



# БИОИНФОРМАТИКА В ЭКОЛОГИЧЕСКОЙ ГЕНЕТИКЕ

А.В. Кочетов, Н.А. Омелянчук,  
Е.В. Игнатъева, В.А. Лихошвай,  
Ю.Г. Матушкин

Институт цитологии и генетики  
СО РАН; Новосибирский  
государственный университет,  
г. Новосибирск

✿ В обзоре рассмотрены информационные ресурсы (базы данных генных сетей, транскрипционных и посттранскрипционных сигналов экспрессии, программы для распознавания функционально-активных сайтов и предсказания их активности) и обсуждается их возможное использование в генной инженерии для оптимизации экспрессии чужеродных генов и создания организмов с новыми свойствами.

✿ Ключевые слова: трансгенез, оптимизация экспрессии, генетически-модифицированные организмы, биоинформатика, базы данных.

## БИОИНФОРМАТИКА И ТРАНСГЕНЕЗ — СОЗДАНИЕ ОРГАНИЗМОВ С НОВЫМИ СВОЙСТВАМИ

Генная инженерия является одной из наиболее активно развивающихся областей современной биотехнологии. Генетически модифицированные организмы (ГМО), не так давно вошедшие в практику, широко применяются в сельском хозяйстве и их удельный вес в производстве сои, хлопчатника и многих других культур постоянно увеличивается. В настоящее время культивируются генетически модифицированные растения, несущие один или два трансгена и их преимущества уже достаточно очевидны. По-видимому, перенос больших комплексов генов не только расширит возможности улучшения существующих сортов, но и позволит изменять характеристики растений качественно — вплоть до получения новых форм с заранее определенными свойствами.

В настоящее время существует ряд нерешенных проблем, затрудняющих получение ГМО. Для эффективного трансгенеза необходимо обеспечивать адекватное управление экспрессией перенесенного гена. Чужеродные гены часто содержат комбинации нуклеотидов, которые могут распознаваться в клетках организма-реципиента как сигналы экспрессии. Даже при ограниченном использовании чужеродной ДНК — когда переносится только белок-кодирующая часть — в ней могут содержаться различные ложные сигналы, например, сайты полиаденилирования [7, 10] или сайты сплайсинга [13, 35], могут содержаться неоптимальные (редкие) кодоны [22, 41] и т. д. Для того чтобы экспрессия трансгена осуществлялась по заданному типу, необходимо уметь выявлять негативные элементы и вносить в нуклеотидную последовательность соответствующие модификации [24]. Это особенно существенно в тех случаях, когда планируется перенести несколько трансгенов, характеристики экспрессии которых должны быть точно сопряжены, или экспрессия трансгенов должна коррелироваться с определенными метаболическими состояниями клеток или фазами роста [6, 9]. Очевидно, что здесь возможности генной инженерии принципиально ограничены уровнем знаний в молекулярной генетике и молекулярной биологии, поскольку охарактеризована только некоторая часть сигналов экспрессии. Однако использование методов и ресурсов биоинформатики способно существенно увеличить возможности планирования генно-инженерных экспериментов и — в определенных случаях — отчасти заменить отсутствие точных знаний проверяемыми гипотезами. В статье рассмотрены некоторые информационные ресурсы (базы данных генных сетей, транскрипционных и посттранскрипционных сигналов экспрессии, программы для распознавания функционально-активных сайтов и предсказания их активности) и обсуждается их возможное использование в генной инженерии для оптимизации экспрессии чужеродных генов.

## БАЗЫ ДАННЫХ СИГНАЛОВ ЭКСПРЕССИИ

Для обеспечения необходимого характера экспрессии перенесенного гена необходимо использовать адекватные регуляторные последовательности и сигналы экспрессии. В настоящее время активно используется ограниченный набор промоторов, что, очевидно, недостаточно для решения большинства сложных биотехнологических задач. В научной литературе опубликовано огромное количество экспериментальных данных о промоторах и сигналах транскрипционного контроля различных генов. Следует отметить, что в статьях приводятся данные о транскрипционной активности производных промоторов — например, делеционных мутантах, промоторах с измененным порядком следования элементов и т. д., характер транскрипции которых может отличаться от исходного варианта. Используя эту информацию можно с большой вероятностью подобрать промотор с нужным типом транскрипции даже в тех случаях, когда необходимо обеспечить высокую специфичность экспрессии трансгена.

