

ПРИРОДА

№ 3, 2001 г.

Ратнер В.А.

Молекулярно-генетическая система управления

© “Природа”

**Использование и распространение этого материала
в коммерческих целях
возможно лишь с разрешения редакции**



Сетевая образовательная библиотека “VIVOS VOCO!”
(грант РФФИ 00-07-90172)

vivovoco.nns.ru
vivovoco.rsl.ru
www.ibmh.msk.su/vivovoco

Молекулярно-генетическая система управления

В.А.Ратнер

Весь опыт молекулярной генетики показывает, что главные молекулярные компоненты клетки — нуклеиновые кислоты и белки. С ними связаны все важные процессы и свойства клеток — самовоспроизведение, наследование, контроль метаболизма, ферментативный катализ, построение морфологических структур, транспорт веществ, развитие, дифференцировка, иммунитет и т.д.

И нуклеиновые кислоты, и белки обладают общими фундаментальными свойствами. Все они построены из стандартных мономеров (нуклеиновые кислоты — из нуклеотидов четырех типов, белки — из аминокислот 20 типов), количество, состав и порядок которых определяют функции и свойства конкретных макромолекул. Клеточная система этих кодирующих биополимеров при помощи собственных внутрисистемных средств (ферментов репликации, транскрипции, трансляции, а также рибосом и т.д.) способна к самовоспроизведению. Исполняющие элементы основных генетических процессов сами построены из таких макромолекул и кодируются в генах той же клеточной системы. Перечисленные «устройства» выполняют в клетках фундаментальные генетические процессы (репликацию, транскрипцию, трансляцию, репарацию, рекомбинацию, дегра-

© В.А.Ратнер



***Вадим Александрович Ратнер**, доктор биологических наук, Соросовский профессор. Заведующий теоретическим отделом и лабораторией молекулярно-генетических систем Института цитологии и генетики СО РАН. Основные интересы связаны с математической генетикой и теорией молекулярной эволюции молекулярно-генетических систем.*

дацию, сегрегацию и др.). Используя компьютерную терминологию, можно сказать, что ДНК и мРНК представляют собой «программное обеспечение» — инструкции, полученные клеткой от родительской клетки. Белки и каталитические РНК составляют «аппаратное обеспечение» — физические механизмы, осуществляющие хранящуюся в памяти программу.

Мы назвали всю систему биополимеров клетки молекулярно-генетической системой управления (МГСУ), которая базируется на структурно-функциональном подходе. При информационно-лингвистическом описании такой системы на передний план выходят принципы организации и управления, са-

мовоспроизведение, информационные процессы, помехоустойчивость, кодирование, память, языки и т.п., а структурные, физико-химические свойства отходят на второй план. В этом случае мономеры считаются символами базового алфавита; макромолекулы задаются последовательностями символов, или генетическими текстами; системы взаимодействующих генов характеризуются схемами их молекулярных взаимодействий, т.е. конструкциями, или генетическими сетями; геномы — последовательностями символов генов, знаков пунктуации и управления и других функциональных единиц, т.е. генетическими картами; вся система управления в целом задается замк-

нутой конструкцией, т.е. схемой функциональных взаимодействий.

Таким образом, молекулярные свойства, отношения, функции, записанные в генетических текстах, можно считать генетической информацией, а правила и закономерности ее записи — генетическим языком. Определяются также и другие понятия информационно-лингвистического характера: генетическая память, информационные процессы, архив генетической информации (геном) и др. Значит, МГСУ клетки — это совокупность ее генетических текстов и молекулярных элементов, выполняющих различные генетические процессы. Такой подход оказывается весьма продуктивным — на его основе формулируются и решаются десятки и сотни принципиальных задач организации и эволюции МГСУ.



Схема основных подсистем МГСУ.

