

УДК 577.2;577.21

# Flow-seq-метод: особенности и применение в изучении бактериальной трансляции

Е. С. Комарова<sup>1</sup>, О. А. Донцова<sup>2,3,4,5</sup>, Д. В. Пышный<sup>6</sup>, М. Р. Кабилов<sup>6</sup>, П. В. Сергиев<sup>1,2,3,4</sup>

<sup>1</sup>Московский государственный университет имени М.В. Ломоносова, Институт функциональной геномики, Москва, 119234 Россия

<sup>2</sup>Московский государственный университет имени М.В. Ломоносова, химический факультет, Москва, 119234 Россия

<sup>3</sup>Сколковский институт науки и технологии, Центр наук о жизни, Сколково, 121205 Россия

<sup>4</sup>Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский институт физико-химической биологии имени А.Н. Белозерского, Москва, 119234 Россия

<sup>5</sup>Институт биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова РАН, Москва, 117997 Россия

<sup>6</sup>Институт химической биологии и фундаментальной медицины СО РАН, Центр коллективного пользования «Геномика», Новосибирск, 630090 Россия

\*E-mail: petya@genebee.msu.ru, kabilov@niboch.nsc.ru

Поступила в редакцию 04.10.2022

Принята к печати 11.11.2022

DOI: 10.32607/actanaturae.11820

**РЕФЕРАТ** Метод Flow-seq основан на использовании библиотек репортерных конструкций, где определенный элемент, регулирующий экспрессию генов флуоресцентных репортерных белков, представлен во многих тысячах вариантов. Библиотеки репортерных конструкций внедряют в клетки, сортируют по уровню флуоресценции, а затем проводят их высокопроизводительное секвенирование. Таким образом, появляется возможность в одном эксперименте на десятках и сотнях тысяч репортерных конструкций выявлять закономерности, определяющие эффективность экспрессии. Этот метод нашел активное применение в оценке эффективности биосинтеза белка множеством вариантов мРНК одновременно. Однако этим возможностями метода Flow-seq не исчерпываются. В представленном обзоре проведен сравнительный анализ метода Flow-seq и других методов, используемых для оценки эффективности трансляции мРНК, а также рассмотрены особенности применения и результаты, получаемые с помощью Flow-seq.

**КЛЮЧЕВЫЕ СЛОВА** Flow-seq, NGS, высокопроизводительное секвенирование, проточная цитометрия, трансляция, бактерии.

**СПИСОК СОКРАЩЕНИЙ** TIR – Translation Initiation Region (участок инициации трансляции); RBS – Ribosome Binding Site (сайт посадки/связывания рибосомы на мРНК); SD – Shine-Dalgarno sequence (последовательность Шайна–Дальгарно); 5'-UTR – 5'-untranslated region (5'-нетранслируемая область); ORF – open reading frame (открытая рамка считывания); NGS – next-generation sequencing (высокопроизводительное секвенирование/секвенирование нового поколения/глубокое секвенирование); Flow-seq – flow cytometry and next-generation sequencing (метод, совмещающий проточную цитометрию и высокопроизводительное секвенирование).

## ВВЕДЕНИЕ

Трансляция – ключевой процесс в жизнедеятельности всех организмов, в ходе которого в клетках с помощью сложного макромолекулярного рибонуклеопротеидного комплекса, называемого рибосомой, синтезируются белки. Рибосома декодирует информацию, записанную в мРНК, и переводит ее в последовательность аминокислот, образующих белок [1].

При этом мРНК не только участвует в данном процессе в качестве пассивного носителя информации, но и предопределяет эффективность трансляции [2].

Одним из участков мРНК, определяющих эффективность ее трансляции, является 5'-нетранслируемая область (5'-UTR) (рис. 1А) [3]. В 5'-UTR находится сайт посадки рибосомы (RBS), содержащий в канонических мРНК последовательность

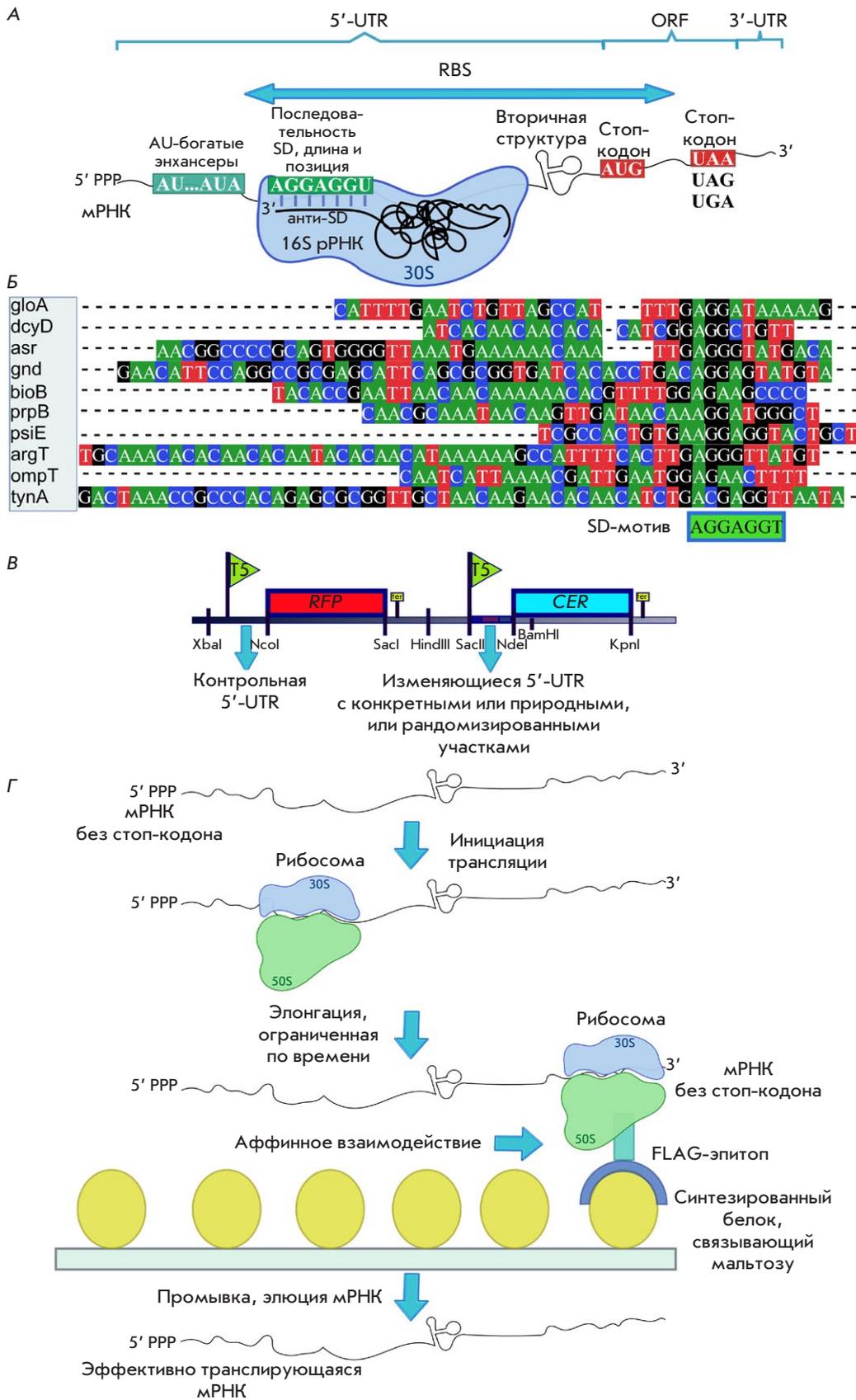
Шайна–Дальгарно (SD) [4–13], комплементарную 3′-концу 16S рРНК [14, 15]. Для высокой эффективности биосинтеза белка SD должна располагаться на оптимальном расстоянии от стартового кодона и иметь оптимальную длину [16–18]. Иногда в одной 5′-UTR можно найти несколько последовательностей Шайна–Дальгарно [2, 17]. Для эффективной трансляции участок инициации трансляции (TIR) должен быть либо полностью одноцепочечным, либо укладываться во вторичную структуру, которую легко разрушить [19–22]. Известны и другие элементы, способные влиять на эффективность трансляции, такие, как богатая остатками аденина и урацила (AU-богатая) область мРНК, с которой взаимодействует рибосомный белок bS1 [23–25], а также начальный участок кодирующей области, следующий сразу за стартовым кодоном [26–28]. Для 5′-UTR эффективно транслируемых мРНК характерно низкое содержание остатков цитидина и наличие повторов с пуриновыми основаниями (AG-повторов) [2].

На сегодняшний день известны различные методы, позволяющие изучать функциональную значимость отдельных участков мРНК для биосинтеза белка. В число этих методов входят сайт-направленный мутагенез [29] или рандомизация [30, 31] участков в 5′-UTR, как правило, перед геном флуоресцентного белка и оценка интенсивности его свечения *in vitro* (или *in vivo*), отражающей эффективность трансляции. С применением термодинамического моделирования *in silico* [18, 32–36], оценивающего силу молекулярных взаимодействий комплекса 30S с мРНК-транскриптом и предсказывающего скорость инициации трансляции, можно определить величины, связанные с эффективностью трансляции. Результаты моделирования можно выборочно проверить экспериментально с использованием репортерных конструкций. Появление такого метода, как проточная цитометрия, позволило оценивать одновременно различные параметры *in vivo* большого количества клеток и по сходству определенных параметров, например, по уровню экспрессии гена флуоресцентного белка, выделять отдельные фракции [37]. Развитие высокопроизводительного секвенирования (NGS) способствовало разработке новых сложных подходов к изучению геномов, определению корреляции между генотипом и фенотипом, в частности таких, как полногеномное секвенирование, секвенирование библиотек плазмидных ДНК, РНК-секвенирование для профайлинга клеточного транскриптома и выделения эффективно транслирующихся мРНК, ChIP-секвенирование для определения сайтов связывания ДНК-ассоциированных белков [38, 39].

## РАЗНООБРАЗИЕ ПОДХОДОВ К ИЗУЧЕНИЮ РОЛИ 5′-UTR В ЭФФЕКТИВНОСТИ ТРАНСЛЯЦИИ

Широкомасштабный анализ генов *E. coli* показал, что большинство мРНК содержат последовательность Шайна–Дальгарно (SD) (рис. 1B), открытую в 1970-х в нескольких бактериальных мРНК [4] и существенной для эффективной инициации трансляции [16–18]. Последовательность SD является самым изученным регуляторным элементом. SD располагается на расстоянии 5–8 нуклеотидов перед стартовым кодоном (8–11 нуклеотидов, если считать от центрального G в SD [7]) и служит местом взаимодействия бактериальной 30S субчастицы, в отличие от эукариотической рибосомы, связывающейся на 5′-конце мРНК для начала сканирования [6]. Разные мРНК *E. coli* содержат SD разной длины, варьирующей в пределах 4–8 нуклеотидов. При этом наиболее вероятный состав последовательности SD agGa.