Для подбора подходящего промотора можно использовать специализированные базы данных — например, TRRD [21], PLACE [14], Plant CARE [26], TRANSFAC [40]. TRRD (Transcription Regulatory Regions Database; <http://www.bionet.nsc.ru/trrd/>), разработанная в Институте цитологии и генетики СО РАН, представляет собой один из наиболее продвинутых информационных ресурсов этого типа и включает комплексное описание регуляции транскрипции генов эукариот во всех органах, тканях, и типах клеток организма и клеточных линиях [21]. Каждая карточка БД соответствует определенному гену и содержит данные о локализации и функционировании транскрипционных регуляторных районов: сайтов связывания транскрипционных факторов, регуляторных единиц (промоторов, энхансеров, сайленсеров), *locus control regions* (LCR), а также данные по экспрессии гена с указанием типа и названия регуляторной последовательности, обеспечивающей реализацию конкретного характера экспрессии. Информация распределена в TRRD по шести взаимосвязанным базам данных (TRRDSITES, TRRDUNITS, TRRDLCR, TRRDFACTORS, TRRDEXP, TRRDGENES):

— TRRDSITES содержит сведения о регуляторных элементах первого уровня, которыми являются сайты связывания транскрипционных факторов;

— TRRDUNITS описывает следующий уровень организации — регуляторные единицы (промоторы, энхансеры, сайленсеры), включающие функционально связанные наборы сайтов; регуляторные единицы могут быть локализованы в различных областях гена — 5'-фланкирующих, 3'-фланкирующих, экзонах, а также интронах;

— TRRDLCR включает структурно-функциональные характеристики следующего регуляторного уровня — LCR;

— TRRDFACTORS включает данные о транскрипционных факторах, взаимодействующих с сайтами;

— TRRDEXP включает данные о качественных особенностях экспрессии генов в зависимости от стадии развития организма, стадии клеточного цикла, вида клеток и этапа их дифференцировки, внешних воздействий и т. д., которые накапливаются в форме паттернов экспрессии;

— TRRDGENES содержит информацию, идентифицирующую ген, а также иерархически организованное представление регуляторных элементов всех уровней.

Седьмая база TRRDBIB содержит библиографические данные.

Отличительной особенностью базы данных TRRD является то, что она содержит только экспериментально подтвержденную информацию. Данные заносятся в TRRD на основе аннотации статей, содержащих описание результатов экспериментов различных типов. Базовым программным средством, обеспечивающим поиск и навигацию по TRRD, а также интеграцию TRRD с внешними информационными и программными ресурсами является *Sequence Retrieval System* (SRS). Предусмотрена возможность поиска нуклеотидной последовательности по гомологии с помощью BLAST. В TRRD поддерживаются браузеры по видам и названиям генов, а также тематические разделы.

В тематическом разделе Plant-TRRD представлена информация по структурно-функциональной организации транскрипционных регуляторных районов генов растений. Раздел составлен на основе анализа 728 экспериментальных статей и содержит данные о 276 генах растений, 430 регуляторных единицах (промоторах, энхансерах, сайленсерах и др.) и 977 сайтов связывания транскрипционных факторов, локализованных в пределах этих регуляторных единиц. В Plant-TRRD содержится 1166 паттернов экспрессии, описывающих экспрессию генов растений в различных клетках и тканях на разных стадиях развития. На рис. 1 представлен фрагмент интерфейса стандартной формы запроса к базе TRRD через поисковую систему SRS. Пользователь может осуществлять поиск регуляторных элементов с заданным типом транскрипционной активности, формируя запросы по интересующим его ключевым словам (например, в информационных полях «*tissue*» и «*organ*» таблицы TRRDEXP).

Помимо сигналов транскрипционного контроля, важное значение может иметь оптимизация экспрессии трансгена на посттранскрипционном уровне. В составе эукариотических мРНК часто содержатся сигналы, контролирующие эффективность трансляции или цитоплазматическую стабильность матрицы (как общую, так и

*separate multiple values by & (and), | (or),*

Tissue	▼	meristem
Organ	▼	flower
ExpressionPatternAC	▼	
ExpressionPatternAC	▼	

*retrieve entries offy*

Use predefined view  \* Names only \*

Create your own view

Select fields to display:

- Organ
- Tissue
- Cells
- StageCellDiff
- ExpressionLevel
- IndReprName
- InductionTime

Рис. 1. Фрагмент интерфейса стандартной формы запроса к базе TRRD в поисковой системе SRS: можно видеть информационные поля, содержащие информацию о паттерне экспрессии транскрипционно-активной последовательности

стадии-, ткане- или стресс-специфическую). Применение таких сигналов в дизайне трансгена может существенно увеличить эффективность трансгенеза [3, 9]. Кроме того, в составе чужеродной ДНК могут содержаться элементы, сходные с посттранскрипционными сигналами экспрессии растительных генов: обнаружение и при необходимости элиминация таких сигналов может оказаться необходимым условием для достижения нужного характера экспрессии трансгена.

В настоящее время существует несколько БД, содержащих информацию об эукариотических посттранскрипционных сигналах экспрессии. Например, структура БД TransTerm [12] и UTRsite [33] позволяет проводить поиск нескольких десятков сигналов в мРНК, интересующих пользователя. Однако в настоящее время известна полная структура очень небольшого числа таких сигналов, что существенно ограничивает применимость таких информационных ресурсов. В ИЦиГ СО РАН была разработана БД TRSIG [17], созданная на основе другого подхода. В этой БД (<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/>; mirror site: <http://dna01.bse.kyutech.ac.jp/jouhou/trsig/trsig.html>) собрана информация о потенциальных сигналах — участках мРНК, которые, согласно экспериментальным данным, способны влиять на эффективность

трансляции или на цитоплазматическую стабильность матрицы. Таким образом, в TRSIG содержатся не только сигналы, структура которых была досконально изучена, но и нуклеотидные последовательности — регуляторы трансляции, активные элементы которых еще не определены.

