Web-Based Visualisation of the Transcriptional Control Network of *Escherichia Coli*

Nestor Sosa[a], Andreas Kremling[b], Esther Ratsch[a] and Isabel Rojas[a,*]

[a]*EML Research gGmbH, Schloß-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany*
[b]*Max-Planck-Institute for Complex Technical Systems, Magdeburg, Germany*

ABSTRACT: Transcription is one of the basic processes of gene expression, controlled by a complex network of biochemical reactions. Despite its importance, most work on the visualisation of biochemical networks focuses on the representation of metabolic pathways. The visualisation of the complex networks controlling transcription requires the implementation of a hierarchical approach that allows the display of the structure of each regulatory region with its transcription factors and regulated operons. This paper presents a web-based application for the visualisation of transcriptional control networks. It takes as case study the organism *Escherichia coli*. The definition of the visual components implemented is mainly based on those proposed by Shen-Orr *et al.*, 2002, slightly extended to visualise complex networks.

KEYWORDS: Transcription networks, visualisation

INTRODUCTION

Cells can adapt very quickly to changes of environmental conditions. This is mainly based on signal transduction units (see Fig. 1) which sense the environmental conditions with sensor elements (often membrane bound proteins). These sensors then transform the external stimulus into an intracellular signal by activating transmitter elements, e.g. a kinase which is able to autophosphorylate. Via direct or indirect interaction with a number of other elements a receiver element is activated, which enables a transcription factor (transcriptor or regulator) to bind to special DNA binding sites (control sequences). The binding of transcription factors increases or decreases the activity of the RNA polymerase and therefore determines the transcription efficiency. In this way, cells can respond to environmental changes by changing the expression of genes and thus the types and amounts of proteins synthesized.

Since more than one regulator protein is able to bind to one control sequence, a complex pattern of interacting regulator proteins emerge, which is further complicated by the fact that the regulated genes may code for elements of the signal transduction unit, like sensors or transcription factors.

Despite of the importance of transcription networks for the cell's functionality, visualisation of bio-chemical networks is mainly restricted to the presentation of metabolic pathways. In the case of *E. coli*,
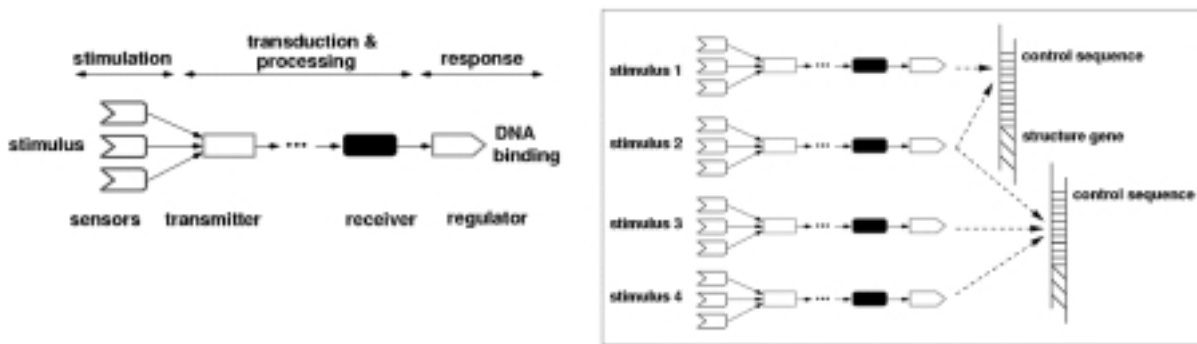

Fig. 1. Schematic description of transcriptional control networks.

for example, there are various databases available which provide data on many aspects of the organisms. Amongst them we can find, for example: EcoCyc [Karp *et al.*, 2002], KEGG [Kanehisa and Goto, 2000] which although it mainly contains information about metabolic recations it also provides some information about signal transduction, and RegulonDB [Salgado *et al.*, 2001] with data on control of gene expression. EcoCyc offers the visualisation of all regulator proteins for a given gene as a graph. Moreover, searching for a regulator protein, results in a list of graphs of all genes which are under control of the regulator protein. However, there are no specific visualisation techniques or layouts for transcriptional networks.

