

УДК: 579:252

REP-элементы генома *Escherichia coli* и сигналы транскрипции: позиционный и функциональный анализ

Маркелова Н.Ю.^{1,2}, Масулис И.С.^{1,2}, Озолин О.Н.*^{1,2}

¹*Институт биофизики клетки, Российская академия наук, Пущино, Московская область, Россия*

²*Пущинский государственный естественно-научный институт, Пущино, Московская область, Россия*

Аннотация. В межгенных участках генома кишечной палочки находится 356 участков, содержащих от 1 до 12 повторяющихся последовательностей с вырожденным консенсусом, названных REP-элементами. Их биологическая роль мало понятна, но множественность в геноме, преимущественная локализация между конвергентными генами и способность формировать шпильчатые структуры послужили основанием для предположения об участии REP-элементов в терминации транскрипции или стабилизации соответствующих РНК, хотя прямые эксперименты не подтвердили способность модельных REP-элементов останавливать синтез РНК. В данной работе позиционный и функциональный анализ, выполненный для всей совокупности аннотированных REP-элементов, позволил установить факт снижения эффективности сквозного синтеза РНК для многих REP-модулей. Однако часть из них не влияла на процессивность транскрипции, что предполагает возможность регуляторного действия REP-элементов в составе РНК. Были также обнаружены REP-последовательности, после которых синтез РНК усиливался и найдены перекрывающиеся с ними промоторы. Выявлен особый характер расположения REP-элементов относительно *промоторных островков*, предполагающий вклад повторяющихся модулей в транскрипционную изоляцию *островков* и их функциональную автономию, а также биологическое значение *островковых* РНК.

Ключевые слова: структурные элементы бактериального генома, REP-элементы, промоторы, *промоторные островки*, механизмы регуляции транскрипции.

ВВЕДЕНИЕ

Повторяющиеся элементы длиной 25–35 нуклеотидных пар (н. п.), содержащие инвертированные последовательности, способные формировать стабильные вторичные структуры, были обнаружены в геномах эубактерий в начале 1980-х и названы REP (Repetitive Extragenic Palindromic)-элементами [1]. Для них был предложен консенсус: GC(g/t)GATGGCG(g/a)GC(g/t)...(g/a)CG(c/t)СТТАТC(c/a)GGCCTAC. Число таких повторов в геноме кишечной палочки по разным оценкам варьирует от 500 до 1000. Преимущественная ассоциация REP-элементов с концами генов вызвали к жизни множество предположений относительно их физиологической роли, механизмов возникновения и эволюции [2]. Так, способность формировать шпильки в РНК в случае присутствия REP-элемента в транскрибируемом участке генома, позволяла

*Ozoline@rambler.ru

рассматривать этот мотив в качестве потенциального аттенюатора или терминатора транскрипции. Такая функция была проверена экспериментально и показано, что сама по себе последовательность REP-элемента в составе матрицы для синтеза РНК не обладает способностью блокировать элонгацию [3]. Позднее были получены убедительные данные, свидетельствующие о способности этих мотивов инициировать р-зависимую аттенюацию [4]. Тем не менее, вопрос о способности REP-элементов влиять на процессивность транскрипции остаётся в значительной степени открытым.

Большое число работ было посвящено роли REP-элементов в стабилизации мРНК и предотвращении их деградации с участием 3'-5' экзонуклеаз [5-8]. Рациональную основу для рассмотрения стабилизирующей роли структурообразующего мотива составляло предположение о том, что шпилька в составе РНК является стерическим препятствием для миграции экзонуклеазы в направлении 3'-5'. Она же является мишенью для РНКаз, разрушающих двухнитевые РНК, и РНКазы E, атакующей однонитевые петли в шпильчатых структурах, что освобождает 3'-конец РНК для экзонуклеолиза. Поэтому предположение о протекторной функции REP-элементов изначально было спорным. Тем не менее, как *in vitro*, так и *in vivo* были получены доказательства сокращения времени жизни РНК при удалении REP-элемента в 3'-концевой нетранслируемой области [6]. Поэтому представление о преимущественной функции REP-последовательности в качестве протектора мРНК оставалось доминирующим на протяжении двух десятилетий.

Наличие взаимно-комплементарных мотивов в структуре REP-элементов, имеющих сродство к ДНК-гиразе, послужило основанием для предположения об их участии в процессах модуляции топологического состояния ДНК [9]. Взаимодействуя с ними, ДНК-гираза может вызывать релаксацию положительных супервитков в ДНК, закономерно накапливающихся в 3'-концевых областях генов в результате транскрипции и ингибирующих считывание следующего гена [10]. Специфически связывающие топоизомеразу REP-элементы, расположенные между последовательными генами, могут снижать это напряжение, стимулируя синтез РНК [11].

По мере накопления данных о полногеномных последовательностях широкого круга микроорганизмов появилась возможность поиска аналогичных элементов у других бактерий и их сравнительного эволюционного анализа. Такой анализ, выполненный для повторяющихся последовательностей в геномах 66 бактерий, позволил классифицировать мотивы с инвертированными повторами по консервативным тетрауклеотидным блокам на 5'-концах (семейства GTAG и CGTC) [12]. При этом последовательность самого структурообразующего палиндрома оказалась гораздо более вариабельной. Геномы эубактерий обогащены повторами GTAG-типа, причем их количество у патогенных штаммов *E. coli* почти вдвое меньше, чем у свободноживущих непатогенных бактерий [12]. Это различие может указывать на участие REP-элементов в адаптации микроорганизмов к вариабельным условиям внешней среды, что обычно сопряжено не только с модуляцией экспрессии имеющихся генов, но и с возможностью ассимиляции новых генов. В этой связи важно отметить, что механизм появления в геномах REP-элементов остаётся наименее понятным, но предположение об их экспансии внутри генома за счёт генной конверсии, с определённой частотой возникающей в процессе рекомбинации, представляется вполне реалистичным. Однако оно пока не подкреплено экспериментальными данными и не объясняет специфической локализации REP-элементов практически только в межгенных локусах.

Таким образом, экспериментальный материал и результаты теоретического моделирования пока не позволили сформулировать общепризнанное представление о биологической роли REP-элементов, механизмах их возникновения, функционирования и эволюции. Поэтому в данной работе предпринята попытка спроецировать новые

данные экспрессионного и геномного анализа на уже имеющуюся информацию, акцентируя внимание не только на взаимном расположении REP-элементов с концами генов, но и на их локализации относительно промоторных областей разных типов.

МЕТОДЫ И АЛГОРИТМЫ

Геном кишечной палочки и исследованные наборы REP-последовательностей

Для анализа был использован геном *Escherichia coli* K12 MG1655 (*E. coli*), карта расположения генов и координаты REP-элементов, взятые из GenBank (NC_000913.3). Позиционный анализ относительно соседних генов был выполнен для всех 356 аннотированных REP-элементов и для выборки из 224 элементов, наиболее близких к последовательности GCCGGATGCGGCGTGAACGCCTTATCCGGCCTACGA ($2e^{-14} \leq E \leq 3e^{-4}$). Размер аннотированных REP-модулей варьирует в очень широком диапазоне от 14 до 767 н. п. Размер последовательностей во второй выборке варьировал от 25 до 36 н. п. Во всех случаях оценивалось расстояние между ближайшими границами REP-элемента и началом или концом соседнего гена.

