

Information on Transcriptional Regulation and Signal Transduction of Escherichia coli K-12 Integrated in the Database RegulonDB.

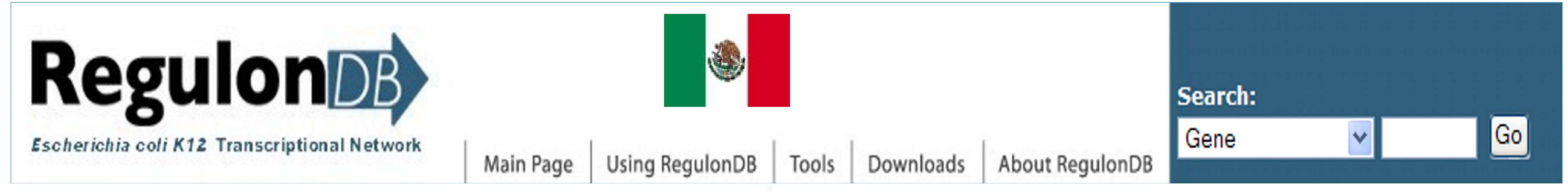
<u>Gama-Castro S.</u>, Santos-Zavaleta A., Peralta-Gil M., Weiss V., Salgado H., Solano H. and Collado-Vides J.

Programa de Genómica Computacional, Centro de Ciencias Genómicas, Universidad

Nacional Autónoma de México, A.P. 565-A, Cuernavaca, Morelos 62100, México.

