



УДК 577.214.625.087.1

**СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ НУКЛЕОТИДНОЙ
ПОСЛЕДОВАТЕЛЬНОСТИ ПРОКАРИОТИЧЕСКОГО ПРОМОТОРА.
ЭЛЕМЕНТЫ СТРУКТУРЫ, ОПРЕДЕЛЯЮЩИЕ ЭФФЕКТИВНОСТЬ
ПРОТЕКАНИЯ ЭТАПОВ НАЧАЛА ТРАНСКРИПЦИИ**

Артемов И. В., Васильев Г. В., Гуревич А. И.

*Институт биоорганической химии им. М. М. Шемякина
Академии наук СССР, Москва*

Статистическими методами с помощью ЭВМ изучены нуклеотидные последовательности 188 участков ДНК, содержащих промоторы. Найдено, что характерной особенностью обобщенных структур промоторов является наличие undекануклеотида NTT(G/C)TTGACA(A/T) или (G/C)TT(G/C)A(G/C)A(A/T)TT(G/T) (сайт узнавания) и гептануклеотида RTATATR или TATAATR (сайт инициации), между которыми находится участок в 12–19 п.о. Промоторы способны функционировать, если достигается минимальный уровень соответствия имеющихся у них сайтов узнавания и инициации обобщенным структурам (значение функции соответствия общей структуры не ниже 0,61; в наиболее эффективных промоторах эта величина достигает 1). Старт транскрипции находится на расстоянии 3–9 пар оснований за сайтом инициации (наиболее эффективный старт на расстоянии 4–7 пар оснований). Транскрипция может начинаться с любого нуклеотида, но предпочтительно с А или G, причем старт с А наиболее эффективен, если это звено входит в состав триплексида САТ или САС.

Эффективность промотора увеличивают дополнительные структурные факторы: протяженный А·Т-богатый участок непосредственно перед сайтом узнавания; целые промоторные структуры или несколько участков связывания РНК-полимеразы в предшествующей нуклеотидной последовательности. Характерной особенностью промотора является наличие элементов симметрии второго порядка в районах старта и сайта узнавания, а также в промежуточном участке между сайтами узнавания и инициации или же А·Т-богатого участка в последнем районе.

В нуклеотидной последовательности промотора содержится необходимая информация, определяющая точку старта и эффективность транскрипции РНК-полимеразой. Среди прокариотических промоторов наиболее изучены промоторы РНК-полимеразы *E. coli*. Их структура охватывает 50–80 п.о., что согласуется с размерами фермента (РНК-полимеразы), взаимодействующего с промотором, возможно, в двух положениях — в участке связывания (узнавания) и, после перемещения, в участке инициации транскрипции [1]. В этой структуре можно выделить три района с характерной (до некоторой степени) нуклеотидной последовательностью: 1) вблизи старта транскрипции, начинающейся обычно с остатка А или G (значительно реже с С) [1, 2] (здесь и далее приведены нуклеотиды той цепи ДНК, которая после точки старта идентична образующейся РНК); 2) гептануклеотидная последовательность за 5–7 п.о. до старта (последовательность Приблпоу [3], TATRATR); 3) 10–15 п.о., расположенных приблизительно за 25–40 п.о. до старта (последовательность Гилберта [2], TGTTG, или Шаллера [4], GTTGACACTTTA). Третий район соответствует участку связывания (узнавания).

Ранее неоднократно делались попытки суммировать структуру различных районов промотора в обобщенной нуклеотидной последовательности. Шерер и др. вывели структуру «идеального» промотора [5] на основе статистической обработки данных по 17 промоторам, исходя из вероятности нахождения того или иного нуклеотида на определенном расстоянии от старта транскрипции*:

* В приведенных обобщенных последовательностях обозначения строчными, прописными или прописными буквами жирным шрифтом даны в соответствии с возрастающей вероятностью нахождения соответствующих нуклеотидов в данном положении относительно старта транскрипции (отмечен точкой).

acc-t-gttGTTGAcATTT(t----tggcGGTTATATTg---cCAt-----a-----ttt.
 t a a ATA a g

Несколько иная обобщенная структура промотора была предложена Зибенлистом [6] *:

t---TGTTGACAA(ТТ-Т(N)₃₋₅t-t-TGgTATAATg-----c--T,
 t

Он отметил, что расстояние между районами более строгого соответствия (последовательностями Гилберта и Прибноу) может колебаться.