Shen-Orr *et al.*, 2002, have shown that transcriptional (control) networks can be decomposed in a very efficient way. The authors define 'network motifs' as patterns of interconnections between regulator proteins and targets. In the proposed framework the transcriptional network of *Escherichia coli* can be described with three main motifs: feed-forward loop, single input module, and dense overlapping regulon (DOR). From the available data, six DORs reflecting six regulation regions are defined for the biochemical network of *Escherichia coli* (anaerobic/aerobic metabolism, carbon utilisation, osmotic stress, stationary phase, DNA metabolism, superoxide). Inspection of the relevant transcription factors revealed that 13 regulator proteins take part in more than one DOR; we therefore call them 'global' regulators. Transcription factors that are present in only one DOR are named 'local' regulators.

In this paper we present a web-based application that allows the interactive visualisation of transcriptional (control) networks. The definition of visual components is based on those proposed by Shen-Orr *et al.*, 2002, slightly extended to visualise complex networks. The system works with an underlying database containing the data related to the transcriptional networks in a structured and easy to query manner.

Within the frame of the development of a web-based application for the dynamic visualisation of transcriptional networks the following requirements for the software were defined:

– Display the hierarchical structure of each regulation region (DOR) with all transcription factors and regulated operons;
– Display the hierarchical regulation structure of each transcription factor;
– Display regulation regions regulated by global transcription factors;
– Display only operons regulated by selected transcription factors.

Following the given brief introduction to the problem and the proposed solution the paper will offer a more in-depth explanation of the approach taken, presenting the database model, on overview of the visualisation approach and algorithms, and the implementation of the application. The paper finalises

with some concluding remarks, highlighting the main advantages of the solution proposed and future directions and extensions of this work.

THE APPLICATION'S COMPONENTS

This section describes in more detail the components of the application developed and its implementation. Due to space restrictions only a high level version of the database model is presented, avoiding details about the cardinalities of the relationships and table attributes.

Database

In order to support and simplify the process of the interactive visualisation of transcriptional networks, a database to store the related data was designed and implemented. The main objective of having this database is to facilitate and accelerate the task of data extraction for the creation of the graphs to be displayed. The problem of the visualisation of transcriptional networks lies not only in the graph layout, but depends strongly in the detection of motifs and networks in the raw data. Using the database avoids having to recur to the raw data and their analysis each time the system requires drawing a graph. The database was populated with data from *Escherichia coli*, the case study organism for this work. The core of the model was built based on the above given description of the transcription regulation process, taking the concepts of *Operon*[1] and *Transcriptor* (transcription factor) as the main participants of the transcriptional networks. An operon codes for one or many proteins of a metabolic pathway and often also for their respective transcription factors. A transcription factor regulates one or more operons of various regulation areas (DORs). Several transcription factors together with the respective regulated operons build *Networks* which can be of varying complexity. With these concepts and relationships the general aspects of transcriptional networks can be described. However, considering the transcriptional network diagrams proposed by Shen-Orr *et al.*, 2002, it was necessary to include new concepts and relationships to represent the relationships between the different layers of regulation networks, namely: *Transregulation, Subnetwork, Supernetwork, Networkregulation* and *Autoregulation*. Figure 2 shows the definition of these new concepts based on the original diagrams proposed by Shen-Orr *et al.*, 2002. Transregulation represents a regulation process between transcription factors. A *Subnetwork* is a network of operons which are all regulated by transcription factors organised in another network. An arrangement of *Subnetworks* forms a *Supernetwork*. The *Networkregulation* relationship handles the regulations between networks. With all these new relationships and objects the task of visualisation can be simplified, by avoiding complex data manipulation and extraction, and even by avoiding complicate graph algorithms.

