserved histidine residues. The LevR EBP in *B. subtilis* activates transcription of *levDEFG*, encoding a PTS EII complex for the transport and phosphorylation of fructose. This EII complex “fine-tunes” activity of LevR, stimulating it by phosphorylation at His585 and suppressing it by phosphorylation at His869 (40).

Genome sequencing projects have revealed several more PRD-containing EBPs in addition to LevR. In most cases these are in gram-positive bacteria (*Bacillus* spp., *Clostridium* spp., *Listeria* spp., and *Thermoanaerobacter tencongensis*), but homologues have also been detected in *Salmonella enterica* serovar Typhimurium and *Klebsiella pneumoniae* (63). In all cases, genes encoding PRD-containing EBPs are found situated adjacent to genes encoding PTS components, often with an obvious σ^{54} -dependent promoter. Interestingly, in the food-borne pathogen *Listeria monocytogenes* the LevR homologue ManR activates an EII complex that confers susceptibility to the bacteriocin mesentericin Y105 (13).

Our survey of the domain organization of EBPs in recently sequenced genomes (Table 1) revealed some novel PRD-containing EBPs. For example, CP2358 and CAC3088 in *Clostridium perfringens* each contain an HPr domain in addition to a PAS domain (see below) and the σ^{54} interaction module. HPr is a small cytoplasmic phosphocARRIER protein that is a central component of the general PTS. A conserved histidine residue in HPr serves as an acceptor for the phosphoryl group from EI. This phosphoryl group can then be transferred to a number of regulatory targets such as EII components. This situation is not totally without precedent, as domains with homology to PTS components are sometimes found within larger modular proteins. For example in *E. coli* the FrwA and FryA proteins each contain HPr, EI, and EII^A domains (67). In the clostridial species, genes that encode HPr-containing EBPs are found adjacent to genes encoding PTS systems.

In the thermophilic anaerobe *T. tencongensis*, protein TTE0180 is a PRD-containing EBP that also contains an EII^A domain that probably forms a galactitol-specific EII complex with the TTE0182 and TTE0183 gene products. Lying between the TTE0180 and TTE0182 open reading frames is TTE0181, encoding a PTS EI, which may therefore form part of the phosphotransfer chain with the galactitol-specific EII.

EBPs CONTAINING GAF, PAS, AND ACT DOMAINS

A protein domain found in cyclic GMP-specific and stimulated phosphodiesterases, adenylate cyclases, and *E. coli* FhlA has been defined as the GAF domain (3). GAF domains (Pfam

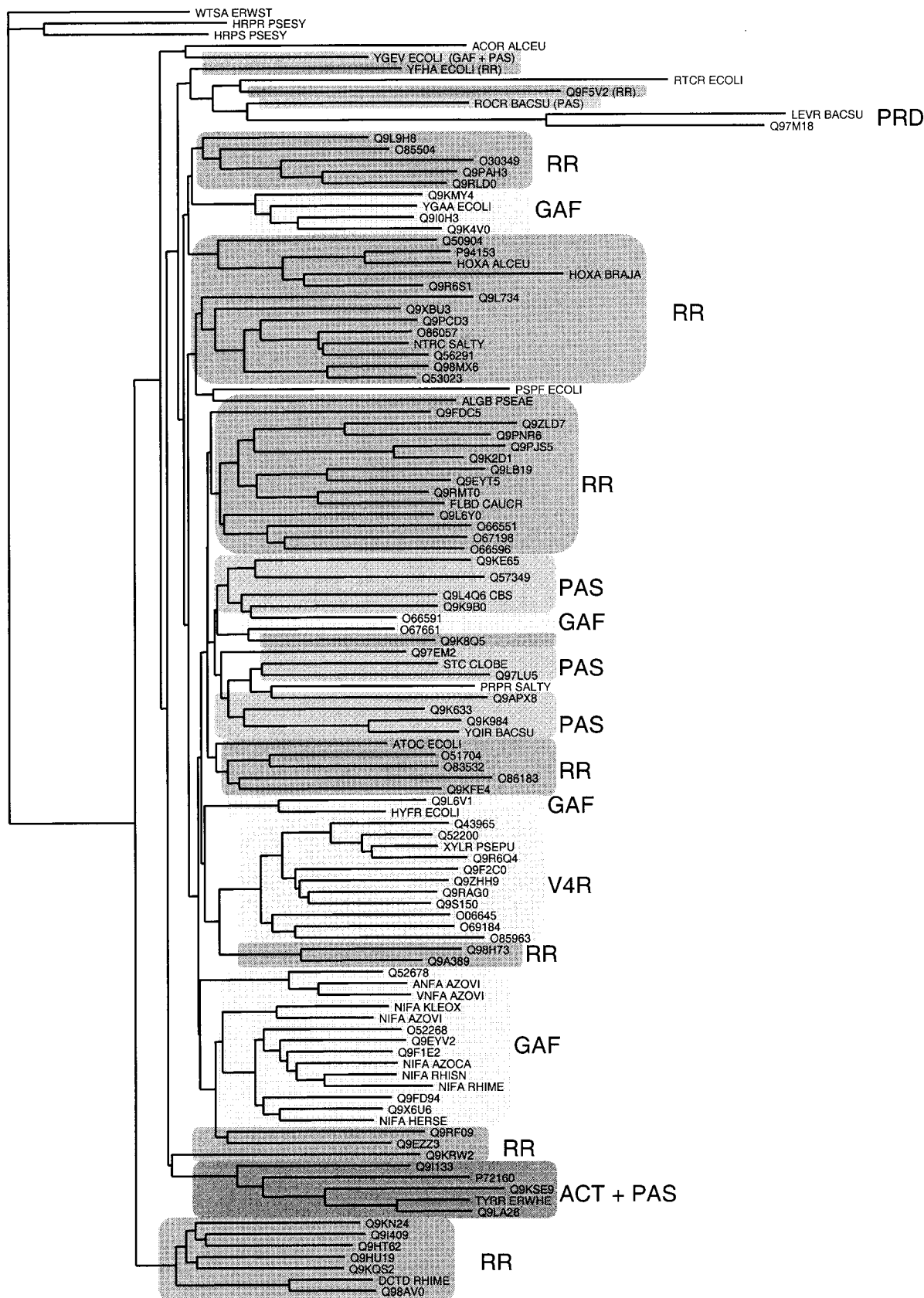


FIG. 2. Phylogenetic analysis of bacterial EBPs. A neighbor-joining tree of representative EBP sequences from the SwissProt/trEMBL databases was constructed by using the Phylip computer package. The root of the tree is unknown, and the location of the root was chosen arbitrarily for clarity of the figure. Those sequences predicted to contain PRD, response regulator (RR), PAS, GAF, V4R, or ACT domains are indicated by shading. Note that PSPF_ECOLI, HRPR_PSESY, HRPS_PSESY, and W TSA_ERWST lack an N-terminal sensory domain. RTCR_ECOLI, ACOR_ALCEU, and PRPR_SALTY have N-terminal sequences that do not closely match any known domains.