Зависимость эффективности биосинтеза белка от длины последовательности SD и ее расстояния от старт-кодона изучали различными методами. Например, с помощью двойной генно-инженерной конструкции (рис. 1B), содержащей гены двух флуоресцентных белков, где первый, RFP (красный флуоресцентный белок), служил внутренним контролем, а второй, CER (циановый флуоресцентный белок), выступал в качестве сенсора эффектов, связанных с изменением элементов 5′-UTR мРНК [17]. При этом вычисляли отношение измеренных интенсивностей флуоресценции двух белков CER/RFP *in vivo*, что позволяло нивелировать эффекты, обусловленные размером бактериальной клетки и флуктуациями копийности репортерной плазмиды. Этот подход, основанный на молекулярном клонировании, в котором используются 16 репортерных конструкций с четырьмя последовательностями SD (2, 4, 6 и 8) разной длины, расположенными на разных расстояниях от инициаторного кодона гена белка CER (7, 10, 13 и 16), и еще одна контрольная конструкция без комплементарных к анти-SD участков, позволил экспериментально изучить влияние длины SD, расстояния от SD до стартового кодона и их сочетаний на биосинтез белка CER. Таким образом показано, что эффективность трансляции мРНК с SD длиной 8 нуклеотидов снижается с увеличением расстояния от старт-кодона до SD, для SD длиной 6 нуклеотидов оптимально расстояние в 10 нуклеотидов. В случае SD средней длины (4 нуклеотида) наблюдается такая же зависимость, как у длинной (8 нуклеотидов). В случае короткой SD (2 нуклеотида) влияние расстояния несущественно, при этом роль этой SD в эффективности биосинтеза белка сохраняется: обеспечивает на порядок большую эф-



**Рис. 1.** А – особенности строения мРНК у бактерий. 5'- и 3'-UTR, 5'- и 3'-нетранспируемые области соответственно. RBS – сайт посадки рибосомы. ORF – открытая рамка считывания, содержащая белок-кодирующую последовательность. SD и анти-SD – последовательности Шайна–Дальгарно и анти-Шайна–Дальгарно соответственно. Б – пример выравнивания последовательностей 5'-UTR мРНК, используемого при широкомасштабном анализе нетранспируемых областей генов с выделением SD-мотива. В – пример двойной репортерной конструкции с контрольной 5'-UTR перед геном флуоресцентного белка RFP и изменяющейся 5'-UTR перед геном второго флуоресцентного сенсорного белка CER для оценки влияния особенностей варибельного участка на эффективность трансляции. Г – схема аффинного выделения рибосом с эффективно транскрибирующейся мРНК. Отбор проводился лимитированием времени *in vitro* трансляции. мРНК содержит 5'-UTR, кодирующую область, включающую участок, кодирующий FLAG-эпитоп, который взаимодействует с синтезируемым белком, связывающим мальтозу, и TolA, который позволяет эпитопу выйти из рибосомного туннеля и правильно свернуться. В конструкции нет стоп-кодона, поэтому рибосома остается на мРНК. Рисунок выполнен в программе Inkscaper

фективность, чем контрольная конструкция без SD. Варьирование этих параметров дает возможность изменять уровень трансляции вплоть до четырех порядков, что указывает на их важность в определении уровня многих белков в клетке [17].

Для быстрого и достаточно простого количественного анализа экспрессии генов *in vitro* можно использовать многочисленные варианты участка в 5'-UTR, получаемые с помощью сайт-направленного мутагенеза, на основе полимеразной цепной реакции (ПЦР). При этом продукт ПЦР, содержащий T7-промотор, тестируемый вариант 5'-UTR и ген флуоресцентного белка eGFP, напрямую используется в системе сопряженной транскрипции-трансляции *in vitro* из клеток *E. coli* [29]. Эффективность трансляции в такой системе можно оценить по интенсивности флуоресценции белка eGFP. Таким образом были получены 54 варианта последовательностей 5'-UTR, из них 18 и 36 с измененными SD- и AU-богатыми последовательностями соответственно, которые обеспечили диапазон относительных уровней экспрессии от 0.1 до 2.0 и показали влияние различных сайтов посадки рибосомы (RBS) на эффективность трансляции [29]. Однако такой точный подход сильно ограничен небольшим набором проверяемых вариантов, не позволяя приблизиться ко всему разнообразию природных 5'-UTR, предшествующих генам, число которых у *E. coli* составляет около  $4 \times 10^3$  [8].

Позднее разработали экспериментальную систему (*рис. 1Г*) [30], основанную на трансляции *in vitro*, которая позволила отбирать наиболее эффективно транслируемые мРНК из большой выборки искусственных последовательностей. Для этого использовали модельную мРНК, содержащую 5'-UTR из 81 нуклеотида, 18 из которых, локализованных перед стартовым кодоном, были полностью рандомизированы, что помогло получить библиотеку из  $\sim 6.9 \times 10^{10}$  различных последовательностей. Модельная мРНК кодирует гибридный белок, содержащий примерно в середине домен, связывающий мальтозу, и FLAG-эпитоп, позволяющий проводить аффинную очистку рибосом, синтезирующих этот гибридный белок. После домена, используемого для аффинной очистки, находится фрагмент белка TolA, служащий исключительно спейсером, достаточным для того, чтобы по окончании синтеза полноразмерного гибридного белка из пептидного туннеля был экспонирован аффинный домен. Эта мРНК не содержит стоп-кодон, поэтому она оставалась связанной с рибосомой после завершения синтеза. Таким образом, мРНК могла быть экстрагирована из аффинно-связанных рибосом и в дальнейшем амплифицирована. Ключевым параметром

отбора мРНК служило ограниченное время трансляции: только мРНК, транслируемые быстро, могли быть аффинно очищены и использованы в следующем раунде селекции [30]. Удивительно, что из отобранных последовательностей, обеспечивающих наиболее быструю трансляцию в *in vitro* системе, 76% не содержали SD и имели С-богатые короткие участки, комплементарные 16S рРНК. Однако мРНК с такими С-богатыми последовательностями не экспрессировались на высоком уровне *in vivo*, возможно, из-за разного среднестатистического соотношения в *in vitro* и *in vivo* системах рибосом и мРНК, конкурирующих, наряду с С-богатыми RBS, за связывание с рибосомами [30]. Тот же самый эксперимент проведен с использованием библиотеки более коротких мРНК с 5'-UTR длиной 40 нуклеотидов [31], наиболее частой в мРНК *E. coli* [40, 41]. Применение секвенирования нового поколения и статистических методов позволило определить мотивы связывания мРНК и рибосомы. Отобранные по скорости трансляции мРНК из библиотеки с более короткими 5'-UTR чаще содержали SD наряду с G/U-богатыми последовательностями [31]. Полученные результаты также указывают на влияние длины 5'-UTR на эффективность инициации биосинтеза белка.

Последовательность 5'-UTR мРНК может определять различную укладку участка перед стартовым кодоном. Связь между стабильностью вторичных структур в TIR и эффективностью трансляции подтвердили крупномасштабным вычислительным анализом [19], указывающим на тенденцию прокариотических и эукариотических генов, особенно с высоким уровнем экспрессии, к дестабилизации вторичной структуры мРНК вблизи инициаторного кодона [20]. Варьирование стабильности ( $< -12$  ккал/моль) шпильчатой структуры, содержащей RBS, с помощью сайт-направленного мутагенеза и последующий анализ *in vivo* выхода белка показали, что чем выше стабильность вторичной структуры, содержащей участок связывания рибосомы, тем ниже эффективность трансляции. Таким образом, показана возможность 500-кратного изменения экспрессии при замене только одного нуклеотида, приводящей к стабилизации вторичной структуры мРНК. Как следствие, наблюдалась полная зависимость начала трансляции от спонтанного разворачивания всего участка инициации мРНК [21]. Однако эта спонтанность была обусловлена присутствием всех необходимых элементов комплекса инициации [22]. Такого рода анализ 12 мРНК с разной стабильностью вторичной структуры и с SD разной длины либо без SD выявил необходимость самой SD, старт-кодона, инициаторной тРНК с форми-

лированным метионином, фактора инициации IF2, связанного с GTP, в комплексе с 30S субчастицей рибосомы для разворачивания вторичных структур мРНК. Вклад каждого индивидуального элемента в нарушение укладки TIR мРНК оценен с использованием константы диссоциации фрагмента мРНК с последовательностью SD из 6 нуклеотидов [22]. Далее методом FRET проведен анализ этого же фрагмента с 5'- и 3'-концами, меченными Су3 и Су5, при добавлении 30S субъединицы вместе со всеми элементами, необходимыми для инициации трансляции. Оценка проводилась относительно контрольной мРНК без SD, но с близкой по стабильности вторичной структурой. В результате анализа была выявлена значимость последовательности SD в разворачивании вторичной структуры мРНК. При этом показана высокая эффективность FRET для свернутой мРНК, концы которой образовывали комплементарное взаимодействие SD с анти-SD, и слабая для развернутой формы мРНК [22].

Эффективность связывания субъединиц рибосомы с той или иной последовательностью 5'-UTR мРНК оценивают с помощью так называемого метода тоупринтинга/тоупринта (рис. 2А). Этот метод основан на использовании флуоресцентно или изотопно меченных праймеров, комплементарных 3'-концу мРНК. После сборки инициаторного комплекса на мРНК проводится реакция обратной транскрипции с последующим электрофоретическим анализом удлиненных кДНК в реакционной смеси. Обратная транскриптаза доходит до 5'-конца мРНК, если мРНК не связана с рибосомой, и образует более короткие продукты при остановке ревертазы, встретившей рибосому. По соотношению длинных и укороченных тоупринтов можно судить о доле мРНК, связавшихся с рибосомой [42, 43].

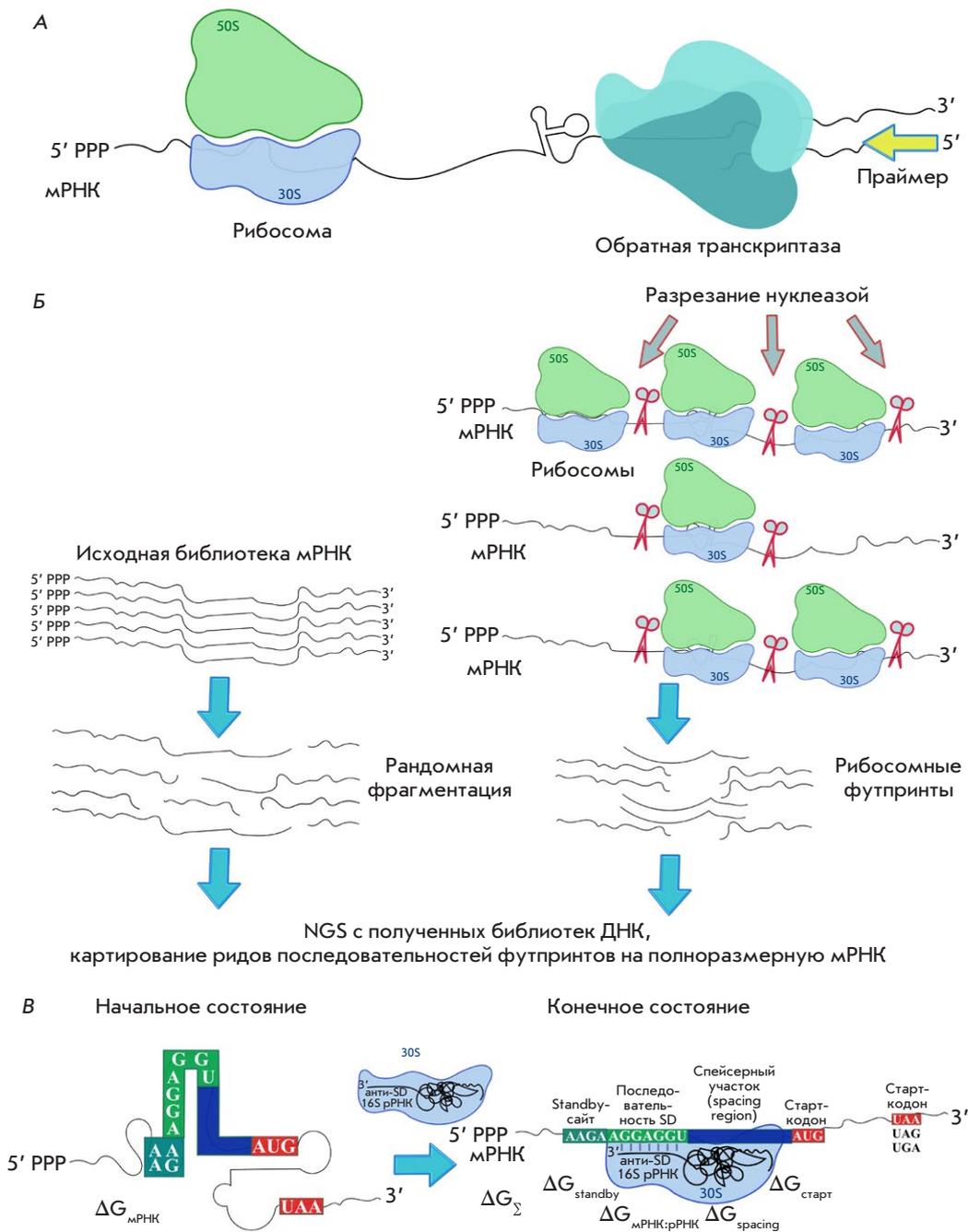
По мере накопления экспериментальных результатов и развития методов их анализа все большую роль начинают играть биоинформатические подходы, позволяющие работать с большими массивами данных. Наблюдаемая в ходе экспериментов инициация трансляции прокариотических мРНК, в 5'-UTR которых не определена последовательность SD, происходит независимо от взаимодействий с анти-SD и под влиянием рибосомного белка bS1. При этом биоинформатический анализ показал снижение стабильности вторичных структур таких последовательностей 5'-UTR, что облегчает образование инициаторного комплекса, компенсируя отсутствие взаимодействий SD и анти-SD [44, 45].