Внутреннее строение

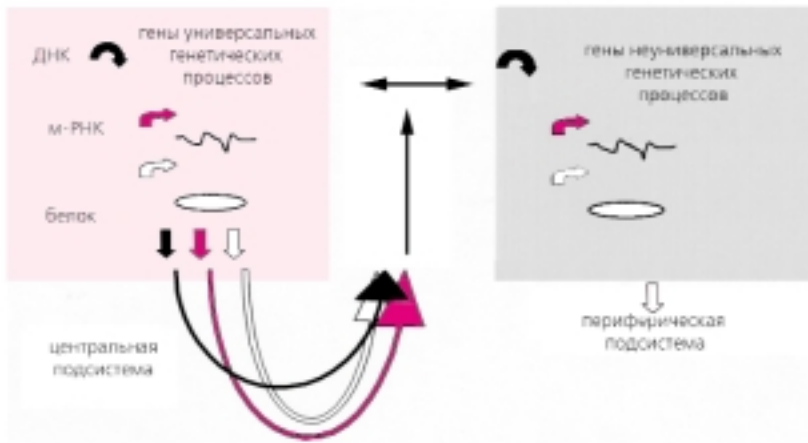
Молекулярно-генетическая система управления достаточно сложна: содержит от 500 до 100 000 генов, примерно столько же белков; она контролирует сотни и тысячи метаболических процессов, а также дифференцировку, развитие, иммунитет и многое другое. В то же время эта сложность не чрезмерна. Отдельные молекулярные и более сложные функции и свойства сосредоточены в конкретных модулях организации, имеющих определенную внутреннюю информационную структуру. В целом реальные клетки и организмы, а также их системы управления имеют иерархическое блочно-модульное строение, возникшее эволюционно [1]. Блоком (модулем) системы называют автономную подсистему с определенной функцией. Иерархичность всей системы управления подразумевает, что модули более высоких ярусов состоят из комбинаций блоков предыдущих ярусов и соединительных элементов, а выделяемые ярусы, блоки и подсистемы имеют информационную природу. Наиболее известные модули нижних ярусов — кодоны, знаки пунктуации

и управления, гены, транскрипты, белки, опероны, репликоны, мобильные элементы и др. Автономность этих модулей определяется дискретностью генов.

Если двигаться от сложных свойств и функций к более простым, можно выделить несколько модулей верхних ярусов. Еще Э.Шредингер в 1944 г. указывал, что «хромосомные структуры являются одновременно и архитектором, и строителем» [2]. В наших терминах [3] это означает, что молекулярно-генетическая система управления имеет центральную подсистему МГСУ, содержащую исполняющие устройства: в репликации — это фермент ДНК-зависимая ДНК-полимераза и комплекс других ферментов; в транскрипции — ДНК-зависимая РНК-полимераза и остальные ферменты; в трансляции — рибосома, ферменты кодирования, т-РНК и др. и гены основных генетических процессов, т.е. универсальных в клетке. Сюда относятся модули репликации, транскрипции, трансляции, репарации, рекомбинации, сегрегации и т.д. Центральная подсистема составляет базу организации клетки, т.е. она — «строитель» по Шредингеру.

Все остальные, неуниверсальные модули объединены в периферическую подсистему, отвечающую за разнообразие других неуниверсальных функций, но которые в совокупности тоже обслуживают всю клетку и обеспечивают ее жизнедеятельность, энергетику и материальную автономность. Фактически в периферической подсистеме сосредоточена информация о специфической «архитектуре» молекулярно-генетической системы управления (по Шредингеру).

«Ядро» центральной подсистемы образует группа взаимодействующих блоков репликации, транскрипции, трансляции и сегрегации, которые в совокупности и во взаимодействии обеспечивают самовоспроизведение системы [4]. Этот модуль мы назвали сайзером (YSER — SYstem of SElf Reproduction), т.е. универсальной системой самовоспроизведения. Белки репликации воспроизводят гены всех четырех блоков, а также любые другие гены, имеющие такую же пунктуацию. Белки транскрипции переписывают гены всех четырех блоков, а также любые другие гены с такой же пунктуацией. Белки трансляции передают



Блочное-модульное строение молекулярно-генетической системы управления клетки. Слева — центральная подсистема, где каждый генетический процесс имеет исполняющее устройство и соответствующую ему группу генов (генетическая память). Справа — периферическая, которая отвечает за разнообразие других функций МГСУ.

информацию от м-РНК всех четырех блоков, а также любых других генов с такой же пунктуацией. Белки сегрегации обеспечивают равное разделение всех генов между дочерними клетками. Иначе говоря, именно универсальность белков и знаков пунктуации этих процессов образует прямые и обратные связи между блоками, что стабилизирует всю МГСУ и обеспечивает ее самовоспроизведение.

