Cyclonet – a database on cell cycle regulation [?]

© Kolpakov F.A.[1,2], Deineko I.[3], Zhatchenko S.A.[1,2]and Kel A.E.[1,4]

[1]Biosoft.Ru/DevelopmentOnTheEdge.com, Novosibirsk, Russia
[2]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia
[3]Institute of Cytology and Genetics, Novosibirsk, Russia
[4]BIOBASE GmbH, Wolfenbuettel, Germany
fedor@biosoft.ru, ake@biobase.de

Abstract

Cyclonet is a database on cell cycle regulation in eukaryotes. It contains information about specific genes, proteins ant their complexes, models of cell cycle regulation and results of their analysis, microarray data, literature references and other related resources.
Using Cyclonet database we are demonstrating that BioUML and BeanExplorer Enterprise Edition technologies can be used for formal description, visualisation and simulation of complex biological systems. BioUML workbench is used to query and edit the database content as well as for visual modelling. BeanExplorer Enterprise Edition provides web interface for access to Cyclonet database.
Availability: http://cyclonet.biouml.org

1 Introduction

Success of the biologically reasonable modelling of cellular systems depends on the completeness of our knowledge and on integration of all fundamental molecular processes, such as signal transduction, regulation of gene expression and metabolism. Mammalian cell cycle is a good example of system where such natural integration of all the molecular processes plays very important role in their regulation. In the last few years numerous modelling approaches have been developed and applied in order to understand molecular mechanisms of cell cycle (for example, [15]). But still major control mechanisms of cell cycle progression and cell cycle exit remains unclear. Despite of the massive development of various biological databases, no specialized repository was created so far that would integrate all knowledge on cell cycle.

We are constructing the Cyclonet database oriented towards the most important problems related to cell cycle control. The first problem that we address is how the control of the checkpoints G1/S and G2/M can be lost thus resulting in uncontrolled rapid proliferation leading to tumour development. Nowadays it is widely accepted that certain interruptions of cell cycle regulatory mechanisms may initially cause tumour development. Another problem to deal with is the mechanisms of cell switch to specific ontogenetic programs (such as activation-induced proliferation, terminal differentiation and apoptosis). In Cyclonet we are collecting computer models that are able to suggest hypotheses about the very first steps on the road to disregulated proliferation as well as about molecular mechanisms of ontogenetic switches.

Gene expression data coming from microarray expression experiments on cell cycle studies are systematically collected in Cyclonet. This data serve us to built real life computer models of the dynamic of cell cycle and to estimate unknown parameters of the model. This is done by methods of reverse engineering when parameters and structure of the model is optimised to fit the experimentally observed dynamics of gene expression changes. Promoter analysis and prediction of targets sites for various transcription factors plays an important role in such modelling closing the gap of knowledge between signal trunsduction reactions in the proliferating cells and gene transcription regulation.

Such computer models systematically produce experimentally testable hypotheses about new molecular mechanisms of gene expression regulation and functions of the genes within regulatory networks. It can assist in discovery of new target genes for medical applications.

2 Cyclonet system architecture

Cyclonet system architecture is shown on Figure 1. Cyclonet data is stored on the server side in relational databases (Fig. 2). MySQL database [14] is used for this purpose.
There are two ways how user can get access to the database content.

Using conventional web browser user can query and edit the database content as well as view diagrams (Fig. 3). BeanExplorer Enterprise Edition [1] technology is used for this purpose. BeanExplorer generates complete user interface of web application directly from the database metadata without the need for manual programming of database and interface objects.

Proceedings of the 6th Russian Conference on Digital Libraries RCDL2004, Pushchino, Russia, 2004

