Regulation and Evolution of the Malonate and Propionate Catabolism in the Proteobacteria

I.A. Suvorova1, D.A. Ravcheev1,2, M.S. Gelfand1,3*

Institute for Information Transmission Problems (the Kharkevich Institute), RAS. Bolshoi Karetny pereulok. 19, Moscow, 127994, Russia1
Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, United States of America2
Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University. Vorobievy Gory 1-73, Moscow, 119992, Russia3

Running title: Malonate and propionate catabolism

* Corresponding author. Mailing address:
Institute for Information Transmission Problems RAS (the Kharkevich Institute), Bolshoi Karetny per. 19-1, Moscow, 127994, Russia. E-mail: gelfand@iitp.ru (M.S.G.).

Abstract

Bacteria catabolize malonate via two pathways, encoded by the mdc and mat genes. In various bacteria, transcription of these genes is controlled by the GntR-family transcription factors MatR/MdcY or/and a LysR-family transcription factor MdcR. Propionate is metabolised via the methylcitrate pathway comprising enzymes encoded by the prp and acn genes. PrpR, the Fis-family sigma54-dependent transcription factor, is known to be a transcriptional activator of the prp genes. Here, we report a detailed comparative genomic analysis of the malonate and propionate metabolism and its regulation in the Proteobacteria. We characterize genomic loci and gene regulation, identify binding motifs for four new TFs, and also new regulon members, in particular, TRAP transporters. We describe restructuring of the genomic loci and regulatory interactions during the evolution of the Proteobacteria.

Introduction

Malonate can be used as a carbon source by a variety of bacteria, such as Acinetobacter calcoaceticus, Klebsiella pneumoniae, Pseudomonas fluorescens and Pseudomonas putida [15]. It is a competitive inhibitor of succinate dehydrogenase. Malonate was found in symbiotic
legumes and in developing rat brains [15,16]. Malonate metabolism is essential in symbiotic nitrogen metabolism, since mutant bacteria with deleted malonate metabolic genes lose its symbiosis ability [13]. In addition, malonate metabolic genes have been used for generation of the industrial strain of Streptomycetes for the production of antibiotics [13].

Two groups of structural genes involved in the malonate metabolism have been characterized. The first group was described for *Rhizobium leguminosarum* [16]. It comprises three clustered malonate metabolic genes, *matA*, *matB* and *matC*, and the divergently transcribed gene of the malonate regulator, *matR* (Table 1). The MatR transcription factor (TF) belongs to the FadR subfamily of the GntR family.

Another system for malonate metabolism was described for *Acinetobacter calcoaceticus* KCCM 4090 [14,15]. This system is encoded by the operon of structural genes *mdcLMACDEGBH* and the divergently transcribed regulator gene, *mdcY* (Table 1). Like MatR, MdcY is also a FadR-family TF. Some Gammaproteobacteria have a different malonate transporter gene (compared to *A. calcoaceticus*), *mdcF* instead of *mdcLM* [19]. The gene organization of the malonate operon of *A. calcoaceticus* is similar to that of *K. pneumoniae* and *P. putida* [14,19], but these and many other Gammaproteobacteria have another TF gene adjacent to the malonate operon, encoding the MdcR transcription factor from the LysR family. MdcR activates expression of the *mdc* genes and represses its own transcription [14,19].

Propionate can also serve as a single carbon source for many bacteria. Its metabolism is strongly connected to the malonate metabolic pathway and the central metabolism, e.g. the TCA cycle (Fig. 1). Propionate is converted to pyruvate via the methylcitrate pathway, enzymes for which are encoded by the *prpBCDE* and *acnB(acnM)/acnD* genes [3,4]. AcnB is a bifunctional enzyme that also belongs to the TCA cycle and the glyoxylate pathway [3,4]. Propionate also participates in the citramalate cycle comprised by the products of the *mutB/mcm*, *meaB*, *pccBA* and *epi/mmce* genes [11,17] (Table 1). PrpR, a sigma54-dependent TF belonging to the Fis family, is known to be a transcriptional activator of the *prp* genes in *Escherichia coli* and *Ralstonia eutropha* HF39 [3]. Moreover, it is known that a GntR-family regulator gene is co-localized with the *prp* cluster in the *Pseudomonas* spp. and *Vibrio cholerae* and it is likely one more propionate regulator [3].

