

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Комарова Екатерина Сергеевна

**Изучение особенностей 5'-нетранслируемой области
бактериальных мРНК, влияющих на эффективность трансляции,
с помощью библиотек репортёрных конструкций**

1.5.3. Молекулярная биология

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата химических наук

Москва – 2023

Работа выполнена в институте функциональной геномики МГУ имени М.В. Ломоносова.

- Научный руководитель** – *Сергиев Пётр Владимирович, доктор химических наук, член-корреспондент РАН*
- Официальные оппоненты** – *Малыгин Алексей Аркадьевич, доктор химических наук, доцент, заведующий лабораторией структуры и функции рибосом, ФГБУН Институт химической биологии и фундаментальной медицины СО РАН*
- Лябин Дмитрий Николаевич, доктор биологических наук, руководитель группы регуляции биосинтеза белка, ФГБУН Институт белка РАН*
- Первушин Дмитрий Давидович, кандидат физико-математических наук, доцент центра молекулярной и клеточной биологии, АНОО ВО «Сколковский институт науки и технологий»*

Защита диссертации состоится «28» ноября 2023 г. в 15 ч. 00 мин. на заседании диссертационного совета МГУ.014.2 Московского государственного университета имени М.В. Ломоносова по адресу: 119234, г. Москва, Ленинские горы, д. 1, стр. 40, НИИ ФХБ, ауд. 501.

Е-mail: agapkina@belozersky.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на портале:

<http://dissovet.msu.ru/dissertation/014.2/2732/>

Автореферат разослан «__» октября 2023 г.

Ученый секретарь
диссертационного совета,
кандидат химических наук

Агапкина Ю.Ю.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

В ходе трансляции рибосома осуществляет декодирование информации, записанной в матричной РНК (мРНК), и перевод её в последовательность аминокислот, формирующую белок. мРНК участвует в этом процессе не только как пассивный посредник, но и предопределяет эффективность трансляции благодаря наличию функциональных элементов. Один из них представлен 5'-нетранслируемой областью (5'-UTR) перед белок-кодирующей областью.

Анализ разнообразия бактериальных 5'-UTR выявил в них ряд особенностей, влияющих на уровень трансляции. Наиболее изученной из них является последовательность Шайна-Дальгарно (SD), расположенная в сайте связывания рибосомы (RBS) в 5'-UTR и комплементарная 3'-концу 16S рибосомной РНК (рРНК) малой субчастицы рибосомы. Способствует трансляции и AU-богатая область 5'-UTR, взаимодействующая с белком bS1 малой 30S субчастицы. Участок инициации трансляции (TIR) должен быть одноцепочечным или иметь легко разрушимую укладку вторичной структуры. На уровень биосинтеза белка также влияет начальный участок кодирующей области.

Не для всех мРНК эффективность трансляции можно объяснить, исходя из перечисленных выше особенностей. Важно уметь настраивать определённые уровни биосинтеза белков для биотехнологии. Хотя известны способы предсказаний *in silico* скорости инициации трансляции отдельной мРНК, их результаты не всегда подтверждаются экспериментом. Остаётся актуальным вопрос о том, что обеспечивает быструю инициацию и эффективную трансляцию мРНК.

Данная работа посвящена изучению особенностей синтетических и природных 5'-UTR бактериальных мРНК, влияющих на эффективность трансляции, с помощью библиотек репортёрных конструкций. Основным в этой работе является метод Flow-seq, сочетающий в себе флуоресцентную проточную цитометрию (FACS) и секвенирование нового поколения (NGS).

Цели и задачи исследования

Целью данной работы является изучение особенностей 5'-UTR бактериальных мРНК, влияющих на эффективность трансляции, с помощью библиотек репортёрных конструкций.

Для достижения поставленной цели определены следующие задачи:

1. Создание библиотек репортёрных конструкций с различными синтетическими и природными 5'-UTR перед геном флуоресцентного белка.
2. Адаптирование метода Flow-seq для более подробного изучения влияния особенностей 5'-UTR бактериальных мРНК на эффективность трансляции.
3. Анализ полученных данных и сравнение их с результатами других работ с применением схожих или отличных методов.
4. Поиск новых эффективных сайтов посадки рибосомы в 5'-UTR бактериальных мРНК и

определение основных элементов таких сайтов, обуславливающих высокий уровень биосинтеза белка.

Объект и предмет исследования

В данном исследовании объект для изучения представлен 5'-нетранслируемой областью бактериальных мРНК, одной из регуляторных элементов в обеспечении разной эффективности трансляции. Предметом настоящего исследования является длина и нуклеотидный состав 5'-UTR бактериальных мРНК наряду со стабильностью и особенностями их вторичной структуры, способностью взаимодействовать с 3'-концом 16S рРНК 30S субчастицы рибосомы.

Научная новизна исследования

В данной работе впервые проводится подробный анализ *in vivo* более 24,5 тысяч разных вариантов 5'-UTR бактериальных мРНК, как синтетических, так и природных, в репортёрных конструкциях перед генами флуоресцентных белков CER (голубого) или RFP (красного). Такое разнообразие мРНК позволило обеспечить изменение эффективности трансляции в пределах 4-х порядков между самым высоким и самым низким уровнем биосинтеза белка. В представленной работе продемонстрированы успешная валидация и применение метода Flow-seq для разделения клеток с библиотеками репортёрных конструкций с помощью FACS и анализа их участков с помощью NGS. Полученные результаты показали возможность использования Flow-seq для широкомасштабной оценки эффективности трансляции.

Для полученных библиотек 5'-UTR разной длины провели анализ их нуклеотидного состава, стабильности их вторичной структуры, их способности взаимодействовать с 3'-концом 16S рРНК 30S субчастицы рибосомы. В итоге выделили ряд особенностей, свойственных набору вариантов 5'-UTR, наиболее эффективных в трансляции. Среди них все известные детерминанты: SD с оптимальными длиной и расстоянием от старт-кодона, AU-богатые участки перед SD, низкая стабильность вторичной структуры 5'-UTR мРНК. Дополнительно найдены ранее не описанные особенности, способствующие увеличению эффективности трансляции: низкая доля остатков цитидина, AG-повторы и множественные SD. AG-повторы можно идентифицировать как новый энхансер трансляции. Мутационный анализ некоторых вариантов 5'-UTR, обеспечивающих высокий уровень биосинтеза белка, позволил отметить аддитивный эффект нескольких SD-мотивов и старт-кодонов. Опираясь на результаты представлены рекомендации по выбору 5'-UTR для систем экспрессии в бактериях.

Для библиотеки природных 5'-UTR впервые провели сравнение результатов, полученных с помощью методов Flow-seq и Ribo-seq (рибосомного профилирования) для широкомасштабной оценки эффективности трансляции, и выявили хорошее соответствие между ними.

Теоретическая и практическая значимость исследования

Полученные результаты интересны с фундаментальной точки зрения, так как раскрывают

новые детали трансляции, позволяя лучше понять, как особенности 5'-UTR бактериальных мРНК влияют на её эффективность. С практической точки зрения результаты работы проливают новый свет на возможности регуляции уровня биосинтеза белка при решении задач биотехнологии. Их можно использовать для создания систем регулируемой экспрессии генов в клетках *E. coli*, в том числе, если необходима точная настройка уровней экспрессии белков, входящих, например, в состав гетеромультимерного комплекса в заданном стехиометрическом соотношении.

