Available online at www.sciencedirect.com
ScienceDirect
Journal of Biotechnology 129 (2007) 191–211
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

The transcriptional regulatory network of the amino acid producer *Corynebacterium glutamicum*

Karina Brinkrolf [a,b], Iris Brune [a], Andreas Tauch [a,*]

[a] *Institut für Genomforschung, Centrum für Biotechnologie, Universität Bielefeld, Universitätsstraße 25, 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*

Received 8 September 2006; received in revised form 14 November 2006; accepted 4 December 2006

Abstract

The complete nucleotide sequence of the *Corynebacterium glutamicum* ATCC 13032 genome was previously determined and allowed the reliable prediction of 3002 protein-coding genes within this genome. Using computational methods, we have defined 158 genes, which form the minimal repertoire for proteins that presumably act as transcriptional regulators of gene expression. Most of these regulatory proteins have a direct role as DNA-binding transcriptional regulator, while others either have less well-defined functions in transcriptional regulation or even more general functions, such as the sigma factors. Recent advances in genome-wide transcriptional profiling of *C. glutamicum* generated a huge amount of data on regulation of gene expression. To understand transcriptional regulation of gene expression from the perspective of systems biology, rather than from the analysis of an individual regulatory protein, we compiled the current knowledge on the defined DNA-binding transcriptional regulators and their physiological role in modulating transcription in response to environmental signals. This comprehensive data collection provides a solid basis for database-guided reconstructions of the gene regulatory network of *C. glutamicum*, currently comprising 56 transcriptional regulators that exert 411 regulatory interactions to control gene expression. A graphical reconstruction revealed first insights into the functional modularity, the hierarchical architecture and the topological design principles of the transcriptional regulatory network of *C. glutamicum*.

Keywords: *Corynebacterium glutamicum*; Transcriptional regulator; Regulatory network; Network motif; Topological design; Systems biology

1. Introduction

One of the present challenges in genome research is the organization of experimental data originating from high-throughput technologies, such as DNA sequencing and transcriptomics. New generations of automated

* Corresponding author. Fax: +49 521 106 5626.
E-mail address: andreas.tauch@genetik.uni-bielefeld.de (A. Tauch).

0168-1656/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2006.12.013

DNA sequencers enable the ultra-fast sequencing of microbial genomes (Margulies et al., 2005; Goldberg et al., 2006), and subsequent DNA microarray analysis permits global transcriptional profiling of the bacterial cell (Lucchini et al., 2001). The huge amount of experimental data generated by these technologies leads to a rapid growth of information about a microorganism of interest. In particular, the knowledge of a complete genome sequence allows, together with high-density microarray technology, the monitoring of information flow leading to the modulation of specific cellular functions in response to the corresponding environmental stimuli (Herrgård et al., 2004). One way how to organize this information from genome-wide transcriptional profiling is to form networks of interactions between the respective cellular entities (Reed and Palsson, 2003). There are three basic components that are crucial to reconstruct the regulatory interactions in a bacterial cell: (i) the DNA-binding transcriptional regulators, (ii) the DNA-binding sites (operators) of the regulatory proteins in the genome sequence and (iii) the regulated target genes (Stormo and Tan, 2002). Defining the repertoire of regulatory genes within a bacterial genome sequence is a relatively easy task, since the deduced proteins can be classified into regulatory protein families on the basis of their amino acid sequence similarity (Marchler-Bauer et al., 2005). Moreover, most of the transcriptional regulatory proteins belong to the helix-turn-helix (HTH) family of DNA-binding proteins, which can be recognized by specific amino acid sequence signatures (Gough et al., 2001; Aravind et al., 2005). On the other hand, genome-wide detection of the cognate DNA-binding sites of a transcriptional regulator is a demanding task, but it is a prerequisite to reveal target genes and to deduce thereof the topology of the gene regulatory network. DNA-binding sites can be discovered either by a variety of computational methods (Stormo and Tan, 2002; Tompa et al., 2005) or by experimental techniques including, for instance, global transcriptional profiling in combination with pattern-recognition methods to obtain the set of co-regulated genes that are under direct transcriptional control by a specific regulatory protein (Stormo and Tan, 2002; Herrgård et al., 2004). This global mapping of transcriptional regulatory interactions provides information on the associations between distinct cellular entities and allows the creation of a diagram of directional connections between the transcriptional regulators and their target genes. When considering additionally the physiological role of the regulatory protein as either activator or repressor of gene expression, a qualitative indicator on how the target genes are regulated (positively or negatively) can be included into the diagram, resulting in a regulatory network topology with qualitative directional connections (Rice et al., 2005).

The currently best-characterized system of regulatory interactions in bacteria is the transcriptional regulatory network of *Escherichia coli* that provides valuable insights into the topological organization and evolution of a bacterial gene regulatory network (Shen-Orr et al., 2002; Madan Babu and Teichmann, 2003; Martínez-Antonio et al., 2006; Lozada-Chávez et al., 2006). By using a combination of computer-assisted methods, 314 genes encoding potential transcriptional regulators were defined, apparently comprising 43% repressors, 35% activators and 22% dual regulators (Pérez-Rueda and Collado-Vides, 2000). Bioinformatics evaluation of these data revealed a multi-layer hierarchical architecture of the gene regulatory network, lacking any feedback regulation at the transcriptional level (Martínez-Antonio and Collado-Vides, 2003; Ma et al., 2004a,b). The majority of direct transcriptional regulatory interactions in *E. coli* are composed of conserved network motifs, including the feed-forward loop, the single input motif, dense overlapping regulons, and the bi-fan motif (Shen-Orr et al., 2002; Dobrin et al., 2004). In addition, autoregulation, the multi-input motif and regulatory cascades by so-called regulator chains are known motifs in gene regulatory networks (Yu et al., 2003). Each conserved network motif has a specific function in determining gene expression and thus the cellular response to fluctuating internal or external signals (Shen-Orr et al., 2002; Yu et al., 2003; Mangan and Alon, 2003). A further feature of the *E. coli* gene regulatory network is the modularity (Resendis-Antonio et al., 2005). A regulatory module comprises parts of a network structure that perform a common physiological function and may be linked by regulatory interactions. Modularity contributes to the robustness of the entire network, by confining damage to a distinct part and thereby preventing the spread of damage into other parts of the network (Aderem, 2005). Different functional modules of the *E. coli* gene regulatory network are preferably linked by bi-fan motifs, whereas feed-forward loops tend to