Biochemical and Biophysical Research Communications 415 (2011) 88–93

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Genomic distribution and heterogeneity of MocR-like transcriptional factors containing a domain belonging to the superfamily of the pyridoxal-5′-phosphate dependent enzymes of fold type I

E. Bramucci, T. Milano, S. Pascarella *

Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma La Sapienza, 00185 Roma, Italy

ARTICLE INFO

Article history:
Received 2 October 2011
Available online 12 October 2011

Keywords:
MocR
Pyridoxal-5′-phosphate
Type I fold
2-Aminoadipate aminotransferase
Bacterial proteome
Hidden Markov models

ABSTRACT

Bacterial proteins belonging to the MocR/GabR family are chimeric proteins incorporating a short N-terminal helix-turn-helix containing domain with DNA-binding properties, and a long C-terminal domain belonging to the superfamily of the pyridoxal-5′-phosphate enzymes of fold type I. The first purpose of this report is to give an overview of the distribution of these factors among the different taxonomical bacterial divisions and to determine the degree of conservation of the main structural features of the PLP binding domain. Complete proteomes of bacteria phyla were scanned with a hidden Markov model representative of the MocR family. Results indicate that presence of MocR factors is heterogeneous even within the single bacterial phylum: some species miss completely the factors, while others possess one or even more regulators. Absence of MocR factors is distinctive of some phyla such as *Chlamydiae*. The genomic distribution of MocR is, as expected, highly correlated to the size of the genome. At variance, phyla missing MocR regulators generally are characterized by compact genomes, of the order of 1.0–2.0 Mb, such as the case of *Mollicutes* or *Chlamydiae*. Apparently, the minimum genome size compatible with the presence of MocR genes is around 2.0–2.5 Mb. Conservation of the residues corresponding to those involved in the interaction with the cofactor pyridoxal-5′-phosphate in the homologous 2-aminoadipate aminotransferase, was analyzed in the multiple sequence alignments of MocR within each phyla considered. In the vast majority of cases, residues are conserved or conservatively replaced. This result suggests that, in most cases, MocR factors preserve at least ability to bind the cofactor and very likely some catalytic abilities.

1. Introduction

Bacterial proteins belonging to the MocR/GabR family [1] are chimeric proteins incorporating a short N-terminal helix-turn-helix (HTH)-containing domain with DNA-binding properties, and a long C-terminal domain belonging to the superfamily of the pyridoxal-5′-phosphate (PLP) enzymes of fold type I [2]. The HTH-containing regions of these proteins vary from 60 to 120 residues in length and are similar to the winged HTH regions of bacterial transcriptional regulators of the broad GntR family [1]. In general, the MocR/GabR family is poorly characterized and many aspects of their structure and function are still unexplored. Two members of the family have been characterized in more detail, namely GabR that positively regulates expression of the *gabTD* operon responsible for utilization of γ-aminobutyric acid [3] and PdxR involved in the regulation of the expression of the divergently orientated *pdxST* genes coding for the subunits of pyridoxal-5′-phosphate synthase [4]. However, their structural and functional properties still remain largely undefined.

The C-terminal domain of the MocR/GabR belongs to the well known superfamily of the fold type-I pyridoxal-5′-phosphate (PLP) dependent enzymes that are mainly involved in the metabolism of amino acids [2,5]. For this superfamily, a wealth of structural and functional information is available [6]. Therefore, understanding the role of such domains within the chimeric MocR/GabR regulators and their differences/similarities to the parent freestanding enzymes appears particularly intriguing and attracting. Under these perspectives, the first purpose of this report is to give an overview of the distribution of the MocR/GabR factors among the different taxonomical bacterial divisions trying to assess the degree of conservation of a few distinguishing features of the active site of type I PLP-binding domain. Therefore this work is meant to give a contribution to the studies aimed at clarifying the function of these regulators which may have important biotechnological and/or biomedical implications.

* Corresponding author. Fax: +39 06 49917566.
E-mail address: Stefano.Pascarella@uniroma1.it (S. Pascarella).

