

УДК 577.2:602.6

СТРАТЕГИИ ОПТИМИЗАЦИИ СИНТЕЗА РЕКОМБИНАНТНЫХ БЕЛКОВ В РАСТИТЕЛЬНЫХ КЛЕТКАХ: КЛАССИЧЕСКИЕ ПОДХОДЫ И НОВЫЕ НАПРАВЛЕНИЯ

© 2019 г. С. М. Розов^{а, *}, Е. В. Дейнеко^{а, б}

^аФедеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия

^бТомский государственный университет, Томск, 634050 Россия

*e-mail: rozov@bionet.nsc.ru

Поступила в редакцию 23.07.2018 г.

После доработки 25.09.2018 г.

Принята к публикации 26.09.2018 г.

В настоящее время фармакологически значимые белки синтезируют в различных системах экспрессии — от бактериальных до культур клеток млекопитающих и насекомых. Растительные системы экспрессии, особенно суспензионные культуры клеток, выгодно сочетают в себе простоту и низкую стоимость бактериальных систем со способностью к посттрансляционным модификациям белков по эукариотическому типу. Одним из недостатков растительных систем экспрессии остается низкий (по сравнению с бактериальными системами) выход целевого рекомбинантного белка. В обзоре рассмотрены разработанные за последние два десятилетия способы повышения уровня экспрессии рекомбинантных генов и способы предотвращения сайленсинга, вызванного случайным встраиванием целевого гена в гетерохроматиновые участки. Появление технологий CRISPR/Cas привело к созданию нового подхода к повышению уровня экспрессии генов — направленного встраивания генов “фармацевтических” белков в конкретные, заведомо транскрипционно активные участки генома. Такими участками могут быть локусы генов домашнего хозяйства растительной клетки, активно экспрессируемые в течение всей интерфазы. Подробно рассмотрена организация некоторых генов домашнего хозяйства, наиболее перспективных для переноса в их локусы генов рекомбинантных белков.

Ключевые слова: рекомбинантные гены, оптимизация экспрессии, фармацевтические белки, суспензионные культуры клеток растений, растительные системы экспрессии, экспрессионные кассеты, эффект положения, генный сайленсинг, CRISPR/Cas, направленный мутагенез, редактирование генома, гены домашнего хозяйства

DOI: 10.1134/S0026898419020149

ВВЕДЕНИЕ

В настоящее время большое число белков медицинский назначения получают не из природных источников, а синтезируют их рекомбинантные аналоги в различных системах экспрессии (*Escherichia coli*, дрожжи, клетки животных и др.). К преимуществам прокариот в качестве системы экспрессии относятся простота генно-инженерных манипуляций, быстрый рост, высокий уровень продукции рекомбинантных белков и возможность масштабного производства белка в биореакторах. Тем не менее, существует целый ряд посттрансляционных модификаций, которые

прокариотическая клетка не в состоянии осуществить. Некорректные модификации или их отсутствие могут значительно изменить свойства синтезируемого белка, в том числе его биологическую активность и фармакокинетику. Поэтому прокариотические системы экспрессии используют преимущественно для синтеза относительно простых терапевтических белков, таких как инсулин, интерферон или гормон роста человека. Более сложные белки чаще получают в системах, основанных на культурах клеток млекопитающих [1], но и эти системы имеют свои слабые места. Это, прежде всего, высокая стоимость культиви-

Сокращения: UTR (UnTranslated Regions) — нетранслируемые области генов; IGS (InterGenic Spacer) — межгенный спейсер; ETS (External Transcribed Spacer) — внешний транскрибируемый спейсер; ITS (Internal Transcribed Spacer) — внутренний транскрибируемый спейсер; CGI (CG Islands) — CpG-островки; IRES (Internal Ribosome Entry Site) — внутренний сайт посадки рибосомы; 3'-CITE (3'-Cap Independent Translation Enhancer) — 3'-кепнезависимые энхансеры трансляции; MAR (Matrix Attachment Regions) — матрикссвязывающие районы хроматина; IME (Intron Mediated Enhancement) — интрон-опосредованное усиление экспрессии; RMCE (Recombinase Mediated Cassete Exchange) — рекомбиназный обмен кассет.

рования, сложность масштабирования процесса и потенциальная зараженность вирусами. Суспензионные культуры клеток растений лишены этих недостатков, они считаются чрезвычайно перспективными для синтеза рекомбинантных белков, поскольку сочетают способность к пост-трансляционной модификации по эукариотическому типу с простотой и дешевой бактериальных систем экспрессии. Однако, несмотря на успехи в использовании суспензионных культур клеток растений для коммерческого получения терапевтических белков [2, 3], остается достаточно много нерешенных проблем, наиболее важная среди которых — недостаточно высокий уровень экспрессии рекомбинантного белка.

Чаще всего проблема повышения продукции рекомбинантного белка в растительных клетках решается за счет оптимизации генетических конструкций, включающих целевой ген, и условий культивирования растительных клеток, а также отбора наиболее “благоприятных” событий интеграции трансгена в растительный геном. Сюда относятся ставшие уже классическими методы оптимизации частот кодонов, подбор конститутивных промоторов, встраивание в вектор 5'-нетранслируемых областей (5'-UTR), интронзависимое усиление экспрессии целевого гена, транспорт рекомбинантного белка в органеллы, апопласт или культуральную среду, коэкспрессия с ингибиторами протеаз [4–7]. Кроме того, выход рекомбинантного белка может сильно зависеть от выбора продуцента и условий его культивирования [4, 5, 7].

При использовании классических методов интеграции рекомбинантной ДНК целевой ген встраивается в случайный участок генома. Разные области генома могут значительно отличаться по своей транскрипционной активности: от высокоактивных (эухроматин) до практически молчащих (гетерохроматин). Состояние хроматина определяется множественными эпигенетическими факторами: расположением нуклеосом, негистоновыми белками, вариантами и модификациями гистонов, метилированием ДНК [8]. Значительная вариабельность уровня и стабильности экспрессии целевого гена, наблюдаемая в трансгенных линиях [9], обусловлена тем, что встраивание целевого гена в гетерохроматиновый район резко уменьшает его экспрессию и может приводить к сайленсингу [10]. Поэтому практически всегда приходится проводить интенсивный скрининг трансгенных линий и отбирать наиболее стабильные и высокопродуктивные из них [11].

В последние годы появляются новые стратегии доставки целевых генов в геном продуцента, позволяющие с высокой вероятностью встраивать их в эухроматиновые регионы. С этой целью используют транспозоноподобные векторные конструкции, которые встраиваются преимущественно

но в эухроматин [12], или векторы, где целевой ген с обеих сторон фланкирован протяженными эухроматиновыми участками [13]. К сожалению, пока эти подходы реализованы только на клетках животных. Появление и быстрое развитие CRISPR/Cas-технологий [14] позволило относительно легко и без серьезных затрат адресно доставлять целевые гены почти в любые нужные конститутивно транскрибируемые участки генома. В дальнейшем можно будет отказаться от трудоемкого скрининга большого числа трансгенных линий и сразу получить стабильный и высокий выход целевого белка.

Для оптимизации экспрессии рекомбинантных генов с использованием CRISPR/Cas-технологий необходимо выявить круг генов, которые конститутивно и высокоэффективно транскрибируются в растительной клетке. Прежде всего, это гены домашнего хозяйства, представленные в геноме несколькими копиями или тысячами копий. Сюда относятся гены рРНК, гены гистонов и ряд других.

В нашем обзоре кратко рассмотрены классические стратегии оптимизации синтеза рекомбинантных белков и новые подходы, связанные с встраиванием целевых генов в эухроматиновые участки генома. Особое внимание уделено организации и структуре генов домашнего хозяйства растительной клетки, являющихся потенциальными сайтами интеграции целевых генов, экспрессию которых необходимо интенсифицировать.

ДИЗАЙН ЭКСПРЕССИОННЫХ КАССЕТ

Экспрессионные кассеты — важная часть рекомбинирующего вектора, содержащая не только целевой ген и промоторные элементы, но и целый ряд энхансерных и регуляторных последовательностей, включая UTR (5'- и 3'-нетранслируемые районы) и внутренние сайты посадки рибосом (IRES). Экспрессионные кассеты могут быть классифицированы по числу цистронов (рекомбинантных генов) в единой соответствующей им мРНК. Моноцистронные кассеты, используемые для синтеза отдельных белков, содержат регуляторные элементы генов этих белков. При необходимости одновременного синтеза двух или более белков в одной клетке их кассеты могут быть помещены в одну или несколько плазмид, но во втором случае резко уменьшается вероятность одновременной трансформации клетки всеми компонентами. С другой стороны, размещение нескольких моноцистронных кассет в одном векторе значительно увеличивает его размеры, что может уменьшить вероятность встраивания в растительный геном, привести к неполной интеграции и потере одного или нескольких трансгенов [15].

Бицистронные кассеты обычно содержат IRES вирусного происхождения, что позволяет избежать 5'-кепазависимого уменьшения скорости трансляции и дает возможность связывать факторы ее инициации. Часто в бицистронные кассеты помещают вирусную 2А-последовательность, которая позволяет инициировать трансляцию в клетках растений в 9 раз эффективнее, чем IRES. 2А-последовательность функционирует котрансляционно, терминируя трансляцию, и позволяет собирать в одной кассете до 9 генов [16, 17]. На основе полицистронных кассет можно сформировать различные опероноподобные структуры. Такие кассеты очень привлекательны для метаболической инженерии растений, например гликоинженерии, так как позволяют за 1 раз вносить в растительный геном целые гетерологичные метаболические пути [17]. Подробно строение экспрессионных кассет и *cis*-факторы, влияющие на их экспрессию, рассмотрены в обзорах [18, 19].

Промоторы

Для экспрессии рекомбинантных белков в растительных клетках используется широкий круг промоторов как растительного происхождения, так и экзогенных. Все эти промоторы можно разделить на конститутивные, тканеспецифичные и индуцибельные. Каждая из этих категорий также делится на промоторы, специфичные для двудольных и однодольных растений [7, 19]. Принято считать, что растительный промотор состоит из двух частей: коровой (кор-промотор), расположенной непосредственно перед геном, и дистальной, расположенной на расстоянии до тысяч пар нуклеотидов перед ним, и содержащей разнообразные энхансеры. Обе части промотора содержат *cis*-элементы, к которым присоединяются различные белки, известные как факторы инициации транскрипции [20]. Подробно строение растительных промоторов и их *cis*-регуляторных элементов рассмотрено в обзоре [19].

Для увеличения экспрессии рекомбинантных белков в растительных клетках очень важен подбор промотора, способного оптимально взаимодействовать с *транс*-факторами транскрипции [21]. База растительных промоторов (<http://ppdb.agr.gifu-u.ac.jp/ppdb/cgi-bin/index.cgi>) содержит большой набор нуклеотидных последовательностей кор-промоторов и групп регуляторных элементов *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa* и *Physcomitrella patens* [22]. Тем не менее, в векторах для трансформации растений очень часто используются экзогенные промоторы. Например, очень широко применяемый конститутивный промотор 35S CaMV вируса мозаики цветной капусты, хорошо совместим с транскрипционной машиной двудольных растений и используется обычно для дизайна компактных экспрессионных кассет,

хотя его активность может снижаться в некоторых тканях и клетках [19]. Активность полиубиквитиновых промоторов PvUbi1 и PvUbi2 *Panicum virgatum* (просо) в несколько раз выше активности 35S CaMV и актинового промотора риса OsAct1 [23]. Тканеспецифичные промоторы могут использоваться для накопления рекомбинантного белка в отдельных органах и тканях. Например, промоторы генов зеинов кукурузы [24], генов глутелина, глобулина и проламинов риса [25] специфичны для эндосперма и определяют синтез белка только в клетках эндосперма зерновок, имеющих достаточно низкую активность протеаз. В результате возрастает не только стабильность и количество рекомбинантного белка, но и все растение не испытывает возможных токсических эффектов от накопления чужеродного белка. Свое применение находят и индуцибельные промоторы. Недавно на основе индуцируемого β -эстрадиолом промотора создана коллекция TRANSPLANTA из 1636 линий *A. thaliana*, каждая из которых экспрессирует на высоком уровне один из 634 факторов транскрипции [26]. Сконструирован промотор, включаемый красным светом ($\lambda = 660$ нм) и выключаемый инфракрасным (740 нм) [27]. В последнее время активно разрабатываются искусственные промоторы — как конститутивные, тканеспецифичные и индуцибельные, так и подходящие для разных видов растений [28].

UTR и опосредованное интронами усиление экспрессии (IME)

Недавние исследования показали, что нуклеотидные последовательности 5'-UTR мРНК разных генов *A. thaliana* значительно варьируют, в результате чего эффективность их трансляции существенно изменяется (более чем в 200 раз). Наиболее благоприятным нуклеотидом в 5'-UTR в позициях $-1...-21$ от точки инициации трансляции оказался остаток А. В частности, для максимальной эффективности трансляции требуется А в позициях $-1...-5$. Напротив, остаток Т в этих позициях обеспечивает самый низкий уровень трансляции. Влияние участка 5'-UTR, расположенного между позициями -1 и -21 , на эффективность трансляции прослеживается и у других видов растений [29]. Механизм этого влияния не совсем понятен, но ранее было показано, что шпилечные структуры, образуемые мРНК в районе стартового кодона, могут сильно влиять на интенсивность трансляции [30]. Более того, этот консервативный район необходим для связывания факторов инициации трансляции и посадки субъединиц рибосомы [31].