Here, we report a detailed comparative genomic analysis of the malonate and propionate metabolism and its regulation in the Proteobacteria. We characterized genomic loci and gene regulation, identified binding motifs, new regulon members and predicted novel TFs. We also described rearrangements of the genomic loci and regulatory interactions during the evolution of the Proteobacteria.
Materials and Methods

Computational analysis of regulons

The genomic sequences of the analyzed Proteobacteria were obtained from GenBank [2], the genomes are listed in Supplementary Table S1. Orthologs of TFs in bacterial genomes were identified by PSI-BLAST [1] searches with default parameters and confirmed by construction of the phylogenetic tree for identified homologs and by co-localization with genes of the corresponding metabolic pathways. Amino acid sequence alignment was performed using the MUSCLE package [8]. Phylogenetic trees were constructed with the PHYLIP package, using the protdist program for the calculation of distances and the maximum likelihood method with default parameters for the tree construction [9]. For all bacterial species that had any of the TFs described in the ‘Introduction’ section, the comparative genomics-based reconstruction of the malonate and propionate regulons was performed. Nucleotide positional weight matrices (profiles) for all TF-binding sites (Table S2) were constructed by the SignalX program [18], using training sets of upstream regions of genes from the analyzed bacteria. Score thresholds for identification of sites were selected so that candidate sites upstream of functionally relevant genes were accepted, while the fraction of genes preceded by candidate sites did not exceed 5% in each studied genome. At that, for long, conserved motifs, the number of candidate sites per genome did not exceed 50.

Candidate binding sites were confirmed by phylogenetic footprinting [21]. In summary, we manually analyzed alignments of upstream regions of orthologous genes and identified groups of consecutive conserved positions, relying on the assumption that binding sites are more conserved than adjacent intergenic regions. Computational search for candidate TF-binding sites in upstream gene regions (~350 to +50 nt relative to the start codon) was performed using the Genome Explorer package [18]. The threshold scores for all types of TF binding sites are given in Supplementary Table S2. Weaker sites (with score 10% less than the threshold) were also taken into account, if their position was similar to positions of stronger sites upstream of orthologous genes and there were no stronger competing sites in the same intergenic region. New candidate members were assigned to a regulon if they were preceded (as a single gene or as a part of an operon) by candidate TF-binding sites in at least four genomes. The requirement that a gene should be preceded by a candidate site in at least four genomes was established based on empirical evidence. A stricter requirement would lead to the loss of known regulon members, such as MatR/MdcY-regulated matC, that are present in only four bacteria in the studied regulon.
A weaker criterion of three genomes yields candidate regulon members with clearly irrelevant function. In most cases, including all non-trivial predictions, the actual number of orthologs preceded by candidate sites was considerably higher (e.g., TRAP transporters, see Supplementary Table S3). At that, it was also required that candidate sites were observed in sufficiently distant species, so that the site conservation could be explained by residual sequence similarity.

Genes were considered to belong to one operon if they were transcribed in the same direction with intergenic distances not exceeding 200 nt. Motif logos were constructed using WebLogo [7].

**Results and Discussion**

**Phylogenetic distribution of the malonate metabolism genes and their regulators between bacteria**

*MatR/MdcY*. MatR and MdcY have been experimentally identified in different bacteria [14,16], and this seems to be the only reason for the different names. The amino acid sequences of MatR and MdcY are very similar (49% identity), these TFs are closely related according to the phylogenetic analysis, and their previously predicted binding sites are also similar [14,16,20]. Moreover, there is no evident separation between TFs whose genes are co-localized in genomes with either *mat* or *mdc* genes. Hence, this TF will be further referred to as MatR/MdcY.

