Computational Biology and Chemistry 58 (2015) 55–61

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

The aspartate aminotransferase-like domain of *Firmicutes* MocR transcriptional regulators

Teresa Milano [a], Roberto Contestabile [a], Alessandra Lo Presti [b], Massimo Ciccozzi [b], Stefano Pascarella [a,*]

[a] Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università La Sapienza, 00185 Roma, Italy
[b] Dipartimento di Malattie Infettive, Parassitarie e Immunomediate, Istituto Superiore di Sanità, 00185 Roma, Italy

ARTICLE INFO

Article history:
Received 12 January 2015
Received in revised form 9 May 2015
Accepted 14 May 2015
Available online 21 May 2015

Keywords:
MocR
Pyridoxal-5′-phosphate
Fold type-I
Bioinformatics
Evolution
Bacterial transcriptional regulation

ABSTRACT

Bacterial MocR transcriptional regulators possess an N-terminal DNA-binding domain containing a conserved helix-turn-helix module and an effector-binding and/or oligomerization domain at the C-terminus, homologous to fold type-I pyridoxal 5′-phosphate (PLP) enzymes. Since a comprehensive structural analysis of the MocR regulators is still missing, a comparisons of *Firmicutes* MocR sequences was undertook to contribute to the understanding of the structural characteristics of the C-terminal domain of these proteins, and to shed light on the structural and functional relationship with fold type-I PLP enzymes. Results of this work suggest the presence of at least three subgroups within the MocR sequences and provide a guide for rational site-directed mutagenesis studies aimed at deciphering the structure-function relationships in this new protein family.

1. Introduction

The GntR family of bacterial transcriptional regulators has been recently described (Haydon and Guest, 1991) and named after the repressor of the gluconate operon in *Bacillus subtilis*. GntR regulators possess an N-terminal domain containing a conserved helix-turn-helix (HTH) module able to recognize specific DNA motifs and a C-terminal effector-binding and/or oligomerization domain (Eb/O). According to current models, binding of the effector molecule to the Eb/O domain would induce a conformational change that influences the DNA-binding properties of the HTH domain. Within the GntR family, the structure of the C-terminus domain is heterogeneous and, depending on its nature, several subfamilies can be identified (Hoskisson and Rigali, 2009). The MocR subfamily is particularly interesting because it is distinguished by a domain homologous to fold type-I pyridoxal 5′-phosphate (PLP) enzymes (Paiardini et al., 2004; Percudani and Peracchi, 2009; Schneider et al., 2000). The type-I archetypal protein is aspartate aminotransferase (AAT), the first PLP-dependent enzyme of fold type-I whose three-dimensional structure has been elucidated (Kirsch et al., 1984). AAT-like PLP-enzymes were until recently deemed to exist mainly as stand-alone homodimers or as higher order complexes. It is now emerging, after scrutiny of genomic sequence data from prokaryotes, that fold type-I proteins can be found also in multidomain contexts such as nonribosomal peptide synthetases (Milano et al., 2013). The MocR regulators themselves represent a multidomain framework hosting AAT-like modules.

MocR regulators are widespread among eubacteria genomes, although with heterogeneity (Bramucci et al., 2011; Hoskisson and Rigali, 2009): apparently, they are rare or even absent in bacteria phyla characterized by compact genomes (size of about 1.0–2.0 Mb) such as *Chlamydiae* and *Mollicutes*, while in other species are present with a variable number of paralogous genes. For example, one MocR-like protein has been found only in the *Mollicutes bacterium* HR1 (genome size 1.88 Mb) out of 52 *Mollicutes* complete proteomes tested. Likewise, a single MocR regulator has been found in the bacterium *Opitutus terrae* (genome size 5.9 Mb) out of 20 *Chlamydiae* complete proteomes scrutinized (Bramucci et al., 2011). Only a few members of the MocR family have been structurally and functionally characterized: for example, GabR that regulates the utilization of γ-aminobutyric acid (GABA) in *B. subtilis* (Belitsky, 2004) and PdxR from *Listeria monocytogenes* and *Corynebacterium glutamicum*, involved in the regulation of PLP biosynthesis (Belitsky, 2014; Jochmann et al., 2011). Recently, the crystallographic structure of the GabR protein, deposited in the Protein Data Bank (PDB) with the codes 4MGR, 4N0B, and 4TV7 (Edayathumangalam et al., 2013; Okuda et al., 2015), confirmed the

* Corresponding author. Tel.: +39 06 49917694; fax: +39 06 49917566.
E-mail address: Stefano.Pascarella@uniroma1.it (S. Pascarella).

http://dx.doi.org/10.1016/j.compbiolchem.2015.05.003
1476-9271/

presence of a fold type-I domain at the C-terminus of the regulator and opened new perspectives in deciphering the mechanism of action of MocRs. GabR is a domain swap homodimer in which the HTH domain of one subunit interacts with the PLP-binding domain of the other subunit; this is possible thanks to a very long linker that connects the two domains.

Currently, a comprehensive structural analysis of the MocR family is missing. The aim of this work was to give a contribution to the understanding of the structural characteristics of the AAT-like domain of these regulators, and to shed light on its structural and functional relationship with PLP-dependent enzymes. For ease of data analysis, our work has been restricted to the *Firmicutes* phylum. Our results also provide a rational guide for site-directed mutagenesis studies aimed at deciphering the structure–function relationships in these new and intriguing family of regulators.

