72–74 Nucleic Acids Research, 2001, Vol. 29, No. 1

RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12

Heladia Salgado, Alberto Santos-Zavaleta, Socorro Gama-Castro, Dulce Millán-Zárate, Edgar Díaz-Peredo, Fabiola Sánchez-Solano, Ernesto Pérez-Rueda, César Bonavides-Martínez and Julio Collado-Vides*

Centro de Investigación sobre Fijación de Nitrógeno. UNAM A.P. 565-A Cuernavaca, Morelos 62100, México

Received October 2, 2000; Revised and Accepted October 11, 2000

ABSTRACT

RegulonDB is a database on mechanisms of transcription regulation and operon organization in *Escherichia coli* K-12. The current version has considerably increased numbers of regulatory elements such as promoters, binding sites and terminators. The complete repertoire of known and predicted DNA-binding transcriptional regulators can be considered to be included in this version. The database now distinguishes different allosteric conformations of regulatory proteins indicating the one active in binding and regulating the different promoters. A new set of operon predictions has been incorporated. The relational design has been modified accordingly. Furthermore, a major improvement is a graphic display enabling browsing of the database with a Java-based graphic user interface with three zoom-levels connected to properties of each chromosomal element. The purpose of these modifications is to make RegulonDB a useful tool and control set for transcriptome experiments. RegulonDB can be accessed on the web at the URL: http://www.cifn.unam.mx/Computational_Biology/regulondb/

INTRODUCTION

RegulonDB is a relational database containing information on mechanisms at the level of transcriptional initiation as well as operon organization with their terminator signals in the *Escherichia coli* K12 chromosome. The database is updated constantly by searching in original publications, and it is complemented by computational predictions. All this information is mapped in the genome sequence. This accumulated knowledge includes 2996 MEDLINE and PubMed links to the literature, plus 2451 links to GenBank. Computational predictions and their methods have been published previously (1,2). Previous publications in this yearly issue explain the initial relational design and subsequent modifications (3–5). The relational design keeps a constant evolution given the enriched expansion of the type of information that is added.

This paper describes the major updates and changes during the previous year to the database. There has been an increase in the number of all objects, such as promoters, binding sites and operons. What we consider the complete repertoire of 314 DNA-binding transcriptional regulators, both known and predicted, has been incorporated (6). The set of predicted operons has been updated (7). Furthermore, we have modified the database to incorporate different conformations of regulatory proteins that depend on their allosteric or covalent modifications, and that enable addition of not only effector metabolites involved in the allosteric interactions, but also the changes in the conditions that provoke the regulated changes in the cell. Furthermore, an innovation in this new version is a set of Java-based graphic user interfaces starting from the whole genome, to a region encompassing on average 18 genes, down to selecting individual operons with their transcription units, and their associated regulatory elements. All these graphic descriptions enable navigation to the text describing the properties of the different chromosomal elements. Table 1 lists links to Supplementary Material illustrating several aspects mentioned here.

OVERVIEW OF THE CURRENT DATA

As in previous versions, the database contains both information gathered from the literature and with the associated MEDLINE, PubMed and/or GenBank link to the original literature. Important additions include a total of 314 transcriptional DNA-binding regulators out of which 165 have experimental evidence and the rest have been predicted based on their helix–turn–helix DNA binding motif (6); we have also changed the complete set of predicted operons, to 578 predicted operons, resulting from a well-defined method based on distances in-between genes and their functional classes when available (7). Relevant information about binding sites and promoters published in a specialized *Escherichia coli* book on regulation (8) were all obtained and included in the database, involving additional searches to complement their relative position or regulated promoters in many cases.

