Journal of Biotechnology 143 (2009) 239–246

Contents lists available at ScienceDirect

Journal of Biotechnology

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

The GlxR regulon of the amino acid producer *Corynebacterium glutamicum*: Detection of the corynebacterial core regulon and integration into the transcriptional regulatory network model

Thomas A. Kohl[a,b], Andreas Tauch[a,*]

[a] *Institut für Genomforschung und Systembiologie, Centrum für Biotechnologie, Universität Bielefeld, Universitätsstraße 27, D-33615 Bielefeld, Germany*
[b] *International NRW Graduate School in Bioinformatics and Genome Research, Centrum für Biotechnologie, Universität Bielefeld, Universitätsstraße 25, D-33615 Bielefeld, Germany*

ARTICLE INFO

Article history:
Received 29 April 2009
Received in revised form 30 July 2009
Accepted 4 August 2009

Keywords:
Corynebacterium glutamicum
Comparative genomics
Transcriptional regulatory network
Transcription factor binding site
Global regulator
GlxR

ABSTRACT

Corynebacterium glutamicum is an industrially important producer of amino acids and an emerging model system for the *Corynebacterineae*. The *glxR* gene of *C. glutamicum* ATCC 13032 encodes a DNA binding transcription factor of the Crp-Fnr protein family. Available data indicated a prominent role of GlxR in the transcriptional regulatory network of *C. glutamicum*. We have used recently published whole-genome annotations of several corynebacteria to derive further evidence for GlxR-mediated transcriptional regulation from a comparative genomics approach. A bioinformatic strategy detected 22 genes belonging to a proposed GlxR core regulon in corynebacteria. Binding of purified GlxR protein to 40-mer oligonucleotides representing predicted GlxR binding sites in the upstream region of core genes from *C. glutamicum* ATCC 13032 was verified *in vitro* by electrophoretic mobility shift assays. Based on current data, a reliable genome-scale prediction of GlxR binding sites was performed, indicating that about 14% of the annotated *C. glutamicum* genes might be under direct transcriptional control by GlxR. Integration of GlxR-mediated regulatory interactions with the data stored in CoryneRegNet enabled the reconstruction of a refined transcriptional regulatory network model for *C. glutamicum*. This network model exhibited a hierarchical and modular structure and is characterized by the presence of master regulators in each functional module, with GlxR serving as the dominating regulatory hub.

1. Introduction

Cyclic adenosine monophosphate (cAMP) is a common and important signaling molecule in bacteria (Botsford and Harman, 1992; Pesavento and Hengge, 2009). A prominent mediator of cAMP-dependent signaling is the cAMP receptor protein (Crp), which is enabled by bound cAMP to recognize specific operator sequences in the regulatory region of its target genes. Crp-dependent transcriptional regulation is a common mechanism of carbon catabolite repression enabling the sequential utilization of preferred carbon sources (Deutscher, 2008). The first and best studied member of the Crp branch of the Crp-Fnr protein family, herein named $Crp_{E.coli}$, was identified in *Escherichia coli* (Zubay et al., 1970). In this organism, the synthesis of cAMP by adenylate cyclase is stimulated by a component of the phosphoenolpyruvate:sugar phosphotransferase system, and cAMP levels are coupled to cellular energy levels (Park et al., 2006; Bettenbrock et al., 2007). $Crp_{E.coli}$ has consistently been recognized as global regulator in the transcriptional regulatory network of *E. coli*, being involved in the regulation of about 400 genes in distinct cellular functionality (Freyre-González et al., 2008; Martínez-Antonio et al., 2008). In this way, $Crp_{E.coli}$ serves as regulatory hub, connecting several functional modules to the network and mediating the cellular response to a high-level signal conveyed by the varying concentrations of cAMP (Kolb et al., 1993).

