

УДК: 579:252

## Мультиспецифичные промоторные островки как участки генома с необычными структурными и функциональными свойствами

Панюков В.В.<sup>1</sup>, Киселев С.С.<sup>2,3</sup>, Шавкунов К.С.<sup>2,3</sup>,  
Масулис И.С.<sup>2,3</sup>, Озолин О.Н.<sup>2,3\*</sup>

<sup>1</sup>Институт математических проблем биологии, Российская академия наук,  
Пушино, Московская область, 142290, Россия

<sup>2</sup>Институт биофизики клетки, Российская академия наук, Пушино, Московская  
область, 142290, Россия

<sup>3</sup>Пуцинский государственный естественно-научный институт, Пушино,  
Московская область, 142290, Россия

**Аннотация.** В работе исследуются необычные участки генома кишечной палочки – мультиспецифичные промоторные островки - имеющие высокую плотность потенциальных точек инициации транскрипции. Они предсказаны алгоритмом PlatPromU без учёта консервативных элементов, распознаваемых  $\sigma$ -субъединицами РНК-полимеразы, и поэтому должны взаимодействовать с полимеразой, в состав которой входят разные  $\sigma$ -факторы. Было установлено, что по структурным и функциональным свойствам эти *островки* очень похожи на *промоторные островки*, обнаруженные ранее алгоритмом PlatProm, который был адаптирован к поиску промоторов основного сигма-фактора кишечной палочки –  $\sigma^{70}$ . Так, подобно *промоторным островкам*, большинство *мультиспецифичных островков* находится в регуляторной области генов, приобретённых *E. coli* в результате горизонтального переноса. Для двойной спирали *островков* характерны большая степень изогнутости и закрученности, чем для обычных промоторов. Оба типа *островков* лучше, чем обычные промоторы, формируют комплексы с РНК-полимеразой, но хуже их начинают синтез полноразмерных мРНК. В значительной степени эта супрессия объясняется повышенной способностью *островков* взаимодействовать с гистоноподобным ингибиторным белком H-NS. Однако синтез РНК с *промоторного островка*, ассоциированного с геном *arrY*, возрастал при его переносе в плазмиду pET28b-eGFP и оказался зависимым не только от наличия мест связывания H-NS, но и от пространственной конфигурации соседних участков. Это значит, что конформационные изменения генома могут увеличить транскрипционную активность *островков*, предоставив продукты рядом лежащих генов для метаболизма бактериальной клетки.

**Ключевые слова:** промоторные островки, структура ДНК, инициация транскрипции, H-NS, горизонтальный перенос генов, бактериальная эволюция.

### ВВЕДЕНИЕ

Экспрессия бактериальных генов осуществляется ферментом РНК-полимеразой, оснащённой сменными  $\sigma$ -факторами (RNAP), и контролируется регуляторными белками, которые кооперируют с ней в инициации транскрипции, взаимодействуя со

\*ozoline@rambler.ru

специальными участками генома – промоторами. Большинство имеющихся в геноме промоторов расположено перед генами, но и внутри кодирующих последовательностей генов часто обнаруживаются сигналы инициации транскрипции, которые могут контролировать синтез антисмысловых или альтернативных РНК [1–4]. Биологическая роль таких РНК не всегда ясна, но их синтез из внутригенных областей подтверждён современными методами, картирующими точки инициации транскрипции (TSP) непосредственно в живой клетке [5–7]. Кроме указанных двух типов промоторов с одной или несколькими стартовыми позициями, все исследованные нами бактериальные геномы содержат необычные участки с аномально высокой плотностью потенциальных точек инициации транскрипции [1, 8, 9]. Так, например, наша программа PlatProm, предназначенная для поиска  $\sigma^{70}$ -зависимых промоторов в геноме *E. coli*, обнаружила 78 участков длиной от 300 н.п. и выше, у которых любое скользящее окно размером 100 н.п. содержало не менее 8 стартовых точек транскрипции. Такие участки были названы нами «*промоторными островками*» (PIs) [1, 9]. Анализ опубликованных данных полногеномного картирования мест связывания RNAP [10, 11] и мест инициации транскрипции [6] показал, что все PIs взаимодействуют с РНК-полимеразой [1] и иницируют синтез коротких abortивных РНК [9]. Т.е. они способны реализовать первую стадию инициации транскрипции, но синтез полноразмерных РНК прерывается и РНК-полимераза не освобождает промотор. Кроме этого, структурный анализ трёхмерной геометрии, выполненный с помощью программного пакета aSHAPE [12], выявил повышенную степень изогнутости *промоторных островков* по сравнению с обычными промоторами, которые формировали анизотропные изгибы двойной спирали ДНК чаще, чем фрагменты непромоторной ДНК [9]. Эта избыточная кривизна может стабилизировать RNAP-промоторные комплексы в *островках*, удерживая их в состоянии синтеза abortивных РНК.

Было также обнаружено, что обычные промоторы не отличаются от непромоторной ДНК по усреднённому твистовому углу, но величина этого параметра для PIs оказалась существенно больше. Это, по-видимому, означает, что в ходе эволюции PIs оказались приспособленными к выполнению какой-то дополнительной функции, кроме взаимодействия с RNAP. Руководствуясь этим предположением, мы проанализировали доступные профили распределения белков бактериального нуклеоида вдоль всей хромосомы [13–17], и обнаружили, что PIs являются мишенями предпочтительного взаимодействия с H-NS [9]. Хорошо известно, что H-NS является одним из основных белков, конденсирующих бактериальный геном. При этом он специфически подавляет экспрессию горизонтально приобретённых генов, связываясь с их регуляторными участками. Поэтому мы исследовали взаимное расположение *промоторных островков* и чужеродной ДНК и обнаружили, что 75 из 78 PIs (96%) находятся рядом с горизонтально приобретёнными генами. Однако ожидаемое число чужих генов у кишечной палочки на порядок больше, чем число *промоторных островков*, обнаруженных PlatProm. Так, Lawrence et al. [18] нашли в её геноме 706 чужих генов, а по оценке Nakamura et al. [19] и Price et al. [20] их больше тысячи. Но если высокая плотность потенциальных промоторов действительно является характерным свойством горизонтально приобретённых генов, то в их число, наряду с промоторами  $\sigma^{70}$ -RNAP, вполне могут входить промоторы других  $\sigma$ -факторов. Для того, чтобы определить, насколько учёт дополнительных сигналов транскрипции сократит разницу между числом потенциально чужих генов и количеством *промоторных островков*, геном *E. coli* был просканирован унифицированным алгоритмом поиска промоторов PlatPromU [21], который игнорирует весовые матрицы, учитывающие присутствие  $\sigma$ -специфических модулей, и поэтому может найти промоторы, распознаваемые всеми семью  $\sigma$ -субъединицами РНК-полимеразы *E. coli*. Если это так, то различие в количестве *островков* и чужих генов может сократиться. В данной работе мы

обнаружили 434 *мультиспецифичных промоторных островка* (MPIs), оценили степень их ассоциации с чужими генами, а также охарактеризовали их структурные и функциональные свойства.

## МЕТОДЫ

### Геном *E. coli*, промоторы и промоторные островки

Нуклеотидная последовательность генома *E. coli* K12 MG1655 и координаты генов были взяты из NCBI GenBank (NC\_000913). Для поиска участков ДНК, удовлетворяющих критерию структурно-модульного сходства с промоторами различной специфичности, использовалась программа PlatPromU [21], которая просматривала обе нити ДНК и вычисляла степень соответствия каждого нуклеотида потенциальному старту транскрипции, что численно выражалось как *вес*. Нуклеотид рассматривался как предсказанный старт транскрипции, если его *вес* превышал пороговый уровень на величину, равную четырём стандартным отклонениям (StD); порог же оценивали по геномному распределению *весов* так, как это описано в [21]. Критерии отбора MPIs были такими же, как это было предложено ранее для PIs (см. выше), но минимальная длина *островка* была увеличена до 360 н.п., что обеспечило присутствие всех PIs в новом наборе, содержащем 434 MPIs. Для сравнительного анализа структурно-функциональных свойств MPIs была использована выборка (всего 351 геномная область) биохимически охарактеризованных регуляторных участков хромосомы *E. coli*, которые содержат один или несколько известных промоторов, не перекрываются с MPIs и, по крайней мере, один из промоторов распознаётся  $\sigma^{70}$ -RNAP. Координаты точек инициации транскрипции этих промоторов были взяты из RegulonDB [7].

