High-resolution detection of DNA binding sites of the global transcriptional regulator GlxR in Corynebacterium glutamicum

Britta Jungwirth,1 Claudia Sala,2 Thomas A. Kohl,1 Swapna Uplekar,2 Jan Baumbach,3 Stewart T. Cole,2 Alfred Pühler1 and Andreas Tauch1

1Institut für Genomforschung und Systembiologie, Centrum für Biotechnologie, Universität Bielefeld, Bielefeld, Germany
2Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
3Computational Systems Biology, Max-Planck-Institut für Informatik, Saarbrücken, Germany

The transcriptional regulator GlxR has been characterized as a global hub within the gene-regulatory network of Corynebacterium glutamicum. Chromatin immunoprecipitation with a specific anti-GlxR antibody and subsequent high-throughput sequencing (ChIP-seq) was applied to C. glutamicum to get new in vivo insights into the gene composition of the GlxR regulon. In a comparative approach, C. glutamicum cells were grown with either glucose or acetate as the sole carbon source prior to immunoprecipitation. High-throughput sequencing resulted in 69 million reads and 2.6 Gb of genomic information. After mapping of these data on the genome sequence of C. glutamicum, 107 enriched DNA fragments were detected from cells grown with glucose as carbon source. GlxR binding sites were identified in the sequence of 79 enriched DNA fragments, of which 21 sites were not previously reported. Electrophoretic mobility shift assays with 40-mer oligomers covering the GlxR binding sites were performed for validation of the in vivo results. The detection of new binding sites confirmed the role of GlxR as a regulator of carbon source metabolism and energy conversion, but additionally revealed binding of GlxR in front of the 6C non-coding RNA gene and to non-canonical DNA binding sites within protein-coding regions. The present study underlines the dynamics within the GlxR regulon by identifying in vivo targets during growth on glucose and contributes to the expansion of knowledge of this important transcriptional regulator.

INTRODUCTION

In the age of massive sequencing of bacterial genomes, transcriptional regulatory networks (TRNs) are interesting to study in order to understand how the regulation of mRNA synthesis contributes to the dynamic expression of biological functions. These networks are composed of transcriptional regulatory proteins, their target genes and the regulatory interactions exerted (Babu et al., 2004). Transcriptional regulatory proteins play a key role in the control of gene expression in prokaryotic cells, as they integrate different signals from the environment or the cellular metabolism to adapt the expression levels for an adequate response to changing conditions (Pollack & Iyer, 2002).

Various approaches exist to elucidate the structure and function of a prokaryotic TRN (Babu et al., 2009). For in vivo investigations, DNA microarray experiments are used to compare distinct strains of a bacterial species or different growth conditions, thereby identifying regulated targets and transcriptional regulators involved (Dharmadi & Gonzalez, 2004; Rodionov, 2007). More recently, chromatin immunoprecipitation (ChIP) with specific antibodies against RNA polymerase or bacterial transcriptional regulators has been successfully applied (Laub et al., 2002). The combination of ChIP with the detection of enriched DNA fragments on a DNA microarray (ChIP-chip) led to the global identification of binding regions for nucleoid-associated and transcriptional regulatory proteins in bacteria (Grainger & Busby, 2008; Wade et al., 2007). Using high-throughput sequencing technologies, it is now possible to investigate the binding sites

1Present address: Molekulare Mycobakteriologie, Forschungszentrum Borstel, 23845 Borstel, Germany.

Abbreviations: ChIP, chromatin immunoprecipitation; ChIP-seq, ChIP with sequencing; COG, Clusters of Orthologous Groups; CRP, cAMP receptor protein; EMSA, electrophoretic mobility shift assay; IP, immunoprecipitation; PWM, position weight matrix; TRN, transcriptional regulatory network.

Three supplementary tables are available with the online version of this paper.
of DNA-interacting proteins by sequencing the immunoprecipitated fragments (ChIP-seq) (Park, 2009). The major advantage of ChIP-seq, when compared with ChIP-chip, is the higher resolution of binding site detection. Recently, this method was used for the first time in Escherichia coli to study the genome-wide distribution of the nucleoid-associated proteins Fis and H-NS (Kahramanoglou et al., 2011). There are few further ChIP-seq studies in prokaryotes so far, reporting on binding of transcriptional regulators and a sigma factor of Mycobacterium tuberculosis, Pseudomonas syringae and Vibrio cholerae (Blasco et al., 2012; Butcher et al., 2011; Davies et al., 2011; Markel et al., 2011; Smollett et al., 2012).