Методология диссертационного исследования

Для достижения цели и решения поставленных задач в данной работе применили как классические, так и современные методы, активно применяемые в мире для проведения такого рода исследований в области молекулярной биологии, биоинформатики. Классические методы представлены манипуляциями с ДНК, клеточными культурами *E. coli*, основами статистики. Среди современных методов можно выделить Flow-seq, включающий FACS и NGS. Благодаря ему появилась возможность с высокой точностью разделять на фракции клетки с библиотеками конструкций и определять их варианты во фракциях. Анализ с применением актуальных биоинформатических подходов позволяет выделить закономерности между определёнными вариантами и эффективностью их трансляции, оценить значимость их влияния.

Основные положения, выносимые на защиту

1. Метод Flow-seq успешно адаптирован для высокопроизводительного определения эффективности трансляции синтетических и природных 5'-нетранслируемых областей, а следовательно и участков посадки рибосомы, многократно превышающих их природное разнообразие у *E. coli*.
2. Высокая эффективность трансляции некоторых мРНК не может быть объяснена только наиболее изученными особенностями: взаимодействием рибосомы с последовательностью Шайна-Дальгарно либо наличием AU-богатого энхансера на 5'-конце мРНК, или отсутствием ингибирующего влияния вторичной структуры мРНК.
3. Эффективность трансляции уменьшается с увеличением доли остатков цитозина в последовательности 5'-нетранслируемой области.
4. Дополнительный старт-кодон и множественные последовательности Шайна-Дальгарно имеют аддитивное влияние на эффективность трансляции.
5. AG-повторы в 5'-нетранслируемой области представляют собой энхансер трансляции.
6. мРНК с природными 5'-нетранслируемыми областями менее вариабельны по эффективности трансляции, чем рандомизированные последовательности. В более длинных, 30-звенных 5'-нетранслируемых областях чаще встречаются благоприятные для трансляции варианты из случайных последовательностей, чем в 20-звенных.

Степень достоверности результатов

Достоверность результатов данного исследования подтверждена их воспроизводимостью, статистической обработкой полученных данных. Поставленные эксперименты с использованием современного оборудования и реактивов от ведущих мировых компаний соответствуют цели и задачам работы.

Апробация работы

Результаты работы представлены на заседании Учёного совета института функциональной геномики МГУ имени М.В. Ломоносова и на заседании кафедры химии природных соединений химического факультета МГУ имени М.В. Ломоносова и на 4 международных конференциях.

Публикации по теме диссертации

По теме диссертационной работы опубликовано 4 научных труда, все из них статьи в рецензируемых научных журналах с индексированием в базах данных Web of Science и Scopus.

Личный вклад автора

Поиск и анализ литературных данных, постановка задач, планирование и проведение экспериментов, анализ и представление полученных результатов, апробация их на научных конференциях, подготовка и написание научных статей осуществлялись автором лично или при его непосредственном участии.

В обзорной статье [1] анализ литературных данных, особенностей и применения метода Flow-seq в изучении бактериальной трансляции выполнен полностью автором лично. В статье [2] с природными вариантами – при определяющем вкладе автора в рамках дипломной работы выполняла Слесарчук А. Н. В статьях [3-4] с рандомизированными и природными библиотеками 5'-UTR автором лично проведены все экспериментальные процедуры. Исключение составляют анализ и сортировка клеток на проточном цитометре [2-4], выполненные совместно с профессором Рубцовой М.П., и секвенирования нового поколения, произведённые совместно с руководителем ЦКП «Геномика» СО РАН Кабиловым М. Р. [2] и научной группой академика РАН Говоруна В.М. [3-4]. Биоинформатический анализ полученных данных проводился коллегами: руководителем ЦКП Кабиловым М. Р. [2], Червонцевой З. С. [3], Евфратовым С. А. и научной группой профессора Гельфанда М.С. [4], при активном участии автора.

Структура и объём диссертации

Данная кандидатская диссертация изложена на 177 страницах и состоит из следующих основных разделов: «Список сокращений», «Введение», «Обзор литературы», «Результаты и обсуждение», «Заключение», «Выводы», «Материалы и методы», «Список литературы». Также в работу включены 3 приложения, благодарности. В работе представлены 72 рисунка, 8 таблиц, отдельно раздел «Материалы и методы» содержит 1 рисунок и 1 таблицу, а в приложении данные приведены в 5 таблицах. Библиография включает список из 261 источников литературы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Для высокопроизводительного анализа эффективности трансляции мРНК с различными 5'-UTR мы использовали библиотеки репортёрных конструкций на основе генов флуоресцентных белков RFP и CER.

1. Проверка возможности использования Flow-seq для анализа эффективности трансляции

Клеточный сортер позволяет с высокой точностью разделять в ходе FACS клетки с разными вариантами репортёрных конструкций в зависимости от соотношения интенсивностей флуоресценции белков CER/RFP. Для сравнения его значений, измеренных флуориметром и клеточным сортером, и оценки эффективности сортировки провели контрольный эксперимент. Для этого смешали равное количество клеток, трансформированных 6 ранее использовавшимися репортёрными плазмидами с известным соотношением CER/RFP (Табл. 1). В ходе FACS клетки с контрольным набором плазмид разделили на 8 фракций по соотношению CER/RFP (Рис. 1А-Б). Для всех фракций путём секвенирования определили доли клеток с каждой плазмидой (Рис. 1Б) и спектрофлуориметром среднее соотношение CER/RFP (Рис. 1В).

Таблица 1. Репортёрные конструкции с известным соотношением CER/RFP.

Конструкция	Длина SD (нт)	Расстояние от SD до старт-кодона (нт)	Эффективность трансляции*	Номер фракции FACS**
8/7	8	7	17	F8
6/10	6	10	13	F8
4/7AU	4	7	3,7	F6-F7
4/7	4	7	0,8	F4-F5
2/7	2	7	0,03	F2
0	0	0	0,001	F1

* Эффективность трансляции оценена как соотношение флуоресценций белков CER/RFP
 ** Номер фракции FACS с наибольшей пропорцией клеток с этой конструкцией

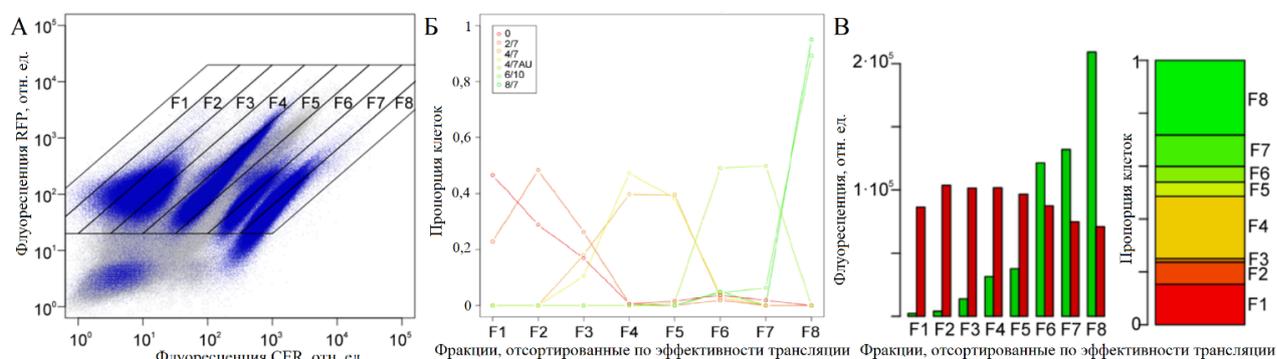


Рисунок 1. Оценка эффективности клеточной сортировки для контрольного набора конструкций. А. FACS клеток с контрольным набором конструкций. Фракции F1-F8 отсортированы по увеличению соотношения CER/RFP и ограничены рамками. Б. Распределение по фракциям клеток с отдельными плазмидами. Соответствие цветов и конструкций в легенде. В. Слева измеренные флуориметром во фракциях значения общей флуоресценции CER (зелёные столбцы) и RFP (красные). Справа – пропорции клеток с конструкциями во фракциях.