РНК-содержащие вирусы растений, не имеющие 5'-кепа, приобрели кепнезависимый механизм, позволяющий им связывать факторы инициации трансляции и субъединицы рибосом. Их

3'-UTR содержат кепнезависимые трансляционные энхансеры (3'-SITE), которые образуют T-, Y-подобные или иные шпильчатые структуры. Эти шпильки пространственно взаимодействуют с 5'-концом вирусной РНК, имитируя 5'-кеп [31]. В системе транзientной экспрессии, основанной на вирусе мозаики вигны, присутствие Y-подобного 3'-SITE повышает уровень трансляции на 30% [32]. Усилить экспрессию можно также с помощью введения 5'-UTR, которые действуют как трансляционные энхансеры. Подробно влияние UTR на эффективность трансляции у растений и его механизмы рассмотрены в обзоре [33].

Еще один способ оптимизации синтеза целевых белков в растении — использование опосредованного интроном усиления экспрессии (IME) [6]. Добавление последовательностей интронов в 5'-UTR или другие участки целевых генов может значительно усиливать их экспрессию на транскрипционном, посттранскрипционном и трансляционном уровнях. Однако механизм IME изучен недостаточно. IME наблюдается не только у растений, но и у животных и даже у дрожжей. Общее у всех интронов, проявляющих эффект IME, — правильная ориентация, локализация в транскрибируемой последовательности и нахождение вблизи сайта инициации транскрипции [6]. Уровень транскрипции при IME повышается в 2–3 раза, но накопление мРНК и содержание белкового продукта может увеличиться в 20–30 раз [34]. Встраивание интрона гена *rubi3* риса в 5'-UTR привело к повышению экспрессии целевых генов в трансгенных растениях риса в 29 раз [35]. Интрон гена убиквитина (*Ubi1*) кукурузы в сочетании со своим промотором обеспечивает очень существенное увеличение экспрессии целевых генов как в трансгенных растениях кукурузы [36], так и в растениях риса [37].

Частым осложнением биотехнологических манипуляций является то, что чужеродный рекомбинантный ген, управляемый сильным конститутивным промотором или встроенный в геном в большом числе копий, имеет тенденцию “замолкать” по механизму косупрессионного сайленсинга [38]. Обойти проблему сайленсинга можно с помощью более слабых или тканеспецифичных промоторов. Так как IME влияет на все уровни экспрессии, включая стабильность мРНК, интроны могут быть отличным инструментом усиления экспрессии генов в сочетании со слабыми промоторами. Более того, эффект повышения экспрессии интронами даже возрастает при использовании слабых промоторов. Поскольку IME каждого интрона может преимущественно влиять на разные фазы экспрессии, интрон с низким уровнем усиления транскрипции может стабилизировать мРНК или усиливать трансляцию, или вызывать и то, и другое [6].

ОПТИМИЗАЦИЯ ИСПОЛЬЗОВАНИЯ КОДОНОВ

В связи с вырожденностью генетического кода одна и та же аминокислота может кодироваться несколькими кодонами (от двух до шести). При этом два первых нуклеотида кодона крайне консервативны, а основная вариабельность приходится на нуклеотид в третьей позиции. Частоты встречаемости синонимических кодонов различаются у разных организмов. Они декодируются рибосомами с разной скоростью, так как соответствующие им тРНК также присутствуют у разных организмов в разных количествах [39]. Поэтому уровень трансляции рекомбинантного белка можно значительно повысить, заменив редкие кодоны на предпочитаемые данным продуцентом. Уровень экспрессии гена можно предсказать исходя из частоты использования кодонов с помощью различных методик и компьютерных программ [59]. База Codon usage database (<http://www.kazusa.or.jp/codon/>) предоставляет возможность узнать частоты использования синонимических кодонов и процент встречаемости GC в генетическом коде более чем у 35000 организмов.

Подходы, включающие замену редких кодонов на частые, в сочетании с уменьшением общего количества GC, приводящего к формированию стабильной вторичной структуры мРНК, часто используются биотехнологическими компаниями и исследовательскими группами для оптимизации экспрессии гетерологичных белков, но, к сожалению, с переменным успехом [41, 42]. В некоторых случаях оптимизация использования кодонов способна существенно увеличить выход белка (до 1000 раз) [41]. Действительно, в хлоропластах *Clamydomonas* и табака экспрессия целевых генов возросла в 75 [43] и 80 [44] раз соответственно. В опытах с хлоропластами табака и латука эффект составил 5–7 и 22–28 раз [45]. К сожалению, оптимизация использования кодонов может не только повышать, но и снижать экспрессию генов гетерологичных белков. Например, после оптимизации кодонов уровень экспрессии антигена SAG178–322 *Toxoplasma gondii* в листьях табака не изменился, тогда как экспрессия антигена SAG1 снизилась в 5–10 раз [46]. Однако чаще оптимизация использования кодонов приводит к усилению экспрессии рекомбинантных белков, но в ряде случаев образуются биологически неактивные нерастворимые агрегаты, которые приходится подвергать рефолдингу для восстановления их вторичной и третичной структуры и биологической активности [41, 47]. Остается неясным, каким образом синонимичные замены кодонов влияют на фолдинг белков, но показано, что не только множественные, но даже и единичные замены могут влиять на активность белков, взаимодействие с лекарствами и ингибиторами, профиль фосфорилиро-

вания, чувствительность к протеолизу [48, 49] и агрегационную способность [50]. В отдельных случаях синонимичные замены могут привести к образованию совершенно иной третичной структуры белка [46, 51]. Важно заметить, что подобные эффекты характерны не только для бактериальных, но и для эукариотических систем экспрессии. Эти структурные и функциональные нарушения могут быть связаны с тем, что фолдинг полипептидной цепи происходит одновременно с трансляцией. Оптимизация кодонов приводит к тому, что участки, ранее содержавшие редкие кодоны и синтезировавшиеся значительно медленнее, начинают быстро покидать рибосому. В результате ключевые районы полипептидной цепи, отвечающие за правильную третичную структуру, не успевают пространственно сблизиться для взаимодействия друг с другом, что приводит к неправильному фолдингу. Таким образом, синонимические замены являются как бы вторичным кодом, определяющим фолдинг белка [46].

ПРЕДОТВРАЩЕНИЕ ВНУТРИКЛЕТОЧНОЙ ДЕГРАДАЦИИ РЕКОМБИНАНТНЫХ БЕЛКОВ

Цитозоль — не лучшее место накопления рекомбинантных белков, прежде всего из-за большого числа протеаз. В эндоплазматическом ретикулуме протеаз меньше, но при накоплении большого количества чужеродного белка в нем включается защитный механизм ответа клетки на белки с неправильной конформацией (Unfolded Protein Response — UPS), сопряженный с убиквитинзависимой протеасомной деградацией [52]. В ряде случаев удается увеличить время жизни белка, сливая его с другими стабильными белками или их доменами, например с убиквитинсвязывающим доменом UBA2 рецептора убиквитина RAD23a [53]. Поскольку создать растение, лишенное протеаз, вряд ли удастся, то для сохранения уже синтезированного белка обычно используют два независимых подхода: коэкспрессию с генами ингибиторов протеаз и транспорт рекомбинантного белка в органеллы и другие внутриклеточные структуры. Кроме того, предпринимаются попытки модифицировать сам механизм UPS у растений, очень консервативный и достаточно хорошо изученный у эукариот [54]. Первый подход, более подробно рассмотренный в обзоре [7], был успешно применен на картофеле с использованием ингибитора протеаз катепсина D томатов [55]. В случае второго подхода к белкам добавляют небольшие сигнальные последовательности, задерживающие их в эндоплазматическом ретикулуме или обеспечивающие их транспорт во внутриклеточные структуры, содержащие относительно мало протеаз: вакуоли, апопласт или пластиды. При использовании ткане-

специфичных промоторов рекомбинантные белки могут синтезироваться также в семядолях (двудольные) или в эндосперме (однодольные), где могут поступать в масляные или белковые запасные тельца, производные эндоплазматического ретикулума. Подробнее это направление рассмотрено в обзорах [5, 7].

Основные достоинства и недостатки описанных методов оптимизации экспрессии рекомбинантных генов, ставших уже классическими, кратко суммированы в сводной табл. 1.

ИСКЛЮЧЕНИЕ НЕГАТИВНОГО ВЛИЯНИЯ ГЕТЕРОХРОМАТИНА НА ЭКСПРЕССИЮ ТРАНСГЕНОВ

Разные районы генома могут значительно различаться по своей транскрипционной активности: от очень активных (эухроматин) до практически молчащих (гетерохроматин). Классическая агробактериальная и биобаллистическая трансформация приводят к случайному встраиванию рекомбинантного гена в геном, что влечет за собой высокую вариабельность его экспрессии в полученных трансгенных линиях [9], нестабильность уровня экспрессии и даже сайленсинг [10]. Здесь проявляется давно известный генетикам эффект положения: активность гена в значительной степени определяется его хромосомным окружением. Состояние гетерохроматина зависит от посттрансляционной модификации гистонов, метилирования ДНК и постоянно поддерживается большим набором факторов, присутствующих в нем, поэтому эффект снижения транскрипционной активности имеет тенденцию распространяться далеко за пределы самого гетерохроматина [56]. Поэтому трансгены, встроившиеся вблизи гетерохроматиновых регионов, склонны к сайленсингу. В последние два десятилетия проводится активный поиск способов избежать негативного влияния гетерохроматина на экспрессию трансгенов.

Барьерные инсуляторы

Один из возможных способов избежать распространения влияния гетерохроматина на близлежащие участки генома заключается во фланкировании трансгена генетическими элементами, блокирующими влияние гетерохроматина, что позволяет трансгену экспрессироваться с высокой эффективностью. Примером таких элементов могут быть барьерные инсуляторы, которые, как считается, играют важную роль в организации хроматина, разделяя его на функциональные домены [57]. Барьерные инсуляторы способны локально влиять на ацетилирование и метилирование гистонов, что, в конечном итоге, может активировать хроматин [58]. В число барьерных инсуляторов входят матрикссвязывающие районы

Таблица 1. Классические методы оптимизации уровня экспрессии целевых белков

Метод	Достоинства метода	Недостатки метода
• Подбор промоторов	<ul style="list-style-type: none"> • Сильные промоторы повышают экспрессию в несколько раз • Тканеспецифичные промоторы могут обеспечить синтез белка в органах с низкой протеазной активностью • Возможность конструировать промоторы с различными <i>цис</i>-энхансерами 	<ul style="list-style-type: none"> • Сильные промоторы могут приводить к быстрому сайленсингу гена • Не оказывает влияния на скорость трансляции
• Введение в ген нетранслируемых районов (UTR)	<ul style="list-style-type: none"> • 5'-UTR могут изменить скорость трансляции в 200 раз • 3'-UTR (3'-CITE – кепнезависимые трансляционные энхансеры) ускоряют трансляцию на треть 	<ul style="list-style-type: none"> • Влияет на экспрессию гена только на уровне трансляции
• Опосредованное интроном усиление экспрессии (IME)	<ul style="list-style-type: none"> • Могут усиливать экспрессию на транскрипционном, посттранскрипционном и трансляционном уровнях • Уровень транскрипции может повышаться в 2–3 раза, накопление мРНК и белка до 30 раз • Позволяет использовать слабые промоторы, значительно усиливая уровень экспрессии 	<ul style="list-style-type: none"> • Малая предсказуемость метода • Неясность механизма усиления экспрессии
• Оптимизация использования кодонов	<ul style="list-style-type: none"> • Способен увеличить выход белка в десятки раз, в некоторых случаях до 1000 	<ul style="list-style-type: none"> • Не всегда приводит к положительным результатам • Может нарушить фолдинг и привести к изменению биологических свойств белка
• Предотвращение деградации белка	<ul style="list-style-type: none"> • Коэкспрессия с ингибиторами протеаз увеличивает выход белка почти в 2 раза • Транспорт белков во внутриклеточные структуры может увеличить его концентрацию до 10 раз 	<ul style="list-style-type: none"> • Не связан с уровнями транскрипции и трансляции

хроматина (Matrix attachment region – MAR) – области ДНК, прикрепляющиеся к ядерному матриксу, создающие хроматиновые петли и образующие границы между разными доменами хроматина [59]. MAR экранируют рекомбинантные гены от негативного влияния соседнего гетерохроматина, усиливают их экспрессию [60] и предотвращают сайленсинг [61]. MAR выделены из геномов большого числа эукариот, в том числе растений [62], и часто используются для фланкирования экспрессионных кассет с целью уменьшения вариабельности экспрессии трансгена [58]. Использование MAR-элемента из гена лизоцима А курицы для фланкирования трансгенов, экспрессирующихся в клетках табака и риса, существенно уменьшало вариабельность экспрессии в трансгенных линиях путем подавления сайленсинга [63]. Но этот же MAR-элемент не влиял на сайленсинг у трансгенного арабидопсиса [64]. На неэффективность и даже на негативный эффект

некоторых MAR на экспрессию трансгена указывали и другие данные [65, 66]. Возможно, MAR это гетерогенная группа элементов, которые объединяет только способность связываться с ядерным матриксом. Кроме того, MAR могут противостоят только *цис*-сайленсингу (распространение эффекта гетерохроматина), но не могут предотвратить *транс*-сайленсинг (например посттранскрипционный) [67]. MAR представляют собой довольно протяженные участки ДНК, которые могут содержать не только сайты связывания с матриксом, но и другие регуляторные элементы.