TABLE 1. Species distribution of enhancer-dependent transcription in bacteria^a

Group	Example species	RpoN present	No. of EBPs
<i>Actinobacteria</i>	<i>Streptomyces coelicolor</i> A3	No	0
<i>Aquificae</i>	<i>Aquifex aeolicus</i>	Yes	5
<i>Bacteroidetes-Chlorobi</i> group	<i>Chlorobium tepidum</i>	Yes	3
<i>Chlamydiae-Verrucomicrobia</i> group	<i>Chlamydia pneumoniae</i> AR39	Yes	1
<i>Chloroflexi</i> (green nonsulfur bacteria)	<i>Chloroflexus aurantiacus</i>	No	0 (+TyrR) ^b
<i>Chrysiogenetes</i>	NA ^c	NA	NA
<i>Cyanobacteria</i> (blue-green algae)	<i>Synechocystis</i> sp. strain PCC 6803	No	0 (+TyrR)
<i>Deferribacteres</i>	NA	NA	NA
<i>Deinococcus-Thermus</i>	<i>Deinococcus radiodurans</i>	No	0
<i>Dictyoglomi</i>	NA	NA	NA
<i>Fibrobacteres-Acidobacteria</i> group	<i>Fibrobacter succinogenes</i>	Yes	7? ^d
<i>Firmicutes</i> (gram-positive bacteria)	<i>Bacillus subtilis</i>	Yes	3
	<i>Lactococcus lactis</i>	No	0
	<i>Staphylococcus aureus</i>	No	0
	<i>Thermoanaerobacter tencongensis</i>	Yes	10
<i>Fusobacteria</i>	<i>Fusobacterium nucleatum</i>	No	2
<i>Nitrospirae</i>	NA	NA	NA
<i>Planctomycetes</i>	<i>Planctomyces limnophilus</i>	Yes	? ^e
<i>Proteobacteria</i> (purple bacteria)	<i>Escherichia coli</i>	Yes	12 (+TyrR)
<i>Spirochaetes</i>	<i>Borrelia burgdorferi</i>	Yes	2
<i>Thermodesulfobacteria</i>	NA	NA	NA
<i>Thermomicrobia</i>	NA	NA	NA
<i>Thermotogae</i>	<i>Thermotoga maritima</i>	No	0

^a Each of the major taxonomic groups of bacteria is listed, as defined by the National Center for Biotechnology Information taxonomy browser (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy>). Where complete genome sequences were available for example members of a group, BLAST searches were performed against the genome sequences with the amino acid sequences PSPF_ECOLI and RPON_ECOLI as probes. Statistically significant hits were aligned against known sequences and manually checked for likely biological significance.

^b TyrR is homologous to the EBPs but does not activate σ^{54} -RNA polymerase.

^c NA, no data are available for this phylum.

^d The genome sequence is not yet completed.

^e No complete genome sequence is available.

accession number PF01590) were first described as noncatalytic cyclic GMP-binding domains in cyclic nucleotide phosphodiesterases (12). Subsequently they were detected in the phytochrome regulatory photoreceptors in plants and cyanobacteria, and they have since been detected in diverse bacterial and eukaryotic signaling proteins. GAF domains occur in EI of the PTS in *E. coli* as well as in the N termini of NifA-related EBPs. In *E. coli*, the FhlA EBP contains two N-terminal GAF domains, which bind formate, leading to transcriptional activation of the formate-hydrogen lyase system (26).

NifA, regulating nitrogen fixation, and related EBPs have been well studied in several organisms. The N-terminal GAF domain of *Azotobacter vinelandii* NifA has been shown to bind 2-oxoglutarate and modulates the response to the regulatory protein (5, 37). According to environmental conditions, NifL forms a repressive complex with NifA analogous to PspA-PspF and HrpV-Hrp/HrpS interactions (43), although NifL shows no homology to either PspA or HrpV. In contrast, the NifA proteins from *Herbaspirillum seropedicae* and *Azospirillum brasilense* detect the nitrogen status via an N-terminal GAF domain, which may interact directly with the PII signal transduction protein (60). It has also been demonstrated that the *A. vinelandii* NifL-NifA system interacts directly with the PII protein (36, 38, 55, 57).

Other, less well characterized EBPs containing a single GAF domain include YgaA in *E. coli*, AQ_218 and AQ_1792 in *Aquifex aeolicus*. YgaA (also known as NorR) has recently been shown to activate transcription of a nitric oxide detoxifi-

cation system (21, 28). The stimuli detected by AQ_218 and AQ_1792 are unknown, but AQ_218 was shown to bind upstream of the *glnBA* operon for nitrogen assimilation (65).

Another regulatory domain commonly found in bacterial EBPs is the Per, ARNT, and Sim (PAS) domain (49), which, despite sharing little sequence similarity, has a three-dimensional structure similar to that of GAF, indicating common ancestry (24). PAS domains (Pfam accession number PF00989) are used as signal sensors in many signaling proteins in archaea, bacteria, and eukaryotes (66, 74). PAS domain proteins often detect a signal by way of a bound cofactor such as heme, flavin, or a 4-hydroxycinnamoyl chromophore.