0006-291X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2011.10.017

2. Materials and methods

Complete bacterial and archaeobacterial proteomes were retrieved from the UniProt database [7] version August, 2011. Only complete proteomes were utilized during all the subsequent analyses. MocR sequences were identified in these proteomes using the method of profile hidden Markov model (HMM) [8], a probabilistic model intended to describe the structural features of a protein family. The model can be calculated starting from a seed multiple sequence alignment representative of the protein family to be modeled. The profile HMM was chosen as the working methodology for its superior sensitivity and specificity. In the reported analysis, the HMMER 3.0 [8] package was utilized under Linux (Ubuntu 10.4) operating system.

The required seed multiple sequence alignment was calculated with the program MUSCLE [9] using two MocR sequences selected from each phylum of Bacteria (e.g., *Actinobacteria*, *Firmicutes*, etc.). Bacteria classification was taken from the taxonomy division of the UniProt databank. HMM searches were carried with the HMMSEARCH module in the proteomes of each phylum. Phyla populated by only one complete proteome were not taken into consideration. Output was parsed and analyzed with Perl and Bash scripts. Statistics were collected using Perl scripts, Origin8 software (OriginLab, Corporation, Northmapton, MA, USA) and Microsoft Office Excel (Microsoft corporation, Redmon, WA, USA). Genome size data were retrieved from the Entrez Genome Project portal of the web site of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genome/).

Multiple sequence alignments of the MocR sequences retrieved from UniProt databank were calculated with MUSCLE [9] while the phylogenetic analysis was carried out with the package PHYLIP [10] in the EMBOSS suite [11] or through the site http://mobyle.pasteur.fr. Phylogenetic trees were built using the neighbor-joining method and were validated with the bootstrap procedure. Sequence editing, display, analysis relied on Seaview [12] or Jalview [13] programs.

3. Results

3.1. Taxonomical distribution of MocR sequences

A portion of the seed multiple sequence alignment utilized to build the query HMM is displayed in Fig. 1. HMM searches were carried out on the complete proteomes of each phylum reported in the taxonomy division of UniProt databank (Table 1). Since the query HMM was able to match also proteome sequences containing only the PLP-binding domain, such as type-I aminotransferases, or the HTH domain of other GntR members, a filter was applied to the program output to select *bona fide* MocR sequences. In particular, a sequence was considered belonging to the MocR family if (a) it displayed a significant *E*-value, (b) it was longer than 420 residues, and (c) it overlapped the query HMM for at least 420 residues. These criteria assured the co-presence of HTH and PLP-binding domains on the same sequence. The resulting pattern of taxonomical distribution indicated interesting trends (Table 1): 7 out of 26 phyla tested possess no MocR/GntR regulators. In these cases, absence of MocR was further tested and confirmed through HMM searches within the entire UniProt set of protein sequences belonging to the bacteria of the same phylum, regardless of completeness of the corresponding proteomes. In all cases, observed lack of regulators was supported. Within the MocR-containing proteomes, various patterns can be observed (Table 1). For example, 70% of *Actinobacteria* proteomes contain MocR-like sequences, while only 20% of the *Bacteroidetes* are positive. Statistics on phyla for which only a few complete proteomes are available (e.g. *Planctomycetes* or *Verrucomicrobia*), are obviously affected by significant statistical noise and should not be considered as definitive. The average number of MocR sequences per proteome (i.e. how many MocR sequences are present in a given proteome) unveils marked differences (Table 1), ranging from 4.2 in the *Proteobacteria* to 1.0 in the *Cyanobacteria*. *Proteobacteria* phylum is very heterogeneous and it is the most populated with complete proteomes. For this reasons, statistics were collected separately for each subdivision, namely *Alpha-*, *Beta-*, *Gamma-*, *Delta-* and *Epsilonproteobacteria*. Among these, the *Betaproteobacteria* have the highest frequency of positive proteomes (80%) and the highest average number of MocR proteins per proteome (8.4) while the *Epsilonproteobacteria* possess the lowest values (10% and 1.3, respectively). An example of the variability of the number of MocR sequences in the species within each phylum is reported in Fig. 2. The number of MocR sequences found in a proteome was significantly correlated to the corresponding genome size, as shown in Fig. 3 for the three most populated phyla.

Complete proteomes from Archaea were also examined and no convincing MocR sequence could be found. The same search has therefore been carried out on the entire set of archaeobacterial proteins in UniProt databank and the negative result was confirmed.

