

DYMICROBIOTECH: A SOFTWARE MODULE FOR AUTOMATIC RECONSTRUCTION OF FRAME-BASED DYNAMIC MODELS OF MICROBIAL GENE NETWORKS

S. Lashin, F. Kazantsev, T. Lakhova, Yu. Matushkin

Kurchatov Genomic Center of the Institute of Cytology and Genetics SB RAS,
630090, Novosibirsk, Russia
Institute of Cytology and Genetics SB RAS,
630090, Novosibirsk, Russia
Novosibirsk State University,
630090, Novosibirsk, Russia

DOI: 10.24412/2073-0667-2024-4-56-68

EDN: QVDKVVH

Modern genetic technologies are used in industrial biotechnology to design microbial strains-producers with target characteristics for based on close integration of experimental and information-computer approaches. The increasing availability of genomic data and methods of their functional annotation requires the development of new methods of systems biology, in particular, methods of reconstruction of gene networks and metabolic pathways controlling target processes and characteristics of microorganisms based on information about sequenced genomes, as well as methods of building mathematical models of these networks and pathways. This paper presents the DynMicrobiotech software module for automatic reconstruction of frame-based mathematical models based on the generalized chemical-kinetic modeling method. The input data for the module are the annotation and markup of the genome, while the output data are the generated model in the form of a system of ordinary differential equations written in SBML format.

Key words: generalized chemical-kinetic method of modeling, differential equations, gene networks.

References

1. GOODWIN S., MCPHERSON J. D., MCCOMBIE W. R. Coming of age: ten years of next-generation sequencing technologies // *Nat. Rev. Genet.* 2016. V. 17. N 6. P. 333–351.
2. QUAIL M. ET AL. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers // *BMC Genomics.* 2012. V. 13. N 1. P. 341.
3. COWAN A. E., MENDES P., BLINOV M. L. ModelBricks—modules for reproducible modeling improving model annotation and provenance // *npj Syst. Biol. Appl.* 2019. V. 5. N 1.
4. GILBERT D. ET AL. Towards dynamic genome-scale models // *Brief. Bioinform.* 2019. V. 20. N 4. P. 1167–1180.
5. KARR J. R. ET AL. A Whole-Cell Computational Model Predicts Phenotype from Genotype // *Cell.* 2012. V. 150. N 2. P. 389–401.

The study is supported by the budget project of Institute of Cytology and Genetics SB RAS No FWNR-2022-0020.

6. KIM W. J., KIM H. U., LEE S. Y. Current state and applications of microbial genome-scale metabolic models // *Curr. Opin. Syst. Biol.* 2017. V. 2. P. 10–18.
7. AKBERDIN I. R. ET AL. In Silico Cell: Challenges and Perspectives // *Math. Biol. Bioinforma.* 2013. V. 8. N 1.
8. DEMIN O., GORYANIN I. *Kinetic Modelling in Systems Biology.* 2008.
9. HELLERSTEIN J. L. ET AL. Recent advances in biomedical simulations: a manifesto for model engineering // *F1000Research.* 2019. V. 8. P. 261.
10. OCONE A., MILLAR A. J., SANGUINETTI G. Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics // *Bioinformatics.* 2013. V. 29. N 7. P. 910–916.
11. FUNAHASHI A. ET AL. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks // *Proc. IEEE.* 2008. V. 96. N 8. P. 1254–1265.
12. KING Z. A. ET AL. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models // *Nucleic Acids Res.* 2016. V. 44. N D1. P. D515–D522.
13. LLOYD C. M. ET AL. The CellML Model Repository // *Bioinformatics.* 2008. V. 24. N 18. P. 2122–2123.
14. MALIK-SHERIFF R. S. ET AL. BioModels—15 years of sharing computational models in life science // *Nucleic Acids Res.* 2019.
15. HENKEL R., WOLKENHAUER O., WALTEMATH D. Combining computational models, semantic annotations and simulation experiments in a graph database // *Database.* 2015. V. 2015. P. 1–16.
16. KIRK P. D. W., BAPTIE A. C., STUMPF M. P. H. Systems biology (un)certainties // *Science (80-.)*. 2015. V. 350. N 6259. P. 386–388.
17. STANFORD N. J. ET AL. The evolution of standards and data management practices in systems biology // *Mol. Syst. Biol.* 2015. V. 11. N 12. P. 851–851.
18. BEAL J. ET AL. Communicating Structure and Function in Synthetic Biology Diagrams // *ACS Synth. Biol.* 2019. V. 8. N 8. P. 1818–1825.
19. BRUGGEMAN F. J., WESTERHOFF H. V. The nature of systems biology // *Trends Microbiol.* 2007. V. 15. N 1. P. 45–50.
20. LIKHOSHVAI V. A. ET AL. Generalized chemokinetic method for gene network simulation // *Mol. Biol.* 2001. V. 35. N 6. P. 919–925.
21. PALSSON B. The challenges of in silico biology Moving from a reductionist paradigm to one that views cells as systems will necessitate // 2000. V. 18. P. 1147–1150.
22. KURATA H. ET AL. BioFNet: Biological functional network database for analysis and synthesis of biological systems // *Brief. Bioinform.* 2013. V. 15. N 5. P. 699–709.
23. RATNER V. A. A molecular genetic control system // *Priroda = Nature.* 2001. V. 3. P. 16–22.
24. MOODIE S. ET AL. Systems Biology Graphical Notation: Process Description language Level 1 Version 1.3 // *J. Integr. Bioinform.* 2015. V. 12. N 2.
25. NORSIGIAN C. J. ET AL. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree // *Nucleic Acids Res.* 2019. V. 48. N D1. P. D402–D406.
26. ZHANG F. ET AL. Systems biology markup language (SBML) level 3 package: multistate, multicomponent and multicompartments species, version 1, release 2 // *J. Integr. Bioinform.* 2020. V. 17. N 2–3. P. 0–74.
27. LIKHOSHVAI V. A. ET AL. A generalized chemical-kinetic method for modeling complex biological systems. Computer model of bacteriophage Lambda ontogenesis // *Vychislitelnye texnologii = Journal of Computational Technologies.* 2000. V. 5. N Special issue dedicated to the 10th anniversary of the Laboratory of Theoretical Genetics of the Institute of Cytology and Genetics SB RAS. P. 87–99.
28. KAZANTSEV F. V. ET AL. System of automated generation of mathematical models of gene networks // *Informatsionnyj vestnik VOGIS.* 2009. V. 13. N 1. P. 163–169.

