

СИСТЕМА АВТОМАТИЗИРОВАННОЙ ГЕНЕРАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ГЕННЫХ СЕТЕЙ

Ф.В. Казанцев¹, И.Р. Акбердин¹, К.Д. Безматерных¹, В.А. Лихошвай^{1,2}

¹ Учреждение Российской академии наук Институт цитологии и генетики
Сибирского отделения РАН, Новосибирск, Россия, e-mail: kazfdr@bionet.nsc.ru;

² Новосибирский государственный университет, Новосибирск, Россия

Одной из актуальных проблем системной биологии является изучение законов функционирования генных сетей как динамических систем, обеспечивающих формирование фенотипических, морфологических и иных характеристик живых организмов. Для решения этой проблемы необходимо разрабатывать средства компьютерного анализа динамики функционирования генных сетей с учетом их структурно-функциональной организации. В работе мы представляем систему MGSgenerator, которая, опираясь на информацию о строении генной сети, автоматически генерирует математическую модель в терминах биохимических реакций и обобщенных функций Хилла. Входной информацией для представленной системы служит база данных GeneNet. Готовая модель может быть экспортирована в форматы SiBML, SBML и Step+ для проведения численного анализа.

Ключевые слова: генные сети, математические модели, автоматическая генерация, обобщенные функции Хилла.

Введение

Одним из центральных вопросов системной биологии является выяснение молекулярно-генетических механизмов функционирования живых систем. Для изучения их динамических характеристик все шире применяются методы математического и компьютерного моделирования. Математические модели строятся на основании информации о структурно-функциональной организации генных сетей (ГС) и их динамических свойствах, которая сосредоточена по сотням и тысячам научных публикаций. В связи с этим получили развитие компьютерные технологии, позволяющие накапливать экспериментальные данные в специализированных базах данных: KEGG (Kanehisa, Goto, 2000), Ecosys (Keseler *et al.*, 2005), Sabio-RK (Wittig *et al.*, 2006), Biomodels (Novère *et al.*, 2006). Это позволило стандартизировать хранение и представление многих типов экспериментальных данных и разрабатывать методы их системного анализа. Одной из таких компьютерных технологий является система GeneNet,

ориентированная на реконструкцию генных сетей (Ananko *et al.*, 2005). Данная система позволяет аккумулировать информацию из научных публикаций о структурно-функциональной организации молекулярно-генетических систем и представлять ее пользователю в форме графических диаграмм. Представленная в GeneNet информация отражает статическую часть сведений о генных сетях. Однако генные сети являются динамическими системами. Их состояние меняется с течением времени, и эти изменения во многом определяются структурно-функциональной организацией генных сетей. Эффективным инструментом для установления взаимосвязей между структурно-функциональной организацией генных сетей и их динамическими свойствами является метод математического моделирования. Разработка систем автоматической генерации математических моделей на основе информации, хранящейся в базах по структурно-функциональной организации генных сетей, является актуальной проблемой системной биологии. В 2008 г. Draeger с соавторами разработали одну из

первых таких систем – SBMLsqueezer (Draeger *et al.*, 2008). Данная система осуществляет автоматическую генерацию математических моделей генных сетей, реконструированных в среде CellDesigner (Funahashi *et al.*, 2003), используя стандарт SBML. Текущая версия стандарта SBML не позволяет хранить полную информацию о структурно-функциональной организации ГС (тип объекта, вид регуляторного взаимодействия и др.), что является необходимым условием для автоматического построения адекватной математической модели. В результате SBMLsqueezer настроен на работу только с ГС, реконструированными в системе CellDesigner.

В настоящей работе мы представляем программу MGSgenerator, которая автоматизирует процесс построения модели в терминах обыкновенных дифференциальных уравнений на основе исходной (первичной) информации о структурно-функциональной организации генной сети, хранящейся в базе данных GeneNet. Программа MGSgenerator анализирует элементарные подсистемы генной сети, генерирует их математические модели и на их основе с учетом структурной организации генной сети строит общую математическую модель. Модель может быть экспортирована в форматы SiBML (Лихошвай и др., 2001), SBML (Hucka *et al.*, 2003) и Step+ (Fadeev *et al.*, 2006) для проведения численного анализа.

