Bacterial Transcriptional Regulators for Degradation Pathways of Aromatic Compounds

David Tropel¹ and Jan Roelof van der Meer¹,²*

Swiss Federal Institute for Environmental Science and Technology (EA&WAG), Dübendorf,¹ and Department of Fundamental Microbiology, University of Lausanne, Lausanne,² Switzerland

INTRODUCTION ................................................................. 474

LysR FAMILY OF TRANSCRIPTIONAL REGULATORS ................................................................. 475
Structure and Conformation ................................................................. 476
Mechanisms of Activation ................................................................. 478

IdiR FAMILY OF TRANSCRIPTIONAL REGULATORS ................................................................. 480
Catabolic Operons Controlled by IdiR Regulators ................................................................. 480
Structure and Conformation ................................................................. 481
Mechanism of Transcription Activation ......................................................... 481

AraC/XylS FAMILY ................................................................. 482
Catabolic Operons Controlled by AraC/XylS-Type Regulators ................................................................. 482
Structure and Conformation ................................................................. 483
Mechanisms of Activation ................................................................. 483

GntR-TYPE REGULATORS ................................................................. 484
Catabolic Operons Controlled by GntR-Type Regulators ................................................................. 484
Structure and Conformation ................................................................. 485
Mechanisms of Regulation ................................................................. 486

TetR-, MarR-, AND FNR-TYPE REGULATORS ................................................................. 486
TetR Family ................................................................. 486
MarR-Type Regulators ................................................................. 487
FNR-Type Regulators ................................................................. 488

TWO-COMPONENT REGULATORY SYSTEMS ................................................................. 488
Involvement in Regulation of Catabolic Pathways ................................................................. 489
Structure and Conformation of the Sensor Kinase Component ................................................................. 489
Structure and Conformation of the Response Regulators ................................................................. 490

XylR/NtrC-TYPE TRANSCRIPTIONAL REGULATORS INVOLVED IN CATABOLIC PATHWAYS ................................................................. 490
Catabolic Operons Controlled by XylR/DmpR Subclass Regulators ................................................................. 490
Domain Organization ................................................................. 491
Structure and Configurations ................................................................. 492
DNA Binding during the Activation Process ................................................................. 493

PROSPECTS ................................................................. 494
ACKNOWLEDGMENTS ................................................................. 494
REFERENCES ................................................................. 494

INTRODUCTION

Human activities have resulted in the release and introduction into the environment of a plethora of aromatic chemicals. Although the input of these synthetic chemicals may be smaller than the total amount of aromatic compounds released from decaying plant material, their novel structures or their quantities as single pure molecules induce major changes in microbial communities, which are the major recyclers of organic chemicals in nature (76). For example, large quantities of a single organic solvent offer a very specific metabolizable carbon source for specialized groups of microorganisms (161) and thus can act selectively for enrichment of those groups in the environment (87, 154). Compounds such as chlorinated solvents, herbicides, and pesticides often carry uncommon chemical structures, side chains, or functional groups, which can have toxic effects or provide new carbon sources for bacteria which have adapted their metabolism to degrade the compounds (191, 207, 261). The interest in discovering how bacteria are dealing with hazardous environmental pollutants has driven a large research community and has resulted in important biochemical, genetic, and physiological knowledge about the degradation capacities of microorganisms (40, 43, 76, 146, 191, 206, 207). A large variety of metabolic pathways have been discovered in very different microorganisms, fueling up-to-date online databases such as the Biocatalysis/Biodegradation Database (57). Knowledge about the biodegradation capacities of microorganisms is being applied directly in bioremediation practice (240), and individual biotransformation reactions are potentially very interesting for incorporation into chemical synthesis (230).
Less obviously associated with bioremediation applications, green chemistry, or production of pharmacy synthons have been studies of the regulatory mechanisms, which govern the expression of specific catabolic pathways (105). However, regulation of catabolic pathway expression has attracted the interest of numerous different groups, who have tried to unravel the molecular partners in the regulatory process, the signals triggering pathway expression, and the exact mechanisms of activation and repression. More recently, information about regulatory systems has attracted interest, with the potential of these systems being used as sensory mechanisms in whole-cell living bioreporters, genetically modified bacteria which can be used as sensors to measure the quality of aqueous, soil, and air environments (49, 96, 242). It soon was discovered that a large diversity of regulatory systems existed for mediating the expression of catabolic pathways. Furthermore, many related catabolic pathways did not carry the same regulatory system, which led to the hypothesis that regulatory systems and their target operons do not necessarily coevolve but seem to become associated independently (23, 39). As genomic, genetic, and biochemical data accumulated, regulatory proteins for catabolic gene expression were classified into other protein families, showing that they (as expected) were not unique to typical biodegradation pathways per se. In fact, many aromatic degradation pathways, typically those involved in the metabolism of aromatic amino acids, hydroxylated benzoates and phenylpropionic acids, ubiquinones, and aromatic amines, are very widespread among microbial species (43), leaving perhaps only the catabolic pathways for toxic and xenobiotic compounds to the groups of microorganisms usually considered for biodegradation (i.e., pseudomonads, sphingomonads, *Rhodococcus* spp., *Ralstonia* spp., and *Burkholderia* spp.). One would thus have to conclude that the capabilities of regulatory proteins to react specifically to substrates or intermediates of catabolic pathways have evolved many times independently; this is explained in more detail below.

Although excellent reviews exist on the diversity of catabolic pathways (76, 206), on evolutionary aspects (99, 262, 271), and on different details of specific regulator families (71, 139, 224, 233), no concise overview of the different regulatory systems involved in biodegradation pathways (or catabolic pathways, as we will refer to them often) exists in the literature, although a shorter treatise was published by Diaz and Priets (44). In our opinion, the topic of biodegradation is relatively coherent, and thus treatment of the different regulatory systems involved in catabolic pathways also deserves a coherent overview—although it is clear that there are no special “catabolic” regulatory proteins. In addition, several three-dimensional structures of regulatory proteins have recently been resolved, which allows a much more detailed prediction of how the actual activation and repression mechanisms of regulatory proteins take place. This might solve many questions which have arisen from more traditional genetic and biochemical approaches to transcription activation and repression. The current view of regulatory proteins is that of a set of sophisticated protein-DNA-RNA polymerase interaction machineries with sufficient diversity to promote or inhibit DNA transcription by RNA polymerase. What remains largely elusive, however, is how the machinery is turning.

The structure of this review is relatively straightforward. The various known regulatory families are treated individually, as far as possible with examples of catabolic pathways. Specific attention has been given to describing the genetic organization of the regulatory genes, promoters, and target operon(s); this is followed by a discussion integrating the knowledge of signaling molecules, DNA binding properties and operator characteristics, regulator-DNA structure, and evidence from regulatory mutants to arrive at a possible mechanism of transcription activation. What is not specifically treated in this review is how regulatory systems themselves are embedded in the host’s physiology and controlled by other global regulatory systems. For information about these aspects, readers are referred to other specific overviews, such as references 23, 232, and 265. For sake of the overview, we have refrained from too many in-depth details of very specific individual transcription regulators.

### LysR FAMILY OF TRANSCRIPTIONAL REGULATORS

**Catabolic Operons Controlled by LysR-Type Regulators**

LysR-type transcriptional regulators (LTTTRs) comprise the largest family of prokaryotic regulatory proteins identified so far (84, 224). The family has expanded to over 100 members that have been identified in diverse bacterial genera. This diversity is also reflected in LTTTRs associated with degradation pathways of aromatic compounds (Table 1) (44). A large group of LTTTRs regulates a single target operon only, such as CatR controlling *catBCA* expression for catechol metabolism in *Pseudomonas putida* (215). CICR controlling the *cicABDE* operon of plasmid pAC27 (33). TcbR controlling the *tcbCDEF* operon in plasmid pP51 of *Pseudomonas* sp. strain P51 (263), and CbnR controlling the *cbnABC* operon for chlorocatechol metabolism in *Ralstonia eutropha* (163). TidR (and/or its identical twin TidS) also regulate an operon coding for chlorocatechol metabolism (i.e., the *tfdDCEF* genes), but the same regulators also control the expression of two other operons, *tfdB* and *tfdG*, all of which are involved in degradation of 2,4-dichlorophenoxyacetate in *R. eutropha* JMP134 (and are located on plasmid pJP4 [Fig. 1A]) (127). Similarly, CatR of *P. putida* can coregulate the expression of the *pheBA* operon for phenol degradation and of the *catBCA* genes when this operon is provided on an additional plasmid (104). Like TidR and TidS, the NahR protein is a master regulator for the regulon of naphthalene degradation and acts by controlling expression from both the *nah* operon, required for the metabolism of naphthalene to salicylate and pyruvate, and the *sal* operon, encoding the enzymes for salicylate conversion (225, 228, 274). In this case, both operons are present on the NAH7 plasmid. Two LTTTRs have been implicated in the transcriptional control of benzoate degradation in *Acinetobacter* sp. strain ADP1. In this microbe, BenM regulates expression from the *benABCDE* and *benPK* operons, which encode the enzymes for conversion of benzoate to catechol. CatM also regulates expression from *benPK* but in addition regulates expression from the *catBCIJFD* operon, leading to further metabolism of *cis,cis*-muconate to the tricarboxylic acid cycle (Fig. 1B). The *catA* gene, encoding catechol 1,2-dioxygenase, which converts catechol to *cis,cis*-muconate, is not present in the two operons but is also regulated by both CatM and BenM (34, 212). CatM...
**Table 1. Regulatory proteins from the LysR family involved in expression control of pathways for degradation of aromatic compounds**