Like *R. capsulatus* NtrC, the *E. coli* TyrR protein is clearly homologous to the σ^{54} -dependent EBPs but lacks the GAFTGA motif required for interaction with σ^{54} -RNAP. Consequently, it is unable to activate transcription from σ^{54} -dependent promoters (48) but nevertheless regulates σ^{70} -dependent promoters for genes involved in degradation of aromatic amino acids. TyrR has an additional sequence at its N terminus, in addition to the PAS domain. This additional sequence appears to encode an aspartokinase, chorismate mutase, and TyrA (ACT) domain (4). ACT domains (Pfam accession number PF01842) are found in a wide range of metabolic enzymes that are regulated by amino acid concentration, with pairs of ACT domains binding specifically to a particular amino acid, leading to regulation of the linked enzyme. In solution TyrR normally exists as a dimer, but it undergoes a reversible conformational change to a hexamer in the presence of ATP and

tyrosine (18, 71). In *P. aeruginosa*, the PA2449 sequence (trEMBL accession number Q9I133) appears to have a PAS domain and an ACT domain at its N terminus, and it has an intact GAFTGA motif and so is potentially able to activate transcription of the nearby predicted σ^{54} -dependent promoter upstream of *gcvH2*, encoding a glycine cleavage protein. Another protein (PhhR, PA0873), which has the sequence GAFEGA for its motif, performs the function of TyrR in *P. aeruginosa* but requires σ^{54} -RNAP for activation of the *phh* operon (59). This protein is therefore competent to regulate σ^{70} -dependent promoters but interacts with the σ^{54} holoenzyme to activate transcription.

EBPs CONTAINING V4R DOMAINS

The V4R, or vinyl 4 reductase, domain (Pfam accession number PF02830) is predicted to bind small molecules such as hydrocarbons and is found in proteins involved in chlorophyll biosynthesis as well as in regulators of phenol degradation (2). One distinct phylogenetic clade of EBPs is comprised of proteins containing one or more V4R domains (Fig. 2). Those members of this group that have been investigated regulate the catabolism of aromatic hydrocarbons and phenols. For example, the prokaryotic EBP XylR is the central regulator of the toluene degradation pathway in *Pseudomonas* species. Genetic and biochemical data indicate that the N-terminal domain of the protein interacts directly with *m*-xylene, which renders the protein competent as a transcriptional activator. Three-dimensional models that go some way to explaining the molecular basis for interaction with substrate have been generated for the V4R domains of XylR and DmpR (17).

NOVEL AND UNEXPECTED DOMAIN CONTEXTS

In addition to the response regulators and GAF, PAS, PRD, and V4R domains, additional domains are found in a few bacterial EBPs. For example the *Clostridium sticklandii* Q9L4Q6 protein, also called PrdR, contains two cystathionine- β -synthase (CBS) domains in addition to a PAS domain at its N terminus (Fig. 1). CBS domains (Pfam accession number PF00571) are found, usually in pairs, in a wide range of proteins and are assumed to have a regulatory role (7). Mutations in the CBS domains of cystathionine- β -synthase lead to homocystinuria in humans. Among bacteria, CBS-containing proteins include IMP dehydrogenase (in *E. coli* and *B. subtilis*), glycine betaine transport ATP-binding protein (in *E. coli* and *B. subtilis*), acetoin utilization protein AcuB (in *B. subtilis*), and numerous proteins of unknown function. PrdR is believed to activate transcription of *prdB*, encoding the 26-kDa subunit of the selenoenzyme D-proline reductase in *C. sticklandii* (31, 32).

Pseudomonas putida protein O87788 has a fragment of an IclR domain (PF01614) at its N terminus. This domain is otherwise known only in the IclR family of transcriptional regulators, which includes the glyoxylate shunt repressor in *E. coli* (39). It is unclear whether the IclR domain is functional in protein O87788.

The BkdR protein (YqiR) is a σ^{54} -dependent EBP that activates expression of an isoleucine and valine degradation pathway in *B. subtilis*. As Debarouille et al. (15) point out, there are two PAS domains in the N-terminal portion of BkdR.

However, Pfam also detects a fragment of a domain normally found in the enzyme acetaldehyde dehydrogenase (Pfam accession number PF02396).

SPECIES DISTRIBUTION OF ENHANCER-DEPENDENT TRANSCRIPTION IN BACTERIA

Now that complete sequences are available for about 70 bacterial genomes, representing many of the major phyla, the pattern of occurrence of enhancer-dependent transcription is becoming clear (Table 2). Genes encoding the σ^{54} -RNAP subunit and EBPs have been identified in numerous proteobacteria, including most of those whose genomes have been sequenced. The only exceptions are highly specialized parasites that have undergone reductive evolution. Genes for enhancer-dependent transcription are widespread, even among organisms with relatively small genomes. For example, in several species of *Chlamydia*, σ^{54} is one of only two alternative σ factors present (33).

In contrast, the green nonsulfur bacterium *Chloroflexus aurantiacus* and the cyanobacterium *Synechocystis* sp. strain PCC 6803 do not encode σ^{54} but contain homologues of TyrR, a protein that is clearly very closely related to the EBPs but that does not activate transcription at σ^{54} -dependent promoters (48). The complete genome sequences of several high-GC gram-positive bacteria (actinobacteria) contain no obvious homologues of σ^{54} or any obvious EBPs, indicating a rather patchy distribution of the enhancer-dependent transcription system among bacteria (64).

It is noteworthy that the genome of the oral pathogen *Fusobacterium nucleatum* (34) does not appear to encode σ^{54} , yet there are two open reading frames, FN1321 and FN1831, with homology to σ^{54} -dependent EBPs and with response regulator domains at their N termini. This indicates that *F. nucleatum* must have either lost its σ^{54} or gained the EBP-like sequences by horizontal transfer, suggesting that the EBP-dependent regulon is subject to significant selection pressure. These EBPs contain the sequences GAFLGA and GSFEGA in place of the conserved GAFTGA motif. Site-directed mutagenesis of the GAFTGA regions of well-characterized EBPs (22, 70) suggests that these proteins may not be competent to activate transcription by σ^{54} -RNAP. Thus, like TyrR and *R. capsulatus* NtrC, in which the GAFTGA sequence appears to have been deleted rather than replaced, these EBPs may have evolved to regulate transcription by the housekeeping σ^{70} -RNAP holoenzyme.