We applied ChIP-seq to the industrially relevant actinobacterium Corynebacterium glutamicum to characterize the regulon of the global transcriptional regulator GtxR. The Crp–Fnr family regulator GtxR is composed of an amino-terminal cAMP-binding domain and a carboxy-terminal DNA recognition helix–turn–helix motif, and interacts with its binding sites with the consensus sequence 5'-TGTGANNNTANNTACA-3' in a cAMP-dependent manner (Han et al., 2007; Jungwirth et al., 2008; Kohl & Tauch, 2009; Kohl et al., 2008). Many bacteria use CAMP as a second messenger in diverse signalling pathways (Botsford & Harman, 1992; Gomelsky, 2011; Rickemberg, 1974). In C. glutamicum, intracellular CAMP levels change depending on the metabolized carbon source. It has been shown that the amount of cAMP in C. glutamicum cells increases during growth on glucose and decreases during growth on acetate (Kim et al., 2004).

The role of GtxR as an important hub in the TRN of C. glutamicum has been analysed by electrophoretic mobility shift assay (EMSA) in vitro or in artificial in vivo systems, such as promoter probe assays (reviewed by Schröder & Tauch, 2010). As a result, a large number of DNA binding sites have been identified. The reference database CoryneRegNet currently lists 151 GtxR-regulated genes (as of September 2012), including the autoregulation of GtxR (Baumbach et al., 2009; Pauling et al., 2012). A total of 215 potential GtxR binding sites were predicted bioinformatically based on current experimental data (Kohl & Tauch, 2009). Recently, a ChIP-chip experiment confirmed the global role of GtxR in transcriptional regulation, detecting 209 potential target regions genome-wide (Toyoda et al., 2011). Among these, 84 regions contain previously described GtxR binding sites, and, based on motif discovery, 94 new binding motifs were identified. About one quarter (53/209) of the detected binding regions are presumably situated in coding regions or in intergenic regions between convergently organized genes (Toyoda et al., 2011).

In the current study, we used ChIP-seq for the detection of DNA binding sites in vivo, and thereby confirmed that GtxR acts as a global transcriptional regulator in response to changing growth conditions, i.e. when comparing growth of C. glutamicum on glucose with that on acetate. Our data underline the role of GtxR as a regulator of alternative carbon source metabolism during growth on glucose and its implication in the control of many important cell functions, such as central metabolism, cell cycle control, cell wall turnover and stress responses.

**METHODS**

**Bacterial cultivation.** C. glutamicum ATCC 13032 cells were grown under aerobic conditions at 30 °C in minimal medium CGXII (Keilhauer et al., 1993) containing 30 mg protecatacic acid l⁻¹ and 420 µg thiamine l⁻¹. The carbon source (40 g glucose l⁻¹) was replaced by 5 g acetate l⁻¹ where appropriate. E. coli JM109 cells containing plasmid pETCRP (Lekt et al., 2006) were grown under aerobic conditions at 37 °C in Luria–Bertani (LB) medium (Sambrook et al., 1989) with 50 µg kanamycin ml⁻¹. Growth was monitored by measuring the OD₆₀₀ using a spectrophotometer (Eppendorf).

**Protein purification and antibody production.** Recombinant GtxR protein was expressed in E. coli JM109 cells carrying plasmid pETCRP and purified as described elsewhere (Lekt et al., 2006). The protein was washed and concentrated using Amicon Ultra-4 10K Centrifugal Filter units (Millipore). This preparation was used for commercial polyclonal antibody production in rabbits and for affinity purification of the anti-GtxR antiserum produced (Eurogentec).