29. AKBERDIN I. R. ET AL. “Electronic cell”: problems and perspectives // *Mathematicheskaya Biologiya i Bioinformatika Mathematical Biology and Bioinformatics*. 2013. V. 8. N 1. P. 287–307.
30. ZHABOTINSKY A. M. Concentration auto oscillations., 1974. P. 1–179. M.: Nauka.
31. HUCKA M. ET AL. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core // *J. Integr. Bioinforma.* 2015. V. 12. N 2. P. 382–549.
32. LAKHOVA T. N. ET AL. Algorithm for the Reconstruction of Mathematical Frame Models of Bacterial Transcription Regulation // *Mathematics*. 2022. V. 10. N 23. P. 4480.
33. KANEHISA M. Enzyme Annotation and Metabolic Reconstruction Using KEGG., 2017. P. 135–145.
34. McDONALD A. G., BOYCE S., TIPTON K. F. ExplorEnz: the primary source of the IUBMB enzyme list // *Nucleic Acids Res.* 2009. V. 37. N Database. P. D593–D597.
35. WITTIG U. ET AL. SABIO-RK–database for biochemical reaction kinetics // *Nucleic Acids Res.* 2012. V. 40. N D1. P. D790–D796.
36. KAZANTSEV F. V. ET AL. MAMMOTH: A new database for curated mathematical models of biomolecular systems // *J. Bioinform. Comput. Biol.* 2018. V. 16. N 01. P. 1740010 (16 pages).
37. OTASEK D. ET AL. Cytoscape Automation: empowering workflow-based network analysis // *Genome Biol.* 2019. V. 20. N 1. P. 185.
38. HOOPS S. ET AL. COPASI—a COMplex PATHway SIMulator // *Bioinformatics*. 2006. V. 22. N 24. P. 3067–3074.
39. COCK P. J. A. ET AL. Biopython: freely available Python tools for computational molecular biology and bioinformatics // *Bioinformatics*. 2009. V. 25. N 11. P. 1422–1423.

DYNMICROBIOTECH: ПРОГРАММНЫЙ МОДУЛЬ ДЛЯ АВТОМАТИЧЕСКОЙ РЕКОНСТРУКЦИИ ФРЕЙМОВЫХ ДИНАМИЧЕСКИХ МОДЕЛЕЙ ГЕННЫХ СЕТЕЙ МИКРООРГАНИЗМОВ

С. А. Лашин, Ф. В. Казанцев, Т. Н. Лахова, Ю. Г. Матушкин

Курчатовский геномный центр Института цитологии и генетики СО РАН,
630090, Новосибирск, Россия
Институт цитологии и генетики СО РАН,
630090, Новосибирск, Россия
Новосибирский государственный университет,
630090, Новосибирск, Россия

УДК 573.22, 57.052

DOI: 10.24412/2073-0667-2024-4-56-68

EDN: QVDKVN

Современные генетические технологии используются в промышленной биотехнологии для конструирования микробиологических штаммов-продуцентов с целевыми характеристиками на основе тесной интеграции экспериментальных и информационно-компьютерных подходов. Все большая доступность геномных данных и методов их функциональной аннотации требует разработки новых методов системной биологии, в частности, методов реконструкции генных сетей и метаболических путей, контролирующих целевые процессы и характеристики микроорганизмов, на основе информации о секвенированных геномах, а также методов построения математических моделей этих сетей и путей. В данной работе представлен программный модуль DynMicrobiotech для автоматической реконструкции фреймовых математических моделей на основе метода обобщенного химико-кинетического моделирования. Входными данными для модуля являются аннотация и разметка генома, выходными — сгенерированная модель в виде системы обыкновенных дифференциальных уравнений, записанная в формате SBML.

Ключевые слова: обобщенный химико-кинетический метод моделирования, дифференциальные уравнения, генные сети.

Введение. Современные генетические технологии используются в промышленной биотехнологии для конструирования микробиологических штаммов-продуцентов с целевыми характеристиками на основе тесной интеграции экспериментальных и информационно-компьютерных подходов.