<table>
<thead>
<tr>
<th>Bacterial genus</th>
<th>Pathway substrate</th>
<th>Regulated operon(s)</th>
<th>RBS</th>
<th>Accession no. (reference)</th>
<th>Position of RBS</th>
<th>Inducers accessions RBS</th>
<th>Regulatory protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatR</td>
<td>Catechol (phenol)</td>
<td>AGACC-N1-5-GGTAT</td>
<td>/H11002</td>
<td>A35118 (199)</td>
<td>79 to 26</td>
<td>2-Chloro-cis-muconate, cis-Muconate</td>
<td>CatR</td>
</tr>
<tr>
<td>ClcR</td>
<td>3-Chlorocatechol</td>
<td>ATAC-N7-GTAT</td>
<td>/H11002</td>
<td>A40641</td>
<td>79 to 53</td>
<td>2-Chloro-cis-muconate, cis-Muconate</td>
<td>ClcR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudomonas tcbCDEF</td>
<td>Benzoate (catechol)</td>
<td>ATAC-N7-GTAT</td>
<td>/H11002</td>
<td>A38861 (179)</td>
<td>78 to 50</td>
<td>Benzoate and cis-Muconate</td>
<td>TfdR/S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ralstonia cbnABCD</td>
<td>3-Chlorocatechol</td>
<td>TTACG-N2-CGTAA</td>
<td>/H11002</td>
<td>A40641</td>
<td>80 to 50</td>
<td>2-Chloro-cis-Muconate, cis-Muconate</td>
<td>cbnABCD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TfdR/S</td>
<td>4-Chlorocatechol</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td>3- or 4-Chlorocatechol, 2- or 3-chlorobenzoate</td>
<td>TfdR/S</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ralstonia tfdC</td>
<td>3-Chlorocatechol</td>
<td>ATAC-N7-GTAT</td>
<td>/H11002</td>
<td>A40641</td>
<td>72 to 50</td>
<td>2-Chloro-cis-Muconate, cis-Muconate</td>
<td>tfdC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tfdC</td>
</tr>
<tr>
<td>TfdT</td>
<td>4-Chlorocatechol</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td>3- or 4-Chlorocatechol, 2- or 3-chlorobenzoate</td>
<td>tfdT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tfdT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>tfdT</td>
</tr>
</tbody>
</table>

and BenM also interfere with expression of the pobA gene for 4-hydroxybenzoate metabolism, perhaps by preventing the expression of the pcaK uptake system (16). Although LTTRs are often associated with ortho-cleavage pathways of catechol, they are also involved in a large variety of other degradation pathways (Table 1). For example, NtdR controls the genes encoding nitroarene dioxygenase in *Acidovorax* sp. strain JS42 (125); the TsaR protein controls the expression of the tsaMBCD genes, which encode the first steps in the degradation of p-toluene sulfonate in *Comamonas testosteroni* T-2 (253); and BphR2 is involved in the regulation of a biphenyl catabolic gene cluster of *Pseudomonas pseudoalcaligenes* KF707 (268). Preliminary data for regulation of the biphenyl pathway indicate that BphR2 is activating the expression of a first operon [bpbaA2(orf3)A3A4] whereas a second regulatory gene (bphR1, a regulator of the GntR family) is activating the expression of the second operon (bpbaX2X3D). LTTRs have also been identified in y-hexachlorocyclohexane degradation in *Sphingomonas paucimobilis* (148), in aniline degradation in *P. putida* (67), and in catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12 (42). A longer list of LTTRs involved in transcriptional control of aromatic compound degradation is presented in reference 44.

In general, the gene for LTTRs lies upstream of its target-regulated gene cluster and is transcribed in the opposite direction (Fig. 1A and B), but exceptions to this rule exist. For example, the gene for the SalR regulator in *Acinetobacter baumannii* strain ADP1 is present within the transcriptional unit salRA. The salA gene is coding for salicylate hydroxylase, which converts salicylate to catechol. Experimental evidence suggests that SalR is regulating the expression of itself and of salA (101). The phrS gene (coding for a further uncharacterized LTTR) is cotranscribed as the first gene of an operon including the phrFEDAL catabolic genes for naphthalene and phenanthrene degradation in *Burkholderia* sp. strain RP007 (119).

All identified LTTRs involved in aromatic degradation pathways act as transcriptional activators for their target metabolic operons in the presence of a chemical inducer, which is usually a pathway intermediate (33, 128, 212, 215, 227). In some cases, the effectors are (substituted) muconates which have lost their aromatic character, whereas in other systems aromatic compounds act as effectors (e.g., salicylate, catechol, and nitrilotoluene [Table 1]). There is also experimental evidence for competition of several compounds on the regulatory protein. For example, fumarate reversibly inhibits the formation of the clcA transcript in vitro transcription assays in the presence of purified ClcR and 2-chloro-cis,cis-muconate (138). The presence of cis,cis-muconate decreases the affinity of the BenM protein for benzoate (31). All LTTRs repress their own expression, and both autorepression and activation of the catabolic operon promoter are exerted from the same binding site, which is called the regulator or repressor binding site (RBS) (20, 216, 227). Autorepression was not influenced by the presence of an inducer in the case of BenM (20), but expression of the clcR and tcbR promoters was slightly enhanced in the presence of an inducer (33, 263). Relatively few data exist on autorepression mechanisms, since most studies on LTTRs have focused on the mechanisms of target gene activation.

*ND, not determined.*
FIG. 1. Typical examples of the genetic organization of pathways for degradation of aromatic compounds. The regulatory genes are indicated in grey. (A) The tfd operons of R. eutropha JMP134(pJP4). The LysR-type transcription regulator TfdR/S regulates the tfdA, tfdC, and tfdDII promoters. (B) The ben and cat operons of Acinetobacter sp. strain ADP1. benM and catM code for LysR-type transcription regulators, which act on four promoters: benP, benA, catA, and catB. (C) The pca and pob clusters of Acinetobacter sp. strain ADP1. The genes pcaU and pobR code for IclR-type regulators, acting on the pcaI and pobA promoters. (D) The hpa cluster of E. coli with HpaA as the XylS/AraC-type regulator. (E) The xyl operons for toluene degradation present on the TOL plasmid of P. putida mt2. XylR is the NtrC-type regulator acting on Pr and Ps, whereas XylS is the exemplar for the XylS/AraC family and acts on the Pm promoter. (F) The paa cluster in E. coli strain W, with the GntR-type regulator PaaX regulating the paaA promoter. (G) The aph cluster in C. testosteroni strain TA441, encoding the aphS (GntR-type) and aphR (XylR-type) regulators. (H) The cym and cmt clusters in P. putida F1 with the TetR-type regulator cymR. (I) The tod cluster of P. putida DOT-T1 with the two-component regulatory system todST. Straight arrows represent transcripts produced after specific regulatory action of the indicated regulator. Parabolic arrows point to the site of action of a specific regulator in cases when multiple regulators are involved. Hooked arrows indicate specific promoter names, where necessary.
Structure and Conformation

LTTRs involved in the degradation of aromatic compounds are composed of 394 to 403 amino acid residues with a molecular mass of between 32 and 37 kDa. All the evidence presented so far points to tetramers being the active form of LTTRs. ClcR and CatR have been identified as dimers in solution (32, 181), but two dimers are needed to bind DNA (141). CatM, BenM, and CbnR were found to form tetramers (20, 30, 155), and NahR probably also acts as a tetrameric form (226). The well-studied LTTR CysB, which is not involved in the activation of pathways for aromatic-compound degradation, also forms tetramers in solution (88, 145). Only TsaR remains as monomer in solution (253). LTTRs have a conserved domain organization, which has been determined from mutagenesis studies and sequence alignments (224). A DNA binding region with a predicted helix-turn-helix (HTH) motif is located in the 66 N-terminal amino acid residues of the protein, two regions located between residues 95 to 173 and residues 196 to 206 are involved in inducer recognition, and one region between residues 227 and 253 is supposed to be involved in multimerization (Fig. 2A).

Recently, the first complete LTTR has been crystallized (CbnR) (155). Crystals of BenM and CatM devoid of their N-terminal DNA binding domain have also been obtained, but their structures have not been resolved yet (29). CbnR crystallized as a tetramer (Fig. 3) with two main parts, the four DNA binding domains (residues 1 to 58) and a central body (155). The four DNA binding domains have no interactions with each other, whereas the central body of the tetramer is composed of four intertwined regulatory domains (residues 88 to 294 of each subunit). The tetramer structure can be regarded as a dimer of a dimer, whereby each dimer is composed of two subunits in different configurations. The two subunits in each dimer are connected through the coiled-coil linker (residues 59 to 87), which at the same time separates the DNA binding domain and the central body (Fig. 3). Subunit A of the AB dimer interacts with the regulatory domains of both subunits Q and P of the other (PQ) dimer, whereas subunit B interacts only with P (Fig. 3). The structure seems to allow easy transmission of conformational changes, for example in the central body, the DNA binding domains. The DNA binding motifs in the tetramer are presented in such a way as to form a V-shaped structure, which matches exactly with the distance and configuration of the two DNA binding sites (Fig. 3). Although the mode of action and the location of the effector binding pockets still need better definition, this structural model beautifully fits and explains previous experimental evidence with the LTTRs.

FIG. 2. Schematic domain organization of the different regulator families. The domain containing the HTH DNA binding motif is indicated in black. The domain bound by the chemical inducer is indicated in white.

FIG. 3. Tetrameric structure of CbnR in side (A) and top (B) views (155). Subunits A, B, P, and Q are shown in yellow, magenta, cyan, and green, respectively. The symmetry axis in the molecule is shown as a red arrow. The location of the coiled-coil linkers (see the text) and DNA binding domains (DBD) are indicated, which would impose a 60°C bending angle on the binding site on the DNA. Reprinted from reference 155 with permission from the publisher and from the authors.
Mechanisms of Activation

Due to their tetrameric form, LTTRs interact with several sites on the DNA of the promoter region. Classically, DNA interactions are shown by DNase I footprinting techniques and purified regulatory protein. Interestingly, DNase I cleavage patterns of the protein-bound nucleotide regions were similar for ClcR, CatR, TcbR, CbnR, and, to a lesser extent, NahR. The different interactions are explained in more detail for ClcR. ClcR binds the clcA promoter irrespective of the presence of inducer (32, 141, 182). However, in the absence of inducer, ClcR protects a 27-bp region (RBS) from –37 to –41 (141), suggesting that the bending angle is relaxed on interaction with the effector. Conformational changes on effector binding could be detected for benzoate and cis,cis-muconate binding to BenM (31). Measurements of the bending angle of the clcA promoter in the absence (71°) and in the presence (55°) of effector corroborate this idea of bending relaxation (140). Similar bending angles were reported for CbnR at the cbnA promoter in the absence and presence of inducer, although for CbnR no changes in DNase I footprints were observed as for ClcR (163). The role of the RBS and ABS is not completely clarified. Although ClcR and CbnR contacted both the RBS and ABS in the absence of inducer, CatR did not (215, 216). It was concluded that the regulatory proteins have a higher binding affinity to the RBS than to the ABS, since a fragment with only the ABS is not bound by ClcR or CatR (139, 181). However, contacts to the ABS are supposed to be necessary for mediating interactions with RNA polymerase (RNAP), and fragments with only the RBS do not lead to in vitro transcript production in the presence of ClcR, RNAP, and inducer (139).