**ChIP-seq.** The ChIP-seq protocol was based on previous experiments with M. tuberculosis (Sala et al., 2009), and was modified for use on C. glutamicum cells. Exponentially growing C. glutamicum glucose and acetate cultures were treated with 1 % formaldehyde (final concentration). After 10 min of incubation at 30 °C with gentle shaking, glycine was added to a final concentration of 125 mM. Cells were harvested by centrifugation, washed twice with ice-cold PBS (pH 7.4) and stored at −20 °C. Pellets were resuspended to a final concentration of 5 × 10⁹ cells ml⁻¹ in PBS with protease inhibitors (Mini Complete, Roche), and cell disruption was achieved using a RiboLyser (Precellys, Peqlab). Cell debris were removed by centrifugation and DNA in the supernatant was sheared to an average size of 200–500 bp by sonication in a water bath (Bioruptor, Diagenode). An input sample of 10 µl was taken and stored at −20 °C until RNase treatment. For immunoprecipitation (IP), 1 % Triton X-100, 0.1 % SDS and 0.1 % sodium deoxycholate (final concentrations) were added, and the samples were incubated overnight at 4 °C on a rotating wheel with 5 µg custom rabbit polyclonal anti-GtxR antibody ml⁻¹ (Eurogentec). Protein–DNA complexes were immunoprecipitated with 50 µl Dynabeads Protein G (Invitrogen) for 4 h at 4 °C. The magnetic beads were washed twice with IP buffer, once with IP buffer plus 500 mM NaCl, once with buffer III (10 mM Tris/HCl, pH 8, 250 mM LiCl, 1 mM EDTA, 0.5 % Nonidet-P40, 0.5 % sodium deoxycholate) and once with Tris-EDTA buffer (pH 7.5). For elution, beads were resuspended in 100 µl elution buffer (50 mM Tris/HCl, pH 7.5, 10 mM EDTA, 1 % SDS) and incubated at 65 °C for 30 min. IP and input samples were diluted with one volume of water and treated with RNase A. Proteinase K was added, and cross-linking was reversed by incubation at 50 °C for 2 h and 65 °C for 8 h. DNA was extracted with phenol/chloroform/isooamyl alcohol (25:24:1), precipitated and resuspended in water. DNA solutions were concentrated in a SpeedVac concentrator. Sequencing libraries with an average fragment size of 250 bp were prepared according to the Illumina ChIP-seq protocol and sequenced in a 38 bp single read run on an Illumina Genome Analyser IIX sequencing machine.

**Read mapping and data visualization.** Quality filtered reads were mapped to the C. glutamicum ATCC 13032 genome sequence (Kalinowski et al., 2003) with SORMAN (Blom et al., 2011), allowing...
for up to two errors per read (insertion, deletion and/or mismatch). Mapped sequence data were visualized using VAMP (Hackl et al., 2011).

**Peak detection.** The in-house tool CISA (Chromatin Immunoprecipitation Sequence Analyzer) was used for peak detection. As a first step in the algorithm, for each sample the number of reads that mapped to a specific genomic position was counted for both strands separately. The absolute difference in read counts between a sample and the background and the enrichment factors (quotient of signal and background) were derived for each base position. Next, the absolute difference in the read counts curve was smoothed by performing a convolution with a Gaussian filter with a variance of 40 bp.

Local maxima with a threshold of at least 40 counts above background were identified from these smoothed signals on both strands. These maxima served as seed positions in the algorithm from which possible bi-modal enrichment patterns were identified afterwards. Therefore, for each peak on the forward or reverse strand, corresponding peaks on the other strand within a window of 60–155 bp downstream or upstream were identified, constrained to exhibit approximately equal height and enrichment factors. Both the forward/reverse and reverse/forward peak pairs were then intersected and reported as enriched DNA fragments.

Additional filtering was performed by comparing the positions of the maxima on the differential expressed signal maps with the positions of maxima on the original smoothed signal for each strand, discarding bi-modal patterns with deviations larger than 16 bp.

Two CISA runs were performed with the glucose IP sample using the input sample and the acetate IP sample as background, respectively. Only enriched fragments detected in both runs within a window of 20 bp were taken into account. A total of 14 peaks were removed manually because of their localization within rrr operons or the tRNA^Glu and tRNA^Asp genes. These regions show a high variation in read counts independently of IP due to the fact that their sequences are not unique in the genome.

The sequences of the detected enriched DNA fragments were extracted, taking the genomic positions of the two peaks on each strand as starting points and adding 60 bp on each side of the enriched DNA fragment to cover the fragment entirely.