Наборы промоторных участков и позиционный анализ

Для сравнительного позиционирования аннотированных REP-элементов относительно мест инициации транскрипции было использовано три набора промоторных областей. Один из них был составлен из 404 промоторов, стартовые точки которых были картированы индивидуально классическими биохимическими методами. Чтобы размер этой компиляции был сопоставим с числом REP-элементов, из более чем 800 известных промоторов *E. coli* в неё были отобраны только такие, которые распознаются σ^{70} -РНК-полимеразой и, согласно специализированной базе данных Regulon [13], не имеют рядом других промоторов с такой же σ -специфичностью. В процессе сравнительного позиционирования REP-элементы искали в диапазоне ± 300 н. п. вокруг стартовой точки транскрипции. Вторая компиляция содержала 51 промоторную область, включающую от 2 до 6 промоторов, начинающих синтез РНК в противоположных направлениях (набор дивергентных промоторов). К сожалению, в неё не вошли две области, находящиеся между дивергентными генами и содержащие REP-элементы, потому что в них картирована точка старта только для одного промотора. Третий набор был составлен из 434 смешанных промоторных островков, отобранных ранее [14] по наличию не менее 8 потенциальных точек инициации транскрипции в каждом скользящем окне размером 100 н. п. на протяжении ≥ 360 н. п. Для поиска потенциальных мест инициации транскрипции в этом случае был использован алгоритм PlatPromU [15], игнорирующий контекст консервативных элементов, распознаваемых σ -субъединицами РНК-полимеразы, и поэтому способный обнаруживать промоторы всех σ -факторов. Значимыми считали потенциальные промоторы, рассчитанный показатель промотор-подобия для которых превышал фоновый уровень на 4 StD ($p < 0.00004$). Средний размер этих областей составляет 660 н. п. Поиск REP элементов для двух последних наборов осуществлялся внутри промоторных областей и вдоль 300 н. п., фланкирующих их по обоим концам. Поиск потенциальных промоторов, распознаваемых σ^{70} -РНК-полимеразой для генов *sapA*, *utjA* и *riuP* был осуществлён алгоритмом PlatProm, как описано ранее [16].

Функциональный анализ

Для того чтобы оценить способность транскрипционного комплекса преодолевать потенциальный барьер, созданный шпильчными структурами REP-элементов, были использованы экспрессионные данные, полученные в работе [17] на микроматрицах

высокой плотности. Зонды этих микроматриц имеют длину 50 нуклеотидов и представляют обе нити генома кишечной палочки с шагом в 25 н. п. Эффективность экспрессии любого локуса генома можно оценить по интенсивности сигнала гибридизации между соответствующими зондами и помеченными флуоресцентной меткой ДНК-копиями клеточных РНК (E_{exp}). Для нормировки использовали аналогичные сигналы гибридизации, полученные с фрагментированной ДНК (E_{gen}). В качестве показателя эффективности транскрипции использовали общепринятый параметр $\log_2 R$, где $R = E_{\text{exp}}/E_{\text{gen}}$. Влияние REP-элементов на уровень транскрипции оценивали по количеству в клетках РНК-матриц для кДНК, гибридизующихся со всеми зондами, соответствующими участку в 1250 н. п. перед концом REP-элемента и 250 н. п. после него. Для генов и *промоторных островков* короче 1250 н. п. учитывали гибридизационные сигналы от начала генов или *островков*. Полученные профили гибридизации были выровнены по дальней по ходу транскрипции границе REP-элемента.

Поиск участков, гомологичных REP-элементам

Множественная представленность REP-элементов в геноме предполагает возможность формирования ими полностью или частично комплементарных дуплексов с похожими последовательностями в других РНК и в регуляторных участках генов. Для поиска таких мишеней вблизи обычных промоторов и *промоторных островков E. coli* был использован Microbial BLAST (NCBI) [18], на который в качестве запроса подавалась полная нуклеотидная последовательность каждого REP-элемента и заданные по умолчанию параметры. Однако в силу особенностей искомым последовательностей, во время сканирования не отключали анализ участков низкой сложности (low complexity regions). Анализ подвергался только геном *E. coli* K12 MG1655. Полученные гомологичные последовательности проецировали на геном. Значимыми считали гомологичные участки длиной ≥ 14 н. п. (минимальная длина аннотированного REP-элемента), попадающие в исследуемые промоторные и припромоторные области.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

REP-элемент, фланкирующий конец гена *utmA* транскрибируется в антисмысловом направлении

Ранее было установлено, что REP-элементы чаще всего располагаются в непосредственной близости от конца соседнего гена. Классический REP-модуль, в частности, находится за геном *utmA* (рис. 1,А). Он имеет последовательность GCCAGATGCGGCGTGAACGCTTTATCCGGACAAC, которая хорошо соответствует консенсусу, предложенному Stern et al. в 1984 г. (рис. 1,В) и формирует стабильную вторичную структуру со свободной энергией фолдинга в -10 кКал/моль (рис. 1,С). Особенностью гена *utmA* является наличие второго REP-элемента, расположенного в соседней межгенной области. Оба REP-элемента относятся к семейству GATG. В геноме кишечной палочки имеется только 23 гена, окружённых REP-элементами с обеих сторон. И хотя REP2 может быть атрибутом гена *riuP*, его влияние на экспрессию *utmA* тоже вполне вероятно. Длина REP2 немного меньше, чем REP1, а уровень соответствия консенсусной последовательности почти такой же (рис. 1,В), но минимальную свободную энергию фолдинга (-6 кКал/моль) имеет конфигурация, заметно отличающаяся от предыдущей (рис. 1,С). Конформационные свойства известных REP-элементов могут, следовательно, существенно отличаться.

Особый интерес к исследованию этой области генома был обусловлен наличием потенциального промотора P_{an} , способного инициировать синтез либо новой РНК из

межгенной области *ymjA-sapA*, заканчивая его на REP-элементе, либо синтез регуляторной антисмысловой РНК к мРНК *YmjA*. В последнем случае последовательность REP1 окажется в составе РНК-продукта. Обе возможности фундаментально значимы, поэтому фрагмент генома, содержащий межгенный участок *sapA/ymjA* и всю кодирующую область *ymjA* (амплифицирован между праймерами 1 и 2, рис. 1,А), был клонирован в экспрессионный вектор pGEMAX [19]. В препаратах РНК, полученных из клеток *E. coli* K12 Mg1655, трансформированных этим вектором, была обнаружена РНК, гибридизующаяся с радиоактивно меченным зондом, комплементарным антисмысловому продукту (волнистая линия на рис. 1). Длина этой РНК оказалась больше 400 н. п., что означает сквозную транскрипцию регуляторной области между генами *ymjA* и *sapA* и, следовательно, включение REP1 в РНК-продукт.