БД TRSIG реализована на платформе SRS и состоит из четырех баз данных: 1) БД объектов (содержит общую информацию о типе трансляционного сигнала); 2) БД экспериментов (описывает условия эксперимента и активность экспериментальных последовательностей); 3) БД экспериментальных нуклеотидных последовательностей (фрагменты мРНК, активность которых изучалась в экспериментах); 4) БД полноразмерных экспериментальных мРНК (предназначена для реконструкции протяженных вторичных структур). Карточки этих БД содержат перекрестные ссылки, что при использовании возможностей SRS позволяет находить нужные сведения. В настоящее время БД содержит информацию о 25 видах трансляционных сигналов, 87 экспериментах и 161 экспериментальных нуклеотидных последовательностях. TRSIG может использоваться для получения как общей, так и углубленной информации о сигналах экспрессии, локализованных в мРНК. В TRSIG также предусмотрена возможность поиска гомологичных участков между фрагментами мРНК, содержащими посттранскрипционные сигналы экспрессии (БД экспериментальных нуклеотидных последовательностей), трансляционная активность которых была определена экспериментально (и описана в БД экспериментов), и мРНК, интересующей пользователя. С этой целью использован классический BLAST алгоритм. Наличие локальной гомологии с участком, содержащим пост-транскрипционный сигнал, можно интерпретировать исходя из имеющихся в TRSIG совокупности данных об относительной активности мутационных вариантов этой последовательности.

Таким образом, компьютерные ресурсы TRRD и TRSIG могут быть использованы в качестве источника информации о служебных нуклеотидных последовательностях, обеспечивающих адекватный контроль экспрессии трансгена на транскрипционном и посттранскрипционном уровнях.

#### ОПТИМИЗАЦИЯ СТРУКТУРЫ ТРАНСГЕНА

Помимо выбора подходящих регуляторных элементов (промотора и сигналов посттранскрипционного контроля), для достижения эффективного взаимодействия с аппаратом экспрессии организма-реципиента часто бывает необходимо модифицировать структуру трансгена. Обычно в трансгенезе используют только белок-кодирующую часть (БКП), хотя в ряде случаев было показано, что интрон (как правило, выделенный из одного

из генов организма-реципиента и перенесенный в БКП трансгена) существенно усиливает уровень экспрессии [25]. Однако БКП чужеродного гена также может обладать негативными характеристиками, снижающими эффективность экспрессии. Можно выделить два вида таких характеристик: 1) параметры мРНК, влияющие на эффективность взаимодействия с аппаратом трансляции организма-реципиента; 2) элементы нуклеотидной последовательности, сходные с сигналами экспрессии организма-реципиента и неправильно воспринимаемые как такие сигналы (например, ложные сайты сплайсинга или полиаденилирования).

Известно, что некоторые параметры растительных мРНК влияют на трансляционную активность [1]. Показано, что 5'-НТП опосредует взаимодействие аппарата инициации трансляции и рибосомы [23]. Лидерный район может содержать как негативные (стабильные шпильки, AUG триплеты и рамки считывания), так и позитивные (трансляционные энхансеры) сигналы. К числу известных сигналов общего характера относят контекст стартового кодона трансляции, влияющий на его распознавание рибосомами [29, 30, 36]. В ряде случаев 5'-НТП содержит сигналы, опосредующие инициацию трансляции по механизмам внутренней инициации трансляции [34]. По-видимому, лидерная последовательность может также влиять на цитоплазматическую стабильность мРНК [3, 11]. В настоящее время критерии, позволяющие определять функциональную активность лидерной последовательности, исследованы недостаточно.

Для обеспечения эффективной инициации трансляции мРНК трансгена часто в качестве 5'-НТП используют трансляционные энхансеры (информация о них представлена в TRSIG). Кроме этого, необходимо оптимизировать контекст стартового кодона трансляции [29, 36] и оценить стабильность вторичной структуры 5'-концевого района мРНК (включающего 5'-НТП и начало CDS). Ранее нами было показано, что характеристики мРНК высокоэкспрессирующихся генов оптимизированы для обеспечения эффективной трансляции [18]. На основе выявленных закономерностей была создана компьютерная система *Leader\_RNA*, позволяющая оценивать трансляционную активность 5'-НТП мРНК растений и млекопитающих [19]. Эта система может быть также использована для оценки функциональной активности 5'-НТП мРНК трансгена в клетках двудольных и однодольных растений. На рис. 2 изображено окно результатов программы *Leader\_RNA*; в качестве последовательности для анализа была использована 5'-НТП мРНК одного из высокоэкспрессирующихся генов человека. В данном случае предсказанная активность этой лидерной последовательности в клетках двудольных растений оказалась низкой, что иллюстрирует важность учета видоспецифических осо-