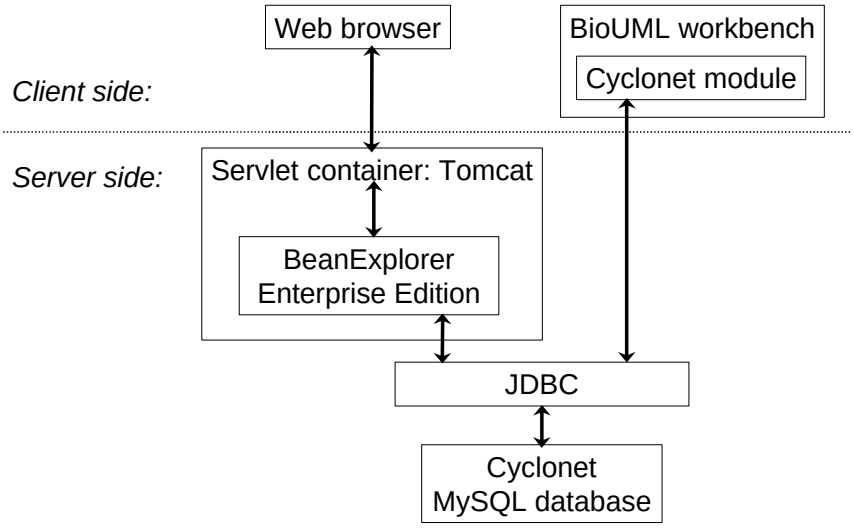

Figure 1. Cyclonet system architecture.

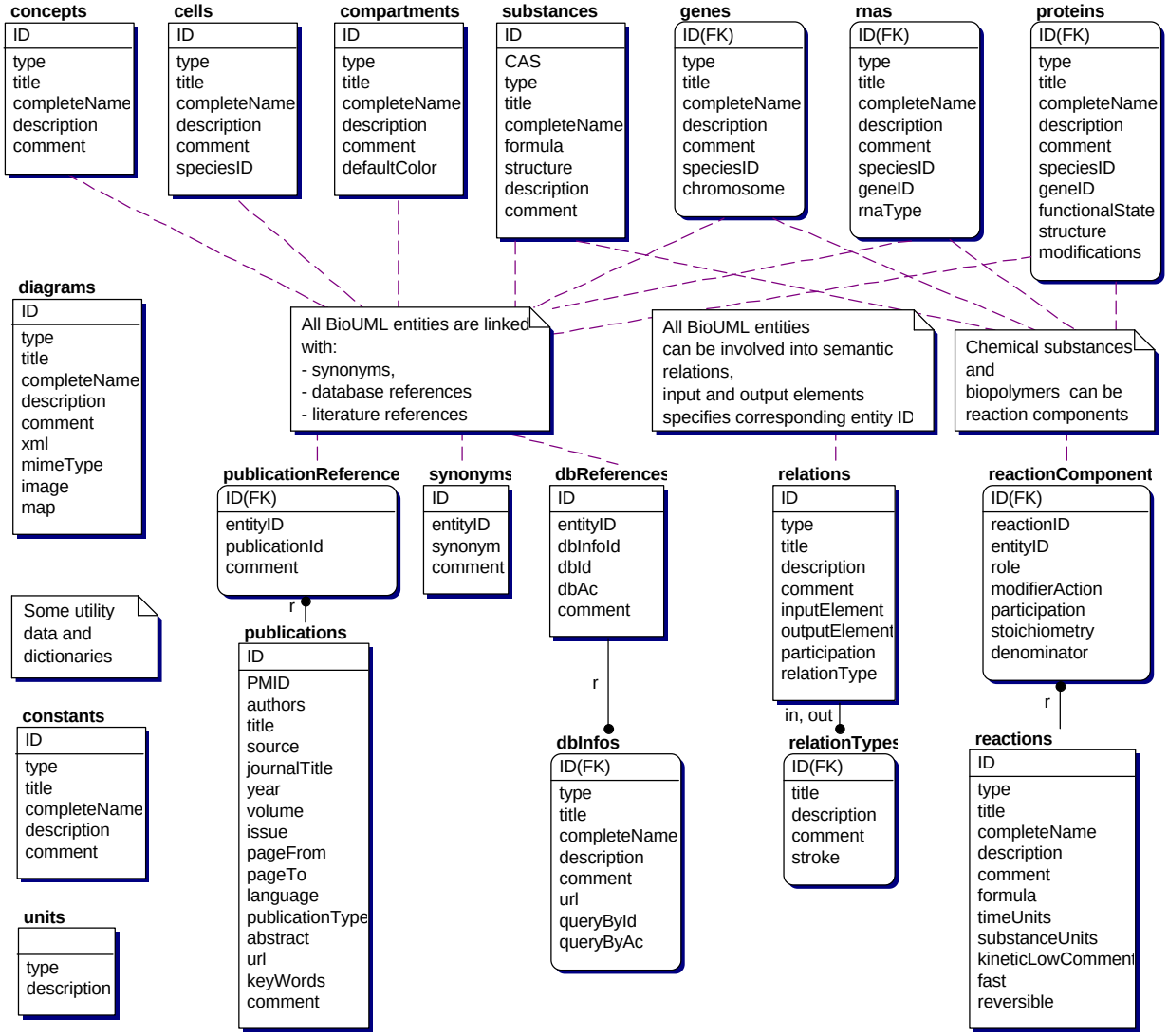

Figure 2. Structure Cyclonet database. Only part used by BioUML workbench is shown.