Известно, что существуют так называемые безлидерные прокариотические мРНК, которые не содержат 5'-UTR, а следовательно, и SD. При этом широкомасштабный *in silico* анализ макроэволюции

показывает уменьшение числа таких генов у бактерий в ходе эволюции. Рассмотрены участки инициации трансляции всех генов 953 бактериальных и 72 архейных геномов, распределенных по группам, согласно их расстояниям до корня (между бактериями и археями) на филогенетическом древе 16S рРНК. Была рассчитана средняя доля «генов без лидера» в каждой группе: сначала она быстро падает, а затем колеблется на низком уровне [46].

Активное развитие методов секвенирования нового поколения и накопленные навыки работы с системой трансляции позволили разработать метод рибосомного профилирования/профайлинга (Ribo-seq) (рис. 2Б), основанного на высокопроизводительном секвенировании фрагментов мРНК, защищаемых транслирующей рибосомой [47]. Этот подход оказался эффективным для изучения экспрессии генов на уровне и транскрипции, и трансляции одновременно, в том числе в ответ на различные воздействия [48–50]. Метод Ribo-seq дает информацию о местоположении рибосом на мРНК с однонуклеотидным разрешением. Такая точность позволяет не только выявлять трансляцию участков мРНК вне аннотированной рамки считывания, но и трансляцию перекрывающихся рамок считывания и смысловое декодирование стоп-кодонов. С помощью Ribo-seq транслируемые рамки считывания выявлены в РНК, ранее считавшихся некодирующими. Также оказалось возможным оценивать влияние разных условий и факторов на трансляцию мРНК в клетках, например, разных сред, модификаций исследуемых белков, антибиотиков [51–56].

Активное использование метода Ribo-seq позволило выявить ряд сложностей и артефактов, связанных с методологией эксперимента и обработкой результатов [57–59]. Многообещающая технология рибосомного профилирования для изучения скорости декодирования рибосом характеризуется наличием нечастых высоких пиков плотности рибосомного футпринта и длинных промежутков при выравнивании соответствующих последовательностей мРНК. Для уменьшения влияния неоднородности данных разработан способ нормализации, эффективный при наличии гетерогенного шума, позволивший выявить существенные различия в распределении ридов (прочитанных последовательностей) по мРНК, и детерминант частот рибосомного футпринта в 30 общедоступных наборах данных рибосомного профайлинга, тем самым ставя под сомнение надежность данного метода как точного представления локальной плотности рибосом без предварительного контроля качества [57]. Такое наблюдение указывает на неполное понимание того, как параметры протокола влияют на плотность рибосомных футпринтов.



**Рис. 2.** А – принцип метода тоупринт. Стабильные рибосомные комплексы останавливают обратную транскриптазу на определенной позиции на мРНК, генерируя короткие кДНК-продукты определенной длины. В качестве праймеров для обратной транскриптазы могут служить радиоактивно или флуоресцентно меченные. Б – схема метода рибосомного профилирования /профайлинга (Ribo-seq). После инициации трансляции проводят разрезание мРНК в местах, где она не защищена рибосомами, специфичной нуклеазой. Параллельно готовят библиотеку исходной мРНК для секвенирования путем рандомной фрагментации. Она будет использоваться в качестве референсных последовательностей. Все полученные рибосомные футпринты используют для приготовления библиотеки ДНК, которую в дальнейшем глубоко секвенируют. По результатам NGS проводят картирование ридов последовательностей футпринтов на полноразмерную мРНК. В – термодинамическая модель бактериальной инициации трансляции. Изменение свободной энергии в течение стадии инициации определяется пятью типами молекулярных взаимодействий, определяющих начальное и конечное положение системы. Рисунок выполнен в программе Inkscape

Возможно, наиболее известная причина этого наблюдения заключается в смещении последовательностей, возникающем во время создания библиотеки рибосомных футпринтов и ее конверсии в кДНК, и последующего секвенирования [58]. Перечисленные стадии вовлекают ряд реакций с использованием ферментов, обладающих сиквенс-специфичностью, как например, нуклеазы [60]. В то же время некоторые антибиотики, которыми рибосомы обрабатывают перед профилированием, обладают такой же сиквенс-специфичностью [61–63], что необходимо учитывать при постановке эксперимента.

С применением метода рибосомного профайлинга на бактериях показано существенное увеличение покрытия рибосомами после последовательностей Шайна–Дальгарно, встречающихся по случайным причинам в кодирующей области [64]. В то время как SD, предшествующие старт-кодону, играют хорошо охарактеризованную роль в инициации трансляции, полученные данные свидетельствуют о том, что элонгация замедляется временным образованием пар оснований между мотивами SD внутри открытых рамок считывания и последовательностью анти-SD в 16S рРНК, причем такие паузы составляли более 70% сильных остановок по всему геному, что привело к выводу о том, что такие паузы были основной детерминантой трансляционных остановок у бактерий [64].

Позднее с использованием усовершенствованного варианта метода Ribo-seq с улучшенным разрешением показали, что ранее наблюдаемое обогащение покрытия мотивов SD рибосомами можно объяснить паузами на глициновых кодонах и невозможностью выделить всю популяцию фрагментов мРНК, защищенных рибосомой. Сделан вывод о том, что мотивы SD не являются, вероятно, основной причиной множества трансляционных остановок *in vivo* [65].

Биофизические модели позволяют оценить эффективность взаимодействия биомолекул, в том числе и мРНК с рибосомой. Примером может служить термодинамическая модель (рис. 2B) [32], которая одновременно оценивает силу молекулярных взаимодействий 30S комплекса с мРНК-транскриптом, рассчитывает свободную энергию Гиббса для каждого элемента в составе конкретной мРНК и предсказывает скорость инициации трансляции: чем большую энергию надо затратить для разворачивания элементов мРНК, тем ниже скорость инициации трансляции. Представленная модель может быть использована как для предсказания относительной скорости инициации трансляции существующей 5'-UTR с определенным RBS, так и для дизайна последовательности RBS, обеспечивающей требуемую скорость инициации трансляции [18, 32].

Применение метода Flow-seq для библиотеки плазмид, содержащих гены флуоресцентных белков, первый из которых служит внутренним контролем, а экспрессия второго изменяется в зависимости от влияния последовательностей, полученных в результате рандомизации 30 нуклеотидов в кодирующей области гена сразу после стартового кодона, позволило разделить полученную библиотеку (более  $30 \times 10^3$  вариантов мРНК) по эффективности трансляции [28]. Дальнейший анализ показал снижение эффективности трансляции мРНК, содержащих SD-подобные последовательности, и уменьшение доли таких мРНК в наборе эффективно транслирующихся, демонстрируя отрицательное влияние SD в данной области мРНК на биосинтез белка и поддерживая, в свою очередь, более ранние наблюдения, полученные на ограниченном наборе модельных мРНК [66].

Примечательно, что распределение энергий связывания анти-SD-последовательностей среди мРНК с эффективной трансляцией сходно с распределением у природных генов *E. coli*. Более того, созданы отдельные конструкции с SD в скользящем окне начального кодирующего участка сразу после старт-кодона и со схожей энергией укладки вторичной структуры, эффективность трансляции которых оценена. Таким образом, получены результаты, согласующиеся с результатами анализа данных после применения метода Flow-seq [28].

### СХЕМА МЕТОДА FLOW-seq, ОСОБЕННОСТИ И РЕЗУЛЬТАТЫ ЕГО ПРИМЕНЕНИЯ

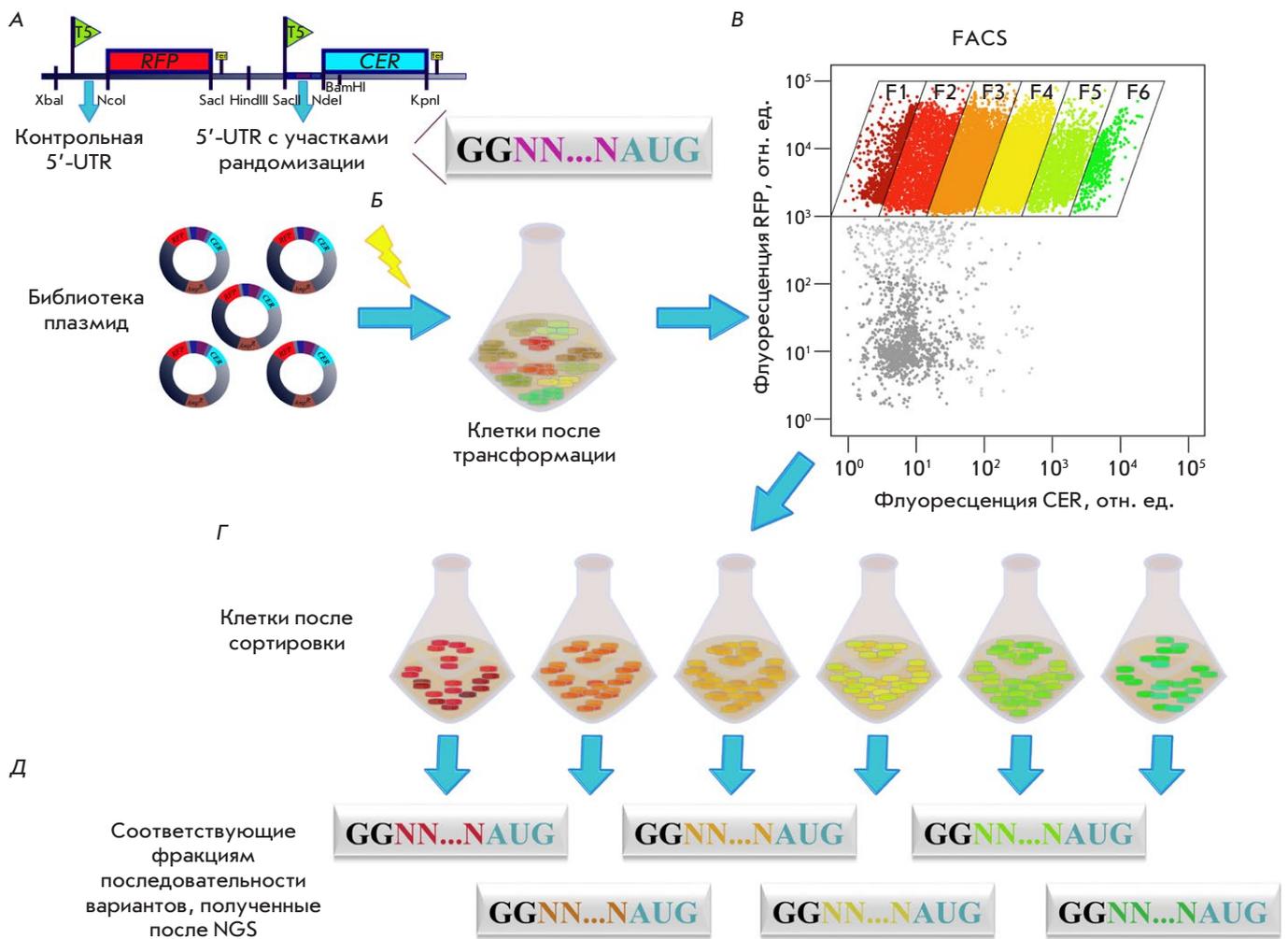
С целью определения влияния какого-либо фактора или набора факторов на уровень экспрессии того или иного гена часто используют тысячи репортерных конструкций, перебирающих различные варианты промоторов, 5'-нетранслируемых областей, а также отдельных участков в них, включая сайты посадки рибосомы (RBS), предшествующие им области (standby sites) или следующие за ними спейсерные участки, начальных «разгонных» участков кодирующей области (ramp region), по отдельности или одновременно (табл. 1). Как правило, такие плазмиды содержат два гена флуоресцентных белков: первый из них выступает в роли сенсора, экспрессия которого чувствительна к переменным участкам, а второй используется в качестве неизменного внутреннего контроля. Полученными наборами конструкций трансформируют штамм бактерий, подходящий для последующей экспрессии и сортировки. Далее с помощью проточной цитометрии оценивают интенсивности флуоресценции двух белков в пуле клеток и формируют группы/фракции клеток с приблизительно одинаковым отношением измеренных