Результаты контрольного эксперимента показали возможность применения Flow-seq для точной и надёжной оценки эффективности трансляции в виде соотношения CER/RFP на уровне отдельной клетки, а измеренные им значения совпадают с данными флуориметра для клеточной культуры.

2. Изучение влияния спейсерного участка в 5'-UTR на эффективность трансляции

Для оценки влияния спейсерной области 5'-UTR мРНК на эффективность трансляции создали библиотеку репортёрных конструкций с участком рандомизации длиной 4 нт (4N) (Рис. 2). В качестве вектора взяли рRFP-CER с двойной репортёрной конструкцией, содержащей гены флуоресцентных белков RFP (красного) и CER (голубого). В данной конструкции 5'-UTR перед геном белка RFP не меняется, а синтез RFP служит внутренним контролем. Перед геном CER находится 5'-UTR с участком 4N (Рис. 2). По соотношению флуоресценции белков CER/RFP можно судить об эффективности трансляции мРНК CER с рандомизованным участком в 5'-UTR.

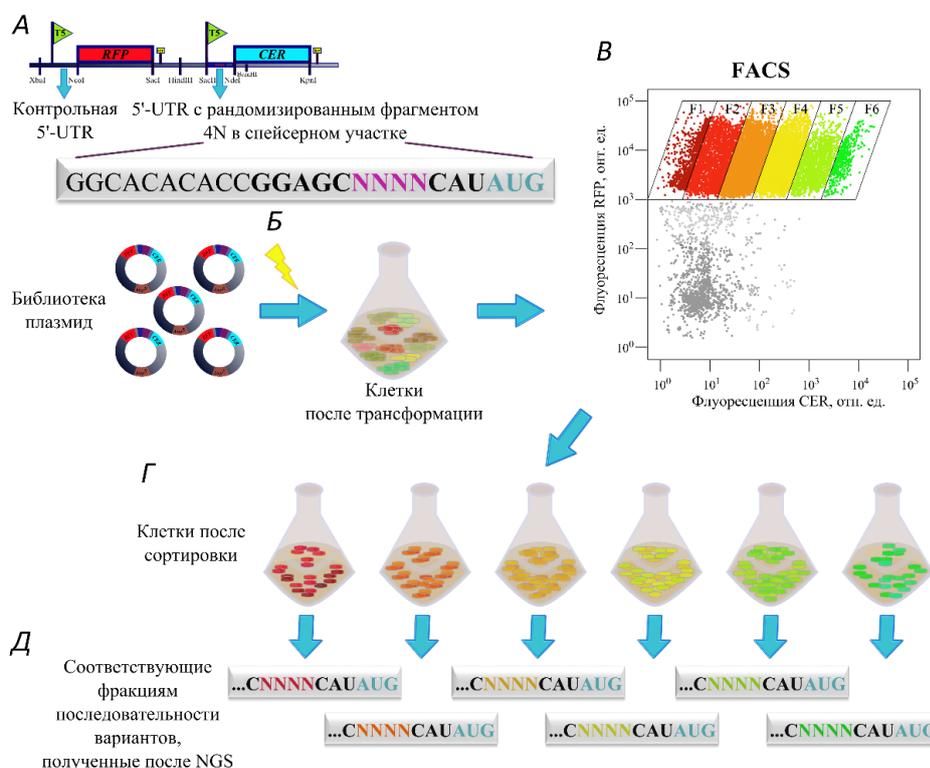


Рисунок 2. Схема метода Flow-seq. А. Схема репортёрных конструкций. Отмечены промоторы, терминаторы, 5'-UTR, кодирующие области флуоресцентных белков CER и RFP. Ниже схемы – последовательность 5'-UTR перед геном белка CER: SD выделена красным, 4N-фрагмент – фиолетовым, старт-кодон – голубым. Б. Электропорация клеток *E. coli* библиотекой плазмид. В. Сортировка клеток по соотношению CER/RFP. Приведена реальная карта, выдаваемая FACS, для клеток с библиотекой плазмид с 4N-фрагментом в спейсерном участке 5'-UTR. Г. Сбор фракций клеток F1-F6 с конкретным соотношением CER/RFP. Д. Секвенирование фрагментов с 4N-участками, полученных после выделения плазмид из клеток фракций и ПЦР-амплификации.

При сортировке с помощью FACS клетки, трансформированные библиотекой плазмид, разделили на 6 фракций по соотношению CER/RFP (Рис. 2В). После выделения плазмидной ДНК

из клеток фракций (Рис. 2Г) проводили ПЦР-амплификацию фрагмента с 4N-участком и секвенировали. NGS позволило установить соответствие между клетками с плазмидами, несущими определённые варианты вставок, и их распределением по фракциям, разделённым по соотношению CER/RFP, в итоге отдельной конструкции присвоили конкретное значение эффективности трансляции (Рис. 2В-Д). Эффективность трансляции полученных конструкций отличалась в пределах 100 крат. С учётом относительно узкого наблюдаемого распределения её мРНК разделили на 2 группы с высокой и низкой эффективностью трансляции.

Во всех позициях области рандомизации аденозин оказался наиболее предпочтительным для трансляции, а цитидин неблагоприятным (Рис. 3А). Это может быть объяснено минимизацией вторичной структуры или возможным стекинг-взаимодействием рибосомы с аденинами 5'-UTR. Расчёт минимальной свободной энергии MFE (Рис. 3Б) показал достоверность разницы в стабильности между эффективно и плохо транслируемыми мРНК ($p\text{-value} = 10^{-15}$).

Все варианты 5'-UTR в нашем наборе данных содержали SD из 4 нт на расстоянии 8 нт перед старт-кодоном. Для оценки влияния на эффективность трансляции добавочных участков, комплементарных 3'-концу 16S рРНК, рассчитали свободную энергию гибридизации фрагмента анти-SD (CACCUCCU) на 3'-конце 16S рРНК с участком 5'-UTR, содержащим 4N-фрагмент (Рис. 3В). Наблюдаемые распределения энергии существенно не отличались ($p\text{-value} > 0,05$), но в отличие от плохо транслируемых мРНК набор эффективно транслируемых вариантов содержал несколько 5'-UTR с дополнительным участком в спейсере, комплементарным 3'-концу 16S рРНК.