Exhaustive BLAST search identified a number of MatR/MdcY TFs, mostly in the Alphaproteobacteria and several Beta- and Gammaproteobacteria (Supplementary Table S3; Supplementary Fig S1).

Phylogenetic footprinting of regions upstream of the *mat* and *mdc* genes revealed the MatR/MdcY binding motif with the consensus TTGTATACAA [14,16,20] (Supplementary Fig. S7A). This motif coincides with the one previously predicted and confirmed by the DNase I footprint assay [14,16].

The distribution of the *mdc* and *mat* operons in bacterial taxa is very flexible. Some bacteria have both *mat* and *mdc* genes, while others have either *mdc* or *mat* operon. In most bacteria, the *matR/mdcY* gene is co-localized with the regulated genes. Most bacteria possess only one malonate regulator of the MatR/MdcY type, although the *Methylobacterium* spp. have two copies of this TF, one clustered with the *mdc*-operon, and the other one, with TRAP genes and, in some cases, like in *Methylobacterium* sp. 4-46, with TRAP and *matAB* genes.

*TRAP transporters*. The tripartite ATP-independent periplasmic (TRAP) transporters are characterized by the usage of electrochemical ion gradient as the driving force for the solute
accumulation. The best characterized TRAP is the high-affinity C4-dicarboxylate transport (Dct) system formed by three proteins: extracytoplasmic solute receptor subunit (DctP), small (DctQ) and large (DctM) integral membrane proteins [10,12,22]. Representatives of the TRAP family are present in a wide range of eubacteria and archaea. Some organisms possess a single TRAP-system (E. coli), while others have several TRAPs (Pseudomonas aeruginosa, Bacillus halodurans). Probable substrates of the TRAP transporters are L-xylulose, gluconate, mannionate, succinate etc [10,12,22].

Many Alphaproteobacteria and some Betaproteobacteria and Gammaproteobacteria have C3-dicarboxylic acid TRAP genes co-localized with genes encoding TFs or enzymes of the malonate metabolism, either forming an operon with them or preceded by their own candidate binding sites. In Sinorhizobium meliloti, these TRAP components were named matPQM [6]. Null mutants for each of the matPQM genes were shown to be unable to grow on the minimal medium containing malonate as the sole carbon source [6]. In this study, in most cases the presence of TRAP genes in the regulated operons correlated with the absence of known malonate transporters mdcLM, mdcF and matC. Among completely sequenced genomes, orthologs of these TRAPs were found only in bacteria that had the malonate catabolism. These observations suggest that these TRAP transporters are involved in the malonate transport and belong to the MatR/MdcY regulon.

MdcR. MdcR orthologs were found in the Gammaproteobacteria and in the Burkholderiales among the Betaproteobacteria, but not in the Alphaproteobacteria. In the Gammaproteobacteria, MdcR, when present, almost always is the only malonate regulator, whereas in the Betaproteobacteria it often accompanies other malonate regulators. The phylogenetic tree of all found MdcR TFs is shown in Supplementary Fig. S2.

In most Gammaproteobacteria MdcR controls a single operon, mdcABCDEGHLM or mdcABCDEFGH. In the Betaproteobacteria, the mdc genes have a diverse arrangement, usually as a single operon, or sometimes in two operons (in Burkholderia phytofirmans PsJN, Ralstonia eutropha JMP134) with mdcLM genes in a different locus. The mdcLM transporter genes are typical for the MdcR regulon. In several Gammaproteobacteria (e.g., Citrobacter koseri ATCC BAA-895, Enterobacter sp. 638) they are replaced with mdcF, and in some Betaproteobacteria (Delftia acidovorans SPH-1, Methylibium petroleiphilum PM1), with matC (Supplementary Table S3).