Материалы и методы

Конструирование моделей проводится на основе информации о строении генных сетей, накапливаемой в базе данных GeneNet (<http://wwwmgs.bionet.nsc.ru/mgs/gnw>). Входная информация подается в виде XML-файла, автоматически генерируемого для каждой генной сети средствами GeneNet. Для конструирования моделей используется представление в GeneNet генных сетей в виде совокупности элементарных событий (подсистем), связанных между собой входами и выходами. Элементарные подсистемы в GeneNet описывают молекулярно-генетические (транскрипция, сплайсинг, транспорт, трансляция и т. д.) и биохимические (ферментативные реакции, фосфорилирова-

ние, формирование комплексов белков и т. д.) процессы, протекающие в различных компартаментах организма. Для идентификации модели, соответствующей элементарному событию в генной сети, поданной на вход, производится анализ наборов входных и выходных реагентов, а также наборов регуляторных взаимодействий, которые описаны в стандарте базы GeneNet.

Математические модели генных сетей генерируются на основе обобщенного химико-кинетического подхода (Лихошвай и др., 2001).

На первом шаге каждой элементарной подсистеме S_p , $I = 1, \dots, n$ ставится в соответствие собственная математическая модель – система обыкновенных дифференциальных уравнений, которая описывает закон изменения концентраций управляемых веществ в зависимости от текущих концентраций управляющих веществ и параметров системы:

$$\frac{dX_i}{dt} = V_i(Y_i, K_i), \quad (1)$$

где X_i – вектор управляемых веществ, Y_i – вектор управляющих веществ, K_i – вектор параметров.

Вид функций V_i конструируется в терминах обобщенных функций Хилла на основе информации о структурно-функциональной организации элементарной подсистемы, представленной в системе GeneNet. Обобщенные функции Хилла являются естественным обобщением стационарных биохимических моделей, поэтому структура ОФХ отражает механизм, а параметры – кинетические характеристики участников описываемого процесса. Более того, метод ОФХ эффективен при недостатке кинетических экспериментальных данных и может быть применен при построении математических моделей молекулярно-генетических систем разных уровней детализации (Likhoshvai, Ratushny, 2007). Конструирование функций V_i проводится на основе оригинального алгоритма, разработанного в рамках настоящей работы.

Генерация модели M из моделей элементарных подсистем осуществляется по правилу суммирования скоростей протекания элементарных процессов, которое формулируется исходя из биохимической природы молекулярно-генетических процессов (Лихошвай и др., 2001):

$$\begin{array}{c}
 S_1 : \frac{dX_1}{dt} = V_1(Y_1, K_1), \quad S_2 : \frac{dX_2}{dt} = V_2(Y_2, K_2), \quad \dots \quad S_n : \frac{dX_n}{dt} = V_n(Y_n, K_n) \\
 \swarrow \quad \downarrow \quad \dots \quad \searrow \\
 M : \begin{cases} \frac{dX}{dt} = V(Y, K), & V(Y, K) = \sum_{i=1}^n V_i(Y_i, K_i), \\ X = \cup X_i, Y = \cup Y_i, K = \cup K_i. \end{cases}
 \end{array}$$

Рис. 1. Правило суммирования скоростей по всем непрерывным моделям элементарных подсистем Si.

Результаты и обсуждение

Представленная в настоящей работе программа MGSgenerator использует в качестве входных данных структурно-функциональную организацию генных сетей, реконструированных в системе GeneNet. В настоящей версии программы алгоритм автоматической генерации математической модели реализован в интерактивном варианте, так как на определенных этапах требуется принятие экспертного решения. Связано это с тем, что алгоритм реализован для ограниченного набора элементарных подсистем, который не исчерпывает возможного их разнообразия. Каждую такую подсистему программа автоматически распознает в составе генной сети и генерирует для нее математическую модель. Правила генерации описаны в следующих разделах статьи. Если в генной сети встречается подсистема, для которой построение элементарной модели невозможно, то выдается сообщение об этом. В этом случае пользователь может либо модифицировать структуру генной сети и подать ее на вход программе MGSgenerator повторно, либо вручную добавить в конструируемую модель математическую модель элементарной подсистемы.