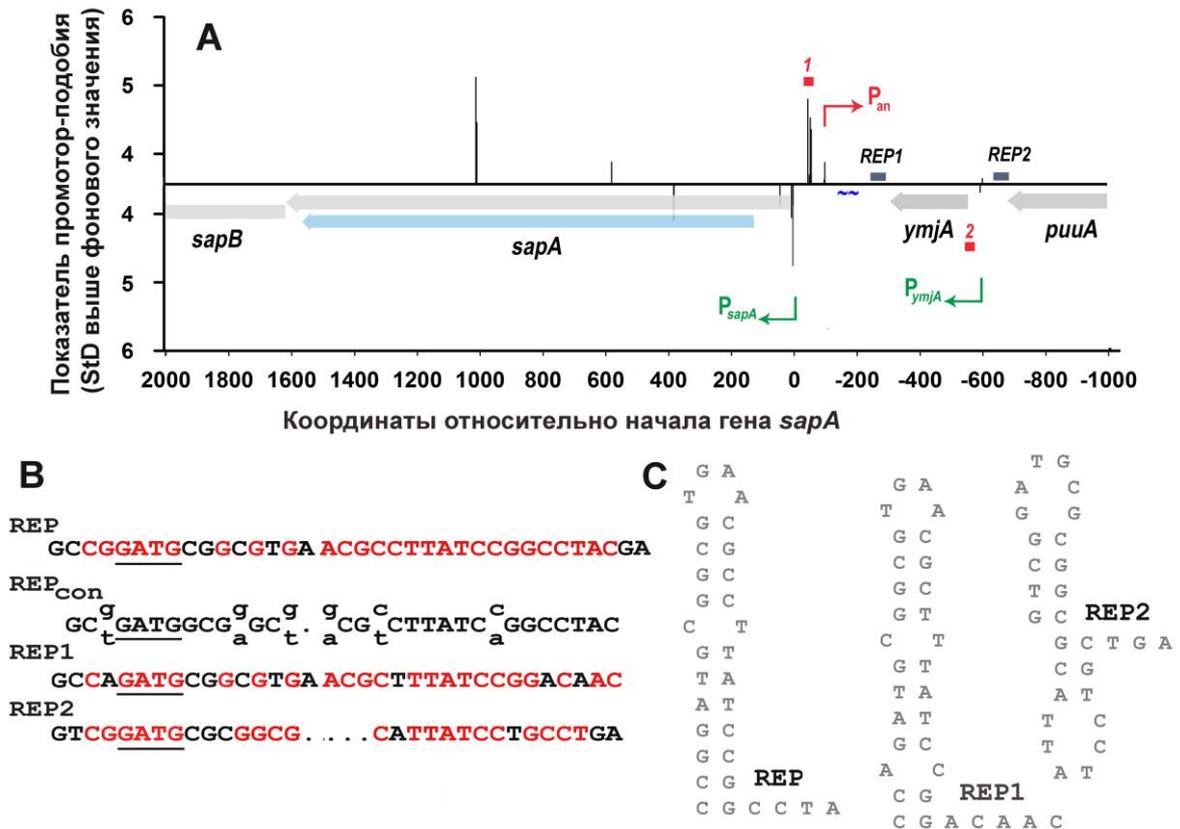


Рис. 1. А. Распределение потенциальных точек инициации транскрипции (столбики), предсказанных алгоритмом PlatProm [16]. Высота столбиков выше и ниже оси X соответствует степени промотор-подобия, рассчитанной для верхней и нижней нитей генома, соответственно. Широкими стрелками показано расположение генов и направление их транскрипции. Голубой стрелкой для *sapA* показана альтернативная открытая рамка считывания (ОРС). Тёмными прямоугольниками показаны места расположения REP1 и REP2. Красными квадратиками обозначены праймеры, использованные для амплификации фрагмента ДНК, клонированного в вектор pGEMAX, волнистой линией обозначен зонд для гибридизации. В. Последовательность наиболее представленного в геноме REP-элемента (REP), консенсус REP-последовательностей, предложенный Stern et al. [1] (REP_{con}), а также контекст REP1 и REP2. Красным выделены нуклеотиды, совпадающие с REP_{con}. Мотив GATG подчеркнут. С. Шпилечные структуры, предложенные программой RNA Structure [20] для REP, REP1 и REP2.

REP-элементы не случайно расположены относительно начала и конца генов

Для нуклеотидной последовательности в 380 н. п., гарантированно попадающей в транскрибируемую с P_{an} область, был выполнен поиск гомологичных участков с использованием Microbial Nucleotide BLAST [18], но никаких других мест множественного подобия, кроме удлинённого на две нуклеотидные пары REP1

(рис. 1,В), обнаружено не было. Этот удлинённый мотив имеет в геноме более 200 гомологичных участков и самая близкая к нему последовательность (REP, рис. 1,В), с ещё большим числом гомологов была выбрана в качестве шаблона. Она входит в состав трёх аннотированных REP-элементов, а последовательность первых 33 пар 44 раза повторена в геноме. В первой половине REP есть отличия от REP_{con}, предложенного Stern et al. [1]. При использовании REP в качестве запроса было отобрано 224 гомологичных мотива и определена их удалённость от концов соседних генов (рис. 2).

В соответствии с опубликованными данными [21], значительная часть этих мотивов оказалась расположенной на расстоянии 10–20 н. п. от конца ближайшего гена (рис. 2). На основании этого ранее была предложена модель, в которой обсуждается участие палиндромного мотива REP-элемента в регуляции трансляции мРНК за счет остановки элонгирующего комплекса с последующим привлечением РНКазы R и деградацией терминального участка мРНК [21].

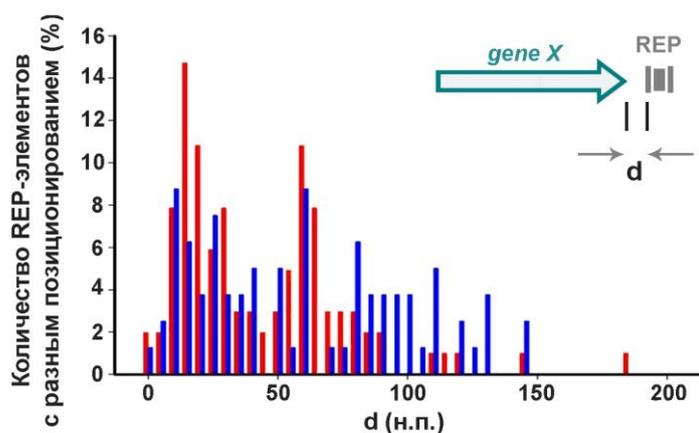


Рис. 2. Распределение расстояний (d) от 3'-концов кодирующих последовательностей для последовательностей, гомологичных REP (рис. 1,В). Красным цветом представлены повторы, расположенные между конвергентными генами, синим – между коллинеарными генами. Для тандемных повторов принималось в расчет наименьшее расстояние до 3'-конца кодирующей последовательности, в случае конвергентной ориентации фланкирующих генов использовалось только меньшее значение из двух возможных. Ориентация мотива по отношению к направлению транскрипции не учитывалась.

Однако, такой механизм действия оказался возможен только для повторов, удаленных от конца гена на расстояние, не превышающее 15 нуклеотидов [21]. Но значительная часть исследованных нами мотивов оказалась более удаленной от терминирующего кодона и формирует второй пик на расстоянии 60–65 нуклеотидов от конца гена (рис. 2). В случае коллинеарных генов, например, на расстояниях > 15 н. п. находится $\sim 80\%$ REP-мотивов (синие столбики на рис. 2). Поэтому этот анализ был повторён для другого набора, в который вошли все 356 аннотированных в геноме *E. coli* REP-элементов (рис. 3,В). Для конвергентных генов в этом случае оценивались расстояния до обоих соседних генов, а границами REP-элемента считались указанные в геноме координаты первой и последней пары, а не концы консервативного мотива. Явно доминирующими оказались расстояния от 1 до 20 н. п. (52.5%), а позиционирование с расстоянием в 15 н. п. и меньше встречалось в 45% случаев. Тем не менее, и более удалённое расположение REP-элементов с расстоянием больше 70 н. п. тоже наблюдалось. Таким образом, значительная часть этих модулей может быть задействована в аттенюации трансляции и деградации мРНК с помощью РНКазы R по механизму, предложенному в работе [21], но биологическое значение половины исследованных повторов ещё нуждается в концептуальном осмыслении.

Согласно последней аннотации генома *E. coli*, размер спейсера между коллинеарными генами, содержащими REP-элементы, варьирует в широком диапазоне (47–6176 н. п.) В среднем (375 н. п.) они больше, чем средняя длина межгенной области в геноме (241 н. п.). Отчасти это обусловлено естественным ограничением на присутствие REP-элементов внутри оперонов, где расстояния между генами очень короткие, а иногда и отрицательные из-за перекрывания стоп-кодона TGA с иницирующим кодоном ATG. Размеры многих межгенных областей с REP-элементами вполне допускают наличие в них новых генов или промоторов для соседних генов. Поэтому был исследован характер распределения REP-элементов относительно иницирующих кодонов соседних генов и установлено, что ближняя граница REP-элемента в таких областях, как правило, находится в потенциальной промоторной области (рис. 3,А). (Стартовые точки ~ 90% известных промоторов удалены от кодирующей последовательности не более чем на 250 н. п. [22], а максимум приходится на первые 40 н. п.). Это значит, что REP-элементы могут быть не случайным образом позиционированы не только относительно конца генов, но и относительно их начала.