In addition to the above described concepts and relationships, the database should include data associating the regulating genes to their corresponding regulation regions (as proposed in Shen-Orr *et al.*, 2002). As a consequence, a new relationship *Memberofnetwork* was added. Operons and transcription factors are assigned to networks, and these networks are related to DOR's. Another relationship added was *Sourcelink*, that links each regulation to a hyperlink containing information about the regulation process. Figure 3 shows the resulting database model for the transcriptional networks, containing all relations needed to visualise the transcriptional regulation process. The table *Regulation* contains all

---

[1] Table and relationship names are identified by starting with a capital letter and by being in italics.
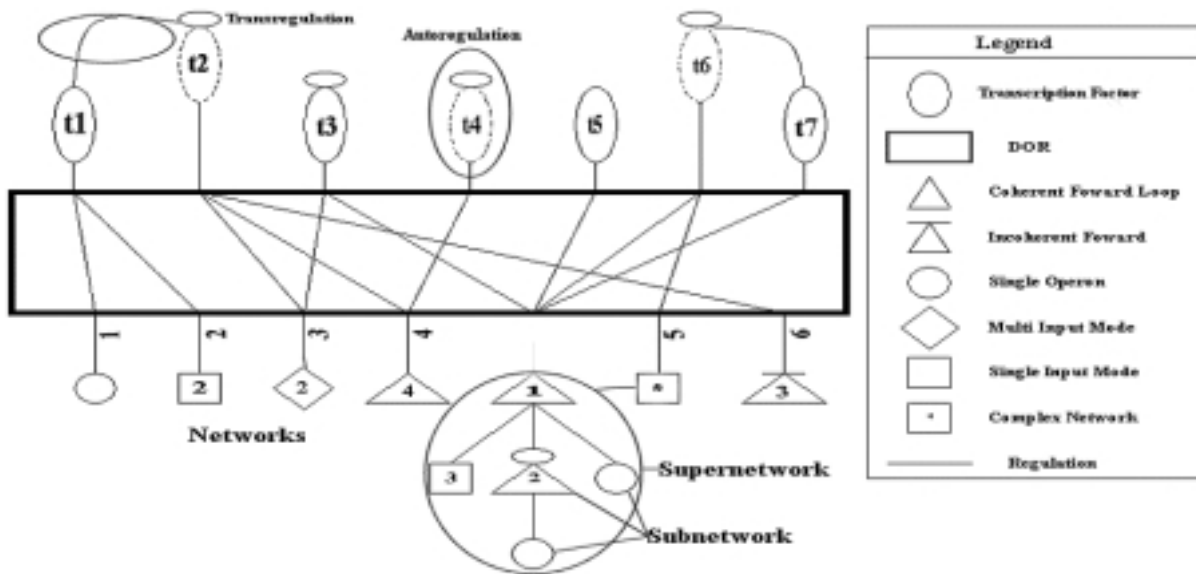
Fig. 2. Example of a DOR view showing the different patterns and relationships that can be visualised. A complex network (number 5) can contain a number of *Subnetworks* to a *Supernetwork*, although there is only one *Supernetwork* in this example.
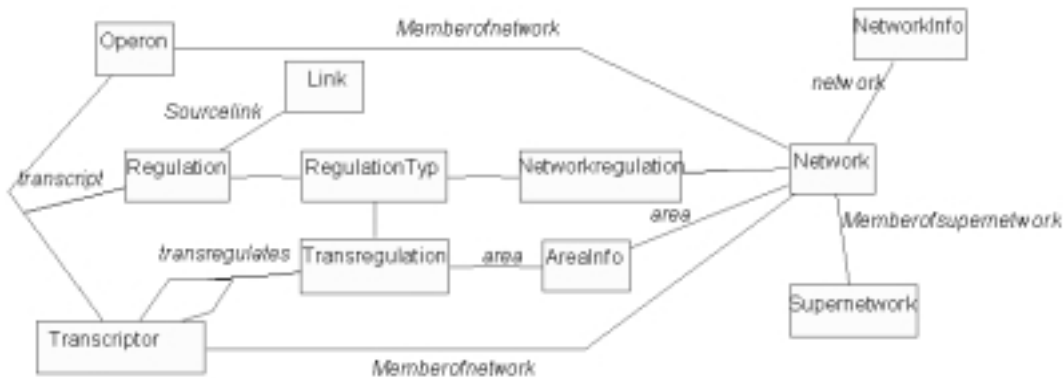

Fig. 3. Database schema for transcriptional networks.

regulation interactions between transcription factors and operons without considering the regulation networks in which they are organised. Additional tables (*Areainfo, Regulation Type*, and *Networkinfo*) were incorporated to store information about the valid types and areas of regulations, and of valid motifs.