2. Materials and methods

2.1. Data collection

Complete *Firmicutes* proteomes were retrieved from the UniProt database release 2013–11 (UniProt, 2014). Putative MocR sequences were identified using the software HMMER 3.0 (Finn et al., 2011) while the required seed multiple alignment was calculated with the program MUSCLE version 3.8 (Bramucci et al., 2011; Edgar, 2004). Data analysis was assisted by in-house written Perl, Python and Bash scripts. The retrieved putative MocR sequences were compared to the models of the Pfam (Protein families) database 27.0 (Finn et al., 2014). The AAT-like domains were extracted from the parent sequences according to the Pfam assignment (code PF00155). Redundancy was removed with the program BlastClust release 2.1.2 (Altschul et al., 1990) using 50% sequence identity cutoff. NCBI CDD (Conserved Domain Database version 3.10) search website tools (Marchler-Bauer et al., 2003) were used to confirm correct assignment of AAT-like fold to the selected domains. Alignment editing and visualization relied on Jalview editor version 2.8 (Waterhouse et al., 2009). Logos were calculated with the WebLogo server (Crooks et al., 2004). Nucleotide sequences coding for the selected MocR proteins were extracted from GenBank release 198 (Benson et al., 2013).

2.2. Phylogenetic analysis

Protein sequence alignments were analyzed with the software MEGA version 6.0 (Tamura et al., 2013). Maximum likelihood (ML) phylogenetic reconstruction using Whelan and Goldman (Whelan and Goldman, 2001) substitution model and elimination of positions with less than 90% site coverage in the input multiple alignment, were utilized. Final tree was tested with 1000 random bootstrap replicates (Felsenstein, 1988). The nucleotide sequence alignment obtained with MUSCLE (Edgar, 2004) was manually edited with BioEdit software version 7.2.5 (Ciccozzi et al., 2011). ModelTest version 3.7 (Posada and Buckley, 2004) was used to select the simplest evolutionary model that best fitted the

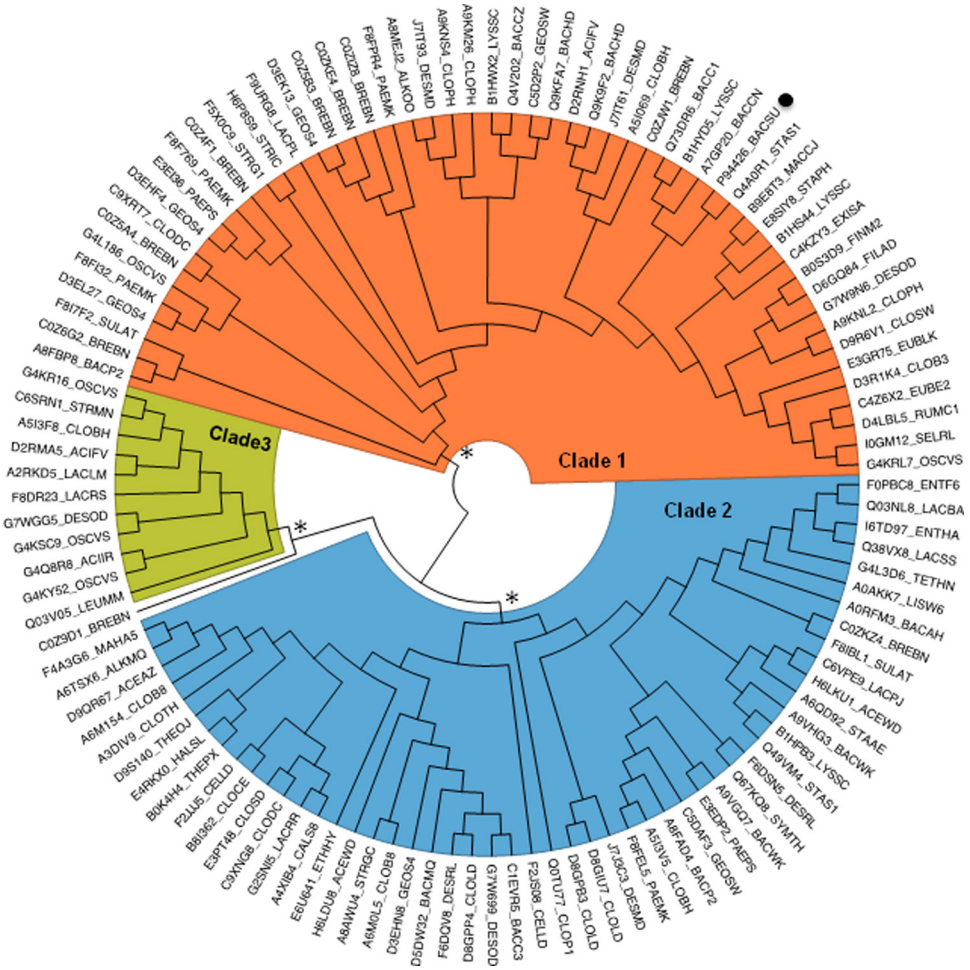

Fig. 1. Phylogenetic tree. Asterisks indicate nodes supported by at least 50% of the bootstrap replicas. Subfamily branches were distinguished by areas of different colors. GabR is marked with a grey circle. Sequences are labeled with the UniProt identifier. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)