3.2. Sequence conservation

Multiple MocR sequence alignments were obtained for each bacterial division with the program MUSCLE [9] and the degree of conservation of the most significant residues potentially interacting with PLP at the active site of the fold-type I domain was analyzed. To label the residues putatively interacting with the PLP cofactor in the MocR sequences, a sequence alignment between the seed alignment and a type-I PLP domain with three-dimensional structure solved was calculated. To this purpose, the seed HMM was utilized to scan the subset of UniProt databank containing sequences of proteins with three-dimensional structure solved. The top structures matching the MocR PLP-binding domain were: the putative uncharacterized protein PH0207 from *Pyrococcus horikoshii* (PDB ID: 1XOM), the multiple substrate aminotransferase from *Thermococcus profundus* (PDB ID: 1WST), and 2-aminoadipate aminotransferase from *Thermus thermophilus* (PDB ID: 2EGY). The latter protein was chosen as a structural reference because better characterized than the former two [14] (Fig. 1).

Attention was focused on the positions encompassing the residues Asp345 (numbering system refers to Fig. 1) and Lys382 which H-bonds to the nitrogen atom of the pyridine ring and reacts with the cofactor aldehyde group at the 5′ position of PLP forming a Schiff-base, respectively. The Asp345 appears to be very much conserved, although in several cases it is replaced by Asn, Glu, Tyr or, more rarely, by Ser, Leu, or Val especially in regulators from the *Betaproteobacteria*. Substitutions have not been observed in *Chloroflexi*, *Cyanobacteria*, *Deinococci*, and *Spirochaetales* which, however, are the least populated phyla. Therefore a sampling effect can explain the apparent conservation. Lys382 is also conserved, but once again many exchanges have been observed with residues such as Ser, Thr, His, Ala, Arg, Gln. The Lys382 appears to be more variable in the *Actinobacteria* division. Fig. 4 reports, as an example, the sequence logos [15] calculated for multiple sequence alignment of the Mocr/GabR of the phyla *Actinobacteria*, *Firmicutes* and *Gammaproteobacteria* in the regions of interest. It should be noted that other residues involved in PLP binding tend to be conserved. In particular, Ser/Thr379, Ser381, and Arg389 which, in 2-aminoadipate aminotransferase, H-bond to the phosphate group of the cofactor are mostly identically conserved or conservatively replaced.

Overall, the extent of average sequence conservation among the PLP binding domain of the MocR regulators within the single