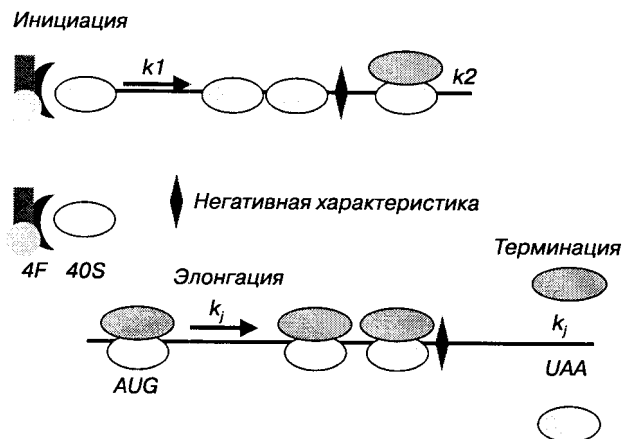


Рис. 2. Модель трансляции эукариотических мРНК.

Представлены стадии инициации (согласно модели лимитирующего звена), элонгации и терминации. Обозначение:  $k_1$ ,  $k_2$ ,  $k_3$  — константы скоростей различных этапов трансляционного процесса; 4F — фактор инициации трансляции eIF-4F; 40S — субъединица рибосомы

бенностей организации генов растений в планировании генно-инженерного эксперимента.

Известно, что кодонный состав влияет на скорость элонгации трансляции у многих прокариот и дрожжей [31]. Однако для мРНК растений и млекопитающих роль этих факторов в определении скорости элонгации трансляции до настоящего времени не изучена. Получены некоторые данные о взаимосвязях между кодонным составом генов растений и их функцией или уровнем экспрессии [5, 8], но в целом эта проблема также исследована недостаточно и выявленные закономерности достаточно слабы. По-видимому, существенным фактором может оказаться взаимосвязь между локализацией редких кодонов и структурой белка [22] или стабильностью мРНК [41].

Несмотря на отсутствие доказательств взаимосвязи между кодонным составом и эффективностью экспрессии, оптимизацию кодонного состава активно используют для повышения уровня экспрессии трансгенов [3, 16]. При этом в качестве оптимальных кодонов используют наиболее часто встречающиеся синонимические кодоны генов организма-реципиента (таблицы средних частот кодонов доступны, например, в БД *TransTerm* [12]). Считается, что такие модификации могут повысить эффективность экспрессии трансгена. Однако реорганизация контекстных характеристик БКП может влиять и на другие параметры экспрессии трансгена: например, при этом могут элиминироваться ложные сигналы сплайсинга и полиаденилирования. Это весьма вероятно, поскольку чужеродные гены часто существенно отличаются от генов организма-реципиента по нуклеотидному составу.

По-видимому, перспективным подходом для предсказания трансляционной активности мРНК является ком-

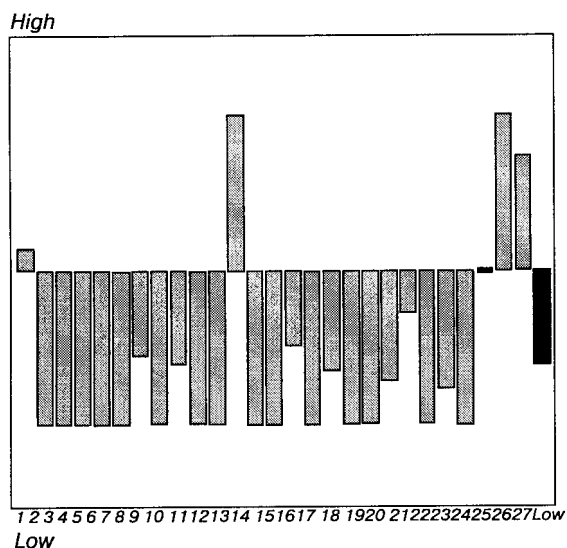


Рис. 3. Фрагмент интерфейса окна результатов программы *Leader\_RNA*.

Программа анализирует нуклеотидную последовательность 5'-НТП мРНК, введенную пользователем, и сравнивает ее характеристики и характеристики лидерных районов высоко- и низкоэкспрессирующихся генов двудольных и однодольных растений или млекопитающих (по выбору пользователя). Каждая характеристика представлена на диаграмме в виде отдельного столбца, значение которого варьируется от -1 (низкая экспрессия) до +1 (высокая экспрессия). Общая оценка трансляционной активности (крайний правый столбец диаграммы) производится по методу «усредненного распознавания»

пьютерное моделирование трансляции [2, 27, 39]. Однако до настоящего времени не разработаны математические модели, позволяющие описывать динамику процесса трансляции с учетом контекстных и структурных особенностей эукариотических мРНК и рассматривающие три основных этапа трансляции: инициацию, элонгацию и терминацию, каждая из которых характеризуется специфическим набором параметров (рис. 3). Вероятно, применение обобщенного химико-кинетического подхода позволит создать адекватную модель [28]. В рамках данного подхода процессы инициации, элонгации и терминации трансляции описываются в терминах элементарных процессов, таких как моно- и бимолекулярные реакции, уравнения Михаэлиса–Ментен и т. д. Параметрами модели являются кодонные составы всех рамок трансляции, константы скоростей протекания элементарных процессов, например параметры процесса размещения заряженной изоакцепторной тРНК в А-сайте рибосомы, или параметры сборки трансляционного комплекса на индивидуальном сайте инициации трансляции индивидуальной мРНК и т. д.