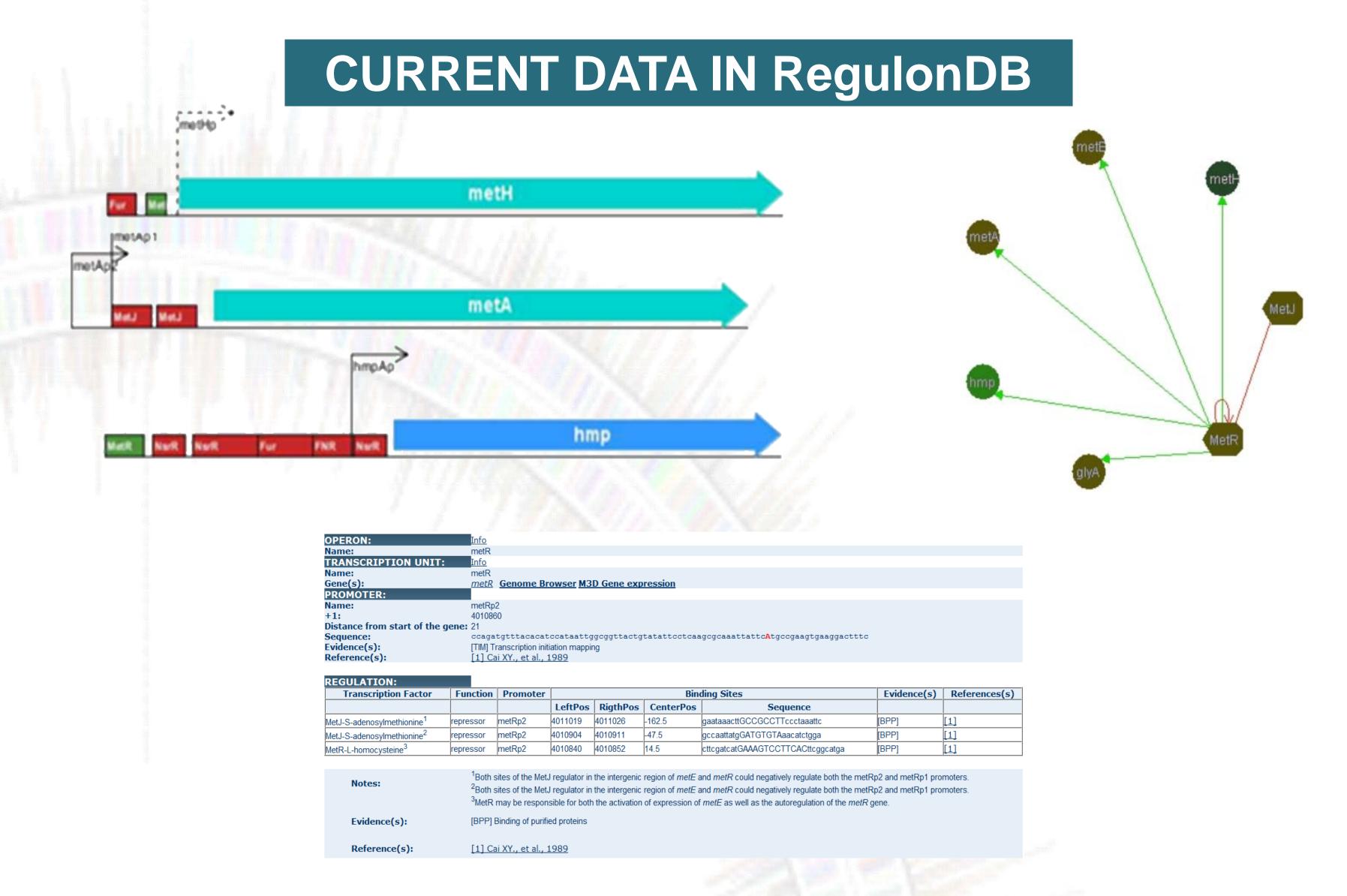


Centro de Ciencias Genómicas



ABSTRACT

Since its inception, RegulonDB (http://regulondb.ccg.unam.mx/) has been a database that compiles information about the regulation of transcription initiation of Escherichia coli K-12. However, we are aware that transcriptional regulation is not an isolated process; instead, it is the response to the different environmental conditions that trigger a series of concatenated reactions that end in transcriptional regulation, and it implies an adequate response in terms of induced and repressed gene products. We are working now to include all these new data in RegulonDB. As a consequence, transcriptional regulation in RegulonDB will be part of a unit that initiates with the signal, continues with the signal transduction to the core of regulation to modify expression of the affected set of target genes, and ends with an adequate response. We refer to these units as genetic sensory response units, or geSorgans.



The inclusion of geSorgans will bring a dramatic change and expansion of RegulonDB, due to the fact that we will be adding several new types of reactions and interactions. We started to collect data about signal transduction of the sigma factors, the two-component systems, of some transcription factors involved in carbon source utilization, and of genes involved in the synthesis of amino acids. We plan a high-level curation with super-pathways summarizing concatenated sets of reactions linked to those other databases that curate such information, while enabling with RegulonDB a compilation of complete geSorgans.

In addition, the number of DNA binding sites for some transcription factors has grown considerably, and therefore we decided to review systematically those sites whose lengths range from 40 to 60 bp with orientation and consensus sequences that are not easy to identify. The current version of RegulonDB is the beginning of a higher-level curation of gene regulation information, and eventually our database will include all regulatory mechanisms and their regulated genes.

NEW DATA: GENETIC SENSORY RESPONSE UNITS (geSorgans)

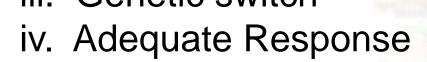


Integration of RegulonDB Data and other databases to build **Sensory Response Units**