Увеличение числа структур промоторов, подвергнутых статистическому анализу, до 46 привело Розенберга и Курта [7] к еще менее определенной обобщенной последовательности:

tt--tgTTGGACA-ttt-(N)₆₋₉—atttgtTATAATg-(N)₄₋₇CAt--(Pu)
 tg

Наконец, Зибенлист, Симпсон и Гилберт пришли к выводу, что обобщенный промотор имеет следующую структуру [8]:

AAA-T----CTTGACA(N)₁₀₋₁₂---T--TATAAT(N)₃₋₇CAt
 T G

При этом на примере двух промоторов (генов *lac UV5 E. coli* и АЗ фага Т7) они экспериментально показали (см. также работы [9, 10]) наличие контактов между РНК-полимеразой и участками промотора в районе последовательности Гилберта, в районе «—16» (находящемся на расстоянии 24 п.о., т. е. двух витков двойной В-спирали от последовательности Гилберта и, следовательно, расположенном на той же стороне спирали), а также в районе последовательности Прибноу и в районе старта транскрипции, между которыми две нити ДНК оказались разделенными.

В то же время очевидно, что в структурах промоторов кроме элементов сходства, обеспечивающих узнавание этих участков ДНК РНК-полимеразой, должны существовать различия, которые обуславливают различную эффективность старта транскрипции. Действительно, сила промоторов, определяемая по скорости реакции взаимодействия с РНК-полимеразой, может различаться на 3 порядка [11], и это, вероятно, коррелирует с реальной эффективностью транскрипции соответствующих генов.

Считая необходимым учитывать при получении обобщенной структуры промотора критерий его эффективности (силы), мы провели статистический анализ известных нуклеотидных последовательностей промоторов, разбив их на группы в соответствии с этим критерием.

Известно, что транскрипция ранних генов бактериофагов, а также генов РНК, участвующих в репликации плазмид, протекает много эффективнее, чем транскрипция генов бактериальных белков (см., например, работы [11, 12, 13]). В бактериальной хромосоме к числу наиболее активно транскрибируемых относятся гены рибосомных белков и в особенности гены рибосомных РНК [14]. Поэтому в первую группу мы включили промотор гена РНК, участвующей в репликации плазмид Col E1 и pBR322, промоторы ранних генов фагов Т5 и Т7 и промоторы генов рибосомных РНК *E. coli* как наиболее сильные из описанных. Ко второй группе мы отнесли сильные промоторы фагов λ, λ_{imm} 434, φ 29, а также малых бактериофагов fd, H, M13, φX174, G4 и рибосомных белков *E. coli*, к третьей — промоторы поздних генов бактериофагов и различных белков *E. coli* и плазмид, к четвертой — структуры слабых промоторов генов *E. coli*, а также промоторов, эффективность которых существенно снижена в результате мутационных изменений, к пятой — структуры, напоминающие промоторные, но с неизвестной функциональной активностью. В шестую группу были объединены структуры тех промоторов, которые в результате делеций или иных изменений утратили способность к инициации транскрипции.

Такое условное деление, как показал дальнейший анализ, в основном совпало с делением на группы по структурным признакам. Однако при наличии в структуре какого-либо отдельного промотора особенностей, характерных для промоторов другой группы, мы считали возможным перемещать структуры из одной группы в другую. Приведенная в таблицах группировка структур промоторов отражает этот подход.

При анализе нуклеотидных последовательностей промоторов мы исходили из определения, что в каждой группе промоторов должны быть участки узнавания (связывания) и образования комплекса (закрытого), образования открытого комплекса и инициации транскрипции, а также (за точкой старта) участок диссоциации σ -субъединицы РНК-полимеразы. В промоторах различной эффективности эти участки могут различаться как по положению (расстояние от точки старта транскрипции), так и по нуклеотидной последовательности. В соответствии с этим в качестве первого критерия общности нуклеотидных последовательностей промоторов мы изучили подобие участков, находящихся на примерно одинаковом расстоянии от старта транскрипции с теми, для которых экспериментально установлены [8—10] контакты с РНК-полимеразой в закрытом и открытом комплексах (районы от -5 до -15 , от -15 до -20 , от -25 до -40).

В качестве второго критерия мы использовали распределение в промоторах и в предшествующих им последовательностях А·Т-богатых участков, поскольку экспериментально показано, что РНК-полимераза предпочтительно связывается с легкоплавкими районами ДНК *E. coli*, колифагов и плазмид [15].

Исходя из предположения, что обратные повторы (палиндромы) вблизи старта транскрипции могут облегчать образование и стабилизировать открытые комплексы с расплетенными участками двойной спирали, мы ввели третий критерий и изучили распределение в структуре промоторов и прилежащих к ним участков ДНК элементов симметрии второго порядка.