Сопоставление математических моделей элементарным подсистемам. В MGSgenerator генная сеть рассматривается как совокупность взаимодействующих элементарных подсистем S_1, \dots, S_n . В представленной версии программы элементарными подсистемами являются элементарные события (реакции), описанные в стандарте базы данных GeneNet. Сопоставле-

ние определенной математической модели для каждой выделенной элементарной подсистемы осуществляется на основе анализа наборов входных и выходных реагентов и наборов управляющих воздействий. Данная информация экстрагируется из стандарта представления реакций базы данных GeneNet, и на ее основе определяются тип подсистемы, а также структура модели.

В текущей версии программы MGSgenerator генерация математических моделей поддерживается для следующих типов элементарных подсистем ГС: репликация, транскрипция, обратная транскрипция, трансляция, ферментативный синтез, транспорт, мультимеризация, сплайсинг, фосфорилирование, ацетилирование и метилирование, а также подсистем притока веществ и их деградации.

Для генерации математических моделей элементарных подсистем всех типов используется формула общего вида, состоящая из 4 сомножителей:

$$\begin{aligned}
 V_i(Y_i, K_i) &= \\
 &= k \cdot V_{i1}(Y_{i1}) \cdot V_{i2}(Y_{i2}, K_{i2}) \cdot V_{i3}(Y_{i3}, K_{i3}) \\
 Y_i &= Y_{i1} \cup Y_{i2} \cup Y_{i3} \\
 K_i &= \{k\} \cup K_{i2} \cup K_{i3}.
 \end{aligned} \tag{2}$$

Первый сомножитель является константой скорости синтеза и имеет размерность, обратную времени. Второй сомножитель V_{i1} соответствует входному веществу, которое определяет тип подсистемы и имеет размерность концентрации. Сомножитель V_{i1} может принимать одно из двух значений, в зависимости от типа подсистемы:

$$V_{i1}(Y_{i1}) = \begin{cases} 1, & Y_{i1} = \emptyset, \text{ если процесс – приток вещества,} \\ y, & Y_{i1} = \{y\}, \text{ иначе.} \end{cases} \tag{3}$$

Здесь y – концентрация ключевого вещества подсистемы. В случае транскрипции или репликации – это ген, в случае трансляции или сплайсинга – РНК, а в случае ферментативного синтеза – фермент. Третий (V_{i2}) и четвертый (V_{i3}) сомножители являются безразмерными. Сомножитель V_{i2} описывает влияние регуляторов на скорость протекания рассматриваемого процесса. В слу-

чае отсутствия последних множитель V_{i2} полагается равным 1. Если же список регуляторов не пуст, то вид данного сомножителя зависит от их типа. Различаются регуляторы, выполняющие функцию активаторов (a_j), и регуляторы, выполняющие функцию ингибиторов (r_l). В соответствии с конкретной спецификацией выписывается формула вида:

$$V_{i2}(Y_{i2}, K_{i2}) = \left(\delta + \sum_j \left(\frac{a_j}{k_{1,a,j}} \right)^{h_{1,a,j}} \right) / \left(1 + \sum_j \left(\frac{a_j}{k_{2,a,j}} \right)^{h_{2,a,j}} + \sum_l \left(\frac{r_l}{k_{r,l}} \right)^{h_{r,l}} \right)$$

$$Y_{i2} = (\{a_j\}_{j=1,\dots,n}, \{r_l\}_{l=1,\dots,m}) \quad (4)$$

$$K_{i2} = (\delta, \{k_{1,a,j}, k_{2,a,j}, h_{1,a,j}, h_{2,a,j}\}_{j=1,\dots,n}, \{k_{r,l}, h_{r,l}\}_{l=1,\dots,m}),$$

где k_{1a} , k_{2a} – константы активатора, k_r – константа ингибитора, h_{1a} , h_{2a} , h_r – коэффициенты Хилла.

Множитель V_{i3} описывает зависимость скорости реакции от концентраций субстратов и продуктов. В случае таких подсистем, как репликация, транскрипция, трансляция, сплайсинг или транспорт, полагается, что $V_{i3} = 1$.