Прогресс в развитии экспериментальных технологий совершил революцию в биологии и позволил исследователям генерировать огромное количество многофакторных данных с высокой степенью детализации [1, 2]. Развитие и доступность высокопроизводительных

Исследование выполнено за счет средств бюджетного проекта ФИЦ ИЦиГ СО РАН «Системная биология и биоинформатика: реконструкция, анализ и моделирование структурно-функциональной организации и эволюции генных сетей человека, животных, растений и микроорганизмов» № FWNR-2022-0020.

вычислительных ресурсов и технологий хранения данных дали новый толчок в развитии математического моделирования биологических систем [3–6]. Сегодня математическое моделирование все чаще рассматривается как основной инструмент для интеграции и анализа экспериментальных данных и системного исследования динамических биологических систем [7–9]. Математические модели динамических биологических систем могут быть сформулированы в терминах систем обыкновенных дифференциальных уравнений (ОДУ), уравнений в частных производных или дифференциальных уравнений с запаздывающим аргументом, стохастических моделей, а также в терминах дискретных операций и, наконец, сочетая перечисленные выше техники, в терминах т. н. гибридных подходов [5, 10]. Выбор подходящего формализма обычно зависит от сложности рассматриваемой биологической системы в смысле точности и доступности экспериментальных данных.

На текущий момент доступно более 1000 различных опубликованных моделей комплексных биомолекулярных систем, которые разработаны для нескольких организмов [6, 11–14]. Количество таких моделей постоянно растет, однако повторное использование, расширение или модификация таких готовых математических моделей довольно нетривиальная задача [15–17]. Модель часто представлена в виде законченной, жестко структурированной в некотором формализме системы (например, в виде системы уравнений или кода программы на языке высокого уровня). Для таких моделей отсутствуют средства, позволяющие модифицировать их структурное и/или функциональное содержание в автоматическом или полуавтоматическом режиме. Однако, декомпозиция моделей на отдельные «элементарные» подсистемы является альтернативным решением озвученной проблемы. Такие подсистемы могут рассматриваться как «строительные блоки», которые используются для описания/представления конкретных биомолекулярных функций и/или элементарных процессов биологических систем [3, 18–21]. *Моделью элементарной подсистемы* (МЭП) мы будем называть минимальную модель подсистемы, которой достаточно для осуществления заданной функции [22–23].

Подход представления моделей в виде комбинации *элементарных подсистем* является естественным в силу самой природы организации биологических систем. Имеются естественные уровни иерархии в организации организмов — естественное разбиение на функциональные подсистемы. Именно такой подход является перспективным в создании комплексных многоуровневых моделей систем живой природы, поддерживающий методы размещения подсистем по компартментам, организации модельных структур в органеллы и ткани. Такой подход в декомпозиции, каталогизации, интеграции, реконструкции и исследовании моделей приводит к вовлечению в работу современных информационно-вычислительных технологий по хранению и доступу к данным (реляционным, объектным и NO-SQL СУБД) и технологий высокопроизводительного вычислительного анализа. Практический интерес к такому подходу обусловлен тем, что с его помощью возможно формировать коллекцию моделей элементарных подсистем, играющих роль «строительных блоков» для последующего использования в новых составных моделях, а также разрабатывать инструменты, которые становятся стандартами в обмене информацией, инструментах анализа и представления данных [24–26].

В данной работе мы описываем программный модуль DynMicrobiotech для автоматической реконструкции фреймовых математических моделей на основе обобщенного метода химико-кинетического моделирования [27]. DynMicrobiotech представляет развитие MGSGenerator, ранее разработанного авторами [28]. В основе подхода лежат декомпозиция графа геной сети на подсистемы и независимая генерация формального математическо-

го описания для каждого процесса, представленного в геной сети. Входными данными для модуля являются аннотация и разметка генома, выходными — сгенерированная модель в виде системы обыкновенных дифференциальных уравнений, записанная в формате SBML.

1. Алгоритм реконструкции математических моделей геной сети. Реконструкция математических моделей на основе информации из аннотированного генома проводится в несколько этапов. В рамках процесса реконструкции естественным подходом организации моделей является блочно-модульный подход моделирования биологических систем [20, 22]. Одним из эффективных методов, реализующих этот подход, является обобщенный химико-кинетический метод моделирования (ОХКММ), когда модель исследуемого явления представляется комбинацией из самостоятельных моделей ее подсистем. Такими подсистемами могут быть метаболические пути, их комбинации или отдельные процессы (ферментативный синтез, транспортные процессы молекул и др.).

Модель молекулярно-генетической системы (МГС) строится на основе уравнения химической кинетики вида (1) и представляет собой различные способы представления *кинетического закона действующих масс* и обобщенного химико-кинетического метода моделирования (ОХКММ), предложенного В. А. Лихошваем [27]:

$$\frac{dX}{dt} = V(Y, K), \quad (1)$$

где X — вектор (список) управляемых переменных, Y — вектор (список) управляющих переменных, K — список параметров.

Допускается вхождение одних и тех же переменных в оба списка, но, в общем случае, списки X и Y не совпадают и могут вообще не пересекаться. Переменные обычно имеют смысл концентраций веществ или вероятностей реализации выделенных состояний веществ. Переменные из списка Y , не входящие в список X , являются параметрами для текущей элементарной модели. Функционал V описывает закон скоростей изменения концентраций веществ из списка X .