Нами была выполнена одна из реализаций модели трансляции такого вида. Для компьютерной реализации использовались численные методы решения систем дифференциальных уравнений, а также имитация стохастических

процессов. Последний выбор объясняется тем, что для адекватного описания процесса элонгации на индивидуальной мРНК необходимо учитывать тот факт, что рибосомы в процессе движения вдоль мРНК не могут менять порядок следования друг за другом. Это приводит к тому, что количество переменных, требуемых для описания процесса элонгации в терминах дифференциальных уравнений, вытекающих из химической кинетики, достигает фантастической величины. Например, чтобы описать элонгацию кодирующего участка мРНК из 300 кодонов требуется около  $10^{30}$  переменных. Таким образом, аппарат дифференциальных уравнений оказывается малоприменимым для описания процессов трансляции, если ставится задача научиться вычислять параметры трансляции мРНК с учетом ее нуклеотидного и кодонного состава.

Поэтому нами был применен комбинированный подход. Описание процессов трансляции проводилось в терминах химической кинетики, а компьютерное моделирование проводилось в терминах стохастики. Этот подход, во-первых, позволяет сохранить основное достоинство обобщенного химико-кинетического метода моделирования — его гибкость, без которой невозможно решение основной задачи — построение адекватной математической модели трансляции. Во-вторых, снимается проблема размерности модели. В этом случае количество переменных, требуемых для описания процесса элонгации трансляции на одной молекуле мРНК, зависит от ее длины линейно.

На данном этапе развития модель описывает процесс элонгации индивидуальной молекулы мРНК с учетом ее кодонного состава. Среднее время экспонирования изоакцепторной мРНК в А-сайте рибосомы зависит от конкретного вида экспонированного в нем кодона, точнее от частоты его использования (рис. 4). Запрещено сближение транслирующих рибосом на расстояние, меньшее чем стерический размер рибосомы. Данный вариант модели уже позволяет рассчитывать многие важные характеристики элонгации индивидуальных эукариотических мРНК с учетом их кодонного состава, интенсивности поступления рибосом на мРНК и интенсивности их диссоциации с мРНК: рассчитывать среднюю скорость элонгации, оценивать среднее расстояние между рибосомами в составе полисомы, выявлять места кластеризации рибосом, рассчитывать элонгационные профили индивидуальных мРНК, сравнивать элонгацию различных мРНК между собой, в том числе проводить оценку влияния синонимичных замен, выявлять лимитирующие звенья и т. д. В результате последовательного развития модели будет достигаться все более полное описание процесса трансляции, что приведет к расширению круга задач, которые можно будет ставить и решать с применением математического моделирования.

Поскольку контекстные характеристики трансгена характерны для организма-донора и могут отличаться от ти-

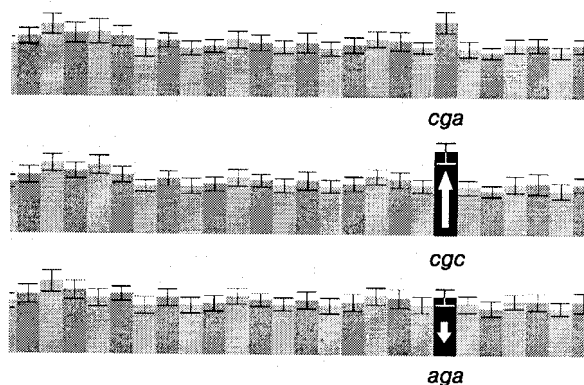


Рис. 4. Стохастическое моделирование процесса трансляции мРНК. 25-й кодон CGA (кодирующий аргинин) имеет низкую скорость прохождения рибосомы.

С помощью моделирования показано, что замена кодона cga на синонимические варианты — кодоны cgc (болсс редкий) и aga (наиболее часто встречающийся) повышают эффективность трансляции мРНК в целом

личных характеристик генов организма-реципиента, то некоторые комбинации нуклеотидов могут восприниматься системой экспрессии клеток организма-хозяина как сигналы экспрессии. Поэтому одна из задач, решение которой часто необходимо для эффективного трансгенеза, заключается в элиминации таких ложных сигналов сплайсинга [13, 35] и полиаденилирования [7, 10]. В настоящее время существует несколько компьютерных программ, сделанных на основе информации о геноме арабидопсиса и позволяющих предсказывать сайты сплайсинга [4, 38], однако эффективность их применения для других видов растений может быть ниже. К сожалению, методы для распознавания потенциальных сайтов полиаденилирования для генов растений пока не разработаны.