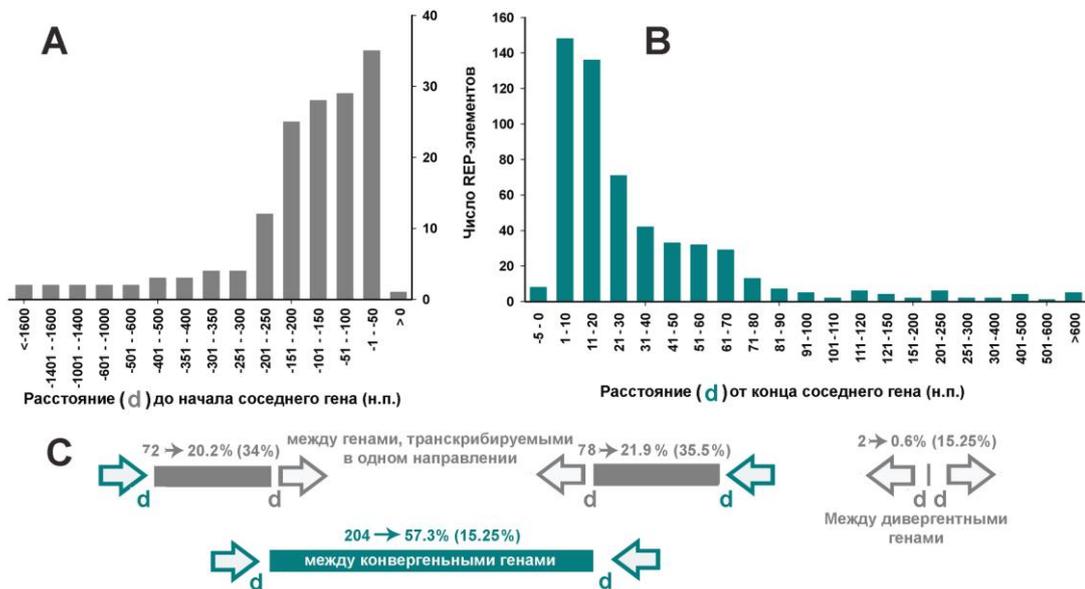


Рис. 3. Распределение расстояний (*d*) аннотированных REP-элементов относительно начала (А) и конца (В) соседних генов. Во всех случаях оценивалось расстояние до ближней границы REP-элемента. Каждый аннотированный в геноме REP-элемент рассматривался как единый модуль, даже если он состоял из нескольких повторяющихся мотивов. Между конвергентными и дивергентными генами учитывались оба расстояния до соседних генов. Ориентация мотива по отношению к направлению транскрипции не учитывалась. На рисунке С серыми и бирюзовыми стрелками обозначены гены, фланкирующие REP-элементы во всех возможных взаимных ориентациях. Межгенные области обозначены серыми прямоугольниками. Их длина пропорциональна количеству пар каждого типа в исследованной выборке межгенных участков, содержащих REP-элементы. Над схемами указано число соответствующих пар и их процент в выборке. В скобках указан процент соседних генов с аналогичной ориентацией в геноме.

REP-элементы могут подавлять сквозную транскрипцию или не влиять на неё

Большая часть REP-элементов (57.3%, рис. 3,С) находится между конвергентными генами, и число таких межгенных участков в 3.8 раз превышает процент генов со встречной транскрипцией в геноме. Эта явная разница вполне закономерно послужила основанием для предположений об участии REP-элементов в терминации транскрипции и/или последующем процессинге РНК-продуктов. В соответствии с этим, REP-элементы обнаружены только в двух промоторных участках между

дивергентными генами (рис. 3,С), что в 25 раз меньше ожидаемого. Но число REP-элементов, лежащих между коллинеарными генами, где терминаторная или аттенуаторная функция могут быть не менее востребованы, чем между конвергентными генами, тоже оказалось в 1.65 раз меньше ожидаемого (рис. 3,С). То есть, значимость терминаторной функции для REP-элементов как таковых не очевидна.

Поэтому на следующем этапе мы использовали результаты полногеномного экспрессионного анализа, полученные на микроматрицах высокой плотности [17] с тем, чтобы исследовать профиль транскрипции генов, расположенных перед REP-элементами, самих REP-элементов и коротких участков генома, фланкирующих эти элементы. Продуктивность транскрипции оценивали по интенсивности сигнала гибридизации между зондами микроматрицы и помеченными флуоресцентной меткой ДНК-копиями клеточных РНК (E_{exp}). Зарегистрированные сигналы нормировали на сигналы гибридизации с этими же зондами, полученными с фрагментированной ДНК, помеченной другой флуоресцентной меткой (E_{gen}). Об эффективности транскрипции судили по величине $E = \log_2 R$, где $R = E_{exp}/E_{gen}$. При этом нулевое значение E означает присутствие одной копии соответствующей РНК на клетку.

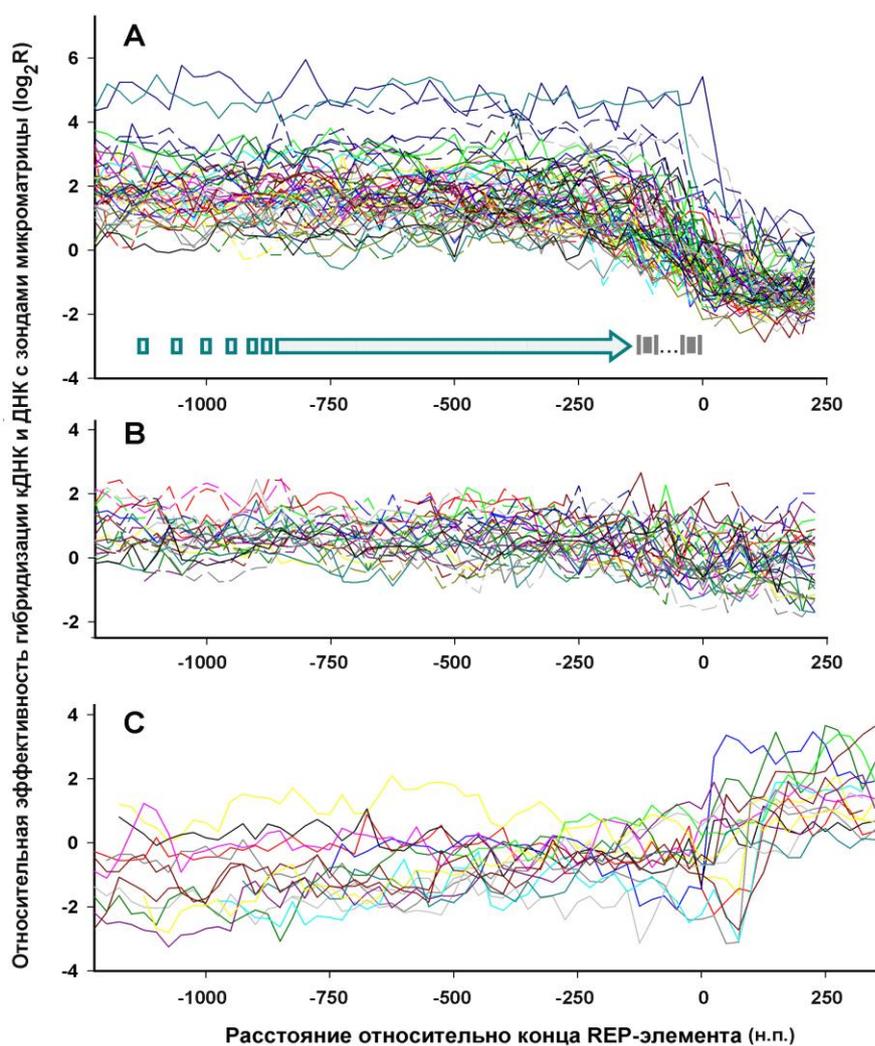


Рис. 4. Суперпозиция профилей транскрипции генов, расположенных перед REP-элементами, и прилегающих к ним участков генома (см. схему на панели А) по данным эксперимента «А» работы [17]. Все траектории выровнены по концу REP-элемента и сгруппированы на панелях А, В и С в зависимости от типа профиля. Для генов короче 1250 н. п. приведены нормированные сигналы гибридизации с зондов, расположенных между началом гена и концом REP-элемента. Для более длинных генов показаны профили вдоль последних 1250 н. п.