Кинетический закон действующих масс выводится на основе теории соударений. Пусть имеется биохимическая реакция:



Тогда скорость $V_{A+B \rightarrow C}$ формирования комплекса C в текущий момент времени t равна $V_{A+B \rightarrow C}(t) = k_1[A][B]$, а скорость $V_{A+B \rightarrow C}$ распада комплекса C на составляющие A и B в текущий момент времени t равна $V_{C \rightarrow A+B}(t) = k_2[C]$.

Если имеется конкретная биохимическая схема реакций вида (2), то мгновенная скорость изменения концентрации любого вещества равняется сумме локальных скоростей изменения концентрации данного вещества в каждой реакции, в которой данное вещество участвует [28–29]. Это простое правило позволяет легко выписать итоговую систему дифференциальных уравнений, описывающих целевую биохимическую схему, используя при этом в качестве правых частей уравнений системы (1) только полиномы первого и второго порядка.

Теоретической основой для этого служит важнейшая для моделирования химической кинетики теорема Корзухина, которая утверждает: «Для любой совокупности неотрицательных кривых, заданных на конечном интервале времени, и любой заданной точности,

Таблица 1

Некоторые формальные блоки ОХКММ (по: [27])

Обратимая биомолекулярная реакция: $x_1 + x_2 \rightleftharpoons x_3$	$\vec{X} = (x_1, x_2, x_3), \vec{P} = (k_1, k_2),$ $\vec{F} : \frac{dx_1}{dt} = k_2 \cdot x_3 - k_1 \cdot x_1 \cdot x_2, \frac{dx_2}{dt} = \frac{dx_3}{dt} = -\frac{dx_3}{dt}.$
Необратимая мономолекулярная реакция: $x \xrightarrow{k} y_1 + y_2 + \dots + y_n$	$\vec{X} = (x, y_1, y_2, \dots, y_n), \vec{P} = (k),$ $\vec{F} : \frac{dx}{dt} = -\frac{dy_1}{dt} = -\frac{dy_2}{dt} = \dots = -\frac{dy_n}{dt} = -k \cdot x, n \geq 0.$
Конститутивный синтез: $\xrightarrow{k} x_1 + \dots + x_n$	$\vec{X} = (x, x_1, \dots, x_n), \vec{P} = (k),$ $\vec{F} : \frac{dx_i}{dt} = k, i = 1, \dots, n, n \geq 0.$
Обобщенная схема 1 Михаэлиса-Ментен	$\vec{X} = (x_2, \dots, x_m, y_1, \dots, y_n),$ $\vec{P} = (m_1, k_i, k_d, a_1, \dots, a_m, b_1, \dots, b_n),$ $\vec{F} : \frac{dx_j}{dt} = a_j \cdot Z, j = 1, \dots, m, \frac{dy_l}{dt} = b_l \cdot Z, l = 1, \dots, n,$ где $Z = \frac{k_d k_i x_1 \dots x_m}{(k_d + x_1) \dots (k_d + x_m) - x_1 \dots x_m}, m \geq 2, n \geq 0.$
Обобщенная схема 2 Михаэлиса-Ментен	$\vec{X} = (e, x_1, \dots, x_m, y_1, \dots, y_n),$ $\vec{P} = (m, k, s_1, \dots, s_m, a_0, a_1, \dots, a_m, b_1, \dots, b_n),$ $m \geq 1, n \geq 0,$ $\vec{F} : \frac{de}{dt} = -a_0 \cdot Z, \frac{dx_j}{dt} = -a_j \cdot Z, j = 1, \dots, m,$ $\frac{dy_l}{dt} = b_l \cdot k \cdot Z, l = 1, \dots, n,$ где $Z = \frac{k s e x_1 \dots x_m}{(s+e) \dots (s_1+x_1) \dots (s_m+x_m) - e x_1 \dots x_m},$ $s = \frac{\frac{1}{x_1} + \dots + \frac{1}{x_m}}{\frac{1}{x_1} + \dots + \frac{1}{x_m}}.$
Обобщенная схема 3 Михаэлиса-Ментен	$\vec{X} = (e_1, \dots, e_l, x_1, \dots, x_m, y_1, \dots, y_n),$ $\vec{P} = (l, m, k, s_1, \dots, s_m, a_0, a_1, \dots, a_m, b_1, \dots, b_n), m \geq 1, n \geq 0,$ $\vec{F} : \frac{de}{dt} = -a_0 \cdot Z, \frac{dx_j}{dt} = -a_j \cdot Z, j = 1, \dots, m,$ $\frac{dy_l}{dt} = b_l \cdot k \cdot Z, l = 1, \dots, n,$ где $Z = \frac{k e_1 \dots e_l x_1 \dots x_m}{(s_1+x_1) \dots (s_m+x_m)}.$
«реакция» $x_1 + \dots + x_m \rightleftharpoons y_1 + \dots + y_n$	$\vec{X} = (x_1, \dots, x_m, y_1, \dots, y_n),$ $\vec{P} = (m, k_1, k_2, a_1, s_{x_1}, s_{y_1}, \dots, a_m, s_{x_m}, s_{y_m}, b_1,$ $r_{x_1}, r_{y_1}, \dots, b_n, r_{x_n}, r_{y_n}),$ $\vec{F} : \frac{dx_i}{dt} = -s_{x_i} k_1 x_1^{a_1} \dots x_m^{a_m} + s_{y_i} k_2 y_1^{b_1} \dots y_n^{b_n}, i = 1, \dots, m,$ $\frac{dy_j}{dt} = r_{x_j} k_1 x_1^{a_1} \dots x_m^{a_m} - r_{y_j} k_2 y_1^{b_1} \dots y_n^{b_n}, l = 1, \dots, n.$
Суммирование: $x = x_1 + \dots + x_n$	$\vec{X} = (x, x_1, \dots, x_n),$ $\vec{F} : \dot{x} = x_1 + \dots + x_n.$
Пороговый автомат	$\vec{X} = (x, x_1, x_2), \vec{F} : \dot{x} = x_1, \text{ если } x_1 \geq x_2.$
Стохастический автомат	$\vec{X} = (x, x_1, y_1, y_2), \vec{F} : \dot{x} = x_1, \text{ если } y_1 \geq y_2,$ где y_1 — случайная величина, равномерно распределенная на отрезке $[0,1],$ y_2 — число из интервала $[0,1].$