**Motif analysis and genome-wide search for GlxR binding sites.** For de novo motif discovery within the significantly enriched DNA fragments, their genomic sequences were submitted to MEME (Bailey & Elkan, 1994). Parameters were set to search for zero or one palindromic motif of 16 bp width per sequence. A genome-wide search for GlxR binding sites was performed using the PoSSuMsearch algorithm (Beckstette et al., 2006), as described previously (Kohl et al., 2008), but without restriction to intergenic regions. The position weight matrix (PWM) model of the binding motif was built based on all DNA binding sites, for which in vitro binding of GlxR using 40 bp oligomers has been reported (Jungwirth et al., 2008; Kohl & Tauch, 2009; Kohl et al., 2008).

Based on the manually curated annotation of the *C. glutamicum* ATCC 13032 genome sequence (Kalinski et al., 2003) and the distribution of 452 (as of September 2012) regulator binding sites in *C. glutamicum* (Pauling et al., 2012), upstream regions were defined for all coding sequences (CDSs) ranging from +20 to −600 bp relative to the translation start.

**Functional analysis of GlxR targets.** All genes for which GlxR binding was detected within the upstream or the coding region were considered to be target genes. They were classified according to their Clusters of Orthologous Groups (COG) categories. The distribution was compared with the COG distribution of all *C. glutamicum* genes. To test for statistical significance, two-sided *P* values were calculated using a Fisher’s exact test as defined in Agresti (1992). The significance level was set to *P* ≤ 0.05.

**DNA band shift assays.** EMSAs were carried out as established and validated previously (Jungwirth et al., 2008; Kohl & Tauch, 2009; Kohl et al., 2008), with the exception that 0.05 pmol labelled double-stranded DNA oligonucleotides were used. Oligomers were designed with the predicted GlxR binding motif in the centre and 12 bp of flanking genomic sequence on both sides. The oligonucleotide sequences for the binding sites with the genomic positions 804771–804786 and 2305775–2305790 were modified at the 3’ or 5’ ends of the flanking genomic regions to avoid secondary structure formation, as described previously (Kohl et al., 2008).

**RESULTS**

**IP of GlxR–DNA complexes and high-throughput sequencing**

A specific anti-GlxR antibody was used in ChIP-seq experiments with the objective of gaining more detailed insights into the GlxR regulon. We investigated the influence of varying cAMP levels on GlxR binding in *C. glutamicum*, since the homologous Crp protein from *E. coli* has been reported to lose its DNA binding capacity under low-cAMP conditions (Grainger et al., 2005). In *C. glutamicum*, intracellular cAMP levels are increased during growth on glucose and decrease during growth on acetate (Kim et al., 2004); thus, these growth conditions were selected for ChIP-seq experiments. Comparison of protein extracts from glucose- and acetate-metabolizing cells in Western blots revealed similar amounts of GlxR protein (data not shown). ChIP samples were prepared from exponentially growing cultures metabolizing each of the two carbon sources, and DNA fragments were sequenced. Additionally, an input DNA sample was sequenced from glucose-metabolizing cells before IP. Each sample was sequenced on one lane of a 38 bp single-read run. Of the 22.7–23.8 million reads (862–902 Mb) sequenced, 22.0–23.7 million reads were uniquely mapped to the *C. glutamicum* ATCC 13032 genome sequence (Table 1). The mapped reads provided the basis for the detection of enriched DNA fragments and candidate GlxR binding sites.

**Detection of enriched DNA fragments**

The characteristics of an enriched DNA fragment are exemplified in Fig. 1 for the promoter region of the acetaldehyde dehydrogenase gene *ald* containing a known GlxR binding site. The distinctive structure of a pair of (local) maxima of read counts or peaks is observed in the profile of the glucose IP sample, but not in those of the acetate IP or the input samples. The representations of two profiles in one diagram together with a third curve showing the calculated absolute difference between them (Fig. 1b) clearly demonstrates the enrichment in the glucose IP sample.
Table 1. Sequencing and mapping data for ChIP-seq with GlxR of C. glutamicum