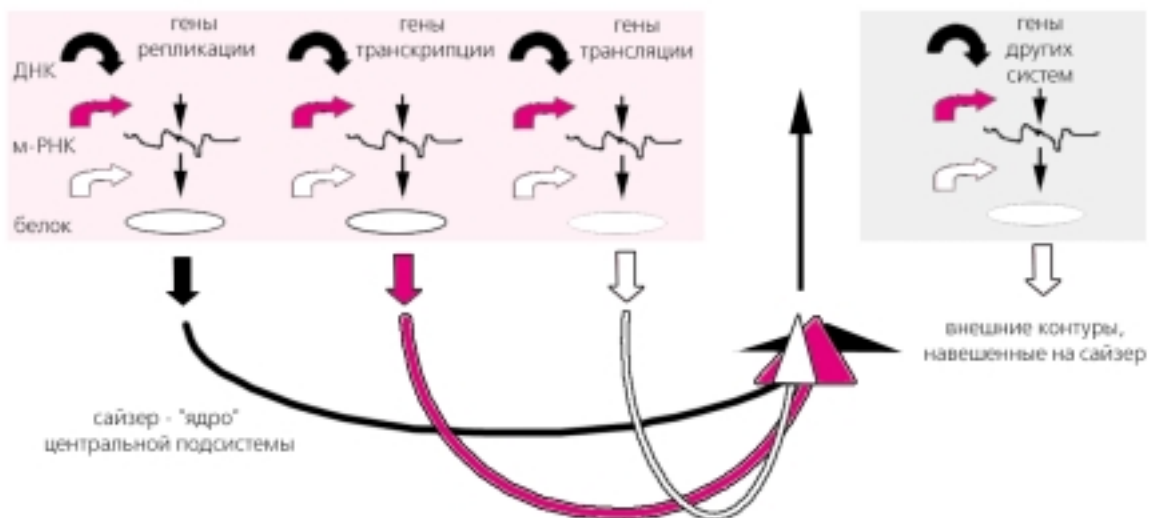
Можно увеличить внутреннюю сложность сайзера или ввести дополнительные, внешние блоки

и контуры МГСУ с такой же пунктуацией; при этом способность к самовоспроизведению сохранится. В начале 80-х годов нам удалось построить математические модели ядра и впервые показать его богатые динамические свойства и ключевую роль в молекулярно-генетической организации [5].

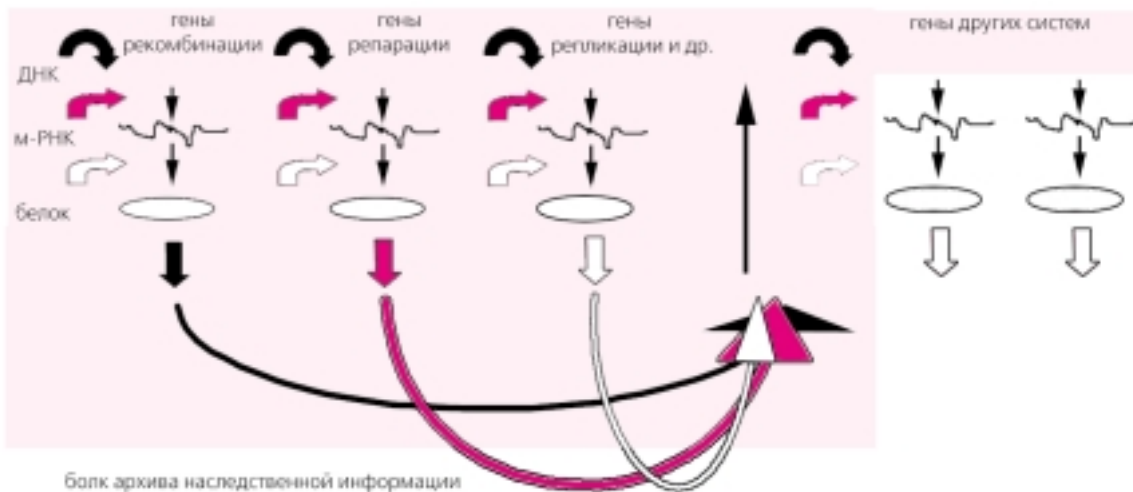
Удивительно, что вопреки ожиданиям схема самовоспроизведения оказалась достаточно простой. Насколько я могу судить, в природе и технике пока нет других систем,

обладающих в полной мере таким самовоспроизведением.