The candidate MdcR binding motif was predicted by the phylogenetic footprinting of the mdcA upstream region. It is a 23 nt palindrome with the consensus
ATCATTACCCTgAgggTAATGAT (Supplementary Fig. S7B). In most Gamma- and Betaproteobacteria, mdcR is not autoregulated. The exceptions are Ralstonia picketti 12J and Psychromonas ingrahamii 37, where mdcR is located in a divergon with other mdc genes and thus shares the candidate binding site.

PrpR. PrpR is a transcriptional activator of the prp genes in some Gamma- (Enterobacteriales and Xanthomonadales) and Betaproteobacteria (Burkholderiales). The phylogenetic tree of all found PrpR TFs is shown in Supplementary Fig. S3.

All Enterobacteriales have propionate genes organized in the operon prpRBCDE. In most Xanthomonadales and Betaproteobacteria, the propionate operon has the prpRBC-acnD-prpF structure. Some Betaproteobacteria have diverse propionate regulon structure and several other TFs regulating propionate metabolism (see below; Supplementary Table S3).

The candidate PrpR binding motif was identified using phylogenetic footprinting of the prpB and prpR upstream regions. The predicted binding motif is a 16 nt palindrome with the consensus rTTTCAwwwwTGAAAy (Supplementary Fig. S8A). PrpR is a sigma54-dependent transcription activator, and indeed, candidate sigma54 promoters were identified upstream of all propionate gene clusters belonging to the PrpR regulon.

New regulators of the malonate and propionate metabolism

MlnR* (FadR subfamily of the GntR family). Some bacteria from several families of the Beta- (Alcaligenaceae, Burkholderiaceae, Comamonadaceae, Rhodocyclaceae) and Gammaproteobacteria (Ectothiorhodospiraceae, Xanthomonadaceae) have another GntR-family TF adjacent to the malonate metabolism genes. This TF is related to MatR/MdcY but is not its ortholog (confirmed by PSI-BLAST and the phylogenetic tree; Supplementary Fig. S1). This protein was named MlnR* (here and further star denotes a newly given name).

The MlnR* TF was predicted to control the malonate utilization and a part of the citramalate cycle. In some Betaproteobacteria, two paralogous copies of MlnR* were found. In that case, one copy is co-localized with the matAB genes, whereas the other one is clustered on the chromosome with genes of the citramalate cycle, mutB/mcm, meaB, pccBA and epi. In bacteria having only one copy of this TF, the mlnR* gene is clustered with both mat and the citramalate cycle genes (e.g. Azoarcus sp. BH72, Dechloromonas aromatica, Polaromonas sp. JS666), or co-localized with either malonate metabolic genes (e.g. Bordetella spp., Ralstonia metallidurans, Cupriavidus taiwanensis) or genes of the citramalate cycle (e.g. Delftia acidovorans, Leptothrix cholodnii, Polaromonas naphthalenivorans). In Xanthomonadacea
Alkalilimnicola ehrlichii MLHE-1 has the \textit{mlnR*} gene co-localized with \textit{matAB} and the malonate TRAP genes \textit{matPQM}, and such gene organization resembles the one in the MatR/MdcY regulon in many Alphaproteobacteria.

The phylogenetic footprinting of the \textit{mlnR*} upstream regions revealed two types of candidate binding motifs. The first one with the consensus, TTATTCATAATTATGAATAA (Type 1, Supplementary Fig. S7C), was found in the Betaproteobacteria. The second type of the predicted MlnR* binding motif was found in the Xanthomonadaceae and \textit{Alkalilimnicola ehrlichii} MLHE-1. Its consensus, ATAATTACGATGTAATTAC (Type 2, Supplementary Fig. S7D), partially coincides, after a shift, with the Betaproteobacterial Type 1 motif. The Type 1 and Type 2 motifs contain a common, short palindromic motif, RTAATTAY, with two repeats in the Type 2 motif and only one copy in the Type 1 motif (Fig. 2). The Type 1 and Type 2 profiles cross-recognize some sites.