In recent years, transcription regulators orthologous to Crp have been identified in several bacterial species (Körner et al., 2003). Characterized orthologs are connected with control of important cellular functions and virulence, e.g., Clp (Hsiao et al., 2005), Crp_{Mtub} (Bai et al., 2005; Rickman et al., 2005), Crp_{Mbov} (Bai et al., 2007), Crp_{Scoel} (Piette et al., 2005), SdrP (Agari et al., 2008), Vfr (Kanack et al., 2006), and GlxR (Kim et al., 2004). These studies revealed the apparent conservation of both, the effector molecule cAMP and the operator sequence TGTGA-N_6-TCACA for the majority of the listed regulatory proteins. GlxR, the Crp homolog in *Corynebacterium glutamicum*, has been intensively investigated in the last years (Kim et al., 2004; Letek et al., 2006; Han et al., 2007; Han et al., 2008; Jungwirth et al., 2008; Kohl et al., 2008). *C. glutamicum* is an

* Corresponding author. Tel.: +49 521 106 8703; fax: +49 521 106 89041.
E-mail address: tauch@cebitec.uni-bielefeld.de (A. Tauch).

0168-1656/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2009.08.005

industrially important producer of amino acids and an emerging model system for the suborder *Corynebacterineae*. Biotechnological production of amino acids by fermentation with *C. glutamicum* represents a large market volume estimated annually at 1,500,000 metric tons L-glutamate and 900,000 metric tons L-lysine.

The genome of *C. glutamicum* ATCC 13032 contains three genes coding for transcription regulators of the Crp-Fnr protein family, *glxR*, *cg1327* and *cg3291* (Brune et al., 2005). A BLASTP analysis indicates *cg1327* and *cg3291* to be potentially paralogous genes, sharing 55% identity and 73% similarity on the deduced amino acid level, while they differ in G+C content by more than 15%. Both genes have not been characterized yet, and among sequenced corynebacteria only the closely related *C. efficiens* features a homologous protein encoded by *ce1525* (Nishio et al., 2003). In contrast, *glxR* belongs to the small core set of regulatory genes that are conserved in all hitherto sequenced corynebacteria (Brinkrolf et al., 2007). Initially, GlxR has been characterized as transcription regulator of genes of the glyoxylate bypass (Kim et al., 2004), genes involved in gluconate metabolism (Letek et al., 2006), and the resuscitation promoting factor 2 gene (Jungwirth et al., 2008). The concentration of the effector molecule cAMP seems to be dependent on the carbon source in *C. glutamicum* (Kim et al., 2004; Polen et al., 2007). Recently, we reported a global survey of GlxR binding sites in the genome of *C. glutamicum* ATCC 13032, applying a combined approach consisting of *in silico* detection and *in vitro* verification by electrophoretic mobility shift assays (Kohl et al., 2008). Thus, we were able to confirm the suggested GlxR involvement in the regulation of individual genes of *C. glutamicum* and to reveal GlxR as global regulator in the transcriptional regulatory network. GlxR is apparently involved in regulating carbohydrate metabolism, aromatic compound degradation, respiration, glutamate and nitrogen metabolism, fatty acid biosynthesis, resuscitation, and the cellular stress response (Kohl et al., 2008).

In this study, we aim at further expansion of the GlxR regulon to enable a more detailed reconstruction of the transcriptional regulatory network of *C. glutamicum*. By performing comparative genomics among sequenced corynebacteria, a proposed core regulon of GlxR and its orthologs was detected, providing reliable DNA binding sites and target genes to expand the known GlxR regulon. The resulting data was integrated into the transcriptional regulatory network of *C. glutamicum*, revealing a hierarchical and modular structure of the model, with GlxR representing the central regulatory hub.

