

РЕГУЛЯЦИЯ ЭКСПРЕССИИ ГЕНОВ И ЭЛЕКТРОННЫЕ МОДЕЛИ ЖИВОЙ КЛЕТКИ

М.С.Гельфанд

Институт проблем передачи информации, Российская академия наук, Москва, Россия

Поступила в редколлегию 1.2.2005

Аннотация—Дан краткий обзор моделей регуляции метаболизма живых клеток. Приведены возможные постановки задач для учёта регуляторных взаимодействий в потоковых моделях, обсуждены источники данных для такого анализа.

1. ОБЗОР

1.1. Модели клетки

В настоящее время разрабатываются два основных типа моделей клеточного метаболизма и физиологии. Потоковые модели рассматривают статическое состояние, при этом строится такое распределение потоков вещества через отдельные реакции, которое максимизирует производство какого-либо вещества (например, АТФ) или смеси веществ (биомассы, заданной процентными долями определённых аминокислот, липидов, нуклеиновых кислот и т.п.). Физиология организма задаётся списком происходящих в нём химических реакций (в том числе транспортных реакций переноса метаболитов через клеточную мембрану), производным от списка генов, а предметом изучения являются распределения потоков в зависимости от состава среды и/или набора реакций (моделирование мутаций). Поскольку вычислительная сторона модели состоит в использовании эффективного алгоритма линейного программирования, возможно моделирование полных геномов, что ограничивается лишь доступностью сведений о функциях генов. Списки реакций и другие параметры для ряда моделей, в т.ч. дрожжей, доступны по адресу (<http://systemsbiology.ucsd.edu/>). Потоковые модели применяются для направленного конструирования штаммов дрожжей [Carlson-02].

Кинетические модели в явном виде учитывают зависимость интенсивности реакции от концентраций метаболитов и кофакторов. Интенсивности реакций описываются системой дифференциальных уравнений. Помимо вычислительных сложностей при анализе больших систем и проблем, связанных с возможной неустойчивостью построенных систем, применимость таких моделей сильно ограничена тем, что в большинстве случаев многие параметры систем не известны. С другой стороны, нет принципиальных сложностей в учете концентраций ферментов. В настоящее время развивается модель дрожжевой клетки (проект “Silicon Cell”, <http://www.siliconcell.net>); база данных кинетических моделей поддерживается на (<http://jjj.biochem.sun.ac.za>).

1.2. Учёт регуляции в потоковых моделях

До последнего времени потоковые модели разрабатывались без учета регуляторных взаимодействий. Хотя были построены модели митохондрии “вообще” [Vo-04] и человеческой [Ramakrishna-01] и частичная модель дрожжевой клетки [Föster-03; Duarte-04], существенно нетривиальные результаты были получены только при моделировании метаболизма кишечной палочки, организма, геном которого изучен наиболее подробно.

Было показано, что потоковые модели относительно успешно предсказывают фенотип мутантных штаммов с нокаутированными метаболическими генами [Edwards-00; Edwards-00a], и что учет

регуляторных взаимодействий улучшает качество предсказаний [Covert-02]. В последнем случае регуляция учитывалась наиболее простым способом, как тотальное включение или выключение гена. Оказалось, что такая модель улучшает предсказания последствий инактивирующих мутаций (на уровне выживет/не выживет), а также в ряде случаев позволяет качественно описать поведение популяций в клеточных культурах [Edwards-01]. Другой путь улучшения предсказаний – учёт ограничений, полученных из анализа экспериментов по введению изотопной метки [Blank-04].

В то же время, было показано, что экспериментальное распределение потоков в диком штамме восстанавливается потоковыми моделями удовлетворительно, а в мутантных – плохо [Segrè-02]. Одним из возможных объяснений этого является то, что экспрессия генов в ходе эволюции оказалась отрегулирована таким образом, чтобы максимизировать обеспечивать протекание реакций с оптимальными интенсивностями. В то же время, в мутантном штамме оптимальное распределение потоков другое, и существующие регуляторные системы не в состоянии его воспроизвести. Действительно, оказалось, что модификация алгоритма линейного программирования, позволяющая найти субоптимальное распределение потоков после нокаута, близкое к распределению дикого типа, позволяет удовлетворительно воспроизвести экспериментальные данные. Если это объяснение верно, можно ожидать, что распределение потоков при выключении (групп) генов за счёт известных регуляторных механизмов, при внешнем сходстве последних с деактивирующими мутациями, может быть успешно воспроизведено в рамках первоначальной модели.

