

УДК 577:004.4:004.94

НОВЫЕ ВОЗМОЖНОСТИ СИСТЕМЫ MGSmodeller

© 2012 г. **Ф.В. Казанцев¹, И.Р. Акбердин¹, Н.Л. Подколотный¹, В.А. Лихошвай^{1,2}**

¹ Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: kazfdr@bionet.nsc.ru;

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 15 июля 2012 г. Принята к публикации 15 августа 2012 г.

Актуальной проблемой системной биологии является моделирование сложноорганизованных молекулярно-генетических систем и их анализ, что требует разработки специальных подходов, позволяющих рассматривать эти системы как совокупность динамически взаимодействующих подсистем с более простой структурой. В данной работе освещается подход, направленный на ускорение процесса реконструкции и комплексного анализа математических моделей молекулярно-генетических систем с использованием высокопроизводительного кластера.

Ключевые слова: молекулярно-генетические системы, математические модели, MGSmodeller, высокопроизводительные вычисления.

ВВЕДЕНИЕ

Одним из центральных вопросов системной биологии является выявление молекулярно-генетических механизмов функционирования живых систем. Для изучения их динамических характеристик все больше применяются методы математического и компьютерного моделирования, эффективное использование которых невозможно без учета специфики строения молекулярно-генетических систем (МГС). К таким особенностям относятся: линейный характер кодирования информации, при котором важное значение для моделирования наблюдаемой динамики имеют взаимное расположение генов, промоторов, терминальных и других генетических элементов на молекуле ДНК; топология молекулы ДНК в разные периоды клеточного цикла; существование обратимых состояний белковых молекул, меняющих их свойства; явления полиаллельности гена; анизотропия пространственных компартментов живых систем, разделяющая процессы биохимического синтеза; эпигенетическая передача наследственной информации. Эффективным методом решения описанных выше сложно-

стей является модульный подход, при котором последовательно описываются комплексные подсистемы клетки через их более простые (элементарные) составляющие (Ратнер, 1966; Лихошвай и др., 2001; Karr *et al.*, 2012).

Для корректного моделирования динамики функционирования МГС необходимо разрабатывать подходы, объединяющие элементарные подсистемы в рамках одного стандарта, и средства вычислительного эксперимента, позволяющие исследовать модели живых систем большого размера (100–1000 элементарных подсистем) (Karr *et al.*, 2012). Использование высокопроизводительных вычислительных комплексов дает возможность проводить полномасштабный вычислительный эксперимент для исследования такого рода моделей, включая численный анализ особенностей функционирования МГС.

Существующие подходы моделирования направлены на разделение функциональности между специализированными программными системами. Сначала проводят графическую реконструкцию модели МГС в одном инструменте (Hucka *et al.*, 2003; Funahashi *et al.*, 2003; Sauro *et al.*, 2003), затем осуществляется численный анализ модели в других программных

инструментах, таких, как Jws (Olivier, Snoep, 2004), Cellerator (Shapiro *et al.*, 2002), COPASI (Hoops *et al.*, 2006), Matlab. Интеграция систем осуществляется на базе комплекса программ конверторов. Разделение функциональности между программными модулями позволяет распределять задачи и использовать узкоспециализированные решения на каждом из этапов моделирования. В результате разработка новых технологических решений в виде компьютерных систем для создания и анализа математических моделей на базе существующих специализированных программных модулей обеспечивает богатый инструментарий как для численного исследования модели, так и для репрезентации результатов моделирования (Karr *et al.*, 2012).

В данной работе представлено развитие программной среды моделирования молекулярно-генетических систем MGSmodeller (Kazantsev *et al.*, 2008). Богатый функционал системы по реконструкции, исследованию и повторному использованию математических моделей МГС был расширен высокопроизводительными методами их сборки и анализа.