2. Material and methods

2.1. Protein purification and DNA band shifts

Purification of hexahistidyl-tagged GlxR (GlxR-His_6) and subsequent electrophoretic mobility shift assays (EMSAs) were carried out as described previously (Kohl et al., 2008; Letek et al., 2006). Briefly, *E. coli* expression strain ER2655 (New England Biolabs) was transformed with a vector-encoded GlxR-His_6 gene present on pETCRP (Letek et al., 2006), grown at 37 °C in LB medium supplemented with 50 μg/ml kanamycin, and set to 17 °C after induction with 1 mM IPTG. GlxR-His_6 was purified by His-tag affinity chromatography, and the obtained protein was analyzed by SDS gel electrophoresis separation, followed by MALDI-TOF mass spectrometry for identification (Hansmeier et al., 2006). EMSAs contained 1 pmol Cy3-labeled 40-mer oligonucleotides with the 16 bp GlxR binding motif in its center, 60 pmol purified GlxR, 200 μM cAMP, 13% (v/v) glycerin, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, and 1 μg herring sperm DNA in a 20-μl volume. After 20 min incubation at 30 °C, assays were separated on 2% agarose gels, and resulting band shifts were visualized using a Typhoon 8600 imager (Amersham Biosciences). Both electrophoresis buffer and gel chamber were precooled to 4 °C to prevent early dissolution of the protein–DNA complex.

2.2. Comparative genomics

To extend the *C. glutamicum* GlxR regulon by means of comparative genomics with other sequenced corynebacteria, we first analyzed their genomes for potential orthologs to *C. glutamicum* ATCC 13032 proteins. The genomes of seven strains were included in this study, *C. glutamicum* ATCC 13032 (Kalinowski et al., 2003), *C. glutamicum* R (Yukawa et al., 2007), *C. efficiens* YS-314 (Nishio et al., 2003), *C. diphtheriae* NCTC 13129 (Cerdeño-Tárraga et al., 2003), *C. jeikeium* K411 (Tauch et al., 2005), *C. kroppenstedtii* DSM 44385 (Tauch et al., 2008a), and *C. urealyticum* DSM 7109 (Tauch et al., 2008b). Identification of potential orthologs was achieved by a bi-directional best blast analysis (Huynen and Bork, 1998). We used BLASTP with an *E*-value threshold of 10^{-5} to compare protein coding genes of the respective genome with those of *C. glutamicum* ATCC 13032, yielding amino acid sequence pairs, so-called bi-directional best hits. These hits represented the reciprocal best alignments of protein sequences, and respective genes were considered as putative orthologs.

2.3. Detection of GlxR binding sites

The prediction of orthologous genes was combined with an analysis of their upstream regions for the presence of GlxR binding sites. DNA binding sites were detected using the PoSSuMsearch algorithm (Beckstette et al., 2006) with all upstream regions of the organism, extracted from the genomic sequence as described previously for *C. glutamicum* (Kohl et al., 2008). The position weight matrix (PWM)-based model of the GlxR binding motif used in the search was derived from all verified binding sites in *C. glutamicum* ATCC 13032. Upstream regions were defined as the non-coding sequence in front of a protein coding region, ranging from +20 relative to the start codon up to a maxium length of 540 bp. A regulatory interaction by GlxR was considered conserved if a potential ortholog of a target gene was specified by the presence of a motif instance of the GlxR binding motif in its upstream region (Baumbach et al., 2009). Furthermore, if regulation by GlxR of orthologous genes was predicted for at least six out of the seven genomes included in the analysis, the genes were considered members of the GlxR core regulon.

3. Results and discussion

3.1. Detection of highly conserved GlxR target genes in corynebacteria

A homolog of the transcription regulator GlxR was detected in each hitherto sequenced corynebacterial genome (Brune et al., 2005; Tauch et al., 2008a,b). Conservation of GlxR-mediated regulation was analyzed in seven corynebacteria to identify members of a putative GlxR core regulon and to extend the GlxR regulon of *C. glutamicum* ATCC 13032. An overview of the employed workflow is presented in Fig. 1. The implemented motif detection pipeline, extraction of upstream sequences, position weight matrix modeling, and motif detection was performed for all sequenced strains, i.e., *C. glutamicum* ATCC 13032, *C. glutamicum* R, *C. efficiens* YS-314, *C. diphtheriae* NCTC 13129, *C. jeikeium* K411, *C. kroppenstedtii* DSM 44385, and *C. urealyticum* DSM 7109. All hitherto characterized GlxR binding sites, 54 motif instances in total (Jungwirth et al., 2008; Kohl et al., 2008), were used to build the motif model. Extraction of upstream sequences was executed with a maximum length of 540 bp. The *E*-value threshold for PoSSuMsearch was set to 400,