### Связь MPIs и обычных промоторов с горизонтально перенесёнными генами

Было исследовано взаимное расположение MPIs и обычных промоторов относительно чужих генов, предсказанных пятью исследовательскими группами [18–20, 22, 23]. Объекты считались ассоциированными с чужими генами, если они располагались внутри длинных «*геномных островков*», предсказанных GIST [22] или IslandViever [23], а также внутри регуляторных областей чужих генов [19–21]. Кроме этого, в группу ассоциированных объектов отбирались такие MPIs, которые целиком лежали в кодирующей области чужого гена, либо перекрывались с ней не менее чем на 100 н.п.

### Анализ данных, полученных методом специфической иммунопреципитации хроматина

Для оценки способности RNAP и H-NS взаимодействовать с MPIs были использованы опубликованные данные [10, 13–15], полученные методом специфической иммунопреципитации ДНК-белковых комплексов (ChIP-on-chip). Этот метод позволяет обнаружить места связывания конкретного белка во всём бактериальном геноме в живой клетке. Для этого его комплексы с ДНК с помощью специфических антител извлекаются из выделенного из клеток и фрагментированного ультразвуком хроматина (ChIP), а экстрагированная из них ДНК гибридизуется с микроматрицами (chip). В случае RNAP эффективность определялась как  $\log_2$  величины отношения гибридизационных сигналов на микроматрице, полученных с ДНК после специфической иммунопреципитации комплекса, и сигналов, полученных с контрольной ДНК, не подвергнутой специфической иммунопреципитации (эксперимент «В» в [10]). Геномный участок считали взаимодействующим с РНК-полимеразой, если среднее значение  $\log_2$  для всех проб внутри MPIs или вблизи

( $\pm 150$  н.п.) TSP нормального промотора было положительным. Для оценки эффективности взаимодействия с H-NS были использованы данные, опубликованные Kahramanoglou et al. [13] и Grainger et al. [14, 15]. В первом случае анализируемые геномные области считали мишенями для взаимодействия с H-NS, если они перекрывались с опубликованными сайтами связывания по крайней мере на 20 н.п. Во втором случае – если они содержали хотя бы один зонд с величиной отношения гибридизационных сигналов  $\geq 1.5$ .

### Дифференциальный анализ экспрессии

Транскрипционная активность MPIs и обычных промоторов оценивалась тем же способом, что и в [9], который заключается в следующем. На вход программы RNAMatcher подавался массив из 8967903 последовательностей клеточных РНК, прочитанных с 5'-конца (дополнение в [6]). Каждая из этих последовательностей (образцов) имела длину 44 нуклеотида. Программа находила позиции, соответствующие каждому образцу в геноме, и подсчитывала общее число одинаковых образцов в массиве. При полном совпадении с геномом (44 нуклеотида) образец рассматривался как результат продуктивной транскрипции. Если же обнаруживалось совпадение только по 10-ти или 11-ти 5'-концевым нуклеотидам и образец имел адаптерную последовательность на 3'-конце, его считали результатом abortивного синтеза.

### Структурный анализ

Сравнение пространственной структуры MPIs со структурой обычных промоторов (фрагментов ДНК размером 300 н.п.) осуществлялось с помощью программного пакета aSHAPE [9, 12] в рамках трёхмерной модели ДНК, предоставляемой сервером *DNA Tools* ([http://hydra.icgeb.trieste.it/dna/model\\_it.html](http://hydra.icgeb.trieste.it/dna/model_it.html)) [24]. Пакет вычислял структурные параметры фрагментов двойной спирали ДНК, используя два вида конформационных цепочек: *углеродные* и *фосфорные*. Вершинами *углеродной* цепочки были геометрические центры отрезков, соединяющих C<sub>8</sub> пуринов с C<sub>6</sub> пиримидинов в каждой паре оснований анализируемого фрагмента. Вершинами *фосфорной* цепочки были центры сегментов, соединяющих атомы фосфора комплементарной пары (подробнее см. [9, 12]). Свободную энергию стэкинга для фрагментов ДНК вычисляли как сумму энергий стэкинга, рассчитанную для входящих в эти фрагменты динуклеотидов. Числовые значения этого параметра были взяты из таблицы DiProDB (<http://diprodb.fli-leibniz.de/ShowTable.php>) [25, 26].

### Функциональный анализ промоторного островка, ассоциированного с геном *appY*

Три фрагмента ДНК, содержащие различные части *промоторного островка*, покрывающего область между генами *tfaX* и *appY*, а также кодирующую последовательность гена *appY* в геноме *E. coli* MG1655, амплифицировали с праймерами:

5'-GATAAGATCTGCAAGTAAAAATGATACTC-3' (F1),

5'-ATGCAGATCTTCCSTGATTATGATTGTG-3' (F3),

5'-TCCATCTAGAACCCSTATCATAAAATTA-3' (R1) и

5'-CCSTTCTAGATTTGTGCGSTTACAATAAA-3' (R2).

Полученные фрагменты ДНК расщепляли рестриктазами *Bgl*III (сайт рестрикции AGATCT) и *Xba*I (сайт рестрикции TCTAGA) и встраивали в плазмиду pET28b между сайтами *Bgl*III и *Xba*I в качестве промотора гена *gfp*, кодирующего зелёный флуоресцирующий белок. Все ферментативные реакции проводились с использованием ферментов фирмы Fermentas в соответствии с протоколом изготовителя.

Флуоресценцию GFP измеряли в колониях клеток *E. coli* OmniMAX, трансформированных полученной плазмидой и выращенных на стандартной агаризованной среде Luria–Bertani с добавлением 60 мкг/мл канамицина. В качестве контроля были использованы колонии клеток, трансформированных плазмидой с беспромоторным геном *gfp*. Уровень флуоресценции регистрировали на флуоресцентном микроскопе Leica (длины волн возбуждения/эмиссии 480/510 нм), и количественно оценивали программой ImageJ (<http://rsbweb.nih.gov/ij/>).

## РЕЗУЛЬТАТЫ

### PlatPromU обнаружил 434 мультиспецифичных промоторных островка в геноме *E. coli*

Программа PlatProm оценивает соответствие каждой позиции генома потенциальному старту транскрипции в  $\sigma^{70}$ -специфичных промоторах. Для определения их *веса* она оперирует позиционными весовыми матрицами, построенными для гексануклеотидов, взаимодействующих с  $\sigma^{70}$ , и использует более 50-ти каскадных матриц, отражающих специфику структурных модулей, способствующих адаптивной изомеризации промоторов в транскрипционных комплексах. Но эволюционный консерватизм аппарата транскрипции позволяет осуществлять поиск промоторов, узнаваемых и альтернативными  $\sigma$ -факторами, при помощи унифицированной версии программы (PlatPromU) [21], которая работает только с каскадными матрицами. С использованием порогового уровня для *веса* показателя, который обеспечивает достоверность  $p < 0.00004$  (определён, как описано в [21]), в геноме *E. coli* было найдено несколько тысяч потенциальных промоторных участков, включая такие, которые обнаруживаются PlatProm. Далее, используя те же критерии для поиска *промоторных островков*, которые описаны в статьях [1] и [9] (наличие  $\geq 8$  TSPs в каждом скользящем окне размером 100 н.п. на протяжении  $\geq 300$  н.п.), было найдено 600 *мультиспецифичных островков*. Все 78 ранее охарактеризованных PIs оказались в этом наборе, и их минимальный размер был 360 н.п. Это отличие от минимально приемлемого размера означает, что промоторы с разной  $\sigma$ -специфичностью могут формировать MPIs. Для того чтобы уменьшить размер этого набора до величины, сопоставимой с уже сформированным набором из 351 промоторного участка, мы увеличили пороговое значение для длины MPIs до 360 н.п. В результате был получен новый набор, содержащий 434 MPIs. Средний размер *островков* в этом множестве равен 661 н.п., тогда как самый длинный *островок* имеет протяжённость 7598 н.п. (таблица 1).