МАТЕРИАЛЫ И МЕТОДЫ

Для реконструкции математических моделей МГС используется система MGSmodeller (Kazantsev *et al.*, 2008; http://modelsgroup.bionet.nsc.ru/?page_id=491). Математические модели реконструируются в формате и по правилам стандарта SibML в рамках обобщенного химико-кинетического подхода (Лихошвай и др., 2001). Анализ результатов моделирования производится средствами системы MGSmodeller и программами Matlab, Gnuplot. Модули компиляции и численного исследования реализованы на языке Fortran. Модули аннотации и редактирования языка SibML, а также постобработки результатов реализованы на языке Java. Расчеты в системе MGSmodeller производились с использованием вычислительного кластера ЦКП «Биоинформатика» СО РАН (bioinformatics.bionet.nsc.ru, www2.sssc.ru/НКС-30Т/НКС-30Т.htm).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В настоящей работе представлена версия программы MGSmodeller для работы на вы-

сокопроизводительном кластере. Как и версия для персонального компьютера (Kazantsev *et al.*, 2008), система предназначена для моделирования МГС в терминах стандарта SibML (Акбердин и др., 2009). Изменению подверглись модули компиляции моделей и модули анализа результатов численного исследования. Переход на высокопроизводительные технологии был связан с необходимостью изучения масштабных комплексных молекулярно-генетических систем (Mironova *et al.*, 2012), что, в свою очередь, невозможно было сделать при использовании мощностей персональных компьютеров за приемлемое время.

Постановка задачи компиляции модели

Математические модели в компьютерной среде MGSmodeller представлены в рамках стандарта SibML как совокупность элементарных подсистем молекулярно-генетических систем. Их реконструкция в рамках среды моделирования производится на основе блочного принципа (Лихошвай и др., 2001). Сначала производится декомпозиция исследуемого объекта до уровня элементарных подсистем, которыми могут быть реакции ферментативного синтеза, подсистемы регуляции экспрессии генов, системы сплайсинга, транспорта, трансляции, процессы созревания и модификации белков, деградации макромолекул и др. Далее описываются математические модели каждой подсистемы, из которых формируется база элементарных моделей. На этой основе исследователь конструирует из элементарных моделей, как из строительных блоков, модель исследуемого объекта. Для этого описывается сценарий сборки модели – файл, содержащий заданную структурно-функциональную организацию модели целевого объекта (Лихошвай и др., 2001), в котором указывается система отношений компартментов (структурный уровень организации целевого объекта) и для каждого компартмента указываются подсистемы, которые должны быть включены в него (функциональный уровень организации объекта).

Предложенный подход позволяет эффективно строить модель исходного целевого объекта, а также модели систем с вариацией структурно-функциональной организации, необходимость рассмотрения которых дикту-

ется текущим исследованием и/или задачами изучения влияния перестроек на уровне компартментной организации системы, изучения влияния мутаций, делеций, вставок и других модификаций, протекающих на генетическом уровне; необходимостью изучения альтернативных механизмов функционирования подсистем и многими другими потребностями, которые могут возникнуть в процессе моделирования.

Этап сборки модели осуществляется как процесс «доопределения» моделей элементарных подсистем – атрибутами, которые осуществляют их однозначную привязку к компартментам; атрибутами, несущими дополнительную информацию о функционировании элементарных моделей в составе комплексной модели целевого объекта. Для комплексных моделей, обладающих сложной структурой, эта операция является ресурсоемкой, требующей постоянного сравнения строковых идентификаторов объектов моделирования. При использовании высокопроизводительного кластера данная операция может осуществляться параллельно для разных частей сценария сборки модели, ускоряя процесс в несколько раз (Казанцев и др., 2012). Например, при исследовании распределения фитогормона ауксина, одного из ключевых морфогенов в регуляции роста и развития растений, по двумерному ансамблю клеток (Mironova *et al.*, 2012) время сборки модели было уменьшено с десятка часов до десятка минут средствами рассматриваемого подхода.

Постановка задачи численного исследования

При постановке задачи численного исследования модели в среде MGSmodeller пользователь может задать временные точки для вывода промежуточного результата на печать.

Для моделирования молекулярно-генетических систем со сложной структурой начальные концентрации и значения параметров модели можно задавать по шаблону. В частности, для многокомпаратментных систем возможно одной операцией установить одинаковую концентрацию одноименных веществ во всех компартментах и только для выбранных компартментов задавать индивидуальное значение концентрации, тем самым производя спецификацию параметра локализации (см. табл. 1).

Для интегрирования дифференциальных уравнений при численном исследовании модели в MGSmodeller используется метод Гира. Также реализована возможность экспорта модели в форматы других систем моделирования (Matlab, Mathematica, SBML), что позволяет проводить дополнительный анализ и сравнение результатов моделирования алгоритмами других систем.