Таблица 1. Применение метода Flow-seq для анализа эффективности трансляции

Элементы мРНК	Количество вариантов в полученных библиотеках в итоге после Flow-seq	Типы вариантов	Результаты	Ссылка
Промоторы и сайты посадки рибосомы (RBS) в 5'-UTR	11894 (94%) из 12653 возможных с комбинациями 114 промоторов и 111 RBS (одна комбинация привела к несовместимому сайту рестрикции)	Взяты из доступных баз данных и сгенерированные с помощью RBS Calculator	Диапазон изменения экспрессии $10^4$ раз. Выбор промотора оказывает наибольшее влияние на уровень РНК и в меньшей степени на уровень белка, так как на эффективность его трансляции влияет и выбор сайта посадки рибосомы, и, возможно, другие факторы. 55% из нескольких сотен проверенных отдельных колоний безошибочно определены при Flow-seq-анализе	[67]
Промоторы и сайты посадки рибосомы (RBS) в 5'-НТО	~ 500 комбинаций из 14 промоторов и 22 RBS для двух детектируемых флуоресцентных белков и более 1200 комбинаций из рандомизированной библиотеки	Конкретные варианты и с участками рандомизации в изучаемых элементах	Динамический диапазон экспрессии $10^3$ раз. Получаемые комбинации приводят к экспрессии произвольного гена в пределах двукратного изменения уровня с надежностью 93%	[75]
Шесть нуклеотидов в спейсерном участке за SD в 5'-UTR и перед старт-кодоном и первые шесть нуклеотидов, следующих за ним (кодоны в положениях +2 и +3 кодирующей области (CDS)) ...-SD-GAC-6N-AUG-6N <sup>syn</sup> -...	13914 (56%) для одного белка и 25861 (53%) для другого из 24576 и 49152 возможных соответственно	Рандомизированные спейсерные участки и кодоны в положениях +2 и +3 с синонимическими заменами, не изменяющими кодирующую последовательность двух сенсорных белков	Диапазон изменения экспрессии $10^3$ раз. Низкое содержание GC-нуклеотидов и уменьшение стабильности вторичной структуры изучаемых элементов важны для высокого уровня экспрессии, не ограниченного данными детерминантами. Распределение уровней флуоресценции белка, измеренных в нескольких десятках колоний с помощью планшетного ридера, совпадает с результатами Flow-seq	[71]
Четыре нуклеотида в спейсерном участке за SD в 5'-UTR и перед старт-кодоном ...-SD-C-4N-CAU-AUG-...	249 (97%) из 256 возможных	Рандомизированные	Диапазон изменения экспрессии $10^2$ раз. Преобладание доли аденозинов и уменьшение доли цитидинов в эффективно транскрибирующихся вариантах. Низкое содержание GC-нуклеотидов и уменьшение стабильности вторичной структуры изучаемых элементов важны для высокого уровня экспрессии. SD-подобные последовательности также встречаются только в вариантах с высокой экспрессией	[39]
SD длиной 6 нуклеотидов в 5'-UTR	4066 (99%) из 4096 возможных	Рандомизированные	Измеренные уровни белков (флуоресцентного и пяти природных) для 91% вариантов последовательностей были в пределах двукратного изменения предсказанного уровня экспрессии в EMOPEC, учитывающего контекстное окружение SD, сводящего к минимуму изменения вторичной структуры	[76]
Standby-сайты различной длины (20–164 нуклеотида) перед SD, дистальные в 5'-UTR	136 5'-UTR с разной длиной и вторичной структурой, формой и числом модулей	Смоделированные варианты	Диапазон изменения эффективности трансляции $10^2$ раз. Скорость инициации трансляции мРНК контролируется площадью поверхности одноцепочечных участков, частичным разворачиванием структуры РНК для минимизации штрафа по свободной энергии связывания рибосомы, отсутствует кооперативное связывание и, возможно, скольжение рибосомы в рассматриваемом участке. Разработана и проверена экспериментально биофизическая модель предсказания скорости инициации трансляции. Рибосома может легко связываться с отдаленными от старт-кодона модулями standby-сайтов и обеспечивать высокую эффективность трансляции	[34]
Сайт посадки рибосомы (RBS) в 5'-UTR с фиксированной последовательностью SD (5 нуклеотидов) и изменяющимся standby-сайтом (4 нуклеотида) и спейсерным участком из 6 нуклеотидов RRRV-AGGAG-R-6N-AUG (R: A/G, V: A/G/C, N:A/U/C/G)	Более 20000 (10%) из ~ 200000 возможных вариантов для двух флуоресцентных белков	Рандомизированные и частично конкретные позиции с неполными вариациями	Диапазон изменения эффективности трансляции $10^4$ раз. На эффективность трансляции значительно влияет консервативность SD, тогда как AC-богатый спейсерный участок слабо зависит от контекста. Для высокой экспрессии наблюдалась низкая стабильность вторичной структуры изучаемого участка. Замена репортерного белка другим зачастую не влияла на общую тенденцию распределения последовательностей, задающих тот или иной уровень биосинтеза белка	[74]

Таблица 1 (продолжение).

Элементы мРНК	Количество вариантов в полученных библиотеках в итоге после Flow-seq	Типы вариантов	Результаты	Ссылка
Почти полная последовательность 5'-UTR длиной 22 или 32 нуклеотида GG-20N/30N-AUG...	11692 (10 <sup>-6%</sup> из возможных вариантов), 11889 (10 <sup>-12%</sup> ) для 20N и 30N, соответственно, 48 вариантов природных с вариациями	Рандомизированные, природные, конкретные	Диапазон изменения эффективности трансляции 10 <sup>4</sup> -кратный. Наблюдались низкая стабильность вторичной структуры и консервативность последовательности SD у вариантов с высокой экспрессией. Наличие AU-богатых энхансеров на 5'-конце в standby-сайте, низкая доля цитидиновых оснований, множественные SD, AG-повторы в 5'-UTR мРНК в отдельных случаях обеспечивают высокую эффективность трансляции	[2]
Последовательности 5'-UTR длиной 2–60 нуклеотидов первых генов оперонов <i>E. coli</i> с GG на самом 5'-конце, сохранившихся при транскрипции GG-natural 5'-UTR	648 (91%) из 713 возможных длиной 2–60 нуклеотидов, (45%) из всех 1451 природных 5'-UTR первых генов оперонов	Природные	Диапазон изменения эффективности трансляции 30-кратный. Наблюдалось влияние вторичной структуры РНК и последовательности SD на эффективность трансляции, но с меньшей вариабельностью по сравнению с рандомизированными библиотеками. Низкая стабильность вторичной структуры и консервативность последовательности SD в вариантах с высокой экспрессией. Результаты оценки эффективности трансляции в случае с теми или иными 5'-UTR коррелировали с результатами, полученными с помощью рибосомного профилирования	[77]
Участки в промоторной области, standby-сайт длиной 10/20/30 нуклеотидов, спейсерный участок из 8 нуклеотидов 10N/20N/30N-SD-8N	~ 12000 (очень маленький процент из возможных)	Рандомизированные	Диапазон изменения эффективности трансляции 10 <sup>3</sup> раз. При высоком уровне экспрессии наблюдалась низкая стабильность вторичной структуры изучаемого участка	[72]
Промоторы, сайты посадки рибосомы (RBS), первые 13 аминокислот кодирующей области белка	14234 комбинации из двух промоторов, четырех сайтов посадки рибосомы (RBS) и последовательностей N-концевых пептидов, соответствующих первым 13 аминокислотам 137 природных генов <i>E. coli</i>	Природные	Диапазон изменения эффективности трансляции более 10 <sup>2</sup> . Использование редких кодонов на N-конце может увеличивать экспрессию в 14 раз независимо от RBS, обеспечивающих ту или иную эффективность трансляции. Уменьшение стабильности вторичной структуры, а не сама редкость кодонов отвечает за увеличение экспрессии	[78]
Первые шесть кодонов, следующих за старт-кодоном, в кодирующей области	10	Природные	Уменьшение стабильности вторичной структуры, а не сама редкость кодонов отвечает за увеличение эффективности трансляции. Редкие кодоны часто A/T-богаты в третьем положении, что чаще коррелирует с увеличением экспрессии, чем синонимичные кодоны, оканчивающиеся на G/C	[81]
Первые 10 кодонов, следующих за старт-кодоном, в кодирующей области	Более 30000	Рандомизированные	Уменьшение стабильности вторичной структуры, а не сама редкость кодонов отвечает за увеличение эффективности трансляции. Сильное влияние на экспрессию оказывают кодоны ближе к стартовому. Дополнительные старт-кодоны в рамках считывания способствуют трансляции. Наличие в N-концевой области белка аминокислот, на синтез которых клетка тратит много ресурсов, отрицательно сказывалось на эффективности синтеза таких белков	[28]

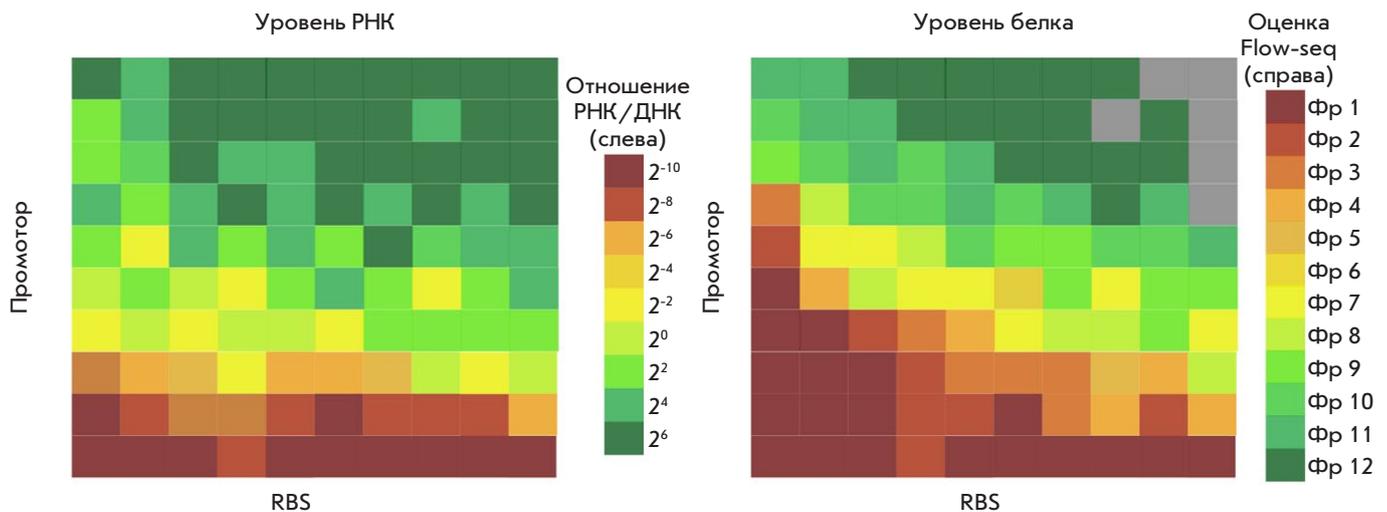


**Рис. 3.** Схема метода Flow-seq (на примере работы с рандомизированными 5'-UTR перед геном белка CER и контрольной 5'-UTR перед геном белка RFP). Представлены этапы конструирования библиотеки плазмид, трансформации, сортировки и секвенирования. А – клонирование рандомизированного ДНК-фрагмента в репортерный вектор перед геном белка CER. Перед геном белка RFP сохраняется константная 5'-UTR. Б – электропорация целой библиотеки плазмид в клетки *E. coli*. В – разделение клеток на основе отношения интенсивностей флуоресценции белков CER/RFP клеточным сортером. Г – сбор фракций с клетками (например, F1–F6) согласно соотношению CER/RFP. Д – выделение ДНК и амплификация рандомизированного участка с последующим высокопроизводительным секвенированием (NGS). Рисунок выполнен в программе Inkscape

уровней флуоресценции этих белков. После наращивания количества собранных клеток из них выделяют плазмиды, варибельный участок которых амплифицируют и подвергают высокопроизводительному секвенированию с целью определения последовательностей ДНК/РНК, попавших в конкретную фракцию и обеспечивающих тот или иной уровень экспрессии репортерного гена (рис. 3).