существует такая (может быть не одна) биохимическая схема, составленная только из биомолекулярных и мономолекулярных реакций, что математическая модель, построенная по данной биохимической схеме, приближает заданную совокупность кривых с заданной точностью» [30].

Важнейшим элементом построения более сложных моделей МГС с помощью ОХКММ [27] является *правило суммирования локальных скоростей* протекания биохимических реакций: общая скорость изменения концентраций компонентов системы является суммой скоростей изменения концентрации данного компонента во всех элементарных процессах, подробнее см. табл. 1.

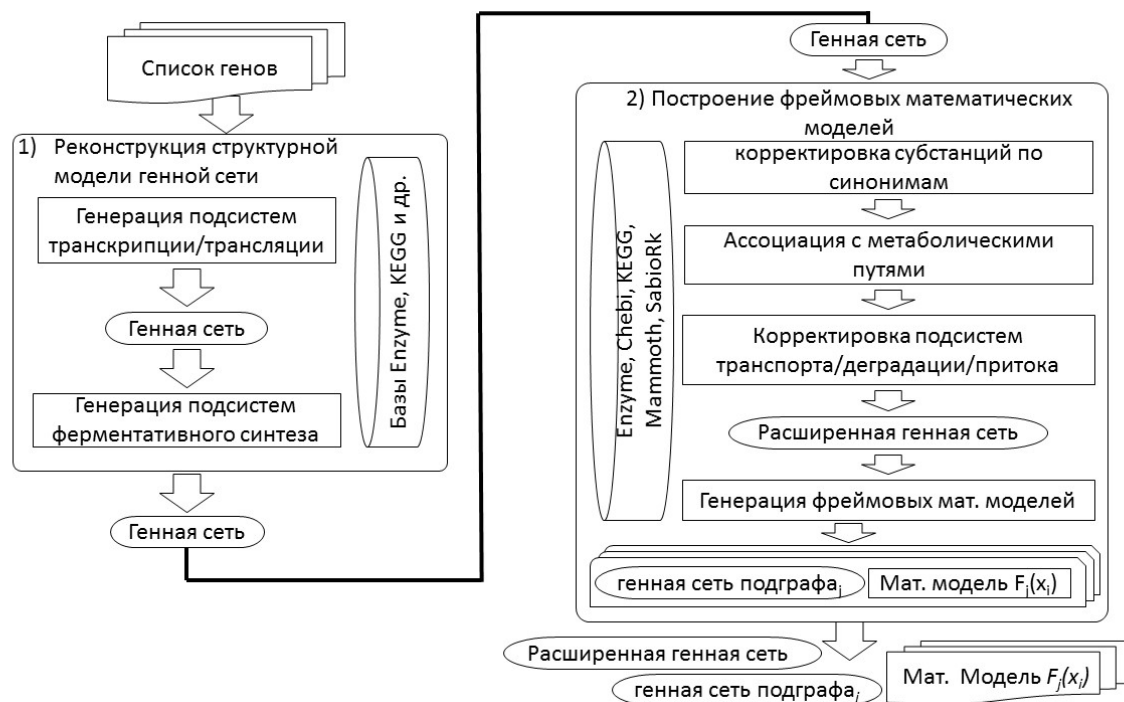


Рис. 1. Схема технологии реконструкции математических моделей по аннотированному геному

Опираясь на принципы ОХКММ, развиваются стандарты представления математических моделей молекулярно-генетических систем. В частности, подход к блочно-модульной организации математических моделей реализован в формате SBML [31], ставшим де-факто стандартом в сообществе системных биологов. Этот формат поддерживает более 200 программных инструментов анализа разных аспектов функционирования биологических систем.

2. Вычислительный конвейер. В рамках работы, опираясь на описанную выше теоретическую и алгоритмическую базу, разработан конвейер обработки данных, который, анализируя входной список генов и их продуктов, восстанавливает 1) структурную модель генной сети организма и 2) строит рамочные математические модели по этой структуре. Принципиальная схема решения задачи реконструкции математических моделей по аннотированному геному представлена на рис. 1.

