DESCRIPTION DATA FILE

1. GENERAL INFORMATION.

Title: *CRP Data Set*

Reference:

Shimada T, Fujita N, Yamamoto K, Ishihama A. (2011). Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS One*. 6(6):e20081. PMID: 21673794

Contact person for this data set:

Questions concerning the content of the data set that are raised by users of RegulonDB will be forwarded to this person. We would appreciate receiving a copy of the response to the user, so we can keep track of taking care of user requests.

Person: RegulonDB staff

Email address: regulondb@ccg.unam.mx

2. DATA SET DESCRIPTION.

Summary:

CRP Data Set.

CRP (cAMP receptor protein), the global regulator of genes for carbon sources utilized in the absence of glucose, is the best-studied prokaryotic transcription factor. Genomic SELEX screening was performed to identify the whole set of CRP regulation targets. A total of 275 CRP-binding peaks were identified via the SELEX methodology, of which 92 are known and 183 were identified as new targets of CRP (265 new regulatory interactions). (It is important to note that some of the CRP targets carry multiple CRP box sequences.) In addition, a computational analysis was carried out by using a collection of 500-bp sequences centered on each peak and the program BioProspector (http://ai.stanford.edu/_xsliu/Bio-Prospector/) to identify consensus recognition sequences of CRP. This program searched for the CRP box sequence by using the whole set of 323 CRP-binding sequences from a total of 275 targets (183 novel plus 92 known targets) identified in this study. A 16-bp CRP box motif sequence (5'-TGTGA-N6-TCACA-3') for all 323 CRP-binding sites was identified.

It has been concluded that CRP plays a key regulatory role in the whole process for the selective transport of carbon sources. It has been estimated that the total number of operons under the direct control of cAMP-CRP ranges from at least 378 to at most 500.

Here we show a list with 265 new regulatory interactions associated with the CRP transcriptional dual regulator based on genomic Selex (gSELEX) and consensus sequence evidence.

Experiment:

The following information is essential to facilitate curation, availability, and classification of your data set within RegulonDB:

- Version of *E. coli* ´s genome used in the experiment: *Escherichia coli* K-12 W3110
- Sequence identifier: Purified His-tagged CRP was mixed with a collection of *E. coli* genome fragments of 200–300 bp in length, and CRP-bound DNA fragments were affinity purified for the identification of CRP recognition sequences.
- Experimental conditions: Cells were grown in LB medium at 37ºC under aeration with constant shaking at 140 rpm. Cell growth was monitored by measuring the turbidity at 600 nm.
- Number of technical replicates: A cutoff level of 4.0 was used for the SELEX-chip pattern.

Methods:

SELEX method

"The genomic SELEX method was carried out as previously described [28]. A mixture of DNA fragments of the *E. coli* K-12 W3110 genome was prepared after sonication of purified genome DNA and cloned into a multicopy plasmid, pBR322. In each SELEX screening, the DNA mixture was regenerated by PCR. For SELEX screening, 5 pmol of the mixture of DNA fragments and 10 pmol of each His-tagged CRP were mixed in a binding buffer (10 mM Tris-HCl, pH 7.8 at 4°C, 3 mM magnesium acetate, 150 mM NaCl, and 1.25 mg/ml bovine serum albumin) and incubated for 30 min at 37°C. For the effector (+) reaction of CRP, 10 insert.symbol mmM adenosine 29,39-cyclic monophos phate (cAMP) (Sigma) was added during the binding reaction and also during the subsequent isolation procedure for CRP-DNA complexes. The DNA transcription factor mixture was applied to a Ni-NTA column, and after washing out unbound DNA with the binding buffer containing 10 mM imidazole, DNA-protein complexes were eluted with an elution buffer containing

200 mM imidazole. DNA fragments recovered from the complexes were PCR amplified."

SELEX-clos

"For SELEX-clos (cloning-sequencing) analysis, PCR products were cloned into the pT7 Blue-T vector (Novagen) and transformed into *E. coli* DH5a. Sequencing of each clone was carried out using the T7 primer (5'-TAATACGACTCACTATAGGG-3')."

SELEX-chip

"For SELEX-chip (DNA tiling array chip) analysis, PCR-amplified products of the isolated DNA-protein complexes obtained in the presence or absence of the effector were labeled with Cy5 and mixed with the original mixture of genome DNA fragments labeled with Cy3. The fluorescent-labeled DNA mixtures were hybridized to a DNA microarray consisting of 43,450 species of a 60-bp-long DNA probe, which were designed to cover the entire *E. coli* genome at 105-bp intervals (Oxford Gene Technology, Oxford, UK). The Cy5 fluorescence intensity of the test sample with each probe was normalized to that of the corresponding peak of Cy3 fluorescence of the original library. The Cy5/Cy3 ratio was measured and plotted along the *E. coli* genome."