Further contacts seem to occur between the regulatory protein complex and the C-terminal domain of the α-subunit (α-CTD) of RNAP. This was concluded from the lack of mRNA synthesis in in vitro transcription experiments with CatR (or ClcR), the catB (clcA) promoter, cis,cis-muconate, and RNAP devoid of the α-CTD (28, 139). A direct interaction of NahR with the α-CTD of RNAP has been demonstrated by using a yeast two-hybrid system (178), and this interaction was not influenced by the presence of salicylate. Chugani et al. (28) postulated that the regulatory complex was directing the α-CTD of RNAP to a region between the RBS and the ABS, which they called the UP-motif (Fig. 4). Interaction with the UP-motif would increase the affinity of RNAP to the promoter (28). However, exactly how this process would lead to transcription activation is not known. The degree of DNA supercoiling has also been implicated in controlling the level of transcription activation from LTTR-dependent promoters. This has been concluded from in vitro transcription experiments with CatR, in which supercoiled templates produced no mRNA transcript except in the presence of cis,cis-muconate whereas linearized templates did (28). Perhaps the regulatory protein can overcome the torsional constraint on the DNA in the presence of inducer, thereby facilitating open-complex formation by RNAP.

In some cases, a third DNA binding site further downstream of the transcriptional start site has been identified. For example, CatR binds with low affinity to a region between +162 and +196 of the catB gene (27). This site has been named the inhibitor binding site (IBS) and is supposed to prevent excessively high expression from the catBCA operon by titrating CatR protein. The affinity of binding of CatR to the IBS increased at higher inducer concentrations, and catB promoter fragments containing the IBS had a three- to fourfold lower expression than those without IBS (27). IBS regions have also been identified in the pheBA operon and clcABD operon (27).

BenM has the unique peculiarity (for the moment) that the inducers benzoate and cis,cis-muconate can have synergistic
effects on the activation process compared to the effects of each inducer alone (20, 31, 34). In addition, BenM represses benA transcription in the absence of inducer. Benzoate and cis,cis-muconate alone had very different effects on the patterns observed in DNase I footprints of BenM on the benA promoter than did both inducers simultaneously. In the absence of inducer, BenM bound two areas with dyad symmetry, site 1 (ATAC-N7-GTAT) at positions $\text{H11002}^57$ to $\text{H11002}^71$ and site 3 (ATTC-N7-GTAT) at positions $\text{H11002}^5$ to $\text{H11002}^19$. Several hypersensitive sites were detected (at positions $\text{H11002}^50$, $\text{H11002}^45$, $\text{H11002}^39$, $\text{H11002}^36$, $\text{H11002}^34$, $\text{H11002}^29$, and $\text{H11002}^24$), suggesting again a clear bending of the DNA imposed by the BenM regulatory complex. In the presence of cis,cis-muconate or benzoate, BenM still protected site 1 but no longer protected site 3. When both inducers were present simultaneously, the number of hypersensitive sites was reduced and instead a region called site 2 (which also contained a dyad symmetry motif, ATAC-N7-GTGT, located between $\text{H11002}^36$ and $\text{H11002}^50$) was protected from DNase I cleavage (20). The authors hypothesized that BenM binding to site 3 (which overlaps with the $\text{H11002}^10$ region) causes the observed repressive effect in the absence of inducer. In the presence of inducer, is BenM released from site 3, enabling RNAP to access the promoter, whereas BenM binding to site 2 would lead to more productive contacts to RNAP and a higher transcriptional output.

IclR FAMILY OF TRANSCRIPTIONAL REGULATORS

Catabolic Operons Controlled by IclR Regulators

IclR-type regulators have a similar structure as the LysR-type regulators (224), but rather dissimilar amino acid sequences distinguish the two families. IclR-type regulators are generally transcriptional repressors (86, 132, 156, 245); however, those which control catabolic pathways have all been described as activators (Table 2). For example, PcaU of Acinetobacter sp. strain ADP1 (75), PcaR of Pseudomonas putida (213), PcaR of Agrobacterium tumefaciens, and CatR and PcaR of Rhodococcus opacus 1CP (58, 59) are activators for the ortho-cleavage pathways which they regulate. However, in the absence of inducer they may still act as repressors, as was shown for PcaU (254). Further IclR members are MhpR, which activates the meta-cleavage pathway in 3-(3-hydroxyphenyl)propionic acid degradation by E. coli (61); PobR, which is the activator for the 4-hydroxybenzoate degradation pathway in Acinetobacter sp. strain ADP1 (45), and OhbR, controlling the genes for the oxygenolytic ortho dehalogenation of halobenzoates (256).

In general, the gene for the IclR-type regulator lies upstream of its target gene cluster and is transcribed in the opposite direction (Fig. 1C). An exception is the pcaR gene of P. putida (160). PcaR actually controls four distinct gene clusters required for the degradation of protocatechuic acid to tricarboxylic acid cycle intermediates and of the 4-hydroxybenzoate transporter PcaK (160). The pcaR gene lies upstream of pcaK and is transcribed in the same direction (90, 159, 213). PcaR from P. putida and PobR and PcaU from Acinetobacter sp. strain ADP1 repress their own expression (46, 81, 254). It is thought that the mechanism of autorepression is different among IclR-type regulators, since not all of them bind at the

<table>
<thead>
<tr>
<th>TABLE 2. Major IclR family members regulating degradation pathways of aromatic compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regulatory protein</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>PcaU</td>
</tr>
<tr>
<td>PobR</td>
</tr>
<tr>
<td>PcaR</td>
</tr>
<tr>
<td>MhpR</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Position relative to the transcription start site.
same position on the promoter region (75, 81, 213) and since for some of them the addition of effectors changes the expression of the regulatory gene itself. For example, expression of pcaR and pobR is independent of the usual effectors for PcaR and PobR (46, 81) but pcaU expression increases in the presence of effectors for PcaU (75).

**Structure and Conformation**

The size of IclR-type regulators is around 238 to 280 amino acid residues (25 to 30 kDa) (59, 61, 82, 112, 256). IclR-family members have an HTH DNA binding motif in the N-terminal domain (45, 61) and a C-terminal domain involved in subunit multimerization and effector binding (112). This was confirmed by the crystal structure of IclR from *Thermotoga maritima* 0065 (Fig. 5) (275), which showed that the amino acid residues previously identified to be involved in PobR inducer specificity (112) are present in IclR and comprise a pocket in the C-terminal domain (275). Although PcaU and PcaR from *P. putida* formed dimers in solution (81, 193), IclR crystallized as a dimer of a dimer with an asymmetric configuration (275).

The two subunits within one IclR dimer interact solely at the interface of their DNA binding domains. As a consequence, the distance between the HTH motifs within one dimer is relatively short and results in a structure favorable for binding relatively short (12- to 14-bp) palindromic DNA sequences with specific contacts predominantly in the major groove of the DNA. The C-terminal domains do not contact each other in the dimer, but they do bridge with the C-terminal domains from the neighboring dimer, which is oriented in asymmetric fashion (Fig. 5). It is not really clear how the tetramer interacts with the DNA. Mass spectrometric data revealed that four IclR subunits are present per DNA containing one single palindrome (see below). The presumed ligand binding region is close to the region involved in tetramerization, suggesting that ligand binding and tetramerization may be linked. Moreover, the tetramer seems to be the active DNA binding form since, in mass spectrometric studies, four IclR subunits bound one synthetic DNA containing one single palindrome. The stoichiometry of two protein subunits for one DNA molecule was never detected (48).

**Mechanism of Transcription Activation**

There is no clear consensus on the binding site for IclR members (Table 2). For example, the MhpR binding site is formed by one 15-bp palindrome that lies 50 to 66 nucleotides upstream of the *mhpA* transcription start site (252), but those of PcaU and PobR are three perfect 10-bp sequence repetitions that lie between 53 and 94 bp upstream of the *pobA* and *pcaI* transcription start sites, respectively (46, 75, 193). Two of the 10-bp sequence repetitions form one palindrome, and the third repeat is oriented opposite to the second one but separated by another 10 bp (193). In contrast to those, the PcaR binding site is formed by a series of 15 nucleotides present twice in the *pcaI* promoter and overlapping with the −10/−35 promoter region (81). One might thus conclude that when the regulators conform to their binding sites, they must be different between MhpR, PobR, PcaU, and PcaR.

IclR-type regulators bind their promoter DNA in the absence of effector, and adding effector molecules had no effect on the affinity of the protein-DNA interactions displayed by purified PobR, PcaU, PcaR, and MhpR (46, 75, 81, 252). When, however, the chemical effector was added to a mixture of regulator (in this case PcaR), purified σ^70^-RNAP, and a *pcaI* promoter fragment, the formation of a PcaR-RNAP-DNA complex was enhanced compared to the situation without effector (81). The authors suggested that the role of the regulatory protein might be to favor the recruitment of RNAP to the promoter, perhaps by optimizing the critical distance between the −35 and −10 elements in the *pcaI* promoter from 16 to 17 bp (81). The same effect of enhanced complex formation was also found for SoxR and MerR (6, 85). Also, in the *mhpA* promoter the −35 and −10 elements are separated by 16 bp (252). However, the *pobA* promoter (regulated by PobR) already has a spacing of 17 bp (75), and thus the generality of this promoter distance optimization as an activation mechanism for IclR-type regulators is debatable.
AraC/XylS FAMILY

Catabolic Operons Controlled by AraC/XylS-Type Regulators

For years the XylS protein was the only member of the AraC family which was involved in expression control of a catabolic operon, namely, the meta-cleavage pathway operon for degradation of meta-toluic acid located on the TOL plasmid of *P. putida* (92). Recent alignment studies have shown that more than 300 proteins, some of which may be involved in the control of catabolic pathways, contained a typical stretch in the C-terminal part of about 100 amino acids that would classify them as AraC/XylS-type regulators (251). The remaining part of the proteins can be very different, though (see below). AraC/XylS-type regulators for catabolic operons generally act as transcription activators in the presence of a chemical effector molecule (Table 3). Some of the more recently discovered catabolic gene regulators of the AraC/XylS family are PobC of *P. putida* WCS358 (13), PobR of *Azotobacter chroococcum* (198), and PobR of *Pseudomonas* sp. strain HR199 (173). PobC and PobR regulate the expression of a *pobA* gene for *p*-hydroxybenzoate hydroxylase. HpaA is a XylS-type protein found in *E. coli* and regulates the expression of the hpaCB genes for *p*-hydroxyphenylacetate hydroxylase (195). Other family members identified in degradation pathways are CbdS, controlling the 2-halobenzoate dioxygenase genes in *Burkholderia* sp. strain TH2 (246); BenR, controlling the benzoate 1,2-dioxygenase operon in *P. putida* (36); CadR, controlling the genes for 2,4-dichlorophenoxyacetic acid degradation in *Bradyrhizobium* sp. strain HW13 (109); and IpbR, which regulates the expression of isopropylbenzene degradation in *P. putida* RE204 (51).