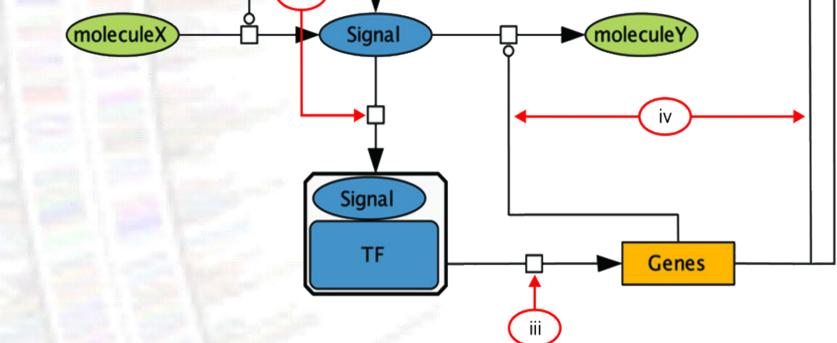


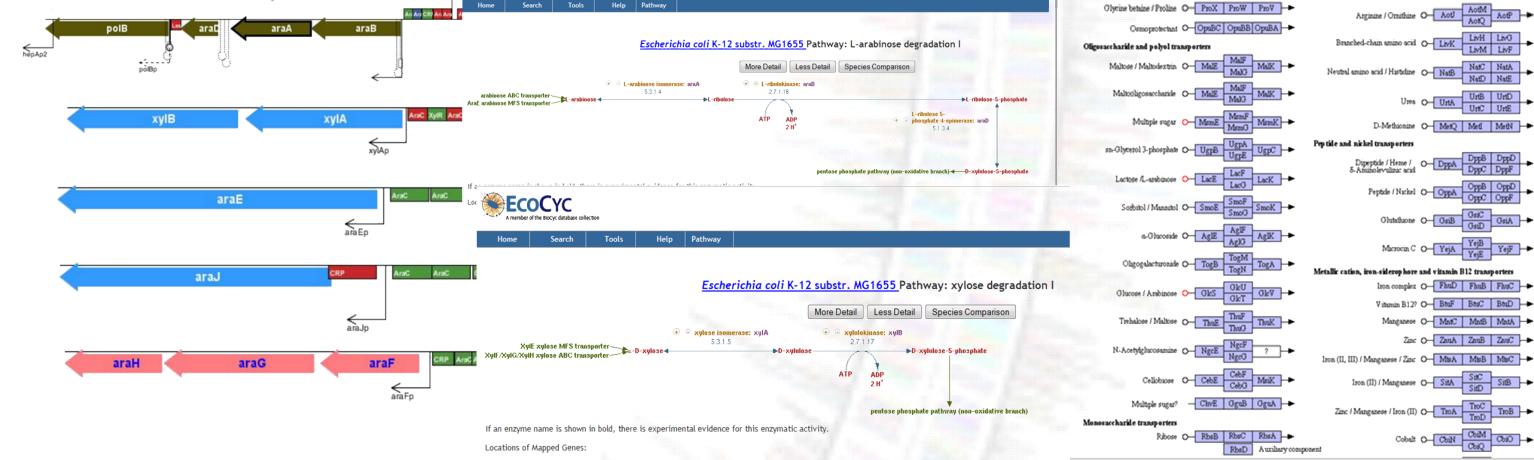




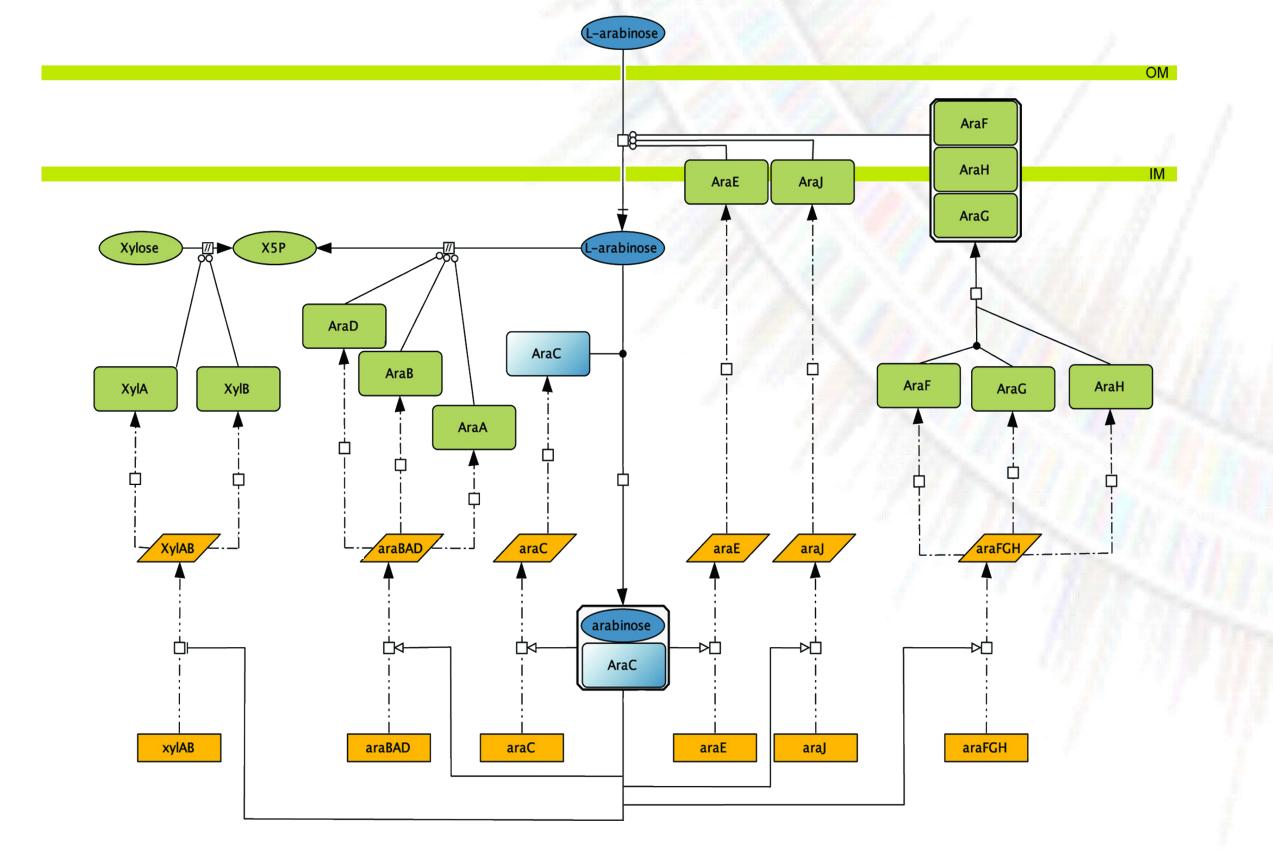








AraC-L-arabinose Sensory Response Unit

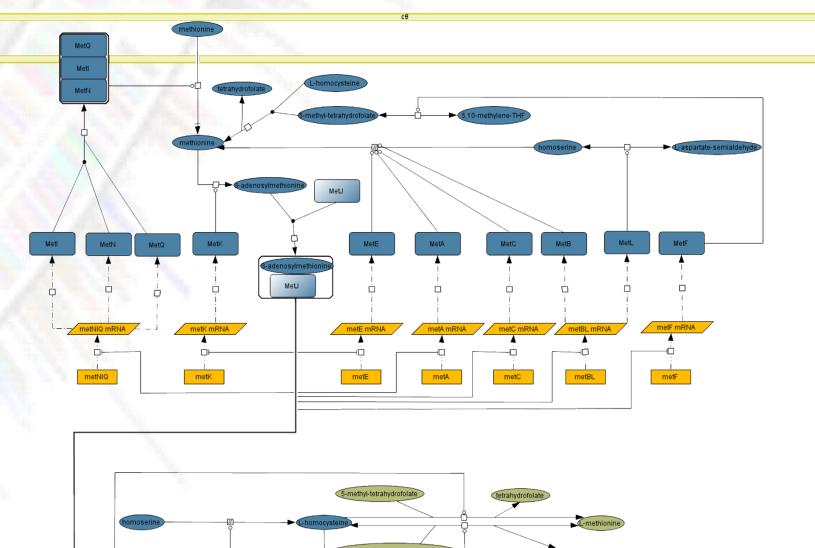


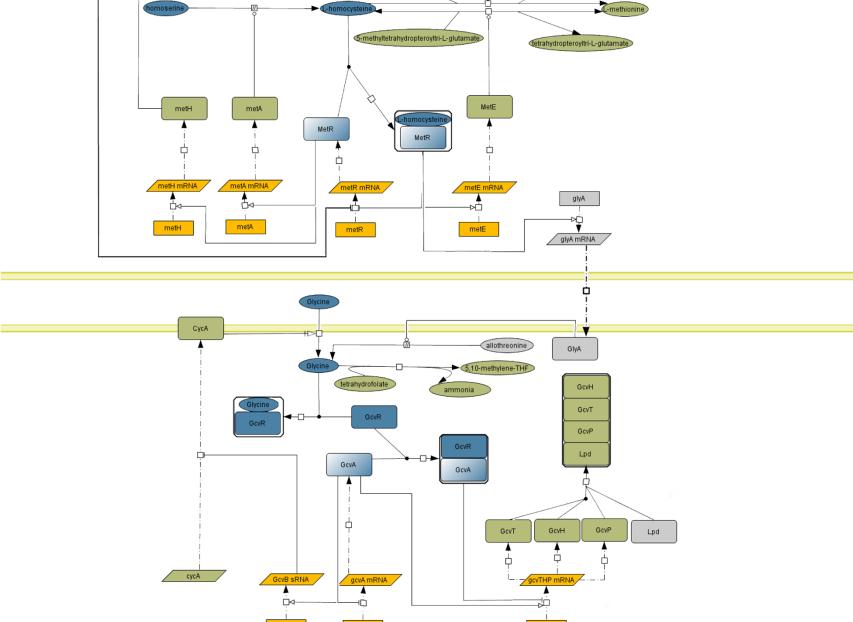
Reactions of a geSorgan

eaction # eaction name Reaction Descriptio ansport of L-arabinose Reaction type browth on L-arabinose and absence of ducose. In the absence of L-arabinose AraC acts as a repressor of the araBAD and araC operor Growth condition Cyc: TRANS-RXN-10 ABC-2-RX eaction Descriptio lex formation of AraC-arabinos Growth condition under which the reaction is active sence of alucos Cvc: RXN0-29. nRNA araBAD -> AraD anslation of the operon mRNA araBA eaction Descrip NA araBAD -> Aral eaction Descript ranslation of the operon mRNA araBAD Reaction # Reaction name nRNA araBAD -> AraA eaction Description ranslation of the operon mRNA araBAD Reaction Description arabinose degradatior Reaction type Reaction Descriptio Activation of the transcription of the operon araC by AraC-arabir eaction type slation of the operon mRNA a