Visualisation

The layout patterns implemented are based on the diagrams presented by Shen-Orr *et al.*, 2002. To display the various regulation cases, the visualisation process was broken down into three main views: global, DOR and subnetwork.

The global view (see Fig. 4) shows all or selected global transcription factors with their corresponding regulated DORs. By clicking on a transcription factor the regulations of other global transcription factors will be hidden, such that its regulated DORs can be easily identified. Pop up windows displaying related
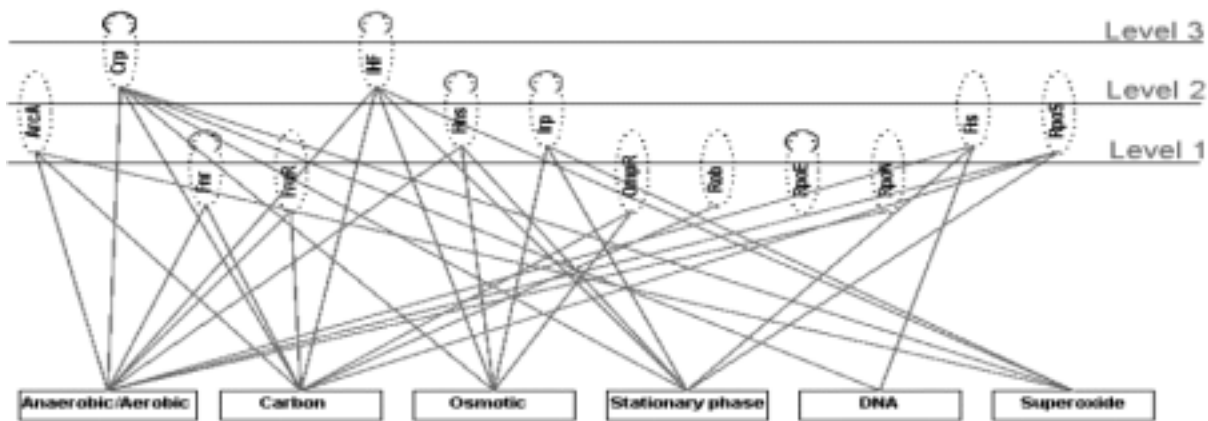
Fig. 4. Global view: Transcription factors are assigned to a level according to the number of DORs they control. The top level is represented by the regulators Crp and IHF which are involved in the control of almost all DORs.

information about the participating elements can be activated when locating the mouse over any element.

The transcription factors are positioned in the y coordinate according to the amount of DORs they regulate, so that three levels of regulation exist (1–2 DORs, 3–4 DORs and 5–6 DORs).

The DOR view (see Fig. 5) derives from the network view proposed in Shen-Orr *et al.*, 2002. It displays the regulations inside a DOR and one level of regulation below the DOR. For each global transcription factor its regulated DORs are queried from the database, displayed and the connection lines drawn. The DOR view is more complex than the global view because it contains all transcription factors of a regulation area with the operons they regulate. Complex networks, with multi level regulations, are represented by a new symbol (a green rectangle). These networks do not follow any of the patterns already discussed.

The layout process of the DOR view also requires more queries than the global view. First, all transcription factors of the area must be requested and their interactions by transregulation checked. If more than one transregulation level is found, the transcription factors are positioned in the y coordinate considering the level of transregulation. Moreover, in the case of one or more transregulation interactions the transcription factors must be reordered to later be properly drawn in such a way that the regulation chains are placed one next to the other. Second, all networks (including single operons) regulated by each transcription factor are requested and their operons determined. Finally, regulation relationships and their regulation types (positive or negative regulation) are requested and drawn as connection lines. For the DOR viewwe have developed an extended search tool which allows the query for operons which are all regulated by a selected set of transcription factors (see Fig. 6) .