**Таблица 1.** Координаты мультиспецифичных промоторных островков в геноме *E. coli* K12 MG1655, их длина и связь с горизонтально приобретёнными генами

участки 1-109			участки 110-218			участки 219-327			участки 328-434		
координата	длина	источник данных *	координата	длина	источник данных *	координата	длина	источник данных *	координата	длина	источник данных *
17025	501	--n--	1278619	452	-----	2283999	462	-----	3530410	425	----p
29065	639	-----	1292968	425	--n1-	2301499	404	--n--	3537787	379	-----
41821	610	----p	1297308	672	-----	2310651	867	-----	3550522	598	-----
58659	756	--n1-	1298357	680	-----	2341466	522	g-n-p	3579578	1910	g-nlp
83820	563	g-n-	1307512	826	--n1-	2342045	736	g-n-p	3582164	742	g-n1-
85243	443	g---	1314018	515	--n1-	2362246	433	g---p	3595538	462	-----
89011	649	--n--	1332751	426	-----	2362997	370	g-nlp	3620924	1702	g-nlp
121653	405	----p	1341256	362	---lp	2363578	445	g-n-p	3629188	731	g-n1-
154703	620	gin1-	1358825	383	--n--	2380550	690	gin1-	3630044	1023	g-n1-
156811	519	g-n1-	1369587	403	--nlp	2381326	406	gin1-	3631295	994	g-n1-

233981	381	ginl-	1388697	372	---p	2383618	405	ginlp	3632356	734	g-nlp
237008	491	ginl-	1390680	598	--nl-	2384587	506	ginlp	3648820	598	g-nl-
238001	939	ginl-	1393485	497	--nlp	2385450	784	-inlp	3649513	733	g-nl-
252508	367	--nl-	1410853	372	--nl-	2386354	570	-inl-	3651226	827	g-nlp
252957	438	--nl-	1411493	681	-inl-	2403103	645	-----	3652495	539	g-nl-
259058	463	-----	1421610	582	-inl-	2404930	644	-----p	3653800	731	g-nlp
262055	441	-----	1422963	403	-inlp	2411059	362	--n-p	3654706	376	g-nlp
284280	377	-i-lp	1431570	1668	g-nl-	2453522	759	g----	3655630	1258	--nl-
291978	892	-inlp	1434606	767	g----	2460800	584	--nl-	3662512	421	--nl-
292911	1453	-inlp	1462958	512	g---p	2461780	627	-inl-	3663561	687	--nl-
296057	553	-in-p	1473307	403	--nlp	2466817	1773	ginlp	3667209	418	-----
310143	921	-in-p	1500175	528	--n--	2474418	389	--n--	3669849	380	--n--
312157	474	-i---	1518777	502	--n--	2478590	648	-inl-	3694146	368	--nl-
312947	678	-inlp	1524817	1053	g-nlp	2479567	521	-inl-	3694645	392	--n--
317665	361	-inlp	1527769	969	g-nlp	2480771	370	-inl-	3718242	504	--nl-
320379	534	--n-p	1529011	425	g-nl-	2481225	1105	-inl-	3735006	416	--n-p
324288	438	--nl-	1529586	478	g-nl-	2482830	438	-inl-	3743785	493	----p
330759	792	--n--	1541737	2674	g-nl-	2483388	999	-inl-	3749670	812	-----
343046	929	--nl-	1565114	468	----p	2488645	405	-inlp	3752332	496	----p
345477	560	--nl-	1570010	526	----p	2491166	563	---lp	3755772	430	g----
347632	374	-----	1577222	1414	ginlp	2492456	376	--nl-	3764123	1099	g-nlp
383775	742	--nlp	1580426	586	ginlp	2494555	496	-----	3765749	1000	g-nl-
389004	483	--nlp	1581411	908	ginl-	2522857	412	----p	3767002	1081	g----
393590	559	--nl-	1584482	418	g-nl-	2531270	420	----p	3787153	479	--n--
400274	418	--n--	1596053	540	--n--	2557957	572	-inl-	3787887	495	--n--
418266	507	----p	1613695	524	--nl-	2559897	383	-inl-	3790880	737	ginl-
450848	492	-----	1622416	502	--n--	2576296	415	-----	3794844	3642	ginlp
479168	1335	--nl-	1630172	1610	--nl-	2588895	523	-----	3798515	2750	ginlp
522049	488	g-nlp	1634966	1101	ginl-	2626549	428	g-n--	3801914	1517	ginlp
526294	897	g-nlp	1636533	609	gi-l-	2627521	404	g-nl-	3805844	756	ginlp
527695	809	g-nl-	1638669	1326	ginlp	2651326	565	-----	3834585	419	--n--
544455	397	----p	1640088	460	-----	2698323	384	-----	3841692	479	----p
545590	425	----p	1644594	386	-inl-	2755495	421	ginl-	3873193	420	--nl-
557105	399	-inlp	1650572	653	--nl-	2757618	1049	ginlp	3886238	583	-----
557872	419	-inlp	1671208	476	-----	2763165	415	ginlp	3903481	399	--n-p
562926	442	-inlp	1702736	415	-----	2771215	2017	ginl-	3904506	410	--n-p
567750	843	ginl-	1710105	803	-----	2781386	816	ginl-	3920483	735	--n--
569055	1132	ginl-	1752558	436	g-n-p	2782245	1079	ginl-	3965542	466	--n--
570511	709	ginl-	1753096	442	g---p	2783819	1101	ginl-	3983964	587	----p
576007	594	ginl-	1762391	516	-----	2785550	638	ginl-	4000416	1011	--nl-
578302	708	ginl-	1768241	504	--n-p	2786386	627	ginl-	4041929	549	--nl-
582252	1804	ginlp	1771642	382	----p	2796362	863	--n-p	4044492	484	-----
584711	518	ginl-	1776585	424	--nlp	2802496	503	-----	4060020	527	--n-p
585353	445	ginl-	1785089	419	-----	2882114	524	ginlp	4067116	514	----p
635732	457	--nl-	1789980	390	--n--	2884888	475	ginlp	4076478	908	--nl-
636502	592	--nl-	1800863	739	--n--	2885512	362	ginlp	4077482	374	--nlp
637951	427	-----	1802156	560	-----	2898106	794	g---p	4090814	387	--n-p
650912	719	----p	1810478	976	--nl-	2901483	610	-----	4158769	373	----p
655045	929	--n-p	1816517	425	----p	2902283	586	g----	4218284	571	g-nl-
659421	423	--n--	1818909	550	----p	2903410	533	g-nlp	4219093	391	g-nl-
660408	703	--n--	1823589	468	-----	2925686	484	-----	4219964	827	g-nl-
675647	439	--nlp	1829942	464	-----	2931729	572	--n-p	4233455	551	--n--
678496	386	--nlp	1842543	494	--nl-	2966921	511	--n--	4240158	413	-----
707074	530	-----	1851755	574	--nlp	2985071	1571	g-nl-	4248582	1975	g-nlp
715656	671	--nlp	1855560	418	--nl-	2986811	7598	g-nlp	4257986	1081	g-nl-
719924	438	--nlp	1868202	750	g-nl-	2996542	498	g-nl-	4259212	577	g-nl-
727973	490	-i---	1876742	368	-----	3003014	538	--nl-	4266343	1135	g---p
732771	637	ginlp	1891663	381	--nl-	3003731	578	--nl-	4273131	396	--n--
735163	1053	ginl-	1903183	360	--n-p	3013557	605	--nl-	4279950	1329	g-nlp
736988	551	ginl-	1905775	445	--nl-	3028904	410	--n-p	4285394	479	g-n-p
751962	520	g-nl-	1943986	454	--nlp	3048753	438	-----	4304478	404	-----
769872	741	-----	1956024	424	--n-p	3065024	392	--n--	4308668	419	----p
799381	712	--nlp	1976097	428	--n-p	3077172	505	--nlp	4310980	501	--nlp
819812	530	--nlp	1977264	482	--nlp	3085926	405	----p	4324734	462	--nlp
848056	446	-----	1984126	905	g-n-p	3098690	453	-----	4335751	656	--nlp
849273	490	-----	1993347	545	-----	3117038	617	----p	4338464	369	----p
871972	643	-----	2009130	1425	--nl-	3131908	427	--nlp	4346775	605	----p
872825	395	--nl-	2021494	489	--nl-	3134231	539	-----	4358053	481	--nl-
915205	531	----p	2022180	452	--nl-	3169846	625	--nl-	4359906	426	--nl-
931240	410	-----	2031411	824	g-n--	3181498	377	-----	4366380	492	----p