Кроме MAR, некоторые инсуляторы также способны защищать трансгены от эффекта положения в геноме растений. Например, 5'-инсулятор гена арилсульфатазы морского ежа усиливает экспрессию трансгенов и уменьшает вариабельность их экспрессии в клетках табака [68]. Инсулятор *gypsy Drosophila melanogaster* также по-

вышает уровень экспрессии трансгенов и снижает ее вариабельность в клетках *A. thaliana*, но только при коэкспрессии белка Su(Hw) (супрессор hairy-wings) *D. melanogaster*, взаимодействующего с *gypsy* [69]. Известные нуклеотидные последовательности барьерных инсуляторов мало сходны между собой, однако появляется все больше данных, что они привлекают к себе гистон-ацетилазную и гистон-метилазную активности. Например, инсулятор 5'-HS4 гена β -глобина курицы взаимодействует с белками, осуществляющими опосредованное HS4 метилирование и ацетилирование гистонов H3 и H4, свойственное эухроматину [70] и препятствует метилированию ДНК [71]. Предполагается, что опосредованные HS4 модификации гистонов предотвращают метилирование H3K9 и присоединение HP1, что останавливает распространение гетерохроматина и локально поддерживает хроматин в активном состоянии [70].

Другие факторы, активирующие хроматин

Первый фактор, обладающий способностью активировать хроматин на любом участке генома, UCOE (Ubiquitous Chromatin-Opening Element), был выделен при изучении района двух генов домашнего хозяйства человека: гена TATA-связывающего белка (*TBP*) и расположенного поблизости гена *PSMB1*, кодирующего субъединицу протеасомы [72]. *TBP* и *PSMB1* транскрибируются в разных направлениях с помощью двойного двунаправленного промотора, расположенного между ними и окруженного с обеих сторон неметилированными CpG-островками (CGI). Экспрессия фрагмента *TBP-PSMB1* в фибробластах мыши оказалась стабильной в течение как минимум 60 дней, т.е. окружающий промотор CGI защищает его от эпигенетического сайленсинга и обеспечивает продолжительную и стабильную экспрессию [72]. Еще один фрагмент генома с такой же архитектурой — локус *HNRPA2B1-CBX3*, также содержит гены двух белков домашнего хозяйства — РНП A2/B1 (*HNRPA2B1*) и белка гетерохроматина 1Hs- γ (*CBX3*). CGI (2.6 т.п.н.), содержащий двунаправленный промотор, который частично перекрывает первые альтернативные экзоны обоих генов. Этот элемент был назван A2UCOE. Локусы *TBP-PSMB1* и *HNRPA2B1-CBX3*, включая их двунаправленные промоторы, окруженные неметилированными CGI, способны активировать хроматин и обеспечивать стабильную экспрессию даже при интеграции в участки центрального конститутивного гетерохроматина [73]. Исследование экспрессии A2UCOE в культуре мононуклеарных клеток крови показало, что в районе сайтов инициации транскрипции генов *HNRPA2B1* и *CBX3* гистоны H3 и H4 метилированы и ацетилированы по эухроматическому типу [74]. Кассеты, содержащие A2UCOE и eGFP с

промотором hCMV, обеспечивали не только 20-кратное увеличение экспрессии eGFP, но также предотвращали сайленсинг и поддерживали стабильность экспрессии в клетках СНО яичников китайского хомячка в течение 100 поколений [73]. Еще один UCOE обнаружен в гене рибосомного белка S3 мыши (*Rps3*). В отличие от предыдущих UCOE, *Rps3* содержит один ген домашнего хозяйства под управлением однонаправленного промотора, окруженного CGI из 358 п.н. Элемент *Rps3* размером 3 т.п.н. в сочетании с промотором hCMV обеспечивает высокую и стабильную экспрессию eGFP в клетках СНО [75]. К сожалению, попытки обнаружения и использования UCOE в растительных системах экспрессии пока не принимались. Подробно использование UCOE в биотехнологии и генной терапии рассмотрено в обзорах [76, 77].

Стабилизирующие антирепрессорные элементы (Stabilizing Anti Repressor Elements — STAR) были обнаружены в результате скрининга библиотеки геномных фрагментов ДНК человека по их способности блокировать ассоциированные с хроматином репрессоры транскрипции, что позволяло трансгенным клеткам выживать на селективной среде в результате подавления вызываемого этими репрессорами сайленсинга генов антибиотикоустойчивости. Выделено 10 элементов с различной антирепрессорной активностью. Элементы STAR7 и STAR40 показали одинаково высокую активность и были активны как в клетках человека, так и в клетках СНО [78]. STAR40, расположенный на хромосоме 22 недалеко от гена *IL17R*, состоит из уникальной некодирующей последовательности и не имеет CpG-островков. Векторы, содержащие целевые гены, управляемые разными промоторами и фланкированные STAR40, позволяют получить большее число стабильно трансформированных клеточных линий и повышают уровень экспрессии пропорционально числу встроенных копий, т.е. уровень экспрессии не зависит от места встраивания [78]. Однако несмотря на многообещающие возможности, это направление не получило дальнейшего развития.

Эухроматин-содержащие векторы экспрессии

Совершенно другой подход к решению проблемы негативного влияния гетерохроматина на экспрессию трансгенов заключается во фланкировании экспрессионной кассеты протяженными заведомо эухроматическими участками. С этой целью в качестве векторов для трансформации клеток СНО использовали искусственные бактериальные хромосомы (BAC) [79], которые представляют собой большие векторы с протяженными участками активного хроматина, способными вместить до 200–300 т.п.н. трансгенного генетического материала. Поэтому в них можно

встраивать большие локусы со всеми их управляющими элементами. ВАС широко используются при получении трансгенных мышей, так как транскрипционная активность встроенных в ВАС генов не зависит от эффекта положения и пропорциональна числу встроенных копий [80]. На основе ВАС получена серия векторов с различными комбинациями локусов активного хроматина, промоторов и других регуляторных элементов, экспрессирующих фрагмент IgG1 человека (IgG-Fc), CN54gp140, сильно гликозилованный полипептид HIV-1 и PG9, антитела против HIV-1. Не все комбинации ВАС, промоторов и локусов активного хроматина привели к значительному увеличению выхода белка при стабильной экспрессии в клетках СНО, но уровень экспрессии был пропорционален числу встроенных копий, что говорит о нечувствительности системы к эффекту положения. Наилучшие результаты — 10-кратное повышение экспрессии — получены с использованием вектора Caggs:VAC^{Rosa26} (ВАС с геном *Rosa26* мыши и промотором Caggs) [79]. Понятно, что использовать ВАС для оптимизации экспрессии в растительных клетках невозможно, но сам принцип окружения экспрессионной кассеты протяженными эухроматическими районами можно с успехом применять при конструировании векторов для растительных систем экспрессии.

Транспозоны как векторы, селективно нацеленные на эухроматин

Другой подход к решению проблем, связанных с гетерохроматином, основан на повышении вероятности встраивания целевого трансгена в эухроматические районы генома. ДНК-транспозоны, перемещающиеся по геному по принципу “разрезал и вклеил”, имеют тенденцию с большей долей вероятности встраиваться в районы регуляторных областей активных генов, где нарушена строгая нуклеосомная структура, и где ДНК более доступна для их специфических транспозаз, и в транскрипционно активные районы, богатые CpG-островками [81]. Векторы, сконструированные на основе таких транспозонов, после первичного встраивания способны перемещаться в районы эухроматина, что повышает и стабилизирует экспрессию трансгена. Такие транспозоны уже давно используются в биоинженерии клеток мыши, человека и насекомых. Полусинтетический транспозон *Sleeping Beauty* некоторое время использовали для инсерционного мутагенеза у мышей [82], однако его применение ограничивала концентрация перемещений недалеко от сайта первичного встраивания и малая генная емкость в качестве вектора для переноса трансгенов [83]. Этих недостатков лишен транспозон *PiggyBac* (PB), первоначально выделенный из генома капустной совки *Trichoplusia ni* [84]. Инсерция PB в

геном зависит от PB-транспозазы, которая доставляется в геном либо самим транспозоном, либо отдельным вектором-помощником, что дает возможность управлять перемещениями транспозона. Сайты инсерции PB имеют высокую тенденцию располагаться в транскрипционно активных участках генома, богатых CGI. PB встраивается по тетра-нуклеотидной последовательности TТАА, а при эксцизии не оставляет за собой следов, восстанавливая исходную последовательность. Кроме того, генный “багаж”, переносимый PB, может превышать 200 т.п.н., что делает его эффективным в случае большого набора генов [81, 85]. PB успешно используется для высокоэффективного трансгенеза клеток человека, мыши и СНО. Подробно его применение описано в обзоре [85]. Весьма перспективной для оптимизации экспрессии трансгенов в растительных клетках представляется разработка систем биоинженерных манипуляций на основе PB-подобных транспозонов растений.

Рекомбиназный кассетный обмен (RMCE)

Опосредованный рекомбиназами кассетный обмен (RMCE) основан на использовании рекомбиназ, широко распространенных у бактерий и дрожжей. С этой целью используют два типа рекомбиназ, тирозиновые и сериновые, которые различаются активным аминокислотным остатком в каталитическом центре. Сайт-специфическая рекомбинация происходит по сайтам узнавания, специфичным для каждой рекомбиназы, и может привести к инсерции, делеции или инверсии фрагмента ДНК, что определяется расположением и ориентацией сайтов узнавания в геномной ДНК и в RMCE-кассете [86]. Из тирозиновых рекомбиназ чаще всего используют две системы: Cre/loxP бактериофага P2 и Flp/FRT дрожжей. Cre и Flp являются рекомбиназами, а loxP и FRT их специфическими сайтами узнавания [87, 88]. Так как рекомбиназы осуществляют обмен только по своим специфическим сайтам узнавания, то сначала в случайные области генома хозяина неспецифически интегрируют участки ДНК, фланкированные вариантами loxP или FRT, в зависимости от используемой системы. Донорную кассету, содержащую целевые и селективные гены, фланкированные теми же вариантами сайтов узнавания тирозиновой рекомбиназы, доставляют в клетки хозяина в составе плазмиды, несущей ген самой рекомбиназы. Возможен вариант с использованием вспомогательной плазмиды с геном рекомбиназы, что позволяет контролировать кассетный обмен. После замещения геномных кассет донорными в результате двойного кроссинговера проводится отбор клеточных линий по селективным маркерам и по стабильному и высокому уровню экспрессии целевых генов

Таблица 2. Методы оптимизации уровня экспрессии целевых белков, связанные с исключением негативного влияния гетерохроматина

Метод	Достоинства метода	Недостатки метода
<ul style="list-style-type: none"> Барьерные инсуляторы (MAR и др.) 	<ul style="list-style-type: none"> Способность усиливать экспрессию и уменьшать ее вариабельность, подавлять генный сайленсинг 	<ul style="list-style-type: none"> Не всегда приводит к положительному эффекту Значительные размеры Подавление только <i>цис</i>-сайленсинга
<ul style="list-style-type: none"> Элементы, активирующие хроматин (UCOE) 	<ul style="list-style-type: none"> Возможность усиления экспрессии гена до 20 раз Способность обеспечивать высокий уровень экспрессии гена и предотвращать генный сайленсинг, даже если ген интегрирован в прицентромерный гетерохроматин 	<ul style="list-style-type: none"> Не известны UCOE для растительных систем экспрессии
<ul style="list-style-type: none"> Стабилизирующие антирепрессорные элементы (STAR) 	<ul style="list-style-type: none"> Обеспечивают высокий уровень экспрессии независимо от места интеграции гена 	<ul style="list-style-type: none"> Не известны для растительных систем экспрессии
<ul style="list-style-type: none"> Содержащие эухроматин экспрессионные векторы 	<ul style="list-style-type: none"> Увеличение уровня экспрессии до 10 раз Уровень экспрессии не зависит от эффекта положения 	<ul style="list-style-type: none"> Значительное утяжеление встраиваемой конструкции Не использовались в растительных системах экспрессии
<ul style="list-style-type: none"> Использование транспозонов как векторов, селективно нацеленных на эухроматин 	<ul style="list-style-type: none"> Возможность контролировать перемещение транспозона с помощью вектора-помощника, несущего транспозазу Возможность переноса значительных объемов целевых конструкций 	<ul style="list-style-type: none"> Транспозонные системы для растений не разработаны
<ul style="list-style-type: none"> Рекомбиназный кассетный обмен (RMCE) 	<ul style="list-style-type: none"> Возможность повторного использования линии с высокой экспрессией трансгена для внесения новых целевых генов Возможность создавать батареи стабильно экспрессирующихся тесно сцепленных генов 	<ul style="list-style-type: none"> Необходимость предварительного отбора линий с сайтами рекомбинации, встроенными в активные участки хроматина

[86, 89]. Примечательно, что отобрав однажды клеточную линию с высокой и стабильной экспрессией трансгена, можно в дальнейшем использовать ее для замещения кассетами с другими и/или дополнительными трансгенами, создавая батареи стабильно экспрессирующихся тесно сцепленных генов [90]. Сериновые рекомбиназы PhiC31 и Vxb1 в отличие от тирозиновых имеют разные сайты узнавания в геномных (*attP*) и донорных (*attB*) кассетах, которые изменяются после рекомбинации, что, с одной стороны, не позволяет использовать их повторно, а с другой, делает обмен необратимым [86]. Показано, что в трансгенных линиях риса, табака и других растений, полученных методом RMCE, наблюдается более стабильная экспрессия трансгенов, сайленсинг происходит реже, чем при случайной интеграции, и устраняются негативные последствия эффекта положения. Подробно механизмы RMCE и его

использование для редактирования геномов растений рассмотрены в обзорах [86, 88, 89].