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sequencing data</th>
<th>Mapping data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster (PF*)</td>
<td>Mapped reads</td>
</tr>
<tr>
<td></td>
<td>Sequenced kilobases</td>
<td>Unmapped reads</td>
</tr>
<tr>
<td>ChIP-seq glucose</td>
<td>77.8 ± 3.9 %</td>
<td>22 845 356</td>
</tr>
<tr>
<td></td>
<td>902 506</td>
<td>904 801</td>
</tr>
<tr>
<td>ChIP-seq acetate</td>
<td>80.6 ± 4.2 %</td>
<td>22 041 173</td>
</tr>
<tr>
<td></td>
<td>862 373</td>
<td>652 853</td>
</tr>
<tr>
<td>Input control</td>
<td>80.7 ± 4.3 %</td>
<td>22 188 776</td>
</tr>
<tr>
<td></td>
<td>862 509</td>
<td>508 829</td>
</tr>
</tbody>
</table>

*PF, Passed filter.

Fig. 1. Visualization of the sequencing results for the ald promoter region. The numbers of reads per basepair were counted for each of the three samples and are represented as profiles (a). The forward strand is shown on the upper scale and the reverse strand on the lower scale of each diagram. On the right-hand side, sequencing results for two samples are shown in one diagram as well as the absolute difference between them (b). The results from the glucose ChIP sample are represented as a dark-grey curve, the results from the second sample in light grey, and the absolute difference between the two in black. The curves are superposed with dark grey in the background, light grey in the middle and black in the foreground.
Inspection of the profiles at the genomic positions of previously known binding sites (Pauling et al., 2012; Toyoda et al., 2011) was performed in order to identify which sites had been enriched by IP with anti-GlxR antibody. The characteristic two-peak structure was present in the glucose IP sample for a subset of these sites. GlxR binding was identified for 32 DNA motifs that had shown interaction in EMSAs previously (Pauling et al., 2012) and for 57 loci detected by ChIP-chip (Toyoda et al., 2011). No or significantly less enrichment was observed in the acetate IP sample at GlxR binding sites. The data from the acetate IP sample were therefore used as a second background control for the automated peak detection with CISA, in addition to the input sample.

With CISA a total of 239 DNA fragments (see Table S1 available with the online version of this paper) were detected, which were enriched in the glucose IP sample over the input and the acetate IP sample. Enrichment factors were calculated as the quotient of read numbers in the glucose IP and the input sample, revealing a broad range of enrichment from 1.1- to 113.3-fold. Further analyses were focused on the 107 DNA fragments that were at least twofold enriched (Table S2). Out of the 107 DNA fragments, two were also enriched in the acetate IP sample over the input. They were localized in the adl promoter region (Fig. 1), which was 15.6-fold enriched in the glucose IP sample and 2.7-fold in the acetate IP sample, and in the intergenic region between the pta–ack operon and the fpr1 gene. This region, containing a predicted binding site (Kohl & Tauch, 2009), was 3.4-fold enriched in the acetate IP sample.

**De novo motif search within enriched DNA fragments**

The sequences of the 107 DNA fragments were submitted to the motif-based sequence analysis tool MEME (Bailey & Elkan, 1994) to detect conserved DNA motifs. The best common motif identified by MEME is shown in Fig. 2(a). The 16 bp palindromic sequence was found in 20 instances (18.7 % of cases) with an E-value of 3.6. The 16 bp palindromic sequence was found in 20 instances (18.7 % of cases) with an E-value of 3.6. The 16 bp palindromic sequence was found in 20 instances (18.7 % of cases) with an E-value of 3.6.

**Fig. 2.** (a) Weblogo of the motif detected with MEME (Bailey & Elkan, 1994) within the sequences of the 107 identified GlxR binding regions. (b) Distribution of the genomic localizations of the binding sites with regard to the closest situated genes. Binding within a region ranging from −600 bp to +20 bp relative to a translation start site was considered to be upstream of a transcription unit. Correspondingly, intergenic binding sites are situated in the shared upstream region of two divergently transcribed genes. For binding inside a coding sequence, we distinguished between purely intragenic positions (intragenic/none) and intragenic positions upstream of another divergent or convergent transcription unit (intragenic, but upstream).