953644	517	g----	2037211	378	--nl-	3183123	473	g--lp	4371868	728	-----
985994	818	-----	2039564	542	--nlp	3187814	502	-inl-	4382328	362	--n-p
996712	487	--nl-	2040188	415	--nl-	3188374	629	-inlp	4407987	664	--nl-
1030860	382	----p	2042317	684	--nl-	3189868	543	-inlp	4435491	1309	g-nl-
1049790	450	--n--	2050929	462	g----	3214483	377	----p	4460668	462	--n--
1050283	554	--nl-	2054529	684	ginlp	3217041	447	----p	4473278	2052	ginlp-
1063159	690	--nl-	2055487	815	ginlp	3232171	384	--nl-	4477500	1195	g-nlp
1091438	1604	-inl-	2066210	527	-inlp	3250683	426	g-n--	4495848	478	-inlp
1094857	419	-inl-	2083352	399	----p	3255981	383	g--p	4501810	502	ginlp
1095707	407	--nl-	2085039	380	-----	3261326	390	--n--	4502360	1553	ginlp
1102438	908	g-nlp	2100942	1725	ginlp	3264009	385	g-nlp	4504233	479	ginlp
1107417	415	--n-p	2103095	2563	ginlp	3264982	708	g-nl-	4522961	485	-inlp
1118150	513	--nl-	2105951	1900	ginlp	3265752	1971	g-nl-	4523535	388	-inlp
1160698	386	--n-p	2111096	435	ginl-	3272800	721	g--p	4537372	477	ginlp
1165013	367	-----	2112807	373	-----	3281941	363	----p	4538360	681	ginlp
1167982	372	--n--	2128041	443	--n-p	3284977	664	g-nlp	4539534	554	ginl-
1195484	481	ginlp	2135166	769	-----	3352044	528	-----	4540509	690	ginlp
1196064	1971	ginlp	2140999	369	----p	3358701	551	-inl-	4553289	468	g-nl-
1209833	854	ginlp	2148871	468	--nlp	3359881	379	-inlp	4554206	740	g-nl-
1210772	1359	ginl-	2166143	472	--nlp	3372534	441	----p	4569842	730	--nl-
1214579	1748	ginl-	2176533	370	----p	3375472	381	-----	4574875	960	-inlp
1215406	921	ginlp	2185400	550	-inl-	3382306	540	----p	4578345	1119	ginlp
1218508	500	-----	2188911	552	ginl-	3383191	482	---l-	4589314	545	-----
1219391	392	-i---	2189551	1054	ginl-	3411440	518	--n--	4592666	405	--n-p
1222160	538	--nlp	2202149	512	----p	3416118	360	-----	4600907	689	--nl-
1228554	498	--nl-	2226735	446	--n-p	3416614	402	--n--	4638587	410	-----
1229535	406	--n--	2231648	435	--nlp	3453341	545	--nl-	4639554	405	-----
1254990	742	--n--	2257132	653	--n-p	3467648	565	-i--p			
1272774	392	----p	2267416	665	--n--	3510244	373	--n-p			

\*Символами **ginlp** обозначен источник данных:

**g** – «геномные островки», обнаруженные GIST [22] (всего 474 гена в геномной карте *E. coli* K12 U00096.2, согласно последней аннотации в базе данных RegulonDB (ver. 8.0) [7]);

**i** – «геномные островки», предсказанные программой IslandViewer [23] (всего 486 генов в том же геноме);

**n** – чужеродные гены, обнаруженные Nakamura et al. [19] (всего 1041 ген в том же геноме);

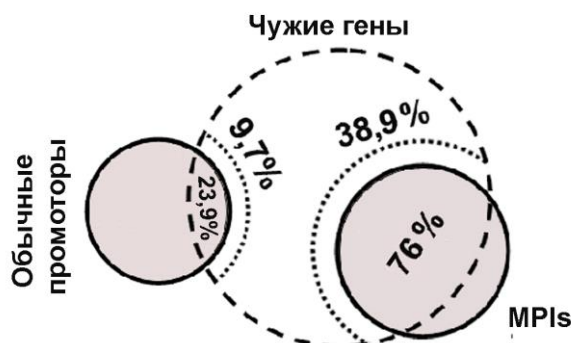
**l** – чужеродные гены, обнаруженные Lawrence и Ochman [18] (всего 714 генов в том же геноме);

**p** – чужеродные гены, обнаруженные Price et al. [20] (всего 1049 генов в том же геноме).

*Островки*, не ассоциированные с чужеродной ДНК, так же как островки, ассоциированные с генами, чей горизонтальный перенос был предсказан только в одной работе, выделены жирным шрифтом.

### Мультиспецифичные промоторные островки ассоциированы с чужими генами

Было установлено, что 370 из 434 MPIs (85%) ассоциированы с генами, чужеродность которых предсказана, по крайней мере, в одной из работ [18, 19, 22, 23], а для 283 островков на эту связь указывают данные, полученные не менее чем двумя коллективами (таблица 1).



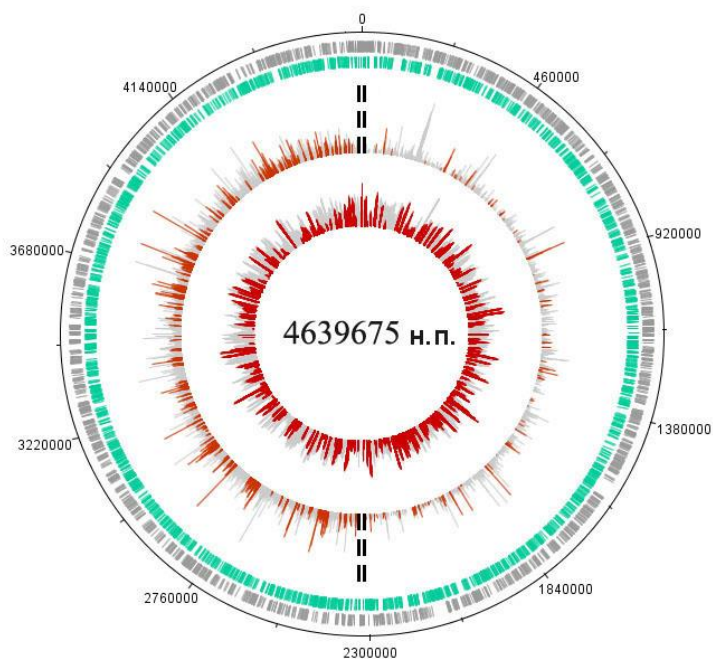
**Рис. 1.** Относительное число обычных промоторов и MPIs (затемнённые окружности слева и справа, соответственно), а также степень их перекрытия с общим набором чужих генов, предсказанных в работах [18, 19, 22] и [23] (пунктирная окружность). Точечные сегменты отсекают доли чужих генов, ассоциированных со сравниваемыми промоторными участками (учтены гены, транскрибируемые дивергентно с общей промоторной области и гены, транскрибируемые в виде полицистронных единиц).



Самая высокая степень перекрытия была обнаружена с перечнем чужих генов, предсказанных Lawrence и Ochman [18] по специфике нуклеотидного состава и профилю использования синонимичных кодонов. Количество MPIs, ассоциированных с чужими генами, в этом случае было в 3.6 раза больше, чем при их случайном распределении в геноме. В то же время степень ассоциации MPIs с чужими генами, предсказанными на основе филогенетического анализа в работе Price et al. [20], оказалось минимальной. Она превышала ожидаемую величину всего в 1.8 раза.

На рис. 1 схематически показана степень взаимного перекрытия между наборами, содержащими 351 обычную промоторную область или 434 MPIs с набором из 1513 чужеродных генов, выявленных четырьмя независимыми группами [18, 19, 22, 23]. Видно, что обычные промоторы, как правило, контролируют экспрессию родных генов кишечной палочки, и только 23.9% из них ассоциировано с чужими генами. Это в 3.2 раза меньше, чем доля таких же MPIs. Тем не менее, только 38.9% чужих генов ассоциировано с *островками*. Т. е., хотя горизонтально приобретённые гены имеют тенденцию накапливать промотор-подобные сайты [22], менее половины из них находится рядом или перекрывается с *промоторными островками*. Более того, около 7% чужих генов, выявленных во всех работах [18–20, 22, 23], вообще не имели промоторов, предсказанных PlatPromU.

Возможно, что эта неоднородность чужих генов отражает динамику их эволюционной адаптации к геному *E. coli*. Если *промоторные островки* действительно являются инструментом эволюционной ассимиляции чужеродного генетического материала, то недавно приобретённые гены могут находиться в состоянии их накопления, в то время как промоторы генов, которые уже интегрированы в регуляторные сети нового хозяина, наоборот, могут быть предметом очищающей эволюции, которая удаляет лишние промоторы.

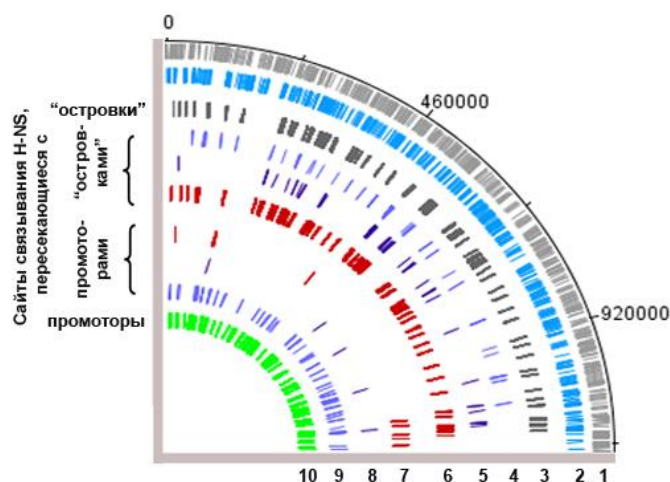


**Рис. 2.** Распределение функциональных сайтов по геному *E. coli*. Две **внешних окружности**: генная карта штамма по обоим нитям. **Третья окружность**: относительное число (N) РНК длиной 10÷11 нуклеотидов (слева) и длиной  $\geq 44$  н. (справа), зарегистрированных Dornenburg et al. [6] и нанесённых на график как  $\log_{10}(N+1)$ . **Четвёртая окружность**: эффективность связывания с RNAP, измеренная Reppas et al. [10], и выраженная как  $\log_2$  отношения сигналов специфичной и контрольной гибридизации. Число РНК-продуктов и гибридизационные сигналы, относящиеся к MPIs, показаны красным цветом.