### ДИЗАЙН ГМО

Генно-инженерные модификации могут быть использованы для решения широкого круга биотехнологических и селекционно-генетических задач. По-видимому, возможности изменения самых разных морфофизиологических или биохимических характеристик растений и животных ограничены только уровнем наших знаний о межгенных взаимодействиях и генных сетях. Планирование генно-инженерного эксперимента должно в первую очередь заключаться в построении схемы модификаций генома, которые должны привести к желаемому эффекту на уровне фенотипа. Это сложная задача, для решения которой необходимо проанализировать большой объем информации в различных областях биологии. Фактически необходимо научиться определять, какие генно-инженерные модификации — будь то перенос трансгенов, инактивация или модуляция характера экс-

прессии генов растения-реципиента или комбинации различных воздействий приведут к желаемому результату. Одним из перспективных подходов, позволяющих решать проблему планирования генно-инженерного эксперимента, являются базы данных генных сетей, сигнальных и метаболических путей [15, 20, 32, 37], в которых накапливаются и систематизируются сведения о регуляции экспрессии генов, путей передачи сигналов и об участии различных ферментов и субстанций в процессе синтеза того или иного вещества в организме. Генная сеть — это группа координировано функционирующих генов, обеспечивающих формирование определенного фенотипического признака организма (молекулярного, биохимического, физиологического, морфологического, поведенческого и т. д.). В качестве образца БД генных сетей можно привести компьютерную систему PlantGeneNet. В настоящее время в PlantGeneNet представлены три генные сети, различающиеся по основным механизмам их регуляции. Так, в основе генной сети фотоморфогенеза ключевую роль играют негативные обратные связи, а генная сеть развития цветка арабидопсиса имеет несколько контуров позитивных обратных связей наряду с каскадной активацией генов (рис. 5). В генной сети образования клубеньков гены растения хозяина и бактерии интегрированы в единую гибридную генную сеть с общими путями передачи сигналов, метаболизмом и координированной регуляцией активности генов у обоих симбиотических партнеров. PlantGeneNet имеет также дополнительную базу PlantGeneNet SDB, где собраны экспериментальные данные по морфогенезу и экспрессии генов у растений как в норме, так и при различных мутациях. Эта реляционная база данных состоит из 5 субБД: база данных по генам, база данных по экспрессии генов, база данных по развитию определенных органов в норме, база данных по фенотипическим эффектам отдельных мутаций в развитии определенных органов и база данных ссылок на публикации. Таким образом, GeneNet SDB в дополнение к GeneNet: 1) предоставляет подробные данные, на которых основана топология генной сети; 2) интегрирует знания о мутации, произошедшей в определенном гене, через изменения активности других генов, как последствия этой мутации, до мутантного фенотипа, как отражения всего комплекса этих изменений.

Очевидно, что БД генных сетей могут быть использованы в качестве источника базовой информации при планировании ГМО, поскольку содержат данные, необходимые для выбора направления генетической модификации. Огромные возможности представляет экспериментаторам изменение паттернов экспрессии генов организма-реципиента с помощью механизмов генетического сайленсинга или косупрессии. Перенос чужеродных генов, кодирующих отсутствующие у организма ферменты, позволяет модифицировать метаболические цепи, изменять спектр и соотношение липидов, углево-

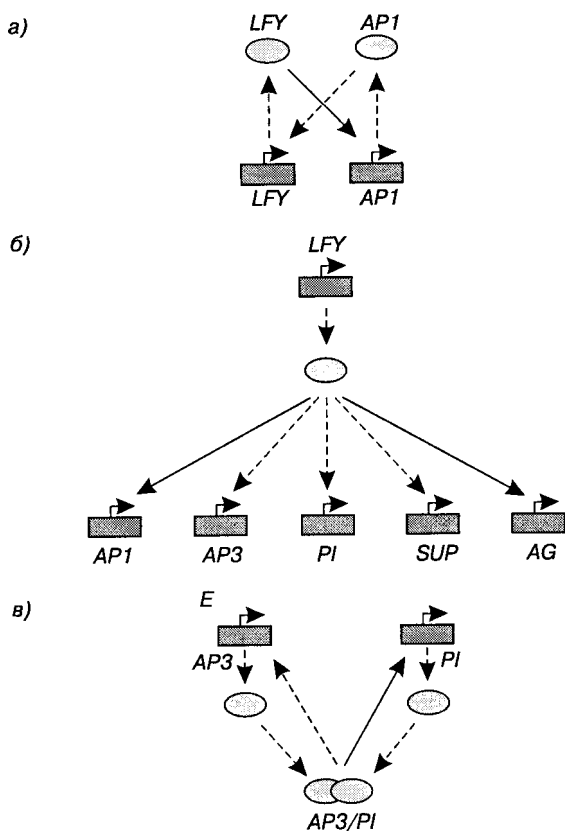


Рис. 5. Основные схемы некоторых контуров позитивных обратных связей и каскадной активации генов в генной сети развития цветка арабидопсиса.