PREDICTED REGULATORY TARGETS OF EBPs IN COMPLETELY SEQUENCED GENOMES

Potential σ^{54} -dependent promoters, the targets of EBPs, are fairly easy to detect in DNA sequences (53, 63, 64) and have a well conserved sequence that can be represented as YTGGC ACGrNNNTTGCW (6). We used the PromScan program (<http://www.promscan.uklinux.net>) to search for close matches to this consensus σ^{54} -dependent promoter sequence. The search used a frequency matrix based on a compilation of 186 known promoters from a range of bacteria (6). Previously, predictions made in this way have often been subsequently confirmed by experimental studies (see, e.g., references 27, 41, and 64). Furthermore, genes encoding regulatory proteins are

TABLE 2. Complete repertoires of EBPs in some example genomes^a

Organism and protein	Domain(s)	Promoter(s)	Probable system controlled by EBP
<i>Chlamydia trachomatis</i> L2 CT468	RR ^b	CT652.1, CT683	Unknown
<i>Borrelia burgdorferi</i> BB0419 BB0763	RR RR	BB420 <i>rpoS</i>	Uncharacterized two-component system Sigma factor σ^s
<i>Chlorobium tepidum</i> CT0108 CT1497 NifA (CT1529)	GAF Unidentified GAF	CT0107 Unknown <i>nifH</i>	Outer membrane protein Unknown Nitrogenase iron protein
<i>Clostridium acetobutylicum</i> CAC0459 CAC3088 CAC0382	PAS HPr, PAS PRD	CAC0458 CAC3083 CAC0383	Permease of unknown function PTS system PTS system
<i>Clostridium perfringens</i> CPE2358	HPr, PAS	CPE2357	PTS system
<i>Escherichia coli</i> K-12 NtrC	RR	<i>ddpXABCDE, glnALG,</i> <i>glnHPQ, glnK-amtB,</i> <i>gluIJKL, nac, potFGHI,</i> <i>yeaGH, yggG, yhdWXYZ</i>	Glutamine synthetase and other functions of nitrogen metabolism
PspF YfhA YgaA HydG (ZraR) PrpR FhlA	None RR GAF RR Unidentified GAF	<i>pspABCDE</i> <i>yfhKGA</i> b2710 <i>zraP, zraSR</i> <i>prpBCDE</i> <i>hyp</i> and <i>hyc</i> operons, <i>hydN-</i> <i>hypF</i>	Phage shock response Uncharacterized two-component system Nitric oxide detoxification system Heavy-metal tolerance system Propionate catabolism Formate hydrogen lyase
RtcR AtoC HyfR YgeV DhaR	Unidentified RR Unidentified, GAF GAF, PAS Unidentified, PAS	<i>rtcBA</i> <i>atoDAEB</i> <i>hyf</i> operon <i>ygeW</i> Unknown	RNA terminal cyclase Acetoacetate catabolism Hydrogenase 4 Carbamoyl transferase Dihydroxyacetone catabolism
<i>Helicobacter pylori</i> 26695 FlgR	RR	<i>flaB, flgE, flgDE', flgBC, flgK.</i>	Flagellar apparatus
<i>Listeria innocua</i> lin2276 lin0778 lin1832	Unidentified PRD PRD	lin2275 lin0777 lin1831	Propionate coenzyme A transferase PTS system PTS system
<i>Listeria monocytogenes</i> lmo0785 lmo1721 lmo2173	PRD PRD Unidentified	lmo0784 lmo1720 lmo2172	PTS system PTS system Propionate coenzyme A transferase
<i>Pseudomonas aeruginosa</i> PhhR (PA0873) FleQ (PA1097) FleR (PA1099) PA1196 PA1335 PA1663 PA1945 PA2005 PA2354 PA2449 YgaA (NorR) PA2665 PA3932 PA4021 AcoR (PA4147) PilR (PA4547) RtcR (PA4581) PA4726 NtrC (PA5125) PA5166 AlgB (PA5483) PA5511	PAS, ACT Unidentified RR Unidentified RR Unidentified Unidentified Unidentified None PAS, ACT GAF None Unidentified Unidentified RR Unidentified RR RR RR RR RR	<i>phh</i> operon <i>flgF, fltD, fltS, fltE</i> <i>flgF, fltD, fltS, fltE</i> PA1195 PA1342 PA1656 PA1934 PA2004 <i>msuE</i> <i>gcvH2</i> Unknown Unknown PA4022 PA4148 <i>pilA</i> PA4582 Unknown <i>glnA</i> PA5167 PA5482 PA5530	Phenylalanine hydroxylase Flagellar motility Flagellar motility Unknown Binding component of ABC transporter Unknown Unknown Unknown NADH-dependent flavin mononucleotide reductase Glycine cleavage system protein H2 Unknown Unknown Aldehyde dehydrogenase Short-chain dehydrogenase Type 4 fimbrial precursor PilA Unknown Unknown Glutamine synthetase C4-dicarboxylate-binding protein Unknown Dicarboxylate transporter
<i>Ralstonia solanacearum</i> RSP1667 RSP0228	RR V4R	RSP1668, <i>rpoN2</i> RSP0229	EBP, σ^{54} Benzaldehyde dehydrogenase oxidoreductase