Оказалось также, что степень соответствия реконструированных потоков экспериментальным данным зависит от среды. Так, модель *E. coli* существенно успешнее воспроизводит потоки при недостатке углерода, чем при недостатке азота [Д. Виткуп, частное сообщение]. Аналогично, потоки при росте на обычных источниках углерода, таких как глюкоза, восстанавливаются лучше, чем потоки при росте на глицероле [Ibarga-02]. Более того, после отбора на рост в необычной среде (40 дней, приблизительно 700 поколений), экспериментально определённые потоки стали совпадать с предсказанными. Как и в предыдущем случае, напрашивается эволюционное объяснение: похоже, что в опыте была воспроизведена эволюция в сторону оптимального использования нового источника углерода, а это последнее и реконструируется в потоковой модели. Наконец, аналогичная эволюция в сторону предсказанных потоков была показана для мутантных штаммов [Fong-03; Fong-04].

Другой, менее трудоёмкий, способ сравнения вычисленных потоков с реальностью - использование косвенных данных, в частности, из которых наиболее массовыми и доступными являются данные об уровнях экспрессии генов в различных условиях (см. список баз данных http://ihome.cuhk.edu.hk/~b400559/arraysoft_public.html и табл. 1). При этом возникают сложности, связанные с тем, что соотношение между концентрацией мРНК и скоростью реакции для соответствующего фермента различно для каждого гена. Для того чтобы их обойти, можно рассматривать пары экспериментов, отличающихся, скажем, составом среды, и сравнивать наблюдаемые отношения концентраций мРНК (как правило, именно эти величины приводятся в базах данных) и отношения вычисленных потоков. Ясно, что для такого анализа можно брать только данные, полученные на хорошо определённых средах.

Следует отметить, однако, что неясно, до какой степени концентрации мРНК и белков даже в штаммах дикого типа коррелируют с предсказанными потоками. Отсутствие такой корреляции для дрожжей объясняют неучётом регуляции при построении моделей [Nielsen-04]. Качественное совпадение изменений уровней экспрессии генов и соответствующих потоков в культурах кишечной палочки, при росте на ацетате и глюкозе было показано в [Oh-02]. С другой стороны, анализ гликолиза в простейших показал отсутствие такой корреляции [terKuile-01]. Одной из причин этого может служить недостаточность данных, используемых для построения моделей даже в хорошо изученных системах [Oh-00]. В [Stelling-02] был разработан метод анализа метаболизма, который вычисляет не оптимальное распределение потоков для какого-то функционала, а среднее распределение по всем элементарным потокам (extreme pathways, elementary modes), возможным в стационарном состоя-

Таблица 1. Основные ресурсы, содержащие данные об экспериментах по измерению уровней экспрессии генов (microarray expression data).

Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo/	NCBI expression data repository
ArrayExpress	http://www.ebi.ac.uk/arrayexpress/	EBI expression data repository
Riken Expression Array Database	http://read.gsc.riken.go.jp/	RIKEN expression data repository
Stanford Microarray Database	http://genome-www5.stanford.edu/	Stanford University expression data repository
Expression Connection	http://db.yeastgenome.org/cgi-bin/expression/expressionConnection.pl	A gene expression database of Saccharomyces genome database (SGD)
yeast Microarray Global Viewer (yMGV)	http://www.transcriptome.ens.fr/ymgv/	A database for yeast gene expression data maintained by Laboratoire de genetique moleculaire, Ecole Normale Supérieure
GermOnline	http://www.germonline.org	Data on expression of genes involved in mitosis and meiosis, gamete formation and germ line development.
ExpressDB	http://twod.med.harvard.edu/ExpressDB/	Data on expression in different conditions from the Church group (Harvard)

нии. Утверждается, что отличия потоков через отдельные реакции, вычисленных таким образом, от уровней экспрессии, сравнимы с экспериментальными ошибками в определении последних.

1.3. Учёт регуляции в кинетических моделях

В кинетических моделях регуляторные взаимодействия учитываются введением в уравнения реакций членов, учитывающих ретроингибирование. Столь же просто ввести коэффициенты, регулирующие концентрацию фермента в зависимости от концентраций метаболитов (и от времени), и тем самым напрямую моделирующие регуляцию экспрессии генов. Однако численные значения соответствующих параметров довольно редко бывают известны. В то же время, когда это удаётся сделать, модель лучше объясняет результаты экспериментов по росту мутантов на различных средах [Goltsov-04; Г.В.Лебедева и О.В.Дёмин, частное сообщение].

