

ЭЛЕКТРОННАЯ БИБЛИОТЕКА ПО СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ ОРГАНИЗАЦИИ ГЕННЫХ СЕТЕЙ: СИСТЕМА ВИЗУАЛЬНОГО МОДЕЛИРОВАНИЯ

^{1,2*}Подколотный Н.Л., ¹Локтев К.А., ²Ткачев Ю.А., ¹Ананько Е.А.,
¹Колчанов Н.А.

¹Институт цитологии и генетики СО РАН,
Новосибирск-90 пр. Лаврентьева, 10

²Институт вычислительной математики
и математической геофизики СО РАН,
Новосибирск-90 пр. Лаврентьева, 6

*Email: pnl@bionet.nsc.ru

DIGITAL LIBRARY ON GENE NETWORKS' STRUCTURAL AND FUNCTIONAL ORGANIZATION: A SYSTEM FOR VISUAL MODELING

^{1,2*}Podkolodny N.L., ¹Loktev K.A., ²Tkachev Yu.A., ¹Ananko E.A.,
¹Kolchanov N.A.

¹Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

²Institute of Computational Mathematics
and Mathematical Geophysics SB RAS, Novosibirsk, Russia

*Corresponding author: e-mail: pnl@bionet.nsc.ru

The system presented in this work is aimed at visual modeling of the structural and functional organization of complex gene networks. The visual model may simplify the gene network description and analysis of its functioning. It could serve as a basis for developing a mathematical model of a gene network describing dynamics of gene networks.

This system is compiled of: (1) gene network editor realized in the environment Windows NT/2000/XP, (2) application server, which provides logic operation in the system's functioning and connection to the database server, (3) database server. Development of the application server and database server was made in the Oracle9i environment.

Введение

В Сибирском отделении РАН ведется разработка информационно-компьютерных технологий нового поколения, интегрированных в компьютерной системе GeneNet [1-4], обеспечивающих решение широкого ком-

плекса проблем компьютерного анализа и моделирования сложных молекулярно-генетических систем (генных сетей, генетически контролируемых метаболических путей, путей передачи сигналов и т.д.) включая: накопление данных и знаний о генных сетях; построение и анализ структурно-функциональных моделей генных сетей; интеграция информации по генным сетям и метаболическим путям; построение математических моделей, описывающих динамику генных сетей; численный анализ моделей генной сети (решений систем дифференциальных уравнений); построение математических моделей генных сетей, описывающих динамику генных сетей и их численный компьютерный анализ; верификация моделей и оценка параметров моделей по экспериментальным данным, экспертным суждениям, экспериментально наблюдаемым траекториям поведения системы; использование разработанных моделей для решения задач накопление вариантов расчетов, анализ и обобщение и формирование базы знаний системы; исследование на молекулярно-генетическом, клеточном и организменном уровнях динамического поведения сложных молекулярно-генетических систем (генных сетей) в норме, при патологических состояниях, а также при воздействии неблагоприятных факторов внешней среды; поиск оптимального управления и коррекции поведения генных сетей при патологических состояниях.

В настоящей работе представлена система визуального моделирования структурно-функциональной организации сложных генных сетей, ввода данных, а также описание их в электронной библиотеке GeneNet. Визуальная модель позволяет упростить описание генной сети, облегчить анализ ее функционирования и может служить основой для создания математической модели генной сети, описывающей динамику генных сетей [5].

Принципы создания системы

В основу разработки программной системы заложены следующие главные принципы:

- (i) использование объектно-ориентированного визуального моделирования, ориентированного на сложные генные сети;
- (ii) возможность произвольного расширения набора объектов, их свойств (атрибутов) и способов их отображения без внесения изменений в саму программу;
- (iii) широкое использование технологии XML/XSL.

Объектно-ориентированное моделирование объединяет в себе процесс объектно-ориентированной декомпозиции, визуальную нотацию для описания логических, физических, статических и динамических моделей описываемой системы.

Для каждой компоненты генной сети задается ее графическое представление, которое может меняться путем редактирования конфигурационных файлов системы. Это позволяет гибко настраивать систему на различные типы визуального представления генных сетей.