The subnetwork view (see Fig. 7) also derives from the network view presented in Shen-Orr *et al.*, 2002. By clicking on a complex network a separate subnetwork view will be displayed. Regulations which are not performed through a DOR are represented by a green connection line. In this view the complex networks are visualised independently of their respective DORs. The view shows the complete regulation of a network as an interconnection between networks. The information about network patterns of the interconnected networks is directly obtained from the table *Networkregulation* in the database, avoiding having to analyse the raw data each time this information is required. Furthermore, for each of the networks in the sub-network view information about its regulating operons will be displayed when scrolling over the respective network.
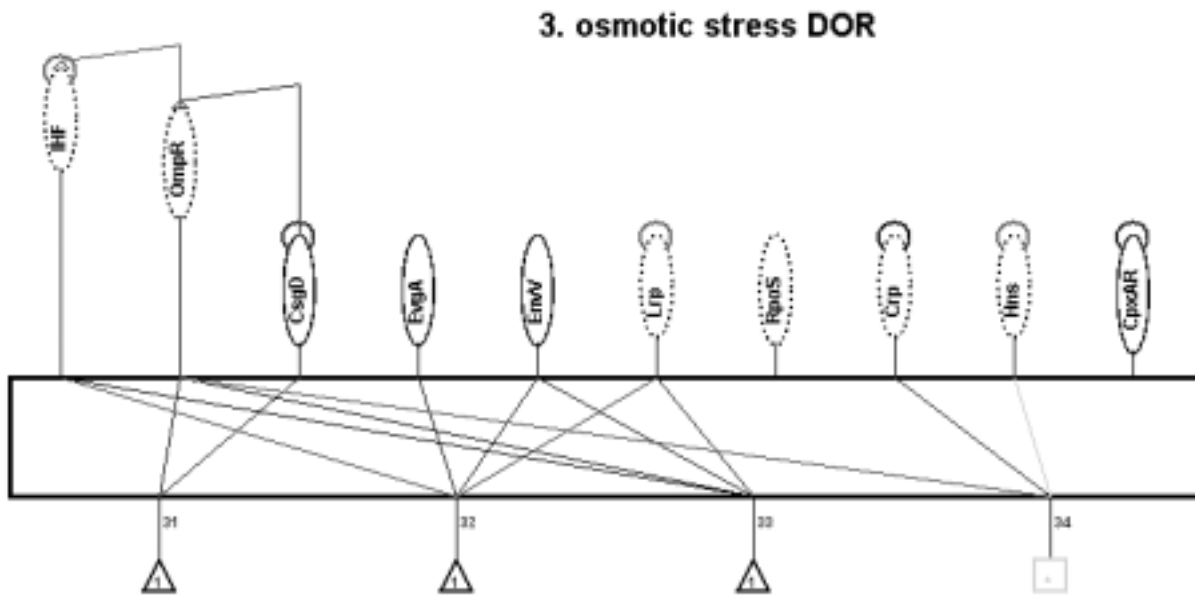
3. osmotic stress DOR

Fig. 5. DOR view of the osmotic osmotic stress, displaying the regulations inside a DOR and one level of regulation below the DOR.
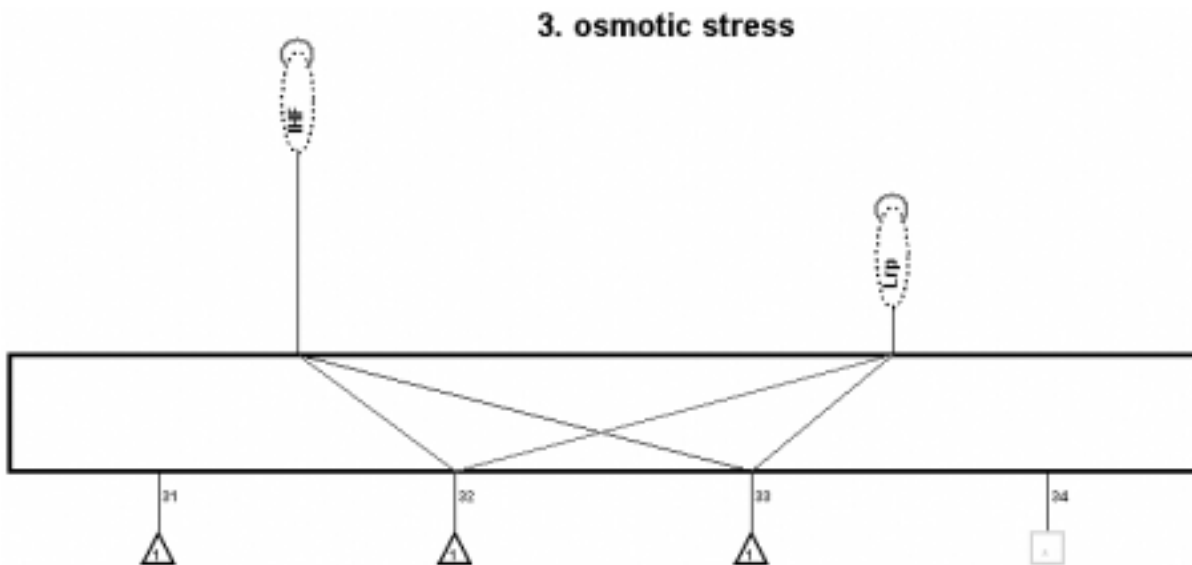
3. osmotic stress