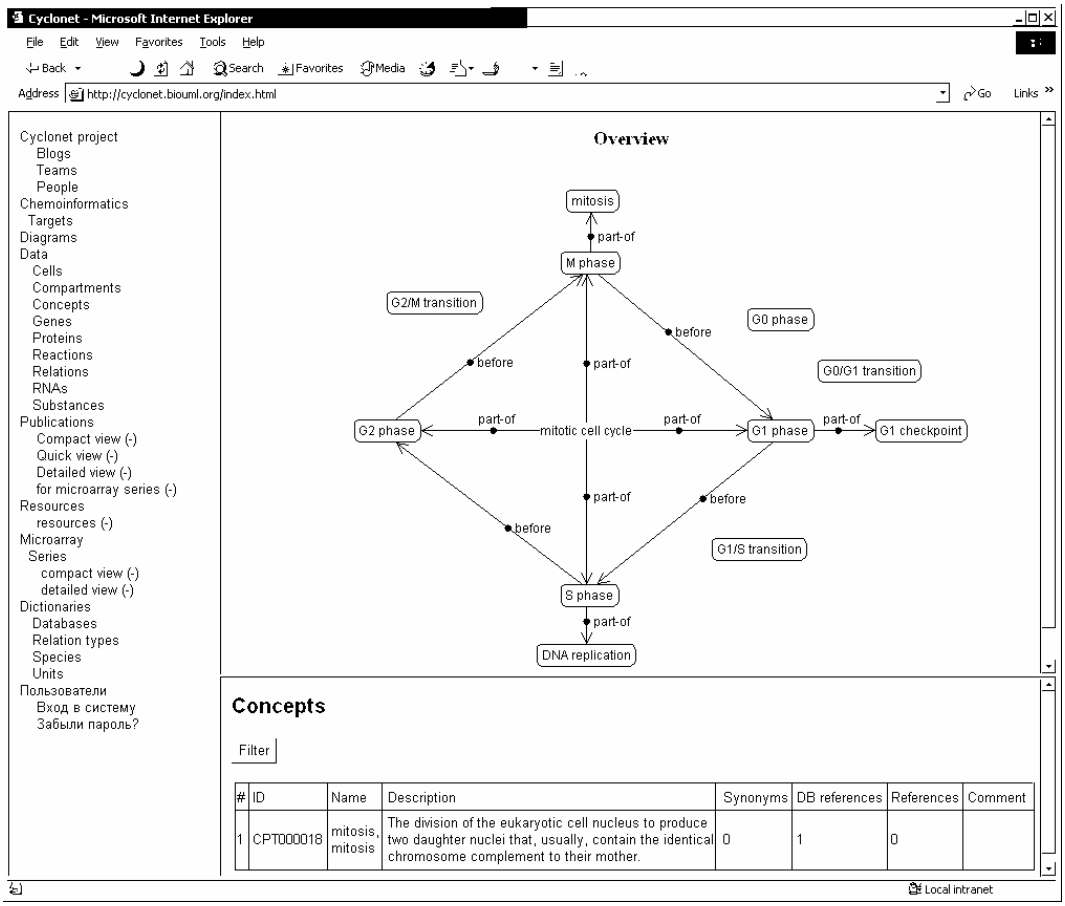

Figure 3. Cyclonet database web interface generated by BeanExplorer EE.

