Cyclonet – AN INTEGRATED DATABASE ON CELL CYCLE REGULATION AND CARCINOGENESIS

***Kolpakov F.*[*1, 2], *Poroikov V.*[3], *Sharipov R.*[1, 2, 4], *Milanesi L.*[5], *Kel A.*[6]**
[1] Design Technological Institute of Digital Techniques, SB RAS, Novosibirsk, 630090, Russia; [2] Institute of Systems Biology OOO, Novosibirsk, 630090, Russia; [3] Institute of Biomedical Chemistry, RAMS, Moscow, Russia; [4] Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; [5] Institute of Biomedical Technologies, CNR, Segrate (MI), Italy; [6] BIOBASE GmbH, Wolfenbuettel, Germany
[*] Corresponding author: e-mail: fedor@biouml.org

Key words: database, cell cycle regulation, cancer, drug design, BioUML

SUMMARY

Motivation: Success of the biologically reasonable modeling of cellular systems depends on the completeness of our knowledge and integration of all fundamental molecular processes, such as signal transduction, regulation of gene expression and metabolism. Mammalian cell cycle is a good example of the system where such natural integration of all the molecular processes plays important role in their regulation. Despite of the massive development of various biological databases, no specialized repository was created so far that would integrate all knowledge on cell cycle.

Results: We have developed the Cyclonet – an integrated database on cell cycle regulation and carcinogenesis. It contains information about known specific genes, proteins and their complexes involved in regulation of cell cycle and carcinogenesis; diagrams of cell cycle regulation and related processes; models of cell cycle and results of their simulation; links on microarray data on cell cycle and on various types of cancer, information on anticancer drug targets as well as their ligands, broad literature references and other related resources.

Availability: http://cyclonet.biouml.org.

INTRODUCTION

The main goal of the Cyclonet database is to integrate information from genomics, proteomics, chemoinformatics and systems biology and provide specialized and comprehensive resource in order to enable molecular biologists working in the field of anticancer drug development to analyze systematically all this data and generate experimentally testable hypothesis (Fig. 1).

METHODS

Novel software technologies were used for development of Cyclonet database:

- BioUML technology (Kolpakov, 2004; http://www.biouml.org) was used for formalized description, visual modeling of eukaryotic cell cycle and for query and editing of the database content. BioUML workbench also allows simulate the described systems behavior using Java or MATLAB simulation engines;

- Biopath database (http://biopath.biouml.org) stores diagrams, description diagram elements (genes, proteins, substances, concepts, reactions and semantic relationships) as well as for storage mathematical models and results of simulation;
- using BeanExplorer Enterprise Edition (http://www.beanexplorer.com) web interface for Biopath database was developed.

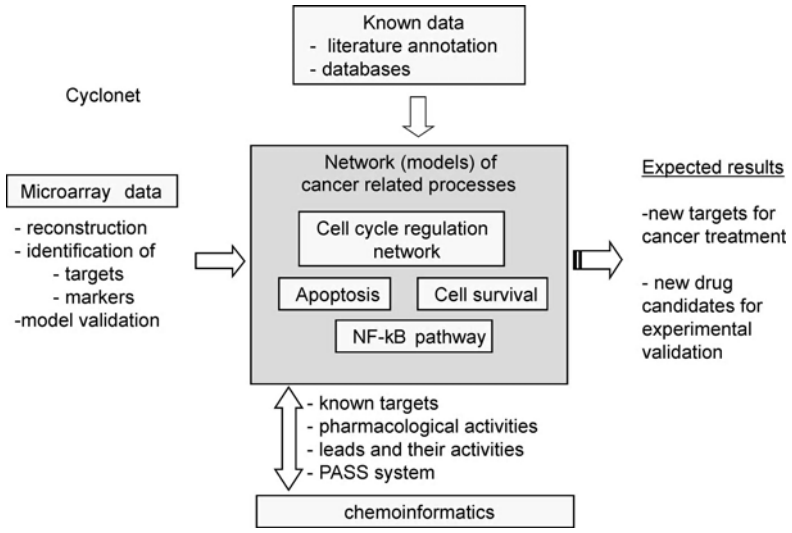

Figure 1. Diagrams (models) of carcinogenesis and related processes as the basis of data integration in Cyclonet database.

