Aromatic metabolism versus carbon availability: the regulatory network that controls catabolism of less-preferred carbon sources in *Escherichia coli*

María A. Prieto *, Beatriz Galán, Begoña Torres ¹, Abel Ferrández ², Cristina Fernández, Baltasar Miñambres ³, José L. García, Eduardo Díaz

Department of Molecular Microbiology, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain

Received 18 February 2004; received in revised form 13 April 2004; accepted 13 April 2004

First published online 7 June 2004

Abstract

The current knowledge on the genetics and biochemistry of the catabolism of aromatic compounds in *Escherichia coli* settles the basis to consider these pathways as a model system to study the complex molecular mechanisms that control the expression of the genes involved in the metabolism of less-preferred carbon sources in this paradigmatic organism. Two different levels of regulation are reviewed: (i) the specific regulatory mechanisms that drive the expression of the catabolic genes when the cognate inducer, i.e., the substrate of the pathway or an intermediate metabolite, is available, and (ii) the global or superimposed regulation that adjust the expression of the catabolic clusters to the general physiological status of the cell.

© 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Catabolism; Aromatic compounds; *Escherichia coli*; Transcriptional regulation; Catabolic pathways; Catabolite repression; CRP; Less-preferred carbon sources

Contents

1. Introduction ......................................................... 504
2. Specific regulatory elements that control the catabolism of aromatic compounds in *E. coli* .
   2.1. The transcriptional activators ................................ 505
   2.2. The transcriptional repressors ............................... 510
3. Superimposed regulation of the aromatic catabolic pathways in *E. coli* .
   3.1. The hca cluster .............................................. 513
   3.2. The mao cluster ............................................ 513
   3.3. The hpa cluster ............................................ 513
   3.4. The mhp cluster .......................................... 514
   3.5. The pau cluster ........................................... 514
   3.6. The pac gene ............................................... 515

[Corresponding author. Tel.: +34-918-373112; fax: +34-915-360432. E-mail address: auxi@cib.csic.es (M.A. Prieto).
1 Present address: Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Campus de Cantoblanco, Madrid 28049, Spain.
2 Present address: DSM Nutritional Products, Biotechnology Department, Building 203/112C, P.O. Box 3255 CH-4002 Basel, Switzerland.
3 Present address: Estación Agrícola Experimental, Consejo Superior de Investigaciones Científicas, Finca Marzanas, Grulleros 24346, León, Spain.

0168-6445/$22.00 © 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.
1. Introduction

*Escherichia coli*, a paradigmatic microorganism in terms of its biochemistry and its genetics, had mainly been regarded as an inhabitant of the warm-blooded animal gut, as well as a pathogen of the enteric, urinary, pulmonary, and nervous systems [1,2]. However, a careful analysis of the ecology of this bacterium revealed that *E. coli* may easily colonize different extracorporeal habitats such as soil, water, sediment, food and it has been also found inside soil protozoa [3,4]. In fact, *E. coli* is presently considered as a highly adaptable microorganism with an extensive repertoire of metabolic and regulatory genes that facilitate the colonization of widely different aerobic and anaerobic environments [3,5]. Either as a host-dependent bacterium or as a free-living organism, *E. coli* has to cope with considerable fluctuations in the availability of nutrients and, therefore, the genes involved in the catabolic machinery are subjected to various types of physiological controls that adjust their expression rates to the environmental conditions [6]. Specific regulatory proteins and regulated promoters are the first key elements that allow catabolic operons to be transcribed only when required and at levels sufficient to guarantee an adequate metabolic return when the particular substrate is abundant and can serve as a nutrient source. Additionally, global regulatory circuits are found superimposed to the specific regulatory mechanisms allowing together a co-ordinated response to the physiological and metabolic state of the bacteria [7,8]. In this study, we address the strategies utilized by *E. coli* to control the expression of the genes responsible of the carbon catabolic pathways that funnel to central metabolism. As a model system of less-preferred carbon source metabolism, we have focused our interest on the ability of *E. coli* to respond to some aromatic compounds (see review in [9]).

Aromatic compounds are highly abundant in soil, water and food and, therefore, they constitute a natural carbon source for *E. coli* in its extraintestinal habitat. Although it is still unknown the complete set of substrates that *E. coli* encounters in the large intestine and the pathways that confer the metabolic advantage to compete with bacteria that share the same habitat, it is likely that aromatic compounds can also be a frequent carbon source for *E. coli* in the animal gut [9]. Ecological and phylogenetic analyses have shown intraspecies variation in the ability of *E. coli* to metabolise different aromatic acids which might reflect a way to increase the fitness and expand the ecological niches of individual *E. coli* cells [10]. All *E. coli* strains tested so far grow using phenylpropionic acid (PP), 3-hydroxyphenylpropionic acid (3HPP), or 3-hydroxycinnamic acid (3HCl) as the sole carbon and energy sources. However, while *E. coli* W is also able to grow on phenylacetic acid (PA), 4-hydroxyphenylacetic acid (4HPA) and 3-hydroxyphenylacetic acid (3HPA), *E. coli* K12 grows on PA but not on 4HPA nor 3HPA. On the contrary, *E. coli* B and *E. coli* C grow on 4HPA or 3HPA but not on PA. None of the tested strains were able to grow on 2HPA, cinnamic acid (Cl) and its 2- or 4-hydroxy derivatives, nor the 2- or 4-hydroxy derivatives of PP [10]. Our understanding of the utilization of these compounds by *E. coli* has leapt forward in the recent years with the molecular characterization of the gene clusters responsible for degradation of 4HPA and 3HPA (hpa) [11], 3HPH and 3HCl (mhp) [12], PP (hca) [13,14], PA (paa) [15], and some aromatic amines (mao) [16,17]. All these gene clusters are structured in transcriptional units or operons that respond to at least one specific regulatory protein [9]. The substrate-dependent expression of catabolic genes is, in turn, subdued to the overall carbon metabolic and energetic status of the cells (superimposed regulation). Fig. 1 shows the gene clusters encoding the aromatic catabolic pathways and their localization in the *E. coli* chromosome. The specific regulators of each transcription unit as well as some global regulators that sense the physiological state of the cells and ensure a productive regulatory network are also shown. The data presented in this work about the PA and 4HPA degradation pathways were performed with the *paa* and *hpa* clusters of *E. coli* W. The *mhp*, *mao* and *hca* clusters of *E. coli* K12 were used to unravel the regulatory systems of the 3HPP, aromatic amines and PP catabolic routes.