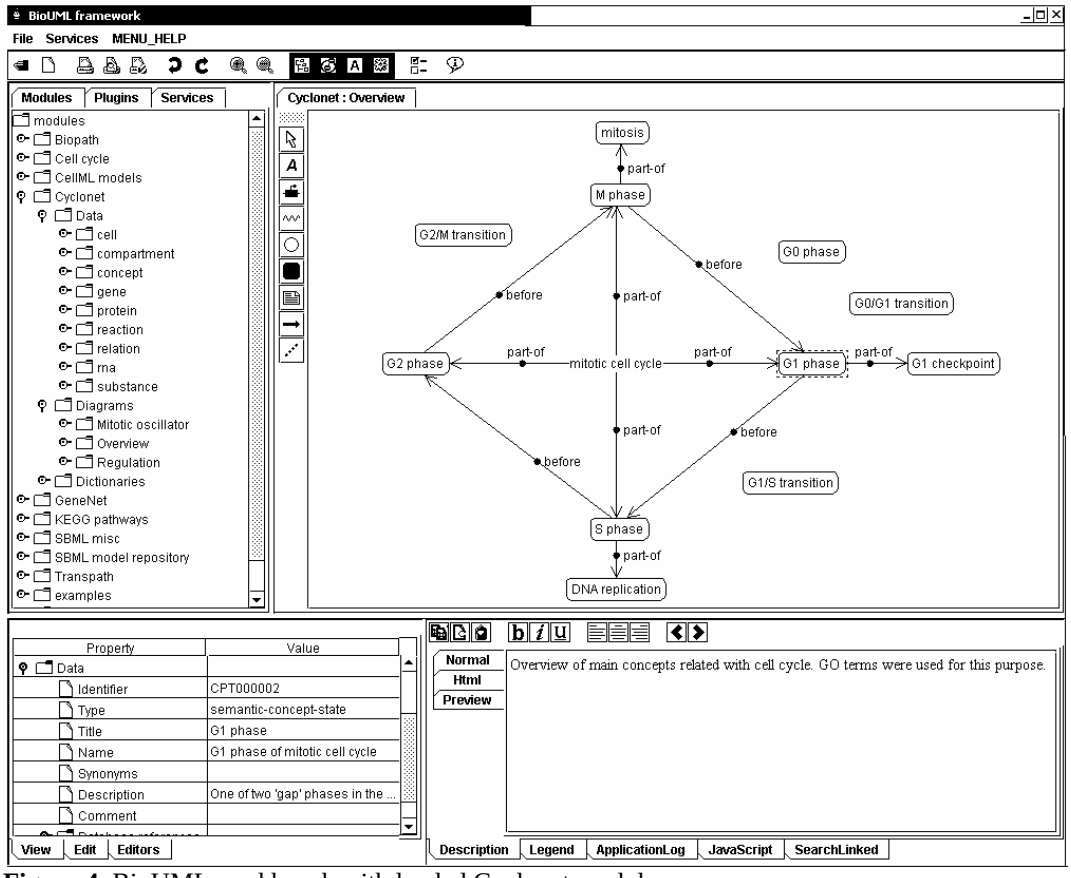

Figure 4. BioUML workbench with loaded Cyclonet module.

BioUML workbench [2] is used to query and edit the database content, view and edit diagrams, analyse and simulate the described systems behaviour using MATLAB or own BioUML simulation engine. Special database module is used to integrate Cyclonet database into BioUML workbench (Fig. 2, 4).

3 Cyclonet database content and structure

The database contains information about cell cycle specific genes, proteins, protein complexes and their interactions, diagrams of cell cycle regulation for vertebrates, models of cell cycle and results of their analyses, microarray data, literature references and other related resources.

The data are compiled from different databases (TRANSPATH[12], TRANSFAC[13], TRANSCompel [9], GeneOntology[6] and other) and from literature annotation. Known cell-cycle models are imported from SBML[7] and CellML[3] model repositories and annotated manually based on literature.

Cyclonet database consists from several sections:

- BioUML data - tables used by BioUML workbench for diagrams and diagram components description (fig. 2);
- microarray - microarray data related with cell cycle and cancer;
- publications - references and full articles concerning cell cycle regulation, cancer and drugs;
- resources - list of miscellaneous useful resources;
- chemoinformatics - issues related with chemoinformatics.

4 BioUML data

Standard BioUML diagram types were used for formal description and simulation of cell cycle regulation:

1) semantic network - describes relationships between main concepts (G1 phase, G1/S transition, mitotic checkpoints) and components of cell cycle regulation (see Fig. 3 as en example);
2) pathway structure - describes structure of cell-cycle regulatory network as a compartmentalised graph. We have divided the network into a number of diagrams that describe some parts of cell-cycle regulatory network in details (for example, network that provides G1/S transition). This diagram type uses graphic notation that is quite similar with GeneNet notation[10];
3) pathway simulation - allows BioUML workbench to generate mathematical models automatically and simulate systems behaviour (see Fig. 5 as an example).

BioUML workbench also defines standard data types that can be used for formal description biological system components (entities, graph nodes) and interactions between them (relationships, graph edges). Each BioUML entity type is mapped into corresponding table of Cyclonet database:

- *concepts* table – describes biological concepts related with cell cycle regulations (for example, G1 phase, mitosis, DNA replication);
- *cells* – provides brief description of cell types and cell lines;
- *compartments* – provides brief description of cellular (for example, nucleus, cytoplasm) and organism compartments (for example, liver, blood).
- *substances* – contains information about chemical substances;
- *genes* – genes description;
- *rnas* – RNAs description;
- *proteins* – protein and their complexes description.