Moreover, we have expanded the design of the database to take into account the different conformations available for regulatory proteins based on their interactions with small metabolites (usually allosteric interactions) or covalent

*To whom correspondence should be addressed. Tel: +527 313 2063; Fax: +527 317 5581; Email: ecoli-t1@cifn.unam.mx
Present address:
Ernesto Pérez-Rueda, Université Libre de Bruxelles av F.D. Roosevelt 50-CP 160/16 B-1050 Bruxelles, Belgium

Table 1. Links to Supplementary Material

Main page of RegulonDB (3.2)	http://www.cifn.unam.mx/Computational_Biology/regulondb
Graphic user interface	http://www.cifn.unam.mx/Computational_Biology/regulondb_graph
Summary table	http://www.cifn.unam.mx/Computational_Biology/regulondb/docs/summary.html
Conditions and effector metabolites	http://www.cifn.unam.mx/Computational_Biology/regulondb/docs/signal_condition.html

modifications (phosphorylation, methylation, etc). By far most regulatory proteins have two allosteric conformations, one capable of binding to DNA and affecting transcription and the other unbinding from the DNA. There are, however, few cases of proteins interacting with more than one metabolite and affecting transcription differentially, such as the case of TyrR interacting with phenyalanine, tyrosine and tryptophan (9–11). The design we implemented is not limited to two conformations for each protein. For each regulatory interaction, however, only one conformation is the 'active' one, that is the one that binds to DNA and has a defined positive or negative effect on a given promoter. This enables encoding in the database cases, for example, of a protein that binds in two conformations and activates and represses a single promoter. For instance, AraC binds to DNA and activates in the presence of arabinose the *araBAD* operon (12)—this is one regulatory interaction—whereas in the absence of arabinose it binds to a similar site but now repressing the same promoter, and therefore supporting a different regulatory interaction as defined in the database. Similar cases are those of Ada binding unmethylated (repressing at high concentration) as well as methylated (activating) *ada* promoter, whereas it only activates the *alkA* promoter (13). Note that what we call 'active' conformation has nothing to do with activators or repressors but with the configuration that is able to bind DNA and have a defined regulatory effect.

To date, RegulonDB contains information that is knowledge supported on the molecular biology of the different systems. We consider that it is useful, especially for analyzing transcriptome experiments, to expand the database in order to include knowledge from genetic studies in the absence of details of the molecular machinery. Thus, knowing for instance, that in the presence of lysine *E.coli* represses *lysP* gene (in the absence of a detailed mechanism) is potentially useful information (14) in the analysis and interpretation of transcriptome experiments. An important effort has been in gathering information on the effector molecules that interact with these proteins, and by these means modify the activation or repression of different genes and operons.

COMPUTATIONAL INFRASTRUCTURE AND GRAPHIC USER INTERFACE

We have, as mentioned above, modified the database to include the expansions described previously. These are basically the addition of a table for conformation previously of regulatory proteins, and the encoding of conditions and their link to a regulatory protein, an effector metabolite, and the effect (inducing or repressing) on the regulated genes. We can then include associations between conditions, i.e. glucose as carbon source, and its effect diminishing cAMP and by these means affecting regulation of CRP-controlled promoters. In less characterized genes there may only be an association between, for instance, high osmolarity in the medium and the repression of *mdoGH* genes (15).

We have spent considerable effort in improving the graphic capabilities of RegulonDB. A Java applet display (Java™ Development Kit JDK™ 1.1.3 Product of Sun Microsystem, Inc.) starts with a circular representation of the whole *E.coli* K-12 genome. In this graph forward and reverse genes are easily identified. This circular representation is clickable and a new window is generated displaying all known and predicted objects within a range of 18 kb centered at the selected position, and with arrows at each extreme that enable 'walking' through the chromosome in either direction. Genes and their clustering into operons are displayed, as well as regulatory elements. We have colored the genes based on their functional classification from Riley (16). An additional window is displayed showing the color code.

Further navigation enables the user to get the detailed information about a particular gene, or to focus on a particular operon. The window on the individual operon displays the operon organization, its regulatory elements and the different transcription units associated to it. The different elements (operons, genes, regulatory elements) are clickable in order to obtain text information about their individual properties. Briefly, this new version of RegulonDB has an increased amount of information and a much more attractive graphic browsing interface. A new server supporting the database will speed its access on the web.