2. Specific regulatory elements that control the catabolism of aromatic compounds in *E. coli*

Four transcriptional activators (HpaA, MaoB, MhpR and HcaR) and two repressors (HpaR and PaaX) drive the specific regulation of the aromatic catabolic pathways in *E. coli*. All these proteins have their own type of inducer molecule. Moreover, the primary structure and, particularly, the DNA binding motifs of the regulators differ among them, suggesting that they
belong to different families of regulatory proteins with the exception of the HpaA and MaoB proteins that are members of the same family. In addition, this global analysis indicates a wide diversity and a divergent evolution of the regulatory systems that control the catabolism of aromatic compounds in *E. coli* [9]. In the following section we will review the genetic and biochemical features of the proteins that drive the substrate-responsive transcription of aromatic compound biodegradative operons in *E. coli*.

### 2.1. The transcriptional activators

#### 2.1.1. The HpaA protein

The HpaA protein (295 aa) is part of the regulatory system of the *hpa* cluster (Figs. 1 and 2). This transcriptional activator controls the *Pbc* promoter for the expression of the *hpaBC* genes which encode the two-component HPA monooxygenase [11,18]. The *hpaA* gene is located upstream of the *hpaBC* operon of *E. coli* W and it is transcribed in the same direction than the catabolic *hpa* genes [18,19] (Fig. 2). The expression of *hpaA* is controlled by at least two tandem promoters, *P*$_{A1}$ and *P*$_{A2}$, that are located within the *hpaX* coding region at 107-bp and 60-bp, respectively, upstream of the translational start codon of *hpaA* [18]. In addition, since the genes *hpaX* and *hpaA* appear to form an operon, the putative promoter of *hpaX* (*P*$_X$) (Fig. 2), may also contribute to the expression of *hpaA*. Moreover, the expression of *hpaA* is also controlled by global transcriptional regulators as shown below.

The HpaA protein is a member of the XylS/AraC family of regulators [19,20]. Proteins of this family consist of a N-terminal non-conserved domain, which seems to be involved in effector/signal recognition and dimerization, and a conserved C-terminal domain, characterized by significant amino acid sequence similarity extending over a 100 residue stretch that contains the characteristic DNA-binding domain of this family. The C-terminal end of HpaA comprises two helix-turn-helix (HTH) DNA-binding motifs, but only the second one is conserved in all proteins of the family [21]. The proposed second HTH motif of HpaA (Val$_{257}$ to Gly$_{281}$) fits the signature pattern of AraC proteins much better than the first (Trp$_{206}$ to Pro$_{234}$). A short highly conserved region found outside the second HTH motif and towards the C-terminal end of the AraC-like proteins [22] is also present in HpaA. According to the characteristics of the XylS/AraC family of regulators, the N-terminal and central regions of the HpaA protein are involved in binding to the effector molecule [23]. In this sense, HpaA shows the highest similarities with regulators that control the catabolism of structural analogs of 4HPA, such as the PobR proteins that activate gene expression for 4-hydroxybenzoate degradation in *Pseudomonas aeruginosa*, *P. putida*, *Rhizobium leguminosarum*, and *Agrobacterium tumefaciens* [8].

Although the HpaA operator within the *Pbc* promoter has not been experimentally described, two direct repeats with similar characteristics to those binding other XylS/AraC family members have been identified upstream of the −35 box of this promoter. 4HPA is the best inducer of HpaA, being the induction produced by 3HPA a 70% of that reached with 4HPA. Despite PA not being a substrate of the HPA hydroxylase, this compound is capable of a low but detectable HpaA-mediated activation of the *Pbc* promoter (35% of that

---

**Fig. 1. Schematic representation of the gene clusters encoding the catabolic pathways for the aerobic degradation of aromatic compounds in *E. coli*.** The gene clusters *mhp*, *mao*, *paa*, *hca*, *hpa*$_u$ (hpa-upper cluster) and *hpa$_m$* (hpa meta cluster) and *pae* gene are indicated by white blocks. The location of the clusters refers to the *E. coli* K-12 linkage map (the *hpa* cluster and the *pae* gene are present in *E. coli* W and absent in *E. coli* K-12). The global regulatory factors involved in the superimposed regulatory systems are indicated at the top. CRP, cAMP receptor protein; IHF, integration host factor. The specific regulatory proteins, transcriptional activators (filled symbols) and repressors (empty symbols), are indicated at the bottom. + and −, mean transcriptional activation and repression, respectively.
achieved with 4HPA). Interestingly, some analogues of 4HPA, such as 2HPA, 3,4-dihydroxyphenylacetic acid (3,4HPA), and other structurally related aromatic compounds, are not able to increase the expression driven by the \( P_{BC} \) promoter [18]. Unravelling the molecular basis of the induction effect of PA on the \( P_{BC} \) promoter and its putative physiological role will require further research.