### 434 MPIs и 78 PIs имеют аналогичные функциональные свойства

Ранее нами было показано, что все 78 PIs, обнаруженные PlatProm, эффективно связываются с RNAP [1], но вместо синтеза полноразмерных РНК обычно запускают abortивную транскрипцию коротких олигонуклеотидов [9]. Окружность 4 на рис. 2 показывает, что MPIs тоже хорошо взаимодействуют с RNAP. Комплексы с ней зарегистрированы для 86% *островков* и 82% обычных промоторов. При этом только 149 из 434 MPIs (34%) начинают синтез РНК длиной  $\geq 44$  нуклеотидов. Это значительно меньше, чем для обычных промоторов – 56%. Более того, общее число РНК-продуктов, соответствующих MPIs, оказалось в два раза меньше ожидаемого количества при их случайном распределении по геному, а в участках, окружающих TSPs обычных промоторов ( $\pm 50$  н.п.), это число было, наоборот, в 27 раз больше. Красная и серая гистограммы на правой половине третьей окружности рис. 2 отражают эту разницу. Зато 97% MPIs и только 71% обычных промоторов ведут синтез abortивных РНК (красная и серая гистограммы на левой половине третьей окружности рис. 2). Таким образом, 434 MPIs проявляют те же свойства, что и 78 PIs [1, 9]. Аналогичное сходство было обнаружено и во взаимодействии с гистонподобным белком H-NS (рис. 3), который специфически ингибирует транскрипцию горизонтально приобретённых генов [28–31].



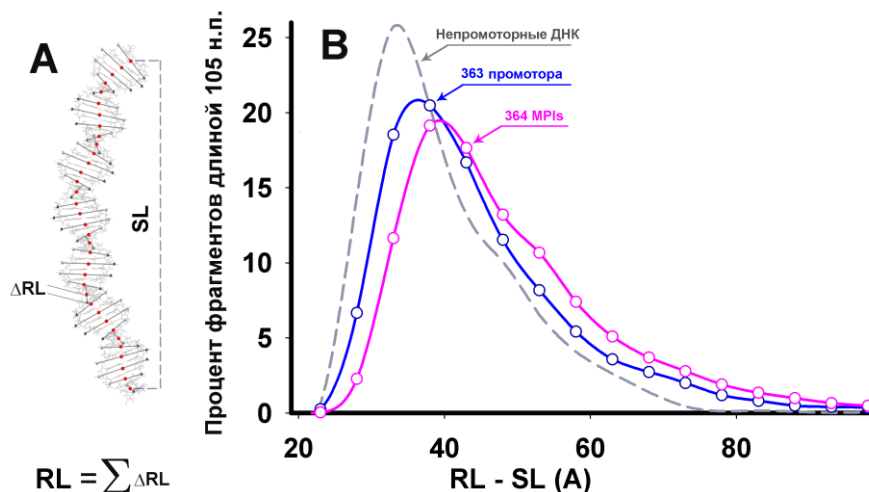
**Рис. 3.** Распределение MPIs (3-й сектор) и обычных промоторов (10-й сектор) в первой четверти генома *E. coli*. Секторы 1 и 2 показывают расположение генов на обоих нитях бактериальной хромосомы. На 6-ом и 7-ом секторах показаны места связывания H-NS, перекрывающиеся с MPIs или обычными промоторами ( $\pm 250$  н.п. вокруг точки старта), соответственно. Комплексы были зарегистрированы в геноме бактериальных клеток, выращенных при стандартных физиологических условиях в среде Luria–Bertani [13, 16]. На 5-ом и 8-ом секторах показаны места связывания H-NS, перекрывающиеся с MPIs или обычными промоторами в геномах клеток, выращенных в среде M9 + фруктоза [14, 15], а на 4-ом и 9-ом – в этой же минеральной среде, но в присутствии салициловой кислоты [14].

На рис. 3 показаны места связывания H-NS, которые выявлены с помощью технологии ChIP-on-chip [13–16] и перекрываются с MPIs (сектора 4–6) или обычными промоторами (сектора 7–9). Анализ показал, что в стандартных условиях роста 425 из 434 MPIs (97.9%) взаимодействуют с H-NS (сектора 5 и 6). Процент обычных промоторов, ассоциированных с H-NS в этих же условиях, был гораздо меньше (21.5%, сектора 7 и 8). Эта разница практически исчезает в условиях кислотного стресса (сектора 4 и 9), когда бактериальный геном подвергается защитной конденсации. Т. е. большинство *промоторных островков* поддерживается клеткой в гетерохроматин-подобном состоянии.

## Структурные особенности MPis способствуют их компактизации

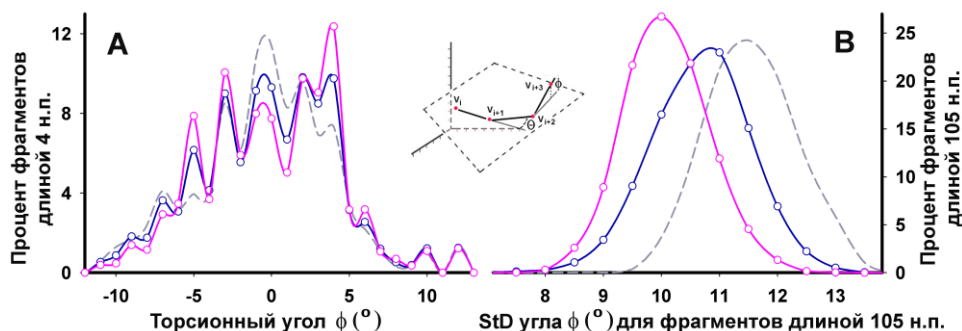
Ранее было обнаружено, что пространственная конфигурация 78 промоторных островков отличается от 3D-структуры обычных промоторов [9]. Островки, выявленные PlatProm, оказались более изогнутыми, в их моделях было больше непланарных тетраплетов, а двойная спираль оказалась «перезакрученной». Эти особенности вполне могут быть причиной необычных функциональных свойств промоторных островков. Если это так, то MPis, имеющие похожие с PIs функциональные свойства, тоже должны иметь характерные структурные особенности. Чтобы это проверить, из общего набора MPis были удалены такие островки, которые содержали PIs. Из оставшихся геномных областей был сформирован набор, содержащий 364 неперекрывающихся фрагмента длиной 300 н.п. Их 3D-модели были построены на сервере *DNA Tools* [24] и сопоставлены с такими же по размеру фрагментами ДНК, взятыми из обычных промоторных участков (всего 363), и с фрагментами непромоторной ДНК (контрольный набор, использованный в [9]). Результаты структурного анализа, представленные на рис. 4–6, свидетельствуют о том, что 364 представителя MPis имеют такие же структурные особенности, как 78 PIs, охарактеризованные ранее.

Так, для оценки глобальной изогнутости двойной спирали ДНК был использован конформационный параметр **RL–SL**. Способ его определения показан на рис. 4,А. Численные значения рассчитывали для всех фрагментов длиной 105 н.п., взятых для каждой модели со сдвигом в 1 н.п. Гистограммы распределения этих величин для всех семейств показаны на рис. 4,В. Они явно отличаются, и среднее значение  $(\mathbf{RL-SL})_{105}$  возрастает в следующем порядке: непромоторные ДНК (37.7 Å) < обычные промоторы (42.6 Å) < MPis (46.6 Å). Это указывает на повышенную способность островков взаимодействовать с белками нуклеоида, предпочтительными мишенями которых являются именно изогнутые участки ДНК.



**Рис. 4.** А: Пример виртуальной модели, созданной с помощью программного обеспечения *DNA Tools* [24], с указанием фосфорной конформационной цепочки (красные точки). В: Гистограмма распределения величин  $(\mathbf{RL-SL})_{105}$ , рассчитанных для всех субфрагментов анализируемых наборов длиной 105 н.п. (интервал разбиения для  $(\mathbf{RL-SL})_{105}$  – 5 Å).