Several auxiliary tables store information about entities synonyms, references to literature and other databases:

- *publicationReferences* – links entity and corresponding publication references;
- *publications* – detailed information about publications;
- *synonyms* – information about synonyms used for corresponding entities;
- *dbReferences* – references to other databases (for example TRANSPATH, GeneOntology);
- *dbInfos* – brief description of external databases used in dbReferences table.

To describe relationships between system components two approaches are used:

- *relations* table contains information on semantic relationships used in semantic network diagrams (Fig. 3);
- *reactions* and *reactionComponents* tables describe chemical reactions used in pathway diagrams (Fig. 5).

Information about diagrams is stored *diagrams* table. Here is brief description of its main fields:

- ID – diagram identifier;
- type – diagram type (semantic network, pathway structure or pathway simulation);
- title – diagram title;
- description – description of the diagram in HTML format;
- xml – diagram structure in DML format [4]. DML – Diagram Markup Language – is XML format developed for BioUML workbench to store diagram structure and layout;
- image – diagram image in GIF format;
- map – diagram image map. It contains areas and hyperlinks for diagram components.

Cyclonet module provides mapping of Cyclonet data into Java objects used by BioUML workbench and vice versa. It also generates diagram image and corresponding map that further can be used in web interface.

5 Application of Cyclonet to model cell cycle

BioUML workbench can be used for visual modelling of biological systems. When the system is described as pathway simulation diagram (Fig. 5) BioUML workbench is able to generate executable model as MATLAB M-file and start MATLAB engine for model simulation and analyses.

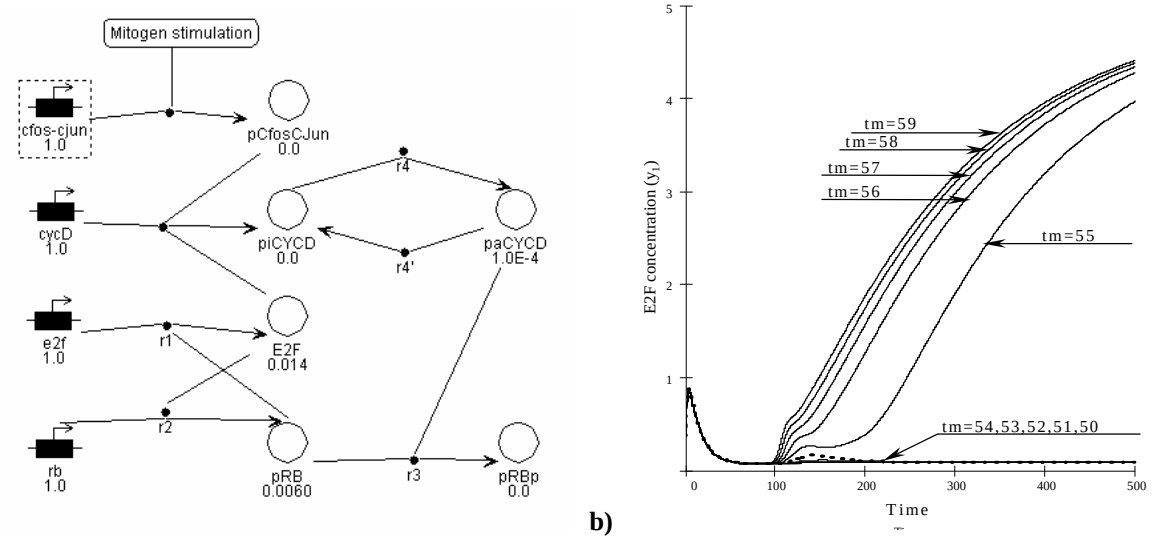

Figure 5. a) G1/S entry model [8] described using BioUML technology;
b) dynamic of E2F-1 concentration in the model corresponding to the two different modes of the system: quiescence or cycle progression.

Computer simulation methods have been applied to study the dynamics of gene networks regulating the cell cycle of vertebrates. The data on the regulation of the key genes obtained from the Cyclonet database have been used as a basis to construct gene networks of different degrees of complexity controlling the G1/S transition, one of the most important stages of the cell cycle. The behaviour dynamics of the model constructed has been analysed. Two qualitatively different functional modes of the system has been obtained. It has been shown that the transition between these modes depends on the duration of the proliferation signal (see Fig. 4). It has also been demonstrated that the additional feedback from factor E2F to genes c-fos and c-jun, which was predicted earlier based on the computer analysis of promoters, plays an important role in the transition of the cell to the S phase.