2.1.2. The MaoB protein

The MaoB protein (301 aa) is a transcriptional regulator of the \( maoA \) and \( padA \) genes that encode the monoamine oxidase and phenylacetaldehyde (PAL)-dehydrogenase, respectively, involved in the catabolism of aromatic amines in \( E. coli \) (Figs. 1 and 3) [16,17]. The \( padA \) gene is transcribed in the opposite direction to that of \( maoA \) and \( maoB \), indicating that these three genes do not form an operon (Fig. 3). MaoB is a member of the AraC family of regulators, sharing only 22% amino acid sequence identity with HpaA, the other AraC-like regulator described above [9].

There is a remarkable regulatory and metabolic link between the catabolic pathway for aromatic amines and those for structurally related aromatic acids such as PA and 4HPA (Fig. 3). Thus, while the inducer molecules of MaoB are tyramine and 2-phenylethylamine (PEA), the transformation of these compounds into 4HPA and PA, respectively, causes also the induction of the \( hpa \) and \( paa \) clusters (Figs. 2–4). Formation of 3,4HPA from dopamine induces also the \( hpa \) cluster (Fig. 3) (see below) [9]. The fact that the \( mao \) genes are located at one end of the \( paa \) cluster in the chromosome of \( E. coli \) (Figs. 1 and 3) suggests that clustering of metabolically related genes, like \( mao \) and \( paa \) genes, might be the result of evolutionary selective forces. In this sense, the \( mao \) cluster can be considered as part of the PA-CoA catabolon, defining catabolon as a complex functional unit integrated by different catabolic pathways, which are, or could be, co-ordinately regulated, and that catalyse the transformation of structurally related compounds into a common catabolite [24] (see below).

2.1.3. The HcaR protein

The \( hcaR \) gene encodes a protein (296 aa) that belongs to the LysR-type transcriptional regulators

---

Fig. 2. Regulation and biochemistry of the \( hpa \) cluster encoding the pathway for the catabolism of HPA (4HPA and 3HPA) in \( E. coli \). (a) The organisation of the catabolic (\( hpaBC/\text{hpa}GEDFHI \)), transport (\( \text{hpaX} \)), and regulatory (\( \text{hpaA} \) and \( \text{hpaR} \)) genes of the \( hpa \) cluster as well as their regulation by \( \text{HpaA} \) and \( \text{HpaR} \) and the global regulators CRP and IHF are represented. Thick arrows indicate the directions of gene transcription. \( P_R, P_G, P_X, P_{A1}, P_{A2}, \) and \( P_{BC} \) are promoter regions. The white square means the active form of the HpaR repressor; the black diamond means the inactive form of HpaA activator. The white circles represent the inducers. + and –, indicate transcriptional activation and repression, respectively. (b) The HPA transport protein (HpaX) is represented by a thick arrow. Ext and Int, indicate outside and inside the cell, respectively. A brief scheme of the biochemistry of the HPA pathway including the first step for the transformation of HPA in 3,4HPA (3,4-dihydroxyphenylacetic acid) by the action of the HpaBC monoxygenase and the final products of the pathway are shown.
(LTTRs) family [8,13,14]. Whereas the N-terminal domain of HcaR contains a HTH motif, the C-terminal domain contains a putative multimerization motif that characterizes LTTRs [25]. The hcaR regulatory gene is transcribed divergently from the corresponding catabolic genes (hcaEFCBD) (Fig. 5) [13]. The corresponding Pr promoter is still unknown but it should be located near or overlapping the Pe regulated promoter of the hca catabolic operon (Fig. 5). The 135-bp intergenic region of the divergently transcribed hcaR and hcaE genes contains a sequence (TAG-N7-CTA) that matches the consensus binding motif of LTTRs [25] and that is located 85 nucleotides upstream of the hcaE translation start site.

By expressing the hca genes in Salmonella typhimurium LT-2, the positive regulatory role of HcaR in the hca cluster and the identification of PP and CI as inducers have been demonstrated [13].

2.1.4. The MhpR protein

The specific regulator of the mhp cluster for the degradation of 3HPP and 3HCl, MhpR (277 aa), is a member of the IclR family of regulators [8]. Usually, the regulators of this family are divergently transcribed...
with respect to the genes that they control, and the corresponding promoters are superimposed in a short DNA segment [8]. A similar genetic arrangement has been observed within the mhp cluster, being the mhpR regulatory gene oppositely oriented to the mhp catabolic genes (Fig. 5). The promoter of mhpR (Pr) and that of the mhp catabolic genes (Pa) are located within the 178-bp intergenic region between the translational start sites of the divergently transcribed mhpR and mphA genes [26]. The transcription initiation site in the Pa promoter was mapped 91 nucleotides upstream of the ATG translation initiation codon of the mphA gene, and putative –10 (TATACT) and –35 (TTG-TAG) boxes typical of 70-dependent promoters were identified. A major transcription initiation site was located for the Pr promoter 107 nucleotides upstream of the ATG translation initiation codon of the mhpR gene. Two putative –10 boxes (AATGAT, TGTAAA) and the absence of consensus –35 regions characterize the Pr promoter. The Pr and Pa promoters in the mhp cluster show, therefore, a peculiar arrangement leading to mRNA transcripts whose first 10 nucleotides are complementary [26].

By using Pr-lacZ and Pa-lacZ translational fusions and gel retardation assays, it has been demonstrated that the mhpR gene product behaves as a 3HPP-dependent activator of the Pa promoter, being the expression from Pr constitutive and MhpR-independent [26]. In this sense, MhpR shows a distinct regulatory feature when compared with other IclR-type regulators of aromatic catabolic pathways, e.g. PcaR from P. putida and PobR and PcaU from Acinetobacter sp. ADP1, all of which act as transcriptional activators of the cognate catabolic genes but they behave as repressors of their own expression [27–29]. Moreover, it is known that some IclR-type regulators, such as PcaU, act on the same promoter both as transcriptional activators, in the presence of the cognate inducer, and repressors, in the absence of the inducer molecule [30]. However, this dual regulatory role was not observed with the MhpR activator and the Pa promoter because no β-galactosidase activity was observed in E. coli cells carrying the Pa-lacZ fusion but lacking the mhpR gene [26].