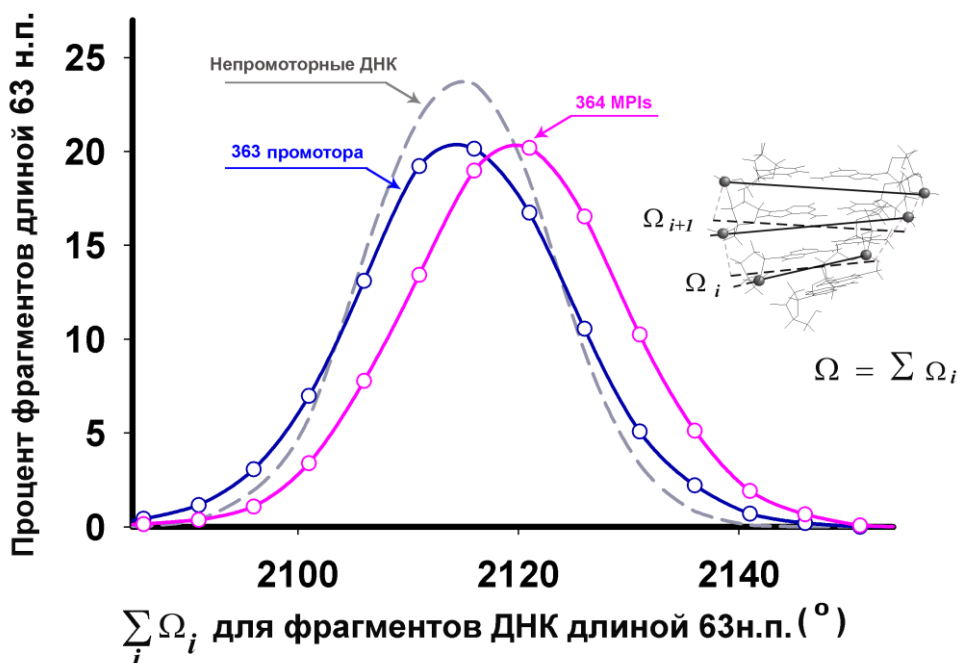
В полном соответствии с предыдущим анализом [9], средние значения угла  $\Theta$  между соседними звеньями углеродной цепочки, соединяющими вершину  $\mathbf{v}_i$  с  $\mathbf{v}_{i+1}$  и вершину  $\mathbf{v}_{i+1}$  с  $\mathbf{v}_{i+2}$  (см. вставку на рис. 5), оказались практически одинаковыми в трёх семействах (5.6°, 5.6° и 5.5°); а значения торсионного угла  $\Phi$  (см. вставку на рис. 5) имели специфическое распределение (рис. 5,А).



**Рис. 5.** Распределение (А) и вариабельность (В) в субфрагментах длиной 105 н.п. угла  $\phi$  (схематически показан на вставке) для виртуальных моделей 364 MPIs (фиолетовые кривые), 363 обычных промоторов (синие кривые) и непромоторных ДНК (пунктирные кривые). Значения угла были рассчитаны для углеродных (А) и фосфорных (В) конформационных цепочек. Интервал разбиения был  $1^\circ$  для величин  $\phi$  (А) и  $0.5^\circ$  для StD (В).

Особый интерес представляет явно пониженная вариабельность угла  $\phi$  в семействе MPIs (рис. 5,В). Ранее эта особенность тоже наблюдалась [9]. Она свидетельствует о некоторой регулярности в трёхмерной траектории ДНК, что может способствовать предполагаемой конденсации *промоторных островков*.

Одним из наиболее дискриминационных свойств *промоторных островков* были большие значения твистового угла  $\Omega$  (схематически показан на вставке рис. 6), нетипичные для обычных промоторов [9]. На рис. 6 приведены гистограммы распределения суммарных углов  $\Omega$  во фрагментах длиной 63 н.п., которые свидетельствуют о такой же закономерности и для MPIs. И хотя суммарный твистовый угол в этом семействе всего на  $5^\circ$  больше, чем у обычных промоторов, избыточная закрученность двойной спирали может препятствовать локальному плавлению ДНК во время образования транскрипционного комплекса и, следовательно, подавлять синтез РНК. В таком случае транскрипционная активность *промоторных островков* должна увеличиваться в генетическом окружении с более высоким уровнем отрицательной суперспирализации, чем в геномной ДНК.

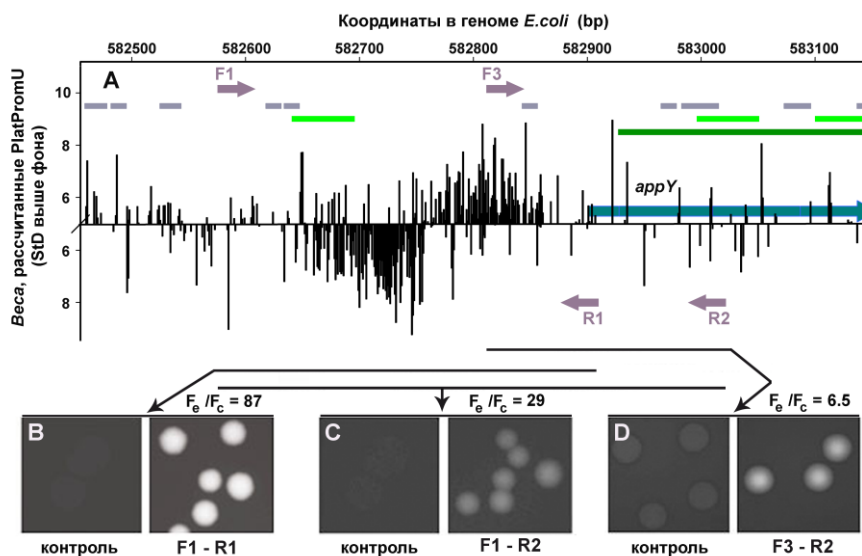


**Рис. 6.** Распределение суммарных углов  $\Theta$ , рассчитанных по *фосфорным цепочкам*, для фрагментов длиной 63 н.п. в трёх изучаемых семействах. Интервал разбиения  $5^\circ$ .

### Промоторный островок, ассоциированный с геном *appY*, способен начинать синтез мРНК гена *gfp*

Чтобы оценить зависимость промоторной активности от свойств генетического окружения, был выбран *островок*, ассоциированный с геном *appY* (рис. 7,А). Его способность начинать синтез длинных РНК до сих пор не была зарегистрирована методами полногеномного скрининга [6, 10], т. е., его промоторная активность в «родном» окружении подавлена. Три разных фрагмента этого *островка* были встроены в плазмиду рЕТ28b-eGFP перед беспромоторным геном зелёного флуоресцирующего белка (*gfp*). В клетках *E. coli*, трансформированных этими плазмидами, наблюдался синтез репортёрного белка (рис. 7,В–D). Таким образом, промоторы *островка*, ассоциированного с геном *appY*, могут, в принципе, функционировать как обычные промоторы.

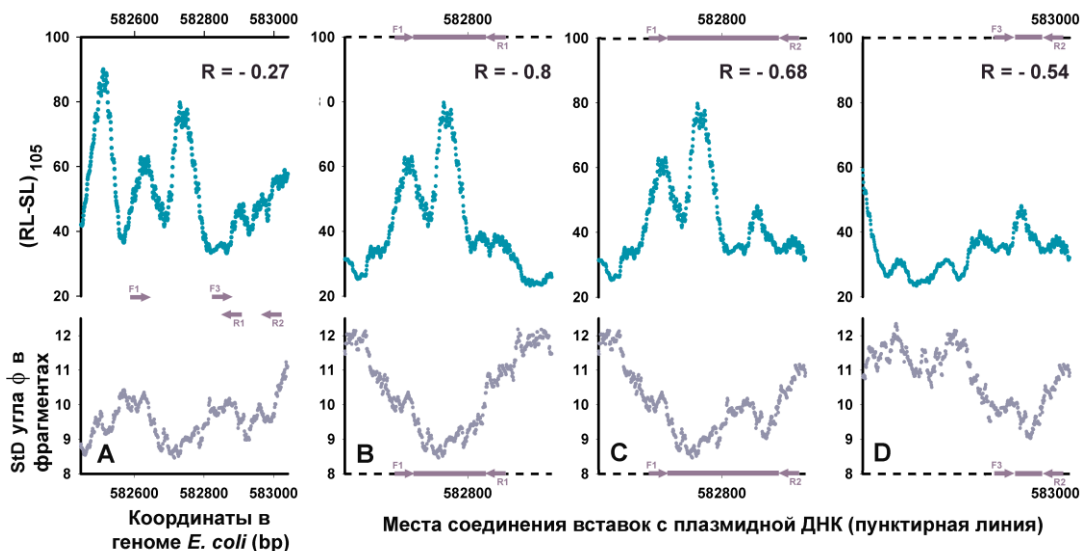
Относительную интенсивность флуоресценции экспериментальных ( $F_e$ ) и контрольных ( $F_c$ ) клеток использовали в качестве меры транскрипционной активности интегрированных в плазмиду промоторных участков. Наибольшую транскрипционную активность, в 87 раз превышающую фоновую экспрессию репортёрного гена, проявила конструкция **F1-R1** (рис. 7,В), которая содержит все промоторы *островка*, лежащие в межгенном пространстве (рис. 7,А). Определённую роль в этом может играть повышенная отрицательная суперспиральность плазмидной ДНК. Кроме того, конструкция **F1-R1** не содержит сайтов связывания ингибитора Н-NS (показаны на рис. 7,А), а более длинная вставка (**F1-R2**) в промоторную область *gfp* увеличивала его транскрипцию всего в 29 раз (рис. 7,С). Это значит, что Н-NS вносит свой вклад в подавление транскрипции *appY* в геномной ДНК. Важно, однако, что удаление активных межгенных промоторов из встроеного в плазмиду фрагмента (конструкция **F3-R2**) снижало его промоторную активность, но сохраняло её на достаточно высоком уровне ( $F_e/F_c = 6.5$ , рис. 7,Д).