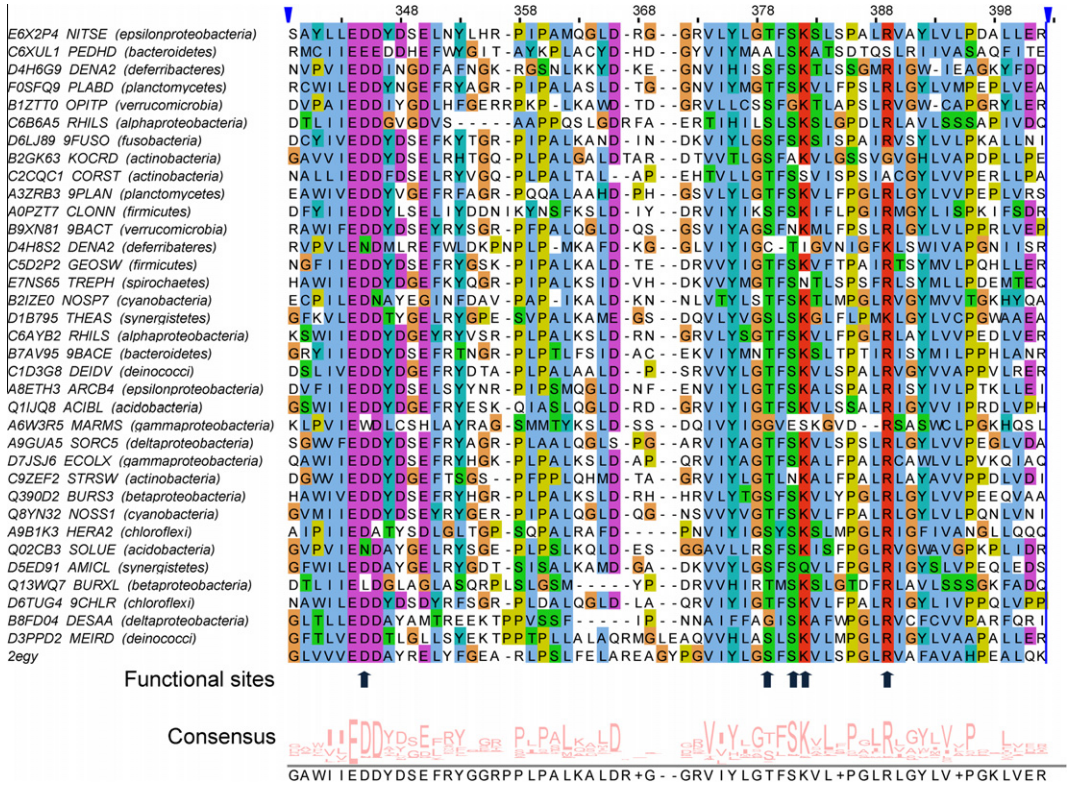

Fig. 1. Portion of the seed multiple sequence alignment used to calculate the HMM profile. Sequences were aligned with MUSCLE and displayed with Jalview. Color code refers to the residue chemical-physical property and indicates conserved columns. Sequences are labeled with UniProt codes and are assigned to the corresponding phylum (in parentheses). Arrows on the line denoted as "Functional sites" indicate residue involved in PLP binding and discussed in the text. Consensus line reports the distribution of the most frequent residues in each column using a logo representation. Dashes denote insertions/deletions in the sequences. The sequence of the PLP-binding domain of the 2-aminoadipate aminotransferase from *Thermus thermophilus* (PDB ID: 2EGY) was incorporated in the alignment to serve as a structural reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
List of bacterial phyla utilized in the work.

	Phylum	No. of complete proteomes	No. of MocR-positive proteomes	No. of total MocR sequences	Fraction of MocR-positive proteomes	Average number of MocR per proteome
1	*Acidobacteria*	5	3	6	0.6	2.0
2	*Actinobacteria*	130	92	276	0.7	3.0
3	*Aquificae*	9	0	0	0.0	0.0
4	*Bacteroidetes*	48	10	29	0.2	2.9
5	*Chlamydiae*	25	0	0	0.0	0.0
6	*Chlorobi*	11	0	0	0.0	0.0
7	*Chloroflexi*	15	2	3	0.1	1.5
8	*Cyanobacteria*	40	6	6	0.1	1.0
9	*Deferribacteres*	3	1	3	0.3	3.0
10	*Deinococcus*	13	6	9	0.5	1.5
11	*Dictyoglomi*	2	0	0	0.0	0.0
12	*Elusimicrobia*	2	0	0	0.0	0.0
13	*Fibrobacteres*	1	1	1	1.0	1.0
14	*Firmicutes*	284	168	532	0.6	3.2
15	*Fusobacteria*	5	2	5	0.4	2.5
16	*Mollicutes*	40	0	0	0.0	0.0
17	*Planctomycetes*	5	1	1	0.2	1.0
18	*Spirochaetes*	27	4	4	0.1	1.0
19	*Synergistetes*	2	2	2	1.0	1.0
20	*Thermotogae*	11	0	0	0.0	0.0
21	*Verrucomicrobia*	4	1	2	0.3	2.0
22	*Alfaproteobacteria*	137	83	238	0.6	2.9
23	*Betaproteobacteria*	96	72	607	0.8	8.4
24	*Deltaproteobacteria*	43	27	56	0.6	2.1
25	*Epsilonproteobacteria*	54	4	5	0.1	1.3
26	*Gammaproteobacteria*	308	211	775	0.7	3.7

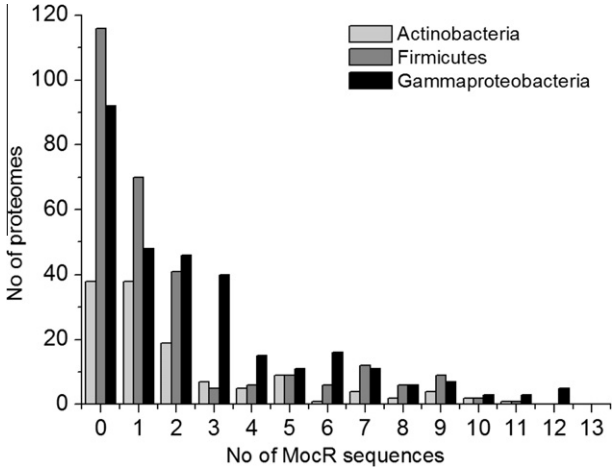

Fig. 2. Histogram reporting the distribution of the number of proteomes (*Y*-axis) containing increasing number of MocR/GntR sequences (*X*-axis). The distribution has been calculated for the three most populated taxonomical divisions.