Для подсистем, имеющих тип «ферментативный синтез», «фосфорилирование», «ацетилирование» и «метилирование», вид функции V_{i3} зависит от концентрации субстратов и продуктов и определяется структурой входного

и выходного списков. Тип распознается по наличию во входном списке сущности «фермент», который осуществляет выбранный процесс. Для подсистем фосфорилирования, метилирования, ацетилирования (в текущей версии системы эти процессы объединены в один класс и не различаются друг от друга, так как нет достаточной информации во входных данных для идентификации) входной список содержит сущность «белок» (s), который подвергается модификации. Сущность, которая отвечает модифицированной форме белка, входит в выходной список (p). В этом случае сомножитель V_{i3} имеет вид:

$$V_{i3}(Y_{i3}, K_{i3}) = \frac{k_1 \cdot x \cdot \frac{s}{k_{ms}} - k_{-1} \cdot x \cdot \frac{p}{k_{mp}}}{1 + \frac{s}{k_{ms}} + \frac{p}{k_{mp}}} \quad (5)$$

$$Y_{i3} = \{p, s, x\}; K_{i3} = \{k_1, k_{-1}, k_{ms}, k_{mp}\},$$

где k_1 – константа прямой реакции, k_{-1} – константа обратной реакции, k_{ms} и k_{mp} – константы Михаэлиса. Для подсистем данного типа по умолчанию полагается, что субстрат и продукт конкурируют между собой за активный сайт фермента.

Для подсистем типа «ферментативный синтез» рассматривается несколько вариантов. Пусть входной список содержит n субстратов (s_j , $j = 1, \dots, n$), из которых синтезируется один продукт, который содержится в выходном списке. Тогда V_{i3} имеет вид:

$$V_{i3}(Y_{i3}, K_{i3}) \equiv \frac{\prod_{j=1}^n \frac{s_j}{k_{ms,j}} - \omega_{p,s} \frac{p}{k_{mp}}}{\prod_{j=1}^n \left(1 + \frac{s_j}{k_{ms,j}} \right) + \frac{p}{k_{mp}}} \quad (6)$$

$$Y_{i3} = \{p, s_j, j = 1, \dots, n\}; K_{i3} = \{\omega_{p,s}, k_{mp}, k_{ms,j}, j = 1, \dots, n\},$$

где k_{ms} и k_{mp} – константы Михаэлиса, $\omega_{p,s}$ – отношение константы прямой реакции к константе обратной реакции. Для подсистем данного подтипа по умолчанию полагается, что субстраты могут независимо связываться с фер-

ментом, но при этом конкурируют с продуктом.

Пусть входной список содержит n субстратов ($s_j, j = 1, \dots, n$), из которых синтезируется n продуктов ($p_j, j = 1, \dots, n$). Тогда применяется формула:

$$V_{i3}(Y_{i3}, K_{i3}) = \frac{k_s \cdot x \cdot \prod_j \frac{s_j}{k_{ms,j}} - k_p \cdot x \cdot \prod_j \frac{p_j}{k_{mp,j}}}{\prod_i \left(1 + \frac{s_j}{k_{ms,j}} + \frac{p_j}{k_{mp,j}} \right)} \quad (7)$$

$$Y_{i3} = \{p_j, s_j, x, j = 1, \dots, n\}; K_{i3} = \{k_s, k_p, k_{ms,j}, k_{mp,j}, j = 1, \dots, n\},$$

где k_s – константа прямой реакции, k_p – константа обратной реакции, k_{ms} и k_{mp} – константы Михаэлиса. В формуле (7) реализовано предположение, что j -й субстрат конкурирует с j -м продуктом за связывание с ферментом, а пары субстрат–продукт не конкурируют между собой.

Пример конструирования модели генной сети. Рассмотрим генную сеть, содержащую ген *sup79b3*, кодирующий транскрипционный фактор, который является автоингибитором инициации транскрипции (рис. 2).