6 Discussion

Cyclonet database is the first database that was created using BioUML and BeanExplorer EE technologies. The other example is database on molecular mechanisms of chronic lung diseases [11].

We believe that suggested combination provides powerful and convenient approach for formal description and simulation of biological and other complex systems. BioUML workbench is used as “thick” client for visual modelling. BeanExplorer EE provides web interface where web browser is used as “thin” client to query and edit information from Cyclonet database. It should be noted that some part of information, for example microarray data, is available only through web interface.

Traditionally SRS system [5] is used for biological databases web publishing. Here we have demonstrated that BeanExplorer EE can be applied for the same task. This technology provides following advantages in comparison with SRS system:

- BeanExplorer EE allows user to edit the database content through web interface;
- different operations (edit or delete selected records, data import and export, etc.) can be associated with different views;
- security and roles support - different user groups can have different roles assigned in the system; different views and operations are available for different roles;
- categories support – this allows user to classify hierarchically the data, for example genes can be classified by their effect;
- data and metadata update without server restart – system developer can change data or user interface and this changes will be immediately available for the users;
- seamless integration with relational databases (Oracle, MS SQL server and other).

References

[1] BeanExplorer Enterprise Edition, DevelopmentOntheEdge.com, 2004. http://www.beanexplorer.com

[2] BioUML - open source extensible workbench for systems biology. Biosoft.Ru, 2004. http://www.biouml.org

[3] CellML - an XML-based language for describing and exchanging models of cellular and subcellular processes. Physiome Sciences, 2001. http://www.cellml.org

[4] DML - Diagrams Markup Language. Biosoft.Ru, 2004. http://www.biouml.org/dml.shtml

[5] T. Etzold, A. Ulyanov., P. Argos. SRS: information retrieval system for molecular biology data banks. In *Methods of Enzymology*, volume 266, pages 114-28, 1996.

[6] Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. In *Nature Genetics*, volume 25, pages 25-29, 2000. http://www.geneontology.org

[7] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, et al., The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models. In *Bioinformatics*, volume 19(4), pages 524-531, 2003. http://www.sbml.org

[8] A.E. Kel, I. Deineko, O.V. Kel-Margoulis, E. Wingender, V. Ratner. Modelling of gene regulatory network of cell cycle control. Role of E2F feedback loops. In *Proceed. German Conf. on Bioinformatics*, Heidelberg, pages 107-114, 2000.

[9] O.V. Kel-Margoulis, A.E. Kel, I. Reuter, I.V. Deineko, E. Wingender. TRANSCompel: a database on composite regulatory elements in eukaryotic genes. In *Nucleic Acids Research*, volume 30, pages 332-334, 2002.

[10] F.A. Kolpakov, E.A. Ananko, G.B. Kolesov and N.A. Kolchanov. GeneNet: a database for gene networks and its automated visualization. In *Bioinformatics*, volume 14(6), pages 529-537, 1998. http://wwwmgs.bionet.nsc.ru/systems/mgl/genenet

[11] F.A. Kolpakov, A.F. Kolpakova. New approaches to understanding of heavy metals role in pathogenesis of chronic respiratory diseases. In *European Respiratory Journal. Abstracts.13th ERS Annual Congress. Vienna, Austria*, page 267, 2003. http://biopath.biouml.org

[12] M. Krull, N. Voss, C. Choi, S. Pistor, A. Potapov, E. Wingender. TRANSPATH: an integrated database on signal transduction and a tool for array analysis. In *Nucleic Acids Research*, volume 31(1), pages 97-100, 2003.

[13] V. Matys, E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A.E. Kel, O.V. Kel-Margoulis, D.U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, E. Wingender. TRANSFAC: transcriptional regulation, from patterns to profiles. In *Nucleic Acids Research*, volume 31, pages374-378, 2003.

[14] MySQL open source database. MySQL AB, 2004. http://www.mysql.com

[15] B. Novak, J.J. Tyson. Modelling the controls of the eukaryotic cell cycle. In *Biochemical Society Transactions*, volume 1(Pt 6), pages 1526-1529, 2003.

? This work was partially supported by INTAS grant Nr. 03-51-5218.