Всего было исследовано 558 участков генома (408 конвергентных и 150 коллинеарных генов), фланкированных REP-элементами. Во многих случаях наблюдался низкий уровень транскрипции рассматриваемого гена ($\Sigma E < 0$) и фланкирующей его области. Это уменьшало достоверность регистрации опосредованных REP-элементом изменений. Поэтому эти профили были выделены в отдельную группу и далее не анализировались. Остальные были разделены на три группы. Первый набор был составлен из транскрибируемых участков, уровень гибридизационного сигнала для которых уменьшался при прохождении REP-модулей (рис. 4,А). Их оказалось 80, причём 21 траектория с REP-ассоциированным угнетением относилась к области, расположенной между коллинеарными генами. Длина REP-элементов в этой группе варьировала от 15 до 281 н. п. Только в одном случае было зарегистрировано угнетение транскрипции на обоих концах REP-модуля, транскрибируемого по обоим нитям ДНК. Он находится между конвергентными генами *metE* и *usgA*. Межгенная область имеет длину всего 41 н. п., а размер находящегося в ней REP-элемента 17 н. п. Он (ниже показан буквами верхнего регистра) имеет некоторую гомологию с REP_{con} (рис. 1), окружен с обоих концов А(Т)-треками и образует идеальную шпильку со свободной энергией фолдинга -18.6 кКал/моль: *aaaatccaaACCGGGTGGTAATAACCAcccggtctttt* (комплементарные участки подчёркнуты). То есть этот REP-элемент входит в состав классического ρ -независимого терминатора транскрипции с двусторонней направленностью. И хотя большое число профилей с подавленной транскрипцией на концах REP-элементов (рис. 4,А), безусловно, свидетельствуют о причастности REP-модулей к остановкам или торможению транскрипции, сам механизм терминации во многих случаях понятен.

Во вторую группу были отобраны 34 профиля, не проявляющие значительной зависимости от REP-элемента (рис. 4,В). Двадцать из них приходятся на области между конвергентными генами и в одном случае (между *ucbL* и *aspC*) стабильная транскрипция была зарегистрирована в обоих направлениях. Это значит, что REP-элементы без терминаторного потенциала могут входить в состав новых РНК. Длина таких повторов варьировала в диапазоне от 21 до 388 н. п., а размер межгенных участков – в диапазоне от 42 до 1238 н. п., что во многих случаях достаточно для кодирования нового продукта.

Самой удивительной оказалась третья группа (рис. 4,С). В неё вошли 16 треков, гибридизационный сигнал в которых увеличивался при прохождении REP-модулей длиной 15–388 н. п. в межгенных участках размером от 56 до 1024 н. п. Такие профили проще всего объяснить наличием перекрывающихся промоторов, которые экспериментально картированы в 7 из 9 геномных участков, находящихся между коллинеарными генами (для генов *fhuA*, *uof*, *sucA*, *poxB*, *phoP*, *tyrB* и *uxuR*) [13]. Для двух остальных генов (*yegP* и *yhaH*) они предсказываются PlatProm [16]. В 8 случаях REP-элементы оказались расположенными слева от стартовой точки транскрипции на расстоянии 60–101 н. п., и только в одном случае этот модуль лежит в ранней транскрибируемой области, начиная с позиции + 25.

Между конвергентными генами экспериментально картированных промоторов нет, но в трёх межгенных участках промоторы предсказаны компьютерным методом. Обнаруженный рост гибридизационных сигналов (рис. 4,С), следовательно, почти во всех случаях можно объяснить наличием рядом расположенных промоторов, и поэтому возник вопрос о возможном влиянии REP-элементов на их активность, которое до сих пор не исследовалась.

Позиционирование REP-элементов относительно известных промоторов не случайно

Для анализа было использовано три набора промоторных последовательностей (рис. 5 и 6). Один из них содержал 404 известных промоторов, активируемых σ^{70} -РНК-полимеразой. Для того чтобы облегчить дальнейшую интерпретацию, мы отобрали в него только «одиночные» промоторы, аналогичные P_{dapB} (рис. 6,А). Нужно, однако, иметь в виду, что новые промоторы в этих областях могут быть обнаружены позднее. Во второй набор была включена 51 промоторная область дивергентных генов, таких, например, как P_{sapA} и P_{an} на рис. 1,А. Эти области включали от 2 до 6 промоторов с разной ориентацией.

Вначале искали аннотированные REP-элементы, левая граница которых попадает в ± 300 н. п. область вокруг стартов транскрипции «одиночных» промоторов. Их оказалось 28 (красные символы на рис. 5,В). Исходя из того, что все REP-элементы находятся в межгенных пространствах и 204 из 356 расположены между конвергентными генами (рис. 3,С), только 152 REP-последовательности могут оказаться в промоторной области. В последней версии генома *E. coli* аннотировано 4318 генов, которые экспрессируются в виде 2628 оперонов. Из них 658 транскрибируются в противоположных направлениях из общей регуляторной области. Это значит, что 152 REP-модуля распределены между 1970 (2628 – 658) промоторными областями (в среднем один REP-элемент на 13 промоторных областей). Ограничение по длине анализируемой области (± 300 н. п.), следовательно, не изменило существенно ожидаемую пропорцию (один REP-элемент на 14 промоторов), а саму выборку промоторов можно считать репрезентативной.