AVAILABILITY

RegulonDB 3.2 can be accessed through the URL: http://www.cifn.unam.mx/Computational_Biology/regulondb/. We kindly ask users of RegulonDB to cite this article.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by grant number 0028 from CONACYT-Mexico, grant R01 RR07861-10 from NIH and DE-FG02-98ER62558 from the US Department of Energy.

REFERENCES

1. Blattner,F.R., Plunkett,G.,III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G. *et al.* (1997) The complete genome sequence of Escherichia coli K-12. *Science*, **277**, 1453–1462.

2. Thieffry,D., Salgado,H., Huerta,A.M. and Collado-Vides,J. (1998) Prediction of transcriptional regulatory sites in the complete genome sequence of Escherichia coli K-12. *Bioinformatics*, **14**, 391–400.
3. Huerta,A.M., Salgado,H., Thieffry,D. and Collado-Vides,J. (1998) RegulonDB: a database on transcriptional regulation in Escherichia coli. *Nucleic Acids Res.*, **26**, 55–59.
4. Salgado,H., Santos,A., Garza-Ramos,U., van Helden,J., Diaz,E. and Collado-Vides,J. (1999) RegulonDB (version 2.0): a database on transcriptional regulation in Escherichia coli. *Nucleic Acids Res.*, **27**, 59–60.
5. Salgado,H., Santos-Zavaleta,A., Gama-Castro,S., Millan-Zarate,D., Blattner,F.R. and Collado-Vides,J. (2000) RegulonDB (version 3.0): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.*, **28**, 65–67.
6. Pérez-Rueda,E. and Collado-Vides,J. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.*, **28**, 1838–1847.
7. Salgado,H., Moreno-Hagelsieb,G., Smith,T.F. and Collado-Vides,J. (2000) Operons in Escherichia coli: genomic analyses and predictions. *Proc. Natl Acad. Sci. USA*, **97**, 6652–6657.
8. Lin,E.C.C. and Lynch,A.S. (eds) (1996) *Regulation of Gene Expression in Escherichia coli*. Chapman and Hall, New York.
9. Wilson,T.J., Argaet,V.P., Howlett,G.J. and Davidson,B.E. (1995) Evidence for two aromatic amino acid-binding sites, one ATP-dependent and the other ATP-independent, in the Escherichia coli regulatory protein TyrR. *Mol. Microbiol.*, **17**, 483–492.
10. Lawley,B., Fujita,N., Ishihama,A. and Pittard,A.J. (1995) The TyrR protein of Escherichia coli is a class I transcription activator. *J. Bacteriol.*, **177**, 238–241.
11. Lawley,B. and Pittard,A.J. (1994) Regulation of aroL expression by TyrR protein and Trp repressor in Escherichia coli K-12. *J. Bacteriol.*, **176**, 6921–6930.
12. Lobell,R.B. and Schleif,R.F. (1990) DNA looping and unlooping by AraC protein. *Science*, **250**, 528–532.
13. Saget,B.M. and Walker,G.C. (1994) The Ada protein acts as both a positive and a negative modulator of Escherichia coli's response to methylating agents. *Proc. Natl Acad. Sci. USA*, **91**, 9730–9734.
14. Neely,M.N. and Olson,E.R. (1996) Kinetics of expression of the Escherichia coli cad operon as a function of pH and lysine. *J. Bacteriol.*, **178**, 5522–5528.
15. Lacroix,J.M., Loubens,I., Tempete,M., Menichi,B. and Bohin,J.P. (1991) The mdoA locus of Escherichia coli consists of an operon under osmotic control. *Mol. Microbiol.*, **5**, 1745–1753.
16. Riley,M. (1993) Functions of the gene products of Escherichia coli. *Microbiol. Rev.*, **57**, 862–952.