**GlxR motif prediction and comparison of in vivo, in silico and in vitro data**

A PWM-based genome-wide search for GlxR binding sites was performed in order to validate the ChIP-seq data by a second in silico approach. This search procedure is advantageous for the detection of more instances of a given motif in the case of degenerate binding sites. In addition, it allows for the comparison of the ChIP-seq data with previous studies (Kohl & Tauch, 2009; Kohl et al., 2008). Subsequently, the 107 enriched DNA fragments were classified according to the occurrence of validated GlxR binding sites from prior EMSA studies or putative motifs predicted by PoSSuMsearch in their sequences (Table S2). Binding sites reported previously to interact with GlxR in EMSAs were located within 32 DNA fragments, including the described autoregulatory binding site (Jungwirth et al., 2008) in the glxR promoter region. Furthermore, 47 fragments contained at least one GlxR binding site predicted by PoSSuMsearch. About one quarter of all DNA fragments (28 out of 107) could not be associated with any DNA motif resembling the GlxR consensus binding site.

Additionally, the fragments were grouped into four categories in relation to the genomic position of the included motifs with regard to closely situated genes (Fig. 2b). For fragments without a motif, the genomic position
of the central 16 bp of the fragment was taken into account. The example of the ald gene (Fig. 1) illustrates GlxR binding upstream of a single transcription unit. Additional binding situations found within the dataset were: (i) intergenic binding upstream of two divergently orientated transcription units; (ii) intragenic binding inside a coding region; and (iii) intragenic binding upstream of a neighbouring gene. Examples of these genomic contexts of GlxR–DNA interaction are shown in Fig. 3. A special case is represented in Fig. 3(b), where protein binding in the intergenic region of the protein-coding gene cg0360 and the small RNA 6C is shown.

Due to the experimental design, all binding sites validated by EMSA in previous studies were located in either upstream or intergenic regions (Table S2, upper part). Interestingly, 59.6 % of the GlxR motifs detected with PoSSuMsearch (28 out of 47) shared these genomic environments (Table S2, middle part). Furthermore, our study validates GlxR binding inside coding regions, which had been suggested by the recent ChIP-chip analysis (Toyoda et al., 2011). With 19 out of 47 PoSSuMsearch-detected motifs and the majority of the fragments without a binding motif (27 out of 28) falling into this category, intragenic binding represented 43 % of the total of 107 GlxR–DNA interaction sites.

EMSAs and functional analysis of the GlxR regulon

To further validate the ChIP-seq results in vitro, EMSAs were designed taking into consideration all possible genomic contexts of GlxR binding in relation to neighbouring genes. Two sets of predicted binding sites were chosen for validation by EMSA: eight motifs with an E-value from PoSSuMsearch smaller than 400 and eight motifs with a higher E-value (Table S2). Gel retardation was observed for all oligomers containing a motif with a higher E-value (Table S2). Gel retardation was observed for all oligomers containing a motif with a higher E-value (Table S2). Gel retardation was observed for all oligomers containing a motif with a higher E-value (Table S2). Gel retardation was observed for all oligomers containing a motif with a higher E-value (Table S2). Gel retardation was observed for all oligomers containing a motif with a higher E-value (Table S2).

Interestingly, 59.6 % of the GlxR motifs detected with PoSSuMsearch (28 out of 47) shared these genomic environments (Table S2, middle part). Furthermore, our study validates GlxR binding inside coding regions, which had been suggested by the recent ChIP-chip analysis (Toyoda et al., 2011). With 19 out of 47 PoSSuMsearch-detected motifs and the majority of the fragments without a binding motif (27 out of 28) falling into this category, intragenic binding represented 43 % of the total of 107 GlxR–DNA interaction sites.

Among the other validated targets, there were genes from carbohydrate metabolism (iolR, iolC–cg0198–iolA–B–D–E–G–H operon, msiK2, prpD2–B2–C2 operon), transporter genes (cg2610, cg1305), a molecular chaperone gene (hscA) and a superfAMILY II DNA/RNA helicase gene (cg1307).