Most of the genes for XylS-type regulators lie upstream of their target operon, but, in contrast to lysR-type genes, they are transcribed in the same direction as the target genes. For example, *cbdS, hpaA, cadR, benR,* and *phcT* are all transcribed in the same direction as the genes they are regulating (36, 109, 196, 246). The *pobR* and *pobC* genes, on the other hand, are transcribed in the opposite direction (13, 173, 198). The *xylS* gene itself is also transcribed in the opposite direction to the *meta*-cleavage pathway operon (Fig. 1E) but is located downstream of it (93).

Expression of the genes for XylS-type regulators is not controlled just by themselves but in many cases by other activators or cascades. The best-studied example is XylS. Expression of *xylS* is strongly dependent on another regulatory protein XylR (see further below). XylR stimulates *xylS*-transcription from a ςx-dependent promoter (called P₉₂) when cells are grown on xylenes (69, 134). In the absence of suitable aromatic inducers, the *xylS* gene is expressed at low constitutive levels from a ςz-dependent promoter called P₉₁ (69). The *hpaA* regulatory gene has the unusual feature of being cotranscribed with the *hpaX* gene, which lies directly upstream of *hpaA* (194, 195). Expression of *hpaA* takes place from two promoters, one located directly upstream in the coding region of *hpaX* and the second located upstream of *hpaX*, in which case a bicistronic transcript is produced (195). Another example is provided by the *phcT* regulatory gene. PhcT is at the top of a regulatory cascade controlling phenol degradation (250). PhcT enhances transcriptional activation of the *phc* operon for phenol degr-
dation by interacting with another regulator, PhcR, but at the same time it represses expression of the phcR gene itself. The dual action of PhcT is possible because its binding site is located in the intergenic region between phcR and the phc phenol operon, which themselves are oriented divergently (250).

Structure and Conformation

XylS/AraC-type regulators involved in degradation control of aromatic compounds are typically between 293 and 322 amino acids in size, with a molecular mass around 35 kDa. One exception is PhcT, which has only 257 amino acid residues with a predicted molecular mass of 28 kDa. As far as has been determined, most AraC members form dimers in solution (21, 24). The solution state of XylS has not been determined directly, since the protein is very difficult to purify. Ruiz et al. found that a peptide containing only the N-terminal end of XylS was able to dimerize (217). The C-terminal end of XylS/AraC-type regulators is the most highly conserved part among this protein family and contains two HTH motifs within a region of 100 or so amino acid residues (Fig. 2C) (71). In XylS the two HTH motifs comprise residues 231 to 252 and 282 to 305 (133), and a truncated XylS polypeptide with only the 112 C-terminal residues was capable of binding and activating transcription from the Pm promoter (103). Two AraC family members (i.e., Rob and MarA) have recently been crystalized (114, 210) (Fig. 6), but they do not match the most common AraC architecture, since MarA is a single-domain protein (with only a DNA binding domain) and Rob has the DNA binding domain located at the N terminus of the protein (52). MarA and Rob were crystallized as one monomer bound to a DNA fragment containing one binding site, CCAC-N7-TAA and GCAC-N7-CAA, respectively (Fig. 6). The two HTH motifs of MarA bind to adjacent segments of the major groove, with the helical axes of the recognition helices almost parallel to the DNA base pairs (Fig. 6). Because the recognition helices in the HTH motifs protrude from the same face of the protein, MarA binds to one face of the DNA. The two HTH motifs are connected by a rigid central linker helix that imposes their orientation and distance on the DNA binding site. Binding results in a bending of the DNA at an angle of 35°, because the two recognition helices are separated by 27 Å whereas the pitch of the DNA B-form is 34 Å (210). The structure of the Rob DNA binding domain is quite similar to that of MarA, but two main differences were observed in the protein-DNA complex (Fig. 6). First, the DNA was not bent in the structure obtained with Rob, and second, only one protein helix was contacting the major groove of the DNA (114). In addition, nonspecifically attached Rob subunits were present in the crystallographic structure on the other side of the target DNA, which could be interpreted as an artifact (114). Martin and Rosner suggested that the presence of the second monomer prevented bending of the DNA and hence interaction of the second protein helix to the major groove of the DNA (137). On the other hand, it has been suggested that the Rob structure represents the state required for transcription activation, since conserved residues of one protein helix in the HTH-motif are not contacting the DNA but are available for interactions with the transcriptional machinery (114). Gallegos et al. had already previously suggested that conserved amino acids in this part of the HTH motif may contact the transcriptional machinery whereas the first helix would be more variable and involved in recognition of the target promoter (71).

The N-terminal end (also called the regulatory domain) of XylS/AraC members is not well conserved. The N-terminal domain of AraC has been crystallized in the absence and in the presence of arabinose, but it has such a low percent identity to XylS (12.5%) and XylS-mediated transcription activation from Pm is so different from AraC and the PBAO promoter (52, 71) that extrapolation of the AraC structure to XylS is doubtful. Structural ideas were retrieved, however, from genetic evidence and different XylS mutants. For example, single point mutations in the N-terminal end of XylS resulted in proteins responsive to other aromatic inducers, which suggests a role for the N-terminal end in effector recognition (Fig. 2C) (144, 218). Furthermore, some mutations in this region impaired the in vitro formation of cross-linking products between XylS peptides containing only the N-terminal part fused to the maltose binding protein. This was interpreted such that the N-terminal part of XylS would be involved in dimerization as well (217, 239). Both C- and N-terminal parts can act independently (218, 219). For example, mutations in both the C- and N-terminal regions of XylS can yield semiconstitutive mutants (203) or suppressors (143). This suggests there is some form of “cross talking” between the two domains, but the mechanism of this remains to be resolved.

Mechanisms of Activation

An understanding of the activation mechanisms by the “catabolic” members of the XylS/AraC family may be biased by the relatively large body of knowledge about XylS compared to others. A first peculiarity of the XylS system is that activation can take place in two modes, i.e., at low xylS expression levels in the presence of an inducer (66, 204) and on overexpression
in the absence of inducer (91, 94, 142, 202, 241), although the latter mechanism may have no physiological meaning for the xyl pathway. At least BenzR seems to behave similarly to XylS in this respect (36).

XylS probably binds as two monomers or one dimer to two direct sequence repeats in the Pm promoter. This can be concluded from crystallographic data for the two AraC members Rob and MarA (114, 210). DNase I footprinting of the XylS recruitment to the Pm promoter showed a protected region from nucleotides −78 to −28 upstream of the transcription start site (102, 103). The protein bound along one side of the DNA, covering four helices and thereby making specific contacts in four adjacent major grooves (102). By earlier site-directed mutagenesis studies, a consensus direct repeat (TGCA-N6-GGNTA), which is present twice in the Pm promoter between −35 and −49 and between −56 and −70, had been identified as the binding determinant (70, 78, 107). In the presence of m-toluate, the degree of protection by XylS on the Pm promoter increased (102, 103), suggesting that the affinity of XylS to its binding site had changed. Subsequent in vivo methylation studies showed an altered methylation protection pattern of the XylS-bound region in the presence of m-toluate, which indicates that conformational changes take place under inducing conditions (78).

Strangely, XylS-mediated transcription activation from the Pm promoter is independent of σ70-RNAP (136) but during exponential growth is dependent on the alternative sigma factor σ32, which normally is involved in the heat shock response (135). This has led to the suggestion that the presence of the chemical effector m-toluate is triggering a similar response for the cell to the heat shock response. In the late exponential and stationary phases, the sigma requirement for the Pm promoter changes to σ32 (135, 136). However, σ38-RNAP uses exactly the same promoter region in the Pm promoter as that used by σ32 (136). Mutants can be obtained with mutations in the N-terminal domain of XylS which change the requirement of XylS for σ38 to σ70 (203).

A large number of XylS/AraC members, including XylS itself, seem to require contact with α-CTD and with parts of the α-factor in order to achieve full activation (15, 53, 116, 117, 131, 220). This was concluded mostly from in vitro transcription experiments and transcriptional reporter gene fusions with mutated RNAPs. A direct interaction between XylS and RNAP has not been demonstrated conclusively (251), but introducing 2 bp between the nucleotides at positions −36 and −37 of the Pm promoter abolished XylS-dependent transcription activation, probably because XylS and RNAP were offset and could no longer interact (77). In vivo dimethyl sulfate footprinting suggested that XylS can retain RNAP on the Pm promoter in the absence of inducer. In the presence of benzate, RNAP was released, with concomitant transcription initiation (147).

What can be concluded about the role of XylS in the process of transcription activation? Apparently, XylS prevents open-complex formation in the absence of inducer but somehow recruits RNAP to bind to the Pm promoter. In the presence of inducer, XylS changes its own conformation and that of its binding site, which allows isomerization of promoter-RNAP to form the open complex (251). Exactly how the effector mediates these changes and how this is transmitted to the DNA and RNAP remain elusive.

### GntR-Type Regulators

#### Catabolic Operons Controlled by GntR-Type Regulators

Regulation by repression is rare for catabolic pathways. Fernandez et al. were the first to report that the product of the paaX gene repressed the expression of two operons for phenylacetic acid (PA) degradation (paaABCDEFGHJK and paaZ [Fig. 1E]) in E. coli strain W (63). These operons code for an aerobic catabolic pathway but have typical features of anaerobic degradation of aromatic compounds, because of activation of PA to phenylacetyl-coenzymeA (PA-CoA) by the action of a PA-CoA ligase (PaaK). By sequence similarity, the paaX gene product was classified as a GntR-type repressor. The paaX gene is located downstream of paaK and is transcribed in the same direction, although it is not part of the same transcript (63). Since then, a few other GntR members have been found for catabolic pathways (Table 4). The repressors PhcS and AphS regulate the expression of the phenol degradation pathway in C. testosteroni strains R5 and TA441 (7, 249). The phcS and aphS genes both belong to a small operon (phcRS and aphRS), of which the second gene is coding for an NtrC/XylR-type regulator (see below) (7). The aphRS and pheS operons are located upstream of and transcribed in the opposite direction from their target catabolic operons (8, 249). The AphS protein acts as a factor for silencing the aph genes for phenol metabolism, whereas AphR is its transcription activator. Knockout mutations in the aphS gene were shown to enhance the transcription of the aph structural genes and of the aphR promoter as well (7). VanR, which represses the vanAB vanillate demethylase genes in Acinetobacter, is also a GntR-type regulator (149). The vanR gene is transcribed in the opposite direction to and overlaps with the vanB gene. Transcriptional regulation of biphenyl degradation in Ralstonia eutropha A5 (152), Pseudomonas sp. strain KKS102 (165), and P. pseudoalcaligenes KF707, is now also assumed to be mediated by GntR-type regulators (BphS) (269). The bphS gene in R. eutropha R5 is oriented in the opposite orientation to the first gene of the biphenyl operon (bphE) (152). In the strain KKS102, bphS is separated from its target operon by an insertion element (165). The third regulatory gene (orf0, now bphR2) present in a biphenyl catabolic regulon is also located upstream of the target operon but transcribed in the same direction (269). The Orf0 (now BphR2) protein is atypical in the GntR group since it is presumed to be an activator rather than a repressor.