В системе моделирования MGSmodeller добавлена возможность численного исследования нескольких моделей (вариантов одной модели) одновременно в параллельном режиме. На этой базе реализована возможность анализа чувствительности математической модели к варьированию параметров с заданным интервалом и шагом изменения (по принципу «параметры изменяются одновременно» или по принципу последовательного варьирования каждого из параметров).

Дополнительные возможности

В результате численного эксперимента для моделей больших размерностей исследователь, как правило, получает большие объемы информации, и возникает проблема их интерпретации, анализа и визуализации. В случае если не хватает возможностей базовых средств визуализации,

Таблица 1

Инструкции к спецификации значений переменных

Объект в SiBML	Семантика
1,5 '<I(Pin3), T(protein)>'	Всем переменным, связанным с функционированием белка <i>Pin3</i> , присвоить значение 1,5
0,5 '<C(c4,05), I(Pin3), T(protein)>'	Переменной, соответствующей концентрации белка <i>Pin3</i> , функционирующего в компартменте c4,05, присвоить значение 0,5

в рамках системы MGSmodeller результаты моделирования представлены в структурированном виде. Организация атрибутов переменных модели, задающих ассоциацию с контекстом моделирования (см. табл. 2), позволяет проводить постобработку данных сторонними программами, в том числе, используя специализированные инструменты визуального анализа (<http://www.gnuplot.info>; Cedilnik *et al.*, 2006).

Возможности программного модуля были успешно протестированы в работе В.В. Мироновой и соавт. (Mironova *et al.*, 2012). Была показана высокая эффективность работы модуля, выраженная в минимальных затратах программных и временных ресурсов исследователей на этапах создания и анализа модели.

Построение конвейеров обработки данных

Скорость сборки модели и удобные принципы получения ее различных вариантов важны на первых этапах моделирования. Когда установлена структура математической модели, определен комплекс задач и сформулированы гипотезы, которые планируется проверить с помощью разработанной модели, наступает этап численного эксперимента и анализа полученных результатов. Применение высокопроизводительного подхода на этом этапе моделирования дает преимущество – при численном анализе можно создавать цепочки технологических процессов, которые могут выполняться параллельно. И на выходе будет получена уже обработанная сводная информация.

Например, в работе В.В. Мироновой и соавт. (Mironova *et al.*, 2012) была использована

следующая последовательность действий в полуавтоматическом режиме: постановка численного эксперимента для варьирования набора параметров – генерация вариантов модели с различными значениями параметров – проведение численных экспериментов – постобработка данных – реорганизация данных – построение различных типов кривых и результирующей поверхности решений для двухмерного массива клеток (рис.). Временные затраты на проведение одного численного эксперимента изменялись от десятка минут до нескольких часов (в зависимости от значения параметров). Было проведено несколько тысяч экспериментов, расчет которых в совокупности занял двое суток.

При исследовании динамики функционирования МГС могут потребоваться методы обработки и визуализации результатов расчетов, отличные от типовых, встроенных в систему. Удобный способ описания результатов расчетов, представленный в статье, существенно уменьшает время создания собственных инструментов и встраивания их в технологическую цепочку процесса обработки данных.

ЗАКЛЮЧЕНИЕ

Исследование молекулярно-генетических механизмов функционирования живых систем с высокой степенью подробности требует рассмотрения и учета особенностей функционирования сотен элементарных подсистем в одно и то же время в различных компартментах системы и т. д. Подготовка данных к этапу анализа в масштабных моделях является отдельной нетривиальной задачей, решение которой заключается в организации модели и данных

Таблица 2

Примеры объектов языка моделирования

Объект в SiBML	Семантика
<C(nucleus), I(CYP79B2), T(gene)>	Переменная, описывающая функционирование гена <i>CYP79B2</i> , локализованного в компартменте nucleus
<C(nucleus), I(CYP79B2), T(rna)>	Переменная, описывающая функционирование РНК гена <i>CYP79B2</i> , находящейся в компартменте nucleus
<T(protein), I(IAMT1p)>	Переменная, описывающая функционирование белка IAMT1p (в однокомпарментной модели, название компартмента опускается)