В разработке используются следующие типы декомпозиции генных сетей и подходы облегчающие восприятие сложных систем:

- (i) По уровням детализации описания генной сети. Это обеспечивает возможность построения многоуровневых иерархических моделей, когда детальное устройство какого-либо компонента системы скрыто от пользователя на данном уровне описания. Кроме того, на каждом этапе аннотирования обеспечивается возможность ввода неполной информации, когда механизм реализации какого-либо процесса неизвестен;
- (ii) Пространственная декомпозиция по различным видам компарментов (клетки, ткани, органы и т.д.);
- (iii) Временная декомпозиция, обеспечивающая выделение подсистем по характерным временам изменения их состояния. В этом случае возникает возможность исследования динамики такого рода подсистем независимо от специфики поведения других подсистем, которые можно описывать в рамках заданных характерных времен как статические системы или системы с пороговым переключением. В общем случае для описания используются гибридные непрерывно-логические модели.
- (iv) Использование тематических слоев (части генной сети), интерпретируемых в терминах предметной области. Множество объектов, определяющее тематический слой, описывается ограничениями на их атрибуты и задается в виде запроса к базе данных. Описания слоев и способ их получения хранятся в базе данных;
- (v) «Взгляд» - часть генной сети, определяющая область интересов пользователя на определенном этапе процесса аннотации. Пользователь явно указывает объект, который необходимо вставить или убрать из поля видимости при описании генной сети, какие объекты и связи будут задействованы при вводе новой информации и расширении генной сети. Эта возможность позволяет пользователю концентрировать внимание только на актуальные в определенный момент фрагменты генной сети. Объекты, находящиеся на заднем плане, не активны и рисуются более бледно. При продолжении работы сохраняется последний «взгляд» пользователя на генную сеть;
- (vi) Дерево иерархии элементов схемы позволяет оперировать многокомпонентными объектами (создавать, копировать, удалять, перемещать и выделять) и существенно облегчает восприятие иерархических отношений между объектами в сложных системах.

Представление данных

Описание основных компонент геномной сети: белки, гены, РНК, реакции, регуляторные воздействия и т.д. легко представляются в реляционно-объектном виде.

Взгляды пользователя на данные и результат запроса к базе данных являются геномной сетью, представление которой может меняться.

Такого рода данные относятся к слабоструктурированным типам данных. Для их представления в настоящей работе используется XML технология.

Таким образом, в базе данных системы имеется несколько уровней представления данных: представление элементарных структур и отношений (взаимодействий); виртуальное представление геномных сетей как взгляда пользователя на сложную молекулярно-генетическую систему.

Конфигурационные файлы системы и все обрабатываемые программой данные представляют собой XML документы либо листы стилей XSL. XML документы хранятся в Базе Данных, или генерируются на лету Сервером на основе информации из Базы Данных.

Кроме этого XML описание используется для выполнения запросов на сохранение, модификацию и получение информации об объектах геномной сети из базы данных и для представления информации, связанной с работой системы в целом (ограничения на значения атрибутов, хранение текстов SQL запросов от сервера к базе данных, конфигурационные файлы компонентов системы и т.д.).

Для отображения геномной сети, полученной из электронной библиотеки GeneNet, с помощью стандартного Web Browser разработана программа GeneNetViewer, в которой используется формат векторной графики SVG (Scalable Vector Graphics). Данный формат позволяет получить качественное масштабируемое изображение.

SVG изображения могут быть динамическими и интерактивными. Объектная модель документа Document Object Model (DOM) для SVG, позволяет создавать эффективную анимацию, основанную на использовании скриптов на XML. Обеспечивается обработка таких событий, как, например, перемещение курсора мыши над объектом или щелчок мыши.

Для преобразования информации, получаемой из базы данных в требуемом формате, трансформации представления геномной сети в формат SVG и при вставке информации в базу данных применяется технология XSLT.

Реализация и результаты

В архитектуре системы можно выделить три уровня: (1) база данных Oracle9i; (2) сервер приложений Oracle9iAS, включающий в себя в качестве компонент Servlet Engine, XML SQL Utility, XML Parser. (3) программы клиенты, которые общаются с сервером по протоколу HTTP посредством обмена XML сообщениями.

База Данных реализована в рамках реляционной СУБД Oracle 9i с использованием технологий объектных типов и вложенных таблиц. Данный подход позволяет получать данные, хранящиеся в реляционных таблицах, в виде XML файлов, легко преобразовывать эти данные (с помощью XSL трансформаций) и передавать их в текстовом виде по протоколу HTTP.