\textbf{GntR- and LysR-family TFs in the Burkholderia spp.} Bacteria from the genus \textit{Burkholderia} spp. have the \textit{mdc} genes, but lack orthologs of MatR/MdcY. MdcR orthologs among genus \textit{Burkholderia} spp. are present only in \textit{Burkholderia multivorans} ATCC 17616, sp. 383 and \textit{phytofirmans} PsJN, moreover, MdcR binding sites upstream of the malonate metabolic genes are found only in \textit{Burkholderia phytofirmans} PsJN. To predict other possible regulators of the malonate metabolism in the \textit{Burkholderia} spp. we searched for TF genes co-localized with the \textit{mdc} genes. Indeed, some \textit{Burkholderia} spp. (Supplementary Table S3) have a gene encoding a different GntR-family TF near the \textit{mdc} genes. This TF is distantly related to MatR/MdcY and is not its ortholog (confirmed by PSI-BLAST and the phylogenetic tree, Supplementary Fig. S1). Other \textit{Burkholderia} spp. contain a gene encoding a LysR-family TF in the \textit{mdc} locus. Orthologs of these GntR- and LysR-family TFs are absent in other genomes studied here. An attempt to find binding motifs of these regulators by the phylogenetic footprinting failed because of high similarity of the entire intergenic regions of these closely related species.

\textbf{PrpR* (\textit{FadR} subfamily of the GntR family).} Multiple representatives of the Gamma- (Alteromonadales, Oceanospirillales, Pseudomonadales, Vibrionales), some Beta- (Burkholderiales), and even Deltaproteobacteria (\textit{Geobacter metallireducens} GS-15) have a GntR-family, FadR-subfamily TF as a regulator of the propionate metabolism. This TF will be further referred to as PrpR*. The phylogenetic tree of all found PrpR* TFs is shown in Supplementary Fig. S4.
Most Alteromonadales and Pseudomonadales have propionate metabolic genes organized in the \(\text{prpR}^*\text{BC-}\text{acnD-}\text{prpFD}\) operon, while some bacteria from these taxa have shorter operons, \(\text{prpR}^*\text{BC-}\text{acnD-}\text{prpF}\) or \(\text{prpR}^*\text{BC-}\text{acnD}\). Some representatives of the \(\text{Pseudomonas}\) spp. also have a second aconitase gene \(\text{acnB}\) in the operon. Most Vibrionales have the \(\text{prpR}^*\text{BC-}\text{acnD-}\text{prpFE}\) operon, while Oceanospirillales have the \(\text{prpR}^*\text{BCD}\) operon. Similar operon organization, \(\text{prpR}^*\text{BDC}\), was observed in \textit{Geobacter metallireducens} GS-15.

\textit{Geobacter metallireducens} is the only member of the Deltaproteobacteria that has any of the regulators studied here. Even congenerics \textit{Geobacter sulfurreducens} and \textit{Geobacter uranireducens}, though having the \textit{prp} metabolic genes, lack orthologous transcription factors and thus were not considered. Among other Delta- and Epsilonproteobacteria, only \textit{Helicobacter hepaticus} has a complete propionate metabolic pathway (\(\text{prpEBCD}\) genes), but lacks any known or predicted propionate TFs.

In the Betaproteobacteria, PrpR* not only is rare, but seems to be not the main propionate regulator, even if it is present in a genome, as most propionate genes are regulated by a HutC-type TF (\(\text{SdhR}^*\), see below, Supplementary Table S3). Among the Betaproteobacteria, PrpR* controls propionate metabolic genes only in \textit{Verminephrobacter eiseniae} EF01-2 and \textit{Bordetella} spp.

The predicted PrpR* binding motif, identified by the phylogenetic footprinting, is a 12 nt palindrome with the consensus ATTGTCGACAAT (Supplementary Fig. S8B).