**Рис. 7.** А: ген *appY* и направление его транскрипции обозначены голубой стрелкой. Вертикальными столбиками выше и ниже оси X показаны точки инициации транскрипции, предсказанные PlatPromU для верхней и нижней нитей ДНК, соответственно. Сайты связывания Н-NS, предсказанные *in silico* [30], изображены толстыми серыми линиями, а места связывания этого белка, зарегистрированные в работах [14, 15] или [13], отмечены зелёными или тёмно-зелёными линиями, соответственно. Серые стрелки указывают позиции праймеров, которые были использованы для амплификации фрагментов ДНК, интегрированных в плазмиду рЕТ28b-eGFP. Относительные размеры фрагментов, амплифицированных с праймерами F1 и R1, F1 и R2 или F3 и R2, показаны горизонтальными линиями. В–D: колонии клеток, трансформированные плазмидой с беспромоторным геном *gfp* (контроль) или плазмидами, содержащими разные фрагменты *промоторного островка* перед геном репортёрного белка.

Это значит, что присутствие мест связывания H-NS в регуляторной области *gfp*, а также самого H-NS и любых других ингибиторов в цитоплазме бактериальных клеток не заблокировало синтез мРНК гена *gfp* (рис. 7,D). Косвенно это указывает на важность структурного фактора в генетическом окружении *островка*, которое отличается в геномной ДНК и в плазмиде, по крайней мере, степенью суперспирализации. К сожалению, влияние этой нелокализованной изомеризации на структурную конформацию промоторной вставки в плазмиде пока невозможно промоделировать.

Но мы имели возможность оценить влияние соседней ДНК на другой зависящий от окружения структурный параметр:  $(\mathbf{RL-SL})_{105}$ . Исследованный фрагмент *промоторного островка*, находящийся между праймерами **F1** и **R2**, в геноме формирует 2 анизотропных изгиба двойной спирали, о чём свидетельствуют максимумы на профиле параметра  $(\mathbf{RL-SL})_{105}$  (верхняя часть рис. 8,A, см. также анимацию F1-R2g.avi в доп. материалах к статье). Кроме того, этот фрагмент находится рядом с ещё одним большим изгибом, более удалённым от гена *appY*. В зависимости от расположения относительно стартовой точки транскрипции, анизотропные изгибы могут активировать или, наоборот, ингибировать синтез РНК. Как и ожидалось, они формируются в участках с малыми вариациями угла  $\phi$  (рис. 8,A, нижний график), который отрицательно коррелирует с параметром  $(\mathbf{RL-SL})_{105}$ . В плазмиде pET28b-EGFP фрагмент **F1-R1** оказывается в плоском окружении (рис. 8,B, см. также анимацию F1-R1.avi в доп. материалах) и демонстрирует максимальную промоторную активность, что предполагает приемлемое для успешной инициации транскрипции расположение двух основных изгибов. Встраивание более длинного фрагмента **F1-R2** в плазмиду приводит к образованию третьего изгиба справа (рис. 8,C, см. также анимацию F1-R2p.avi в доп. материалах). Так как фрагмент **F2-R2**, содержащий только этот небольшой изгиб (рис. 8,D, см. также анимацию F2-R2.avi в доп. материалах), оказался транскрипционно активным, ясно, что РНК-полимераза взаимодействует с этой промоторной областью. Не исключено, что именно это взаимодействие блокирует синтез РНК в конструкции **F1-R2**, который с исключительно высокой эффективностью начинается на межгенных промоторах, расположенных между праймерами **F1** и **R1**.



**Рис. 8. А:** профили конформационного параметра,  $(\mathbf{RL} - \mathbf{SL})_{105}$  (верхний график) и вариации угла  $\phi$  (нижний график) вдоль *промоторного островка*, ассоциированного с геном *appY* в геноме. Структурные параметры были последовательно вычислены для всех субфрагментов размером 105 н.п. со сдвигом в 1 н.п., а полученные значения указаны для их первых позиций. Расположение праймеров, использованных для амплификации, обозначено серыми стрелками. **В–D:** то же самое для разных вставок в плазмиде pET28b-eGFP. Расположение вставок в плазмидной ДНК схематически показано на осях X в виде серых кассет и пунктирных линий, соответственно. Символ **R** обозначает коэффициент корреляции между параметрами  $(\mathbf{RL} - \mathbf{SL})_{105}$  и величин StD.



Таким образом, структурная информация, полученная в данной работе, позволяет предложить несколько механизмов, объясняющих подавленную транскрипционную активность исследованного *островка* в геноме, его активность в плазмиде, и зависимость экспрессии репортёрного гена от особенностей интегрированного промоторного участка.

## ОБСУЖДЕНИЕ

Известно, что протяжённые *геномные островки* (4000 ÷ 48000 н.п.), состоящие из чужеродной ДНК и часто содержащие гены патогенности, отличаются высоким содержанием сигналов инициации транскрипции [22]. Анализируя распределение в геноме промоторов, предсказанных алгоритмом PlatProm в хромосоме *E. coli* K12 MG1655, мы также обнаружили 78 необычных областей с чрезвычайно высокой плотностью потенциальных мест инициации транскрипции, но очень низкой способностью инициировать синтез нормальных мРНК [1, 9]. Оказалось, что большинство из них находится рядом с генами, приобретёнными в процессе горизонтального переноса, в том числе рядом с такими, которые лежат за пределами протяжённых *геномных островков*. Была выдвинута гипотеза [9], что *промоторные островки* являются продуктом ускоренной эволюции, которая необходима для ассимиляции чужеродного генетического материала. Обладая специфической 3D-конформацией и будучи обогащёнными сайтами связывания Н-NS, *промоторные островки* могут выполнять прямо противоположные функции: препятствовать экспрессии нежелательных «пришельцев» и предоставлять подходящие промоторы для чужих генов, если их экспрессия по каким-то причинам становится востребованной. Эта гипотеза требует тщательной проверки. Возник, в частности, вопрос: почему число предсказываемых разными авторами чужих генов у *E. coli* K12 MG1655 гораздо больше (> 1000), чем число обнаруженных ранее *островков* (78)?

Данное исследование было предпринято с надеждой сократить это различие благодаря учёту *островков*, включающих в себя промоторы разной  $\sigma$ -специфичности. Используя алгоритм PlatPromU, игнорирующий весовые матрицы консервативных гексанулеотидов, узнаваемых  $\sigma^{70}$ , мы обнаружили 434 *мультиспецифичных промоторных островка*. Их специфическая локализация в геноме, так же как функциональные (рис. 2 и 3) и структурные (рис. 4–6) свойства, оказались практически такими же, как у 78 охарактеризованных ранее *островков*. Но только ~39% потенциально чужих генов оказалось ассоциировано с новым набором обогащённых промоторами участков (рис. 1). Следует отметить, что этот процент немного занижен, т. к. при поиске MPIs мы использовали более строгие критерии, чем при поиске PIs. При использовании прежних критериев размер этого набора увеличился бы на 166 MPIs, но доля чужих генов, ассоциированных с таким набором, увеличивается только до 50%. Ясно, что снижая пороговые уровни, мы можем ещё увеличить степень оцениваемого перекрытия. Но около 7% чужих генов, выявленных во всех публикациях [18–20, 22, 23], совсем не имеют предсказанных точек инициации транскрипции. Стало очевидным, что некоторые гены, приобретённые за счет горизонтального переноса, не стали накапливать сигналы транскрипции. Пока не ясно, являются ли они псевдогенами, или уровень их экспрессии с самого начала оказался пригодным для *E. coli*. В любом случае, механизмы, лежащие в основе адаптации чужеродного генетического материала к регуляторным сетям клеток-хозяев, остаются неизвестными, и обнаруженные нами *островки* могут стать предметом для целенаправленного исследования адаптивных изменений в регуляторных участках перенесённых генов.