а — взаимодействие между генами APETALA1 и LEAFY в ходе инициации меристемы цветка; б — активация с участием гена LEAFY, генов APETALA1, APETALA3, PISTILLATA, AGAMOUS и SUPERMAN; в — позитивная регуляция активности генов APETALA3 и PISTILLATA с участием AP3/PI гетеродимера

дов, различных вторичных метаболитов. Таким образом, БД генных сетей — по мере их появления и развития — будут использоваться в качестве важного компонента информационного обеспечения генно-инженерного эксперимента, поскольку позволяют эффективно осуществлять выбор генов-мишеней организма-хозяина для модификации экспрессии или трансгенов, которые должны быть перенесены в геном модифицируемого организма.

**Литература**

1. Кочетов А.В., Шумный В.К. Влияние структуры мРНК на процесс инициации трансляции в клетках растений // Успехи совр. биол. — 1998. — Vol. 118. — P. 754–770.
2. Bagnoli F., Lio P. Selection, mutations and codon usage in bacterial model // J. Theor. Biol. — 1995. — Vol. 173. — P. 71–281.
3. Biemelt S., Sonnewald U., Galmbacher P., Willmitzer L. et al. Production of human papillomavirus type 16 virus-like particles in transgenic plants // J. Virol. — 2003. — Vol. 77. — P. 9211–9220.
4. Brendel V., Kleffe J. Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in Arabidopsis

- thaliana genomic DNA // Nucleic Acids Res. — 1998. — Vol. 26. — P. 4748–4757.
5. Chiapello H., Lisacek F., Caboche M., Henaut A. Codon usage and gene function are related in sequences of Arabidopsis thaliana // Gene. — 1998. — Vol. 209. — GC1–GC38.
6. DellaPenna D. Plant metabolic engineering // Plant Physiol. — 2001. — Vol. 125. — P. 160–163.
7. Diehn S.H., Chiu W.-L., De Rocher E.J., Green P.J. Premature polyadenylation at multiple sites within a Bacillus thuringiensis toxin gene-coding region // Plant Physiol. — 1998. — Vol. 117. — P. 1433–1443.
8. Duret L., Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis // Proc. Natl. Acad. Sci. USA. — 1999. — Vol. 96. — P. 4482–4487.
9. Gallie D.R. Controlling gene expression in transgenics // Curr. Opin. Plant Biol. — 1998. — Vol. 1. — P. 166–172.
10. Grec S., Wang Y., Le Guen L. et al. Cryptic polyadenylation sites within the coding sequence of three yeast genes expressed in tobacco // Gene. — 2000. — Vol. 242. — P. 87–95.
11. Gutierrez R.A., Macuntosh G.C., Green P.J. Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control // Trends Plant Sci. — 1999. — Vol. 4. — P. 429–438.
12. Jacobs G.H., Rackham O., Stockwell P.A. et al. TransTerm: a database of mRNAs and translational control elements // Nucleic Acids Res. — 2002. — Vol. 30. — P. 310–311.
13. Haseloff J., Siemering K.S., Prasher D.C., Hodge S. Removal of cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly // Proc. Natl. Acad. Sci. USA. — 1997. — Vol. 94. — P. 2122–2127.
14. Higo K., Ugawa Y., Iwamoto M., Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database:1999 // Nucleic Acids Res. — 1999. — Vol. 27. — P. 297–300.
15. Kanehisa M., Goto S., Kawashima S., Nakaya A. The KEGG databases at GenomeNet // Nucleic Acids Res. — 2002. — Vol. 30. — P. 42–46.
16. Khanna H.K., Raina S.K. Elite Indica transgenic rice plants expressing modified Cry1Ac endotoxin of Bacillus thuringiensis show enhanced resistance to yellow stem borer (Scirpophaga incertulas) // Transgenic Res. — 2002. — Vol. 11. — P. 411–423.
17. Kochetov A.V., Grigorovich D., Kolchanov N.A., Sarai A. Database on mRNA located eukaryotic expression signals influencing translation efficiency and specificity // Genome Informatics. Ser. 12 / Edt: Matsuda et al. Universal Academy Press: Tokyo, 2001. — P. 492–493.
18. Kochetov A.V., Ischenko I.V., Vorobiev D.G., Kel A.E. et al. Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features // FEBS Lett. — 1998. — Vol. 440. — P. 351–355.
19. Kochetov A.V., Ponomarenko M.P., Kisselev L.L., Kolchanov N.A. Prediction of eukaryotic mRNA translational properties // Bioinformatics. — 1999. — Vol. 15. — P. 704–712.
20. Kolchanov N.A., Ananko E.A., Likhoshvai V.A. et al. Gene networks description and modeling in the GeneNet system // Gene Regulation and Metabolism. Postgenomic Computational Approaches / Eds Collado-Vides J., Hofstadt R. Cambridge. London: MIT Press, 2000. — P. 149–179.
21. Kolchanov N.A., Ignatieva E.V., Ananko E.A. et al. Transcription Regulatory Regions Database (TRRD): its status in 2002. Nucleic Acids Res. 2002a. — Vol. 30. — P. 312–317.
22. Komar A.A., Lesnik T., Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation // FEBS Lett. — 1999. — Vol. 462. — P. 387–391.
23. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation // Gene. — 2002. — Vol. 299. — P. 1–34.
24. Larrick J.W., Thomas D.W. Producing proteins in transgenic plants and animals // Curr. Opin. Biotechnol. — 2001. — Vol. 12. — P. 411–418.