phylum is not very high: for example, in *Actinobacteria* average percentage identity is 29% with minimum at 14%; *Alphaproteobacteria* 26% and minimum 14%; *Gammaprotebacteria* 27% and minimum 10%; *Firmicutes* 27% and 8%, respectively.

4. Discussion

MocR sequences belong to the wide GntR family of transcription factors [16] characterized by the presence of an N-terminal DNA-recognition domain of the HTH (helix-turn-helix) group and an effector/oligomerization C-terminal domain. In particular, MocR effector domain is distinguished because of its exceptional average length of about 350 residues and its homology to the fold type I aminotransferases [2]. Several genome-level studies on the distribution of GntR factors have been recently published in the literature [17,18] but none of them specifically addressed the MocR family and the characteristics of the active site of the PLP binding domain. Therefore, the first purpose of this report is to give an overview of the distribution of these factors among the different taxonomical bacterial divisions and to determine the degree of conservation of some of the distinguishing features of the PLP binding domain.

Application of hidden Markov models to an appropriately designed seed aligment, provided an effective tool to retrieve the members of the MocR/GabR family in complete bacterial proteomes. Since the model is able to perceive similarity to the HTH domains and to aminotransferases of fold type I alone, an output filtering procedure was necessary to avoid false positive hits. Our empirical tests suggested as *bona fide* MocR members those sequences longer that 420 residue overlapping the model for more than 420 positions among those obtaining a significant *E*-value.

Our results indicate that presence of MocR factors is rather heterogeneous even within the single bacterial phylum: some species miss completely the factors, while others possess one or even more regulators. Evidently, the presence of the regulators does not overlap with the taxonomical classification of bacteria, whereas the absence is distinctive of some phyla like *Chlamydiae*. In general, the genomic distribution of MocR is, as expected, highly correlated to the size of the genome (Fig. 3) and, consequently, to the complexity of the relative metabolic and regulatory networks. However, wide variations within each phylum can be observed. Indeed, microorganisms with similar genome size possess very different number of MocR sequences and vice versa. At variance, phyla missing