Другой универсальный модуль центральной подсистемы — архив генетической информации. Он содержит частично перекрывающиеся универсальные блоки репарации, общей рекомбинации, репликации и др., а также неуниверсальные блоки транспозиции мобильных генетических элементов, обратной транскрипции, хромосомных перестроек, упаковки хроматина и некоторых других. Архив генетической информации отвечает за



Строение сайзера — блока самовоспроизведения (справа — его внешние контуры). Он состоит из универсальных модулей репликации, трансляции и транскрипции и обеспечивает воспроизведение всех макромолекулярных компонентов клетки и разделение копий в пространстве.



Блок архива наследственной информации. Это система хранения, поддержания, дублирования коррекции и манипулирования, обеспечивающая изменчивость генетической информации в геноме и некоторых клеточных органеллах.

хранение, поддержание, дублирование, коррекцию, манипулирование, т.е. в целом — за помехоустойчивость генетической информации. Он же обеспечивает изменчивость генетической информации, поскольку функция помехоустойчивости дополнительна к функции мутирования, которая возникает в основном из-за ошибок в системах репликации и других процессах.

Архив генетической информации включает сотни и тысячи отдельных секций центральной и периферической подсистем с автономным доступом к ним (гены, опероны и т.д.). Они восприимчивы к специфическим внешним сигналам и через них — к внешнему управлению, а также способны к отклику путем автономного изменения потоков считывания информации. В секциях периферической подсистемы записаны функции контроля метаболизма, онтогенеза, физиологических процессов, иммунитета и т.д. Эти секции находятся на входе соответствующих блоков, отвечающих за специфическое управление секциями архива.

Блок управления метаболизмом состоит из сотен и тысяч автономных систем метаболического контроля, работающих почти неза-

висимо и параллельно или входящих в сложные каскадные системы. Сложность блока соответствует сложности путей метаболизма. Специфическое управление секциями блока осуществляется через прямые и обратные связи между генами, ферментами, регуляторными белками, знаками управления, метаболитами, сигнальными агентами, гормонами и др. Автономное управление может быть введено почти в любом неуниверсальном звене потока информации. Наиболее существенные автономные подсистемы этого блока контролируют базовый метаболизм (включая синтез нуклеотидов и аминокислот), энергетику, транспорт веществ и другие процессы.

По такому же принципу организован и блок управления развитием (онтогенезом), контролирующей временно-пространственную топографию онтогенетических событий. Это особенно важно для многоклеточных организмов, но развитие претерпевают также отдельные клетки (клеточный цикл) и даже вирусы и фаги внутри клеток. Онтогенез завершается самовоспроизведением их систем управления. На входе этого блока находится большая группа автономных секций архива (генов), кодирующих специфические зве-

нья развития — дифференцировку, морфогенез, формирование тканей, клеточное деление и т.п. На промежуточных этапах используются универсальные блоки репликации, транскрипции, трансляции, процессинга и др.

Очевидно, что онтогенез совершается на основе наследственной программы управления развитием, записанной в геноме (архиве). Вопрос состоит в том, каким образом эта программа записана. Содержит ли блок наследственной памяти алгоритм программы?

Опыт молекулярной генетики показывает, что генетическая программа развития действительно записана в геноме, но не в виде последовательного текста [5]. Во всяком случае, генетические карты не проявляют какой-либо существенной упорядоченности генов, коррелирующей с их функциями в онтогенезе. Прямой способ задания программ последовательным текстом присущ компьютерам, но не выполняется в рамках молекулярно-генетической системы управления. Программа онтогенеза представлена в архиве опосредованно, через взаимодействие молекулярных компонент системы. Управляемые секции архива (гены) кодируют отдельные макромолекулярные компоненты (белки, РНК)

с определенными характеристиками — каталитическими, структурными, регуляторными и т.д. В результате взаимодействия этих компонентов на основе специфических отношений катализа, опознавания, матричной активности, физического сцепления, пространственного контакта, генетического управления и т.д. формируется динамическая система развития системы управления, реализующая эту программу. Схемы отношений между компонентами системы обычно изображают в виде генетических сетей.