DISCUSSION

In the present study, ChIP-seq was, for the first time to our knowledge, applied to C. glutamicum. The method has rarely been used for bacteria until now, although it has been routinely exploited in eukaryotic cells for several years. To our knowledge, seven ChIP-seq studies have been published on bacterial transcription factor, sigma factor and nucleoid-associated protein binding sites (Blasco et al., 2012; Butcher et al., 2011; Davies et al., 2011; Jutras et al., 2012; Kahramanoglou et al., 2011; Markel et al., 2011; Smollett et al., 2012), and one dataset was used in a bioinformatics approach (Lun et al., 2009). While epitope-tagged proteins were used in most of the above-mentioned studies, here, the Crp-family protein GlxR was immunoprecipitated with a specific antibody from cells growing on glucose or acetate as the sole carbon source. Sequencing of the ChIP samples in conjunction with automated peak finding led to the detection of 107 DNA fragments that were more than twofold enriched by IP from glucose-metabolizing C. glutamicum cells.

Recently, the importance of improving peak detection tools for ChIP-seq datasets has been underlined (Händstad et al., 2011; Rye et al., 2011; Wilbanks & Facciotti, 2010). In our study, the method for peak detection relies on the characteristic features of a peak, such as the appearance of one read count maximum on the leading followed by a second maximum on the lagging strand. By using two background controls, the present study responds to another suggestion for amelioration (Rye et al., 2011). Furthermore, we evaluated the ChIP-seq results using a de novo and a PWM-based motif discovery approach. This approach facilitated the comparison with existing data and led to the identification and in vitro validation of new binding sites. It might also be useful for the estimation of false positives detected by ChIP-seq, as the presence of a DNA motif resembling the GlxR consensus within an
enriched fragment provides additional evidence for a true binding site. However, the in vivo situation can differ from in silico predictions and in vitro conditions. For detailed analysis of a specific binding site, further in vivo studies are needed, going far beyond the global picture given by ChIP-seq.

In accordance with the fact that GlxR is known to interact with DNA in a cAMP-dependent manner (Han et al., 2007; Jungwirth et al., 2008; Kim et al., 2004; Kohl & Tauch, 2009; Kohl et al., 2008), we observed significantly less or no DNA enrichment in acetate-metabolizing cells where the cAMP level was decreased. Similarly, a significant drop in

Fig. 3. Visualization of sequencing results for four enriched DNA fragments. Representation of (a) the intergenic region between cg1142 and cg1143, (b) the intergenic region between the protein-coding gene cg0360 and the small RNA 6C, (c) the intragenic binding within cg0968 upstream of cg0967, and (d) the intragenic binding within cg0875. The peaks shown in (a) and (b) contain GlxR motifs located upstream of either of the two divergently transcribed genes, and the protein could be implicated in the regulation of both transcription units. The peaks in (c) and (d) are situated inside the coding sequence.
the number of binding regions and the enrichment factors was described in a recent ChIP-chip study with GlxR (Toyoda et al., 2011) after deletion of the cyaB gene, encoding the only known adenylate cyclase of C. glutamicum. Nevertheless, cAMP is still detectable in the cyaB deletion strain (Cha et al., 2010), and the decrease of GlxR binding was not as marked as in our study (Toyoda et al., 2011). The present result is also in accordance with the situation in E. coli, where lowering the intracellular cAMP level has been reported to result in a loss of Crp receptor protein (CRP) binding by ChIP-chip (Grainger et al., 2005).

As the effect observed in our study was very drastic, and only two out of 107 fragments were also enriched after IP from acetate-grown cells, there are probably additional factors involved besides the differences in cAMP levels. An explanation might be the overlap in the function of GlxR with other transcriptional regulators such as RamA and RamB in acetate-metabolizing cells (reviewed by Schröder & Tauch, 2010).

GlxR is also directly connected to several other transcriptional regulator genes. Using ChIP-seq, iolR, whcA, cg0343, gabR, whcE, cg1143, fruR, vanR, genR, cg3388 and the two-component system citAB were identified as GlxR targets during growth on glucose. GlxR is therefore involved in hierarchical regulation, which enables the cell to dynamically control the expression of different metabolic subsystems. This confirms the role of the GlxR protein as a global hub within the TRN of C. glutamicum.