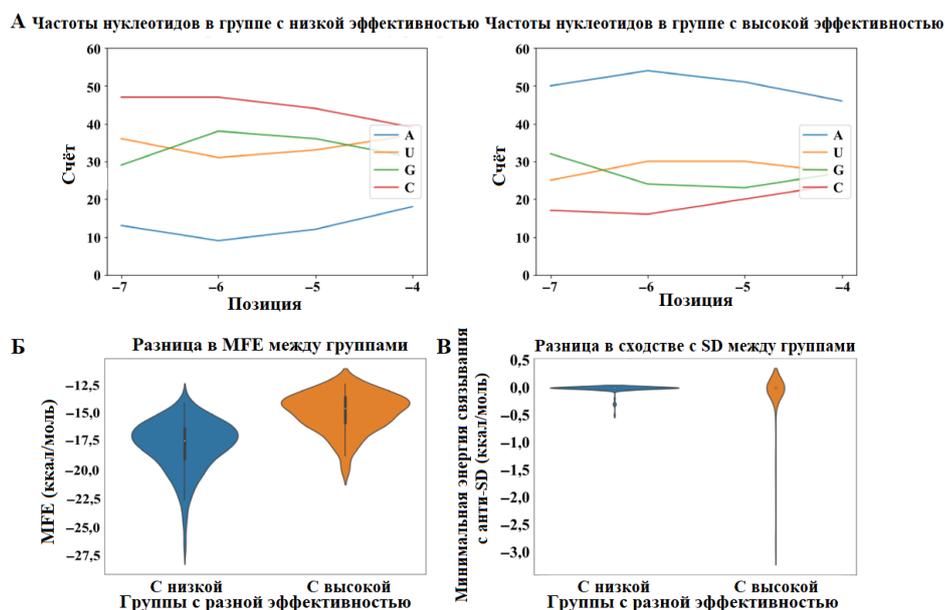


Рисунок 3. Особенности спейсерной области 5'-UTR мРНК с разной эффективностью трансляции. А. Частота нуклеотидов в 4N-участке 5'-UTR мРНК, обеспечивающих низкую (слева) и высокую (справа) эффективность трансляции. Позиции нуклеотидов показаны относительно старт-кодона. Б. Распределения в группах минимальной свободной энергии сворачивания вторичной структуры (MFE). В. Распределения в группах минимальной энергии гибридизации с 3'-концом 16S рРНК спейсера длиной 8 нт 5'-UTR с 4N-участком и фланкирующими 1 нт перед ним и 4 нт после него.

3. Изучение влияния рандомизированных 5'-UTR на эффективность трансляции

Для определения последовательности 5'-UTR мРНК, эффективной для посадки рибосомы и приводящей к наибольшей экспрессии гена, создали набор плазмид с участками рандомизации длиной 20 и 30 нт перед геном *cer*. В результате получили свыше 10^4 конструкций для каждой библиотеки 20N и 30N, сильно превышая по числу разнообразие 5'-UTR природных генов *E. coli*.

Метод Flow-seq применили для разделения клеток *E. coli*, трансформированных набором плазмид с участками рандомизации 20N и 30N в 5'-UTR перед геном *cer*, и анализа соответствия определённого соотношения CER/RFP и задающих его переменных 5'-UTR. Большинство 5'-UTR с 20N и 30N-участками плохо функционировали как сайты инициации трансляции. Только 0,03% клеток с библиотекой 20N попали во фракцию F8 с наибольшим соотношением CER/RFP, в случае библиотеки 30N в F8 вошли 0,45% клеток, их число значительно выше, чем у 20N. С помощью NGS выявили 11692 вариантов 5'-UTR для 20N и 11889 вариантов для 30N-участков 5'-UTR, каждому из которых сопоставили значение соотношения CER/RFP. В итоге получили уникальную библиотеку 5'-UTR, почти в 5 раз более разнообразную, чем коллекция природных 5'-UTR *E. coli*.

Анализ нуклеотидного состава рандомизированных 5'-UTR выявил почти равномерное распределение 4-х типов нуклеотидов в 5'-UTR плохо транслируемых мРНК для обеих библиотек (Рис. 4, панели F1). А более высокая эффективность трансляции привела к большему смещению этого состава в 5'-UTR (Рис. 4, панели F8). Заметное обогащение основаниями А и Г в позициях от -13 до -7 относительно старт-кодона в высоко эффективных 5'-UTR (Рис. 4, панели F8) скорее всего отражает наличие SD-мотивов. Значительное преобладание оснований А и, в меньшей степени, U характерно для областей как до, так и после предполагаемых мотивов SD в 5'-UTR мРНК из фракции с наибольшей экспрессией (Рис. 4, панели F8), что вероятно отражает наличие AU-богатых энхансеров. Участки 5'-UTR мРНК, обеспечивающих высокую эффективность, имеют заметно сниженную долю цитидинов (Рис. 4, панели F8), как при анализе 4N-участка рандомизации спейсера 5'-UTR мРНК.

Сравнение нуклеотидного состава 5'-UTR мРНК, приводящих к наибольшей экспрессии, между библиотеками показало, что эффективная трансляция коротких 5'-UTR из 20N, видимо, достигается только за счёт SD (Рис. 4, панели F8). В более длинных 5'-UTR из 30N ограничения, налагаемые на нуклеотидный состав области SD таких мРНК, не столь выражены. В эффективно транслируемых мРНК с длинными 5'-UTR видно смещение 5'-концевой области по составу в сторону оснований AU, указывая на её важность в обеспечении эффективности трансляции. Большинству коротких 5'-UTR необходимо иметь большую SD для обеспечения эффективной трансляции, тогда как более длинные 5'-UTR могут достигать той же эффективности с SD меньшей длины и разными, например, AU-богатыми энхансерами в 5'-области перед SD.