На первом этапе идет формирование структурной (т.е. графовой) модели для всех процессов, связанных с геномом. В рамках этого этапа формируются список участников моделирования и их роли в процессах. На входе имеется список генов и их продуктов с указанными идентификаторами в публичных базах данных (таких как NCBI, SwissProt и др.). Поэтапно генерируются:

— Реакции транскрипции/трансляции — реакция синтеза белка, где выбранный ген является запускающим условием синтеза, причем в реакции не расходуется. Если имеется информация по позициям транскрипционных факторов для этого гена, то, с применением инструментария *Operon_Equations* [32], генерируется более подробная структурная и математическая модель подсистемы. При наличии дополнительных регуляторных участников процесса математический закон генерируется с использованием инструмента *MGSgenerator* [28].

— Если продукт выбранного гена является ферментом — модель дополняется подсистемой ферментативного синтеза на основе информации, извлеченной из базы ферментативных реакций Enzyme Database [33–34].

Второй этап реконструкции модели — анализ полученных процессов. В этот анализ входят:

— Сравнение элементов модели по синонимам для выявления дублей. Используются публичные базы данных и их API (Enzyme Database и CHEBI). Расширение аннотации элементов модели через кросс-ссылки на сторонние профильные базы данных.

— Анализ процессов на вхождение в известные метаболические пути (используется БД KEGG и ее API).

— Анализ модели на наличие транспортных подсистем для молекул, находящихся в разных компартментах. Дополнительное описание процессов транспорта между компартментами (если необходимо).

— Дополнение модели подсистемами деградации молекул и процессами притока молекул в модель (если они не синтезируются в самой модели).

— Для каждого связанного подграфа генерируется отдельная математическая модель. Для анализа графа используется библиотека networkX.

— При необходимости извлекается информация из открытых баз данных по кинетике процессов, таких как SabioRk [35] и Mammot [36].

Результирующие модели оформляются в следующих форматах:

— SBML — международный формат представления моделей биологических систем.

— GML — формат представления данных в виде графа, в котором каждая вершина и связь между вершинами специфицирована набором собственных атрибутов. Данный формат поддерживается многими инструментами визуализации и анализа графов (например, Cytoscape и yEd).

— CSV — представления данных в виде структурированных текстовых таблиц с фиксированным разделителем. С таким форматом можно работать в инструментах, входящих в офисные пакеты работы с данными, такие как MS Excel и Open Office CALC. Такого рода данные могут быть использованы в инструменте построения графов генных сетей Cytoscape [37].

Так как экспорт модели осуществляется в формате SBML, проведение вычислительных экспериментов можно проводить в любом инструменте, поддерживающем импорт в этом формате. К таким инструментам относятся Copasi [38], CellDesigner [11] и библиотеки моделирования Systems Biology Toolkit для среды моделирования Matlab и библиотека BioPython [39] для языка Python.

3. Пример работы алгоритма генерации фреймовых моделей. Описанный выше протокол реконструкции математических моделей был применен для анализа 150 генов микроорганизмов, секвенированных в ходе выполнения проекта. В данном разделе приведены результаты по реконструкции полногеномной математической модели бактерии *Corynebacterium glutamicum* В-2403. На рис. 2 представлена структурная модель, визуализированная в рамках стандарта SBGN в инструменте yEd (<https://www.yworks.com/products/yed>). Граф является двудольным с количеством узлов 8332 и 9770 связей на 1030 исходных генов ферментов в аннотации. Отдельной актуальной задачей в этой области стоит вопрос планарной раскладки таких графов в условии игнорирования связей ряда простых метаболитов, которые участвуют в большом количестве реакций. Причем

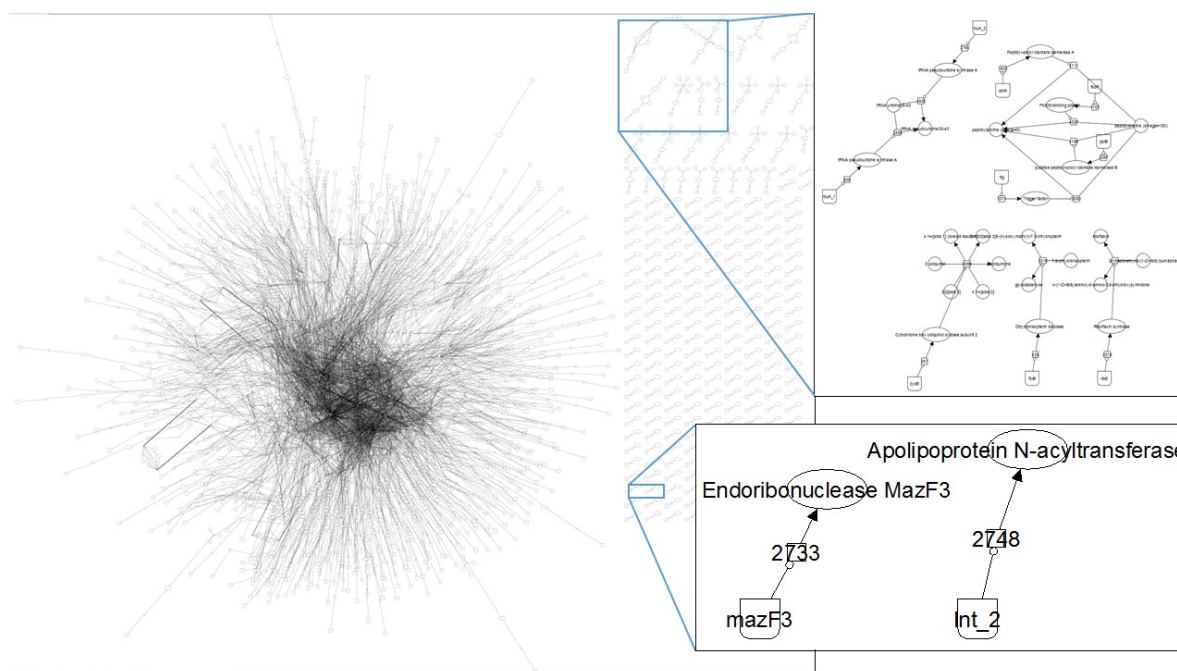


Рис. 2. Реконструированная расширенная структурная модель генной сети, построенная на данных аннотации генома *C. glutamicum* (B-2403), с учетом информации о наличии ферментов. Визуализация выполнена с помощью программы yEd в нотации SBGN