Этот подход применили для создания ряда конструкций с различными комбинациями сайтов посадки рибосомы и промоторов одновременно. Количество РНК и зеленого флуоресцентного белка (GFP), синтезируемых клетками, трансформированными каждой из конструкций, относили к количе-

ству соответствующей ДНК, таким образом определяя эффективность транскрипции и трансляции. Также в конструкцию включали ген флуоресцентного белка mCherry, используемого в качестве внутреннего контроля, с консервативным промотором и сайтом посадки рибосомы (RBS) [67]. В итоге был получен набор из 12653 плазмид с различными комбинациями 114 промоторов и 111 вариантов RBS. Чтобы оценить уровни ДНК и РНК в стационарной фазе проводили глубокое секвенирование ДНК (DNA-seq) и РНК (RNA-seq) из клеток в этой фазе. Для оценки уровней двух флуоресцентных белков клетки сортировали по соотношению интенсивностей флуоресценции GFP/mCherry. Из со-



**Рис. 4.** Схематическое изображение примерных репрезентативных карт уровней эффективности синтеза РНК и белка. Уровни РНК (слева) и белка (справа) для небольшого набора конструкций размещены на сетке согласно идентичности промоторов (ось Y) и сайтов посадки рибосом (RBS, ось X). Промоторы и RBS отсортированы по возрастанию средних значений эффективности синтеза РНК и белка соответственно. Серые клетки указывают на конструкции, соответствующие уровням, лежащим ниже эмпирически заданного порога. Шкалы уровней РНК (отношения РНК к ДНК) и белка (отношения флуоресценции белков GFP (зеленый) к RFP (красный)) показаны справа от соответствующих им карт. Рисунок выполнен, опираясь на источник [67], в программе Inkscaper

вокупностей клеток со сходными соотношениями интенсивности флуоресценции GFP/mCherry выделяли плазмидную ДНК и проводили ее глубокое секвенирование. При этом на экстрагированные последовательности, принадлежащие конкретной группе, «довешивали» специфические для каждой отдельной группы баркодные последовательности, которые служили в дальнейшем для поиска и сортировки последовательностей по ранее определенным группам при анализе ридов, полученных в результате секвенирования. Затем оценивали уровни двух флуоресцентных белков в группах, отношение GFP/mCherry определяли как меру эффективности трансляции, делили клетки по величине этого отношения на три вида: слабые, средние и сильные, и определяли соответствующие им последовательности. При этом, как и ожидалось, клетки в библиотеке содержали примерно одинаковые уровни белка mCherry, интенсивности флуоресценции которого имели нормальное/гауссовское распределение и изменялись в пределах одного порядка, в то время как уровни экспрессии гена *gfp* различались на 4 порядка. С помощью секвенирования были верифицированы 282 отдельные колонии, 55% из которых были правильными, т.е. содержали константные участки без ошибок, и для которых определяли ожидаемые варианты промоторов и сайтов посадки рибосомы без мутаций. Измеряли уровни флуоресценции большинства из этих 55% правильных комбинаций промоторов и RBS и использовали их в дальнейшем в качестве контрольного набора.

Результаты, полученные после крупномасштабного секвенирования ДНК, РНК, измерения уровней экспрессии генов флуоресцентных белков, использовали на следующем этапе в качестве основы для построения репрезентативных карт. При их построении для каждого типа конструкций с конкретными вариантами промотора и сайта связывания рибосомы определяли уровни транскрипции и трансляции (рис. 4). Дальнейший анализ позволил оценить самые эффективные и неэффективные сочетания, содержащиеся в библиотеке полученных конструкций (табл. 2) [67]. Был проведен подробный дисперсионный анализ (ANOVA) [68] уровней РНК и белка, независимо определяемых как промотором, так и сайтом посадки рибосомы. Такой подход помог учесть и эффекты, показывающие зависимость уровня РНК от скорости трансляции.

Для визуализации полученных оценок использовали программы на языках R [69] и Python [70], адаптированные к работе с большими массивами данных. Результаты анализа дисперсий позволили объяснить различия в уровнях РНК в 92.5% случаев выбором промотора, в 3.8% – выбором сайта посадки рибосомы, а оставшиеся 3.7% не удалось объяснить выбором одного из изменяемых элементов. Различия в уровнях белка GFP в 53.8% случаев объясняются выбором промотора, в 29.6% – выбором RBS, остальной процент не определялся ни одним из этих двух вариативных факторов. Таким образом, сделан вывод, что выбор промотора оказывает наибольшее влияние на уровень РНК и в меньшей

**Таблица 2.** Примеры последовательностей промоторов и сайтов посадки рибосомы (RBS), обеспечивающих неэффективную и эффективную экспрессию

№	Эффективность экспрессии	Промотор	RBS
1	Неэффективная экспрессия	<u>GGCGCGCCTCGACATTTATCCCTTGCGGCGA</u> ATACTTACAGCCATAGCAA	CACCATACACATATG
2		<u>GGCGCGCCTGATAGCTAGCTCAGTCCTAGG</u> GATTATGCTAGCAGATG	ATCTTAATCTAGCGCGGGACAGTTTCATATG
3		<u>GGCGCGCCTCGACAATTAATCATCCGGCTCG</u> ATACTTACAGCCATCGATT	TCTAGAGAAAGACCCGAGACACCATATG
4		<u>GGCGCGCCCACGGTGTAGACATTTATCCCTT</u> GCGGCGAATACTTACAGCCATGTGAA	ATCTTAATCTAGCTTTGGAGTCTTTCATATG
5		<u>GGCGCGCCTTGACAGCTAGCTCAGTCCTAGG</u> GATTGTGCTAGCCAATC	TCTAGAGAAAGATTAGAGTCACCATATG
6		<u>GGCGCGCCCACGGTGTAGACAATTAATCAT</u> CCGGCTCGATACTTACAGCCATGATTC	ATCTTAATCTAGCCCGGGAGCATTTCATATG
7		<u>GGCGCGCCTCGACATCAGGAAAATTTTTCTG</u> ATACTTACAGCCATGCGGA	TCTAGAGAAAGACAGGACCCACCATATG
8		<u>GGCGCGCCCACGGTGTAGACATCAGGAAAA</u> TTTTTCTGATACTTACAGCCATCGACC	TCTAGAGAAAGAGCCGACATACCATATG
9		<u>GGCGCGCCTTTATAGCTAGCTCAGCCCTTGGT</u> ACAATGCTAGCGCCTG	ATCTTAATCTAGCCTGGGATCGTTTCATATG
10		<u>GGCGCGCCTTTATGGCTAGCTCAGTCCTAGGT</u> ACAATGCTAGCCATAC	ATCTTAATCTAGCCAGGAACGTTTCATATG
1	Эффективная экспрессия	<u>GGCGCGCCTTGACATCGCATCTTTTTGTACCT</u> ATAATGTGTGGATAGAGT	AATCTCATATATCAAATATAGGGTGGATCA TATG
2		<u>GGCGCGCCAAAAAGAGTATTGACTTCAGGAA</u> AATTTTTCTGTATAATGTGTGGATGTTCA	AATCTCATATATCAAATATAAGGCGGATCA TATG
3		<u>GGCGCGCCAAAAAGAGTATTGACTATTAATC</u> ATCCGGCTCGTATAATAGATTCATTGAAG	ATTAAGAGGAGAAATTACATATG
4		<u>GGCGCGCCTTGACATCGCATCTTTTTGTACCT</u> ATAATAGATTCATGATGA	AAAGATCTTTAAGAAGGAGATATACATATG
5		<u>GGCGCGCCTTGACATAAAGTCTAACCTATAG</u> GATACTTACAGCCATACAAG	AAAGAGGAGAAATTACATATG
6		<u>GGCGCGCCTTGACATCAGGAAAATTTTTCTG</u> TAGATTTAACGTATAGGTA	AATCTCATAAATCAAATATAAGGGGGATC ATATG
7		<u>GGCGCGCCAAAAAGAGTATTGACTTCGCATC</u> TTTTTGTACCTATAATAGATTCATTGCTA	GAATTCATTAAGAGGAGAAAGGTCATATG
8		<u>GGCGCGCCAAAAAGAGTATTGACTTCGCATC</u> TTTTTGTACCCATAATTAATTCATTCACA	AATCTCATATCTCAAATATAAGGGGGATCA TATG
9		<u>GGCGCGCCAAAAATTTATTTGCTTTTTATCC</u> CTTGC GCGATATAATAGATTCATCTTAG	AATCTCATAGATCAAATATAGGGGGGATC ATATG
10		<u>GGCGCGCCAAAAATTTATTTGCTTTTCGCAT</u> CTTTTTGTACCTATAATGTGTGGATAATAA	ATCTTAATCTAGCGGGGAGAAATTCATATG

Примечание: примеры комбинаций последовательностей промотора и сайта посадки рибосомы выбраны с учетом наибольших и наименьших значений уровня РНК и трансляции, соответственно, для эффективной и неэффективной экспрессии белка; подчеркнуты последовательности сайтов рестрикции; последние пять нуклеотидов последовательностей промотора служат уникальным баркодом для определения места начала транскрипции. Последовательности приведены в ориентации 5'→3'.

степени на уровень белка, так как на эффективность трансляции влияет и выбор сайта посадки рибосомы, и, возможно, другие факторы [67].

В ряде работ с применением метода Flow-seq изучено влияние последовательностей 5'-нетранслируемых областей разного размера и их отдельных

участков на эффективность синтеза репортерного флуоресцентного белка [2, 39, 71–74].

Варьирование спейсерных участков, расположенных между последовательностью Шайна–Дальгарно и старт-кодоном, позволило создать библиотеки небольшого размера, где 4 и 6 нуклеотидов данного

участка были рандомизированы. При этом удалось получить 100- [39] и 1000-кратную [71] разницу между самым высоким и самым низким количеством белка на выходе соответственно. В первом случае в число самых эффективных и не эффективных вошли следующие последовательности спейсерного участка: сААААсау, сGAAAсау, сАUAAсау, сАUAUсау и сССGСсау, сCUCUсау, сCGCUсау, сССGUсау соответственно, окруженные с 5'-конца последовательностью SD (GGAG), а с 3'-конца старт-кодоном (AUG). Во втором случае среди последовательностей, следующих за SD (AAGAAGGA), предшествующих старт-кодону (AUG) и обеспечивающих наибольшую экспрессию, можно выделить gacUAGAGC, gacUGUAAG, gacAAAACC, gacGUGGUU. Интересно, что в обоих случаях среди самых эффективных встречается последовательность СААААС.

В первом случае для создания библиотеки использовали одноцепочечные олигонуклеотиды с четырьмя рандомизированными нуклеотидами в спейсерном участке и необходимыми сайтами рестрикции для последующего встраивания в вектор перед старт-кодоном гена флуоресцентного белка CER. Полученный таким образом набор клонов подвергали сортировке, а выделенные вариативные участки плазмид использовали для NGS-секвенирования [39].