Одной из специальных разновидностей кинетических моделей являются модели регуляции, явно учитывающие скорости транскрипции и трансляции и потребность этих процессов в метаболитах [Likhoshvai-02]. В частности, таким образом была построена модель регуляции синтеза аминокислот, в которой были учтены регуляторные механизмы репрессии транскрипции и аттенуации [Elf-01]. Оказалось, что оба вида регуляции весьма чувствительны к концентрациям регулирующего целевого метаболита и в большей части интервала физиологически возможных условий функционируют практически как логические переключатели. Следует отметить, что это не вполне согласуется с множественными биологическими наблюдениями, указывающими, что регуляция в клетках, в частности, регуляция экспрессии генов аминокислотных оперонов, редко является абсолютной.

2. ВОЗМОЖНЫЕ ПОСТАНОВКИ ЗАДАЧ

Одна из возможных задач анализа регуляции в потоковых моделях уже была упомянута – это сравнение адекватности потоковых моделей для нокаутных мутантов и для случая репрессии генов, регулируемых в диком штамме. Гипотеза состоит в том, что реконструируемые потоки будут лучше соответствовать экспериментальным данным во втором случае, нежели в первом.

Если известны параметры, описывающие репрессию или активацию генов, можно напрямую учесть их как ограничения на потоки [Covert-01; Covert-03]. При этом возможна итеративная процедура: вычисляются потоки, а, стало быть, и стационарные концентрации, которые используются для установ-

ления ограничений; вычисляются потоки с учётом этих ограничений, и т.д. [В.А.Любецкий, частное сообщение]. Такого рода приёмы можно применять, даже если ограничения известны не полностью, например, только в нескольких метаболических путях – это равносильно тому, что остальные реакции считаются нерегулируемыми. Кроме того, в первом приближении, по-видимому, достаточно считать, что любая регуляция осуществляется по триггерному механизму: реакция идёт либо не идёт в зависимости от того, превышает ли концентрация регулирующего метаболита некоторый фиксированный порог (ср. обсуждение в последнем абзаце предыдущего параграфа).

С другой стороны, экспрессионные данные могут быть использованы для выставления ограничений на потоки. Например, при моделировании метаболизма на разных стадиях клеточного цикла, можно учитывать, какие гены экспрессируются на каждой стадии. Одной из существенных проблем здесь является то, что данные об уровнях экспрессии, как правило, не являются дискретными: для многих генов и стадий не характерна абсолютная репрессия или активация. В то же время, не ясно, как перевести относительные уровни экспрессии в ограничения для потоковой модели. Даже если считать, сильно огрубляя, что концентрации всех мРНК пропорциональны концентрациям соответствующих белков (тем самым, игнорируя различия в скоростях трансляции и в устойчивости белков), всё равно остаётся неизвестным второй множитель, определяющий скорость реакции, а именно, количество реакций, катализируемых в единицу времени одной молекулой фермента. Тем самым, даже зная концентрации мРНК, не удаётся выставить ограничения на максимальный поток через соответствующую реакцию.

Один из способов обойти это ограничение, по-видимому, состоит в следующем. Проведём моделирование для достаточно широкого спектра условий, варьируя среду, целевой функционал, и набор работающих генов, и рассмотрим максимальный из потоков, наблюдаемых в этих моделях для данной реакции. Сопоставим ему максимальный уровень экспрессии, наблюдаемый в эксперименте для данного гена. Это даст требуемое соотношение между наблюдаемыми (уровень экспрессии/измеренная концентрация мРНК) и вычисляемыми (поток) величинами; ясно, что такое соотношение будет своим для каждого гена.

Помимо экспрессионных данных, информацию о регуляции можно извлекать из литературы и из баз данных регуляторных сигналов: TRANSFAC (<http://www.gene-regulation.com>), в частности, дрожжевой модуль TSM (<http://transfac.gbf.de>) [Matys-03] и SCPD (<http://cgsigma.cshl.org/jian>) [Shu-99]. Далее, можно использовать результаты массовых экспериментов по иммунной ко-преципитации (http://web.wi.mit.edu/young/regulator_network/) [Lee-02]. Наконец, интерес представляют группы потенциально ко-регулируемых генов, получаемые из одновременного анализа различных источников данных, в частности, данных по ко-экспрессии и по белок-белковым взаимодействиям [Тройанская-03].