DNase I footprinting experiments and mutational analysis mapped a MhpR-protected region (OR), centred at position –58 with respect to the Pa transcription start site, which is indispensable for MhpR binding and in vivo activation of the Pa promoter. The OR sequence corresponds to a 17-bp imperfect palindromic motif, GGTGACCTGTTGACACA, with its pseudo-dyad axis through the central T base (underlined) defining two 8-bp half-sites. The interaction of the regulatory protein with its target DNA does not require the Pr
promoter region. Although the MhpR protein binds to OR even in the absence of the 3HPP inducer, increasing concentrations of 3HPP enhanced retardation of the DNA probe, which indicates that even though 3HPP is not indispensable for binding of MhpR to the mhp intergenic region in vitro, it facilitates such interaction [26].

A noticeable peculiarity of the regulatory system driven by MhpR is that this protein seems to be essential for recruiting a second activator, the global cAMP receptor protein (CRP) regulator (see below), to the cognate Pa promoter. This peculiar synergistic transcription activation is a mechanism widely used in eukaryotes, but few examples have been reported in prokaryotes [31], and it has not been observed yet in other aromatic catabolic pathways. The MhpR activator becomes, thus, the first member of the IclR family of transcriptional regulators endowed with such distinctive recruiting skill.

The catabolism of 3HPP and PP in E. coli converges at the common intermediate 2,3-dihydroxyphenylpropionic acid (DHPP) by the action of the MhpA monooxygenase and the HcaEFCD/HcaB dioxygenase/dihydrodiol-dehydrogenase, respectively, as shown.

**Fig. 5.** Regulation and biochemistry of the mhp and hca clusters encoding the pathways for the catabolism of 3HPP and PP, respectively, in E. coli. (a) The organisation of the catabolic (mhpABCDEF and hcaEFCBD), transport (mhpT) and regulatory (mhpR and hcaR) genes as well as their regulation by MhpR and HcaR and the global regulator CRP are represented. The thick arrows indicate the directions of gene transcription. P_k, P_ß and P_ß are promoter regions. The black circle means the inactive form of the MhpR activator. The black pentagon means the inactive form of the HcaR activator. The white circle represents the inducer. +, indicates transcriptional activation. The double-head arrow means synergistic transcription activation by MhpR and CRP. (b) Scheme of the biochemistry of the 3HPP and PP catabolic pathways. The 3HPP transport protein (MhpT) is represented by a thick arrow. Ext and Int, indicate outside and inside the cell, respectively. A brief scheme of the pathways including the final products and the first step for the transformation of 3HPP and PP into the common intermediate DHPP (2,3-dihydroxyphenylpropionic acid) by the action of the MhpA monooxygenase and the HcaEFCD/HcaB dioxygenase/dihydrodiol-dehydrogenase, respectively, is shown.
ionic acid (DHPP) (Fig. 5). Interestingly, the 3HPP and the PP degradation branches of the DHPP meta-cleavage pathway in *E. coli* are controlled by members of different families of regulatory proteins, i.e., the MhpR (IclR family) and HcaR (LysR family) activators, respectively (Fig. 5). Although the physiological meaning of such branch specificity for the regulatory proteins remains still unknown, it might facilitate the interactions between branches and could reflect a kind of hierarchy of pathway utilization in *E. coli* when this bacterium faces 3HPP and PP simultaneously.

2.2. The transcriptional repressors

2.2.1. The HpaR protein

The *hpaR* gene encodes the transcriptional repressor of the 3,4HPA dehydrogenative cluster (*hpaGEDFHII, meta* operon) of *E. coli* W. It is located upstream and divergently oriented with respect to this operon (Figs. 1 and 2) [11,18,32]. Based on sequence comparison analysis, the HpaR protein (148 aa) has been proposed to be a transcriptional regulator belonging to the MarR family [18]. Recently, it was shown that HpaR negatively regulates not only the expression of the *hpa meta* operon but also its own expression. This finding agrees with previous observations reported for other MarR-like proteins and may represent a common feature of this family of regulators [32]. HpaR binds specifically to the OPR1 operator from *Pg* (hpa-meta promoter) and to OPR2 operator from *Pr* (hpaR promoter). Although HpaR binds co-operatively to both operators, OPR1 remains the first operator to be occupied by the HpaR repressor. These data indicate that as soon as the inducers are depleted from the medium, HpaR binds to OPR1 to shut-off the expression of the catabolic enzymes before repressing its own expression by binding to OPR2. DNase I footprinting experiments revealed that both HpaR operators comprise a 27-bp region containing two 9-bp palindromic sequences separated by 4-bp (AAAGTAATCATTAAACATATTAAATGATT for OPR1, and TTCAAAATCATTAAATAGAAACAGTT for OPR2; inverted repeats are underlined). A similar structural design has been described for other promoters controlled by MarR-like regulators [33], and it is consistent with the dimeric structure proposed for such regulators where each subunit binds to one of the two inverted half sites of the operator [34–36]. Analysis of the 3D-structures of two members of the MarR family, i.e., MexR and MarR, suggests that despite a conserved global architecture, the regulatory mechanisms operated by these two regulators are different and they need to be clarified [37].