The data in Cyclonet are compiled mainly from manual literature annotation. Links to the public databases, GeneOntology, RefSeq and Ensembl, are provided from genes, proteins and other corresponding entries. Known cell-cycle models are imported from SBML (Hucka *et al*., 2003; http://www.sbml.org) and CellML (http://www.cellml.org) model repositories and annotated manually based on literature. Cyclonet also contains a vast body of literature references that are arranged by categories.

RESULTS

Cyclonet database integrates data of genomics, proteomics, chemoinformatics, and systems biology for their use in drug design:

- genomics – we have collected and categorized (arranged) links to available microarray experiments (Table 1). During next stage of work (BioUML team is now working on support of microarray data) we will merge microarray data on breast cancer with corresponding BioUML diagrams and analyses tools. Cyclonet includes information about 196 genes related with cell cycle and cancer development.
- proteomics – Cyclonet contains information about 2465 proteins, their complexes and interactions (protein, its modified form – for example, phosphorylation, protein complexes and protein families are considered as a separate entries in protein table).
- chemoinformatics – Cyclonet contains information about 40 key targets for anticancer treatments, 33 related pharmacological activities. 422 ligands known to be related with these activities are placed in database. Chemical and structural formulae are available for these ligands (Fig. 2). This information is used by computer program PASS (Poroikov, Filimonov, 2005; http://www.ibmc.msk.ru/PASS) to predict new ligands with anticancer activities.

- systems biology – Cyclonet contains 351 diagrams describing cell cycle regulation and related systems. Cyclonet also includes 32 mathematical model of cell cycle regulation annotated from literature (Fig. 3). The number of collected diagrams related to cell cycle exceeds the content of any other related databases and resources.

DISCUSSION

New high-throughput methods allow generation of massive amounts of molecular biological data. These, mainly phenomenological, data are often difficult to relate with the activation/inhibition of particular signal transduction pathways and/or transcriptional regulators. A way to facilitate data interpretation is to construct gene regulatory networks that include signal transduction mediators, transcriptional regulators and target genes. This is a complex task, not only because of the huge number of molecules involved, but also because of variations across tissues, developmental stages and physiological conditions. However, these networks hold the key to the understanding of the regulatory processes within a cell.

The aim of the Cyclonet database is, therefore, to develop an integrative approach that will help researchers to understand the cell cycle process through modeling and simulation of gene regulatory networks. It combines and puts results of different high-throughput experimental methods together under the roof of bio- and chemoinformatics, to exploit the full potential of the included methods as well as the generated data.

Table 1. Microarray web resources and microarray series related with cell cycle and cancer

Category	Entries	Comment
		Resources → microarray
data → processed	24	This category includes two groups of resources related to cancer and cell cycle experiments: 1) Resources containing raw or processed published data. 2) Resources containing published papers or description of the experiments and in some cases a link to the site with raw data.
data → web interface	10	Resources in this category do not contain information about published papers concerning experiment and raw data but allow exploring expression of a particular gene in the different conditions.
data	7	Unclassified microarray data.
Total microarray resources	41	
		Series (microarray experiments)
Cell cycle	24	Microarray data related with cell cycle
For modeling	36	Data are believed to be useful for modeling
Treatment	50	Data related with different treatment (chemotherapy, hormones, growth factors, etc.) of normal and cancer tissues or cell lines
Cancer → targets	11	Experiments where some genes were identified as targets for cancer treatment
Cancer → markers	4	Experiments where some genes were identified as cancer markers
Cancer total	310	All cancer data.
Total microarray series	354	
		Publications
Publications, microarray	287	354 microarray series refer to 287 articles