Особый интерес представляет использование техники сравнительной геномики для определения структуры метаболической карты. В настоящее время разложение метаболической карты на отдельные пути в таких базах данных как KEGG и BioСус основано на биохимической традиции. Объективные подходы к определению путей могут быть основаны на предположении, о том, что гены, составляющие путь, будут встречаться в геномах одновременно и экспрессироваться в одних условиях. В самом деле, было показано, что кластеры ко-экспрессируемых генов в дрожжах часто соответствуют линейным метаболическим путям, причём изоферменты, функционирующие в составе различных путей, принадлежат к разным кластерам [Ihmels-04]. Геномные подходы к определению путей основаны на анализе филогенетического распределения [Glazko-04] и позиционной кластеризации [vonMering-03] позволяют идентифицировать гены-кандидаты на заполнение пробелов в геном-специфических метаболических картах. Другой подход состоит в использовании метаболического моделирования: метаболический путь можно определить как набор реакций, потоки через которые полностью или частично обуславливают друг друга [Ederer-03; Burgard-04]. Представляет интерес следующая постановка задачи: найти набор потоков, существенно отличающихся при ро-

сте на минимальной среде и на среде с добавлением какого-либо метаболита (одной аминокислоты, нуклеотида, кофактора) и сравнить этот набор с соответствующим регулоном, т.е. набором генов, регулируемых одним фактором транскрипции [А.А.Миронов, частное сообщение]. Следует отметить, что многочисленные наблюдения показывают, что регулоны не полностью консервативны: наборы генов, регулируемых ортологичными факторами транскрипции, могут отличаться даже в относительно близких геномах. Можно предположить, что пути, идентифицированные путём анализа потоков, будут соответствовать консервативным ядрам регулонов.

Наконец, представляется интересным попытаться учесть в кинетических моделях силу отдельных регуляторных сайтов. В [Zaslaver-04] было показано, что экспрессия генов из одного пути при исключении конечного метаболита из среды (т.е. при включении пути) подчиняется естественному закону: экспрессия гена начинается тем раньше, чем ближе к началу пути находится кодируемый этим геном фермент. Это может быть одним из объяснений для наблюдаемого нами эффекта консервативности неконсенсусных нуклеотидов в ортологичных регуляторных сайтах [Kotelnikova-05]. Поскольку известно, что есть корреляция между силой сайта и его близостью к консенсусу, можно попытаться оценить уровень регуляции для любого сайта, и учесть его в кинетической модели. Интересно также, насколько такие модели будут воспроизводить экспериментальные данные. В качестве предварительного анализа было бы интересно сопоставить (предсказанную) силу сайта и положение на метаболическом пути фермента, кодируемого регулируемым геном.

Автор выражает благодарность Л.Н.Дроздову-Тихомирову, В.А.Любецкому, А.А.Миронову за ценное обсуждение. Эта работа была частично поддержана программами “Молекулярная и клеточная биология” и “Происхождение и эволюция биосферы” РАН.

СПИСОК ЛИТЕРАТУРЫ

1. Blank L.M., Kuepfer L., Sauer U. Deciphering metabolic network robustness. *5th Int. Conf. on Computational Systems Biology ICSB '2004* (Heidelberg, Germany), p. 85.
2. Burgard A.P., Nikolaev E.V., Schilling C.H., Maranas C.D. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.*, 2004, vol. 14, pp. 301–312.
3. Carlson R., Fell D., Sreenc F. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol. Bioeng.*, 2002, vol. 79, pp. 121–134.
4. Covert M.W., Schilling C.H., Palsson B.Ø. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, 2001, vol. 213, pp. 73–88.
5. Covert M.W., Palsson B.Ø. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.*, 2002, vol. 277, pp. 28058–28064.
6. Covert M.W., Palsson B.Ø. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.*, vol. 221, pp. 309–325.
7. Duarte N.C., Herrgard M.J., Palsson B.Ø. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.*, 2004, vol. 14, pp. 1298–1309.
8. Ederer M., Sauter T., Bullinger E., Gilles E.D., Allgöwer F. An approach for dividing models of biological reaction networks into functional units. *Simulation*, 2003, vol. 79, pp. 703–716.
9. Edwards J.S., Palsson B.Ø. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA*, 2000, vol. 97, pp. 5528–5533.
10. Edwards J.S., Palsson B.Ø. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics*, 2000a, vol. 1, pp. 1.
11. Edwards J.S., Ibarra R.U., Palsson B.Ø. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 2001, vol. 19, pp. 125–130.
12. Elf J., Berg O.G., Ehrenberg M.J. Comparison of repressor and transcriptional attenuator systems for control of amino acid biosynthetic operons. *J. Mol. Biol.*, 2001, 313, pp. 941–954.