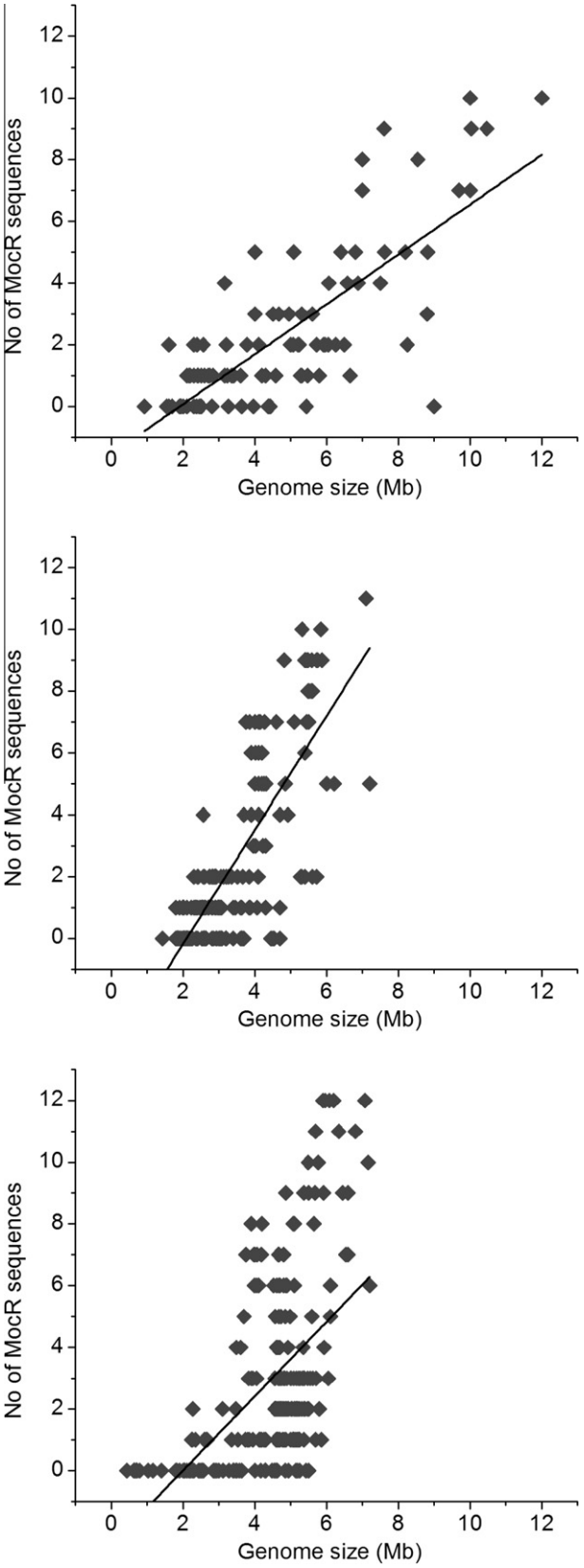

Fig. 3. Correlation between genome size and number of MocR sequences. Each diamond represents one proteome the position of which within the graph indicates the genome size and the number of MocR sequences found therein. Panels A, B, and C refer to *Actinobacteria*, *Firmicutes* and *Gammaproteobacteria* proteomes, respectively. Regression line is reported in each graph. *R* values for distributions in A, B and C are 0.78, 0.79 and 0.58, respectively.

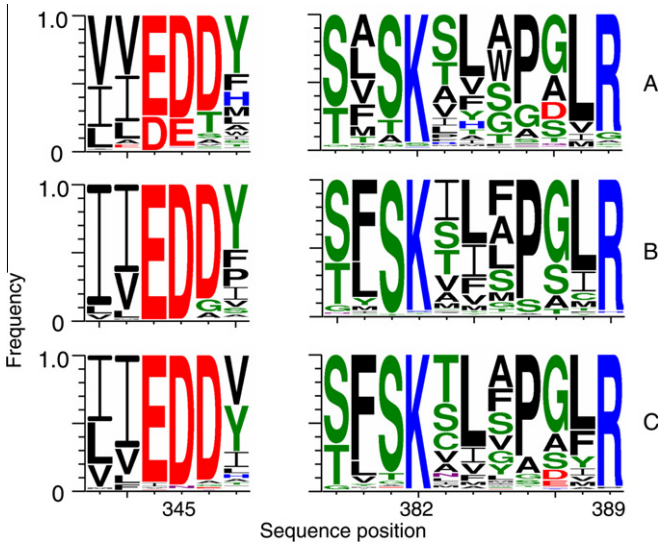

Fig. 4. Logo of the sequence encompassing the sites interacting with the PLP. Logo were calculated with the WebLogo service [15] using the multiple sequence alignment of all the MocR sequences found in *Actinobacteria* (A), *Firmicutes* (B) and *Gammaproteobacteria* (C). The height of each one-letter code character indicating amino acids is proportional to the observed frequency of the corresponding residue in that position. Sequence positions (*X*-axis) use the numbering system of Fig. 1. Color code indicates residue chemical–physical characteristics. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

MocR-like regulators generally are characterized by compact genomes, of the order of 1.0–2.0 Mb, such as the case of *Mollicutes* or *Chlamydiae*. Apparently, the minimum genome size compatible with the presence of MocR genes is around 2.0–2.5 Mb.

Although GntR regulators are reported to be known in *Archaea* proteomes [1], no MocR member was detected in our study. This suggests that these regulators are typical of eubacterial microorganisms.