Результаты структурного анализа, полученные в этой работе, подтвердили все предыдущие наблюдения [9], в том числе довольно неожиданный низкий уровень



вариаций торсионного угла  $\phi$  в моделях *промоторных островков*. Было высказано предположение, что это отражает некоторую регулярность в пространственной конфигурации двойной спирали островковых участков ДНК. На примере *промоторного островка*, ассоциированного с геном *appY*, мы действительно увидели, что локализация изгибов (максимумы параметра **RL-SL**) совпадает с участками, имеющими низкие значения параметра StD для  $\phi$  (рис. 8). Коэффициенты корреляции между этими двумя параметрами очень велики и для трёх фрагментов, встроенных в плазмиду pET28b-EGFP, убывают в следующем порядке:  $|\mathbf{R}_{(F1-R1)}| > |\mathbf{R}_{(F1-R2)}| > |\mathbf{R}_{(F3-R2)}|$ , который в точности соответствует характеру изменений зарегистрированной транскрипционной активности (рис. 7, B–D). Пока не понятно, насколько общей закономерностью является это соответствие. Но наличие изгибов и пониженная изменчивость торсионных углов являются свойствами не только *промоторных островков*, но и обычных промоторов (рис. 4 и 5). А это значит, что способность к активной транскрипции закодирована в структурной организации *островков* и при необходимости может быть реализована.

Работа была поддержана Российским Фондом фундаментальных исследований (гранты № 12-07-00530 и № 13-04-00997), а также Министерством образования и науки Российской Федерации (проект «Множественные промоторы как индикаторы перекрывающейся транскрипции и альтернативного кодирования», Шифр 4.4119.2011).

### СПИСОК ЛИТЕРАТУРЫ

1. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons in genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4419–4431.
2. Huerta A.M., Collado-Vides J. Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J. Mol. Biol.* 2003. V. 333. P. 261–278.
3. Tutukina M.N., Shavkunov K.S., Masulis I.S., Ozoline O.N. Intragenic promoter-like sites in the genome of *Escherichia coli*. Discovery and functional implication. *J. Bioinform. Comput. Biol.* 2007. V. 5. P. 549–560.
4. Тутукина М.Н., Шавкунов К.С., Масулис И.С., Озолин О.Н. Антисмысловая транскрипция в локусе *hns Escherichia coli*. *Молекулярная биология*. 2010. Т. 44. С. 497–506.
5. Mendoza-Vargas A., Olvera L., Olvera M., Grande R., Vega-Alvarado L., Taboada B., Jimenez-Jacinto V., Salgado H., Juarez K., Contreras-Moreira B., Huerta A.M., Collado-Vides J., Morett E. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE*. 2009. V. 4. Article № e7526.
6. Dornenburg J.E., DeVita A.M., Palumbo M.J., Wade J.T. Widespread antisense transcription in *Escherichia coli*. *mBio*. 2010. V. 1. Article № e00024-10.
7. Salgado H., Peralta M., Gama-Castro S., Santos-Zavaleta A., Muniz-Rascado L.J., Garcia-Sotelo J.S., Weiss V., Solano-Lira H., Martinez-Flores I., Medina-Rivera A., Salgado-Osorio G., Alquicira-Hernandez S., Alquicira-Hernandez K., Lopez-Fuentes A., Porron-Sotelo L., Huerta A.M., Bonavides-Martinez C., Balderas-Martinez Y., Pannier L., Olvera M., Labastida A., Jimenez-Jacinto V., Vega-Alvarado L., del Moral-Chavez V., Hernandez-Alvarez A., Morett E., Collado-Vides J. RegulonDB (version 8.0): omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucl. Acids Res.* 2013. P. D203–D213. URL: <http://regulondb.ccg.unam.mx/> (дата обращения: 28.07.2013).
8. Shavkunov K.S., Tutukina M.N., Masulis I.S., Ozoline O.N. Promoter islands: the novel elements in bacterial genomes. *J. Biomol. Struct. Dynam.* 2011. V. 28. P. 1128–1129.

9. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE*. 2013. V. 8. Article № e62601.
10. 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.
11. Herring C.D., Raffaele M., Allen T.E., Kanin E.I., Landick R., Ansari A.Z., Palsson B.O. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol.* 2005. V. 178. P. 6166–6174.
12. Панюков В.В., Назипова Н.Н., Озолинь О.Н. Пакет программ aSHAPE для изучения пространственной конформации участков бактериального генома. *Математическая биология и биоинформатика*. 2011. Т. 6. С. 211–227. URL: [http://www.matbio.org/2011/Panjukov2011\(6\\_211\).pdf](http://www.matbio.org/2011/Panjukov2011(6_211).pdf) (дата обращения: 28.07.2013).
13. Kahramanoglou C., Seshasayee A.S.N., Prieto A.I., Ibberson D., Schmidt S., Zimmermann J., Benes V., Fraser G.M., Luscombe N.M. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucl. Acids Res.* 2011. V. 39. P. 2073–2091.
14. Grainger D.C., Hurd D., Goldberg M.D., Busby S.J.W. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucl. Acids Res.* 2006. V. 34. P. 4642–4652.
15. Grainger D.C., Aiba H., Hurd D., Browning D.F., Busby S.J.W. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. *Nucl. Acids Res.* 2007. V. 35. P. 269–278.
16. Prieto A.I., Kahramanoglou C., Ali R.M., Fraser G.M., Seshasayee A.S.N., Luscombe N.M. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucl. Acids Res.* 2012. V. 40. P. 3524–3537.
17. Cho B.K., Knight E.M., Barrett C.L., Palsson B.O. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res.* 2008. V. 18. P. 900–910.
18. Lawrence J.G., Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA*. 1998. V. 95. P. 9413–9417.
19. Nakamura Y., Itoh T., Matsuda H., Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics*. 2004. V. 36. P. 760–766.
20. Price M.N., Dehal P.S., Arkin A.P. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* 2008. V. 9. Article № R4.
21. Киселев С.С., Озолинь О.Н. Структурообразующие модули как индикаторы промоторной ДНК в бактериальных геномах. *Математическая биология и биоинформатика*. 2011. Т. 6. С. 39–52. URL: [http://www.matbio.org/2011/Ozoline2011\(6\\_39\).pdf](http://www.matbio.org/2011/Ozoline2011(6_39).pdf) (дата обращения: 28.07.2013).
22. Huang Q., Cheng X., Cheung M.K., Kiselev S.S., Ozoline O.N., Kwan H.S. High-density transcriptional initiation signals underline genomic islands in bacteria. *PLoS ONE*. 2012. V. 7. Article № e33759.
23. Langille M.G., Brinkman F.S. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics*. 2009. V. 25. P. 664–665.
24. Vlahovicek K., Kajan L., Pongor S. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucl. Acids Res.* 2003. V. 31. P. 3686–3687. URL: [http://hydra.icgeb.trieste.it/dna/model\\_it.html](http://hydra.icgeb.trieste.it/dna/model_it.html) (дата обращения: 28.07.2013).

25. Friedel M., Nikolajewa S., Suehnel J., Wilhelm T. DiProDB: a database for dinucleotide properties. *Nucl. Acids Res.* 2009. V. 37. D37–D40. URL: <http://diprodb.fli-leibniz.de/ShowTable.php> (дата обращения: 28.07.2013).
26. Sponer J., Gabb H.A., Leszczynski J., Hobza P. Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys. J.* 1997. V. 73. P. 76–87.
27. Oshima T., Ishikawa S., Kurokawa K., Aiba H., Ogasawara N. *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.* 2006. V. 13. P. 141–153.
28. Lucchini S., Rowley G., Goldberg M.D., Hurd D., Harrison M., Hinton J.C.D. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* 2006. V. 2. Article № e81.
29. Dorman C.J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 2007. V. 5. P. 157–161.
30. Munch R., Hiller K., Grote A., Scheer M., Klein J., Schobert M., Jahn D. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics.* 2005. V. 21. P. 4187–4189.

Материал поступил в редакцию 23.07.2013, опубликован 01.09.2013.