GntR family members that control the degradation of aromatic compounds (except Orf0) are transcriptional repressors in the absence of the pathway substrates. In the presence of the pathway substrate, repression is released by interaction of the regulator with the aromatic compounds or one of its metabolites (62, 149, 165, 249). This was interpreted as a loss of DNA binding affinity mediated by the chemical inducer (see below) (62, 165). The mechanism by which Orf0 activates transcription is not known. Strong evidence for its activator function comes from gel shift assays performed with the orf0 promoter region and purified Orf0. In these experiments, Orf0 bound the pro-
moter fragment more strongly in the presence of inducers (2,3-dihydroxybiphenyl and 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate) than in their absence (269).

Structure and Conformation

GntR-like regulators are between 239 and 254 amino acids long with a molecular mass of around 27 kDa. The PaaX protein is considerably bigger, with 316 amino acids and a molecular mass of 35 kDa. Structural information comes from GntR-type regulators that are not involved in pathways for degradation of aromatic compounds. Haydon and Guest first described bacterial regulators of the GntR family as having similar N-terminal DNA binding domains but displaying high heterogeneity among the various C-terminal effector binding and oligomerization domains (Fig. 2D) (83). This view has been largely confirmed by crystallographic structure determination and mutagenesis studies (199, 200, 259, 273). Exemplary for the GntR-type structure is FadR (211), whose structure has been resolved in the presence of its target DNA and its cognate inducer (Fig. 7). FadR exists as a homodimer in solution and binds DNA as a dimer (199, 259, 260, 273). The protein in the crystal structure showed an α/β N-terminal domain (residues 1 to 72) with the HTH motif and a C-terminal domain (residues 79 to 228) that is formed by seven α-helices with short connecting loops. The seven α-helices are packed together to form a bundle, and a large cavity within this bundle makes up the effector (myristoyl-CoA) binding pocket (259, 260). N- and C-terminal domains interact only via a short linker region comprising two short α-helices. The two monomers in the

FIG. 7. Structure of the FadR-DNA complex (273). The paired recognition helices are in the major groove, and the two wings are within the flanking minor grooves. Note that the DNA is bent toward the FadR protein, leaving a slight contraction in the major groove as a result. Reprinted from reference 273 with permission from the publisher and from the authors.

<table>
<thead>
<tr>
<th>Table 4: GntR-type regulators controlling the degradation pathways of aromatic compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regulatory protein</strong></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>BphS</td>
</tr>
<tr>
<td>Orf0 (BphR)</td>
</tr>
<tr>
<td>AphS</td>
</tr>
<tr>
<td>VanR</td>
</tr>
</tbody>
</table>

* Relative to the transcription start site.

ND, not determined.
dimer are packed together in parallel fashion, resulting in reciprocal interactions of the domains and linker regions across the interface (273). Only the tip of the paired α-recognition helices projects orthogonally into the same major groove (259, 273). Each terminal portion of the α-recognition helix interacts in an identical fashion with the DNA. Part of the HTH motif also penetrates the minor groove of the DNA helix to make specific contacts with two bases (259). The DNA has a B-form conformation with a curvature of 20° toward the protein, and this results in a contraction of the central major groove and an expansion of the opposite minor groove (273).

**Mechanisms of Regulation**

The repression mechanism of GntR-type regulators seems to be a simple hindrance for RNAP binding or open-complex formation (Fig. 8). GntR-type proteins bind either the promoter region itself or between the transcription and translation start sites. For example, four regions on the bphE promoter were bound by purified BphS of *Pseudomonas* sp. strain KKS102. The binding sites are located between positions −12 and +3 (called BS₀), +6 and +23 (BS₁), +27 and +40 (BS₂), and +44 and +58 (BS₃) compared to the transcription start site. BphS has more affinity to BS₀ and BS₁ than to BS₂ and BS₃, a difference which was explained by the presence of inverted repeats in BS₀ and BS₁ (Table 4), and no such features in BS₂ and BS₃. In fact, fragments with only BS₀ and BS₁ are sufficient for BphS to mediate repression, whereas transcription from an artificial promoter fragment containing only BS₂ and BS₃ was not repressed to a significant extent. Nevertheless, the strongest repression was obtained when all four binding sites were present (165). BphS loses its ability to bind to the promoter in the presence of the inducer 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid (an intermediate in biphenyl metabolism).

A 15-bp imperfect palindromic sequence was identified in the PaaX-dependent promoters paaZ and paaA. The palindrome was located between positions −30 and −16 in the paaZ promoter and between +3 and +17 in the paaA promoter. DNase I footprinting showed that the PaaX protein actually protects a larger region, which extends from −1 to +51 in the paaA promoter and from −30 to +17 in the paaZ promoter (62). Hypersensitive sites were spaced at approximately 10-nucleotide intervals, suggesting binding of PaaX to one side of the double helix. The fact that PaaX protected the −10 box in the paaZ promoter but not in the paaA promoter suggests a different mechanism of PaaX repression in these two promoters (as shown in the model in Fig. 8). In the presence of phenylacacetate-CoA, binding of PaaX to the paaA and paaZ promoters was released (62). A sequence of events leading to the release of repression can be deduced from the FadR-DNA structure in the presence of myristoyl-CoA (the inducer for FadR) (259). Binding of the effector induces conformational changes in the sensor kink helix α8 (Fig. 7), which then forms two separate helices. As a result, some of the side chains on this helix are pushed toward the neighboring helix, α4 (linker region), which responds by shifting away from helix α8 toward the N-terminal DNA binding domain. This shift introduces additional contacts with the helix α1 in the N-terminal domain, which tilts away and causes a hinge-bending motion of both DNA binding domains of the dimer in opposite direction. This leads to a change in the distance between the two recognitions helices, which now lose their ability to interact with the DNA helix (259).

**TetR-, MarR-, AND FNR-TYPE REGULATORS**

**TetR Family**

Only one example of a TetR-type regulator involved in the regulation of an catabolic pathway for aromatic compounds is known. This protein, CymR, is a repressor for the expression of the genes for p-cymene (*cym*) and p-cumate (*cmt*) degradation in *P. putida* F1 (Fig. 1H; Table 5) (50). The *cym* gene itself is positioned upstream of and transcribed in the same direction as the *cym* operon. Similar sequences have been identified in the *cmt* and *cym* promoter regions (Table 5) and may be the recognition sites for the CymR repressor protein (50, 164). CymR is supposed to impose its repressing effect by inhibiting RNAP access to the promoter. In the presence of the pathway effector, p-cumate, CymR is assumed to no longer bind to its operator site, in analogy to the TetR repressor in the presence of tetracycline (170). Strangely, p-cumene, the substrate for the *cym*-encoded pathway, is not an effector for CymR (50). The CymR protein is 203 amino acids long with a molecular mass of 23 kDa and is supposed to form a dimer in solution, like the TetR protein, for which the structure is known (170). Its N-terminal part contains an HTH motif (predicted for amino acids 19 to 64 in CymR), which is responsible for DNA binding, whereas the C-terminal domain binds tetracycline (170). The C-terminal domains interact in the dimer to form a core. The HTH motifs bind to the (palindromic) operator sequence, with the recognition helices aligned parallel to the major groove. The distance and the relative orientation of the HTH motifs distinguish between the induced and the noninduced status of TetR. Inducer binding triggers a sequence of α-helix displace-
mans, which results in an increasing separation of the recognition helices. The shift of the recognition helices along the major groove disrupts the contacts between the DNA binding domains and the operator, causing dissociation of the TetR operator complex (170).

**MarR-Type Regulators**

The nbz operon for aminophenol degradation on plasmid pNB1 in *P. putida* HS12 is regulated by NbzR, a MarR-type repressor protein (176). The chemical inducer for the pathway has not been identified, and it was proposed that another trans-acting factor might be involved in regulating pathway expression. Other MarR-type regulators involved in catabolic pathways include HpcR (HpaR), the repressor of the *hpc* (*hpa*) operon for 3- and 4-hydroxy- and 3,4-dihydroxyphenylacetic acid (homoprotocatechuate) degradation in *E. coli* strains C (214) and W (68). In fact, current information from genome sequence shows that the pathways for homoprotocatechuate degradation are very common among *Gammaproteobacteria*. CbaR, another MarR-type regulator, controls the *cbaABC* operon for 3-chlorobenzoate degradation present on plasmid pBRC60 in *C. testosteroni* BR60 (197). 3-Chlorobenzoate and protocatechuate are effectors for CbaR, leading to derepression (197). The genes for benzoyl-CoA reductase in *Rhodopsseudomonas palustris* BSII), which are located near the transcription start site. BS I is the BSI DNA fragment with higher affinity. In DNase I footprinting experiments, CbaR bound to a BSI DNA fragment with higher affinity than to BSI, which led to the conclusion that independent CbaR proteins bind to each site. In the presence of 3-chlorobenzoate, benzoate, or protocatechuate, DNA binding by CbaR is disrupted, whereas 3-hydroxy- and 3-carboxybenzoate increased binding affinity by 60-fold. The physiological significance of this phenomenon is not understood (197). HpaR from *E. coli* is binding to two

### TABLE 5. TetR, MarR, and FNR family members and two-component regulators in degradation pathways of aromatic compounds

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Family</th>
<th>Inducer(s) or inducing condition(s)</th>
<th>Accession</th>
<th>Inducer(s) or inducing condition(s)</th>
<th>Accession</th>
<th>Inducer(s) or inducing condition(s)</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aminophenol</td>
<td>TetR</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
</tr>
<tr>
<td>Aminophenol</td>
<td>MarR</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
</tr>
<tr>
<td>Aminophenol</td>
<td>FNR</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
</tr>
<tr>
<td>Aminophenol</td>
<td>TetR</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
</tr>
<tr>
<td>Aminophenol</td>
<td>MarR</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
</tr>
<tr>
<td>Aminophenol</td>
<td>FNR</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
<td>p-Cumate</td>
<td>B43334</td>
</tr>
</tbody>
</table>

**Note:** Relative to the transcription start site.
operator sites in the Pγ promoter for the hpaGEDEFHI operon, one of which overlaps with the transcriptional start site (68). The operator around the transcriptional start site contains a perfect palindromic sequence (AATCATTAA-ATCAA) are found, upstream of the transcription start sites of hbaR and badD, interrupted inverted repeats (TTGAT-N4-ATCAA) are found, which were postulated to be AadR binding sites (55, 56). The cprK gene is part of a cluster for 4-chlorophenol dehalogenase (238). One small transcript which encompasses the cprBA genes for the dehalogenase itself is induced under conditions of dehalorespiration; the remainder of the cpr genes are expressed constitutively. At several points within the cpr cluster, FNR-like binding motifs were detected. Within the cprB promoter, this motif (TTAAT-N4-CTAA) occurs at 41 nucleotides upstream of the transcription start site and is presumed to be a CprK binding site (238).