Одним из первых крупных опытов математического и компьютерного моделирования онтогенетической сети был связан с развитием фага λ [6], излюбленного объекта

исследования 60—70-х годов. На основе огромной информации к 1980 г. нам удалось построить компьютерную модель, которая успешно описывала и прогнозировала различные свойства развития фага λ в клетке. Так концепция

МГСУ проявила свой эвристический потенциал. На ее основе мы смогли сформулировать и решить десятки проблем молекулярно-генетической организации [3, 4]: модели оперонных и полирепликонных систем; системные свойства генетического кода; описание свойств генетического языка; динамические модели ансамблей макромолекул, в том числе сайзера; задачи комплементарного анализа; проблемы теории молекулярной эволюции МГСУ; блочно-модульный принцип организации и эволюции МГСУ; принцип лимитирующих факторов организации МГСУ и многое другое. К сожалению, блок-схема МГСУ фага λ

столь сложна, что ее описание выходит за рамки этой статьи.

Эпоха секвенирования

В 1977—1978 гг. в биотехнологии и молекулярной биологии произошло важное событие: появи-

лись достаточно быстрые и удобные методы расшифровки последовательностей (секвенирования) полинуклеотидов. После этого стал стремительно нарастать вал прочитанных последовательностей генов, вирусных геномов, различных фрагментов и т.д. В начале 80-х годов возникли основные международные банки данных нуклеотидных и полипептидных последовательностей, т.е. генетических текстов. Назрела необходимость использовать этот богатейший материал в рамках теории МГСУ, т.е. в компьютерных методах для извлечения генетической информации из моря данных и ее интерпретации.

Опыт нашей работы в рамках концепции МГСУ позволил нам быстро начать собственные, весьма успешные, разработки. Группа сотрудников (В.В.Соловьев, А.А.Жарких, И.Н.Шиндялов, М.П.Пономаренко, А.Э.Кель, И.Б.Рогозин и др.) под руководством Н.А.Колчанова предложила метод сравнительного анализа последовательностей, поиска генов и знаков управления, восстановления пространственной структуры макромолекул, филогенетического анализа и др. [7]. За два последних десятилетия уже проанализированы сотни коллекций и семейств генов, РНК, белков, знаков пунктуации и управления, построены сотни филогенетических деревьев. Исходные общие предположения о принципах организации этой системы управления подтвердились на большом фактическом материале.

Однако стратегия массового секвенирования не сводилась просто к накоплению разрозненных фрагментов генетических текстов. Стали направленно собираться данные о генах и механизмах управления некоторых достаточно сложных клеточных и межклеточных подсистем: иммунитета, кровотока, отклика на тепловой шок, клеточного цикла, SOS-репарации и др. Такие подсистемы характеризуют генными сетями (схемами взаимодействий), на которых базируются динамические модели функционирования. Сейчас этот

раздел моделирования чрезвычайно актуален и продуктивен, поскольку позволяет выявить режимы работы клеточных подсистем и их возможные нарушения.

Генетические взаимодействия в значительной степени реализуются через регуляторные белки, их функциональные центры, знаки управления генов и отчасти — через регуляторные сигналы (метаболиты, гормоны, модифицирующие воздействия). Именно поэтому в последние годы стремительно развиваются исследования функциональных знаков (сайтов) управления транскрипцией — операторов, сайтов узнавания общеклеточных активаторов, энхансеров, инсуляторов, ткане-специфических и гормон-специфических сайтов и др. Выявить такие знаки управления и замкнуть связи в схемах генных сетей можно с помощью компьютерного анализа регуляторных зон генов [8].

С другой стороны, появились новые экспериментальные технологии, позволяющие автоматически следить за динамикой сотен и тысяч генов и их продуктов в ходе функционирования и развития клеток. Такова технология микрочипов ДНК [9]. Сотни и тысячи клонированных фрагментов ДНК с определенными генами распределяют по микролункам панели и гибридизуют с меченой валовой РНК клетки (через так называемую к-ДНК). Благодаря комплементарности цепей в каждой лунке гибридизуется отдельная фракция к-ДНК. Затем автоматически микрометодами измеряют концентрации гибридных молекул. В результате в каждый момент времени можно выписать все активно транскрибируемые гены. Эти данные легко сравнить с моделями генных сетей.