13. Fong S.S., Marciniak J.Y., Palsson B.Ø. Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale *in silico* metabolic model. *J. Bacteriol.*, vol. 185, pp. 6400–6408.
14. Fong S.S., Palsson B.Ø. Metabolic gene deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genet.*, vol. 36, pp. 1056–1058.
15. Foster J., Famili I., Fu P., Palsson B.Ø., Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.*, 2003, vol. 13, pp. 244–253.
16. Goltsov A.N., Lebedeva G.V., Lavrova A.I., Demin O.V. Modeling of purine nucleotides biosynthesis in *E. coli*. *5th Int. Conf. on Computational Systems Biology ICSB'2004* (Heidelberg, Germany), p. 95.
17. Glazko G.V., Mushegian A.R. Detection of evolutionarily stable fragments of cellular pathways by hierarchical clustering of phyletic patterns. *Genome Biol.*, 2004, vol. 5, p. R32.
18. Ibarra R.U., Edwards J.S., Palsson B.Ø. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 2002, vol. 420, pp. 186–189.
19. Ihmels J., Levy R., Barkai N. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat. Biotechnol.*, 2004, vol. 22, pp. 86–92.
20. Kotelnikova E.A., Makeev V.Y., Gelfand M.S. Evolution of transcription factor DNA binding sites. *Gene*, 2005, in press.
21. Lee T.I., Rinaldi N.J., Robert F., Odom D.T., Bar-Joseph Z., Gerber G.K., Hannett N.M., Harbison C.T., Thompson C.M., Simon I., Zeitlinger J., Jennings E.G., Murray H.L., Gordon D.B., Ren B., Wyrick J.J., Tagne J.B., Volkert T.L., Fraenkel E., Gifford D.K., Young R.A. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 2002, vol. 298, pp. 799–804.
22. Likhoshvai V.A., Matushkin Y.G. Differentiation of single-cell organisms according to elongation stages crucial for gene expression efficacy. *FEBS Lett.*, 2002, vol. 516, pp. 87–92.
23. Matys V., Fricke E., Geffers R., Gossling E., Haubrock M., Hehl R., Hornischer K., Karas D., Kel A.E., Kel-Margoulis O.V., Kloos D.U., Land S., Lewicki-Potapov B., Michael H., Munch R., Reuter I., Rotert S., Saxel H., Scheer M., Thiele S., Wingender E. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, 2003, vol. 31, pp. 374–378.
24. Nielsen J. From gene expression to metabolic fluxes. *5th Int. Conf. on Computational Systems Biology ICSB'2004* (Heidelberg, Germany), p. 23.
25. Oh M.K., Liao J.C. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol. Prog.*, 2000, vol. 16, pp. 278–286.
26. Oh M.K., Rohlin L., Kao K.C., Liao J.C. Global expression profiling of acetate-grown *Escherichia coli*. *J. Biol. Chem.*, 2002, vol. 277, pp. 13175–13183.
27. Ramakrishna R., Edwards J.S., McCulloch A., Palsson B.O. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Regulatory Integrative Comp. Physiol.*, 2001, vol. 280, pp. R695–R705.
28. Segrè D., Vitkup D., Church G.M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA*, 2002, vol. 99, pp. 15112–15117.
29. Stelling J., Klamt S., Bettenbrock K., Schuster S., Gilles E.D. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 2002, vol. 420, pp. 190–193.
30. ter Kuile B.H., Westerhoff H.V. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.*, 2001, vol. 500, pp. 169–171.
31. Troyanskaya O.G., Dolinski K., Owen A.B., Altman R.B., Botstein D. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc. Natl. Acad. Sci. USA*, 2003, vol. 100, pp. 8348–8353.
32. Vo T.D., Greenberg H.J., Palsson B.O. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J. Biol. Chem.*, 2004, vol. 279, pp. 39532–39540.

33. von Mering C., Zdobnov E.M., Tsoka S., Ciccarelli F.D., Pereira-Leal J.B., Ouzounis C.A., Bork P. *Proc. Natl. Acad. Sci. USA*, 2003, vol. 100, pp. 15428–15433.
34. Zaslaver A., Mayo A.E., Rosenberg R., Bashkin P., Sberro H., Tsalyuk M., Surette M.G., Alon U. Just-in-time transcription program in metabolic pathways. *Nat. Genet.*, 2004, vol. 36, pp. 486-491.
35. Zhu J., Zhang M.Q. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics*, 1999, vol. 15, pp. 607-611.