ACTIVE CURATION

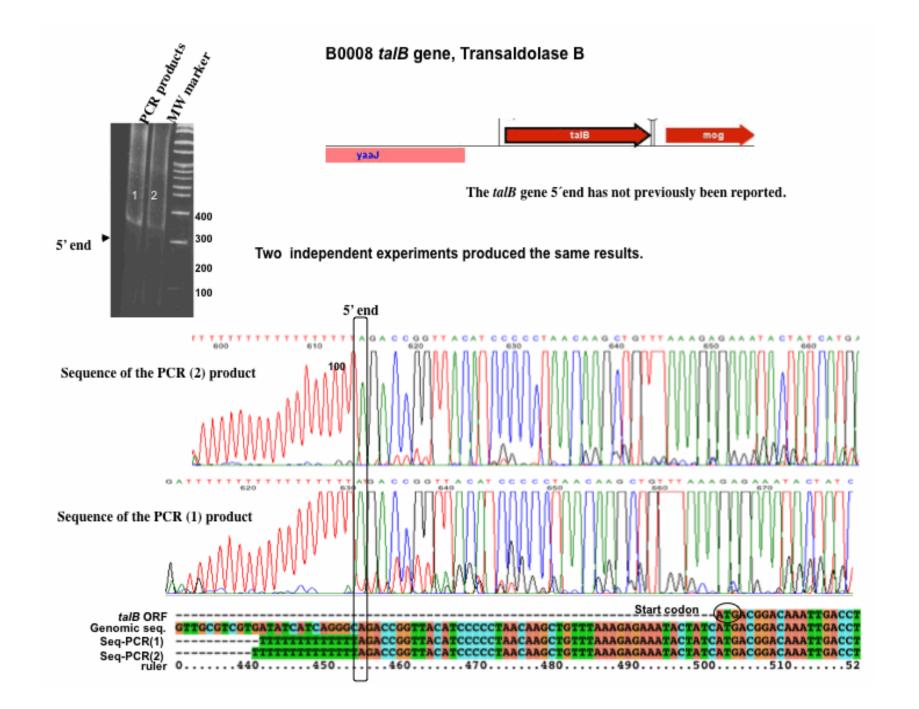
geSorgan Connections



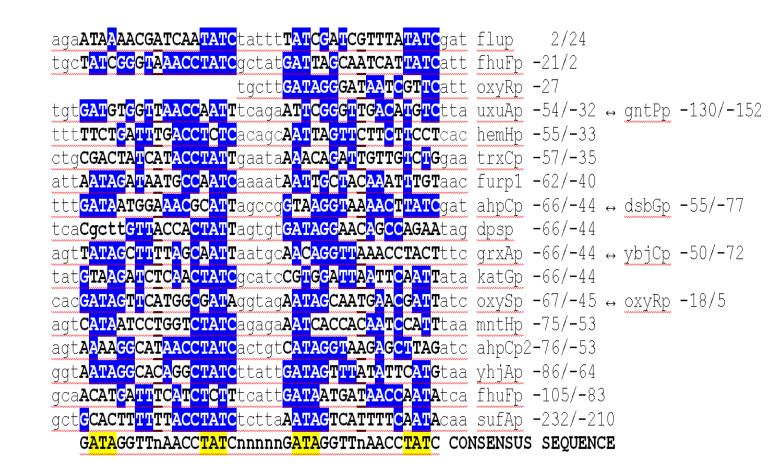


Summary: In the presence of L-arabinose, AraC activates transcription of genes that code for proteins necessary for utilization and transport of L-arabinose.

High-Throughput Mapping of Transcription start sites



Redefining Transcription factor binding sites



Alignment of the binding sequences of OxyR. Capital letters represent the two conserved inverted repeat motifs of 17 base pairs and the conserved motifs in blue. The motifs identified in the literature highlighted in yellow .

Funding

This work is supported by the National Institutes of Health [GM071962, GM077678]