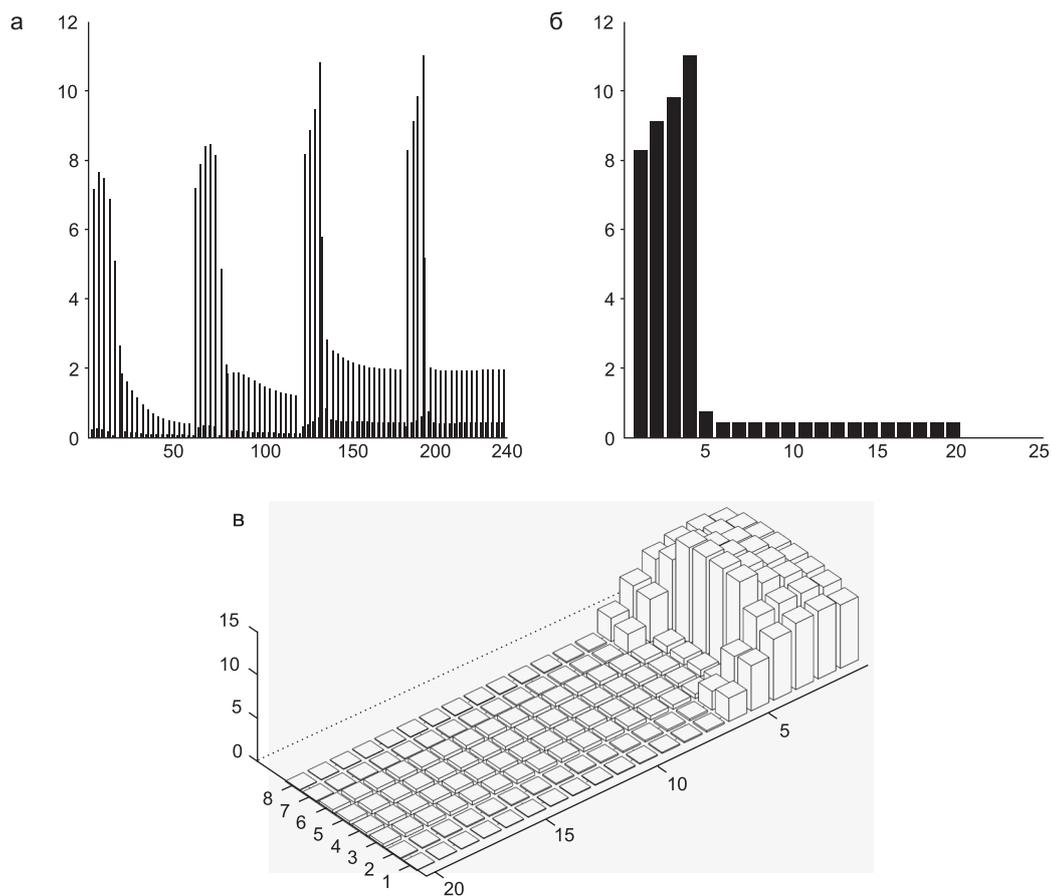


Рис. Примеры сортировки, упорядочения и отображения переменных для ансамбля клеток.

а – вектор всех переменных в конечной точке расчета (по оси X – индекс переменной, по оси Y – ее значение); б – значения переменных, характеризующих концентрацию ауксина в клетках корня вдоль центральной оси (по оси X – индекс клетки, по оси Y – значение переменной в этой клетке); в – интегральная поверхность распределения концентрации фитогормона вдоль осевого среза корня в двухмерном варианте, образованная значениями концентраций ауксина в каждой клетке (по оси X – индекс клетки, по оси Y – индекс клетки, по оси Z – значение переменной в клетке с индексом (X, Y)).

моделирования в форматах, поддерживающих автоматизированный способ их обработки и повторного использования.

Предложен подход к математическому моделированию динамики функционирования биологических систем, который позволяет в процессе реконструкции и исследования моделей сложноорганизованных МГС формировать базы моделей элементарных подсистем для их повторного использования при решении новых задач. Заложенные в программную систему MGSmodeller алгоритмы сокращают время на создание различных вариантов математической модели исследуемого целевого объекта, позволяющих описывать мутации, делеции, вставки и другие молекулярно-генетические модифи-

кации, а также проводить проверку гипотез об альтернативных механизмах функционирования подсистем.

Применение высокопроизводительных вычислений позволяет существенно сократить время на проведение анализа комплексной модели и одновременно исследовать воздействие одних и тех же внешних факторов на динамику функционирования различных вариантов структурно-функциональной организации системы.