As the operator OPR1 is centred at position +2 of *Pg*, i.e., overlapping the transcriptional start site, it could be presumed that HpaR represses transcription from *Pg* by a mechanism based on steric hindrance, that is, by inhibiting binding of the RNA polymerase (RNAP) to the target promoter. However, experimental data suggest a repression mechanism different to the classical steric hindrance. Thus, HpaR might repress transcription from *Pg* by a road-block mechanism that prevents transcription initiation through an upstream displacement of the RNAP from its functional promoter binding site [32]. On the other hand, footprinting analyses have demonstrated that the second HpaR operator, OPR2, is centred at position +47 relative to the *Pr* transcriptional start site. In this case, the location of OPR2 suggests that the repression effect mediated by HpaR might be produced by inhibiting the transcription elongation process from *Pr* [32]. In addition, we have observed that HpaR binding to OPR2 is co-operative with binding to OPR1. Since these HpaR binding experiments were performed in the absence of other regulatory proteins that form part of the *hpa* regulatory system, such as CRP or the integration host factor (IHF) (see below), co-operativity indicates that binding of HpaR to OPR1 stabilizes binding to OPR2 by protein–protein interactions. The dimerization (or oligomerization) of the HpaR dimers cannot be excluded [32]. It was shown by in vivo and in vitro experiments that the activation of *Pg* and *Pr* promoters requires the presence of 4HPA, 3HPA or 3,4HPA as inducers, and that these effectors interact directly with the HpaR repressor. However, the presence of the inducers, even at high (millimolar) concentrations and under different ionic strength conditions, had little effect on the binding of HpaR to the cognate promoters [32].

As mentioned before, MarR, the prototype of the MarR family, is a dimer which contains a typical winged-helix DNA binding motif in each monomer. Up to now, no 3D-structure of a DNA–protein complex is available for a protein of the MarR family. The two recognition helices are supposed to bind two adjacent major grooves of the target DNA whereas the wings might well be positioned to make minor groove or phosphate backbone contacts to the distal parts of the inverted repeat [34]. The 3D-structure of HpaR has been modelled using the crystal co-ordinates of MarR and MexR [32]. Since MarR was crystallized in the presence of the inhibitor salicylate, and the crystal of MexR does not contain any ligand, both 3D-structures were used as models for the ligand-bound and ligand-free conformations, respectively, of HpaR (Fig. 6). It has been proposed that HpaR might bind two molecules of 4HPA per monomer, one between the DNA recognition helix and the “wing” structure (corresponding to the SAL-A site in MarR), and the other between the recognition helix and the a2 helix (corresponding to the SAL-B site in MarR). In both cases, the ligands would be solvent-exposed [32] (Fig. 6). Since binding of salicylate to the SAL-A site in MarR does not cause a significant rearrangement in this re-
region compared to the structure of the ligand-free MexR crystal, it was predicted that binding of 4HPA should also have minimal effects on the structure of HpaR. On the other hand, binding of the second ligand molecule to the SAL-B site of MarR appears to induce a considerable conformational change at the $\alpha_1$ and $\alpha_6$ helices and, to a lesser degree, at the $\alpha_2$ helix when compared to the ligand free MexR structure [35]. Therefore, it was postulated that similar conformational changes occur in HpaR after 4HPA binding (Fig. 6). As $\alpha_1$ and $\alpha_6$ helices are the structures that most contribute to the dimerization surface both in MarR and MexR [34,35], the movement of these helices induced by 4HPA binding would have a dramatic effect on the final configuration of the HpaR dimer and, especially, on the spatial disposition of the DNA-recognition helices. Since the distance between operators OP1 and OP2 is 200 bp, which corresponds to 19 turns of B-DNA helix (assuming a pitch of 10.5 bp per turn), the two dimers of HpaR should be positioned on the same face of the DNA helix. Our model proposes that the two dimers of HpaR bound at OP1 and OP2 could interact to each other leading to the formation of an oligomer which could generate a repression loop. It is also likely that whereas the global regulators CRP and IHF might help the stabilization of the repression loop through induction of DNA bending, 4HPA might disturb the interactions between the two HpaR dimers as well as those within the HpaR–DNA complexes [32].

2.2.2. The PaaX protein

The paaX gene forms part of the paa cluster for PA degradation in E. coli (Figs. 1 and 4). The PaaX protein (316 aa) contains a stretch of 25 residues (aa 39–64) that shares similarity with the HTH motif of transcriptional regulators of the GntR family [38]. Three promoters, Pz, Pa, and Px, driven the expression of genes paaZ, paaABCDEFGHIJK, and paaXY, respectively (Fig. 4), have been identified in the paa cluster [15]. By using genetic and biochemical approaches, the PaaX protein was shown to act as a transcriptional repressor of the catabolic Pa and Pz promoters [15] as well as of the Px promoter (Fernández C., Miñambres, B., Díaz, E. and García, J.L., unpublished results). The region protected by the PaaX repressor in DNaseI footprinting assays is about 50-bp and contains a conserved 15-bp imperfect palindromic sequence motif (WWRTGATTCLGYWT) that was shown, through mutational analyses, to be indispensable for PaaX binding and repression [38]. The region protected by PaaX is located immediately downstream of the transcription start site within the Pa promoter and it spans the +1 and the −10 region in the Pz promoter, probably revealing a different mechanism of transcriptional repression in each of these two promoters. PA-CoA, but not PA, specifically inhibited binding of PaaX to the target sequences, confirming the first intermediate of the pathway (PA-CoA) as the true inducer [38]. Since PA-CoA is a common metabolite in the catabolism of several other aromatic compounds such as styrene, trans-styrylacetic acid, PEA, phenylacetaldehyde, phenylacetyl amides, phenylacetyl esters and n-phenyl-alkanoic acids containing an even number of carbon atoms, in different bacteria, the whole metabolic system has been named PA-CoA catabolon [24].