В базе данных представлены как системные данные, связанные с работой сервера и редактора генных сетей, так и данные о биологических объектах. Данные о биологических объектах имеют три уровня представления:

- (i) Абстрактные классы, описывающие понятия предметной области или типы объектов (например, такие понятия, как белок, ген, РНК и т.д.);
- (ii) Конкретные реализации различных типов объектов абстрактных данных (например, модифицированный белок);
- (iii) Описание объектов типа сущностей, реакций или управлений в контексте конкретной генной сети. Наследует свойства классов 2 уровня. Однако, часть атрибутов для объектов этого уровня имеют значения определенные только для конкретной генной сети.

Основными функциями сервера являются:

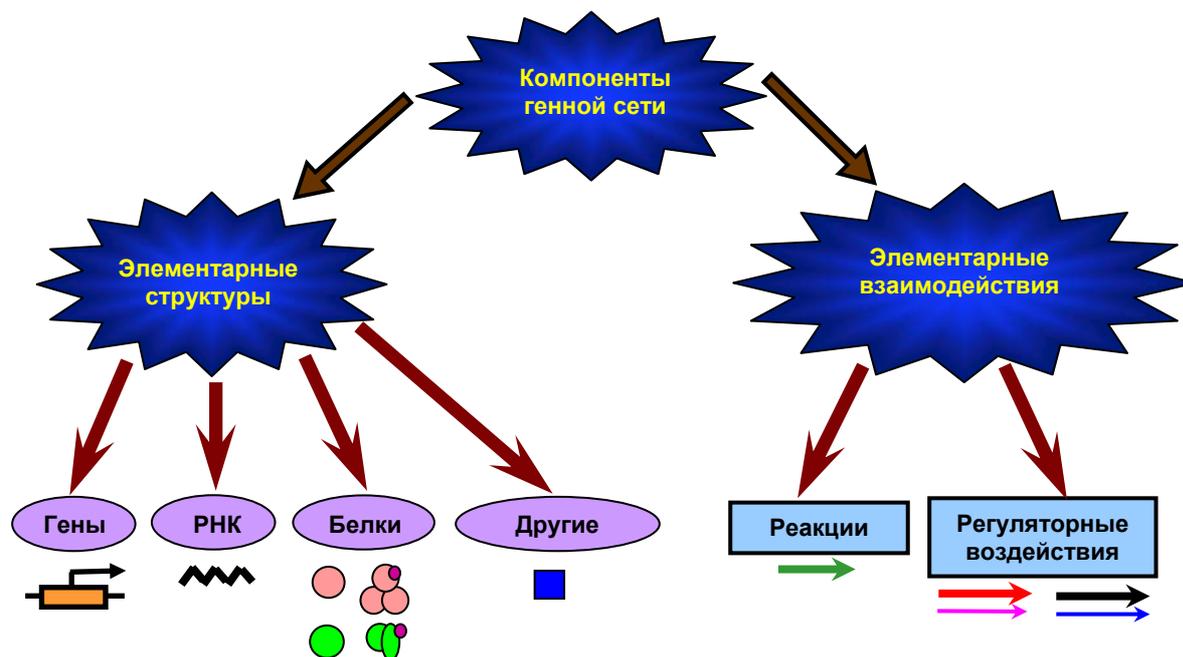
- (i) обеспечение связи нескольких различных клиентов с сервером баз данных;
- (ii) поиск, создание, редактирование и удаление генных сетей, объектов, процессов, управлений и связей в базе данных GeneNet по запросу клиента, администрирование базы данных GeneNet;
- (iii) обеспечение секретности и ограниченности доступа к базе данных;
- (iv) загрузка и сохранение информации в базу данных в формате XML и т.д.;

На Рис.1. представлена иерархия классов в объектном описании генной сети и визуальное представление основных компонент.

Для каждого пользователя определяется категория доступа и права доступа к определенным ресурсам. В реализацию сервера заложена возможность расширения с целью включения модуля анализа генных сетей и моделирования динамики генных сетей. Общение сервера и СУБД происходит с помощью технологии JDBC и языка запросов SQL. Получаемые сервером данные преобразуются в XML формат и передаются клиентам.

В настоящее время реализованы две клиентских программы: редактор генных сетей и вьюер генных сетей.

Рис. 1. Иерархия классов компонент генной сети



Программа визуализации генных сетей (GNViewer) реализована в виде java-applet и ориентирована на использование стандартного Web Browser (типа Internet Explorer 5.5 и выше или Netscape) и Adobe SVG plugin, позволяющие отображать графическое представление генной сети, получаемое от сервера в формате SVG. Использование в Java апплете XML парсера позволяет динамически, на стороне клиента, изменять изображение, выделять слои и другие процедуры облегчающие визуальный анализ генной сети.