The proteins of the FNR family are generally between 233 and 242 amino acids in length with a calculated molecular mass of around 27 kDa. The best-characterized member of the family is FNR, a global regulator which represses the expression of genes involved in aerobic respiration and activates the expression of genes that permit the reduction of alternative electron acceptors under anaerobic conditions. In its active DNA binding form, FNR is a homodimer containing one [4Fe-4S]2+ cluster per subunit (12). Integrity of the cluster is required for FNR dimerization, site-specific DNA binding, and transcription activation (108). The cluster is also essential for redox sensing by FNR. FNR is supposed to interact with the α-C-terminal domain and the sigma factor of the RNAP (121, 131). The amino acid sequence of AadR contains the essential conserved cysteine residues for iron-sulfur coordination and, as such, may be a redox sensor as well. In contrast, HbuR and CprK lack the characteristic N-terminal cysteines and are not specific sensors for anaerobic conditions (56, 238).

**TWO-COMPONENT REGULATORY SYSTEMS**

Two-component signal transduction systems comprise the major mechanisms by which bacteria sense environmental signals and control global cellular processes. They typically consist of two individual proteins, a sensory histidine kinase and a response regulator. In the general model (reviewed in references 5, 54, 175, 209, and 243), signal perception by the N-terminal domain of the sensor kinase catalyzes an ATP-dependent phosphorylation of a conserved histidine in the central region of the protein. Once the histidine is phosphorylated, the
phosphoryl group is transferred from the histidine to a conserved aspartate usually found at the N terminus of the cognate response regulator. Phosphorylation of the regulatory domain catalyzes the effector domain, which is the final mediator for gene expression control. Although this is a basic scheme, it is highly adaptable, and numerous variations have been found.

Involvement in Regulation of Catabolic Pathways

A number of two-component regulatory systems are known to control the expression of catabolic pathways. The first of these is the TodST system, which was identified in *P. putida* F1 and *P. putida* DOT-T1 (118, 151). TodST controls toluene degradation from the *todX* promoter, which lies upstream of the *todXFC1BADEGIH* operon. The *todX* genes are located downstream of but in the same direction as *todH* (Fig. 11). The regulatory genes form their own transcriptional unit. Expression from the *todX* promoter is elicited by growth on toluene (Table 5) (118). In *P. mendocina*, the TmoST proteins regulate the expression of the *tmo* toluene-4-monoxygenase genes. The *tmoST* genes themselves are not located in the direct vicinity of the *tmo* operon (205). Another two-component system regulates the expression of benzylsuccinate synthase in *Azoarcus* sp. strain T and in *Thauera aromatica* strain K172. Benzylsuccinate synthase is the first enzyme in the anaerobic conversion of toluene and is under the control of a two-component system named TdiSR (3, 126). In *T. aromatica* strain T1, two potential two-component systems, *tutCB* and *tutCB*, seem to be involved in the regulatory control of toluene degradation under aerobic and anaerobic conditions (35, 126). A further two-component system named BpdST was identified as the possible sensor/regulator of the biphenyl degradation pathway in *Rhodococcus* sp. strain M5 (115). As with the *todST* genes, *bpdST* is located downstream of the target operon. Transcription activation by BpdST takes place in the presence of biphenyl (115). Finally, a two-component system called StyRS has been described for the *styABCD* degradation pathway of styrene to phenylacetate (Table 5). StyRS-mediated expression from the *styA* promoter is inducible by styrene but repressed by phenylacetate (166). StyRS analogs have been found in *Pseudomonas* sp. strain Y2 (264), in *P. putida* CA-3 (166), in *Pseudomonas* sp. strain VLB120 (174), and in *P. fluorescens* ST (222). In all these bacteria, the *stySR* genes are located upstream of and in the same direction as the *styABCD* operon.

Structure and Conformation of the Same Sensor Kinase Component

The sensor kinase proteins of the two-component systems involved in regulation of degradation pathways of aromatic compounds are structurally dissimilar. One example is the TodS protein, which is 978 amino acids long and has a molecular mass of 108 kDa. TodS has five predicted defined domains, of which the first contained a so-called bZIP motif. The bZIP motif consists of a stretch of several basic amino acid residues, which probably contact DNA directly, and an adjacent region with a heptad repeat of leucine residues (the leucine zipper) that mediates protein dimerization (74, 257). The presence of a bZIP domain in a sensory histidine kinase was very unusual. The TodS bZIP domain alone was shown to bind a DNA fragment upstream of the *todS* start codon. This region contained the sequence 5’-TGACTCA-3’, which is identical to the recognition sequences of the eukaryotic proteins FOS and Jun, which also contain the bZIP motif (118). The possible physiological relevance of the TodS bZIP binding to the DNA is not known. The TodS protein also contains two identical histidine kinase domains, called HK1 and HK2, spanning amino acids 184 to 409 and 756 to 978, respectively. The histidines at positions 190 and 760 are the sites of autophosphorylation. The two domains also carry a peptide motif called the G1 and G2 blocks, which are involved in ATP binding. An intrinsic response regulator domain reminiscent of CheY (266) spans amino acids 443 to 570 and could be potentially phosphorylated by other signal-transducing proteins. Amino acids 34 to 153 contain a hypothetical chemical sensor domain, whereas the region between positions 592 and 735 is supposed to form the oxygen-sensing domain (26, 118). Oxygen-sensing domains are found in various redox and light sensor proteins (248, 276).

The authors concluded that TodS might actually be a dual sensor that is capable of sensing both toluene as a primary signal and oxidative stress in the cytoplasm (118). Isolation of TodS effector specificity mutants proved that the aromatic compounds directly interact with the sensor kinase (26). From amino acid similarity comparisons, it was concluded that TutC, TmoS, and StyS are homologous to TodS (35, 205, 264).

In contrast to TodS, the Tdis protein of *T. aromatica* K172 is only 548 amino acids long and has a molecular mass of 63 kDa (126). Tdis has three distinct predicted domains. Two domains are found at the N terminus (residues 27 to 162 and 182 to 330); they are similar to each other and to other oxygen-sensing domains (see above). The C-terminal domain (residues 349 to 546) is a typical conserved histidine kinase domain. Tdis of *Azoarcus* sp. strain T and TutC1 of *T. aromatica* strain T1 are very similar in organization to Tdis. Leuthner and Heider suggested that the Tdis-type protein could be specific for aerobic toluene metabolism whereas TodS (or StyS) would preferentially be involved in aerobic metabolism only (126). *T. aromatica* strain T1 can metabolize toluene under both aerobic and anaerobic conditions (60) and, interestingly, has two toluene sensor/response regulators, of which TutCB resembles TodST and TutC1B1 resembles TdiST. However, it is not known whether one of them really controls the aerobic pathway only and the other controls the anaerobic pathway.

The last type of sensory histidine kinase is found only in BpdS. BpdS is exceptionally long: 1576 amino acids with a molecular mass of 170 kDa (115). One region in the N terminus (residues 61 to 146) shows sequence similarity to a conserved domain of eukaryotic proteins, which use tyrosine or serine/threonine as acceptors of phosphoryl groups. Between residues 316 and 323, there is a Walker A motif that could bind ATP. The C terminus of BpdS contains the histidine kinase domain. In addition, seven potential hydrophobic transmembrane segments exist, which could mean that the C-terminal histidine kinase is located within the cytoplasm whereas the N-terminal part of the protein would stick in the periplasmic space. In this respect, BpdS resembles receptor or receptor-like kinase proteins, which carry out signal transduction processes in both plants and animals (115).
Structure and Conformation of the Response Regulators

Once the signals are perceived by the sensory kinases of the system, the kinases transmit the signal in form of phosphorylation to the response regulator. The response regulators (TodT, BpdT, etc.) are 206 to 227 amino acids long and have a molecular mass of between 23 and 25 kDa. The response regulators in the two-component systems controlling the aromatic compound degradation pathway have a very similar configuration. They contain an N-terminal regulatory domain (residues 10 to 115 in TodT) with an aspartate phosphorylation site and a C-terminal DNA binding domain (residues 144 to 187) with an HTH motif (118). TodT binds the todX promoter even in the absence of phosphorylation (118). The TodT binding site (the tod box) has been identified as a 12-bp inverted repeat (the tod box) has been identified by speciation as a 12-bp inverted repeat (ATAAAC-N4-GTTTAT) centered at bp −105.5 from the todX transcriptional start site. The dyad symmetry of the binding sites suggests that TodT binds as a dimer (118).

The response regulator StyR of the StyRS system was shown to bind an identical sequence motif to TodT (i.e., ATAAAC-N4-GTTTAT); however, in this case the motif was centered at bp −44.5 from the styA transcription start site (Table 5) (222). The nonphosphorylated form of StyR was shown to be a monomer, but on phosphorylation the two monomers join to form a dimer (124). This is also known to occur in other two-component regulatory systems (37, 64, 89, 129, 243). StyR in dimeric form has a 10-fold-higher affinity for the styA promoter than does the monomer (124), which could mean that dimerization of the response regulator is needed for efficient recruitment of the protein on the promoter and subsequent transcription activation. Interactions of TodT or StyR with RNAP have not been experimentally probed, and it cannot be concluded that the transcription activation mechanisms should be the same.