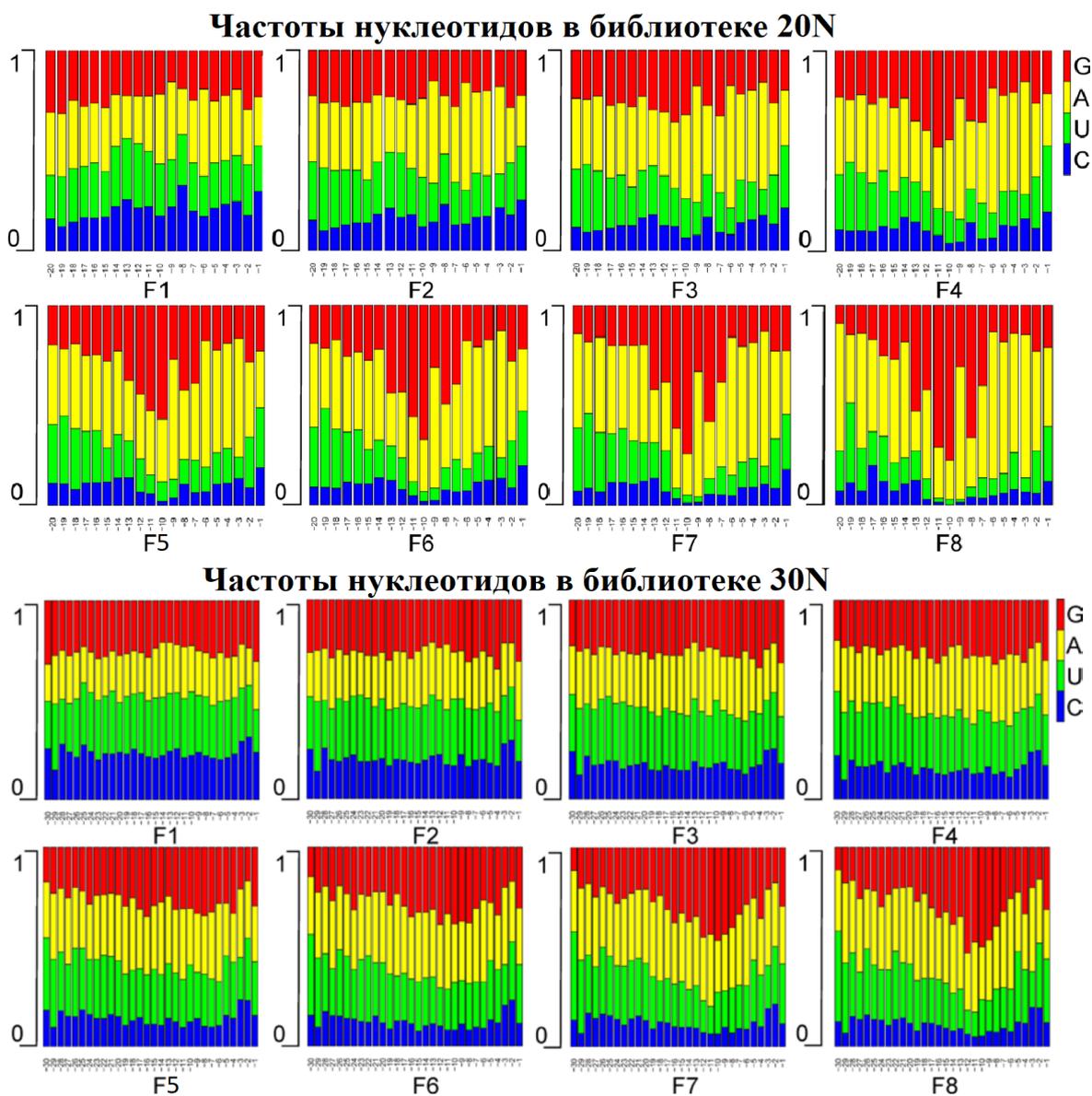


Рисунок 4. Влияние на эффективность трансляции нуклеотидного состава рандомизированных 20N- и 30N-фрагментов 5'-UTR. Панели F1-F8 соответствуют фракциям, отсортированным по увеличению эффективности трансляции. Нуклеотиды имеют цветовую кодировку: G – красный, A – жёлтый, U – зелёный, C – синий (легенда справа от панелей F4). Каждая полоска – доля нт в каждой позиции 5'-UTR относительно старт-кодона AUG (показаны снизу под графиками).

Анализ вторичной структуры всех найденных вариантов 5'-UTR (Рис. 5) показал, что для биосинтеза белка невыгодно образование сильной вторичной структуры 5'-UTR мРНК.

Для оценки важных особенностей 5'-UTR – частоты и длины SD в 5'-UTR мРНК с разным уровнем экспрессии (Рис. 6) рассчитали показатели сходства с канонической SD для каждых 9 нт «скользящего окна» во всех 5'-UTR из обеих библиотек. В итоге в каждой позиции «окна» для всех 5'-UTR из полученных наборов данных (Рис. 6) выявили устойчивое увеличение частоты SD-мотивов при переходе от плохо к эффективно транслируемым мРНК. Более того, в вариантах 5'-UTR мРНК, приводящих к высокому уровню биосинтеза белка, большинство таких мотивов расположено в оптимальных позициях относительно старт-кодона (Рис. 6 сравните F1 и F8).

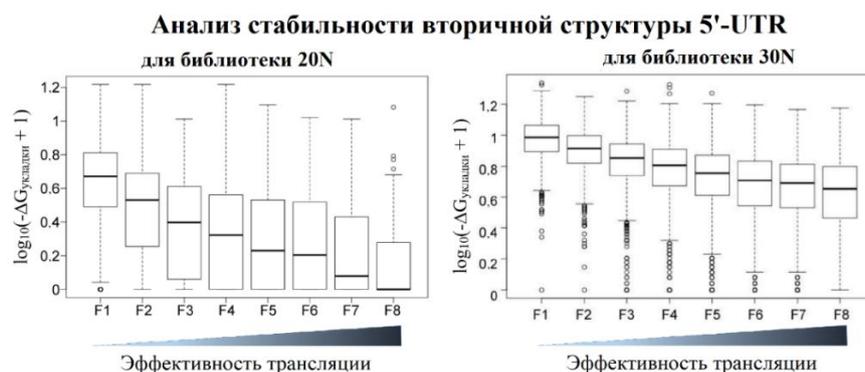


Рисунок 5. Влияние вторичной структуры на эффективность трансляции. Распределение энергии укладки 5'-UTR с 20N- (слева) и 30N- (справа) фрагментами с разным уровнем трансляции. Фракции F1-F8 отсортированы по возрастанию эффективности трансляции и отмечены внизу. Распределение значений энергии укладки, рассчитанной как $\log_{10}(-\Delta G_{\text{укладки}} + 1)$, показано в виде «ящика с усами».

Неожиданно, что среди эффективно транслируемых мРНК чаще встречаются 5'-UTR с несколькими SD-мотивами (Рис. 6). Схожее аддитивное влияние множественных SD-мотивов в 5'-UTR на уровень белка наблюдалось и при рандомизации спейсерного 4N-участка.

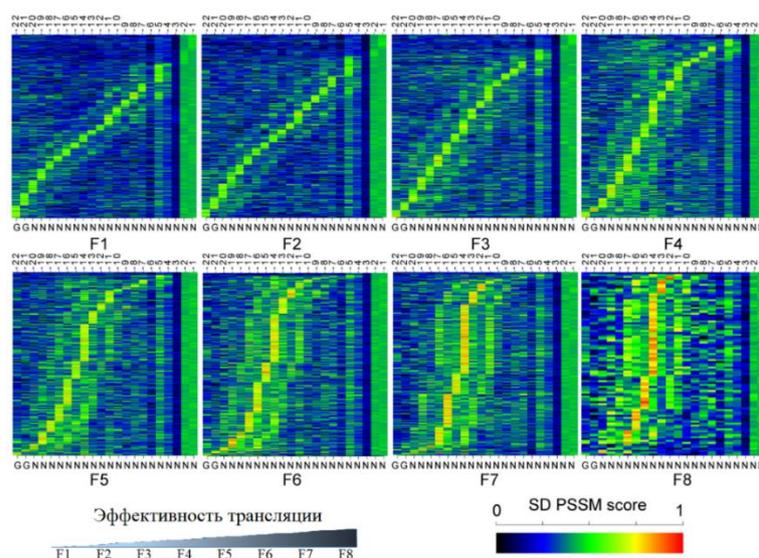


Рисунок 6. Распределение показателей сходства с SD по матрице PSSM (SD PSSM score) для позиций 5'-UTR мРНК с разной эффективностью трансляции из библиотеки 20N. Панели F1-F8 соответствуют фракциям, отсортированным по возрастанию эффективности трансляции от F1 до F8. Позиции относительно старт-кодона в мРНК вверху панелей, схема последовательности 5'-UTR с 20N-участками внизу. Цветные полосы отражают отдельные мРНК, сгруппированные по положению максимального показателя SD PSSM score, а их цвет – сходство фрагмента с началом на конкретном нуклеотиде и до SD, меняясь от чёрного (нет сходства) до красного (наибольшее сходство) по шкале значений SD PSSM score и соответствующих им цветов (внизу справа).