Анализ сложноорганизованных моделей часто требует применения дополнительных специализированных инструментов. Предложенный подход к организации и описанию структуры атрибутов объектов моделирования и результа-

тов численного исследования облегчает процесс разработки такого инструментария.

Работа выполнена при частичной поддержке РФФИ (гранты № 10-01-00717, № 11-04-01254а, 12-04-31119 мол_а); Минобрнауки (госконтракт П857); Президиума РАН (проекты 6.8 и 30.29); гранта НШ-5278.2012.4; СО РАН (междисциплинарные интеграционные проекты № 80, № 130) и фонда «Династия» (грант для молодых биологов).

ЛИТЕРАТУРА

Акбердин И.Р., Казанцев Ф.В., Омелянчук Н.А., Лихошвай В.А. Математическое моделирование метаболизма ауксина в клетке меристемы побега растения // Информ. вестник ВОГиС. 2009. Т. 13. № 1. С. 170–175.
 Казанцев Ф.В., Миронова В.В., Новоселова Е.С. и др. Язык моделирования молекулярно-генетических систем SiBML // Тр. конф. «Параллельные вычислительные технологии (ПаВТ) 2012». Новосибирск, 26–30 марта 2012. Новосибирск: ИВМиМГ СО РАН, 2012. С. 722.
 Лихошвай В.А., Матушкин Ю.Г., Ратушный А.В. и др. Обобщенный химико-кинетический метод моделирования генных сетей // Молекуляр. биология. 2001. Т. 3. № 6. С. 1072–1079.
 Ратнер В.А. Генетические управляющие системы. Новосибирск: Наука, 1966. 181 с.
 Cedilnik A., Geveci B., Moreland K. *et al.* Remote large data visualization in the ParaView framework // Eurographics Parallel Graphics and Visualization. 2006. P. 163–170.
 Funahashi A., Morohashi M., Kitano M., Tanimura N. Cell-

Designer: a process diagram editor for generegulatory and biochemical networks // BIOSILICO. 2003. V. 1. P. 159–162.
 Hoops S., Sahle S., Gauges R. *et al.* COPASI – a COMPLEX PATHWAY SIMULATOR // Bioinformatics. 2006. V. 22. P. 3067–3074.
 Hucka M., Finney A., Sauro H. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models // Bioinformatics. 2003. V. 19. P. 524.
 Karr J., Sanghvi J., Macklin D. *et al.* A whole-cell computational model predicts phenotype from genotype // Cell. 2012. V. 150. I. 2. P. 248–250.
 Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D. *et al.* MGSmodeller – a computer system for reconstruction, calculation and analysis of mathematical models of molecular-genetic system // Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008). Novosibirsk, 22–28 June 2008. Novosibirsk: Inst. Cytol. Genet., 2008. P. 113.
 Mironova V.V., Novoselova E.S., Doroshkov A.V. *et al.* Combined *in silico/in vivo* analysis of mechanisms providing for root apical meristem self-organization and maintenance // Annals Bot. 2012. V. 110. I. 2. P. 349–360. doi:10.1093/aob/mcs069
 Olivier B.G., Snoep J.L. Web-based kinetic modelling using JWS Online // Bioinformatics. 2004. V. 20. P. 2143–2144.
 Sauro H.M., Hucka M., Finney A. *et al.* Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration // OMICS. 2003. V. 7. Issue. 4. P. 355–372.
 Shapiro B.E., Levchenko A., Meyerowitz E.M. *et al.* Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations // Bioinformatics. 2002. V. 19. I. 5. P. 677–678.

NEW FACILITIES OF THE MGSmodeller

F.V. Kazantsev¹, I.R. Akberdin¹, N.L. Podkolodny¹, V.A. Likhoshvai^{1,2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, e-mail: kazfdr@bionet.nsc.ru;

² Novosibirsk National Research State University, Novosibirsk, Russia

Summary

Mathematical modeling and analysis of complex molecular-genetic systems (MGS) are the key challenges in the systems biology era. To solve this task the special technologies and programming approaches considering the MGS as an ensemble of dynamic interconnected subsystems with a more simple structure are necessary to be developed. We have presented the approach that is aimed at acceleration of reconstruction of the complex MGS mathematical models and complex analysis using high performance computation techniques.

Key words: molecular-genetic systems, mathematical models, MGSmodeller, high-throughput computation.