25. *Le Hir H., Nott A., Moore M.J.* How introns influence and enhance eukaryotic gene expression // *Trends Biochem. Sci.* — 2003. — Vol. 28. — P. 215–220.
26. *Lescot M., Dehais P., Thijs G.* et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences // *Nucleic Acids Res.* — 2002. — Vol. 30. — P. 325–327.
27. *Likhoshvai V.A., Matushkin Yu.G.* Theoretical analysis of possible evolutionary trends in codon distribution along the mRNA // *Proc. 1st Intern. Conf. of Bioinformatics of genome regulation and structure / Ed. N.A. Kochanov / Novosibirsk, 1998.* — Vol. 1. — P. 341–344.
28. *Likhoshvai V.A., Matushkin Yu.G., Vatolin Yu.N., Bazhan S.I.* A generalized chemical kinetic method for simulating complex biological systems. A computer model of ? phage ontogenesis // *Computational technol.* — 2000. — Vol. 5. — P. 87–99.
29. *Lukaszewicz M., Feuermann M., Jerouville B., Stas A., Boutry M.* In vivo evaluation of the context sequence of the translation initiation codon in plants // *Plant Sci.* — 2000. — Vol. 154. — P. 89–98.
30. *Matushkin Yu.G., Likhoshvai V.A., Kochetov A.V.* Local secondary structure may be a critical characteristic influencing translation of unicellular organisms mRNA. // *Bioinformatics Of Genome Regulation And Structure / Ed. By N. Kolchanov and R. Hofstaedt. Kluwer Academic Publishers: Boston/Dordrecht/London.* — 2003. — P. 103–114.
31. *McCarthy J.E.G.* Posttranscriptional control of gene expression in yeast // *Microbiol Mol. Biol. Rev.* — 1998. — Vol. 62. — P. 1492–1553.
32. *Mueller L.A., Zhang P., Rhee S.Y.* AraCyc. A Biochemical Pathway Database for Arabidopsis // *Plant Physiol.* — 2003. — Vol. 132. — P. 453–460.
33. *Pesole G., Liuni S., Grillo G.* et al. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002 // *Nucleic Acids Res.* — 2002. — Vol. 30. — P. 335–340.
34. *Pestova T.V., Kolupaeva V.G., Lomakin I.B.* et al. Molecular mechanisms of translation initiation in eukaryotes // *Proc. Natl. Acad. Sci. USA.* — 2001. — Vol. 98. — P. 7029–7036.
35. *Rouwendaal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.* Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage // *Plant Mol. Biol.* — 1997. — Vol. 33. — P. 989–999.
36. *Sawant S.V., Kiran K., Singh P.K., Tuli R.* Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants // *Plant Physiol.* — 2001. — Vol. 126. — P. 1630–1636.
37. *Schacherer F., Choi C., Gotze U.* et al. The TRANSPATH signal transduction database: a knowledge base on signal transduction networks // *Bioinformatics.* — 2001. — Vol. 17. — P. 1053–1057.
38. *Tolstrup N., Rouze P., Brunak S.* A branch point consensus from Arabidopsis found by non-circular analysis allows for better prediction of acceptor sites // *Nucleic Acids Res.* — 1997. — Vol. 25. — P. 3159–3163.
39. *Tomita M.* Whole-cell simulation: a grand challenge of 21<sup>st</sup> century // *Trends Biotechnol.* — 2001. — Vol. 19. — P. 205–210.
40. *Wingender E., Chen X., Fricke E.* et al. The TRANSFAC system on gene expression regulation // *Nucleic Acids Res.* — 2001. — Vol. 29. — P. 281–283.
41. *Van Hoof A., Green P.J.* Rare codons are not sufficient to destabilize a reporter gene transcript in tobacco // *Plant Mol. Biol.* — 1997. — Vol. 35. — P. 383–387.

**Bio-informatics and trans-genesis are creating organisms with new features**

*A.V. Kochetov, N.A. Omelianchuk, E.V. Ignatieva, V.A. Likhoshvai, Yu.G. Matushkin*  
Institute of cytology and genetics CO RAS; Novosibirsk state university.

✿ THE SUMMARY: Bioinformatic resources (databases of the gene nets, transcriptional and posttranscriptional expression signals, programs for recognition of functional sites and prediction of their activities) are reviewed in respect of their utility for gene engineering experiments.

✿ KEY WORDS: transgenesis, optimization of expression, genetically-modified organisms, bioinformatics, databases.