\textit{PrpQ* (XRE family)}. Most Alpha- and some Betaproteobacteria have a gene encoding a XRE-family TF that is likely the transcriptional regulator of the propionate metabolism. This TF will be further referred to as PrpQ*. The phylogenetic tree of all found PrpQ* TFs is shown in Supplementary Fig. S5.

The candidate PrpQ* binding motif was identified by the phylogenetic footprinting of the upstream regions of the \(\text{prpQ}^*\) and \(\text{pccBA}\) operons. The motif is a short 8 nt palindrome with the consensus TTTGCrAA, often present in multiple copies upstream of regulated genes (Supplementary Fig. S8C). In Betaproteobacteria, reconstructed PrpQ* regulon includes genes for the part of methylcitrate pathway (Supplementary Table S3). In Alphaproteobacteria, PrpQ* binding sites were found upstream of genes, encoding enzymes of citramalate and/or methylcitrate pathways, such as the \(\text{prp}\), \(\text{acnD}\), \(\text{pccBA}\) and \(\text{mutB}\) genes (Supplementary Table S3).
SdhR* (HutC subfamily of the GntR family). Multiple Betaproteobacteria (Burkholderiales) have a GntR-family, HutC-subfamily TF that, according to the co-localization, may control expression of genes encoding enzymes involved in the TCA cycle, such as succinate dehydrogenase sdhABCD, citrate synthase gltA, bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase acnB, malate dehydrogenase mdh, and some other catabolic genes, in particular ygfY, functionally connected to sdhB, and tam, encoding trans-aconitate 2-methyltransferase that catalyzes the monomethyl esterification of trans-aconitate at high and cis-aconitate, isocitrate, and citrate at lower velocities and affinities [5]. This TF will be further referred to as SdhR*. The phylogenetic tree of all found SdhR* TFs is shown in Supplementary Fig. S6.

Phylogenetic footprinting of sdhR*, mdh and acnB upstream regions revealed a conserved area that is likely the SdhR* binding site. The predicted binding motif is a 26 nt palindrome with the consensus TCTTATGTCTTATATAAGACATAAGA (Supplementary Fig. S8D). This motif has an intrinsic symmetry, comprising two direct repeats of complementary motifs TCTTAT and ATAAGA in the distal and proximal part of the palindrome, respectively.

In many Betaproteobacteria (Supplementary Table S3), SdhR* also controls a number of genes of the propionate utilization pathway (prpB, prpC, prpD, acnA, prpF) and the glyoxylate shunt (malyl-CoA lyase mcl and isocitrate lyase aceA).

Conclusions

The diversity of the organization of the malonate and propionate regulons in different taxonomic groups leads to the question about the evolution of these regulatory systems.

The most consistent group is the Alphaproteobacteria. The mode of malonate and propionate regulation is the same among these bacteria: the malonate metabolic genes are regulated by MatR/MdcY, and the propionate utilization is under the PrpQ* regulation. Both these TFs are typical for the Alphaproteobacteria and are present in only few Beta- and Gammaproteobacteria, seemingly, as a result of horizontal gene transfer (Fig. 3).