4. Нестандартные 5'-UTR с высокой эффективностью трансляции

Все известные детерминанты в составе 5'-UTR: стабильность вторичной структуры, SD-мотивы, AU-богатые энхансеры на 5'-конце влияют на эффективность трансляции, но только их наличие не объясняет всей наблюдаемой изменчивости в уровнях экспрессии для полученного набора данных. Для поиска новых энхансеров выбрали три 5'-UTR с минимальными SD-

подобными мотивами среди мРНК из фракций с наибольшим соотношением CER/RFP из обеих библиотек. Среди них две 5'-UTR длиной 32 нт: А-богатая и с короткой SD, а третья AG-богатая длиной 22 нт (Табл. 2).

Таблица 2. Эффективность трансляции мРНК с отобранными 5'-UTR и их мутантов.

Конструкция	Последовательность 5'-UTR	Измеренная эффективность трансляции	Предсказанная эффективность трансляции
Контроль	GGAGAAGGAGAUUCAU	1	1
А-богатая 5'-UTR	GGAUUUAAAAAAAAAGGCGGA AAAAUA AUG CAU	6,02	0,68
А-богатая 5'-UTR без AUG	GGAUUUAAAAAAAAAGGCGGA AAAAUAAUCAU	2,59	0,82
А-богатая 5'-UTR -2G	GGAUUUAAAAAAAAAGGCAAA AAAAUA AUG CAU	4,50	0,68
А-богатая 5'-UTR -5G	GGAUUUAAAAAAAAAACAAA AAAAUAAUCAU	0,38	1,57
2/7	CACACAACACCUGA <u>UCA</u> ACU	0,03	0,14
5'-UTR с короткой SD	GGUUUCUUAUUUGGUUCGGA GUG AGAUGCGAU	4,31	0,68
5'-UTR с короткой SD -SD	GGUUUCUUAUUUGGUUCCCU CUGAGAUGCGAU	0,076	0,13
5'-UTR с короткой SD -9UA	GGCCCCCGCCCGGUUCGGA GUG AGAUGCGAU	0,073	1,18
5'-UTR с короткой SD -3U	GGUUUCUUA <u>CCCGGUUCGGA</u> GUG AGAUGCGAU	0,51	0,65
5'-UTR с короткой SD -3'	GGUUUCUUAUUUGGUUCGGA GUCACA <u>UCCCAU</u>	2,35	0,18
5'-UTR с короткой SD +SD	GGUUUCUUAUUUGGUUCCAC AUGGAGUCCCAU	0,41	0,92
5'-UTR с короткой SD -GUC	GGUUUCUUAUUUGGUUCGGA GUCAGAUGCGAU	3,21	0,11
5'-UTR с короткой SD -GCG	GGUUUCUUAUUUGGUUCGGA GCGAGAUGCGAU	4,32	0,56
4/13	CACCCGGAGCAACAACACU	0,06	0,04
4/7	CACACAACCCGGAGCAACU	0,8	0,22
AG-богатая 5'-UTR	<u>GGAGUCUAAAGAGAGAGAGA</u> GU	5,09	2,05
AG-богатая 5'-UTR -AG ₁₋₃	<u>GGAGUCUAAACACACAGAGA</u> GU	0,15	0,13
AG-богатая 5'-UTR -AG ₄₋₆	<u>GGAGUCUAAAGAGAGACACA</u> CU	1,40	0,58
AG-богатая 5'-UTR -G	<u>GGAGUCUAAAGAGACAGAGA</u> GU	1,14	0,31
AG-богатая 5'-UTR +SD1	<u>GGAGUCUAAAGAGAGGAGAA</u> GU	1,52	4,18
AG-богатая 5'-UTR +SD2	<u>GGAGUCUAAAGGAGGAGAGA</u> GU	1,56	12,37
4/16	<u>CCGGAGCACACACAACAACU</u>	0,02	0,04

SD подчёркнута, добавочный старт-кодон показан красным, внесимые замены – жирным.

Для отобранных 5'-UTR провели мутационный анализ, встраивая найденные и изменённые их варианты в репортёрную конструкцию (Табл. 2).

А-богатая 5'-UTR имеет в составе 50% аденозинов, АУ-богатый участок на 5'-конце, 2 короткие SD (AAGG и GGA) на неоптимальном расстоянии от старт-кодона и дополнительный старт-кодон в спейсере и в рамке считывания с геном *cer*, при этом число остатков цитозина заметно уменьшено (2 нт), как и у многих 5'-UTR мРНК с высокой эффективностью трансляции (Рис. 4). Мутация добавочного старт-кодона снижает в 2 раза эффективность (Табл. 2, А-богатая 5'-UTR без AUG. Замена SD GGA на AAA (Табл. 2, А-богатая 5'-UTR -2G) приводит всего лишь к 1,3-кратному падению эффективности трансляции. Мутации добавочного кодона AUG и обеих SD вызывают 16-кратное снижение эффективности (Табл. 2, А-богатая 5'-UTR -5G). Последний мутант содержит не более 2 нт, комплементарных 3'-концу 16S рРНК, но его эффективность трансляции на порядок выше, чем у других репортёрных мРНК с SD из 2 нт (Табл. 2, конструкция 2/7). Высокую эффективность трансляции мРНК с А-богатой 5'-UTR можно объяснить аддитивным эффектом двух SD, вносящих разный вклад, добавочного старт-кодона и ещё не известных элементов 5'-UTR в дополнение к перечисленным.

Другая необычно эффективная мРНК включает SD длиной 4 нт на расстоянии 13 нт от старт-кодона (Табл. 2, 5'-UTR с короткой SD). Схожая конструкция 4/13 имеет на 2 порядка ниже соотношение CER/RFP (Табл. 2). Замена добавочного инициаторного кодона GUG (Табл. 2, 5'-UTR с короткой SD -GUC и -GCG соответственно) не изменяла эффективность трансляции. Мутации SD (Табл. 2, 5'-UTR с короткой SD -SD) приводят к отсутствию трансляции. Изменения в 5'-проксимальной области, не затрагивающие SD, сильно снижают уровень биосинтеза белка (Табл. 2, 5'-UTR с короткой SD -9UA). Даже замена 3 остатков уридина перед SD заметно подавляет трансляцию (Табл. 2, 5'-UTR с короткой SD -3U). Мутации в спейсере лишь умеренно влияют на уровень белка (Табл. 2, 5'-UTR с короткой SD -3'). Перемещение SD ближе к старт-кодону (Табл. 2, 5'-UTR с короткой SD +SD) уменьшает эффективность более, чем в 5 раз, но оно оказалось полезным для трансляции по сравнению с конструкцией 4/13 с SD из 4 нт (Табл. 2). Неожиданно высокую эффективность трансляции 5'-UTR с короткой SD можно объяснить наличием ранее не описанных энхансеров. Примечательно, что уровень биосинтеза CER очень чувствителен к мутациям в ней. Положение SD, удалённое от старт-кодона и неблагоприятное для других 5'-UTR, в этом случае выгодно для трансляции. У-богатый участок перед отдалённой SD служит энхансером трансляции (Табл. 2).