Данная генная сеть содержит 4 элементарные подсистемы: подсистему синтеза и транспорта мРНК из ядра в цитоплазму, подсистему синтеза и доставки белка из цитоплазмы в ядро, подсистемы деградации мРНК и белка (подсистемы деградации не отображаются в

структуре генной сети). В результате разбора информации система MGSgenerator генерирует математическую модель, содержащую два дифференциальных уравнения:

$$\begin{cases} \frac{dr}{dt} = k_r \cdot g \cdot \frac{k_{ba}}{1 + \left(\frac{p}{k_i}\right)^{k_g}} - k_{dr} \cdot r \\ \frac{dp}{dt} = k_p \cdot r - k_{dp} \cdot p \end{cases}, \quad (8)$$

где g – концентрация гена *sup79b3*, r – концентрация мРНК *sup79b3*, p – концентрация белка Cyp79b3, k_r и k_p – константы активации, k_{ba} – константа базальной активности, k_{dr} и k_{dp} – константы деградации, k_i – константа ингибирования, k_g – Хилловский коэффициент.

Программа MGSgenerator была использована для генерации математической модели, контролирующей метаболизм ауксина в меристеме побега *Arabidopsis thaliana* L., и показала свою эффективность при решении поставленных теоретических задач (Акбердин и др., 2009).

Расширяемость программы MGSgenerator. Программа MGSgenerator реализована на языке программирования Java, что обеспечивает ее переносимость на различные платформы. Программа поддерживает две точки расширения функциональности: G – точку подключения источников данных и Σ – точку подключения модулей экспорта готовой математической модели, что позволяет сформировать модульную компьютерную систему моделирования и численного анализа динамики функционирования генных сетей (рис. 3).

В качестве источника данных в представленной компьютерной модульной системе для

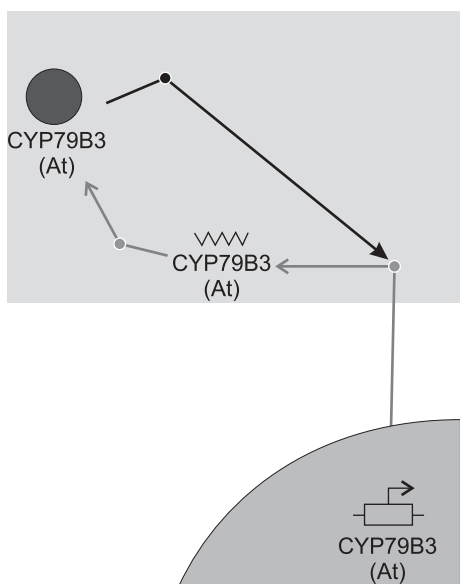


Рис. 2. Структура генной сети процесса автоингибирования.

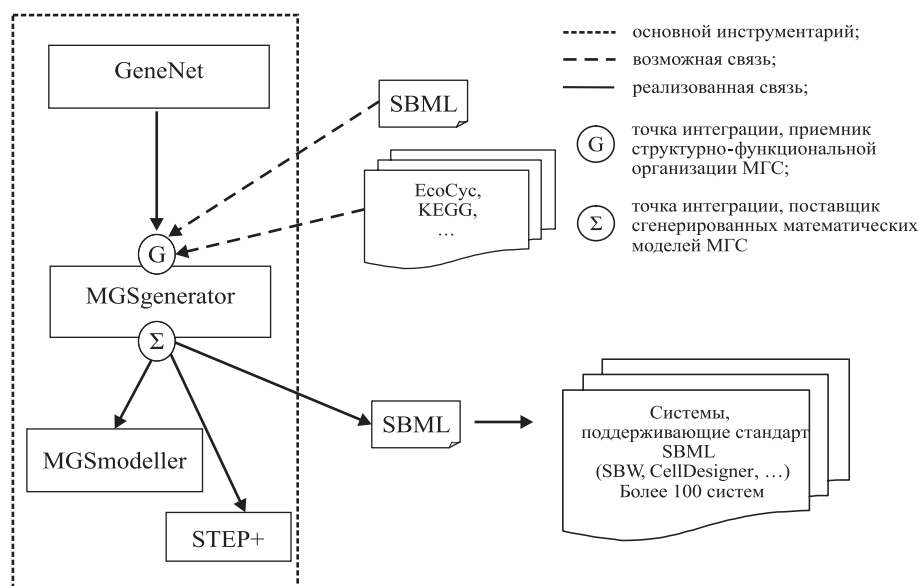


Рис. 3. Схема организации компьютерной модульной системы моделирования на основе программы MGSgenerator.