Основными функциями редактора генных сетей является:

- (i) Отображение, ввод и редактирование генной сети, представленной в виде графа;
- (ii) Поддержка различных типов декомпозиции сложной генной сети;
- (iii) Создание новых объектов и редактирование атрибутов уже существующих;
- (iv) Запись информации о генной сети в базу данных;
- (v) Загрузка информации о генной сети из базы данных;
- (vi) Взаимодействие с сервером системы посредством обмена XML сообщениями по протоколу HTTP;
- (vii) Передача серверу данных для авторизации пользователя.

Редактор генных сетей GENED разработан в среде Windows NT/2000/XP на языке Visual C++ и использует следующие дополнительные компоненты:

- (i) D2Editor - ActiveX общего назначения для манипулирования сложными геометрическими объектами на плоскости (собственная разработка);
- (ii) MSXML parser - COM объект, предоставляющий развитые средства работы с XML и XSL документами (Microsoft);
- (iii) GDIPlus.dll - библиотека объектно-ориентированной двумерной графики для Windows (Microsoft).

Редактор генных сетей в настоящий момент имеет два базовых способа (типа) отображения диаграмм – в виде дерева (иерархии отношений между объектами генной сети) и в виде двумерного изображения (граф генной сети). Хотя типы отображения фиксированы, конкретный вид изображения в зависимости от значений атрибутов объектов может быть произвольно изменен посредством настройки конфигурационных файлов.

К базовым типам объектов, с которыми оперирует система, относятся, например, такой объект как компартмент, то есть объект, в котором могут содержаться другие объекты (компартменты и простые объекты). Имеется возможность расширения набора типов объектов. Все объекты могут иметь произвольно расширяемый набор свойств, который влияет на их отображение на диаграмме. Набор свойств (атрибутов), тип каждого свойства (является оно обязательным или нет; тип значения; значение по умолчанию), допустимые значения этого свойства и т.д. задаются в специальных конфигурационных файлах. Способ влияния свойств объекта на его изображение также содержится в конфигурационных файлах. К базовым (genetic) свойствам объектов относятся, например, координаты на плоскости (для всех объектов) и размеры (для компартментов).

Все обрабатываемые программой данные, а также все конфигурационные файлы представляют собой XML документы либо таблицы стилей XSL. Использование XML стандарта для описания данных предоставляет гибкие возможности для адаптации редактора к конкретным способам визуализации генных сетей. Изображения объекта могут конструироваться из набора таких графических примитивов как прямоугольник, овал, линия. Описание такого изображения выполняется в стандарте XML.