Достоинства и недостатки методов оптимизации экспрессии рекомбинантных генов, основанных на исключении негативного влияния гетерохроматина, кратко суммированы в сводной табл. 2.

Несколько особняком стоит метод транзientной экспрессии рекомбинантных генов, при котором ген не встраивается в геном, а экспрессируется какое-то время в составе свободной плазмиды. Интенсивность транзientной экспрессии не подвержена эффекту положения и может быть на 1–2 порядка выше, чем при перманентном встраивании. Но транзientная экспрессия относительно непродолжительна (до нескольких дней) и фактически неприменима в суспензионных клеточных культурах. Этот метод чаще всего используется, когда требуется быстро получить небольшое количество белка [91, 92].

ВСТРАИВАНИЕ РЕКОМБИНАНТНЫХ ГЕНОВ В ЗАДАННЫЕ РАЙОНЫ ГЕНОМА

Проблема малой эффективности гомологичной рекомбинации у эукариот

Если у прокариот и дрожжей встраивание ДНК легко происходит по механизму гомологичной рекомбинации (HDR), т.е. обмена гомологичными участками между геномной ДНК и кассетой-донором, то у большинства многоклеточных организмов, в том числе и у растений, вероятность такого события крайне низка. По разным оценкам и у разных объектов она составляет от 10^{-3} до 10^{-6} на каждую трансформацию [93]. Поэтому долгое время в клетках эукариот не удавалось осуществить направленный мутагенез, т.е. прямую модификацию определенных генов. Впервые направленный мутагенез был осуществлен в 1987 г. на стволовых клетках эмбриона мыши [94, 95], а годом позже на клетках табака. В протопласты табака, содержащие дефектный ген устойчивости к канамицину, перенесли плазмиду, несущую фрагмент гена устойчивости к канамицину, что привело к восстановлению устойчивости к канамицину в результате HDR [96]. Но частота гомологичных обменов была крайне низкой. Показано, что путем HDR в геном чаще всего встраивается только один конец вектора для направленного мутагенеза, тогда как другой — в результате негомологичного соединения концов (NHEJ). Возможен и альтернативный вариант, когда направленный мутагенез не происходит, а вектор встраивается в случайный сайт генома [97]. Число обнаруживаемых событий гомологичного обмена пытались повысить с помощью особых форм отбора клеток — генспецифической селекции, позитивно-негативной селекции, а также путем изменения экспрессии генов хозяина, ответственных за HDR и внедрения в растительный геном генов, обеспечивающих гомологичные обмены у прокариот. Но только сверхэкспрессия гена *RAD54* дрожжей — члена семейства генов *SWI2/SNF2*, вовлеченного в процессы ремоделирования хроматина, привела к повышению частоты направленного мутагенеза на 1–2 порядка [98]. Подробно эти вопросы рассмотрены в обзорах [91, 93, 97, 99].

Системы редактирования геномов

Решение проблемы низкой частоты гомологичных обменов у эукариот было подсказано самой природой. В мейозе специфические клеточные нуклеазы вызывают двухцепочечные разрывы ДНК, приводящие к гомологичным обменам между хромосомами — кроссинговеру. Увеличение с помощью двухцепочечных разрывов частоты гомологичных обменов на два порядка показано в клетках дрожжей и растений [100]. Это позволило разработать большое число методов направленной

модификации генома растений, прежде всего сайт-специфического мутагенеза и сайт-специфического встраивания трансгенов. Осталось создать нуклеазы, которые можно “настраивать” для высокоспецифичного связывания с определенными, заранее заданными нуклеотидными последовательностями.

В 2000 г. созданы первые химерные нуклеазы, Zinc-finger нуклеазы (ZFN), которые представляли собой гибридный белок, содержащий нуклеазный домен рестриктазы FokI и три–четыре домена так называемых цинковых пальцев, найденных в факторах транскрипции. Каждый домен цинковых пальцев распознает последовательность из трех нуклеотидов, и массив из трех–четырех таких доменов в пересчете на димерную молекулу может распознавать и связываться с последовательностью из 18–24 н. На клетках табака, кукурузы и арабидопсиса показана эффективность использования ZFN для направленного мутагенеза у растений, причем почти во всех экспериментах частота направленного мутагенеза превышала 10%. Однако у метода “цинковых пальцев” обнаружены и серьезные недостатки, такие как не вполне строгое распознавание тринуклеотидов, приводящее к заметному числу расщеплений ДНК в нецелевых участках. Поскольку каждый домен распознает три нуклеотида, не ко всем последовательностям ДНК можно подобрать цинковые пальцы. Кроме того, метод оказался весьма трудоемким и дорогим, так как для каждой определенной нуклеотидной последовательности необходимо создавать свою оптимальную структуру ZFN [101]. Подробно цинковые пальцы и их использование для геномного редактирования рассмотрены в обзорах [102, 103].

Более эффективным средством адресного разрезания ДНК оказались конструкции на основе химерных нуклеаз, известных как TALEN (Transcription Activator-like Effector Nucleases). В этих нуклеазах роль структур, распознающих нуклеотидную последовательность, играют TAL-домены, соединенные в гибридный белок с доменом нуклеазы FokI, каждый из которых опознает один нуклеотид. Эти TAL-домены нашли в 2009 г. у бактерии *Xanthomonas*, паразитирующей в клетках растений, а уже в 2010 г. с использованием TALEN получили первые результаты [104]. TAL-домен распознает только один нуклеотид, поэтому можно синтезировать нуклеазу, вносящую двухцепочечные разрывы в любую заданную последовательность генома [105]. За довольно короткое время с помощью инструментария TALEN адресный мутагенез был осуществлен на большом числе видов растений, и его эффективность доходила до 90%. Более подробно TALEN и их использование для адресной модификации генов растений рассмотрено в обзорах [105–107].

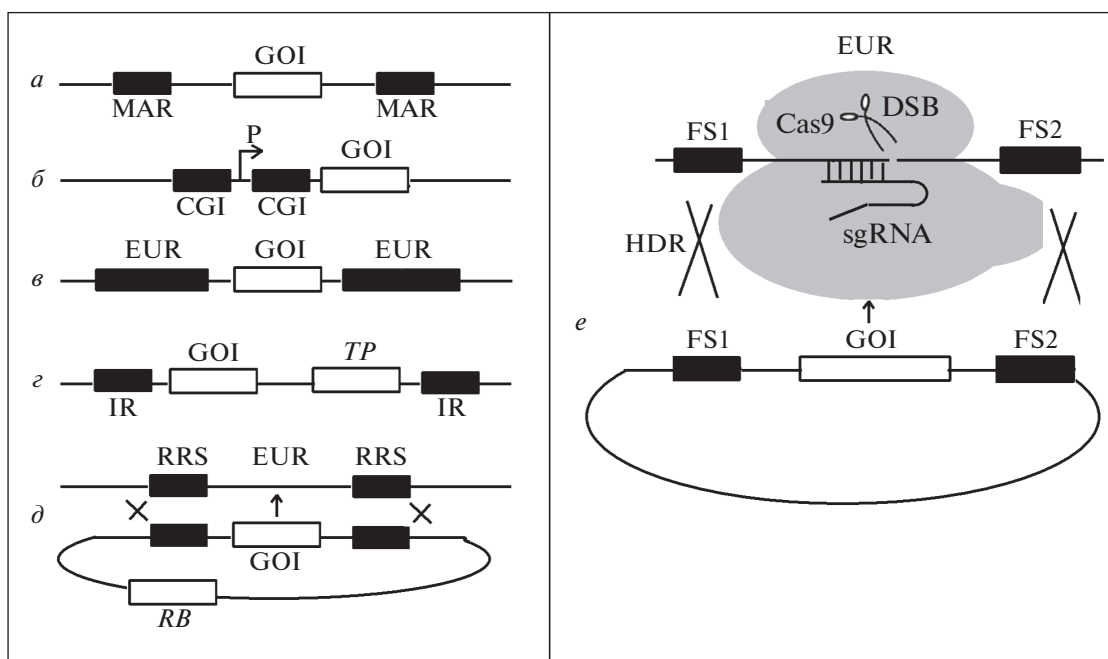


Рис. 1. Классические методы устранения негативного влияния гетерохроматина на экспрессию трансгенов (*a–d*), и схема направленного встраивания рекомбинантных генов в заданный эухроматиновый участок генома с использованием технологии CRISPR/Cas9 (*e*). *a* – Барьерные инсуляторы; *b* – универсальные активаторы хроматина (UCOE); *c* – фланкирование целевого гена протяженными участками эухроматина; *d* – использование транспозоноподобных векторов; *e* – рекомбинантный касетный обмен (RMCE). Обозначения: GOI (Gene Of Interest) – целевой ген; P – промотор; CGI – неметилированные CpG-богатые островки; EUR – эухроматиновые участки; IR (Inverted Repeats) – инвертированные повторы; TP – ген транспозазы; RRS (Recombinase Recognition Sites) – сайты узнавания рекомбиназы; RB – ген рекомбиназы; FS1, FS2 – фланкирующие последовательности; DSB (Double Stranded Break) – двухцепочечный разрыв; HDR (Homology Directed Repair) – гомологичная репарация, приводящая к гомологичному обмену.

В 2012–2013 гг. в области геномного редактирования произошел поистине революционный прорыв: был разработан новый метод CRISPR/Cas адресной модификации генов, открывший принципиально новые возможности для генетических манипуляций на уровне геномов высших организмов [108]. Этот метод чрезвычайно прост, обеспечивает практически точное воздействие на заданные участки ДНК и может использоваться в любой современной молекулярно-биологической лаборатории. Как это часто бывает, ничего особенно придумать не пришлось: природа все сделала сама. Систему CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), ассоциированную с нуклеазой Cas, многие бактерии и археи используют для защиты от заражения фагами и чужеродными плазмидами. К настоящему времени детально описано несколько типов систем CRISPR, функционирующих у различных бактерий [109]. Наиболее удобной для биотехнологического применения оказалась система CRISPR/Cas9 типа II-A, обнаруженная у *Streptococcus pyogenes*. В состав этой системы входит рибонуклеопротеидный комплекс tracrRNA–crRNA–Cas, в котором специфические РНК узнают определенные последовательности ДНК, а нуклеаза

Cas9 разрезает их [110, 111]. В 2012 г. была показана возможность объединения функций tracrRNA и crRNA в единой молекуле РНК, получившей название sgRNA (single guide), что сразу же позволило использовать эту систему для нужд биотехнологии [112].

На основе системы CRISPR/Cas типа II-A созданы конструкции, кодирующие элементы универсального геномного редактора CRISPR/Cas9. sgRNA в этой системе опознает участок ДНК длиной 20 п.н., а нуклеаза Cas9 производит в нем двухцепочечный разрыв. С помощью системы CRISPR/Cas9 можно осуществлять все виды модификаций генома: создавать точечные мутации, встраивать в определенные места новые гены либо, наоборот, удалять крупные участки ДНК, исправлять или заменять отдельные генетические элементы и фрагменты генов [110, 111].

Система CRISPR/Cas9 может эффективно использоваться для адресного встраивания целевых генов в заданные участки активного хроматина. На рис. 1 приведены все ставшие уже классическими методы преодоления негативных эффектов гетерохроматина на экспрессию трансгена (рис. 1*a–d*), а также схема направленного переноса

са гена в заданный эухроматический район генома с использованием эндонуклеазы Cas9 (рис. 1e).

Разработано множество онлайн-инструментов дизайна sgRNA и подбора рестриктаз для встраивания целевого гена практически в любую последовательность генома растений, содержащую короткий PAM (Protospacer Adjacent Motif). Например, CRISPR-P (<http://crispr.hzau.edu.cn/CRISPR2/>) и CRISPR-PLANT (<https://www.genome.arizona.edu/crispr/>) Сравнительный анализ достоинств и недостатков большей части этого инструментария опубликован в 2015 г. в специальном исследовании [113]. Число работ, посвященных редактированию геномов, в том числе растительных, с помощью системы CRISPR/Cas9 растет в геометрической прогрессии. Если в 2013 г. (по данным Академии Google) таких работ было 130, то сейчас их общее число превышает 7500, и только за первые 4 мес. 2018 г. появилось около 1000 новых, в которых упоминается эта система. Более подробно система CRISPR/Cas9 и ее использование для редактирования геномов растений рассмотрены в многочисленных обзорах [14, 110, 111, 114–117].