Во втором случае с целью оптимизации синтеза двух конкретных белков, кодируемых генами *araH<sup>WT</sup>* и *narK<sup>WT</sup>*, их кодирующие последовательности были соединены с участком, кодирующим добавочную последовательность TEV-GFP-His<sub>8</sub>, где TEV – сайт распознавания протеазой вируса гравировки табака (VTM/TEV – tobacco etch virus), His<sub>8</sub> – восьми-гистидиновый таг для дальнейшей очистки. Таким образом, измеряемая флуоресценция GFP может отражать уровни экспрессии интересующих генов. Для создания библиотеки использовали вектор, включающий описанный сложный кодирующий участок под контролем T7-промотора, и два праймера, обратный из которых был константным, а прямой содержал по шесть вариативных нуклеотидов перед и после старт-кодона, причем последние удовлетворяли синонимичным заменам кодонов. Экспрессию индуцировали ИПТГ, после чего клетки сортировали с помощью FACS на отдельные фракции по интенсивности флуоресценции белка GFP. Далее из полученных фракций выделяли библиотеки плазмидных ДНК и подвергали их высокопроизводительному секвенированию [71].

Анализ результатов секвенирования нескольких десятков тысяч различных вариантов мРНК, полученных в двух описанных выше экспериментах,

показал, что низкое содержание GC-нуклеотидов и отсутствие (или минимизация) вторичной структуры мРНК в исследуемом спейсерном участке способствуют увеличению количества синтезируемого белка [39, 71]. Поэтому для увеличения выхода синтезируемого белка кажется разумным использовать олигоадениловые или другие А-богатые спейсеры между последовательностью SD и старт-кодоном, избегая при этом цитидиновых оснований, хотя не стоит исключать некоторые конкретные мРНК с А-богатыми спейсерными участками, которые могут маскировать сайт инициации трансляции в своей вторичной структуре, если, например, начало кодирующей области будет U-богатым.

Эти результаты необходимо учитывать при дизайне репортерных плазмид, когда уровень экспрессии экзогенных генов планируют регулировать в соответствии с конкретной биотехнологической потребностью. Для коэкспрессии генов, продукты которых должны синтезироваться в определенном стехиометрическом соотношении, например, когда белки являются субъединицами гетеромультимерного комплекса, уровни экспрессии этих генов можно отрегулировать путем правильного подбора спейсерных участков.

Определение чувствительности к незначительным изменениям в последовательности регуляторных элементов в 5'-UTR, таких, как последовательность Шайна–Дальгарно, представляет сложную задачу, поскольку небольшие изменения в составе 5'-UTR могут привести к непредсказуемым изменениям в уровне экспрессии гена [34, 75]. Зависимость эффективности трансляции от последовательности 5'-UTR делает возможной эффективную и мультиплексную инженерию при условии, что создаваемые модели способны адекватно предсказывать эти изменения [73].

В качестве еще одного инструмента для проектирования уровня экспрессии генов в биоинженерии разработана ЕМОРЕС, представляющая собой почти полную базу данных измеренных с помощью Flow-seq уровней экспрессии *gfp* в зависимости от присутствия той или иной последовательности SD [76].

Хорошо известно, что влияние конкретной SD в значительной степени зависит от ее генетического контекста [32]. Соответственно, следует соблюдать особую осторожность при повторном применении измеренных уровней экспрессии в биоинженерии метаболических путей или в синтетической биологии, поскольку сайт связывания рибосомы сильно зависит от локальной вторичной структуры мРНК. Однако если последовательности Шайна–Дальгарно можно модифицировать, минимально воздействуя

при этом на вторичную структуру в данной области мРНК, то относительный порядок силы экспрессии конкретной SD вероятно сохранится [73]. Данные особенности учитываются при использовании алгоритма в базе EMOPEC, что позволяет протестировать широкий спектр уровней экспрессии генов при минимальных изменениях в последовательности SD. Это делает возможным параллельное и эффективное редактирование генома, настраивающее уровень экспрессии генов.

Метод Flow-seq многократно использовался для того, чтобы понять, как нуклеотидные последовательности различных участков 5'-UTR влияют на эффективность трансляции. В частности, такому исследованию подвергали сайты посадки рибосомы с фиксированной последовательностью SD [74], 5'-UTR разной фиксированной длины [2] или природные 5'-UTR разной длины [77], а также standby-сайты и спейсерные участки [72]. Анализ десятков тысяч проверенных вариантов позволил получить разницу в пределах 4 и даже 5 порядков в эффективности синтеза репортерного белка. Причем замена одного репортерного белка другим зачастую не влияла на общую тенденцию распределения последовательностей, задающих тот или иной уровень биосинтеза белка, что указывает на то, что эти изменения определяются именно переменными участками мРНК. Схожие наблюдения, связанные с низкой стабильностью вторичной структуры и консервативностью последовательности SD, прослеживались в случае вариантов, определяющих высокую эффективность трансляции [2]. Те же факторы выявлены как значимые для эффективности трансляции репортерного гена, перед которым находился набор природных 5'-UTR, однако в данном случае вариативность эффективности трансляции была значительно меньше, чем в случае библиотеки полностью рандомизованных последовательностей 5'-UTR [77]. Встречались и отдельные случаи, свидетельствующие о присутствии AU-богатых энхансеров на 5'-конце в standby-сайте, низкой доле цитидиновых оснований, множественных SD, AG-повторов в 5'-UTR мРНК, обеспечивающих высокий уровень репортерного белка [2].

Сходный подход использовали и для выяснения влияния редких кодонов в начале кодирующей области мРНК на эффективность трансляции [78]. Согласно наблюдениям, редкие кодоны чаще встречаются в начале кодирующей области природных генов, особенно тех, которые экспрессируются на высоком уровне, что, возможно, важно для обеспечения высокого уровня биосинтеза белка [64, 79–82]. По другим данным, редкость кодонов в начале кодирующей области это просто следствие от-

бора, движущей силой которого служит стремление минимизировать вторичную структуру в начале кодирующей области мРНК [19, 78, 82]. В научной литературе продолжается дискуссия о причинах и следствиях возникновения кластера редких кодонов в начале кодирующих областей и о том, насколько эти кластеры влияют на эффективность трансляции. Возможными причинами расхождения во мнениях могут быть особенности получения данных, на которых эти мнения основаны. В частности, различные научные группы использовали для своих выводов природные [79–84] или синтетические последовательности [80, 85–90], а также несколько различающиеся способы анализа [79–90].

Для выяснения причин повышенной частоты встречаемости редких кодонов в начале кодирующей области генов бактерий и ее функциональной роли на основе массива олигонуклеотидов была создана большая библиотека, насчитывающая 14234 комбинации из двух промоторов (сильного и слабого), четырех сайтов посадки рибосомы (сильного, среднего, слабого и природного) и последовательностей первых 13 кодонов 137 генов *E. coli*. Подобные регуляторные элементы были помещены перед геном эффективно сворачивающегося зеленого флуоресцентного белка (sfGFP) в плазмиде, с которой конститутивно коэкспрессируется белок mCherry [78]. Во всей полученной библиотеке измерили уровни ДНК, РНК и белков с помощью DNA-seq, RNA-seq и Flow-seq соответственно.

Гипотеза о “codon ramp” гласит о медленной трансляции первых N-концевых кодонов кодирующей области, что в дальнейшем способствует уменьшению остановок рибосомы во время биосинтеза белка [79, 88, 89]. Некоторое увеличение эффективности трансляции при наличии редких кодонов в начале кодирующей области можно объяснить изменением вторичной структуры мРНК, а не редкостью кодонов как таковых [78]. Наконец, профили покрытия рибосом показали, что концентрация тРНК, которая, собственно, и предопределяет эффективность использования кодонов, не коррелирует со скоростью трансляции. Специфические редкие кодоны могут создавать внутренние мотивы, похожие на последовательность SD, которые, в свою очередь, могут влиять на эффективность трансляции в клетках *E. coli* [64]. Поиск ассоциации между внутренними SD-схожими мотивами и изменениями экспрессии привел к обнаружению слабой, но статистически значимой связи.

Изучение влияния синонимических мутаций на эффективность трансляции позволило сделать следующий вывод [85]. Присутствие редких кодонов в *E. coli*, часто A/T-богатых в третьем положении,

чаще коррелирует с увеличением экспрессии, чем наличие синонимичных кодонов, оканчивающиеся на G/C, что указывает на связь с вторичной структурой мРНК [85]. Также показано, что уменьшение GC-контента коррелирует с увеличением экспрессии белка [78]. С помощью предсказания вторичной структуры РНК для первых 120 оснований каждого транскрипта с использованием специализированной на укладке нуклеиновых кислот программы NUPACK [91] обнаружено, что увеличение силы вторичной структуры коррелирует с уменьшением уровня экспрессии, что объяснило большую вариацию, чем любое другое изменение, оцененное ранее [78].

Позже проанализировали более  $30 \times 10^3$  вариантов кодонов в положениях 2–11 кодирующей области репортерного флуоресцентного белка, полученных в ходе рандомизации первых 30 нуклеотидов, следующих за старт-кодоном. Ген второго флуоресцентного белка оставался неизменным и служил внутренним контролем. Созданная библиотека плазмид была исследована с помощью метода Flow-seq [28], что позволило подтвердить отрицательное влияние вторичной структуры мРНК на эффективность трансляции, в то время как положительная роль редких кодонов в начале кодирующей области в экспрессии гена не наблюдалась.

В то же время были выявлены следующие закономерности. Некоторые кодоны, находясь в начале кодирующей области, положительно (AUG, AGA, GUA, GCA, CAC, CGA, UAC, AAA, кодирующие дополнительный Met помимо инициаторного, положительно заряженные аминокислоты Arg, Lys, His, гидрофобные алифатические Ala, Val и ароматический Tyr), а некоторые отрицательно (CUC, CCC, CCG, CUG, GGA, GGG, GGC, GCC, кодирующие гидрофобные алифатические аминокислоты и аминокислоты с большей или меньшей конформационной свободой по сравнению с остальными аминокислотами Leu, Pro, Gly, Ala) влияют на уровень экспрессии. Чем ближе соответствующий кодон к инициаторному, тем сильнее его влияние. Дополнительные старт-кодоны в рамке считывания способствуют трансляции. Присутствие в N-концевой области белка аминокислот, на синтез которых клетка тратит много ресурсов, отрицательно сказывалось на эффективности синтеза таких белков в бедной среде.

Применение метода Flow-seq не ограничивается приведенными примерами. Этот метод применяют также для оценки (с помощью репортерных конструкций как биосенсоров в различных штаммах бактерий, в том числе и нокаутных [92]) эффектов, влияющих на гликолитические процессы, для оценки терминаторных последовательностей [93], определения генов, вовлекаемых в изменения

конкретного метаболического пути (с использованием биосенсорных конструкций [94]) и решения других задач, например, исследования сплайсинга [95].

### ВКЛАД FLOW-seq В СИНТЕТИЧЕСКУЮ БИОЛОГИЮ

Синтетическая биология – это недавно появившаяся область науки, связанная с дизайном и созданием живых организмов или отдельных процессов в природных организмах [96–98]. Возникла и развивается данная дисциплина благодаря сочетанию технологий генной инженерии и рекомбинантных ДНК с компьютерным моделированием. Таким образом, синтетическая биология стремится определить поведение организмов и процессов в них, чтобы потом модифицировать и объединять их в сложных конкретных задачах. Для надежной работы синтетических составных систем необходимо, чтобы составляющие систему белки образовывались в заранее заданных соотношениях [97].

Созданы три программы-калькулятора, которые оценивают эффективность трансляции на основе последовательностей 5'-UTR мРНК, так как считается, что общая скорость трансляции пропорциональна скорости инициации. Такие калькуляторы, как показано, дают адекватные оценки уровня биосинтеза белка.