Уже в начале 80-х годов стали известны первые полные геномы вирусов и фагов, среди них — фага λ [10]. В конце 80-х годов нача-

лись крупные международные проекты полного секвенирования клеточных геномов бактерий, грибов, высших животных и растений — кишечной палочки, дрожжей, дро-

зофилы, мыши, пшеницы, человека и др. Сегодня многие из них уже завершены; опубликовано свыше 20 полных клеточных геномов микоплазм, археобактерий, кишечной палочки, возбудителей ряда болезней, а также пекарских дрожжей, маленького червя-нематоды и растения-арабидопсиса. Вероятно, истинное число прочитанных геномов гораздо больше, потому что многие фармацевтические фирмы засекречивают свои результаты, не публикуя их. Совсем скоро станет известен полный геном дрозофилы — классического объекта генетики, появятся окончательные данные о геноме человека. Эти чрезвычайно важные события в молекулярной генетике и биоинформатике означают существенный прорыв знаний, возможностей эксперимента и развития теории МГСУ.

Первые результаты полного секвенирования клеточных геномов наполнили основные модули МГСУ конкретным содержанием. Бактериальные геномы содержат примерно от 500 генов у микоплазм до почти 5 тыс. генов у кишечной палочки. По сравнительным оценкам, клетки в центральной подсистеме должны содержать не менее 300 генов. Анализ генома кишечной палочки выявил 4909 генов, из которых 4288 кодируют белки, но функции 38% из них пока неизвестны. На долю «ядра» приходится свыше 460 известных генов (около 10%), на долю блока контроля метаболизма — свыше 1047 (около 25%), контролирурующих 804 фермента и 988 метаболических реакций, т.е. их сложность сопоставима. Около 20% известных генов связано с клеточными структурами и процессами. Поскольку не все выявленные гены идентифицированы, число их будет возрастать, но доли генов вряд ли существенно изменятся. Кроме того, в геноме кишечной палочки найдено 2584 оперона — управляемых единиц транскрипции (секций архива). Примерно такие же доли генов в основных блоках и у других бактерий с иными размерами геномов.

Подавляющее число генов кишечной палочки входит в автономно управляемые секции архива, имеющие знаки управления и подчиненные регуляторным белкам [11]. На оперон приходится в среднем один-два регуляторных белка и один-два знака управления, т.е. регуляторные зоны генов достаточно просты. В то же время число выявленных и предполагаемых регуляторных белков у кишечной палочки 178, т.е. на порядок меньше, чем число оперонов. Это значит, что опероны скорее образуют ассоциации, управляемые совместно небольшим числом регуляторных белков. Так начинают вырисовываться контуры конструкции клеточных систем управления, в описании которых концепция МГСУ оказалась весьма эффективной.

Сравнительный анализ полных геномов фагов, вирусов и клеточных органелл с клеточными ядерными геномами показывает, что первые обычно не имеют полноценных блоков как контроля метаболизма, так и сайзера, но содержат блок развития и морфологии. Например, у фага ϕ X174 от блока

репликации присутствует только нуклеаза А (блока метаболизма нет), у фага λ — только регуляторные гены (О и Р), но нет блоков транскрипции, трансляции и метаболизма [9]; у фага Т4 есть блок репликации, но нет блоков транскрипции и трансляции [12]. В митохондриях млекопитающих имеется часть системы трансляции, но остальные блоки сайзера и блок метаболизма отсутствуют [13]. В хлоропластах растений присутствует полноценная система трансляции, гены блока транскрипции, но нет системы репликации [14]. Поэтому фаги и органеллы воспроизводятся только с участием клетки-хозяина, они не самостоятельны.