В последнее время активно разрабатывается еще одна система генного редактирования на основе CRISPR из *Francisella novicida* с эндонуклеазой Cas12a (Cpf1). Эта нуклеаза использует T-богатый PAM, в отличие от G-богатого у Cas9, более короткую (~43 н.) направляющую crRNA (~100 н. в sgRNA у Cas9), имеет один, а не два каталитических центра для создания двухцепочечных разрывов в ДНК и образует липкие концы, что облегчает дальнейший гомологичный обмен. Эта система более точно опознает целевую последовательность в геноме, тогда как CRISPR/Cas9 допускает несовпадение нескольких нуклеотидов и может разрезать схожие нецелевые участки [118, 119]. Но сравнительный анализ частоты адресного мутагенеза у кукурузы, проведенного с использованием этих двух эндонуклеаз, показал большую эффективность Cas9: 90–100% и только 0–60% у Cas12a, причем не выявлено ни одной нецелевой мутации в случае Cas9 [120]. Аналогичные исследования, проведенные на растениях риса с использованием Cas12a, также показали, что частота направленных мутаций составляет от 16 до 30% при отсутствии неадресных модификаций [121]. Таким образом, система CRISPR/Cas9 пока остается предпочтительной для растительных объектов.

ТРАНСКРИПЦИОННО АКТИВНЫЕ РАЙОНЫ ГЕНОМА, НАИБОЛЕЕ ПЕРСПЕКТИВНЫЕ ДЛЯ ПЕРЕНОСА В НИХ ЦЕЛЕВЫХ РЕКОМБИНАНТНЫХ ГЕНОВ

Как уже не раз отмечалось, для высокой и стабильной экспрессии в геноме растения целевой трансген должен локализоваться в районе с высокой транскрипционной активностью. Для этого

наиболее подходят участки, содержащие так называемые гены домашнего хозяйства, конститутивно и активно транскрибируемые на протяжении всей интерфазы клеточного цикла. Кроме того, желательно, чтобы такой ген был представлен в геноме не одной копией. Это, во-первых, позволит встроить несколько копий трансгена, а, во-вторых, сохранит функциональность самого гена-мишени за счет его копий, оставшихся интактными. Набор таких генов достаточно велик, и мы рассмотрим только наиболее перспективные из них. Это, прежде всего, гены различных РНК, участвующих в синтезе белков, гены гистонов, актина, тубулина и убиквитина. В табл. 3 представлены данные о числе копий этих генов в геноме *A. thaliana*, их организации, хромосомной локализации и фазе активности в клеточном цикле.

Гены 45S (35S) рРНК

В ядерном геноме высших растений гены, кодирующие 18S, 5.8S и 25S рРНК организованы в единую транскрипционную единицу длиной ~5 т.п.н. и разделены между собой внутренними транскрибируемыми спейсерами (ITS) [122]. По аналогии с генами рРНК животных эту транскрипционную единицу часто называют геном 45S рРНК. Однако у растений этот единый транскрипт короче за счет меньшей длины спейсеров и имеет коэффициент седиментации 35S [123]. Эта транскрипционная единица повторена в геноме растений от 1000 до десятков тысяч раз, формируя протяженные тандемные повторы, расположенные в одном или в нескольких локусах. Транскрипционные единицы в таких повторах разделены более вариabельными нетранскрибируемыми межгенными спейсерами участками (IGS) длиной 5–15 т.п.н. [124].

Гены 35S рРНК *A. thaliana* образуют кластеры в двух локусах в околотеломерных районах хромосом 2 и 4. Каждый из этих локусов содержит ~375 генов, что в сумме составляет приблизительно 4 млн.п.н. [125], т.е. всего в диплоидном геноме *A. thaliana* около 1500 генов 35S рРНК. Каждый генный повтор содержит IGS и транскрипционную единицу 35S, включающую 5'- и 3'-внешние транскрибируемые спейсеры (5'- и 3'-ETS), а также ITS1 и ITS2, удаляемые в ходе процессинга. Все 1500 генов имеют практически идентичные последовательности, кодирующие рРНК, однако однонуклеотидный полиморфизм и небольшие делеции/инсерции в спейсерах приводят к тому, что геном *A. thaliana* содержит более 10 подтипов генов 35S рРНК [126]. Приблизительно половина этих подтипов на определенных этапах развития подвергается сайленсингу, вызванному состоянием хроматина, а другая половина конститутивно работает в течение всего онтогенеза [127]. Активно транскрибируемые подтипы принадлежат локусу хромосомы 4, а молчащие — локусу хромосомы 2

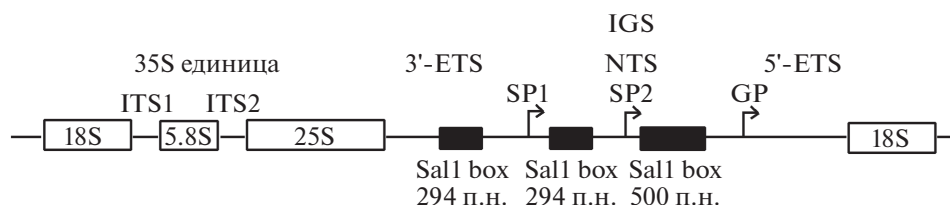
Таблица 3. Гены домашнего хозяйства *Arabidopsis thaliana*, наиболее перспективные для встраивания в них генов целевых рекомбинантных белков

Гены домашнего хозяйства	Число копий в гаплоидном геноме	Организация в геноме	Распределение по хромосомам (локусам)	Фаза активности
Гены 35S рРНК	750	Тандемные кластеры 18S, 5.8S, 25S рРНК	Два локуса в двух хромосомах	Конститутивно интерфаза
Гены 5S рРНК	1000	Тандемные повторы	Четыре локуса в трех хромосомах	Конститутивно интерфаза
Гены тРНК	10–20 копий каждой тРНК	Частично образуют тандемные повторы, гены разных тРНК диспергированы по геному	Во всех хромосомах	Конститутивно интерфаза
Гены гистонов	15	В основном не кластеризованы и не образуют тандемных повторов	В четырех хромосомах (гистон H3)	S-фаза, часть копий конститутивно интерфаза
Гены актина	8–10	Диспергированы по геному	В четырех хромосомах	Конститутивно интерфаза
Гены тубулина	12	Диспергированы по геному	В четырех хромосомах	Конститутивно интерфаза
Гены убиквитина	12, из них пять полиубиквитиновые (три–шесть повторов)	Диспергированы по геному		Конститутивно интерфаза

[126]. Перенос молчащих подтипов с хромосомы 2 на хромосому 4 приводит к их активации, указывая на то, что сайленсинг определяется состоянием прилежащего к ним хроматина [128]. IGS разделяет две соседние транскрипционные единицы и содержит множественные регуляторные элементы [129]. У *A. thaliana* IGS несет два спейсерных промотора, генный промотор и три группы так называемых SalI-повторов, расположенных в тех же позициях, что и энхансеры в IGS *Xenopus* и мыши. SalI-повторы, по-видимому, также являются энхансерами, поскольку они способны функционально заменять энхансеры в ооцитах *Xenopus* [130] (рис. 2).

Представляется очевидным, что ядрышко, содержащее гены 35S рРНК, является наиболее активной частью генома. Поэтому адресное встраивание целевых рекомбинантных генов в области

активного ядрышкового организатора может привести к их стабильному и высокому уровню экспрессии. Наиболее перспективными в этом смысле представляются районы IGS и 5'-ETS. Несмотря на то, что в норме 35S рРНК транскрибируется РНК-полимеразой I, которая не кепирует свои транскрипты, рекомбинантные гены с собственными сильными промоторами для РНК-полимеразы II должны привлекать ее в эти области генома. Кроме того, при расположении целевых генов в 5'-ETS даже транскрипты РНК-полимеразы I могут в дальнейшем транслироваться, если экспрессионная кассета содержит упоминавшиеся ранее энхансерные элементы, 5'-IRES [16] или 3'-CITES [31], которые обеспечивают кепнезависимое привлечение факторов трансляции и связывание с рибосомой.

**Рис. 2.** Строение кластера рибосомных генов *A. thaliana*. NTS (NonTranscribed Spacer) – нетранскрибируемый спейсер; SP1, SP2 – спейсерные промоторы; GP – генный промотор; SalI box – энхансерные повторы.

Гены 5S рРНК и тРНК

У большинства эукариот гены 5S рРНК расположены отдельно от кластера генов 35S рРНК. Тем не менее, у дрожжей, некоторых нематод и членистоногих, а также у некоторых растений (часть семейства Asteraceae, некоторые голосеменные) гены 5S рРНК входят в состав tandemных повторов 35S рРНК внутри IGS и транскрибируются РНК-полимеразой III либо в прямом, либо в обратном направлении [124, 131]. Диплоидный геном *A. thaliana* содержит ~2000 копий генов 5S рРНК [132], образующих длинные tandemные массивы в перичентромерных районах хромосом 3, 4 и 5. Хромосомы 3 и 4 несут по одному локусу генов 5S рРНК, тогда как хромосома 5 содержит на одном плече большой локус генов 5S рРНК и незначительный в другом. Небольшой локус генов 5S рРНК обнаружен на хромосоме 3 не у всех экотипов, а его хромосомная локализация отличается вариабельностью [133]. Повторяющаяся единица генов 5S рРНК содержит транскрибируемый участок ~120 п.н. и IGS длиной ~480 п.н. Экспрессируются только гены хромосомы 4 и большой локус на хромосоме 5, содержащие в сумме ~300 генов (гаплоидный геном) и кодирующие два класса 5S рРНК: основной (82%) и минорный (18%), отличающиеся одной—двумя нуклеотидными заменами в транскрибируемом районе. На начальных стадиях развития проростков доля минорного класса может достигать 25% [134, 135].

Ядерные гены тРНК растений, как и у других эукариот, представлены мультигенными семействами, в которых отдельные копии либо беспорядочно разбросаны по геному, либо образуют кластеры в одном или нескольких сайтах. Первый вариант более характерен для генов тРНК растений. Большинство генов тРНК активно и конститутивно транскрибируются на протяжении всей интерфазы. Обычно геном растений содержит 10–20 копий гена каждой тРНК. Например, ген тРНК^{Tyr} повторен в геноме *A. thaliana* минимум 20 раз, а в геноме *Nicotiana rustica* — по крайней мере 14 раз. Все обнаруженные у *A. thaliana* копии гена тРНК^{Tyr} организованы в tandemные повторы с повторяющейся единицей длиной 1.5 т.п.н., причем эта единица содержит также ген тРНК^{Ser} [136, 137]. Гены тРНК, как и гены 5S рРНК, транскрибируются РНК-полимеразой III [138].

Сайтами, перспективными для встраивания целевых рекомбинантных генов в районы расположения генов 5S рРНК и тРНК, могут быть межгенные спейсеры (в случае tandemно повторенных генов) или непосредственно прилегающие к гену заведомо эухроматиновые области. Так как транскрипты РНК-полимеразы III не подвергаются кепированию, имеет смысл использовать кассеты экспрессии, содержащие 5'-IRES или 3'-CITES.

Гены гистонов

У разных групп живых организмов организация генов гистонов имеет значительные различия. У многих животных (почти все беспозвоночные, рыбы, амфибии) гены гистонов собраны в кластеры, содержащие гены всех пяти типов гистонов (H3, H4, H2a, H2b, H1), а сами кластеры образуют tandemные повторы. У нематод и птиц tandemные повторы кластеров не образуют, а диспергированы в геноме. Геном млекопитающих (мышь, человек) содержит в основном неполные кластеры из генов гистонов двух—трех типов, не образующие tandemных повторов и разбросанные по геному [139]. Об организации генов гистонов растений известно значительно меньше. Гены гистонов кукурузы, пшеницы, сои, ячменя и арабидопсиса обычно не собраны в кластеры, не образуют tandemных повторов и диспергированы по геному [140–142]. У растений наиболее хорошо изучены гены гистона H3. Число копий генов гистона H3 у эволюционно удаленных видов варьирует от 13 до 16 [143]. Геном *A. thaliana* содержит 15 копий гена гистона H3, расположенных на четырех хромосомах [142]. Следует ожидать, что число копий генов гистонов других типов должно быть таким же, так как для формирования нуклеосомной структуры необходимы их эквивалентные количества, однако гистона H1 требуется вдвое меньше. Действительно, в геноме гороха выявлены семь копий гена гистона H1, расположенных в трех локусах на двух хромосомах [144]. Экспрессия большинства генов гистонов зависит от репликации, т.е. происходит в S-фазе клеточного цикла, но малая их часть конститутивно транскрибируется в ходе всей интерфазы [145].

Гены гистонов несомненно являются важнейшими генами домашнего хозяйства, так как обеспечивают необходимую для всех эукариот компактизацию хроматина, принимают участие в эпигенетической регуляции экспрессии генов и располагаются в областях активно транскрибируемого хроматина. Поэтому с целью повышения уровня экспрессии трансгенов весьма перспективным представляется расположение целевых рекомбинантных генов в участках генома, фланкирующих гены гистонов. Особенно эффективным это должно быть в случае суспензионных культур растительных клеток, представленных постоянно делящимися клетками, у которых гены гистонов транскрибируются практически постоянно и активно.