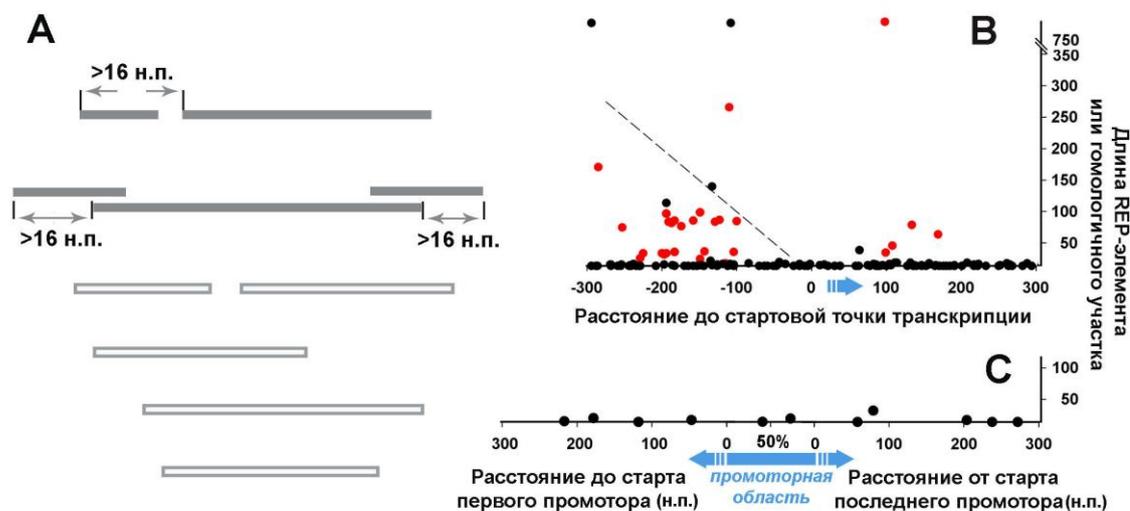


Рис. 5. Распределение известных REP-элементов и гомологичных им последовательностей (красные и чёрные символы, соответственно) относительно стартовых точек транскрипции 404 одиночных (В) и 51 множественных дивергентных промоторов (С). Показаны позиции 5'-концов REP-элементов или их гомологов. Пунктирной линией на В отделены мотивы, не перекрывающиеся с промоторами (слева) или затрагивающие их стартовые точки транскрипции (справа до позиции 0). Голубая стрелка показывает направление транскрипции. На рисунке С голубой двусторонней стрелкой показана область, занятая дивергентными промоторами (средний размер 127 н. п.). Так как длина участка со множественными промоторами отличалась у разных генов, позиционирование гомологов в ней осуществляли в процентах от левой границы. На рисунке А схематически показана стратегия отбора последовательностей, гомологичных REP-модулям. Серые прямоугольники соответствуют последовательностям, учитываемым как независимые гомологи, светлые - перекрывающимся и, поэтому, не учитываемым последовательностям.

Как и предполагалось, исходя из анализа данных рис. 4,С, большая часть известных REP-элементов оказалась расположенной слева от промоторов (красные символы на рис. 5,В). И только один из них с 5'-концом в позиции -110, имея длину 256 н. п., перекрывается с промотором и стартовой точкой транскрипции. В транскрибируемую область попали только 5 из аннотированных REP-элементов. Поиск в исследуемом наборе последовательностей, гомологичных разным REP-элементам, выявил несколько дополнительных REP-подобных участков, включая 2 инсерционных элемента длиной 767 н. п., содержащих REP-модуль (чёрные точки на рис. 5,В), но не изменил эту диспропорцию. Однако, короткие гомологи длиной 14–15 н. п. в равной степени обнаруживаются с обеих сторон от стартовой точки.

Анализ 51 промоторной области, содержащей дивергентные промоторы, с обеих сторон фланкированные транскрибируемыми участками, подтвердил их обеднённость REP-повторами (рис. 5,С). Известных REP-элементов в них не оказалось, а поиск похожих последовательностей позволил выявить всего несколько коротких гомологов. Локализация REP-модулей в промоторной зоне, следовательно, не случайна, с явным их преобладанием в нетранскрибируемом участке, расположенном левее стартовой точки транскрипции

Локализация REP-элементов относительно промоторных островков отличается от их расположения относительно известных промоторов

Промоторные островки в определённом смысле являются антиподами одиночных промоторов и в некотором смысле аналогами промоторных областей с дивергентной транскрипцией. Они содержат огромное число потенциальных точек старта для синтеза РНК (см. пример на рис. 6,А), но в большинстве случаев ведут только ограниченный синтез коротких олигонуклеотидов [14, 16, 23].

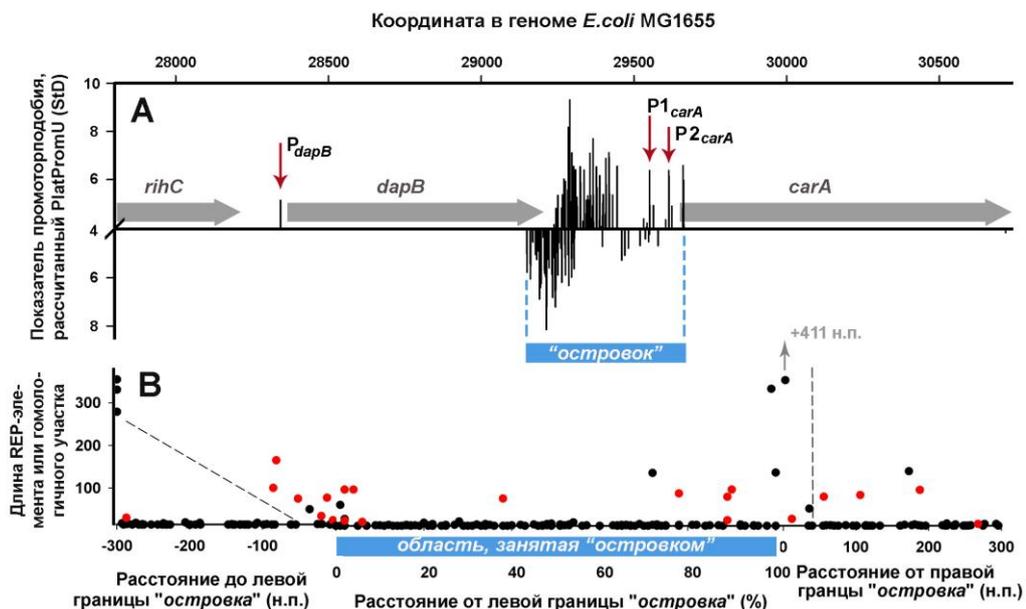


Рис. 6. А. Пример смешанного промоторного островка, выявленного алгоритмом PlatPromU [14]. Красными стрелками указаны стартовые точки одиночного (P_{dapB}) и tandemных ($P1_{carA}$ и $P2_{carA}$) промоторов. В. Распределение известных REP-элементов и их гомологов относительно 434 смешанных островков. Красными и чёрными символами, соответственно, показаны позиции их 5'-концов. Пунктирными линиями отделены мотивы, перекрывающиеся с островковыми промоторами. Голубым прямоугольником схематически показана область, занятая островком (средний размер 660 н. п.). Так как островки отличаются по длине, позиционирование гомологичных последовательностей в них осуществляли в процентах от левой границы. Вертикальной стрелкой отмечена позиция инсерционного элемента длиной 767 н. п., аналогичного трём таким же элементам, показанным на рис. 4.

Низкая продуктивность *островков* предполагает наличие специальных механизмов, подавляющих способность их промоторов инициировать транскрипцию. Некоторые из этих механизмов уже установлены (избыточная позитивная суперскрученность двойной спирали [14, 23] и способность связываться с гистон-подобным белком бактериального нуклеоида H-NS [24]). Однако причастность REP-элементов к терминации транскрипции (рис. 4) и особый характер их ассоциации с промоторами (рис. 5) послужили основанием для более детального исследования взаимного расположения этих элементов и *островков* (рис. 6). В результате было выявлено более 20 REP-элементов или их протяжённых гомологов, находящихся в непосредственной близости от *островков*, но в отличие от обычных промоторов, почти все они находятся в зоне, перекрывающейся с *островковыми* промоторами (на рис. 6 обозначена пунктирными линиями). Теоретически они могли бы оказывать такое же негативное воздействие на процессивность транскрипции, как это показано на рис. 4,А, и вносить независимый вклад в подавленный синтез РНК-продуктов, начавшийся на *островковых* промоторах.