Fig. 6. DOR view of the osmotic osmotic stress having applied an extended search to query for all operons regulated by a selected set of transcription factors, here for the transcription factors IHF and Lrp. Both transcription factors control operons number 32 and 33.

To easily display this case, we do not need to identify any network patterns or motifs using data mining algorithms that work on the regulation interaction. Instead the regulation between networks can be directly obtained from the database using the table *Networkregulation*. Furthermore, it is possible to later determine its regulating operons.
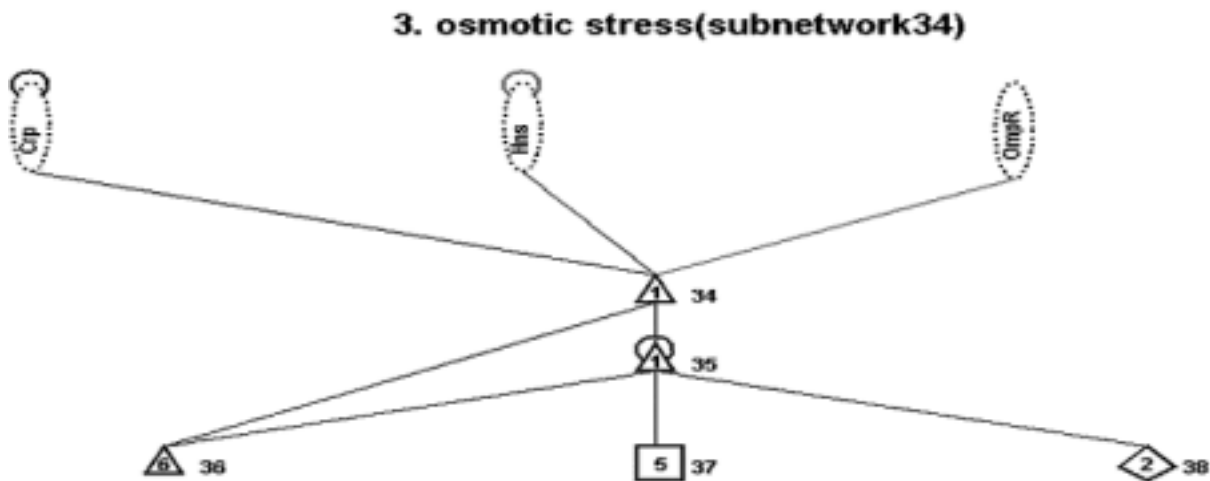
Fig. 7. *Subnetwork* view, showing the complete regulation of a network as an interconnection between networks.

IMPLEMENTATION

To develop the web-based visualisation tool several different web technologies were examined, like for example Java Applets, XML, XSLT, Java Servlets and Java Server Pages. Although Java Applets are very fast because the classes are precompiled and directly executed by the browser, this technology was not chosen for several reasons, including: browser incompatibility, lack of browser support, slow download and unpredictable behavior on different operating systems. In addition, the interaction between the database and the applets requires clients to have very high speed internet access, otherwise the internet connection can be overloaded and response time becomes too high. Another, very good and up to date technology is XML (extensible markup language) together with XSLT (extensible stylesheet language). The power of XML lies in its simplicity and its standard way to structure and package data. XSLT is a transformation language used to convert XML in other formats like HTML. Using this technology it is possible to create dynamic images using rasterised SVG (Scalable Vector Graphics, an XML grammar for stylable graphics). SVG permits the creation of detailed and zoomed-in views without regenerating a new image, while other formats as JPEG or GIF would need to be regenerated every time an image is zoomed or panned. Unfortunately, writing a general XSLT is very difficult (because of coordinate transformations in SVG), therefore a special plug in for the browser must be installed.