Question remains as to whether the PLP binding domains possess catalytic activity or at least conserve the ability to bind PLP or one of the B_6 vitamers [3]. This critical aspect has not been extensively investigated and therefore very few data were reported in the literature. It has been demonstrated [4] that actinobacterial MocR-like regulator from *Corynebacterium glutamicum* denoted as PdxR, involved in the transcriptional regulation of the *pdxS* and *pdxT* genes, possess no aminotransferase activity. Indeed, the PdxR belongs to a MocR subgroup characterized by lack of the Lys382 necessary for the covalent binding of PLP that is replaced by Ser, Thr or His. Interestingly, Asp345 is conserved in the same sequences. Conservation of Thr379, Ser381 supports the idea that such forms may have retained the ability to bind PLP or other forms of the cofactors such as pyridoxamine phosphate, though Arg389 that also bind the phosphate group through a salt bridge, is not conserved. Indeed, substitution of the active site Lys in several PLP-dependent enzymes among which serine hydroxymethyltransferase [19], does not disrupt capability of cofactor binding, though catalytic activity is dramatically impaired. We applied phylogenetic analysis to the multiple sequence alignment of the PLP-binding domain of the Actinobacterial MocR from proteomes containing only single regulators. The results clearly show that the Lys382-missing proteins PdxR from *Actinobacteria* form a subgroup (Supplementary Fig. S1). At variance, in the case of the regulator GabR from *Bacillus subtilis* [3], it has been suggested that catalytic activity of the PLP-binding domain is necessary for its action as a transcriptional activator.

Similar conservation patterns can be seen in other phyla. For example, *Firmicutes* display strong conservation of Lys382 except, for example, in the case of *Bacillus amyloliquefaciens* (UniProt ID: E1UM50_BACAS), *Brevibacillus brevis* (UniProt ID: A7Z1V2_BACA2), *Clostridiales genomosp.* BVAB3 (UniProt ID: D3R1K4_CLOB3) where Gln replaces for Lys382 and *Caldicellulosiruptor kristjanssonii* (UniProt ID: E4S414_CALKI), *C. kronotskyensis* (UniProt ID: E4SB32_CALK2), *C. hydrotermalis* (UniProt ID: E4Q790_CALH1), *C. obsidiansis* (UniProt ID: D9TFM7_CALOO), *C. owensesis* (UniProt ID: E4Q418_CALOW), *C. saccharolyticus* (UniProt ID: A4XIB4_CALS8), and *Anaerocellum thermophilum* (UniProt ID: B9MKZ0_ANATD) where a Thr replaces for the same Lys. In a few cases, like in *Lactobacillus plantarum* (UniProt ID: Q88U48_LACPL), Arg is the replacement for Lys382. At variance with *Actinobacteria*, Arg389 appears more conserved and lack of Lys382 does not correlate with substitution of Arg and viceversa except in the case of *Brevibacillus brevis* (UniProt ID: A7Z1V2_BACA2). Similar replacements can be found in the corresponding domains from other phyla and a few examples are reported in Table 2.

Table 2
Sample of MocR sequences bearing substitutions of the residues interacting with PLP in the vicinity of Asp345 and Lys382.

Reference	Databank code PDB ID: 1EGY	Asp345 VVE**D**DA	Lys382 SF**SK**VLSPGL**R**
Alphaproteobacteria	UniProt ID: Q981V5_RHILO	IIA**N**SM	**SFEC**GLSPQY**G**
	UniProt ID: A4YPA7_BRASO	LIE**D**DY	**SFTA**SLGAGL**R**
Bacteroidetes	UniProt ID: C7PMV7_CHIPD	IIE**D**DY	**SVMH**TLPIS–
	UniProt ID: A4AWD2_MARSH	IIE**D**DY	**KLGQ**SFFPSF**H**
Betaproteobacteria	UniProt ID: C6XB13_METSD	ILA**A**DY	**ALPS**ILPADR**Q**
	UniProt ID: Q13RW3_BURXL	IVE**D**DF	**GYCN**TYGIDL**R**
Firmicutes	UniProt ID: E4S414_CALKI	IIE**D**DF	**SFST**VTMPAL**R**
	UniProt ID: E4RIZ6_HALSL	IIF**E**SY	**ELSN**KVFPGL**D**
	UniProt ID: E3DNZ6_HALPG	IVI**Q**QY	**ALTN**RVFPGL**K**
	UniProt ID: D3R1K4_CLOB3	VIE**D**DY	**TFSQ**TISPSL**R**
Gammaproteobacteria	UniProt ID: F0LUG0_VIBFN	IIE**D**DF	**SFTA**SIAPGL**R**
	UniProt ID: 4ZCZ4_SHEVD	LIE**Y**D-	**DFYD**TISPAI**S**
	UniProt ID: B1J9P9_PSEPW	LLE**N**DL	**SLQA**LVGAEA**P**