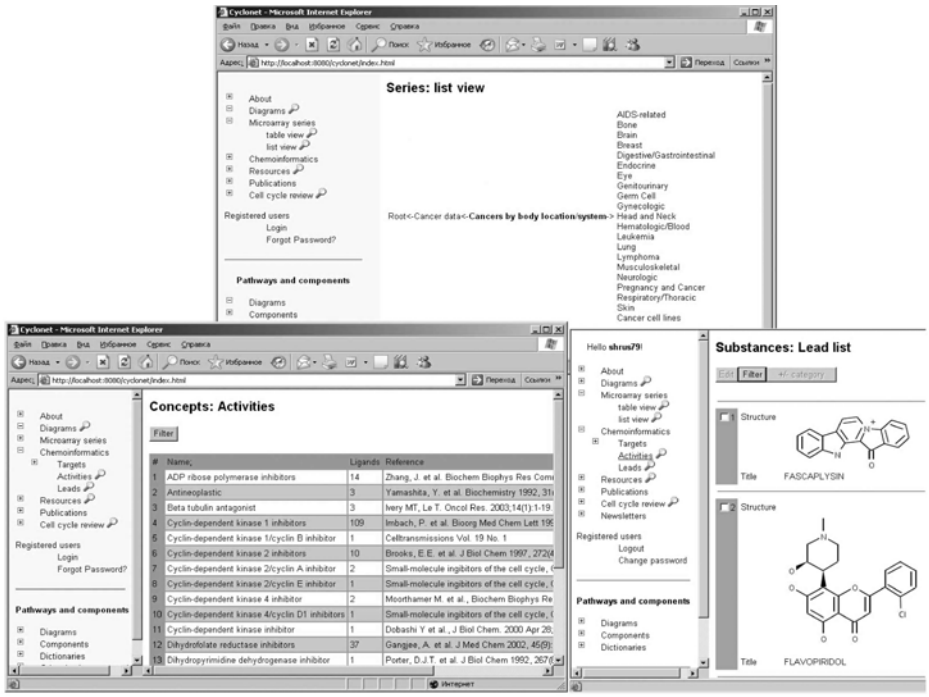

Figure 2. Web interface of Cyclonet database generated with BeanExplorer technology. Top screen displays fragment of microarray series classification in Cyclonet database, bottom left screen demonstrates fragment of list of pharmacological activities for anticancer therapy, bottom right – examples of chemical structure of two cyclin-dependent kinase 4 inhibitors.

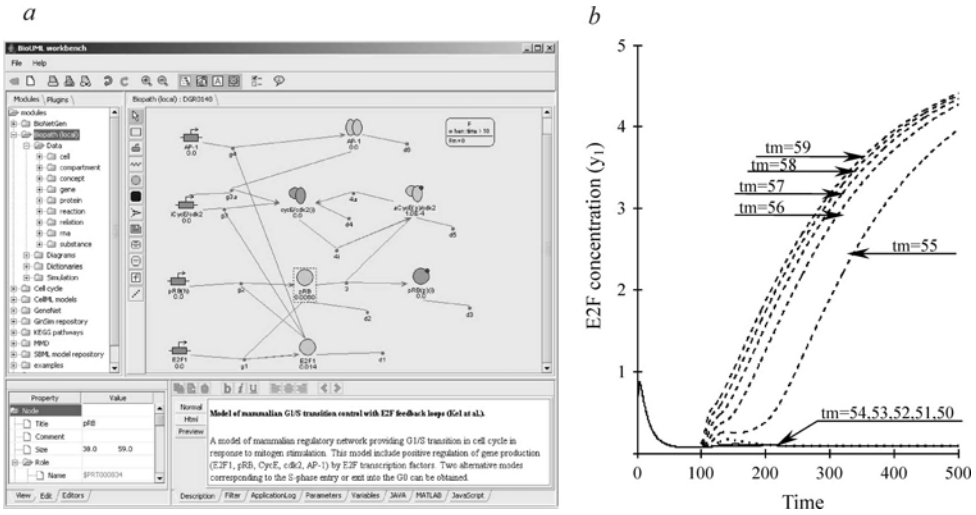

Figure 3. On the left – G1/S entry model (DGR0140, Kel *et al.*, 2000) simulating two different modes of the system: quiescence or cycle progression – described using BioUML technology. On the right – dynamics of E2F-1 concentration – simulation results of the G1/S entry model. tm – duration of system stimulation by mitogen (in seconds).

ACKNOWLEDGEMENTS

This work was supported by INTAS grant No. 03-51-5218, MIUR-FIRB grant No. RBLA0332RH Laboratory for Interdisciplinary Technologies in Bioinformatics and BIOINFOGRID No. 026808. Authors are grateful to V. Komashko and V. Valuev for microarray data annotation and E. Cheremushkina for annotation of some diagrams.

REFERENCES

Hucka M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Kel A. *et al*. (2000) Modeling of Gene Regulatory Network of Cell Cycle Control. Role of E2F Feedback Loops. *Proceedings of the German Conference on Bioinformatics (GCB 2000)*, pp. 107–114.

Kolpakov F.A. (2004) http://www.bioiml.org.

Poroikov V., Filimonov D. (2005) PASS: Prediction of Biological Activity Spectra for Substances. In Helma C. (ed), *Predictive Toxicology*. Taylor & Francis, pp. 459–478.