Continued on following page

TABLE 2—Continued

Organism and protein	Domain(s)	Promoter(s)	Probable system controlled by EBP
RSC0332	RR	<i>dctA1</i>	C4-dicarboxylate transport transmembrane protein
RSP0959	Unidentified	RSP0958	Unknown
RSP1668	Unidentified	RSP1668, <i>rpoN2</i>	EBP, σ^{54}
RSC1186	Unidentified	None	Unknown
RSC1261	RR	<i>glnA1</i>	Glutamine synthetase
RSP0009	RR	Unknown	Unknown
RSC2807	RR	Unknown	Unknown
RSP0123	PAS	Unknown	Unknown
RSC3129	PAS	<i>exaC</i>	Putative NAD ⁺ -dependent acetaldehyde dehydrogenase oxidoreductase protein
RSP1079	Unidentified	RSP1076	Glucose-fructose oxidoreductase oxidoreductase protein
<i>Thermoanaerobacter tencongensis</i>			
TTE0180	PRD, EIIA	<i>ptsN</i> , TTE0180	PTS system
TTE0268	GAF	TTE0269	Unknown
TTE0331	PRD	<i>celB</i>	PTS system
TTE0550	PAS	<i>maoC</i>	Acyl dehydratase
TTE0717	PAS	TTE0718	Acyl-coenzyme A hydrolase
TTE0825	Unidentified	<i>aprE2</i>	Subtilisin-like serine protease
TTE1710	V4R, PAS	<i>cooC</i>	CO dehydrogenase maturation factor
TTE0191	PRD	<i>manX</i>	PTS system
TTE1998	PAS	<i>dak1.2</i>	Dihydroxyacetone kinase
TTE0687	PAS	<i>acoA2</i>	Thiamine pyrophosphate-dependent dehydrogenase, E1 component alpha subunit
<i>Vibrio cholerae</i>			
VCA0117	Unidentified	VCA0105	Unknown
VCA0142	RR	VCA0144	Immunogenic protein
VCA0182	Unidentified	VCA0183	Ferrisiderophore reductase
PgtA (VCA0704)	RR	Unknown	Unknown
LuxO (VC1021)	RR	VC1020	Unknown
TyrR (VC1308)	ACT, PAS	VC1301	Serine transporter
VC1522	RR	VC1518	Unknown
		VC1523	Unknown
		VC1516	Iron-sulfur cluster-binding protein
PspF (VC1679)	None	VC1678	Phage shock protein A PspA
VC1817	Unidentified	Unknown	Unknown
VC1926	RR	VC1929	C4-dicarboxylate-binding periplasmic protein
VC2137	Unidentified	<i>ftiE</i>	Flagellar hook-basal body complex protein
		VC2136	Sensory box histidine kinase
		<i>ftiK</i>	Flagellar hook length control protein
NtrC (VC2749)	RR	<i>glnA?</i>	Glutamine synthetase?
<i>Xanthomonas campestris</i> pv. <i>Campestris</i>			
XCC0189	RR	<i>glnA</i>	Glutamine synthetase
XCC1933	Unidentified	<i>vioA</i>	Nucleotide sugar transaminase
XCC1030	Unidentified	<i>prpB</i>	Carboxyphosphoenolpyruvate phosphonmutase
XCC0563	RR	<i>ybjY</i>	ABC transporter permease
XCC3689	RR	XCC3685	Unknown
XCC2590	RR	None	Unknown
XCC3472	Unidentified	<i>mxAF</i>	Methanol dehydrogenase heavy chain
<i>Xanthomonas axonopodis</i> pv. <i>Citri</i>			
XAC0208	RR	<i>glnA</i>	Glutamine synthetase
XAC0226	RR	XAC0227	Unknown
XAC3238	RR	<i>fimA</i>	Fimbrillin
XAC0845	None	XAC0836	Unknown
		XAC0838	Unknown
XAC1967	Unidentified	<i>vioA</i>	Nucleotide sugar transaminase
XAC1136	Unidentified	<i>prpB</i>	Carboxyphosphoenolpyruvate phosphonmutase
XAC3643	RR, Hisk, HATPase	<i>ybjY</i>	ABC transporter permease
XAC3733	RR	None	Unknown
XAC0654	Unidentified	<i>mxAF</i>	Methanol dehydrogenase heavy chain
<i>Yersinia pestis</i> CO92			
NtrC	RR	<i>glnALG</i>	Glutamine synthetase
PspF	None	<i>pspABCDE</i>	Phage shock response
YfhA	RR	<i>yfhK</i>	Uncharacterized two-component system
FleR	None	<i>flgB</i> , <i>fljD</i>	Flagellar apparatus

^a The complete repertoires of EBPs were inferred from several complete genomes by using a combination of BLAST searches and manual inspection of alignments. Potential σ^{54} -dependent promoters were detected by using the PromScan tool (<http://www.promscan.uklinux.net>). The list of promoters is not necessarily exhaustive. The domain organization of each potential EBP was examined by using the Pfam protein search facility (<http://www.sanger.ac.uk/Software/Pfam/search.html>). The predicted sensory input domains are listed.

^b RR, response regulator.

often situated adjacent to their target genes on the chromosome. Therefore, by combining promoter prediction and positional information, we can glean clues as to the targets of these regulatory proteins. To better understand the patterns of variation in the enhancer-dependent regulon, we took this approach to analyzing the complement of EBPs in several diverse bacteria (Table 1).

The number of EBPs encoded in genomes is rather variable between organisms, as is the repertoire of functions regulated by enhancer-dependent transcription. Recently the complete sets of EBPs and their probable regulatory targets have been described for *E. coli* K-12 (53) and for *S. enterica* serovar Typhimurium, *S. enterica* serovar Typhi, and *Yersinia pestis* (63). However, additional complete genome sequences have become available for additional strains of *E. coli* and *Y. pestis*, allowing us to compare the repertoires of EBPs between closely related organisms and hence make inferences about the recent evolutionary history of this regulon.

Two different biovars of the plague bacillus, *Y. pestis*, have been completely sequenced, representing two different plague pandemics. The genome of *Y. pestis* KIM (16) encodes three EBPs: NtrC (regulator of nitrogen assimilation), PspF (regulator of several stress responses), and YfhA (unknown function). However, *Y. pestis* CO92 encodes an additional EBP, YPO0712 (also annotated as FleQ) (46). Blast searches revealed high similarity to the FleR/FleQ proteins from *Pseudomonas* species, which activate transcription of several genes involved in flagellar motility in a σ^{54} -dependent fashion. Indeed, YPO0712 is situated close to several operons containing genes for components of the flagellar apparatus, and several of these have good candidate σ^{54} -dependent promoter sequences (data not shown). The regulation of motility and chemotaxis has been well studied in *E. coli* and *Salmonella* species and does not directly involve the σ^{54} -dependent regulon. Since close homologues of FleR and FleQ are not found in any of the closely related enteric bacteria, it seems likely that an ancestor of *Y. pestis* CO92 has recently acquired the flagellar transcriptional regulator from a *Pseudomonas*-like organism. It is unclear whether *Y. pestis* KIM has subsequently lost this EBP and the cognate promoters or whether *Y. pestis* CO92 acquired them since diverging from KIM. As Deng et al. (16) point out, the biological significance of genes encoding flagellar components is obscured by the observation that *Y. pestis* is nonmotile. We may just be seeing the remnants of an obsolete motility apparatus, or alternatively, the flagellum might be required in some hitherto-undiscovered stage of the plague bacillus life cycle.