Гены актина и тубулина

Актин и тубулин — основные компоненты цитоскелета клетки, формирующие микрофиламенты и микротрубочки. *A. thaliana*, содержит 8–10 значительно дивергировавших копий генов ак-

тина, разбросанных по четырем хромосомам, как у риса и сорго. У *Medicago truncatula* найдены только четыре гена актина, тогда как у кукурузы их 21 [146, 147]. Аналогичная картина наблюдается и в случае генов тубулина. У *A. thaliana* таких генов 12, из них четыре кодируют α -тубулин, а восемь — β -тубулин. Эти гены располагаются на четырех хромосомах, и только два гена β -тубулина тесно сцеплены между собой [148]. Не все варианты генов актина и тубулина экспрессируются с одинаковой эффективностью в различных органах и тканях: одни более активны в соматических органах, другие в генеративных [147, 149]. Подробно строение и регуляция генов актинов и тубулинов растений рассмотрены в обзорах [147, 150]. Перспективными для встраивания целевых рекомбинантных генов могут оказаться 5'-UTR и, возможно, интроны этих генов.

Гены убиквитина

Убиквитин — короткий белок (76 аминокислотных остатков), который может ковалентно присоединяться своим С-концевым остатком Gly к боковой ε -аминогруппе остатков Lys любого белка, формируя Y-образную структуру. По своей консервативности убиквитин сопоставим с гистоном H4: убиквитинов грибов, животных и растений различаются двумя—тремя аминокислотными остатками [151]. Эффект убиквитинирования не сводится только к хорошо известной протеасомной деградации белков. Присоединяясь к гистонам и факторам транскрипции, убиквитин принимает участие в регуляции разнообразных процессов в клетке, включая транскрипцию [152, 153].

Гены убиквитина уникальны по своей организации: это всегда гибридные гены, где часть, кодирующая убиквитин, соединена голова-к-хвосту (без межгенного спейсера) с частью, кодирующей либо другой белок, либо еще одну копию убиквитина (полиубиквитиновые гены). Всего в геноме *A. thaliana* насчитывается 12 генов убиквитина, пять из которых полиубиквитиновые (содержат три—шесть повторов, кодирующих элементарный убиквитин), пять “слиты” с участками, кодирующими три разных белка (L40, S27a3, S27a1) малой и большой субъединиц рибосомы, и два гена, содержащие кодирующую последовательность убиквитин-подобного белка RUB [151, 152]. Гибридный транскрипт этих генов транслируется как единое целое и только посттрансляционно процессируется — разрезается убиквитин-специфичными протеазами (деубиквитиназами), опознающими Gly75—Gly76 в молекуле убиквитина, причем далее могут находиться любые аминокислотные остатки, кроме Pro. Это свойство деубиквитиназ использовано для конструирования белков с заданными N-концевыми структурами [151,

154, 155], а также для успешного повышения уровня экспрессии целевых генов, слитых с 3'-частью гена убиквитина [156]. Подробно строение генов убиквитина и роль убиквитина в регуляции клеточных процессов рассмотрена в обзорах [151—153]. С учетом уникальных особенностей строения генов убиквитина и возможностей деубиквитиназ рассматривается возможность переноса целевых рекомбинантных генов не только в области, фланкирующие гены убиквитина, но и в дистальные 5'-районы гена белка, слитого с убиквитином, причем в этом случае целевому гену не понадобится собственный промотор, так как транскрипция начнется с убиквитинового промотора, сопоставимого по эффективности с промотором 35S CaMV.

ЗАКЛЮЧЕНИЕ

В последние годы становится очевидным, что растительные системы экспрессии, особенно суспензионные клеточные культуры, сочетающие в себе простоту и низкую стоимость бактериальных систем со способностью ко многим посттрансляционным модификациям эукариотического типа, наиболее перспективны для производства “фармакологических” белков в промышленных масштабах. Но одной из основных проблем, связанных с растительными системами, остается низкий (по сравнению с бактериальными системами) уровень экспрессии целевого рекомбинантного белка. Предложено множество стратегий, направленных на решение этой проблемы, прежде всего на оптимизацию самой экспрессионной кассеты для увеличения уровня транскрипции рекомбинантного гена. Это правильный подбор промоторов, включение в кассету различных регуляторных элементов, 5'-UTR, интронов и других энхансерных последовательностей, а также оптимизация кодонов самого целевого гена. Другое направление связано с сохранением наработанного целевого белка — коэкспрессии с ингибиторами протеаз и активного транспорта белка в протеолитически наименее активные клеточные компартменты. Эти методы оптимизации экспрессии рекомбинантных генов уже долгое время успешно используются и стали классическими. Но, к сожалению, классические методы не всегда и не на всех объектах приводят к ожидаемому результату. Даже такая казалась бы очевидная и понятная методика, как оптимизация кодонов в целевом гене, может повлечь за собой совершенно неожиданные последствия — нарушение фолдинга целевого белка и потерю его функциональности.

Другое очень важное направление оптимизации экспрессии рекомбинантных генов — преодоление негативных эффектов гетерохроматина и предотвращение сайленсинга генов. С этой целью успешно применяют барьерные инсуляторы,

факторы, активирующие хроматин, рекомбиназные кассеты и транспозоноподобные векторы. С появлением простой и удобной системы направленного геномного редактирования CRISPR/Cas появилась возможность развития принципиально нового подхода к проблеме оптимизации экспрессии целевого рекомбинантного белка — направленного встраивания целевого гена в заранее выбранную область активно транскрибируемого хроматина. Такими областями являются прежде всего конститутивно транскрибируемые гены домашнего хозяйства, представленные в геноме не одной копией, что дает возможность встраивания в прилежащие к ним области нескольких копий трансгена, не нарушая при этом работу самого гена домашнего хозяйства. В предложенном обзоре мы попытались кратко охарактеризовать гены домашнего хозяйства растительной клетки, наиболее привлекательные для оптимизации экспрессии трансгенов с использованием этого нового подхода. Один из рассматриваемых подходов, а именно, интеграцию целевого гена в район ядрышкового организатора и район одного из генов гистона H3 с целью повышения выхода рекомбинантного белка в культуре клеток растений, мы изучаем в настоящее время на модельной клеточной линии *A. thaliana*.

Работа выполнена при финансовой поддержке Российского научного фонда (№ 17-14-01099).

СПИСОК ЛИТЕРАТУРЫ

- Zhu J. (2012) Mammalian cell protein expression for biopharmaceutical production, *Biotechnol. Adv.* **30**, 1158–1170.
- Teoah Y., Shulman A., Kizhner T., Ruderfer I., Fux L., Nataf Y., Bartfeld D., Ariel T., Gingis-Velitski S., Hanania U., Shaaltiel Y. (2015) Large-scale production of pharmaceutical proteins in plant cell culture — the protalix experience. *Plant Biotechnol. J.* **13**, 1199–1208.
- Naji-Talakar S. (2017) Plant-derived biopharmaceuticals: overview and success of agroinfiltration. *Trends Capstone*. **2**, 1–12.
- Ullrich K.K., Hiss M., Rensing S.A. (2015) Means to optimize protein expression in transgenic plants. *Curr. Opin. Biotechnol.* **32**, 61–67.
- Habibi P., Prado G.S., Pelegrini P.B., Hefferon K.L., Soccol C.R., Grossi-de-Sa M.F. (2017) Optimization of inside and outside factors to improve recombinant protein yield in plant. *Plant Cell Tissue Organ Culture (PCTOC)*. **130**, 449–467.
- Laxa M. (2017) Intron-mediated enhancement: A tool for heterologous gene expression in plants? *Front. Plant Sci.* **7**, 1977.
- Saberianfar R., Menassa R. (2018) Strategies to increase expression and accumulation of recombinant proteins. In: *Molecular Pharming: Applications, Challenges, and Emerging Areas*. Ed. Kermod A.R. New York: John Wiley & Sons Inc., 119–135.
- Nandi S., Khush G.S. (2015) Strategies to increase heterologous protein expression in rice grains. In: *Recent Advancements in Gene Expression and Enabling Technologies in Crop Plants*. New York: Springer, 241–262.
- Nocarova E., Fischer L. (2009) Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression. *BMC Plant Biol.* **9**, 44.
- Huang N., Yang D. (2005) ExpressTec: high level expression of biopharmaceuticals in cereal grains. In: *Modern Biopharmaceuticals: Design, Development and Optimization*. Ed. Knablen J. New York: Wiley VCH, **3**, 931–947.
- Nandi S., Yalda D., Lu S., Nikolov Z., Misaki R., Fujiyama K., Huang N. (2005) Process development and economic evaluation of recombinant human lactoferrin expressed in rice grain. *Transgenic Res.* **14**, 237–249.
- Stolt-Bergner P., Benda C., Bergbrede T., Besir H., Celie P.H., Chang C., Drechsel D., Fischer A., Geerlof A., Giabbai B., Heuvel van den J., Huber G., Knecht W., Lehner A., Lemaitre R., Norden K., Pardsee G., Racke I., Remands K., Sander A., Scholz J., Stadnik M., Storici P., Weinbruch D., Zaror I., Lua L.H.L., Suppmann S. (2018) Baculovirus-driven protein expression in insect cells: A benchmarking study. *J. Struct. Biol.* **203**(2), 71–80. doi.org/10.1016/j.jsb.2018.03.004
- Mader A., Prewein B., Zboray K., Casanova E., Kurnert R. (2013) Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. *Appl. Microbiol. Biotechnol.* **97**, 4049–4054.
- Luo M., Gilbert B., Ayliffe M. (2016) Applications of CRISPR/Cas9 technology for targeted mutagenesis, gene replacement and stacking of genes in higher plants. *Plant Cell Repts.* **35**, 1439–1450.
- Naqvi S., Farre G., Sanahuja G., Capell T., Zhu C., Christou P. (2010) When more is better: multigene engineering in plants. *Trends Plant Sci.* **15**, 48–56.
- Geier M., Fauland P., Vogl T., Glieder A. (2015) Compact multi-enzyme pathways in *P. pastoris*. *Chem. Commun.* **51**, 1643–1646.
- Ha S.H., Liang Y.S., Jung H., Ahn M.-J., Suh S.-C., Kweon S.-J., Kim D.-H., Kim Y.-M., Kim J.-K. (2010) Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. *Plant Biotechnol. J.* **8**, 928–938.
- Makhzoum A., Benyammi R., Moustafa K., Trémouillaux-Guiller J. (2014) Recent advances on host plants and expression cassettes' structure and function in plant molecular pharming. *BioDrugs*. **28**, 145–159.
- Biłtas R., Szafran K., Hnatuszko-Konka K., Kononowicz A.K. (2016) Cis-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue Organ Culture (PCTOC)*. **127**, 269–287.
- Porto M.S., Pinheiro M.P.N., Batista V.G.L., Cavalcanti dos Santos R., de Albuquerque Melo Filho P., de Lima L.M. (2014) Plant promoters: an approach of structure and function. *Mol. Biotechnol.* **56**, 38–49.