Перспективы

Бурное развитие биоинформатики и геномики последних лет, появление новых комплексных экспериментальных технологий со-

здали новую и весьма продуктивную ситуацию. Полное секвенирование и идентификация генов таких объектов, как вирусы, бактерии, дрожжи, дрозофила, нематода, арабидопсис, человек и др., дает возможность полного моделирования МГСУ клеток, в том числе и в процессе развития. На этой основе ожидается прорыв в медицине, биотехнологии, производстве продуктов питания, фармакологии. Моделирование развития вирусов позволит выявить «слабые» секции, мишени для фармакологического и иммунного воздействия. Модели МГСУ открывают широкие возможности в самых разных направлениях: у бактерий и грибов они помогут найти и оптимизировать пути биосинтеза лекарств, антибиотиков, метаболических продуктов, а также нащупать мишени противодействия многим заболеваниям; у растений — выявить лимитирующие звенья продуктивности и преодолеть их естественными и генноинженерными средствами. У дрозофилы модели МГСУ скорее всего будут незаменимы при изучении фундаментальных процессов генетической организации, онтогенеза, изменчивости и т.п.; у человека они позволят ускорить поиск средств фармакологической и иммунной защиты от заболеваний, генной терапии и т.д. Можно моделировать взаимодействия систем управления паразит—хозяин, патоген—хозяин, симбионт—хозяин, а также онкогенное перерождение.

Среди уже изученных систем управления появляются некоторые экзотические формы, которые могут пролить свет на условия выживания в экстремальных ситуациях: при высоком уровне радиации, высоких или очень низких температурах, в бескислородной среде, в полной темноте, в условиях химического и иного загрязнения и т.д.

В теории молекулярно-генетической системы управления еще не решены многие принципиальные вопросы. Необходимо построение полноценной генетической лингвистики, правил генетического языка с возможным прогнозом мо-

лекулярных функций [3]. Здесь уже имеется ряд крупных открытий, но они не позволяют точно предсказать пространственную структуру белка по его аминокислотной последовательности. Вероятно, кроме ключевых правил («крупных мазков») существует еще «шлейф» весьма специфичных и локальных, число которых очень велико и практически не может быть описано. Поэтому в модели и реальном эксперименте допустим случайный поиск оптимальных решений.

На базе секвенированных геномов и генных сетей необходимо разработать динамические модели онтогенеза, клеточного цикла, канцерогенеза, физиологических функций, иммунного ответа и др. Единую теорию молекулярной эволюции можно построить как теорию эволюции МГСУ и ее подсистем управления, механизмов регуляции, лимитирующих факторов организации и т.д. [4].

Все это означает, что роль теории МГСУ и вообще теоретических методов в молекулярной генетике, а также в смежных областях (теории молекулярной эволюции, генетике развития, вирусологии, генной

инженерии и биотехнологии) и их приложении будет возрастать. Во многих случаях эта теория становится ключом решения краевых проблем и неперенной составляющей каждодневной исследовательской работы. ■

Литература

1. *Ратнер В.А.* // Генетика. 1992. Т.28. №2. С.5—24.
2. *Schrdinger E.* What Is Life? The Physicists Viewpoint. Cambridge, 1944.
3. *Ратнер В.А.* Молекулярно-генетические системы управления. Новосибирск, 1975; *Он же.* Концепция молекулярно-генетических систем управления. Новосибирск, 1993.
4. *Atlan H., Koppel M.* // Bull. Math. Biol. 1990. V.52. №3. P.335—348.
5. *Ratner V.A., Zharkikh A.A., Kolchanov N.A. et al.* Molecular Evolution. Berlin, 1996.
6. *Kananyan G.Kh., Tchuraev R.N., Ratner V.A.* // J. Theor. Biol. 1981. V.90. P.301—315.
7. Modeling and Computer Methods in Molecular Biology & Genetics / Eds V.A.Ratner, N.A.Kolchanov. N.Y., 1992.
8. Proceedings of the 2nd International Conference on Bioinformatics of Genome Regulation and Structure (BGRS-2000). Novosibirsk, 2000.
9. *Risi J.L. de, Iyer V.R.* // Nature Genetics. 1999. V.21. Suppl.1 P.33—37.
10. *Sanger F. et al.* // J. Mol. Biol. 1982. V.163. P.729—773.
11. *Blattner F. et al.* // Science. 1997. V.277. P.1453—1462.
12. Molecular Biology of Bacteriophage T4 / Ed. J.D.Karam. Washington, 1994.
13. *Brooke J.L.* // Nuclear Acids Research. 1999. V.27. P.1767—1780.
14. *Umesono K., Ozeki H.* // Trends in Genetics. 1987. V.3. №10. P.281—287.