Recently, a new role for the PaaX protein in the aromatic metabolism of E. coli has been discovered. It was shown that the PaaX repressor links the catabolism of aromatic compounds with the metabolism of penicillins [39]. Penicillin G acylase (PAC) is one of the most important enzymes used worldwide at industrial scale for the semisynthesis of $\beta$-lactam antibiotics [40]. PAC is able to hydrolyze penicillin G as well as esters and amides of PA and other aromatic acids such as 4HPA [41–45]. This hydrolytic activity and the fact that PAC...
biosynthesis was shown to be activated by PA had favoured the thought of PAC as a scavenger enzyme for natural phenylacetylated compounds. The synthesis of PAC is probably one of the most complex processes so far described for bacterial proteins since it is subject of both transcriptional and translational regulatory controls (see below). However, the regulatory proteins involved in the PA-mediated pac induction remained unknown until the recent finding that PaaX controls negatively the transcription of the pac gene [39] (Fig. 4). This finding not only settles the basis for clarifying the puzzling data reported previously on this complex regulatory system (see below), but definitively supports the implication of PAC in the PA-CoA catabolon. The pac expression depends on the synthesis of PA-CoA, and a palindromic sequence proposed as the PaaX operator is located upstream of the −35 box of the Ppac promoter overlapping a CRP binding site [39]. This unusual location of the operator within the Ppac promoter suggests a novel mechanism of repression by the PaaX protein that will require further studies. Taking into account the role of PAC as a PA scavenger enzyme and the fact that the promoter of the pac gene and those of the catabolic operons of the paa cluster are controlled by the same regulator, the PaaX protein, one might expect that these genes had evolved in a co-ordinated way. However, genomic analyses revealed that the pac gene is located far from the paa cluster in E. coli W (Figs. 1 and 4) and it is absent in the genome of other E. coli strains that contain the paa cluster. These observations suggest that the pac gene has been acquired by E. coli W to encode a peripheral pathway for funnelling PA esters and amides to a resident PA central pathway. It is tempting to speculate that after being recruited by E. coli W, the pac gene was further subjugated to the discipline of PaaX repression by evoking the PaaX palindromic sequences (operator) within the Ppac promoter [39].

3. Superimposed regulation of the aromatic catabolic pathways in E. coli

Bacteria thriving in the environment face with a range of physical and chemical signals that need to be processed to achieve a positive or negative physiological response. For instance, bacteria usually are confronted with alternative carbon sources and they need to decide which of them will be preferentially consumed before metabolising less preferred substrates to guarantee, therefore, a satisfactory metabolic return. To achieve this goal, bacteria have developed a physiological control that governs and adjust the specific regulation of catabolic operons to the physiological and metabolic state of the cells [7]. The classic example of superimposed regulation is the repression of the synthesis of many catabolic enzymes in enteric bacteria by the presence of glucose in the culture medium [46]. This catabolite repression phenomenon, termed the “glucose effect”, regulates the transcription of catabolic operons by modulating transcription factors availability. The prototype system, whose molecular basis has been well characterized, is the glucose–lactose diauxie for the catabolism of lactose in E. coli [47,48]. Firstly, glucose increases the concentration of the inducer-free lac repressor (LacI) by preventing the entry of lactose into the cell (inducer exclusion) through a functional phospho-enolpyruvate-sugar phosphotransferase system (PTS) and the direct involvement of the LacY permease of the lac operon [48]. Secondly, glucose lowers the levels of the CRP–cAMP complex by reducing cAMP levels and decreasing also the rate of transcription initiation at the crp promoter [49]. The glucose effect also extends to the metabolism of aromatic substrates in E. coli, and the molecular basis of this response is being elucidated and it will be presented in this section.

As mentioned above, the uptake of substrates inside E. coli cells could be a major factor of the superimposed regulation. E. coli contains specific transport proteins for some aromatic compounds. Two permeases for the uptake of 4HPA/3HPA (HpaX) and 3HPP/3HCI (MhpT) have been genetically and biochemically characterized in E. coli [9,50]. Additionally, two potential permeases for the uptake of PP (HcaT) and PA can be also identified through in silico analysis of the E. coli genome [9]. The HpaX protein (458 aa) belongs to the major facilitator superfamily (MFS) of transport proteins and it is the best characterized aromatic permease from E. coli [50]. The HpaX transporter was selected to check whether aromatic permeases might be involved in the regulatory systems that prevent utilization of less preferred carbon sources when favourite substrates, such as glucose, are present in the culture medium. It was shown that the glucose effect on the hpa-meta operon was independent of the presence of the HpaX permease [6]. On the other hand, aromatic compounds, and particularly the lipophilic weak aromatic acids such as PA, PP and their hydroxylated derivatives, can enter the cells by passive diffusion when present at high (milimolar) concentrations. All these findings point to the fact that aromatic permeases instead of being involved in catabolite repression of the metabolism of aromatic compounds in E. coli, they increase the efficiency and rate of substrate acquisition and thus may impart a growth advantage in natural environments where these compounds are present at low (micromolar) concentrations [6,50]. The correct functionality of these transport proteins obviously mediates the fine sense of the specific transcriptional regulators [6].

As mentioned below, CRP and cAMP [46] are the major key factors driving catabolite repression of the aromatic catabolic pathways in E. coli. Nevertheless,
other global regulatory factors contribute also to the catabolite repression phenomenon and they are essential for an efficient mineralization of these less preferred carbon substrates when preferred nutrients are not available to E. coli cells [8].

3.1. The hca cluster

It has been reported that when E. coli is exposed to mixtures of glucose (an easily degradable substrate) and PP in batch culture, cells utilised the two compounds sequentially, i.e., the utilization of PP was immediately repressed by glucose, regardless of whether glucose was present in the initial substrate mixture or it was pulsed to cells growing with PP alone [51]. However, PP and glucose were consumed simultaneously in carbon-limited continuous culture resembling the environmental conditions where growth is mostly carbon limited [52,53]. The molecular mechanisms underlying such carbon-dependent hca gene expression remain to be elucidated.