Complete genome sequences are available for both the *E. coli* laboratory strain K-12 (9) and the pathogenic (enterohemorrhagic) *E. coli* strain O157:H7 (23). Furthermore, the complete genome sequence for the very closely related *Shigella flexneri* 2a has recently become available (29). *E. coli* K-12 possesses 12 known σ^{54} -dependent EBPs. Of these, a core of eight are conserved in O157:H7 and *S. flexneri*, three of which are also conserved in *Y. pestis* and *Salmonella* species (53, 63). One of those missing from O157:H7 and *S. flexneri* is AtoC. In *E. coli* K-12, AtoC regulates acetoacetate catabolism and is required for utilization of short-chain fatty acids. The absence of this regulator is consistent with the observation that O157:H7 is unable to utilize short-chain fatty acids (23).

In contrast to the situation in the enteric bacteria described above, the two sequenced strains of the gastric pathogen *Helicobacter pylori* each contain only one EBP-encoding sequence (1, 68). This EBP, called FlgR, is known to play a major role in the transcriptional regulation of flagellar motility (35, 61). However, as well as having reduced motility, an *rpoN* mutant of *H. pylori* also displayed reduced viability and altered cell morphology in stationary phase (20). Interestingly, we have found a potential σ^{54} -dependent promoter sequence (AGGCATTG GGTTTGCG) upstream of the gene encoding the MreB cell shape-determining protein in *H. pylori* 26695 but not in *H. pylori* J99.

It is clear from the data in Table 2 and from other surveys of this regulon (6, 11, 42, 63, 64) that whereas some genes, e.g., *glnA*, are commonly controlled by EBPs, in some other genera EBPs regulate diverse physiological functions in different species (and even different strains within a species). Therefore, the elaborate σ^{54} -EBP mode of transcription has apparently been recruited to provide fine-tuning of specific metabolic processes appropriate to the physiology of the particular host organism. It is presently unclear what selective pressures maintain this plasticity.

PHYLOGENETIC ANALYSIS OF THE BACTERIAL EBPs

To better understand the evolution of domain arrangements in EBPs, we performed a phylogenetic analysis of EBP protein sequences. We obtained an alignment from the Pfam database (8) that contained all of the EBP central region (σ^{54} interaction module) sequences in SwissProt and trEMBL. We manually edited the alignment to remove obvious false positives and reduced redundancy by removing multiple sequences of greater than 80% identity, and we used this alignment to produce the neighbor-joining phylogenetic tree shown in Fig. 2. We also used maximum-likelihood and parsimony methods to build trees, which were broadly consistent with the neighbor-joining tree.

From Fig. 2, it is evident that the sequences generally cluster with sequences from proteins with similar domain architectures. Those EBPs containing a V4R domain and those with PRD domains form single clusters, as do those containing both PAS and ACT domains. However, those EBPs containing a response regulator domain do not fall into a single phylogenetic cluster. Similarly, those EBPs containing GAF or PAS domains are polyphyletic. This suggests that the fusion of these sensory domains with the σ^{54} interaction module may have arisen several times during the evolution of the bacteria.

CONCLUDING REMARKS

Signal transduction proteins are often poorly annotated in sequence databases, largely because of their complex domain organizations. This is particularly true for the bacterial EBPs. For example, *E. coli* YgaA has been misannotated as a response regulator. Therefore we took a domain-oriented approach to surveying the structural diversity of this important family of regulators and revealed several novel features as well as rationally classifying the more familiar members of the family. Furthermore, we have developed the powerful approach of combining positional analysis of genes with detection of bind-

ing sites in DNA to infer novel insights into the enhancer-dependent regulon in diverse completely sequenced organisms.

ACKNOWLEDGMENTS

We thank Alex Bateman and the Pfam group for useful discussions and support.