Первым из трех калькуляторов является RBS Calculator [33, 99]. В основу его создания как метода предсказательного дизайна контролируемой инициации трансляции и белкового синтеза в бактериях [32, 33] легла ранее рассмотренная термодинамическая модель. Использование этого метода позволяет изменять эффективность трансляции в диапазоне 5 порядков величины [33, 34]. Однако предсказания, сделанные с помощью RBS Calculator, не всегда совпадают с экспериментальными данными, полученными как с помощью Flow-seq, так и путем тестирования индивидуальных репортерных конструкций [2].

Еще один вычислительный метод моделирования 5'-UTR, способный предсказывать эффективность трансляции по последовательности мРНК с той или иной 5'-UTR – UTR Designer или UTR Library Designer [100, 101]. В этом методе, сходном с программой RBS Calculator, используется термодинамический параметр, определяемый как разница в свободной энергии Гиббса до и после сборки инициаторного комплекса 30S на мРНК, и учитывается аффинность взаимодействия рибосомы и доступность мРНК и рибосомы. Как и RBS Calculator, эта программа имеет два варианта дизайна: в прямом происходит генерирование 5'-UTR с заданным уровнем эффективности трансляции желаемой последовательности белка. В обратном варианте пред-

сказывается уровень биосинтеза белка с введенной последовательности мРНК с 5'-UTR и первыми 35 нуклеотидами белок-кодирующей области. Принцип работы описанного метода создания библиотеки мРНК с различными 5'-UTR заключается в генерировании последовательностей 5'-UTR с помощью рандомизации нуклеотидов, комбинаторного перебора вариантов конструкций с выбором тех, которые будут обеспечивать желаемый уровень трансляции белка. Более того, можно отметить константную часть в 5'-UTR, присутствие которой обязательно в результирующей последовательности, тогда комбинаторный перебор будет затрагивать только ее окружение. Валидация данного метода проведена на двух библиотеках 5'-UTR, содержащих 16 последовательностей с разными уровнями трансляции в заданном диапазоне, с использованием флуоресцентного репортера; результаты предсказания *in silico* хорошо соответствовали измеренным данным *in vivo* [100]. Однако далеко не всегда предсказания описанного подхода коррелируют с результатами, полученными *in vivo* на других выборках последовательностей 5'-UTR в выбранном диапазоне эффективностей биосинтеза белка.

Последний RBS Designer, как и два предыдущих, рассчитывает свободные энергии, но отличается по методу расчета скорости трансляции. Имея в основе стационарную кинетическую модель, этот калькулятор оценивает вероятность связывания конкретной мРНК с рибосомой (эффективность трансляции), исходя из вероятности доступности представленного участка мРНК, содержащего RBS, и аффинности связывания рибосомы. Каждая из этих программ показала схожую точность предсказания [97].

На данный момент представлено несколько моделей предсказаний, построенных благодаря большому количеству данных, полученных методами крупномасштабного секвенирования, анализу различных библиотек и результатам других генно-инженерных методов. В качестве примера можно привести возможное предсказание сайтов инициации трансляции, полезное для локализации белок-кодирующих участков генов при компьютерной аннотации бактериальных и архейных геномов [102], а также предсказание возможных геномных последовательностей, которые соответствуют функциональным мотивам РНК [103], или же предсказание уровня экспрессии генов с новыми комбинациями генетических элементов [75].

Даже экспериментальная проверка эффективности трансляции, определяемой каким-либо участком связывания в модельной системе, не может гарантировать, что такая же эффективность будет

наблюдаться, если заменить последовательность кодирующей области. Так происходит, например, из-за образования вторичных структур при комплементарности кодирующей области и 5'-UTR. Для того, чтобы сделать уровень экспрессии любого гена, экспрессируемого в гетерологичной системе, более предсказуемым, было проведено исследование с помощью специально созданных бицистронных конструкций. В этом исследовании короткая стандартная рамка считывания предшествовала кодирующей области репортера, эффективность экспрессии которого определяли цитофлуориметрически. Рамки перекрывались с помощью рандомизированного участка реинициации трансляции. Таким образом установлено, что использование реинициации снимает зависимость эффективности трансляции от кодирующей области второго гена. В таком искусственном опероне в качестве второго гена использовали как *gfp*, так и *rfp*. Результирующие уровни экспрессии этих разных генов хорошо коррелировали между собой [75].

Таким образом, экспериментальное определение эффективности экспрессии с помощью цитофлуориметрии или Flow-seq можно прямо и надежно использовать при создании экспрессионных конструкций в синтетической биологии.

## ЗАКЛЮЧЕНИЕ

Flow-seq-метод комбинирует гибкие способы генной биоинженерии с клеточной сортировкой на основе проточной цитометрии и глубокого секвенирования ДНК, чтобы обеспечить всестороннюю оценку взаимосвязей между генотипом и фенотипом. Одно из применений Flow-seq заключается в изучении влияния специфических регуляторных элементов на синтез белка (табл. 1). Дизайн направленных изменений на основе репортерных конструкций с использованием генов флуоресцентных белков позволяет быстро и эффективно выяснить вклад конкретных вариантов регуляторных последовательностей в эффективность синтеза белка. Как и другие методы изучения влияния элементов 5'-нетранслируемой области мРНК на эффективность трансляции, данный подход имеет свои особенности, которые следует учитывать при планировании комплексного и многоступенчатого эксперимента. Несмотря на большие возможности рассматриваемого метода, его применение имеет свои ограничения, обусловленные прежде всего сложностями, возникающими на разных стадиях: при клонировании библиотеки ДНК, при сортировке клеток с разными соотношениями интенсивностей флуоресценции репортерных белков, при высокопроизводительном секвенировании, анализе полученных ридов и в по-

следующих расчетах. В число ограничений входит и использование всего двух флуоресцентных белков или других детектируемых реагентов такого рода, поскольку существует возможность перекрывания спектров флуоресценции этих белков и, как следствие, ошибок при регистрации сигналов. Тем не ме-

нее, метод Flow-seq широко используется в разных областях научных исследований и не теряет актуальность на протяжении многих лет. ●

*Работа выполнена при поддержке грантов РФФИ № 17-00-00369, 17-00-00366, 17-00-00367.*

## СПИСОК ЛИТЕРАТУРЫ

- Saier M.H. Jr. // *J. Bacteriol.* 2019. V. 201. № 15. P. e00091–e119.
- Evfratov S.A., Osterman I.A., Komarova E.S., Pogorelskaya A.M., Rubtsova M.P., Zatsepin T.S., Semashko T.A., Kostryukova E.S., Mironov A.A., Burnaev E., et al. // *Nucl. Acids Res.* 2017. V. 45. № 6. P. 3487–3502.
- Brenneis M., Soppa J. // *PLoS One.* 2009. V. 4. № 2. P. e4484.
- Shine J., Dalgarno L. // *Nature.* 1975. V. 254. P. 34–38.
- Shine J., Dalgarno L. // *Proc. Natl. Acad. Sci. USA.* 1974. V. 71. № 4. P. 1342–1346.
- Kozak M. // *Gene.* 2005. V. 361. P. 13–37.
- Shultzaberger R.K., Bucheimer R.E., Rudd K.E., Schneider T.D. // *J. Mol. Biol.* 2001. V. 313. № 1. P. 215–228.
- Rudd K.E. // *Nucl. Acids Res.* 2000. V. 28. № 1. P. 60–64.
- Ma J., Campbell A., Karlin S. // *J. Bacteriol.* 2002. V. 184. P. 5733–5745.
- Gardner P.P., Eldai H. // *Nucl. Acids Res.* 2015. V. 43. № 2. P. 691–698.
- Schluenzen F., Tocilj A., Zarivach R., Harms J., Gluehmann M., Janell D., Bashan A., Bartels H., Agmon I., Franceschi F., et al. // *Cell.* 2000. V. 102. P. 615–623.
- Kaminishi T., Wilson D.N., Takemoto C., Harms J.M., Kawazoe M., Schluenzen F., Hanawa-Suetsugu K., Shirouzu M., Fucini P., Yokoyama S. // *Structure.* 2007. V. 15. P. 289–297.
- Arenz S., Wilson D.N. // *Cold Spring Harb. Perspect. Med.* 2016. V. 6. № 9. P. a025361.
- Wegmann U., Horn N., Carding S.R. // *Appl. Environ. Microbiol.* 2013. V. 79. № 6. P. 1980–1989.
- Nakagawa S., Niimurab Y., Miurac K.-i., Gojobori T. // *Proc. Natl. Acad. Sci. USA.* 2010. V. 107. № 14. P. 6382–6387.
- Vimberg V., Tats A., Remm M., Tenson T. // *BMC Mol. Biol.* 2007. V. 8. P. 100.
- Osterman I.A., Evfratov S.A., Sergiev P.V., Dontsova O.A. // *Nucl. Acids Res.* 2013. V. 41. P. 474–486.
- Chen H., Bjerknes M., Kumar R., Jay E. // *Nucl. Acids Res.* 1994. V. 22. P. 4953–4957.
- Gu W., Zhou T., Wilke C.O. // *PLoS Comput. Biol.* 2010. V. 6. P. e1000664.
- Gingold H., Pilpel Y. // *Mol. Systems Biol.* 2011. V. 7. P. 481.
- de Smit M.H., van Duin J. // *Proc. Natl. Acad. Sci. USA.* 1990. V. 87. P. 7668–7672.
- Sean M.S., Simpson J. // *Mol. Cell.* 2006. V. 22. P. 105–115.
- Ban N., Beckmann R., Cate J.H., Dinman J.D., Dragon F., Ellis S.R., Lafontaine D.L., Lindahl L., Liljas A., Lipton J.M., et al. // *Curr. Opin. Struct. Biol.* 2014. V. 24. P. 165–169.
- Laursen B.S., Sorensen H.P., Mortensen K.K., Sperling-Petersen H.U. // *Microbiol. Mol. Biol. Rev.* 2005. V. 69. P. 101–123.
- Lauber M.A., Rappsilber J., Reilly J.P. // *Mol. Cell. Proteomics.* 2012. V. 11. P. 1965–1976.
- Stenström C.M., Isaksson L.A. // *Gene.* 2002. V. 288. P. 1–8.
- Gonzalez de Valdivia E.I., Isaksson L.A. // *Nucl. Acids Res.* 2004. V. 32. № 17. P. 5198–5205.
- Osterman I.A., Chervontseva Z.S., Evfratov S.A., Sorokina A.V., Rodin V.A., Rubtsova M.P., Komarova E.S., Zatsepin T.S., Kabilov M.R., Bogdanov A.A., et al. // *Nucl. Acids Res.* 2020. V. 48. P. 6931–6942.
- Park Y.S., Seo S.W., Hwang S., Chu H.S., Ahn J.-H., Kim T.-W., Kim D.-M., Jung G.Y. // *Biochem. Biophys. Res. Commun.* 2007. V. 356. № 1. P. 136–141.
- Barendt P.A., Shah N.A., Barendt G.A., Sarkar C.A. // *PLoS Genet.* 2012. V. 8. P. e1002598.
- Barendt P.A., Shah N.A., Barendt G.A., Kothari P.A., Sarkar C.A. // *ACS Chem. Biol.* 2013. V. 8. № 5. P. 958–966.
- Salis H.M., Mirsky E.A., Voigt C.A. // *Nat. Biotechnol.* 2009. V. 27. № 10. P. 946–950.
- Salis H.M. // *Meth. Enzymol.* 2011. V. 498. P. 19–42.
- Borujeni A.E., Channarasappa A.S., Salis H.M. // *Nucl. Acids Res.* 2014. V. 42. № 4. P. 2646–2659.
- Hofacker I.L. // *Nucl. Acids Res.* 2003. V. 31. № 13. P. 3429–3431.
- Farasat I., Kushwaha M., Collens J., Easterbrook M., Guido M., Salis H.M. // *Mol. Syst. Biol.* 2014. V. 10. P. 731.
- Nakeff A., Valeriote F., Gray J.W., Grabske R.J. // *Blood.* 1979. V. 53. № 4. P. 732–745.
- Solieri L., Dakal T.C., Giudici P. // *Ann. Microbiol.* 2012. V. 63. P. 21–37.
- Komarova E.S., Chervontseva Z.S., Osterman I.A., Evfratov S.A., Rubtsova M.P., Zatsepin T.S., Semashko T.A., Kostryukova E.S., Bogdanov A.A., Gelfand M.S., et al. // *Microb. Biotechnol.* 2020. V. 13. P. 1254–1261.
- Kim D., Hong J.S.-J., Qiu Y., Nagarajan H., Seo J.-H., Cho B.K., Tsai S.F., Palsson B.O. // *PLoS Genet.* 2012. V. 8. № 8. P. e1002867.
- Lesnik E.A., Fogel G.B., Weekes D., Henderson T.J., Levene H.B., Sampath R., Ecker D.J. // *BioSystems.* 2005. V. 80. P. 145–154.
- Gould P.S., Bird H., Easton A.J. // *BioTechniques.* 2005. V. 38. P. 397–400.
- Shirokikh N.E., Alkalaeva E.Z., Vassilenko K.S., Afonina Z.A., Alekhina O.M., Kisselev L.L., Spirin A.S. // *Nucl. Acids Res.* 2010. V. 38. № 3. P. e15.
- Wen J.-D., Kuo S.-T., Chou H.-H.D. // *RNA Biol.* 2021. V. 18. № 11. P. 1489–1500.
- Tzareva N.V., Makhno V.I., Boni I.V. // *FEBS Lett.* 1994. V. 337. P. 189–194.
- Zheng X., Hu G.Q., She Z.S., Zhu H. // *BMC Genomics.* 2011. V. 12. P. 361.
- Ingolia N.T., Ghaemmhami S., Newman J.R., Weissman J.S. // *Science.* 2009. V. 324. № 5924. P. 218–223.
- Andreev D.E., O'Connor P.B., Fahey C., Kenny E.M., Terenin I.M., Dmitriev S.E., Cormican P., Morris D.W., Shatsky I.N., Baranov P.V. // *Elife.* 2015. V. 4. P. e03971.
- Andreev D.E., O'Connor P.B., Zhdanov A.V., Dmitriev R.I., Shatsky I.N., Papkovsky D.B., Baranov P.V. // *Genome Biol.* 2015. V. 16. № 1. P. 90.
- Meydan S., Marks J., Klepacki D., Sharma V., Baranov P.V., Firth A.E., Margus T., Kefi A., Vázquez-Laslop N., Mankin A.S. // *Mol. Cell.* 2019. V. 74. № 3. P. 481–493.e6.
- Brar G.A., Weissman J.S. // *Nat. Rev. Mol. Cell. Biol.* 2015. V. 16. № 11. P. 651–664.