3.2. The mao cluster

The expression of the regulatory maoB gene (see above) is subject to catabolite repression by glucose [16]. By using cya−E. coli mutant strains that express a maoB::lacZ fusion, it was shown that the catabolite repression by glucose was mediated by the cAMP−CRP complex on the PmaoB promoter (Figs. 1 and 3), and a potential CRP-binding site was identified in such promoter region. Thus, the catabolite repression in the catabolism of aromatic amines may reflect the catabolite repression of the regulatory maoB gene rather than being a direct effect on the catabolic PmaoA promoter [9,16].

3.3. The hpa cluster

The regulatory apparatus of the hpa cluster is also subject to catabolite repression by glucose [16]. By using cya−E. coli mutant strains that express a maoB::lacZ fusion, it was shown that the catabolite repression by glucose was mediated by the cAMP−CRP complex on the PmaoB promoter (Figs. 1 and 3), and a potential CRP-binding site was identified in such promoter region. Thus, the catabolite repression in the catabolism of aromatic amines may reflect the catabolite repression of the regulatory maoB gene rather than being a direct effect on the catabolic PmaoA promoter [9,16].

Fig. 7. Role of acetic acid in the superimposed regulation mediated by glucose on the Pg promoter from the hpa meta-cleavage route. Scheme of the production, secretion and reabsorption of acetate during the metabolism of glucose. The acetate produced from the catabolism of glucose via the phosphotransacetylase-acetate kinase pathway (Pta-AcK pathway) (Fig. 7), generating energy and biosynthetic components via the tricarboxylic acid cycle and the glyoxylate shunt, respectively [55]. In this sense, the acetate excreted at the stationary phase acts as an overflow metabolite that can provide the energy to produce cAMP and to adapt the cells rapidly to the utilization of a new less-preferred carbon source such as the aromatic compounds [6]. Therefore, the rapid activation of the Pg promoter at the stationary phase of growth of E. coli W is not solely dependent on the depletion of glucose in the culture medium, but also on the acetate excreted and accumulated in the culture medium (Fig. 7) [6]. The strategy of reusing the excreted acetate to energize the cells provides a clear advantage in the transition from glucose to other carbon sources.
advantage when easy carbon sources are depleted and bacteria need to cope with less preferred substrates through pathways that have been acquired during evolution for starvation emergencies. Although it is evident that after glucose depletion the cells can use 4HPA as sole carbon and energy source, this adaptive process may require a long period of time when the concentration of 4HPA is low [6]. To reduce the length of the adaptive process and, therefore, increase the chance of survival under critical environmental conditions, the cells can reuse the secreted acetate obtaining the energy necessary to adapt rapidly to the catabolism of 4HPA [6].

A high affinity CRP binding site was identified at position −61.5 with respect to the transcription start site of Pg [6], a location that characterizes CRP-dependent activation of class I promoters [46]. On the other hand, an IHF binding site has been mapped around position −103 of Pg. Since IHF and CRP can bind simultaneously to the Pg promoter (Fig. 2) [6], IHF may also act as an activator of the system and might, synergistically with CRP to promote transcription from Pg thus involving an activation mechanism of class III promoters [46]. In fact, CRP and IHF binding sites are separated by 42-bp, i.e., four turns of B-DNA assuming a DNA helix repeat of 10.5-bp, which places in phase the IHF and CRP induced bends in the operator region and might explain the synergistic activation of Pg by CRP and IHF. The A-tracts located between the IHF and CRP binding sites could act as an UP element facilitating the interaction of the RNA polymerase α subunit C-terminal domain (αCTD) with the promoter region [56].

Preliminary data indicated that the hpa regulatory system in cells grown in rich medium was still more complex than previously envisioned. The Pg activity is down-regulated also in rich medium, mimicking the physiological repression described for the aromatic metabolism in Pseudomonads [57–61] but that had not been observed in enterobacteria so far. More interestingly, the repression effect of rich medium on Pg activity is independent of the HpaR specific regulator and it is controlled by an unknown factor(s) that does not involve any of the protein amino acids which form part of the culture medium (Galán, B., Kolb, A., García, J.L. and Prieto, M.A., unpublished results).

3.4. The mhp cluster

Superimposed to the specific MhpR-mediated regulation of the Pa promoter, it was observed a strict catabolite repression control of such promoter carried out by the CRP-cAMP complex. This allows expression of the mhp catabolic genes when glucose is not available and 3HPP is present in the medium (Fig. 5). Gel retardation assays performed with purified CRP in the presence of cAMP showed that CRP was not able to bind to the Pa–Pr DNA region (Fig. 5) unless MhpR was also present in the retardation assay, thus likely generating a ternary DNA–CRP–MhpR complex [26]. Sequence analysis of the Pa promoter revealed the existence of a potential CRP-binding site centred at position −95.5 relative to the transcription start point of Pa. However, the left half-site of such binding motif poorly matches with the consensus sequence for CRP-binding [26], which could explain the lack of interaction of CRP with the Pa promoter and the observation that MhpR is strictly necessary for binding of CRP to its target DNA. Such co-dependence upon two transcription activators is a mechanism widely used in eukaryotes, but few examples have been reported in prokaryotes [31]. MhpR becomes the first regulator of the IclR family that is indispensable for recruiting CRP to the cognate promoter, and no example of such co-dependence has yet been observed for other regulators of aromatic catabolic pathways. The A-tract located between the putative CRP-binding site and the MhpR-binding site [26] may act as an UP element facilitating the interaction of the RNA polymerase α C-terminal domain with the promoter region.