раскладка таких графов, желательно, должна соответствовать стандартным картам метаболических путей.

Итоговая система содержит более 3500 обыкновенных дифференциальных уравнений с более 4500 параметров. В силу блочно-модульной организации модели управление параметрами осуществляется в рамках именно этих блоков, без необходимости выискивания соответствия в векторах параметров.

Заключение. Разработанный программный модуль объединяет в себе цикл задач по созданию модели метаболизма бактерии, включая учет схемы взаимодействия биохимических реакций. Результирующие модели могут быть численно исследованы с использованием стандартного инструментария как широкого профиля (Matlab/Octave, Mathematica и др.), так и узкоспециализированного (среды для моделирования биологических систем COPASI). Модуль был успешно апробирован при анализе 500 геномов микроорганизмов из коллекции Курчатова геномного центра ИЦиГ СО РАН.

Список литературы

1. GOODWIN S., McPHERSON J. D., McCOMBIE W. R. Coming of age: ten years of next-generation sequencing technologies // Nat. Rev. Genet. 2016. Т. 17. № 6. С. 333–351.
2. QUAIL M. и др. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers // BMC Genomics. 2012. Т. 13. № 1. С. 341.
3. COWAN A. E., MENDES P., BLINOV M. L. ModelBricks—modules for reproducible modeling improving model annotation and provenance // npj Syst. Biol. Appl. 2019. Т. 5. № 1.

4. GILBERT D. и др. Towards dynamic genome-scale models // *Brief. Bioinform.* 2019. Т. 20. № 4. С. 1167–1180.
5. KARR J. R. и др. A Whole-Cell Computational Model Predicts Phenotype from Genotype // *Cell.* 2012. Т. 150. № 2. С. 389–401.
6. KIM W. J., KIM H. U., LEE S. Y. Current state and applications of microbial genome-scale metabolic models // *Curr. Opin. Syst. Biol.* 2017. Т. 2. С. 10–18.
7. АКБЕРДИН И. Р. и др. In Silico Cell: Challenges and Perspectives // *Math. Biol. Bioinforma.* 2013. Т. 8. № 1.
8. DEMIN O., GORYANIN I. *Kinetic Modelling in Systems Biology.* , 2008.
9. HELLERSTEIN J. L. и др. Recent advances in biomedical simulations: a manifesto for model engineering // *F1000Research.* 2019. Т. 8. С. 261.
10. OCONE A., MILLAR A. J., SANGUINETTI G. Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics // *Bioinformatics.* 2013. Т. 29. № 7. С. 910–916.
11. FUNAHASHI A. и др. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks // *Proc. IEEE.* 2008. Т. 96. № 8. С. 1254–1265.
12. KING Z. A. и др. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models // *Nucleic Acids Res.* 2016. Т. 44. № D1. С. D515–D522.
13. LLOYD C. M. и др. The CellML Model Repository // *Bioinformatics.* 2008. Т. 24. № 18. С. 2122–2123.
14. MALIK-SHERIFF R. S. и др. BioModels—15 years of sharing computational models in life science // *Nucleic Acids Res.* 2019.
15. HENKEL R., WOLKENHAUER O., WALTEMATH D. Combining computational models, semantic annotations and simulation experiments in a graph database // *Database.* 2015. Т. 2015. С. 1–16.
16. KIRK P. D. W., BAVTIE A. C., STUMPF M. P. H. Systems biology (un)certainties // *Science (80-.)*. 2015. Т. 350. № 6259. С. 386–388.
17. STANFORD N. J. и др. The evolution of standards and data management practices in systems biology // *Mol. Syst. Biol.* 2015. Т. 11. № 12. С. 851–851.
18. BEAL J. и др. Communicating Structure and Function in Synthetic Biology Diagrams // *ACS Synth. Biol.* 2019. Т. 8. № 8. С. 1818–1825.
19. BRUGGEMAN F. J., WESTERHOFF H. V. The nature of systems biology // *Trends Microbiol.* 2007. Т. 15. № 1. С. 45–50.
20. LIKHOSHA V. A. и др. Generalized chemokinetic method for gene network simulation // *Mol. Biol.* 2001. Т. 35. № 6. С. 919–925.
21. PALSSON B. The challenges of in silico biology Moving from a reductionist paradigm to one that views cells as systems will necessitate // 2000. Т. 18. С. 1147–1150.
22. KURATA H. и др. BioFNet: Biological functional network database for analysis and synthesis of biological systems // *Brief. Bioinform.* 2013. Т. 15. № 5. С. 699–709.
23. РАТНЕР В. А. Молекулярно-генетическая система управления // *Природа.* 2001. Т. 3. С. 16–22.
24. MOODIE S. и др. Systems Biology Graphical Notation: Process Description language Level 1 Version 1.3 // *J. Integr. Bioinform.* 2015. Т. 12. № 2.