In general, most of the MocR-like sequences appear to undergo strong selective pressure to maintain the residues essential for PLP or other B_6 vitamers binding, suggesting that this function is important for the physiological role the regulators play in the cell. Therefore, it could be argued that, except in a few cases, correct functionality of MocR-like regulators requires at least ability to bind B_6 vitamers and, perhaps in many cases, catalytic activity as well. Authors [16] have proposed that one of the possible functions of the PLP-binding domain within the MocR factors is to provide dimerization properties, since the parent type-I PLP enzymes are active, in most cases, as dimers. The head-to-tail dimerization ability is likely the predominant function available in those PLP-binding domains where both Lys and Asp are absent. In all the other cases, the vast majority, strong evidence supports the presence of B_6 vitamers binding ability and perhaps catalytic activity as well.

Acknowledgments

This work has been partially funded by the Italian Ministero dell'Istruzione, Università, Ricerca. This work will be submitted by E. Bramucci in partial fulfillment of the requirements of the degree of "Dottorato di Ricerca in Biochimica" at Sapienza, Università di Roma. Authors are indebted to Prof. Alessandro Paiardini for critically reading the manuscript. This paper is dedicated to Prof. Francesco Bossa on the occasion of his 70th birthday.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2011.10.017.

References

[1] P.A. Hoskisson, S. Rigali, Variation in form and function the helix-turn-helix regulators of the GntR superfamily, Adv. Appl. Microbiol. 69 (2009) 1–22.
[2] G. Schneider, H. Käck, Y. Lindqvist, The manifold of vitamin B6 dependent enzymes, Structure 8 (2000) R1–R6.
[3] B.R. Belitsky, *Bacillus subtilis* GabR, a protein with DNA-binding and aminotransferase domains, is a PLP-dependent transcriptional regulator, J. Mol. Biol. 340 (2004) 655–664.
[4] N. Jochmann, S. Götker, A. Tauch, Positive transcriptional control of the pyridoxal phosphate biosynthesis genes *pdxST* by the MocR-type regulator PdxR of *Corynebacterium glutamicum* ATCC 13032, Microbiology 157 (2011) 77–88.
[5] R. Percudani, A. Peracchi, A genomic overview of pyridoxal-phosphate-dependent enzymes, EMBO Rep. 4 (2003) 850–854.
[6] R. Percudani, A. Peracchi, The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families, BMC Bioinform. 10 (2009) 273.
[7] UniProt Consortium, Ongoing and future developments at the Universal Protein Resource, Nucleic Acids Res. 39 (2011) D214–D219.
[8] S.R. Eddy, A new generation of homology search tools based on probabilistic inference, Genome Inform. 23 (2009) 205–211.
[9] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinform. 5 (2004) 113.
[10] J. Felsenstein, PHYLIP – Phylogeny inference package (version 3.2), Cladistics 5 (1989) 164–166.
[11] T.J. Carver, L.J. Mullan, Website update: a new graphical user interface to EMBOSS, Comp. Funct. Genomics. 3 (2002) 75–78.
[12] M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building, Mol. Biol. Evol. 27 (2010) 221–224.
[13] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview Version 2—a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (2009) 1189–1191.
[14] T. Tomita, T. Miyagawa, T. Miyazaki, S. Fushinobu, T. Kuzuyama, M. Nishiyama, Mechanism for multiple-substrates recognition of α-aminoadipate aminotransferase from *Thermus thermophilus*, Proteins 75 (2009) 348–359.
[15] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.
[16] S. Rigali, A. Derouaux, F. Giannotta, J. Dusart, Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies, J. Biol. Chem. 277 (2002) 12507–12515.
[17] V. Vindal, K. Suma, A. Ranjan, GntR family of regulators in *Mycobacterium smegmatis*: a sequence and structure based characterization, BMC Genomics. 8 (2007) 289.
[18] L. Ji, J. Xie, GntR family regulators of the pathogen of fish tuberculosis *Mycobacterium marinum*, Biochem. Biophys. Res. Commun. 410 (2011) 780–785.
[19] D. Schirch, S. Delle Fratte, S. Iurescia, A. Angelaccio, R. Contestabile, F. Bossa, V. Schirch, Function of the active-site lysine in *Escherichia coli* serine hydroxymethyltransferase, J. Biol. Chem. 268 (1993) 23132–23138.