REFERENCES

- Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**:176–180.
- Anantharaman, V., E. V. Koonin, and L. Aravind. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J. Mol. Biol.* **307**:1271–1292.
- Aravind, L., and C. P. Ponting. 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem. Sci.* **22**:458–459.
- Aravind, L., and E. V. Koonin. 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* **287**:1023–1040.
- Barrett, J., P. Ray, and A. Sobczyk. 2001. Concerted inhibition of the transcriptional activation functions of the enhancer-binding protein NIFA by the anti-activator NIFL. *Mol. Microbiol.* **39**:480–494.
- Barrios, H., B. Valderrama, and E. Morett. 1999. Compilation and analysis of σ^{54} -dependent promoter sequences. *Nucleic Acids Res.* **27**:4305–4313.
- Bateman, A. 1997. The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends Biochem. Sci.* **22**:12–13.
- Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30**:276–280.
- Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453–1474.
- Bowman, W. C., and R. G. Kranz. 1998. A bacterial ATP-dependent, enhancer binding protein that activates the housekeeping RNA polymerase. *Genes Dev.* **12**:1884–1893.
- Buck, M., M. T. Gallegos, D. J. Studholme, Y. Guo, and J. D. Gralla. 2000. The bacterial enhancer-dependent σ^{54} (σ^N) transcription factor. *J. Bacteriol.* **182**:4129–4136.
- Charbonneau, H., R. K. Prusti, H. LeTrong, W. K. Sonnenburg, P. J. Mulaney, K. A. Walsh, and J. A. Beavo. 1990. Identification of a noncatalytic cGMP-binding domain conserved in both the cGMP-stimulated and photo-receptor cyclic nucleotide phosphodiesterases. *Proc. Natl. Acad. Sci. USA* **87**:288–292.
- Dalet, K., Y. Cenatiempo, P. Cossart, and Y. Hechard. 2001. A σ^{54} dependent PTS permease of the mannose family is responsible for sensitivity of *Listeria monocytogenes* to mesentericin Y105. *Microbiology* **147**:3263–3269.
- Darwin, A. J., and V. L. Miller. 2001. The *psp* locus of *Yersinia enterocolitica* is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. *Mol. Microbiol.* **39**:429–444.
- Debarbouille, M., R. Gardan, M. Arnaud, and G. Rapoport. 1999. Role of *bkdR*, a transcriptional activator of the *sigL*-dependent isoleucine and valine degradation pathway in *Bacillus subtilis*. *J. Bacteriol.* **181**:2059–2066.
- Deng, W., V. Burland, G. Plunkett III, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles, J. S. Matson, F. R. Blattner, and R. D. Perry. 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**:4601–4611.
- Devos, D., J. Garmendia, V. de Lorenzo, and A. Valencia. 2002. Deciphering the action of aromatic effectors on the prokaryotic enhancer-binding protein XylR: a structural model of its N-terminal domain. *Environ. Microbiol.* **4**:29–41.
- Dixon, M. P., R. N. Pau, G. J. Howlett, D. E. Dunstan, W. H. Sawyer, and B. E. Davidson. 2002. The central domain of *Escherichia coli* TyrR is responsible for hexamerization associated with tyrosine-mediated repression of gene expression. *J. Biol. Chem.* **277**:23186–23192.
- Elderkin, S., S. Jones, J. Schumacher, R. Studholme, and M. Buck. 2002. Mechanism of action of the *Escherichia coli* phage shock protein PspA in repression of the AAA family transcription factor PspF. *J. Mol. Biol.* **320**:23–37.
- Fujinaga, R., T. Nakazawa, and M. Shirai. 2001. Allelic exchange mutagenesis of *rpoN* encoding RNA-polymerase σ^{54} subunit in *Helicobacter pylori*. *J. Infect. Chemother.* **7**:148–155.
- Gardner, A. M., R. A. Helmick, and P. R. Gardner. 2002. Favorubredoxin, an inducible catalyst for nitric oxide reduction and detoxification in *Escherichia coli*. *J. Biol. Chem.* **277**:8172–8177.
- González, V., L. Olvera, X. Soberon, and E. Morett. 1998. In vivo studies on the positive control function of NifA: a conserved hydrophobic amino acid patch at the central domain involved in transcriptional activation. *Mol. Microbiol.* **28**:55–67.
- Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* **8**:11–22.
- Ho, Y. S., L. M. Burden, and J. H. Hurley. 2000. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.* **19**:5288–5299.
- Hoch, J. A. 2000. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**:165–170.
- Hopper, S., and A. Bock. 1995. Effector-mediated stimulation of ATPase activity by the sigma 54-dependent transcriptional activator FHLA from *Escherichia coli*. *J. Bacteriol.* **177**:2798–2803.
- Hubner, A., X. Yang, D. M. Nolen, T. G. Popova, F. C. Cabello, and M. V. Norgard. 2001. Expression of *Borrelia burgdorferi* OspC and DbpA is controlled by a RpoN-RpoS regulatory pathway. *Proc. Natl. Acad. Sci. USA* **98**:12724–12729.
- Hutchings, M. I., N. Mandhana, and S. Spiro. 2002. The NorR protein of *Escherichia coli* activates expression of the flavorubredoxin gene *norV* in response to reactive nitrogen species. *J. Bacteriol.* **184**:4640–4643.
- Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou, and J. Yu. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res.* **30**:4432–4441.
- Jyot, J., N. Dasgupta, and R. Ramphal. 2002. FleQ, the major flagellar gene regulator in *Pseudomonas aeruginosa*, binds to enhancer sites located either upstream or typically downstream of the RpoN binding site. *J. Bacteriol.* **184**:5251–5260.
- Kabisch, U. C. 2001. Ph.D. thesis. Proteinchemische und molekularbiologische Charakterisierung der D-Prolin-Reduktase aus *Clostridium sticklandii*. Martin-Luther-Universität Halle-Wittenberg, Halle, Germany.
- Kabisch, U. C., A. Grantzdorffer, A. Schierhorn, K. P. Rucknagel, J. R. Andreesen, and A. Pich. 1999. Identification of D-proline reductase from *Clostridium sticklandii* as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a propeptide. *J. Biol. Chem.* **274**:8445–8454.
- Kalman, S., W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis, and R. S. Stephens. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nat. Genet.* **21**:385–389.
- Kapatral, V., I. Anderson, N. Ivanova, G. Reznik, T. Los, A. Lykidis, A. Bhattacharyya, W. Bartman, G. Gardner, G. Grechkin, L. Zhu, O. Vasieva, L. Chu, Y. Kogan, O. Chaga, E. Goltsman, A. Bernal, N. Larsen, M. D'Souza, T. Walunas, G. Pusch, R. Haselkorn, M. Fonstein, N. Kyrpides, and R. Overbeek. 2002. Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586. *J. Bacteriol.* **184**:2005–2018.
- Kinsella, N., P. Guerry, J. Cooney, and T. J. Trust. 1997. The *flgE* gene of *Campylobacter coli* is under the control of the alternative sigma factor σ^{54} . *J. Bacteriol.* **179**:4647–4653.
- Little, R., V. Colombo, A. Leech, and R. Dixon. 2002. Direct interaction of the NifL regulatory protein with the GlnK signal transducer enables the *Azotobacter vinelandii* NifL-NifA regulatory system to respond to conditions replete for nitrogen. *J. Biol. Chem.* **277**:15472–15481.
- Little, R., S. Perry, V. Colombo, et al. 2002. Integration of nitrogen, oxygen and redox status by the *Azotobacter vinelandii* regulatory complex, p. 238–242. In T. Finan, M. O'Brian, D. Layzell, et al. (ed.), Nitrogen fixation: global perspectives. CABI International, Wallingford, United Kingdom.
- Little, R., F. Reyes-Ramirez, Y. Zhang, W. C. van Heeswijk, and R. Dixon. 2000. Signal transduction to the *Azotobacter vinelandii* NIFL-NIFA regulatory system is influenced directly by interaction with 2-oxoglutarate and the PII regulatory protein. *EMBO J.* **19**:6041–6050.
- Maloy, S. R., and W. D. Nunn. 1982. Genetic regulation of the glyoxylate shunt in *Escherichia coli* K-12. *J. Bacteriol.* **149**:173–180.
- Martin-Verstraete, L., V. Charrier, J. Stulke, A. Galinier, B. Erni, G. Rapoport, and J. Deutscher. 1998. Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Mol. Microbiol.* **28**:293–303.
- Mathews, S. A., and P. Timms. 2000. Identification and mapping of sigma-54 promoters in *Chlamydia trachomatis*. *J. Bacteriol.* **182**:6239–6242.
- Merrick, M. J. 1993. In a class of its own—the RNA polymerase sigma factor σ^{54} (σ^N). *Mol. Microbiol.* **10**:903–909.