21. Moustafa K., Makhzoum A., Trémouillaux-Guiller J. (2016) Molecular farming on rescue of pharma industry for next generations. *Crit. Rev. Biotechnol.* **36**, 840–850.
22. Hieno A., Naznin H.A., Hyakumachi M., Sakurai T., Tokizawa M., Koyama H., Sato N., Nishiyama T., Hasebe M., Zimmer A.D., Lang D., Reski R., Rensing S.A., Obokata J., Yamamoto Y.Y. (2014) ppdb: plant promoter database version 3.0. *Nucl. Acids Res.* **42**(database issue), D1188–D1192.
23. Mann D.G.J., King Z.R., Liu W., Joyce B.L., Percifield R.J., Hawkins J.S., LaFayette P.R., Artelt B.J., Burrell J.N., Mazarei M., Bennetzen J.L., Parrott W.A., Stewart C.N. (2011) Switchgrass (*Panicum virgatum* L.) polyubiquitin gene (PvUbi1 and PvUbi2) promoters for use in plant transformation. *BMC Biotechnol.* **11**, 74.
24. Zhang J., Martin J.M., Beecher B., Morris C.F., Hannah L.C., Giroux M.J. (2009) Seed-specific expression of the wheat puroindoline genes improves maize wet milling yields. *Plant Biotechnol. J.* **7** (8), 733–743.
25. Qu L.Q., Xing Y.P., Liu W.X., Xu X.P., Song Y.R. (2008) Expression pattern and activity of six glutelin gene promoters in transgenic rice. *J. Exp. Bot.* **59** (9), 2417–2424.
26. Coego A., Brizuela E., Castillejo P., Ruíz S., Koncz C., Del Pozo J. C., Pineiro M., Jarrillo J.A., Paz-Ares J., León J. (2014) The TRANSPLANTA collection of *Arabidopsis* lines: a resource for functional analysis of transcription factors based on their conditional overexpression. *Plant J.* **77**, 944–953.
27. Muller K., Siegel D., Jahnke F.R., Gerrer K., Wend S., Decker E.L., Reski R., Weber W., Zurbriggen M.D. (2014) A red light-controlled synthetic gene expression switch for plant systems. *Mol. Biosyst.* **10**, 1679–1688.
28. Liu W., Stewart C.N. (2016) Plant synthetic promoters and transcription factors. *Curr. Opin. Biotechnol.* **37**, 36–44.
29. Kim Y., Lee G., Jeon E., Sohn E.J., Lee Y., Kang H., Lee D.W., Kim D.H., Hwang I. (2014) The immediate upstream region of the 5'-UTR from the AUG start codon has a pronounced effect on the translational efficiency in *Arabidopsis thaliana*. *Nucl. Acids Res.* **42**, 485–498.
30. Plotkin J.B., Kudla G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42.
31. Simon A.E., Miller W.A. (2013) 3' cap-independent translation enhancers of plant viruses. *Annu. Rev. Microbiol.* **67**, 21–42.
32. Meshcheriakova Y.A., Saxena P., Lomonosoff G.P. (2014) Fine-tuning levels of heterologous gene expression in plants by orthogonal variation of the untranslated regions of a nonreplicating transient expression system. *Plant Biotechnol. J.* **12**, 718–727.
33. Srivastava A.K., Lu Y., Zinta G., Lang Z., Zhu J.K. (2018) UTR-dependent control of gene expression in plants. *Trends Plant Sci.* **23**, 248–259.
34. Laxa M., Müller K., Lange N., Doering L., Pruscha J.T., Peterhänsel C. (2016) The 5' UTR intron of the *Arabidopsis* GGT1 aminotransferase enhances promoter activity by recruiting RNA polymerase II. *Plant Physiol.* **172**, 313–327.
35. Lu J., Sivamani E., Azhakanandam K., Samadder P., Li X., Qu R. (2008) Gene expression enhancement mediated by the 5' UTR intron of the rice rubi3 gene varied remarkably among tissues in transgenic rice plants. *Mol. Genet. Genom.* **279**, 563–572.
36. Scott M.P. (2009) Transgenic maize. *Methods Mol. Biol.* **526**, 6–7.
37. Christensen A.H., Quail P.H. (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res.* **5**, 213–218.
38. Depicker A., Van Montagu M. (1997) Post-transcriptional gene silencing in plants. *Curr. Opin. Cell Biol.* **9**, 373–382.
39. Dana A., Tuller T. (2014) The effect of tRNA levels on decoding times of mRNA codons. *Nucl. Acids Res.* **42**, 9171–9181.
40. Gould N., Hendy O., Papamichail D. (2014) Computational tools and algorithms for designing customized synthetic genes. *Front. Bioeng. Biotechnol.* **2**, 41.
41. Gustafsson C., Govindarajan S., Minshull J. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.* **22**, 346–353.
42. Wu G., Zheng Y., Qureshi I., Zin H.T., Beck T., Bulka B., Freeland S.J. (2007) SGDB: a database of synthetic genes re-designed for optimizing protein overexpression. *Nucl. Acids Res.* **35**, D76–D79.
43. Franklin S., Ngo B., Efuet E., Mayfield S.P. (2002) Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. *Plant J.* **30**, 733–744.
44. Gisby M.F., Mellors P., Madesis P., Ellin M., Laverty H., O'Kane S., Ferguson M.W., Day A. (2011) A synthetic gene increases TGFbeta3 accumulation by 75-fold in tobacco chloroplasts enabling rapid purification and folding into a biologically active molecule. *Plant Biotechnol. J.* **9**, 618–628.
45. Kwon K.-C., Chan H.-T., León I.R., Williams-Carrier R., Barkan A., Daniell H. (2016) Codon optimization to enhance expression yields insights into chloroplast translation. *Plant Physiol.* **172**, 62–77.
46. Komar A.A. (2016) The art of gene redesign and recombinant protein production: approaches and perspectives. In: *Protein Therapeutics*. Eds Sauna Z.E., Kimchi-Sarfati C. Cham, Switzerland: Springer, 161–177.
47. Komar A.A. (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* **34**, 16–24.
48. Kimchi-Sarfati C., Oh J.M., Kim I.W., Sauna Z.E., Calcagno A.M., Ambudkar S.V., Gottesman M.M. (2007) A “silent” polymorphism in the *MDR1* gene changes substrate specificity. *Science*. **315**, 525–528.
49. Yu C.H., Dang Y., Zhou Z., Wu C., Zhao F., Sachs M.S., Liu Y. (2015) Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell.* **59**, 744–754.
50. Kim S.J., Yoon J.S., Shishido H., Yang Z., Rooney L.A., Barral J.M., Skach W.R. (2015) Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science*. **348**, 444–448.

51. Buhr F., Jha S., Thommen M., Mittelstaet J., Kutz F., Schwalbe H., Rodnina M.V., Komar A.A. (2016) Synonymous codons direct cotranslational folding toward different protein conformations. *Mol. Cell.* **61**, 341–351.
52. Liu J.-X., Howell S.H. (2010) Endoplasmic reticulum protein quality control and its relationship to environmental stress responses in plants. *Plant Cell.* **22**, 2930–2942.
53. Jang I.C., Niu Q.W., Deng S., Zhao P., Chua N.H. (2012) Enhancing protein stability with retained biological function in transgenic plants. *Plant J.* **72**, 345–354.
54. Thomas D.R., Walmsley A.M. (2015) The effect of the unfolded protein response on the production of recombinant proteins in plants. *Plant Cell Rep.* **34**, 179–187.
55. Goulet C., Khalf M., Sainsbury F., D'Aous M.-A., Michaud D. (2012) A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast. *Plant Biotechnol. J.* **10**, 83–94.
56. Allshire R.C., Madhani H.D. (2017) Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244.
57. Emery D.W. (2011) The use of chromatin insulators to improve the expression and safety of integrating gene transfer vectors. *Hum. Gene Ther.* **22**, 761–774.
58. Singer S.D., Liu Z., Cox K.D. (2012) Minimizing the unpredictability of transgene expression in plants: the role of genetic insulators. *Plant Cell Rep.* **31**, 13–25.
59. Harraghy N., Gaussin A., Mermoud N. (2008) Sustained transgene expression using MAR elements. *Curr. Gene Ther.* **8**, 353–366.
60. Wang F., Wang T.Y., Tang Y.Y., Zhang J.H., Yang X.J. (2012) Different matrix attachment regions flanking a transgene effectively enhance gene expression in stably transfected Chinese hamster ovary cells. *Gene.* **500**, 59–62.
61. Verma D., Verma M., Dey M., Jain R.K., Wu R. (2005) Molecular dissection of the tobacco Rb7 matrix attachment region (MAR): effect of 50 half on gene expression in rice. *Plant Sci.* **169**, 704–711.
62. Chinn A.M., Comai L. (1996) The heat shock cognate 80 gene of tomato is flanked by matrix attachment regions. *Plant Mol. Biol.* **32**, 959–968.
63. Oh S.-J., Jeong J.S., Kim E.H., Yi N.R., Yi S.-I., Jang I.-C., Kim Y.S., Suh S.-C., Nahm B.H., Kim J.-K. (2005) Matrix attachment region from the chicken lysozyme locus reduces variability in transgene expression and confers copy number-dependence in transgenic rice plants. *Plant Cell Rep.* **24**, 145–154.
64. Butaye K.M.J., Goderis I.J.W.M., Wouters P.F.J., Pues J.M.-T.G., Delaure S.L., Broekaert W.F., Depicker A., Cammue B.P.A., De Bolle M.F.C. (2004) Stable high-level transgene expression in *Arabidopsis thaliana* using gene silencing mutants and matrix attachment regions. *Plant J.* **39**, 440–449.
65. de Bolle M.F.C., Butaye K.M.J., Coucke W.J.W., Goderis I.J.W.M., Wouters P.F.J., van Boxel N., Broekaert W.F., Cammue B.P.A. (2003) Analysis of the influence of promoter elements and a matrix attachment region on the inter-individual variation of transgene expression in populations of *Arabidopsis thaliana*. *Plant Sci.* **165**, 169–179.
66. Torney F., Partier A., Says-Lesage V., Nadaud I., Barret P., Beckert M. (2004) Heritable transgene expression pattern imposed onto maize ubiquitin promoter by maize adh-1 matrix attachment regions: tissue and developmental specificity in maize transgenic plants. *Plant Cell Rep.* **22**, 931–938.
67. Allen G.C. (2009) The role of nuclear matrix attachment regions in plants. *Plant Cell Monogr.* **14**, 101–129.
68. Nagaya S., Yoshida K., Kato K., Akasaka K., Shimmyo A. (2001) An insulator element from the sea urchin *Hemicentrotus pulcherrimus* suppresses variation in transgene expression in cultured tobacco cells. *Mol. Genet. Genomics.* **265**, 405–413.
69. She W., Lin W., Zhu Y., Chen Y., Jin W., Yang Y., Han N., Bian H., Zhu M., Wang J. (2010) The gypsy insulator of *Drosophila melanogaster*, together with its binding protein suppressor of hairywing, facilitate high and precise expression of transgenes in *Arabidopsis thaliana*. *Genetics.* **185**, 1141–1150.
70. Huang S., Li X., Yusufzai T.M., Qiu Y., Felsenfeld G. (2007) USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol. Cell. Biol.* **27**, 7991–8002.
71. Dickson J., Gowher H., Strogantsev R., Gaszner M., Hair A., Felsenfeld G., West A.G. (2010) VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* **6**, e10000804.
72. Harland L., Crombie R., Anson S., deBoer J., Ioannou P.A., Antoniou M. (2002) Transcriptional regulation of the human TATA binding protein gene. *Genomics.* **79**, 479–482.
73. Williams S., Mustoe, T., Mulcahy T., Griffiths M., Simpson D., Antoniou M., Irvine A., Mountain A., Crombie R. (2005) CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol.* **5**, 17.
74. Majocchi S., Aritonovska E., Mermoud N. (2014) Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucl. Acids Res.* **42**, 193–204.
75. Simpson D.J., Williams S.G., Irvine A.S. (2009) Expression elements. US Patent No US7632661 B2. Washington, DC: U.S. Patent and Trademark Office.
76. Neville J.J., Orlando J., Mann K., McCloskey B., Antoniou M.N. (2017) Ubiquitous chromatin-opening elements (UCOE): applications in biomanufacturing and gene therapy. *Biotechnol. Adv.* **35**, 557–564.
77. Saunders F., Sweeney B., Antoniou M.N., Stephens P., Cain K. (2015) Chromatin function modifying elements in an industrial antibody production platform-comparison of UCOE, MAR, STAR and cHS4 elements. *PLoS One.* **10**, e0120096.
78. Kwaks T.H., Barnett P., Hemrika W., Siersma T., Sewalt R.G., Satijn D.P.E., Brons J.F., van Blokland R., Kwakman P., Kruckeberg A.L., Kelder A., Otte A.P. (2003) Identification of anti-repressor elements that

- confer high and stable protein production in mammalian cells. *Nat. Biotechnol.* **21**, 553–558.
79. Zboray K., Sommeregger W., Bogner E., Gili A., Sterovsky T., Fauland K., Grabner B., Stiedl P., Moll H.P., Bauer A., Kunert R., Casanova E. (2015) Heterologous protein production using euchromatin-containing expression vectors in mammalian cells. *Nucl. Acids Res.* **43**, e102–e102.
80. Van Keuren M.L., Gavrilina G.B., Filipiak W.E., Zeidler M.G., Saunders T.L. (2009) Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic Res.* **18**, 769–785.
81. Ahmadi S., Davami F., Davoudi N., Nematpour F., Ahmadi M., Ebadat S., Azadmanesh K., Barkhordari F., Mahboudi F. (2017) Monoclonal antibodies expression improvement in CHO cells by PiggyBac transposition regarding vectors ratios and design. *PLoS One.* **12**, e0179902.
82. Dupuy A.J., Fritz S., Largaespada D.A. (2001) Transposition and gene disruption in the male germline of the mouse. *Genesis.* **30**, 82–88.
83. Horie K., Yusa K., Yae K., Odajima J., Fischer S.E., Keng V.W., Hayakawa T., Mizuno S., Kondoh G., Ijiri T., Matsuda Y., Plasterk R.H.A., Takeda J. (2003) Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. *Mol. Cell. Biol.* **23**, 9189–9207.
84. Fraser M.J., Ciszczon T., Elick T., Bauser C. (1996) Precise excision of TTAA specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of *Lepidoptera*. *Insect Mol. Biol.* **5**, 141–151.
85. Li R., Zhuang Y., Han M., Xu T., Wu X. (2013) piggyBac as a high capacity transgenesis and gene-therapy vector in human cells and mice. *Dis. Models. J. Mech.* **6**, 828–833.
86. Wang Y., Yau Y.-Y., Perkins-Balding D., Thompson J.G. (2011) Recombinase technology: applications and possibilities. *Plant Cell Rep.* **30**, 267–285.
87. Lee G., Saito I. (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene.* **216**, 55–65.
88. Turan S., Kuehle J., Schambach A., Baum C., Bode J. (2010) Multiplexing RMCE: versatile extensions of the FLP-recombinase-mediated cassette-exchange technology. *J. Mol. Biol.* **402**, 52–69.
89. Turan S., Galla M., Ernst E., Qiao J., Voelkel C., Bode J. (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J. Mol. Biol.* **407**, 193–221.
90. Baumann M., Gludovacz E., Sealover N., Bahr S., George H., Lin N., Kayser K., Borth N. (2017) Preselection of recombinant gene integration sites enabling high transcription rates in CHO cells using alternate start codons and recombinase mediated cassette exchange. *Biotechnol. Bioengin.* **114**, 2616–2627.
91. Шешукова Е.В., Комарова Т.В., Дорохов Ю.Л. (2016) Растительная фабрика продукции моноклональных антител. *Биохимия.* **81**, 1392–1409.
92. Komarova T.V., Sheshukova E.V., Dorokhov Y.L. (2018) Plant-made antibodies: Properties and therapeutic applications. *Curr. Med. Chem.* **25**, 1–15.
93. Yamauchi T., Iida S. (2015) Gene targeting in crop species with effective selection systems. In: *Advances in New Technology for Targeted Modification of Plant Genomes*. Eds Zhang F., Puchta H., Thomson J.G. New York: Springer, 91–111.
94. Doetschman T., Gregg R.G., Maeda N., Hooper M.L., Melton D.W., Thompson S., Smithies, O. (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature.* **330**, 576–578.
95. Thomas K.R., Capecchi M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell.* **51**, 503–512.
96. Paszkowski J., Baur M., Bogucki A., Potrykus I. (1988) Gene targeting in plants. *EMBO J.* **7**, 4021–4026.
97. Puchta H., Fauser F. (2013) Gene targeting in plants: 25 years later. *Int. J. Dev. Biol.* **57**, 629–637.
98. Shaked H., Melamed-Bessudo C., Levy A.A. (2005) High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. *Proc. Natl. Acad. Sci. USA.* **102**, 12265–12269.
99. Da Ines O., White C.I. (2013) Gene site-specific insertion in plants. In: *Topics in Curr. Genet.* Eds Renault S., Duchateau P. Springer, **23**, 287–315.
100. Chapman J.R., Taylor M.R., Boulton S.J. (2012) Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell.* **47**, 497–510.
101. Voytas D.F. (2013) Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant. Biol.* **64**, 327–350.
102. Urnov F.D., Rebar E.J., Holmes M.C., Zhang H.S., Gregory P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646.
103. Carroll D. (2011) Genome engineering with zinc-finger nucleases. *Genetics.* **188**, 773–782.
104. Christian M., Cermak T., Doyle E.L., Schmidt C., Zhang F., Hummel A., Bogdanov A.J., Voytas D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics.* **186**, 757–761.
105. Christian M., Voytas D.F. (2015) Engineered TAL effector proteins: Versatile reagents for manipulating plant genomes. In: *Advances in New Technology for Targeted Modification of Plant Genomes*. Eds Zhang F., Puchta H., Thomson J.G. New York: Springer, 55–72.
106. Sprink T., Metje J., Hartung F. (2015) Plant genome editing by novel tools: TALEN and other sequence specific nucleases. *Curr. Opin. Biotechnol.* **32**, 47–53.
107. Barnett P. (2018) Transcription activator like effector nucleases (TALENs): A new, important, and versatile gene editing technique with a growing literature. *Sci. Technol. Libraries.* **37**, 100–112.
108. Cong L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W., Marraffini L., Zhang F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science.* 1231143.
109. Makarova K.S., Haft D.H., Barrangou R., Brouns S.J., Charpentier E., Horvath P., Moineau S.A., Mojica F.J.M., Wolf Y.I., Yakunin A.F., van der Oost J.,