25. NORSIGIAN C. J. и др. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree // *Nucleic Acids Res.* 2019. Т. 48. № D1. С. D402–D406.
26. ZHANG F. и др. Systems biology markup language (SBML) level 3 package: multistate, multicomponent and multicompartments species, version 1, release 2 // *J. Integr. Bioinform.* 2020. Т. 17. № 2–3. С. 0–74.
27. ЛИХОШВАЙ В. А. и др. Обобщенный химико-кинетический метод моделирования сложных биологических систем. Компьютерная модель онтогенеза бактериофага Lambda // *Вычислительные технологии.* 2000. Т. 5. № Специальный выпуск, посвященный 10-летней годовщине Лаборатории теоретической генетики Института цитологии и генетики СО РАН. С. 87–99.
28. КАЗАНЦЕВ Ф. В., ДР. Система автоматизированной генерации математических моделей генных сетей // *Информационный вестник ВОГИС.* 2009. Т. 13. № 1. С. 163–169.
29. АКБЕРДИН И. Р. и др. «Электронная клетка»: проблемы и перспективы // *Математическая биология и биоинформатика.* 2013. Т. 8. № 1. С. 287–307.
30. ЖАВОТИНСКИЙ А. М. Концентрационные автоколебания., 1974. С. 1–179. М.: Наука.
31. НУСКА М. и др. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core // *J. Integr. Bioinforma.* 2015. Т. 12. № 2. С. 382–549.
32. ЛАКНОВА Т. N. и др. Algorithm for the Reconstruction of Mathematical Frame Models of Bacterial Transcription Regulation // *Mathematics.* 2022. Т. 10. № 23. С. 4480.
33. КАНЕИСА М. Enzyme Annotation and Metabolic Reconstruction Using KEGG., 2017. С. 135–145.
34. McDONALD A. G., BOYCE S., TIPTON K. F. ExplorEnz: the primary source of the IUBMB enzyme list // *Nucleic Acids Res.* 2009. Т. 37. № Database. С. D593–D597.
35. WITTIG U. и др. SABIO-RK—database for biochemical reaction kinetics // *Nucleic Acids Res.* 2012. Т. 40. № D1. С. D790–D796.
36. KAZANTSEV F. V. и др. MAMMOTH: A new database for curated mathematical models of biomolecular systems // *J. Bioinform. Comput. Biol.* 2018. Т. 16. № 01. С. 1740010 (16 pages).
37. OTASEK D. и др. Cytoscape Automation: empowering workflow-based network analysis // *Genome Biol.* 2019. Т. 20. № 1. С. 185.
38. HOOPS S. и др. COPASI — a COMplex PATHway SIMulator // *Bioinformatics.* 2006. Т. 22. № 24. С. 3067–3074.
39. СОСК Р. J. A. и др. Biopython: freely available Python tools for computational molecular biology and bioinformatics // *Bioinformatics.* 2009. Т. 25. № 11. С. 1422–1423.



Лашин Сергей Александрович — канд. биол. наук, доцент, ведущ. науч. сотр. сектора биоинформатики и информационных технологий в генетике Института цитологии и генетики СО РАН, E-mail: lashin@bionet.nsc.ru.

Окончил в 2003 г. ММФ НГУ. Специалист в области математического и компьютерного моделирования биологических систем широкого круга — молекулярно-генетических, популяционно-генетических, экологических, разработки биоинформатических методов, программного обеспечения и баз данных.

Graduated in 2003 from MMF NSU. Specialist in the field of mathematical and computer modeling of biological systems of wide range — molecular-genetic, population-genetic, ecological, development of bioinformatic methods, software and databases.



Казанцев Федор Владимирович, канд. биол. наук, старш. науч. сотр. сектора компьютерного анализа и моделирования биологических систем Института цитологии и генетики СО РАН, E-mail: kazfdr@bionet.nsc.ru.

Области научных интересов: системная биология, биоинформатика в разделах математического и компьютерного моделирования генных, белковых, метаболических и сигнальных сетей. Участвовал в разработке серии предметно-ориентированных биологических интернет ресурсов и баз данных.

Research interests: systems biology, bioinformatics in the sections of mathematical and computer modeling of gene, protein, metabolic and signaling networks. Participated in the development of a series of subject-oriented biological Internet resources and databases.



Лахова Татьяна Николаевна — младш. науч. сотр. сектора биоинформатики и информационных технологий в генетике Института цитологии и генетики СО РАН, E-mail: tlakhova@bionet.nsc.ru.

В область научных интересов входят математическое моделирование биологических систем и анализ данных.

Research interests include mathematical modeling of biological systems and data analysis.



Матушкин Юрий Георгиевич — канд. биол. наук, доцент, старш. науч. сотр., зав. лабораторией молекулярно-генетических систем Института цитологии и генетики СО РАН, E-mail: mat@bionet.nsc.ru.

Специалист в области молекулярной эволюции и математического моделирования молекулярно-генетических систем.

Specialist in molecular evolution and mathematical modeling of molecular genetic systems.

Дата поступления — 02.11.2024