Третий вариант 5'-UTR с неожиданно высокой эффективностью трансляции содержит подряд 6 AG-повторов (Табл. 2, AG-богатая 5'-UTR) и дополнительно включает SD состава GGAG на неоптимальном расстоянии: 20 нт от старт-кодона, поэтому функциональность такой SD маловероятна. В соответствии с результатом Flow-seq для этой 5'-UTR соотношение CER/RFP,

измеренное флуориметром, в 5 раз выше, чем у контрольной (Табл. 2). Мутации, меняющие первые 3 и последние 3 AG-повтора (Табл. 2, AG-богатая 5'-UTR -AG₁₋₃ и -AG₄₋₆), приводили к 34 и 3,6-кратному уменьшению эффективности соответственно. Даже замена G на C в середине области с AG-повторами снижает уровень биосинтеза белка в 4,5 раза (Табл. 2, AG-богатая 5'-UTR -G). Возможно участок GAG в AG-повторах, выступает в качестве SD. Создание сильных SD (AGGAG и GGAGG) в этой 5'-UTR (Табл. 2, AG-богатая 5'-UTR +SD1 и +SD2 соответственно) в любом из этих положений снижает экспрессию белка примерно в 3,3 раза. Получается, что участок с AG-повторами служит уникальным энхансерным элементом трансляции. Даже после удаления наиболее функционально значимой его части (Табл. 2, AG-богатая 5'-UTR -AG₁₋₃) остаточный уровень биосинтеза белка сильно выше, чем у конструкции 4/16 с сопоставимой SD.

Для проверки является ли необычно высокая эффективность трансляции артефактным результатом взаимодействия между 5'-UTR и кодирующей областью гена *cer*, то есть путём формирования вторичной структуры большой протяжённости, кодирующие области генов *cer* и *rfp* поменяли для выборки репортерных конструкций. В результате обнаружили одинаковые тенденции в уровнях эффективности трансляции для обоих репортёрных генов *cer* и *rfp*.

При поиске AG-богатых энхансеров трансляции на месте SD по 5'-UTR природных генов *E. coli* нашли схожие вариации 5'-UTR перед генами *ybaB*, *lon*, *rplY* и *lrp*. Их полные 5'-UTR встроили перед геном *cer* в вектор pRFPCER. Для оценки значимости AG-фрагментов природных 5'-UTR для эффективности трансляции их заменяли AC-повторами и параллельно мутировали их на сильные SD. В итоге наблюдали, что внесение SD способствует трансляции лучше, чем исходные AG-повторы, тогда как мутации, приводящие к удалению AG-повторов во всех случаях приводили к снижению на порядок уровня биосинтеза белка. Ввиду чего AG-повторы можно рассматривать как энхансеры.

Полученные результаты анализа выбранных 5'-UTR и их мутантов позволяют отметить следующее. Не все наблюдаемые эффекты можно объяснить, исходя из стабильности вторичных структур вблизи старт-кодона или наличия, либо отсутствия SD с оптимальными длиной и расстоянием от старт-кодона. AU-богатый участок (как и U-богатая область) на 5'-конце и нуклеотидный состав среднего участка 5'-UTR, оказывают значительное влияние на эффективность трансляции. AG-повторы в RBS служат уникальным энхансерным элементом. Аддитивный эффект не одной SD, добавочного старт-кодона приводит к увеличению уровня биосинтеза белка. Внесение в состав 5'-UTR большего числа цитидинов вызывает регулярное снижение эффективности трансляции.

5. Изучение влияния природных 5'-UTR на эффективность трансляции

Выборку из 713 природных 5'-UTR первых генов оперонов *E. coli*, ограниченную вариантами длиной 2-60 нт, адаптировали для использования Flow-seq. Для оценки влияния

выбранных вариантов на уровень биосинтеза белка создали библиотеку конструкций, содержащих перед старт-кодоном гена *cer* 713 вариантов 5'-UTR разной длины.

Полученные плазмиды rFP CER T7 содержали гены флуоресцентных белков, где первый *rfp* находился под контролем промотора T5, а второй *cer* – под контролем T7 промотора, причём оба гена имели свои терминаторы транскрипции. Перед геном *rfp* 5'-UTR не менялась, а перед *cer* встраивали природные 5'-UTR. Полученной библиотекой конструкций трансформировали клетки BL21 (DE3) *E.coli*. С помощью FACS эти клетки разделили по соотношению CER/RFP на 8 фракций F1-F8 так, что F8 включала клоны с наибольшей эффективностью трансляции. Фракцию F8 составили 0,3% клеток, что на порядок больше, чем в случае библиотеки 20N, и близко к доле клеток в ней в библиотеке 30N (0,45%). Общая разница эффективностей между крайними фракциями составила около 30 раз, что значительно меньше диапазона для рандомизированных библиотек. NGS позволило установить соответствие между клетками, несущими определённые варианты вставок в плаزمиде, и их распределением по фракциям, разделённым по соотношению CER/RFP. В результате число природных 5'-UTR в 8 фракциях, представленное к последующему анализу, составило 648 из 713 вариантов, использованных для клонирования.

Для набора полученных природных 5'-UTR проанализировали особенности, важные для трансляции. Сначала оценили число вариантов 5'-UTR во фракциях с разным соотношением CER/RFP, отражающим эффективность трансляции при использовании Flow-seq (Рис. 7А). В итоговой библиотеке большинство 5'-UTR обеспечивало умеренную экспрессию.

Наличие менее стабильных вторичных структур РНК (Рис. 7Б) и более сильного взаимодействия с 3'-концом 16S рРНК (Рис. 7В), как и ожидалось, наблюдали в более эффективно транслирующихся мРНК. Однако диапазон различий по этим признакам между фракциями много меньше, чем наблюдалось для полностью рандомизированных 5'-UTR. Очевидно, природные 5'-UTR эволюционировали таким образом, чтобы среди них не было совсем неэффективных.