Based on the goal of being as general as possible with respect to the browser used, the solution adopted is based on Java Servlets together with Java Server Pages (JSP). Java has a good support for creating windows, graphics, colors, and interfacing with networks. Java Servlets run on the server and all the processing is carried out at the server side. JSP permits the embedment of Java code inside of HTML pages and allows an easy and fast development of interactive web sites. To run Java Servlets and JSP a Servlet engine or Server container is required. There are various implementations of such engines; the one chosen for this application was Tomcat of the Apache Jakarta Project.

The application is a 3-tier one, following the Model View Controller architecture (MCV [Douglass, 1997]). The controller is a Java Servlet intercepting all requests and delegating them to responsible routines. Depending on the request parameters sent by the client, the desired visualisation is chosen and executed. There are Java classes to access the database, others to generate the image and its mapping, and one that implements a connection pool used to speed up the database access through parallel querying.

Once the image file (a GIF file) is generated at the server side the controller forwards the data to the presentation layer, in this case a JSP, which displays the generated graph visualisation on a web page. The GIF file is mapped using a map file, which contains the coordinates of each region of an image. By clicking on one of these regions an URL link or action is activated which might generate another visualisation. Using this mapping technology together with the Cascading Style Sheets (CSS, a simple mechanism for adding style to Web documents) the visualisation becomes more dynamic. The information about each element does not always have to be requested from the database because in many cases it will already be embedded in the map file.

CONCLUSIONS AND FUTURE WORK

This paper describes the development of an application for the dynamic visualisation of transcriptional networks. The basic visualisation components used in the application are based on those proposed by Shen-Orr *et al*., 2002. These were extended with new visualisation elements to allow the display of other types of information and to facilitate the display of information already displayed in the visualisations proposed by Shen-Orr *et al*., 2002. As case study we took *Escherichia coli*. The visualisation application is supported by an embedded database containing the information related to the transcriptional networks in a structured and easy accessible manner. This facilitates and accelerates the process of visualisation of the information.

The main goal of this work was to visualise important aspects and structure of regulation in transcriptional networks, using different views: A global view, showing all global transcription factors and its regulated DORs; a DOR view, displaying the hierarchical structure of the regulation within a regulation area; and a subnetwork view, showing hierarchical motif inter-regulation structures.

The application was developed using Java Servlets and JSP, leaving the processing tasks at the server side. However, due to the use of image maps it is not necessary to load the required data from the database each time it is required.

With slight modifications to the existing database schema (mainly to include information about the organism related to the operons and transcriptors) and to the application, the system can be used for multiple (different) organisms. For example for Yeast, using the data of DORs detected and compiled by Segal *et al*., 2003. Such an extension would also allow the comparison of transcriptional networks of organisms, which would be an interesting research topic.

Access to the application is provided at: http://sabio.villa-bosch.de/motif/index.html.

ACKNOWLEDGEMENTS

REFERENCES

- Douglass, B. P. (1997). Real-time UML: Developing Efficient Objects for Embedded Systems. Addison-Wesley Object Technology Series, Addison-Wesley.
- Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. **28**, 27-30.
- Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Collado-Vides, J., Paley, S. M., Pellegrini-Toole, A., Bonavides, C. and Gama-Castro, S. (2002). The Ecocyc database. Nucleic Acids Res. **30**, 56-58.

- Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millán-Zárate, D., Díaz-Peredo, E., Sánchez-Solano, F., Pérez-Rueda, E., Bonavides-Martínez, D. and Collado-Vides, J. (2001). RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* k-12, Nucleic Acids Res. **29**, 72-74.
- Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat. Genet. **34**, 166-176.
- Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat. Genet. **31**, 64-68.