настоящей версии программы MGSgenerator служит база GeneNet, которая предоставляет информацию о структурно-функциональной организации ГС. В будущем функциональность системы может быть расширена другими источниками данных, в частности, подключением баз KEGG и ЭкоСус. В качестве форматов представления математической модели в настоящей версии программы MGSgenerator используются: формат SiBML системы MGSmodeller (Kazantsev *et al.*, 2008), формат системы STEP+ (Fadeev *et al.*, 2006) и формат SBML (Hucka *et al.*, 2003). В случае необходимости функциональность по представлению математической модели также может быть расширена.

Заключение

Представленная в работе программа MGSgenerator является ключевым элементом при разработке эффективных компьютерных модульных систем, которые в перспективе способны обеспечить полный технологический цикл изучения закономерностей строения и функционирования генных сетей с применением методов математического моделирования. Предложенная версия генератора математических моделей генных сетей является начальным этапом в разработке автоматизированных систем генерации математических моделей. Включенные в нее математи-

ческие модели покрывают лишь незначительную часть возможных механизмов. В дальнейшем планируется существенно расширить базу шаблонных моделей элементарных подсистем, а также возможности представления математических моделей ГС в других форматах.

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.
- 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.
- Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. V. 33. P. D425–D427.
- Draeger A., Hassis N., Supper J. *et al.* SBMLsqueezer: a CellDesigner plug-in to generate kinetic rate equations for biochemical networks // BMC Systems Biol. 2008. V. 2.
- Fadeev S.I., Korolev V.K., Gainova I.A., Medvedev A.E. The package Step+ for numerical study of autonomous systems arising when modeling dynamics of genetic-molecular systems // Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2006). Novosibirsk, 16–22 July 2006. Novosibirsk: Inst. Cytol. Genet., 2006. V. 2. P. 118–120.

- Funahashi A., Morohashi M., Kitano M., Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks // BIOSILICO. 2003. V. 1. P. 159–162.
- 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.
- 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.
- Kanehisa M., Goto S. KEGG: Kyoto encyclopedia of genes and genomes // Nucl. Acids Res. 2000. V. 28. P. 27–30.
- Keseler I., Collado-Vides J., Gama-Castro S. *et al.* EcoCyc: a comprehensive database resource for *Escherichia coli* // Nucl. Acids Res. Database issue. 2005. V. 33. P. D334–D337.
- Likhoshvai V., Ratushny A. Generalized Hill function method for modeling molecular processes // J. Bioinformatics and Comp. Biol. 2007. V. 5. № 2. P. 521–531.
- Novère N., Bornstein B., Broicher A. *et al.* BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems // Nucl. Acids Res. 2006. № 34. P. D689–D691.
- Wittig U., Golebiewski M., Kania R. *et al.* SABIO-RK: Integration and Curation of Reaction Kinetics Data // Proc. of the 3rd Intern. Workshop on Data Integration in the Life Sciences 2006 (DILS'06). Hinxton, UK. Lecture Notes in Bioinformatics. P. 94–103.

THE TOOL FOR AUTOMATIC GENERATION OF GENE NETWORKS MATHEMATICAL MODELS

F.V. Kazantsev¹, I.R. Akberdin¹, K.D. Bezmaternykh¹, V.A. Likhoshvai^{1,2}

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

²Novosibirsk State University, Novosibirsk, Russia

Summary

Nowadays there are a plenty of reconstructed gene networks that describe different systems in live organism critical for forming phenotypical, morphological and others properties of individuals. The analysis of gene networks in terms of dynamical systems is one of the key challenges of system biology. To facilitate this analysis we have proposed MGSgenerator system that provides automatic generation of mathematical models on the basis of gene networks structure. The mathematical models are generated in biochemical and generalized Hill function terms. The GeneNet database is used as data source of gene network structures. The generated model could be exported in SiBML, SBML and Step+ formats for further analysis.