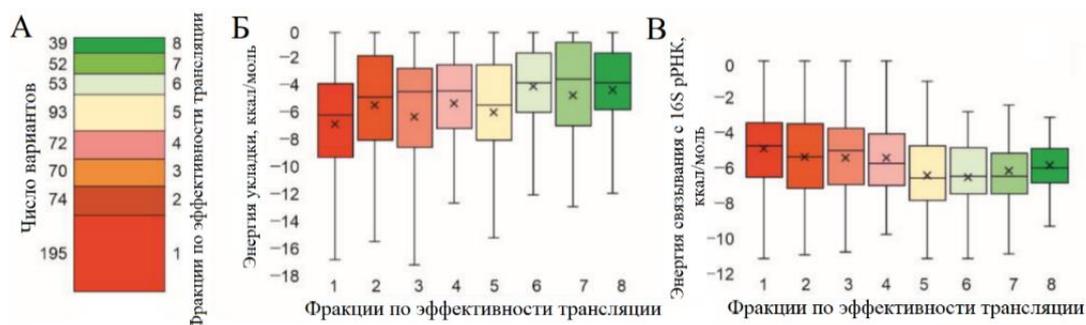


Рисунок 7. Особенности природных 5'-UTR с разной эффективностью трансляции. Номера 1-8 соответствуют фракциям F1-F8, отсортированным по увеличению эффективности. А. Число вариантов 5'-UTR (слева) во фракциях 1-8 (справа). Б. Распределение энергии укладки участка мРНК, включающего 5'-UTR и первые 16 нт кодирующей области гена *cer*, во фракциях. В. Распределение энергии взаимодействия 5'-UTR с участком анти-SD 16S рРНК во фракциях. Для панелей Б-В распределения значений показаны в виде «ящика с усами». Горизонтальная линия соответствует медиане, а крестик – среднему значению.

Эффективности трансляции, определённые методом Flow-seq, в целом согласуются с данными, полученными *in vivo* методом Ribo-seq для природных мРНК.

ЗАКЛЮЧЕНИЕ

Понимание того, как и какие особенности мРНК влияют на эффективность трансляции, формировалось на протяжении десятилетий, начиная с открытия SD в 5'-UTR небольшого набора мРНК у бактерий. Позже оказалось, что есть и другие элементы мРНК, важные в определении уровня экспрессии белка. Однако не все наблюдаемые эффекты в биосинтезе белка объясняются известными детерминантами. С целью изучить особенности 5'-UTR мРНК бактерий, влияющие на эффективность трансляции, метод Flow-seq адаптировали для одновременного определения уровня биосинтеза белка для библиотек репортёрных конструкций из более 24,5 тысяч вариантов синтетических и природных 5'-UTR, многократно превышающих их природное разнообразие у *E. coli*. Анализ полученных наборов мРНК показал наличие известных элементов: SD, AU-богатого энхансера на 5'-конце мРНК, вторичных структур, которые легко расплавить, или их отсутствие в 5'-UTR мРНК, обеспечивающих высокий уровень экспрессии белка. Вместе с тем представили и ранее не описанные особенности 5'-UTR мРНК, способствующие увеличению эффективности трансляции: низкая доля остатков цитозина, множественные SD, AG-повторы. Также отметили, что более эффективно транслируемые варианты мРНК чаще присутствуют в библиотеках 5'-UTR с большей длиной (30 нт), чем в наборе 5'-UTR с длиной 20 нт. Для набора мРНК с природными 5'-UTR показали низкую вариабельность эффективности трансляции, достигнутую при эволюционном отборе в пользу более эффективных вариантов. Впервые проведённый анализ результатов, полученных для 648 природных 5'-UTR перед геном *ser* в составе репортёрных конструкций с помощью Flow-seq и для природных мРНК с тем же набором 5'-UTR генов *E. coli* с помощью Ribo-seq, выявил хорошее соответствие между ними.

Полученные результаты представляют интерес с фундаментальной точки зрения, так как проливают новый свет на то, как последовательность 5'-UTR мРНК бактерий влияет на уровень биосинтеза белка. Вместе с тем они позволяют учесть особенности трансляции на практике.

ВЫВОДЫ

1. Метод Flow-seq успешно адаптирован для высокопроизводительного определения эффективности трансляции синтетических и природных 5'-нетранслируемых областей, а следовательно и участков посадки рибосомы, многократно превышающих их природное разнообразие у *E. coli*.
2. Показано, что высокая эффективность трансляции некоторых мРНК не может быть объяснена исключительно наиболее изученными особенностями, такими как взаимодействие рибосомы с последовательностью Шайна-Дальгарно, наличие AU-богатого энхансера на 5'-конце мРНК и отсутствие ингибирующего влияния вторичной структуры мРНК.

3. Продемонстрировано, что эффективность трансляции уменьшается с увеличением доли остатков цитозина в последовательности 5'-нетранслируемой области.
4. Отмечено, что дополнительный старт-кодон и множественные последовательности Шайна-Дальгарно имеют аддитивное влияние на эффективность трансляции.
5. Обнаружено, что AG-повторы в 5'-нетранслируемой области представляют собой энхансер трансляции.
6. Выявлено, что мРНК с природными 5'-нетранслируемыми областями менее вариабельны по эффективности трансляции, чем рандомизированные последовательности. Замечено, что в более длинных, 30-звенных 5'-нетранслируемых областях чаще встречаются благоприятные для трансляции варианты из случайных последовательностей, чем в 20-звенных.

СПИСОК ПУБЛИКАЦИЙ АВТОРА ПО ТЕМЕ РАБОТЫ

Статьи в рецензируемых научных изданиях, индексируемых в базах данных Web of Science и Scopus, рекомендованных для защиты в диссертационном совете МГУ.014.2 по специальности 1.5.3. Молекулярная биология (химические науки):

1. **Komarova Ekaterina S., Dontsova Olga A., Pyshnyi Dmitry V., Kabilov Marsel R., Sergiev Petr V.** Flow-Seq Method: Features and Application in Bacterial Translation Studies // *Acta Naturae* (англоязычная версия), 2022. Vol. 14. № 4. P. 20-37. **IF 2,0 (Web of Science).**
doi: 10.32607/actanaturae.11820
2. **Komarova Ekaterina S., Slesarchuk Anna N., Rubtsova Maria P., Osterman Ilya A., Tupikin Alexey E., Pyshnyi Dmitry V., Dontsova Olga A., Kabilov Marsel R., Sergiev Petr V.** Flow-Seq Evaluation of Translation Driven by a Set of Natural *Escherichia coli* 5'-UTR of Variable Length // *International Journal of Molecular Sciences*, 2022. Vol. 23. № 20. P. 12293. **IF 5,6 (Web of Science).**
doi: 10.3390/ijms232012293
3. **Komarova Ekaterina S., Chervontseva Zoya S., Osterman Ilya A., Evfratov Sergey A., Rubtsova Maria P., Zatsepin Timofei S., Semashko Tatiana A., Kostryukova Elena S., Bogdanov Alexey A., Gelfand Mikhail S., Dontsova Olga A., Sergiev Petr V.** Influence of the spacer region between the Shine-Dalgarno box and the start codon for fine-tuning of the translation efficiency in *Escherichia coli* // *Microbial Biotechnology*, 2020. Vol. 13. № 4. P. 1254-1261. **IF 5,7 (Web of Science).**
doi: 10.1111/1751-7915.13561
4. Evfratov S.A.¹, Osterman Ilya A.¹, **Komarova Ekaterina S.¹**, Pogorelskaya A.M., Rubtsova M.P., Zatsepin T.S., Semashko T.A., Kostryukova E.S., Mironov A.A., Burnaev E., Krymova E., Gelfand M.S., Govorun V.M., Bogdanov A.A., Sergiev P.V., Dontsova O.A. Application of sorting and next generation sequencing to study 5'-UTR influence on translation efficiency in *Escherichia coli* // *Nucleic Acids Research*, 2017. Vol. 45. № 6. P. 3487-3502. (¹-равный вклад авторов). **IF 14,9 (Web of Science).** **doi: 10.1093/nar/gkw1141**