52. Reid D.W., Shenolikar S., Nicchitta C.V. // *Methods*. 2015. V. 91. P. 69–74.
53. Ingolia N.T., Husmann J.A., Weissman J.S. // *Cold Spring Harb. Perspect. Biol.* 2019. V. 11. № 5. P. a032698.
54. Weaver J., Mohammad F., Buskirk A.R., Storz G. // *mBio*. 2019. V. 10. № 2. P. e02819–18.
55. Meydan S., Klepacki D., Mankin A.S., Vázquez-Laslop N. // *Meth. Mol. Biol.* 2021. V. 2252. P. 27–55.
56. Vazquez-Laslop N., Sharma C.M., Mankin A., Buskirk A.R. // *J. Bacteriol.* 2022. V. 204. № 1. P. e0029421.
57. O'Connor P.B., Andreev D.E., Baranov P.V. // *Nat. Commun.* 2016. V. 7. P. 12915.
58. Andreev D.E., O'Connor P.B., Loughran G., Dmitriev S.E., Baranov P.V., Shatsky I.N. // *Nucl. Acids Res.* 2017. V. 45. № 2. P. 513–526.
59. Glaub A., Huptas C., Neuhaus K., Ardern Z. // *J. Biol. Chem.* 2020. V. 295. № 27. P. 8999–9011.
60. Gerashchenko M.V., Gladyshev V.N. // *Nucl. Acids Res.* 2017. V. 45. № 2. P. e6.
61. Marks J., Kannan K., Roncase E.J., Klepacki D., Kefi A., Orelle C., Vázquez-Laslop N., Mankin A.S. // *Proc. Natl. Acad. Sci. USA*. 2016. V. 113. № 43. P. 12150–12155.
62. Vázquez-Laslop N., Mankin A.S. // *Annu. Rev. Microbiol.* 2018. V. 72. P. 185–207.
63. Svetlov M.S., Koller T.O., Meydan S., Shankar V., Klepacki D., Polacek N., Guydosh N.R., Vázquez-Laslop N., Wilson D.N., Mankin A.S. // *Nat. Commun.* 2021. V. 12. № 1. P. 2803.
64. Li G.W., Oh E., Weissman J.S. // *Nature*. 2012. V. 484. № 7395. P. 538–541.
65. Mohammad F., Woolstenhulme C.J., Green R., Buskirk A.R. // *Cell Rep.* 2016. V. 14. № 4. P. 686–694.
66. Jin H., Zhao Q., Gonzalez de Valdivia E.I., Ardell D.H., Stenström M., Isaksson L.A. // *Mol. Microbiol.* 2006. V. 60. № 2. P. 480–492.
67. Kosuri S., Goodman D.B., Cambray G., Mutalik V.K., Gao Y., Arkin A.P., Endy D., Church G.M. // *Proc. Natl. Acad. Sci. USA*. 2013. V. 110. № 34. P. 14024–14029.
68. Mutalik V.K., Guimaraes J.C., Cambray G., Mai Q.A., Christoffersen M.J., Martin L., Yu A., Lam C., Rodriguez C., Bennett G., Keasling J.D., Endy D., Arkin A.P. // *Nat. Methods*. 2013. V. 10. P. 347–353.
69. Dessau R.B., Pippert C.B. // *Ugeskr. Laeger* 2008. V. 170. P. 328–330.
70. Sanner M.F. // *J. Mol. Graph. Model.* 1999. V. 17. P. 57–61.
71. Mirzadeh K., Martinez V., Toddo S., Guntur S., Herrgard M.J., Elofsson A., Norholm M.H., Daley D.O. // *ACS Synth. Biol.* 2015. V. 4. P. 959–965.
72. Sauer C., van Themaat E.V.L., Boender L.G.M., Groothuis D., Cruz R., Hamoen L.W., Harwood C.R., van Rij T. // *ACS Synth. Biol.* 2018. V. 7. № 7. P. 1773–1784.
73. Klausen M.S., Sommer M.O.A. // *Meth. Mol. Biol.* 2018. V. 1671. P. 3–14.
74. Duan Y., Zhang X., Zhai W., Zhang J., Zhang X., Xu G., Li H., Deng Z., Shi J., Xu Z. // *ACS Synth. Biol.* 2022. V. 11. № 8. P. 2726–2740.
75. Mutalik V.K., Guimaraes J.C., Cambray G., Lam C., Christoffersen M.J., Mai Q.-A., Tran A.B., Paull M., Keasling J.D., Arkin A.P., et al. // *Nat. Methods*. 2013. V. 10. P. 354–360.
76. Bonde M.T., Pedersen M., Klausen M.S., Jensen S.I., Wulff T., Harrison S., Nielsen A.T., Herrgård M.J., Sommer M.O. // *Nat. Meth.* 2016. V. 13. P. 233–236.
77. Komarova E.S., Slesarchuk A.N., Rubtsova M.P., Osterman I.A., Tupikin A.E., Pyshnyi D.V., Dontsova O.A., Kabilov M.R., Sergiev P.V. // *Int. J. Mol. Sci.* 2022. V. 23. № 20. P. 12293.
78. Goodman D.B., Church G.M., Kosuri S. // *Science*. 2013. V. 342. № 6157. P. 475–479.
79. Tuller T., Carmi A., Vestsigian K., Navon S., Dorfan Y., Zaborske J., Pan T., Dahan O., Furman I., Pilpel Y. // *Cell*. 2010. V. 141. № 2. P. 344–354.
80. Allert M., Cox J.C., Hellinga H.W. // *J. Mol. Biol.* 2010. V. 402. № 5. P. 905–918.
81. Pechmann S., Frydman J. // *Nat. Struct. Mol. Biol.* 2013. V. 20. № 2. P. 237–243.
82. Bentele K., Saffert P., Rauscher R., Ignatova Z., Blüthgen N. // *Mol. Syst. Biol.* 2013. V. 9. P. 675.
83. dos Reis M., Savva R., Wernisch L. // *Nucl. Acids Res.* 2004. V. 32. № 17. P. 5036–5044.
84. Shah P., Ding Y., Niemczyk M., Kudla G., Plotkin J.B. // *Cell*. 2013. V. 153. № 7. P. 1589–1601.
85. Kudla G., Murray A.W., Tollervey D., Plotkin J.B. // *Science*. 2009. V. 324. № 5924. P. 255–258.
86. Welch M., Govindarajan S., Ness J.E., Villalobos A., Gurney A., Minshull J., Gustafsson C. // *PLoS One*. 2009. V. 4. № 9. P. e7002.
87. Zhou M., Guo J., Cha J., Chae M., Chen S., Barral J.M., Sachs M.S., Liu Y. // *Nature*. 2013. V. 495. № 7439. P. 111–115.
88. Navon S., Pilpel Y. // *Genome Biol.* 2011. V. 12. № 2. P. R12.
89. Tuller T., Waldman Y.Y., Kupiec M., Ruppin E. // *Proc. Natl. Acad. Sci. USA*. 2010. V. 107. № 8. P. 3645–3650.
90. Subramaniam A.R., Pan T., Cluzel P. // *Proc. Natl. Acad. Sci. USA*. 2013. V. 110. № 6. P. 2419–2424.
91. Zadeh J.N., Steenberg C.D., Bois J.S., Wolfe B.R., Pierce M.B., Khan A.R., Dirks R.M., Pierce N.A. // *J. Comput. Chem.* 2010. V. 32. № 1. P. 170–173.
92. Lehning C.E., Siedler S., Ellabaan M.M.H., Sommer M.O.A. // *Metab. Eng.* 2017. V. 42. P. 194–202.
93. Zhai W., Duan Y., Zhang X., Xu G., Li H., Shi J., Xu Z., Zhang X. // *Synth. Syst. Biotechnol.* 2022. V. 7. № 4. P. 1046–1055.
94. Glanville D.G., Mullineaux-Sanders C., Corcoran C.J., Burger B.T., Imam S., Donohue T.J., Ulijasz A.T. // *mSystems*. 2021. V. 6. № 1. P. e00933–20.
95. Cheung R., Insigne K.D., Yao D., Burghard C.P., Wang J., Hsiao Y.E., Jones E.M., Goodman D.B., Xiao X., Kosuri S. // *Mol. Cell*. 2019. V. 73. № 1. P. 183–194.e8.
96. Andrianantoandro E., Basu S., Karig D.K., Weiss R. // *Mol. Systems Biol.* 2006. V. 2. P. 2006.0028.
97. Reeve B., Hargest T., Gilbert C., Ellis T. // *Front. Bioeng. Biotechnol.* 2014. V. 2. P. 1–6.
98. Chappell J., Jensen K., Freemont P.S. // *Nucl. Acids Res.* 2013. V. 41. № 5. P. 3471–3481.
99. Zhang L., Lin X., Wang T., Guo W., Lu Y. // *Bioresour. Bioprocess.* 2021. V. 8. № 1. P. 58.
100. Seo S.W., Yang J.S., Kim I., Yang J., Min B.E., Kim S., Jung G.Y. // *Metab. Eng.* 2013. V. 15. P. 67–74.
101. Seo S.W., Yang J.S., Cho H.S., Yang J., Kim S.C., Park J.M., Kim S., Jung G.Y. // *Sci. Rep.* 2015. V. 4. № 1. P. 4515.
102. Zhu H., Wang Q. // *Curr. Bioinform.* 2014. V. 9. P. 155–165.
103. Laserson U., Gan H.H., Schlick T. // *Nucl. Acids Res.* 2005. V. 33. № 18. P. 6057–6069.