- Koonin E.V. (2011) Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477.
110. Ran F.A., Hsu P.D., Wright J., Agarwala V., Scott D.A., Zhang F. (2013) Genome engineering using the CRISPR–Cas9 system. *Nat. Protocols.* **8**, 2281–2308.
111. Hsu P.D., Lander E.S., Zhang F. (2014) Development and applications of CRISPR–Cas9 for genome engineering. *Cell.* **157**, 1262–1278.
112. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816–821.
113. Brazelton Jr, V.A., Zarecor S., Wright D.A., Wang Y., Liu J., Chen K., Yang B., Lawrence-Dill C.J. (2015) A quick guide to CRISPR sgRNA design tools. *GM Crops Food.* **6**, 266–276.
114. Ma X., Zhu Q., Chen Y., Liu Y.G. (2016) CRISPR/Cas9 platforms for genome editing in plants: developments and applications. *Mol. Plant.* **9**, 961–974.
115. Puchta H. (2017) Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Curr. Opin. Plant Biol.* **36**, 1–8.
116. Khan M.H.U., Khan S.U., Muhammad A., Hu L., Yang Y., Fan C. (2018) Induced mutation and epigenetics modification in plants for crop improvement by targeting CRISPR/Cas9 technology. *J. Cell. Physiol.* **233**, 4578–4594.
117. Банников А.В., Лавров А.В. (2017) CRISPR/Cas9 — король геномного редактирования. *Молекуляр. биология.* **51**, 582–594.
118. Bayat H., Modarressi M.H., Rahimpour A. (2018). The conspicuity of CRISPR–Cpf1 system as a significant breakthrough in genome editing. *Curr. Microbiol.* **75**, 107–115.
119. Langner T., Kamoun S., Belhaj K. (2018) CRISPR Crops: Plant genome editing toward disease resistance. *Ann. Rev. Phytopathol.* **56**, 479–512.
120. Lee K., Zhang Y., Kleinstiver B.P., Guo J.A., Aryee M.J., Miller J., Malzahn A., Zarecor S., Lawrence-Dill C.J., Joung J.K., Qi Y., Wang K. (2018) Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnol. J.* <https://doi.org/10.1111/pbi.12982>.
121. Li S., Zhang X., Wang W., Guo X., Wu Z., Du W., Zhao Y., Xia L. (2018). Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. *Mol. Plant.* **11**, 995–998.
122. Gerlach W.L., Bedbrook J.R. (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl. Acids Res.* **7**, 1869–1885.
123. Hemleben V., Zentgraf U. (1994) Structural organization and regulation of transcription by RNA polymerase I of plant nuclear ribosomal RNA genes. In: *Plant Promoters and Transcription Factors*. Berlin, Heidelberg: Springer, 3–24.
124. Garcia S., Cortes P.P., Fernández X., Garnatje T., Kovarik A. (2016) Organization, expression and evolution of rRNA genes in plant genomes. In: *Recent Advances in Pharmaceutical Sciences*. Eds Muñoz-Torres D., Domínguez A., Manresa A. Barcelona: Presas, **VI**, 49–75.
125. Copenhaver G.P., Pikaard C.S. (1996) RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant J.* **9**, 259–272.
126. Chandrasekhara C., Mohannath G., Blevins T., Pontvianne F., Pikaard C.S. (2016) Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* **30**, 177–190.
127. Earley K.W., Pontvianne F., Wierzbicki A.T., Blevins T., Tucker S., Costa-Nunes P., Pontes, O.M., Pikaard C.S. (2010) Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev.* **24**, 1119–1132.
128. Mohannath G., Pontvianne F., Pikaard C.S. (2016) Selective nucleolus organizer inactivation in *Arabidopsis* is a chromosome position-effect phenomenon. *Proc. Natl. Acad. Sci. USA.* **113**, 13426–13431.
129. Gruendler P., Unfried I., Pascher K., Schweizer D. (1991) rDNA intergenic region from *Arabidopsis thaliana*. Structural analysis, intraspecific variation and functional implications. *J. Mol. Biol.* **221**, 1209–1222.
130. Doelling J.H., Gaudino R.J., Pikaard C.S. (1993) Functional analysis of *Arabidopsis thaliana* rRNA gene and spacer promoters *in vivo* and by transient expression. *Proc. Natl. Acad. Sci. USA.* **90**, 7528–7532.
131. Drouin G., De Sa M.M. (1995) The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol. Biol. Evol.* **12**, 481–493.
132. Campell B.R., Song Y., Posch T.E., Cullis C.A., Town C.D. (1992) Sequence and organization of 5S ribosomal RNA encoding genes of *Arabidopsis thaliana*. *Gene.* **112**, 225–228.
133. Franz P., Armstrong S., Alonso-Blanco C., Fischer T.C., Torres-Ruiz R.A., Jones G. (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J.* **13**, 867–876.
134. Cloix C., Tutois S., Yukawa Y., Mathieu O., Cuvillier C., Espagnol M.-C., Picard G., Tourmente S. (2002) Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Res.* **12**, 132–144.
135. Mathieu O., Jasencakova S., Vaillant I., Gendrel A.-V., Colot V., Schubert I., Tourmente S. (2003) Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in *Arabidopsis*. *Plant Cell.* **15**, 2929–2939.
136. Maréchal-Drouard L., Weil J.H., Dietrich A. (1993) Transfer RNAs and transfer RNA genes in plants. *Annu. Rev. Plant Biol.* **44**, 13–32.
137. Beier D., Stange N., Gross H.J., Beier H. (1991) Nuclear tRNA Tyr genes are highly amplified at a single chromosomal site in the genome of *Arabidopsis thaliana*. *Mol. Gen. Genet. MGG.* **225**, 72–80.
138. Arimbasseri A.G., Maraiia R.J. (2016) RNA polymerase III advances: structural and tRNA functional views. *Trends Biochem. Sci.* **41**, 546–559.

139. Eirín-López J.M., González-Romero R., Dryhurst D., Méndez J., Ausió J. (2009) Long-term evolution of histone families: old notions and new insights into their mechanisms of diversification across eukaryotes. In: *Evolutionary Biology*. Ed. Pontarotti P. Berlin, Heidelberg: Springer, 139–162.
140. Chaubet N., Philipps G., Gigot C. (1989) Organization of the histone H3 and H4 multigenic families in maize and in related genomes. *Mol. Gen. Genet. MGG.* **219**, 404–412.
141. Kanazin V., Blake T., Shoemaker R.C. (1996) Organization of the histone H3 genes in soybean, barley and wheat. *Mol. Gen. Genet. MGG.* **250**, 137–147.
142. Li M., Fang Y. (2015) Histone variants: the artists of eukaryotic chromatin. *Sci. China Life Sci.* **58**, 232–239.
143. Ingouff M., Berger F. (2010) Histone3 variants in plants. *Chromosoma.* **119**, 27–33.
144. Kosterin O.E., Bogdanova V.S., Gorel F.L., Rozov S.M., Trusov Y.A., Berdnikov V.A. (1994) Histone H1 of the garden pea (*Pisum sativum* L.); composition, developmental changes, allelic polymorphism and inheritance. *Plant Sci.* **101**, 189–202.
145. Henikoff S., Ahmad K. (2005) Assembly of variant histones into chromatin. *Annu. Rev. Cell. Dev. Biol.* **21**, 133–153.
146. McDowell J.M., Huang S., McKinney E.C., An Y.Q., Meagher R.B. (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana*. *Genetics.* **142**, 587–602.
147. Šlajcherová K., Fišerová J., Fischer L., Schwarzerová K. (2012) Multiple actin isoforms in plants: diverse genes for diverse roles? *Front. Plant Sci.* **3**, 226.
148. Silflow C.D., Oppenheimer D.G., Kopozak S.D., Ploense S.E., Ludwig S.R., Haas N., Peter Snustad D. (1987) Plant tubulin genes: structure and differential expression during development. *Genesis.* **8**, 435–460.
149. Oakley R.V., Wang Y.S., Ramakrishna W., Harding S.A., Tsai C.J. (2007) Differential expansion and expression of α - and β -tubulin gene families in *Populus*. *Plant Physiol.* **145**, 961–973.
150. Breviario D. (2008) Plant tubulin genes: regulatory and evolutionary aspects. In: *Plant Microtubules. Plant Cell Monographs*. Ed. Nick P. Berlin, Heidelberg: Springer, **11**, 207–232.
151. Callis J., Carpenter T., Sun C.W., Vierstra R.D. (1995) Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins in *Arabidopsis thaliana* ecotype Columbia. *Genetics.* **139**, 921–939.
152. Callis J. (2014) The ubiquitination machinery of the ubiquitin system. *The Arabidopsis Book*, **12**, e0174.
153. Radici L., Bianchi M., Crinelli R., Magnani M. (2013) Ubiquitin C gene: structure, function, and transcriptional regulation. *Adv. Biosci. Biotechnol.* **4**, 1057–1062.
154. Bachmair A., Finley D., Varshavsky A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science.* **234**, 179–186.
155. Gonda D., Bachmair A., Wunning I., Tobias J., Lane W., Varshavsky A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712.
156. Hondred D., Walker J.M., Mathews D.E., Vierstra R.D. (1999) Use of ubiquitin fusions to augment protein expression in transgenic plants. *Plant Physiol.* **119**, 713–724.

OPTIMIZATION STRATEGIES OF RECOMBINANT PROTEIN SYNTHESIS IN PLANT CELLS: CLASSICAL APPROACHES AND NEW DIRECTIONS

S. M. Rozov^{1, *}, E. V. Deineko^{1, 2}

¹Federal Research Center Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

²National Tomsk State University, Tomsk, 634050 Russia

*e-mail: rozov@bionet.nsc.ru

Nowadays, pharmacological proteins are synthesized in various expression systems including bacterial, mammalian and insect cell cultures. Plant expression systems, especially plant cell suspension cultures, advantageously combine the simplicity and cheapness of bacterial systems and the ability to perform post-translational modifications like mammalian cell cultures do. One of the drawbacks of plant expression systems is the lower yield of the recombinant protein of interest when compared with bacterial systems. This review analyses the techniques developed over the past two decades to increase the expression levels of recombinant genes that have already become classic, and ways to prevent gene silencing caused by accidental insertion of the genes of interest into heterochromatin sites. With the advent of CRISPR/Cas9 technology, it became possible to create a new approach to increase expression levels of the transgenes encoding for pharmacological proteins by targeting these genes into specific genome sites known to be highly transcriptionally active. Such target sites can be the loci of the “housekeeping” genes of the plant cell, which are actively expressed throughout the interphase. The organization of some “housekeeping” genes, the most promising for insertion of the transgenes, is reviewed in detail.

Keywords: recombinant genes, expression optimization, pharmaceutical proteins, plant cell suspension cultures, plant expression systems, expression cassettes, position effect, gene silencing, CRISPR/Cas9, directed mutagenesis, genome editing, housekeeping genes