XylR/NtrC-TYPE TRANSCRIPTIONAL REGULATORS INVOLVED IN CATABOLIC PATHWAYS

In contrast to the above-described regulatory proteins, which control transcription mediated by $\sigma^{34}$-RNAP (except for XylS), XylR/NtrC-type regulators activate RNAP containing the alternative sigma factor $\sigma^{54}$. The $\sigma^{54}$-RNAP holoenzyme forms a stable complex with −12 and −24 promoters but is unable to start transcription without further activation (17, 192, 223). Isomerization to the open complex is strongly stimulated by specific regulatory proteins, generally known as prokaryotic enhancer binding proteins. The best-characterized member is NtrC, and often the family of prokaryotic enhancer binding proteins is referred to as the NtrC family. NtrC-type regulators bind to DNA regions called upstream activating sequences (UASs), which are usually located more than 100 bp upstream from the $\sigma^{54}$-RNAP binding site (18, 19, 79, 150, 208). Interactions between $\sigma^{54}$-RNAP bound to the −12/−24 region and the regulatory protein associated with the UASs are often facilitated by a bend in the intervening DNA, which can be the result of integration host factor binding (190).

Catabolic Operons Controlled by XylR/DmpR Subclass Regulators

Regulation of aromatic-compound degradation is very often mediated by NtrC-type regulators. The best-studied examples of those are the XylR and the DmpR proteins from P. putida (see below). For this reason, the specific subclass of NtrC-type regulatory proteins which act on promoters from catabolic pathways is also often referred to as the XylR/DmpR subclass, and we refer to them as such here as well. Briefly, XylR from P. putida mt-2 is one of the two main regulatory proteins for the xylene and toluene degradation pathway encoded on the TOL plasmid (the other being XylS [see above]) (201). XylR is responsive to chemical effectors such as m- and p-xylene, which are the substrates for what is called the upper pathway. XylR in combination with $\sigma^{54}$-RNAP triggers expression from the P$_r$ promoter (in front of the xylU gene) and from the P$_p$ promoter, driving expression from xylS as well (2, 47, 91, 111, 202). The gene for XylR itself is located downstream of the meta-cleavage pathway operon and of xylS (Fig. 1E). In addition to being a transcriptional activator, XylR represses its own expression. Under inducing conditions, autorepression of xylR increases (14, 134). The reason for the autorepression seems to lie in the fact that the xylR promoter and the XylR binding sites in the intergenic regions for xylS and xylR overlap, which may prevent $\sigma^{54}$-RNAP binding to the xylR promoter.

The other very extensively studied NtrC-type transcriptional regulator involved in the degradation of aromatic compounds is DmpR. DmpR regulates expression from the P$_p$ promoter, which drives transcription from one single large operon for phenol degradation (dmpKLMNOPQBCDEFGHI) that is present on the pV1150 plasmid in Pseudomonas sp. strain CF600 (234). The dmpR gene itself is located directly upstream of and is divergently oriented from dmpK. A large number of highly similar regulators to DmpR have been found, which all control phenol degradation operons (Table 6) (8, 153, 158, 177, 229, 249). A few other examples worth mentioning here are TbuT and TbmR, which control toluene monoxygenase gene expression in Burkholderia picae PK01 (22) and Burkholderia sp. strain JS150 (98), respectively. Two other members are TouR and RsoR, which controls toluene monoxygenase gene expression in Pseudomonas putida mt-2 (23) and P. putida mt-2 (22). Also, AreR, which is involved in the degradation pathway for aryl esters (arcBA) in Acinetobacter sp. strain ADP1 (100). The genes for TbuT and TouR are located downstream of the target operons (arcABC) and touABCDEF, respectively, but are transcribed in the same direction (11, 22). Expression of touT is partly controlled by TouT itself, since a transcript covering both the tou structural genes and touT promoter is synthesized in the presence of toluene, which starts at the TbuT-dependent promoter P$_{touT}$ (22). Also, arcR is transcribed in the same direction as areCBA but is located upstream of it. Finally, HbpR is the regulatory protein which controls two small operons for 2-hydroxybiphenyl degradation (hbpCD) in Pseudomonas azelata HBP1 (97). The intergenic region between hbpR and hbpC looks as if it has been subject to recent duplication and deletion events, since a small 5’ part of a second hbpR copy is still present (96). Expression from hbpR itself is autoregulated by HbpR, which
could be attributed to a secondary set of HbpR binding sites overlapping the hbpR promoter (96).

The types of chemical effectors and cofactors required for XylR/DmpR-mediated activation have been very well studied (Table 6). One of the motivations for these studies has been the urge to modulate the effector spectrum of the regulators by mutagenesis (73, 237). XylR/DmpR-like activators require a chemical effector and ATP as the cofactor. The effector is usually the primary substrate of the target pathway or a compound related to this. Both XylR and DmpR hydrolyze ATP or dATP, which is essential for activation (188, 270). In addition, XylR can hydrolyze GTP, whereas DmpR only binds to but does not hydrolyze GTP, CTP, and UTP (270). Some aromatic compounds such as 2,4-dimethylphenol and 4-ethylphenol can bind DmpR and activate ATP hydrolysis without triggering transcription activation (unproductive effectors). This process is not well understood but may involve a slightly different conformational change required for ATP hydrolysis and for productive contacts with the transcription machinery exerted by productive and unproductive effectors (168, 169). Unproductive effectors can also act as competitive inhibitors for productive effectors (i.e., those which lead to transcription activation) (168). 3-Nitrotoluene can also bind XylR without mediating activation (73, 221). In several cases, inducing compounds which effect transcription activation are not direct substrates for the target pathway, such as 2-aminobiphenyl and HbpR (97).

### Domain Organization

XylR/DmpR-type regulators are typically between 563 and 570 amino acids long (molecular mass of 62 to 68 kDa), with the exception of AcrR, which is 600 amino acids long. Since XylR and DmpR have resisted full purification, there is no direct structural information on those proteins, except ideas which can be extrapolated from genetic experiments. We therefore use structural data from other NtrC-type regulators, which are not involved in the degradation of aromatic compounds, to illustrate possible structural motifs for XylR/DmpR subclass members. Members of the NtrC family in general display an ordered three-domain structure, although the domains themselves are not necessarily very similar among different regulators (150) (Fig. 2E). The C-terminal end (forming the so-called D-domain) contains the HTH motif responsible for DNA binding (162) and in some cases for dimerization (110, 184).

The central or C-domain is the most highly conserved region among NtrC family members, and seven highly conserved regions (C1 to C7) have been identified here (150, 172). Mutagenesis of the regions C1 (containing the Walker A motif), C4 (Walker B motif), and C7 in XylR and DmpR showed they are involved in ATP binding and hydrolysis (188, 270). Mutagenesis of and cross-linking experiments with DctD (an NtrC-type regulator not involved in catabolic pathways) showed the conserved C3 region to be interacting with RNAP through the β and σ34 subunits of the holoenzyme (106, 122, 267). In vitro interactions between σ34 and PspF (a noncatabolic activator) took place between the 1-region of σ34 and the PspF C-domain, most probably involving the 6-amino-acid GAFTGA motif within the C3 region (25).
The so-called A-domain, which is located at the N-terminal end and is very specific for the different NtrC-type regulators, has been implicated in signal perception and hence determines the specificity of the protein (233). The A-domain of DmpR and XylR (211 amino acids long) was shown to interact directly with the inducing aromatic compounds, and various effector specificity mutations have been generated in this region of the protein (38, 73, 167, 168, 183, 221, 235, 236). One inducer binding site is present per monomer, which was demonstrated for DmpR (167) and which could be pinpointed to a subregion between amino acid residues 107 and 186 (237). A structural model of the XylR A-domain generated by a “fold recognition approach” (41) predicted eight α-helices and seven β-strands connected by 14 loops. Based on this model and the different XylR/DmpR effector specificity mutants, five different regions in the A-domain were postulated to form the aromatic compound binding site. These are loops 12 (position 174), 10 (position 142), 4 (position 65), 2 (positions 37), and 14 (position 203). The A-domain was also shown to modulate the ATPase activity of the C-domain. This was concluded from observations that regulatory proteins with their A-domain deleted displayed constitutive ATPase activity and activated transcription to the same extent as the complete protein in the presence of an inducer (9, 185, 229, 236). Successive deletion of the A-domain identified a subregion between residues 160 to 210 which was sufficient to repress XylR-dependent transcription activity in the absence of the inducer (187). Several other lines of evidence (167, 185, 229) have led to the current hypothesis that the A-domain acts as movable domain which, in its closed state, represses the ATPase activity of the C-domain but, in an open state that occurs under inducer binding, exposes the constitutive ATPase activity. The connecting amino acid sequence between the A- and C-domains is provided by a short region called the B-linker, which seems essential for the change from the closed to the open state. DNA binding data with purified HbpR suggested that the protein assembles to the oligomeric form irrespective of the presence of the effector (255).

Structure and Configurations

Based on sequence similarities, NtrC-family members were classified as members of the AAA+ superfamily (ATPases associated with various cellular activities) (157). Recent structural information has confirmed this sequence-predicted classification. The structure of the joined N-terminal regulatory and central ATPase domains of NtrC1 (NtrC1AC) from Aquifex aeolicus and the structure of the isolated central domain (NtrC1C) were recently determined in their ADP-bound form (Fig. 10) (123). The central ATPase domain in NtrC1AC is inactive, whereas that of NtrC1C is capable of oligomer assembly, ATP hydrolysis, and transcriptional activation. The resolution of these structures has been an important step in recognizing the structural differences between the active and inactive states of the central domain. NtrC1AC is a dimer in solution in which the regulatory domain of each monomer is adjacent to the central domain of its partner in the dimer (Fig. 10A). This might also be the state for XylR/DmpR proteins, since, for instance, DmpR without ATP, DNA, and inducer appeared as a dimer in solution as well (270). The main dimerization interface is formed by the B-linker and monomers are held in a front-to-front fashion in the dimer (Fig. 10A). The dimerization interface seems important for maintaining the off-state of the ATPase domain. In fact, mutations in the B-linker of XylR that disrupted the linker’s secondary structure resulted in a much higher background expression of a single-copy chromosomal PγlacZ fusion (72), which seems in agreement with the presumed role of the B-linker.