The ChIP-seq experiment revealed GlxR binding in the upstream regions of genes for alternative carbon source uptake and metabolism. Interestingly, genes involved in the carbohydrate and central metabolism and transport systems (COG categories C, G, E, P and Q) were significantly enriched among GlxR targets as compared with the genome. The regulator thus shows some characteristics observed for homologous Crp proteins from E. coli and other Gram-negative bacteria in relation to carbon catabolite repression (Kolb et al., 1993), even though the regulation of carbon metabolism is markedly different in C. glutamicum. Many substrates are co-metabolized by this organism, and monophasic growth is observed with only a few exceptions. The mixtures glucose/glutamate, glucose/ethanol and acetate/ethanol induce diauxic growth with preferential utilization of glucose or acetate (Arndt & Eikmanns, 2007; Arndt et al., 2008; Kotrbova-Kozak et al., 2007; Krämer et al., 1990; Kromemeyer et al., 1995). Interestingly, the enrichment...
factors for three DNA fragments from the upstream regions of the genes for the acetaldehyde dehydrogenase (ald), the alcohol dehydrogenase (adhA) and the phosphotransacetylase–acetate kinase (pta–ack) operon, all needed for ethanol degradation, were, at 15.6, 11.9 and 18.3, among the highest observed enrichment factors. Moreover, GlxR binding was even detected during growth on acetate in the upstream regions of ald and pta–ack (enrichment factors: 3.4 and 2.7). This could be a clue to explain diauxic growth by a particularly high affinity of GlxR for these sites, leading to (nearly) complete repression in the presence of glucose.

Compared with previous studies including ChIP-chip (Bussmann et al., 2009; Jungwirth et al., 2008; Kim et al., 2004; Kohl & Tauch, 2009; Kohl et al., 2008; Letek et al., 2006; Nishimura et al., 2011; Panhorst et al., 2011; Park et al., 2010; Toyoda et al., 2011), a total of 21 new binding sites and 46 corresponding target genes, including genes in predicted operons (Pauling et al., 2012), were identified with ChIP-seq. Under the chosen growth conditions, GlxR interacted with 40% (31 out of 77) of all DNA motifs from previous EMSA studies with 40 bp oligomers (Jungwirth et al., 2008; Kohl & Tauch, 2009; Kohl et al., 2008), and with one motif validated by EMSA with PCR products (Toyoda et al., 2011). Compared with the recent ChIP-chip study on GlxR binding showed an overlap of 57 fragments identified in both studies. The higher resolution of the ChIP-seq study allowed for the validated identification of intragenic sites for the first time, as the existence of such sites was only suggested by the microarray results. The 28 fragments without a predicted binding motif can be found among the 50 sites exclusively identified by ChIP-seq, pointing out the necessity of combining different approaches and techniques to fully elucidate the GlxR regulon.

The ChIP-seq results provide evidence of a non-protein-coding gene controlled by GlxR. The binding site in the shared upstream region of the gene for the small RNA 6C and the divergently transcribed gene cg0360 was validated by EMSA. With its suggested function in sporulation, dormancy or metabolically inactive cell states (Swiercz et al., 2008), the small RNA fits the profile of a GlxR target well. Further experiments will show the relevance of GlxR binding for the regulation of the 6C RNA. An example of a Crp regulator controlling the expression of a small RNA, cyaR, can be found in E. coli (De Lay & Gottesman, 2009).

We also describe GlxR as a non-canonical regulator, binding in intragenic regions. Beyond the regulon identified so far, there might be new mechanisms of action and targets to be detected outside the search space of promoter regions. For E. coli CRP, a role in chromosome structuring has been discussed based on ChIP-chip results, which revealed the presence of a large number of binding sites with low affinity (Grainger et al., 2005). Other ChIP-chip and ChIP-seq datasets from eukaryotic cells also suggest a role for genome-wide transcription factors binding beyond direct target regulation (MacQuarrie et al., 2011). In addition, at least for some of the GlxR binding regions lacking an evident correlation with a promoter region, new genetic features might be discovered in the future, for example by high-resolution RNA sequencing.

Acknowledgements

B.J. acknowledges the receipt of a grant from the Studienstiftung des Deutschen Volkes. This work was supported by SystemsX.ch and the Swiss National Science Foundation under grant number 31003A-125061.

REFERENCES


Edited by: D. Grainger