3.5. The paa cluster

When E. coli cells are grown in PA-containing minimal medium in the presence of glucose, gene expression from the Pa and Pz promoters is subject to carbon catabolite repression [38]. The operon-specific regulation mediated by the PaaX repressor is subordinated to a superimposed regulation mediated by global regulators such as CRP and IHF (Fig. 4). E. coli crp− strains failed to express the Pa-lacZ and Pz-lacZ translational fusions, indicating that CRP acts as an activator of the gene expression driven by the Pa and Pz promoters. Gel retardation assays confirmed the binding of the cAMP–CRP complex to the Pa promoter region. Since a potential CRP binding site was identified at position −61.5 with respect to the major transcription start site of Pa, this promoter might follow a CRP-dependent activation mechanism similar to that described for class I promoters [38,46]. Although CRP is also necessary for activity of the Pz promoter, no binding of CRP to Pz was observed in gel retardation assays, suggesting that CRP bound to Pa is able to activate the divergent Pz promoter. IHF also binds to the paaZ-paaA intergenic region and stimulates transcription from the Pa and Pz promoters (Fig. 4) [38]. Whether IHF-induced bending
of the paaZ–paaA intergenic region can bring together RNA-polymerase bound to Pz and CRP bound to Pa, is an interesting model that remains to be confirmed.

3.6. The pac gene

The synthesis of PAC protein from E. coli W constitutes a nice example of how different regulatory levels and mechanisms can be coupled to efficiently control peripheral routes of aromatic compounds. It has been mentioned above that the pac gene is repressed by the PaaX regulator but becomes transcribed as a monocis-
tronic unit when PA is present in the culture medium [63] (Figs. 1 and 4). It has been also described that metabolic carbohydrates and polyalcohols (glucose, fructose and glycerol) repressed pac transcription [64]. The repression by glucose appears to be a typical CRP-dependent process [65,66]. Involvement of IHF as an additional regulatory protein in pac transcription was also reported [66], but IHF does not bind directly to the Ppac promoter and, therefore, its effect on pac expression should be pleiotropic [39].

In addition to the transcriptional regulation of the pac gene, PAC synthesis is under a complex transla-
tional and post-translational control. Synthesis of an active PAC protein requires cell growth restricted to temperatures below 30 °C [67]. PAC belongs to a small group of proteolytically processed bacterial enzymes that become active by removing a linker peptide (30–50 aa) in the proenzyme (Fig. 8) [68]. In short, PAC is synthesized as an inactive cytoplasmic precursor of 93 kDa (preproPAC) which becomes catalytically active after its translocation to the periplasmic space with the consequent hydrolysis of the N-terminal signal peptide (proPAC). The proPAC protein suffers an autoproteo-
lytic process that releases an internal peptide and generates the active PAC enzyme, an heterodimer of α (23.8 kDa) and β (62.2 kDa) subunits (Fig. 8) [68].

In summary, the synthesis of PAC is the result of co-
ordinated transcriptional, translational and post-trans-
lational control, and represents an interesting model system to study both specific and superimposed regu-
lation along the complete gene expression scheme.

4. Concluding remarks

Even though E. coli is probably the best known and characterised organism, about 35% of its genome en-
codes unknown functions. To fill this gap, the study of the physiological phenomena that occur under non-
conventional growing conditions, such as the mineralization of aromatic compounds, becomes crucial. The current view on the metabolic abilities of E. coli to cope with aromatic compounds revealed this bacterium as a very useful model system to decipher biochemical, ge-
netic, evolutionary, and ecological aspects of this par-
ticular type of less-preferred carbon source metabolism. The aromatic catabolic pathways of E. coli appear to be particularly interesting because they present a wide structural and functional diversity of regulatory systems that control their expression. Up to now, four activators, HpaA, MaoB, MhpR and HcaR and two repressors, HpaR and PaaX, have been described as pathway-specific regulators that sense the inducer me-
tabolites (the aromatic compounds) and prepare the cell
machinery to use them as carbon and energy sources. Superimposed to the specific regulation, there is a more global control that adjusts the metabolic machinery of each particular catabolic pathway to the physiological status of the cell.

All specific regulators that control the aromatic catabolic pathways in *E. coli*, with the sole exception of HpaA and MaoB, belong to different families of regulatory proteins, and the study of some of them has given rise to unexpected findings. For instance, MhpR appears to be essential for recruiting the global CRP regulator to the catabolic promoter, a peculiar synergistic transcription activation mechanism widely used in eukaryotes, but very infrequent in prokaryotes. On the other hand, the highly strict catabolite repression mechanism mediated by CRP and IHF on the *Pg* promoter of the *hpa-meta* cluster is quite peculiar since this repression control does not require the presence of the specific HpaR repressor and/or the 4HPA permease (HpaX), thus excluding the involvement of a typical inducer exclusion mechanism. The behaviour of this catabolic promoter in cells growing in rich medium indicates that the *hpa* regulatory system is still more complex than previously envisioned, and it constitutes a fascinating model system to unravel a regulatory network governing the non-favourite carbon source metabolism.

The specific transport proteins of aromatic compounds, such as the HpaX permease for the uptake of 4HPA/3HPA, mediate the fine sense of the specific transcriptional regulators of the catabolic pathways and increase the efficiency and rate of substrate acquisition allowing to the cell the use of substrates present in the medium at very low concentrations.

The advent of high-throughput methods for analysis of gene expression (whole-genome DNA microarrays) and function (proteomic techniques) as well as advances in silico biology will allow to obtain a more complete picture of the complex mechanisms that control the expression of genes governing the *E. coli* metabolism. Analysis of the regulatory network underlying the catabolism of aromatic compounds in *E. coli* will certainly contribute to discover new genes/proteins involved in global regulation, thus increasing our current knowledge on the functionality of the genome of this paradigmatic microorganism. An interesting spin-off of this research is the use of the regulatory machinery for the development of biomonitoring devices, biosensors, and biosafety circuits for environmental control [8].

Acknowledgements

We are indebted to Jesús Sanz for the 3D model of the HpaR protein. This work was supported by the Comisión Interministerial de Ciencia y Tecnología (Grants ABM97-603-C02-02, BIO2003-05309-C04-02, GEN2001-4698-C05-02, and CICYT-P4-Optimización de la producción de cefalosporina) and by the EU grant QLRT-2001-02884.

References