Indeed, among the Gammaproteobacteria, the horizontal transfer of matR/mdcY occurred in the Pseudomonadales and Psychromonadaceae (confirmed by the phylogenetic tree; Fig. 5, S1). Ectothiorhodospiraceae and Xanthomonadaceae have the GntR-family TF named MlnR* as a malonate regulator, which is typical for the Betaproteobacteria (see below). It is likely they inherited it from their common ancestor with the Betaproteobacteria (Supplementary Fig. S1). In other Gammaproteobacteria, the malonate metabolism genes are regulated by MdcR, and this TF seems to be the original malonate regulator in this taxonomic group (Fig. 5). The propionate
metabolism in the Gammaproteobacteria is mostly controlled by a GntR-family TF named here PrpR*, but some Enterobacteriales and all analyzed Xanthomonadales use a Fis-family TF, PrpR, as the propionate regulator. This TF is also present in Betaproteobacteria, and, according to the phylogenetic tree (see Supplementary Fig. S3), either Beta- and Gammaproteobacteria inherited it from their common ancestor, or the common ancestor of the Enterobacteriales and Xanthomonadales had it transferred from some ancient Betaproteobacteria. Both scenarios involve multiple losses of PrpR in a variety of lineages. The GntR-family TF PrpR* is widespread among Gammaproteobacteria (Fig. 5). This TF is also present in several Betaproteobacteria and Deltaproteobacteria Geobacter metallireducens GS-15. According to the branch localization in the phylogenetic tree, G. metallireducens got PrpR* as a result of horizontal gene transfer from a genome close to the ancestor of the Alteromonadales. In seems that some Betaproteobacteria had PrpR* transferred from the Pseudomonadales (for example, Bordetella petrii DSM 12804, see Supplementary Fig. S4), while some others inherited it from their common ancestor with the Gammaproteobacteria (e.g., Bordetella bronchiseptica RB50, Verminephrobacter eiseniae EF01-2, Fig. S4). An interesting variant is seen in Bordetella avium 197N that has two genes prpR*, one likely obtained from the Pseudomonadales and the other inherited from the common ancestor with the Gammaproteobacteria. It is plausible that in the Gammaproteobacteria we observe an intermediate stage of replacement of PrpR by PrpR* (Fig. 5).

The most diverse regulation is seen in the Betaproteobacteria (Fig. 4, Table 2). Overall, bacteria from this group possess at least five malonate regulators: MatR/MdcY, which they presumably got from the Alphaproteobacteria via horizontal gene transfer (confirmed by the phylogenetic tree, Supplementary Fig. S1); MdcR, horizontally transferred (according to the phylogenetic tree, Supplementary Fig. S2) from the Gammaproteobacteria or inherited from the common ancestor of Beta- and Gammaproteobacteria; LysR- and GntR-family regulators, found only in the Burkholderia spp. However, most Betaproteobacteria contain a GntR family TF, MlnR*. The propionate utilization in the Betaproteobacteria is also under diverse regulation. Only few Betaproteobacteria, three closely related Comamonadaceae species, have PrpQ*, as a result of the horizontal transfer from the Alphaproteobacteria (confirmed by the phylogenetic tree, Supplementary Fig. S5). Most Betaproteobacteria have SdhR*, either as the only propionate regulator, or accompanied by PrpR or sometimes PrpR* (Fig. 4). It is interesting to note that most Betaproteobacteria that got MlnR*, lack the Fis-family TF PrpR, possibly due to the fact that these TFs control alternative pathways of propionate conversion (Supplementary Table S3).
Overall, we have reconstructed the malonate and propionate regulons, described their flexible and diverse regulation in the Proteobacteria, found new TFs that control the malonate and propionate metabolism, identified their candidate binding sites by positional and sequence analysis. The comparative genomic analysis also yielded new candidate members of the malonate regulon, namely, TRAP transporters.

Acknowledgments

This study was partially supported by The Russian Foundation of Basic Research (grants 09-04-92745 and 10-04-00431), the Russian Academy of Sciences (via program "Molecular and Cellular Biology") and state contracts 14.740.11.0003 and 07.514.11.4007.