The central domain protein NtrC1C without the A-domain crystallized as a ring-like heptamer (Fig. 10B). The ultrastructure of purified PspF stabilized in oligomeric form with the ATP transition analogue ADP-aluminium fluoride was reconstructed as a trimer of dimers (or asymmetric hexamer) from cryoelectronmicrographs (231). The diameter and height of the NtrC1C heptamer rings were approximately 124 and 40 Å, respectively, with a front-to-back packing of the protomers. In this configuration, the ATPase domains contact two others simultaneously on each side. In the NtrC1C dimer, one side of the ATPase domain is contacting the other subunit whereas the second one (β-hairpin and GAFTGA in Fig. 10) is interacting with the B-linker. Because of this, an essential highly conserved Arg residue at position 293 is far away from the bound ATP molecule in the dimer whereas in the heptamer R293 comes close to the ADP of its neighboring protomer. The Arg-293 residue could thus actually be contacting the γ-phosphate of ATP and could catalyze ATP hydrolysis in the heptamer configuration. There is good evidence to assume that
XylR/DmpR-like proteins also form heptamers or hexamers as part of their activation cycle. For example, addition of ATPγS and the inducer 2-methylphenol resulted in mainly hexameric forms of DmpR (270). Although XylR has resisted purification attempts, the effect of ATPγS and ATP on subunit oligomerization was detected with purified XylR-ΔA (186). Transmission electron microscopy shows that XylR-ΔA and a P₅ DNA fragment were forming larger (XylR oligomers) and smaller (dimer) complexes, depending on the presence of ATP (72).

The GAFTGA-loop forms the motif that is assumed to be interacting with the RNA polymerase transcription machinery. In the NtrC³ structure, all GAFTGA-loops pointed toward the middle central pore of the heptamer (Fig. 10B). This would create an extended binding surface for σ⁵⁴-RNAP, although it is not clear where exactly the A-domains would go in the heptamer. For the PspF transcriptional regulator (which, however, lacks an A-domain), the interaction surface to σ⁵⁴-RNAP was formed between the I-region of σ⁵⁴-RNAP and the GAFTGA-loop of PspF (25). The hexameric state observed for PspF (231) and the heptameric NtrC³ structure strongly suggest that XylR/DmpR-type regulators form similar oligomeric complexes. However, very different from PspF and NtrC1, which do not have a specific effector domain, must be the mechanism leading to effector-mediated transcription activation. Garmendia and de Lorenzo have proposed an activation cycle for XylR (72). In the activation cycle, XylR would change between dimer (unbound) and heptamer/hexamer (bound to the DNA), with the disassembly process being triggered by ATP hydrolysis (Fig. 11). For XylR/DmpR-like proteins, which carry an A-domain, there must thus be some sort of mechanical movement of the A-domain in the oligomeric form on induction, probably mediated by the B-linker, thereby perhaps exposing the GAFTGA-motifs in the central part. However, the typical signature stretches of the B-linker are not present in recognizably similar form in some other XylR/DmpR-like proteins members such as HbpR and TbuT (169), which may indicate that the mechanistic “movement” is not a requirement per se or can be achieved by other structures than the B-linker only.

**DNA Binding during the Activation Process**

The current activation model does not include a specific function for the DNA binding site occupied by the XylR/DmpR-subclass regulatory proteins. Two functions have been proposed: (i) to provide a “physical” support for the regulatory protein in the vicinity of σ⁵⁴-RNAP and (ii) to aid in obtaining the active form of the regulatory complex. This was concluded from experiments in which the addition of the DNA fragment containing the binding site increased the ATPase activity of XylR-ΔA or TouR (and TouR-ΔA) (9, 188). However, it has to be said that similar experiments with DmpR showed a small or nonexistent stimulatory effect (270).

XylR/DmpR-like proteins bind to an approximately 60-bp sequence which is located far upstream of the transcription initiation site (the UAS) (Table 6). This was shown from DNase I footprinting analysis for purified HbpR, as well as for DmpR and XylR variants devoid of their N-terminal A-domains (189, 247, 255). Interestingly, the protection patterns were quite similar for the three proteins, and matching DNA sequences forming two imperfect palindromes were present in each of the promoters (189, 247, 255). For XylR, the UAS is localized between bp −127 and −172 from the P₅ transcription start site and between bp −139 and −184 from P₆. The DmpR binding sequences lay between bp −126 and −171 from the P₅ promoter, whereas those for HbpR are located between bp −180 and −227 on the hbpC promoter and between bp −168 and −225 on the hbpD promoter (96, 255). A second set of HbpR-UAS is present in the intergenic region between hbpR and hbpC (bp −494 to −540), which is improperly positioned to act on the hbpC promoter but which was shown to be important for hbpR autoregulation (96). The two palindromic sequences in the promoter are approximately 16 bp long with a spacing of 29 to 42 bp between the centers of the palindromes.

XylR-ΔA bound one UAS on the P₅-DNA with higher affinity than it bound the other (189). At that point, the authors suggested that this could mean that two dimers bound the two binding sites independently. However, in the case of HbpR, both palindromic sequences were simultaneously needed for binding, since fragments containing only one were not maintained in a stable protein-DNA complex in gel mobility shift assays (255). Adding 2-hydroxybiphenyl and ATP resulted in a twofold change in binding affinity of HbpR to the promoter DNA (255). DNase I footprints of the P₅ UAS fragment with purified XylR-ΔA and ATP showed an occupation and slightly expanded protected region of the UAS at lower protein concentrations compared to those in the assay without ATP (186). ATP hydrolysis itself is not required for occupation of the UASs by XylR, since the use of ATPγS, which in contrast to

![FIG. 11. Schematic mode of activation of XylR/DmpR-type regulators.](http://mmbr.asm.org/)
ATP, is not hydrolyzable, gave rise to identical footprinting patterns to those for ATP (186). This would be in agreement with the binding of a higher-order oligomeric form. In vivo footprinting of the $P_e$ promoter showed that the presence of the inducer slightly altered the interaction between XylR and the DNA (1). More recently, Valls and de Lorenzo repeated the in vivo footprinting with a single-copy xylR gene and the $P_e$ promoter (258). Their results showed that XylR interacts only very transiently with its binding site, with hardly any protection visible under both uninduced and induced conditions. In fact, the estimated number of XylR hexamers is around 5 to 30 per individual cell under stationary-phase conditions (65). This would equal a concentration of 10 to 50 nM, which is quite low compared to the 400 nM XylR-$\Delta$A required to protect the DNA in in vitro DNase I footprints. However, it is clear that most static DNA binding studies fully neglect the kinetic aspects of the regulatory cycle in the microbes themselves.

PROSPECTS

As a result of many studies during the past decade, we now have a much better understanding of the behavior of different regulatory proteins in the process of mediating transcription. Although a few pivotal regulatory proteins in classical degradation pathways for aromatic compounds (i.e., XylS, XylR, and DmpR) have resisted purification, the solution state could be deduced for several others, sometimes in the presence of the DNA binding regions. In some cases it has been possible to deduce the structural changes induced in the protein or protein-DNA complex by the chemical effector. In contrast, little is currently known about the interactions between the regulatory protein and the effector compounds. The effector binding pockets and effector binding affinities are not well defined, and it is not clear how effector binding to the regulatory protein transmits to an “activation” signal for RNAP.

Although a relatively clear picture has emerged for the different components of the regulatory machineries, their interactions with the operator DNA and with RNA polymerase or with other auxiliary protein factors, and the types of signaling molecules, the picture is mostly static. Most models suggest a regulatory protein which is bound to its DNA binding site and ready to activate transcription (or disappear after derepression). Clear indications for XylR/DmpR-type regulators point at a cyclic model between the inactive regulator state and the active (multimeric) state, but with an equilibrium toward the inactive state and very little site occupancy by XylR. We might thus be very surprised to find our ideas about the activation (or derepression) processes changing when more kinetic or affinity equilibria data become available.

Our current perspective on regulatory systems is most probably somewhat biased by the choice of traditionally studied microorganisms with degradation capacities, such as Pseudomonas, Acinetobacter,Ralstonia, and Burkholderia, typical members of the Gamma- and Betaproteobacteria. It might very well be that new interesting systems are discovered, such as the two FNR-like and one MarR-like regulatory proteins in Rhodopseudomonas which control benzene degradation under anaerobic conditions. Also, regulatory proteins with possible membrane-spanning domains (such as BpdS) or with association with transport and chemotaxis (such as NahY of P. putida) (80) may provide very new insights in how bacteria accomplish recognition and expression regulation of aromatic degradation pathways. In the end and from an environmental perspective, it will again be necessary to refocus on the strategies used by bacteria to recognize and utilize aromatic compounds (or other polluting chemicals) at very low concentration ranges (micromolar or lower). Although in the laboratory we like to feed our pet microorganisms with much higher doses (millimolar), the organisms in their natural environments have to cope with very low carbon compound fluxes or (at best) fluctuating levels. Are regulatory proteins sufficiently sensitive to react and mediate specific catabolic gene expression under these conditions, or are the organisms using alternative strategies to metabolize low carbon fluxes? Gene expression data for the 2,4-dichlorophenoxyacetate degradation pathway in R. eutropha, for example, have suggested that only very slight bursts of transcription activation take place at 10 $\mu$M 2,4-dichlorophenoxyacetate (127). Similar experiments with S. paucimobilis have not shown any measurable gene induction of the lin pathway at 3 $\mu$M lindane (M. Suar, J. R. van der Meer, K. Lawlor, C. Holliger, and R. Lal, submitted for publication) but, rather, a low constitutive expression. Interestingly, a very clear difference exists between the apparent effector binding affinity of heavy-metal-responsive regulatory proteins (nano- to micromolar range) and those responsive to aromatic compounds (usually micromolar) (95). Obtaining a better view of the differences in the molecular binding affinities of the effector compounds to the regulator and the role of DNA binding in the affinity of effector binding may help to resolve this apparent mystery. This will also be an important challenge for the application of bacterial regulatory systems for the construction of environmental quality sensors, which should have sensitivities below micromolar concentrations.

ACKNOWLEDGMENTS

We thank Victor de Lorenzo for critically reading the manuscript. We also gratefully acknowledge the many detailed suggestions by two unknown reviewers, which have helped to improve the manuscript.

Part of this work was financed under the EU 5th Framework Program, BIOCARTE contract QLK3-CT-2002-01923, and by the Swiss Federal Office for Education and Science.

REFERENCES


207. Reineke, W., and H.-J. Knockmann. 1988. Microbial degradation of halo-