43. Money, T., J. Barrett, R. Dixon, and S. Austin. 2001. Protein-protein interactions in the complex between the enhancer binding protein NIFA and the sensor NIFL from *Azotobacter vinelandii*. *J. Bacteriol.* **183**:1359–1368.
44. Morett, E., and L. Segovia. 1993. The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J. Bacteriol.* **175**:6067–6074.
45. Neuwald, A. F., L. Aravind, J. L. Spouge, and E. V. Koonin. 1999. AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**:27–43.
46. Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**:523–527.
47. Pelton, J. G., S. Kustu, and D. E. Wemmer. 1999. Solution structure of the DNA-binding domain of NtrC with three alanine substitutions. *J. Mol. Biol.* **292**:1095–1110.
48. Pittard, A. J., and B. E. Davidson. 1991. TyrR protein of *Escherichia coli* and its role as repressor and activator. *Mol. Microbiol.* **5**:1585–1592.
49. Ponting, C. P., and L. Aravind. 1997. PAS: a multifunctional domain family comes to light. *Curr. Biol.* **7**:R674–R677.
50. Postma, P. W., J. W. Lengeler, and G. R. Jacobson. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**:543–594.
51. Rabinovich, E., A. Kerem, K. U. Frohlich, N. Diamant, and S. Bar-Nun. 2002. AAA-ATPase p97/Cdc48p, a cytosolic chaperone required for endoplasmic reticulum-associated protein degradation. *Mol. Cell. Biol.* **22**:626–634.
52. Ray, P., K. J. Smith, R. A. Parslow, R. Dixon, and E. I. Hyde. 2002. Secondary structure and DNA binding by the C-terminal domain of the transcriptional activator NifA from *Klebsiella pneumoniae*. *Nucleic Acids Res.* **30**:3972–3980.
53. Reitzer, L., and B. L. Schneider. 2001. Metabolic context and possible physiological themes of σ^{54} -dependent genes in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **65**:422–444.
54. Reizer, J., and M. H. Saier, Jr. 1997. Modular multidomain phosphoryl transfer proteins of bacteria. *Curr. Opin. Struct. Biol.* **7**:407–415.
55. Reyes-Ramirez, F., R. Little, and R. Dixon. 2001. Role of *Escherichia coli* nitrogen regulatory genes in the nitrogen response of the *Azotobacter vinelandii* NifL-NifA complex. *J. Bacteriol.* **183**:3076–3082.
56. Rombel, I., A. North, I. Hwang, C. Wyman, and S. Kustu. 1998. The bacterial enhancer-binding protein NtrC as a molecular machine. *Cold Spring Harbor Symp. Quant. Biol.* **63**:157–166.
57. Rudnick, P., C. Kunz, M. K. Gunatilaka, E. R. Hines, and C. Kennedy. 2002. Role of GlnK in NifL-mediated regulation of NifA activity in *Azotobacter vinelandii*. *J. Bacteriol.* **184**:812–820.
58. Saier, M. H., Jr., and J. Reizer. 1994. The bacterial phosphotransferase system: new frontiers 30 years later. *Mol. Microbiol.* **13**:755–764.
59. Song, J., and R. A. Jensen. 1996. PhhR, a divergently transcribed activator of the phenylalanine hydroxylase gene cluster of *Pseudomonas aeruginosa*. *Mol. Microbiol.* **22**:497–507.
60. Souza, E. M., F. O. Pedrosa, M. Drummond, L. U. Rigo, and M. G. Yates. 1999. Control of *Herbaspirillum seropedicae* NifA activity by ammonium ions and oxygen. *J. Bacteriol.* **181**:681–684.
61. Spohn, G., and V. Scarlato. 1999. Motility of *Helicobacter pylori* is coordinately regulated by the transcriptional activator FlgR, an NtrC homolog. *J. Bacteriol.* **181**:593–599.
62. Stock, A. M., V. L. Robinson, and P. N. Goudreau. 2000. Two-component signal transduction. *Annu. Rev. Biochem.* **69**:183–215.
63. Studholme, D. J. 2002. Enhancer-dependent transcription in *Salmonella enterica* Typhimurium: new members of the σ^N regulon inferred from protein sequence homology and predicted promoter sites. *J. Mol. Microbiol. Biotechnol.* **4**:367–374.
64. Studholme, D. J., and M. Buck. 2000. The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol. Lett.* **186**:1–9.
65. Studholme, D. J., S. R. Wigneshweraraj, M. T. Gallegos, and M. Buck. 2000. Functionality of purified σ^N (σ^{54}) and a NifA-like protein from the hyperthermophile *Aquifex aeolicus*. *J. Bacteriol.* **182**:1616–1623.
66. Taylor, B. L., and I. B. Zhulin. 1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **6**:479–506.
67. Tchieu, J. H., V. Norris, J. S. Edwards, and M. H. Saier, Jr. 2001. The complete phosphotransferase system in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* **3**:329–346.
68. Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, J. C. Venter, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539–547.
69. van Tilbeurgh, H., and N. Declerck. 2001. Structural insights into the regulation of bacterial signalling proteins containing PRDs. *Curr. Opin. Struct. Biol.* **11**:685–693.
70. Wang, Y.-K., J. Lee, J. Brewer, and T. Hoover. 1997. A conserved region in the σ^{54} -dependent activator DctD is involved in both binding to RNA polymerase and coupling ATP hydrolysis to activation. *Mol. Microbiol.* **26**:373–386.
71. Wilson, T. J., P. Maroudas, G. J. Howlett, and B. E. Davidson. 1994. Ligand-induced self-association of the *Escherichia coli* regulatory protein TyrR. *J. Mol. Biol.* **238**:309–318.
72. Xu, H., and T. R. Hoover. 2001. Transcriptional regulation at a distance in bacteria. *Curr. Opin. Microbiol.* **4**:138–144.
73. Zhang, X., M. Chaney, S. R. Wigneshweraraj, J. Schumacher, P. Bordes, W. Cannon, and M. Buck. 2002. Mechanochemical ATPases and transcriptional activation. *Mol. Microbiol.* **45**:895–903.
74. Zhulin, I. B., B. L. Taylor, and R. Dixon. 1997. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.* **22**:331–333.