Tables

Table 1. Genes involved in the malonate and propionate catabolism.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function of product</th>
</tr>
</thead>
<tbody>
<tr>
<td>mdcL</td>
<td>malonate transporter subunit</td>
</tr>
<tr>
<td>mdcM</td>
<td>malonate transporter subunit</td>
</tr>
<tr>
<td>mdcA</td>
<td>malonate/acetyl-CoA transferase, alpha subunit</td>
</tr>
<tr>
<td>mdcC</td>
<td>acyl-carrier protein, delta subunit</td>
</tr>
<tr>
<td>mdcD</td>
<td>malonyl-CoA decarboxylase, beta subunit</td>
</tr>
<tr>
<td>mdcE</td>
<td>a protein involved in the stability of the enzyme complex or co-decarboxylase, gamma subunit</td>
</tr>
<tr>
<td>mdcF</td>
<td>malonate transporter</td>
</tr>
<tr>
<td>mdcG</td>
<td>holo-ACP synthase</td>
</tr>
<tr>
<td>mdcB</td>
<td>triphosphoribosyldephospho-CoA synthase, a protein involved in the formation of the prosthetic group precursor</td>
</tr>
<tr>
<td>mdcH</td>
<td>malonyl-CoA/ACP transacylase</td>
</tr>
<tr>
<td>matA</td>
<td>malonyl-CoA decarboxylase</td>
</tr>
<tr>
<td>matB</td>
<td>malonyl-CoA synthetase</td>
</tr>
<tr>
<td>matC/matP</td>
<td>malonate transporter</td>
</tr>
<tr>
<td>mdcY/matR</td>
<td>GntR-family regulator of malonate metabolism</td>
</tr>
<tr>
<td>mdcR</td>
<td>LysR-family regulator of malonate metabolism</td>
</tr>
<tr>
<td>prpB</td>
<td>2-methylisocitrate lyase</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>prpC</td>
<td>2-methylcitrate synthase</td>
</tr>
<tr>
<td>prpD</td>
<td>2-methylcitrate dehydratase</td>
</tr>
<tr>
<td>prpE</td>
<td>propionyl-CoA synthetase</td>
</tr>
<tr>
<td>prpF</td>
<td>aconitase-accessory protein</td>
</tr>
<tr>
<td>acnB</td>
<td>bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase</td>
</tr>
<tr>
<td>acnA</td>
<td>aconitate hydratase</td>
</tr>
<tr>
<td>prpR</td>
<td>regulator of propionate metabolism</td>
</tr>
<tr>
<td>mutB/mcm</td>
<td>methylmalonyl-CoA mutase</td>
</tr>
<tr>
<td>meaB</td>
<td>putative auxiliary metallochaperone, involved in the protection and assembly of methylmalonyl-CoA mutase (G3E family of P-loop GTPases)</td>
</tr>
<tr>
<td>pccA</td>
<td>propionyl-CoA carboxylase, subunit alpha</td>
</tr>
<tr>
<td>pccB</td>
<td>propionyl-CoA carboxylase, subunit beta</td>
</tr>
<tr>
<td>epi/mmce</td>
<td>methylmalonyl-CoA epimerase</td>
</tr>
</tbody>
</table>

Table 2. Distribution of malonate/propionate transcription factors among various taxa of Proteobacteria

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Number of genera obtaining TFs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MatR/MdcY</td>
</tr>
<tr>
<td>Alpha</td>
<td>12</td>
</tr>
<tr>
<td>Beta</td>
<td>4</td>
</tr>
<tr>
<td>Gamma</td>
<td>2</td>
</tr>
<tr>
<td>Delta</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure Legends**

**FIG. 1.** Malonate and propionate metabolism.
The methylcitrate pathway is marked with red lines, the citramalate cycle – with blue.

**FIG. 2.** Common parts of the two types of MlnR* binding motifs.
The common parts are marked with dashed lines, the repeated part in the Type 1 and Type 2 motifs is set in bold.

**FIG. 3.** Regulation in Alphaproteobacteria.
FIG. 4. Regulation in Betaproteobacteria.

Line color denotes the mode of regulation, using the following color code: MatR/MdcY – yellow; MdcR – orange; MlnR* - pink; SdhR* - light blue; PrpR – dark blue; PrpR* - green; PrpQ* - violet; LysR-family TF (Burkholderia) – light brown; GntR-family TF (Burkholderia) – dark green.

FIG. 5. Regulation in Gammaproteobacteria.

Line color denotes the mode of regulation, using the following color code: MatR/MdcY – yellow; MdcR – orange; MlnR* - pink; PrpR – dark blue; PrpR* - green.

References