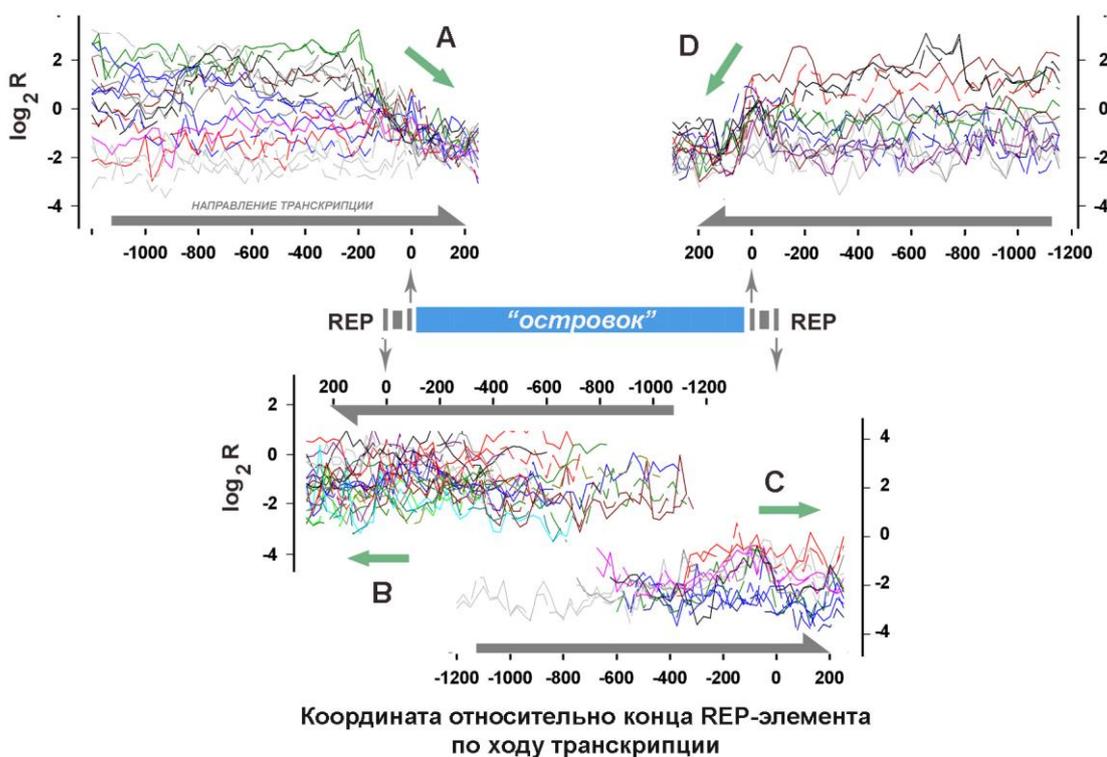


Рис. 7. Суперпозиция профилей транскрипции (эксперименты А и В в работе [17]) обеих нитей ДНК вдоль *промоторных островков* (В и С) и прилегающих к ним участков генома (А и D). Все траектории выровнены по дальним по ходу транскрипции границам REP-элементов (позиции 0). Профили гибридационных сигналов на панелях А и D одинаковы по длине, так как захватывают стандартные по размеру участки, фланкирующие REP-элемент с левой (А) или с правой (D) стороны. Профили гибридационных сигналов на панелях В и С отличаются по размеру, так как охватывают переменную область между границей REP-элемента и правой (В) или левой (С) границей *островка*. Зелёные стрелки акцентируют наличие (А и D) или отсутствие (В и С) изменений в уровне транскрипции при прохождении REP-элемента.

Но прохождение REP-элементов при транскрипции, направленной из *островков* не влияло на уровень гибридационных сигналов (рис. 7,В и 7,С). Зато прохождение этих же REP-элементов при транскрипции, направленной внутрь *островков* сопровождалось явным уменьшением регистрируемого продукта (рис. 7,А и 7,Д). То есть, REP-элементы, ассоциированные с *промоторными островками*, способствуют их изоляции от транскрипции, инициированной за пределами *островка*.

ЗАКЛЮЧЕНИЕ

Несмотря на более чем тридцатилетнюю историю изучения REP-элементов, практически все аспекты их возникновения, внутригеномной экспансии и функционирования остаются малопонятными. Позиционный и функциональный анализ этих элементов, выполненный в данной работе, подтвердил традиционную точку зрения о причастности REP-повторов к терминации транскрипции (рис. 4,А). Однако обнаружение REP-сигналов, игнорируемых транскрипционным аппаратом и таких модулей, синтез РНК после которых увеличивается, а не уменьшается, свидетельствует о том, что терминаторная функция REP-модулей может быть лишь частными случаями их функционального репертуара. Она может быть обусловлена использованием шпилечных структур REP-последовательностей в составе обычных терминаторов и не отражать истинную сущность этих элементов, если таковая имеется. Тем не менее, исходя из типичного для REP-элементов расположения в терминаторной области генов (рис. 2, 3,В и 3,С) в клетках должно быть много разных РНК, имеющих на 3'-концах REP-подобные модули. Не исключено, что такие РНК подвержены особому REP-опосредованному процессингу в зависимости от общего состояния транскриптома.

Регистрация сквозной транскрипции через REP-элементы (рис. 4,В) означает возможность их включения в синтезируемые РНК-продукты и, следовательно, способность формировать регуляторные дуплексы с другими РНК, содержащими аналогичные повторы. Такой РНК, в частности, может быть *SapZ*, транскрибируемая в межгеномном участке *umjA/sapA* в антисмысловом направлении к гену *umjA* (рис. 1,А). Очевидно, что масштабность такого типа регуляторных воздействий определяется числом потенциальных партнёров, а их, по-видимому, немного, так как основная часть REP-элементов находится в области терминаторов или прилегает к промоторам (рис. 3), но в нетранскрибируемой зоне (рис. 5). Тем не менее, функционирование REP-элементов в составе нетранслируемых РНК как основы для РНК-РНК и РНК-белкового узнавания может иметь особое значение. В настоящее время уже известно три малых некодирующих РНК, содержащих REP-элементы (*SroC*, C0362, C0664). Наиболее изучена из них *SroC* [25, 26], которая может связываться с малой РНК *GcvB*, являющейся глобальным посттранскрипционным регулятором многих мРНК, в том числе, переносчиков аминокислот и коротких пептидов, включая АВС-транспортеры (*Opp*, *Dpp*), и основных факторов транскрипции, таких как *Lrp*, *PhoP* и *CsgD*. Установлено, что REP-элемент *SroC* является непосредственным сайтом связывания для *GcvB* и это взаимодействие приводит к направленной деградации *GcvB*, при участии РНКазы E. Поэтому взаимодействие REP-мотивов из разных транскриптов с ключевыми регуляторными РНК может быть основой их активного участия в межмолекулярных коммуникациях.

Самой резонансной частью данной работы может стать неожиданный характер влияния REP-элементов на транскрипцию вблизи *промоторных островков*. Учитывая уже известный факт их низкой транскрипционной активности, наличие элементов с потенциальной терминаторной функцией на границах *островков* было вполне прогнозируемым. Но, вопреки ожидаемому подавлению выходящей с *островков* транскрипции, мы обнаружили ассоциированное с REP-элементами снижение синтеза РНК, направленного к *островкам* с соседних участков. Терминаторная функция шпилечных структур закономерно зависит от того с какой стороны от них имеется тормозящий гомонуклеотидный трек. Поэтому односторонность супрессорного эффекта сама по себе тоже не удивительна, но тот факт, что расположенные рядом с *островками* REP-элементы ориентированы таким образом, чтобы изолировать *островки* от внешней транскрипции, является симптоматичным. Он может свидетельствовать о целесообразности поддержания их автономного транскрипционного статуса.

Исследование выполнено за счёт гранта Российского научного фонда (проект №14-14-00985).