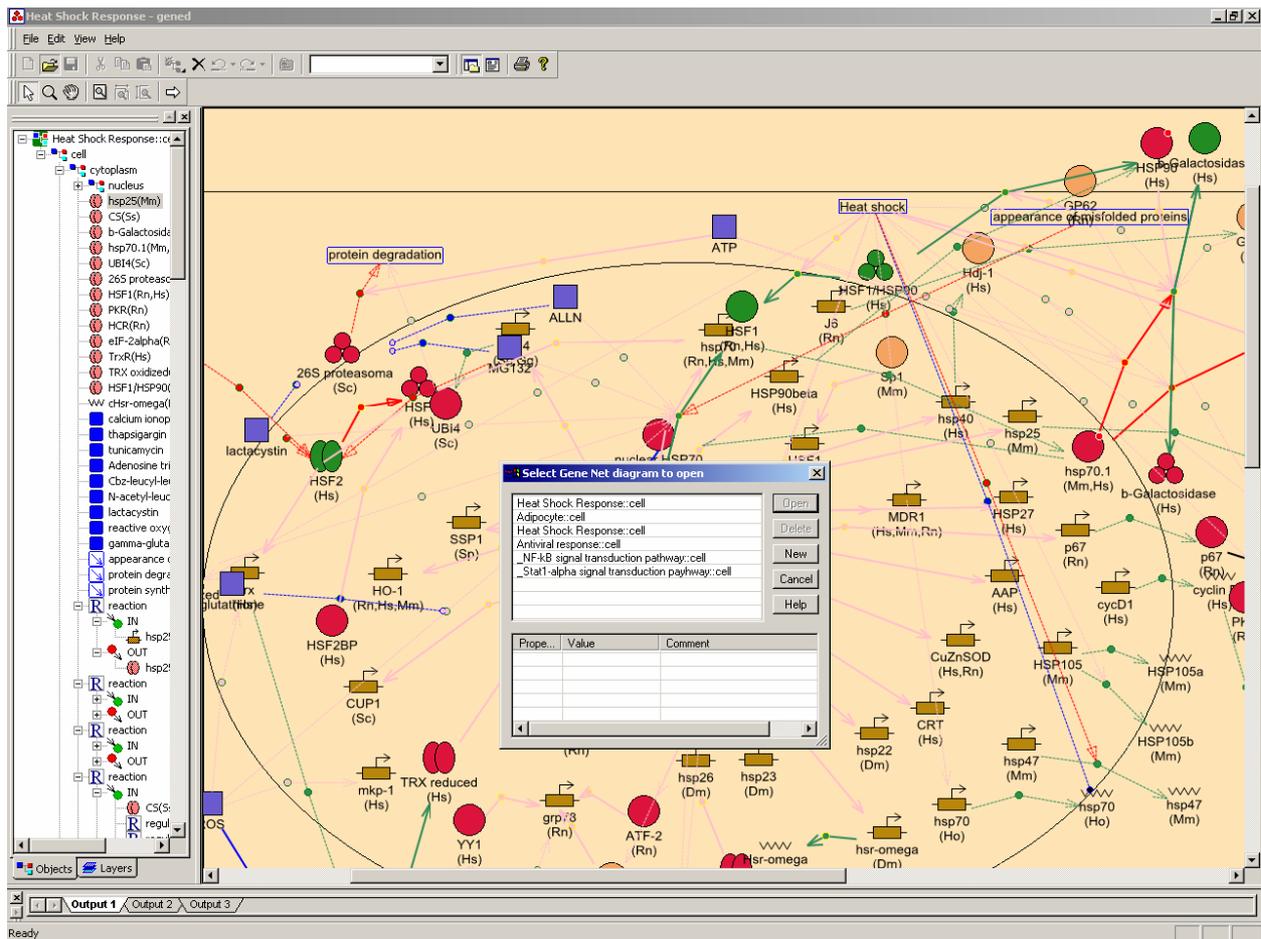


Рис. 2. Screen short интерфейса системы визуального моделирования

Заключение

Разработана система визуального моделирования структурно-функциональной организации сложных генных сетей. Редактор генных сетей позволяет гибко настраиваться на способ визуализации элементов генной сети, расширять множество типов объектов, их атрибутов и ограничений, используемых для верификации данных.

Используемые технологии разработки базы данных и сервера приложений позволяют расширять функциональные возможности системы и обеспечивают возможность работать с большими объемами разнородных данных.

Благодарности

Работа частично поддержана РФФИ (No. 01-07-90376, 02-07-90359), Министерством промышленности, науки и технологий Российской Федерации (№ 43.073.1.1.1501), СО РАН (Интеграционный проект № 65).

Литература

1. Ananko, E.A., Podkolodny, N.L., Ignatieva, E.V., Podkolodnaya, O.A., Stepanenko, I.L., and Kolchanov, N.A. GeneNet system: its status in 2002. *Proc. III Intern. Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2002)*, Novosibirsk, 2002. p. 70-73.
2. Kolchanov N.A., Podkolodny N.L., V.A. Likhoshvai et al. Computer Systemic Biology: Informational and Software Tools for Complex Molecular Biological Systems//*Там же*. p. 140-142.
3. Kolchanov N.A. "Gene Networks Description and Modelling in the GeneNet System". In: *Gene Regulation and Metabolism: Post-Genomic Computational Approaches* eds. Collado-Vides J. and Hofestadt R. MIT Press (Book chapter), 2001.
4. Ананько Е.А., Лихошвай В.А., Колпаков Ф.А., Подколотный Н.Л. и др. Электронная библиотека GeneNet: описание и моделирование генных сетей животных и растений // Вторая Всероссийская научная конференция "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Сборник докладов. Протвино, 26-28 сентября 2000. стр. 66-72.
5. Loktev, K.A., Tkachev, Yu.A., Ananko, E.A., and Podkolodny N.L. A system for visual modeling of gene networks structural and functional.// *Proc. III Intern. Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2002)*, Novosibirsk, 2002. p. 132-135.