СПИСОК ЛИТЕРАТУРЫ

1. Stern M.J., Ames G.F.L., Smith N.H., Robinson E.C., Higgins C.F. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell*. 1984. V. 37. P. 1015–1026. doi: [10.1016/0092-8674\(84\)90436-7](https://doi.org/10.1016/0092-8674(84)90436-7).
2. Higgins C.F., McLaren R.S., Newbury S.F. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? *J. Gene*. 1988. V. 72. P. 3–14. doi: [10.1016/0378-1119\(88\)90122-9](https://doi.org/10.1016/0378-1119(88)90122-9).
3. Stern M.J., Prossnitz E., Ames G.F.L. Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium*: involvement of REP sequences. *Mol. Microbiol.* 1988. V. 2. P. 141–152. doi: [10.1111/j.1365-2958.1988.tb00015.x](https://doi.org/10.1111/j.1365-2958.1988.tb00015.x).
4. Espeli O., Moulin L., Bocard F. Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *Mol. Biol.* 2001. V. 314. P. 375–386. doi: [10.1006%2Fjmbi.2001.5150](https://doi.org/10.1006%2Fjmbi.2001.5150).
5. Merino E., Becerril B., Valle F., Bolivar F. Deletion of a repetitive extragenic palindromic (REP) sequence downstream from the structural gene of *Escherichia coli* glutamate dehydrogenase affects the stability of its mRNA. *Gene*. 1987. V. 58. P. 305–309. doi: [10.1016/0378-1119\(87\)90386-6](https://doi.org/10.1016/0378-1119(87)90386-6).
6. Newbury S.F., Smith N.H., Robinson E.C., Hiles I.D., Higgins C.F. Stabilization of translationally active mRNA by prokaryotic REP sequences. *Cell*. 1987. V. 48. P. 297–310. doi: [10.1016/0092-8674\(87\)90433-8](https://doi.org/10.1016/0092-8674(87)90433-8).
7. Bachellier S., Clement J.M., Hofnung M. Short palindromic repetitive DNA elements in enterobacteria: a survey. *J. Res. Microbiol.* 1999. V. 150. P. 627–639. doi: [10.1016/S0923-2508\(99\)00128-X](https://doi.org/10.1016/S0923-2508(99)00128-X).
8. Khemici V. and Carpousis A.J. The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required *in vivo* for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol.* 2004. V. 51. P. 777–790. doi: [10.1046/j.1365-2958.2003.03862.x](https://doi.org/10.1046/j.1365-2958.2003.03862.x).
9. Ton-Hoang B., Siguier P., Quentin Y., Onillon S., Marty B., Fichant G., Chandler M. Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *J. Nucleic Acids Res.* 2012. V. 40. P. 3596–3609. doi: [10.1093/nar/gkr1198](https://doi.org/10.1093/nar/gkr1198).
10. Moulin L., Rahmouni A.R., Bocard F. Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of *Escherichia coli*. *Mol Microbiol.* 2005. V. 55. P. 601–610. doi: [10.1111/j.1365-2958.2004.04411.x](https://doi.org/10.1111/j.1365-2958.2004.04411.x).
11. Messing S.A., Ton-Hoang B., Hickman A.B., McCubbin A.J., Peaslee G.F., Ghirlando R., Chandler M., Dyda F. The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease. *J. Nucleic Acids Res.* 2012. V. 40. P. 9964–9979. doi: [10.1093/nar/gks741](https://doi.org/10.1093/nar/gks741).
12. Di Nocera P.P., De Gregorio E., Rocco F. GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes. *J. BMC Genomics.* 2013. V. 14. P. 522. doi: [10.1186/1471-2164-14-522](https://doi.org/10.1186/1471-2164-14-522).
13. Salgado H., Peralta-Gil M., Gama-Castro S., Santos-Zavaleta A., Muñoz-Rascado L., García-Sotelo J.S., Weiss V., Solano-Lira H., Martínez-Flores I., Medina-Rivera A. et al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* 2013. V. 41. P. D203. doi: [10.1093/nar/gks1201](https://doi.org/10.1093/nar/gks1201).
14. Панюков В.В., Киселев С.С., Шавкунов К.С., Масулис И.С., Озолинь О.Н. Мультиспецифичные промоторные островки как участки генома с необычными

- структурными и функциональными свойствами. *Математическая биология и биоинформатика*. 2013. Т. 8. № 2. С. 432–448. doi: [10.17537/2013.8.432](https://doi.org/10.17537/2013.8.432).
15. Киселев С.С., Озолин О.Н. Структурообразующие модули как индикаторы промоторной ДНК в бактериальных геномах. *Математическая биология и биоинформатика*. 2011. Т. 6. № 1. С. 39–52. doi: [10.17537/2011.6.39](https://doi.org/10.17537/2011.6.39).
 16. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons from genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4919–4931. doi: [10.1093/nar/gkp490](https://doi.org/10.1093/nar/gkp490).
 17. Reppas N.B., Wade J.T., Church G.M., Struhl K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell*. 2006. V. 24. P. 747–757. doi: [10.1016/j.molcel.2006.10.030](https://doi.org/10.1016/j.molcel.2006.10.030).
 18. NCBI Microbial Nucleotide BLAST. URL: http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=MicrobialGenomes (дата обращения: 07.06.2015).
 19. Igarashi K., A. Ishihama. Bipartite functional map of the *E. coli* RNA polymerase α subunit: involvement of the C-Terminal region in transcription activation by CAMP-CRP. *Cell*. 1991. V. 65. P. 1015–1022. doi: [10.1016/0092-8674\(91\)90553-B](https://doi.org/10.1016/0092-8674(91)90553-B).
 20. Mathews D.H. RNA secondary structure analysis using RNAstructure. *Current Protocols in Bioinformatics*. 2014. V. 46. P. 12.6.1–12.6.25. doi: [10.1002/0471250953.bi1206s46](https://doi.org/10.1002/0471250953.bi1206s46).
 21. Liang W., Rudd K.E., Deutscher M.P. A Role for REP sequences in regulating translation. *Mol. Cell*. 2015. V. 58. № 3. P. 431–439. doi: [10.1016/j.molcel.2015.03.019](https://doi.org/10.1016/j.molcel.2015.03.019).
 22. Brok-Volchanski A.S., Masulis I.S., Shavkunov K.S., Lukyanov V.I., Purtov Yu.A., Kostyanicina E.G., Deev A.A., Ozoline O.N. Predicting sRNA genes in the genome of *E. coli* by the promoter-search algorithm PlatProm. In: *Bioinformatics of Genome Regulation and Structure II*. Eds. Kolchanov N., Hofestaedt R. Springer, 2005. P. 11–20. doi: [10.1007/0-387-29455-4_2](https://doi.org/10.1007/0-387-29455-4_2).
 23. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE*. 2013. V. 8. P. e62601. doi: [10.1371/journal.pone.0062601](https://doi.org/10.1371/journal.pone.0062601).
 24. Purtov Yu.A., Glazunova O.A., Antipov S.S., Pokusaeva V.O., Fesenko E.E., Preobrazhenskaya V.V., Shavkunov K.S., Tutukina M.N., Lukyanov V.I., Ozoline O.N. Promoter islands as a platform for interaction with nucleoid proteins and transcription factors. *J. Bioinformatics and Computational Biology*. 2014. V. 12. № 2. P. 322–331. doi: [10.1142/S0219720014410066](https://doi.org/10.1142/S0219720014410066).
 25. Vogel J., Bartels V., Tang T.H., Churakov G., Slagter-Jäger J.G., Hüttenhofer A., Wagner E.G. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* 2003. V. 31. P. 6435–6443. doi: [10.1093/nar/gkg867](https://doi.org/10.1093/nar/gkg867).
 26. Miyakoshi M., Chao Y., Vogel J. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J.* 2015. V. 34. P. 1478–1492. doi: [10.15252/embj.201490546](https://doi.org/10.15252/embj.201490546).

Материал поступил в редакцию 16.06.2015, опубликован 24.06.2015