Наконец, мы изучили расположение в структуре ДНК, предшествующей промотору, дополнительных промоторов или участков связывания РНК-полимеразы, подобных частям промотора.

1. Подобие участков нуклеотидных последовательностей промотора

При получении обобщенных структур промоторов был использован принцип максимальной вероятности в итерационном процессе получения оптимальной структуры.

Структурное подобие было исследовано в двух районах: между нуклеотидами -15 и -4 , а также между нуклеотидами -50 и -30 . При получении обобщенной структуры мы использовали промоторы первых двух групп, считая, что среди сильных промоторов лучше проявится соответствие в исследуемых районах обобщенным структурам. Анализ показал возможность выбора начального приближения как совокупности двух гептануклеотидов в первом районе и двух ундекануклеотидов во втором изучаемом районе. На основе анализа вычисленной функции соответствия для двух указанных районов автоматически с помощью ЭВМ были выделены сайты узнавания и инициации. В случаях неоднозначности нуклеотидов на конкретных местах в начальном приближении уточнение нуклеотидной структуры обобщенного промотора было получено на основании принципа максимальной вероятности.

В районе между нуклеотидами -15 и -4 получен образ, состоящий из двух близких к гептануклеотиду Прибноту обобщенных структур сайта инициации: RTATATR или TATAATR. Среднее значение функции соответствия в этом районе в первой группе промоторов $-0,87$, во второй группе $-0,80$.

Во втором районе получен образ, состоящий из двух близких к последовательности Гилберта обобщенных структур сайта узнавания: NTT(G/C)TTGACA(A/T) и (G/C)TT(G/C)A(G/C)A(A/T)TT(G/T). При

Параметры функции соответствия
P — среднее значение, *D* — дисперсия

Номер группы	Сайт инициации		Сайт узнавания		Суммарная структура	
	<i>P</i>	<i>D</i>	<i>P</i>	<i>D</i>	<i>P</i>	<i>D</i>
1	0,87	0,12	0,73	0,09	0,79	0,05
2	0,80	0,12	0,79	0,11	0,79	0,07
3	0,76	0,12	0,71	0,10	0,73	0,07
4	0,69	0,09	0,71	0,10	0,70	0,06
5	0,69	0,17	0,67	0,14	0,68	0,04
6	0,66	0,07	0,69	0,04	0,68	0,05

этом среднее значение функции соответствия в этом районе в первой группе промоторов — 0,73, во второй группе — 0,79. Эти две ундекануклеотидные последовательности являются как бы частями одного тридекануклеотида: TT(G/C)TT(G/C)A(G/C)A(A/T)TT(G/T).

В изученных нуклеотидных последовательностях участки с высоким значением функции соответствия полученным обобщенным структурам были обнаружены неоднократно, иногда они накладывались друг на друга, и для отбора единой обобщенной структуры промотора мы ввели ограничение длины участка, разделяющего сайты узнавания и инициации, отбросив варианты со слишком близко и слишком далеко расположенными друг от друга сайтами. В промоторах первых двух групп минимальная длина промежуточного участка в результате составила 12 п.о., максимальная — 19 п.о. В этих промежуточных участках не было найдено района какого-либо подобию, и, следовательно, экспериментально обнаруженные контакты с РНК-полимеразой в этом районе, по-видимому, не связаны с непосредственным «узнаванием» нуклеотидной последовательности.

Значение функции соответствия суммарного промоторного участка по первой группе промоторов — 0,79, по второй группе промоторов — 0,79.

Незначительное сходство в нуклеотидных последовательностях отмечалось вблизи точки старта транскрипции. Почти у всех промоторов первой группы стартовым нуклеотидом является А (90%), и ему предшествует С (76%), а за А в большинстве случаев следует Т или С. Лишь в двух промоторах (1.7 и 1.9) * С предшествует стартовому нуклеотиду G. При этом участок между точкой старта и сайтом инициации включает 3—8 п.о. и в большинстве случаев — 4—7 п.о. Для пяти промоторов гgpP2 (1.14—1.18) с длиной промежуточного участка 3—4 нуклеотида ранее было найдено, что транскрипция начинается с А в значительно меньшей степени, чем с лежащего на два нуклеотида дальше С или с лежащего еще далее G [16]. По-видимому, для эффективного старта транскрипции длина промежуточного участка должна быть более 4 п.о.

Стартовым нуклеотидом большинства промоторов второй группы является G (64%), и ему, как правило, предшествует С или G. У некоторых промоторов (2.2, 2.5, 2.6, 2.8, 2.9, 2.19, 2.21, 2.23, 2.24, 2.25, 2.31) транскрипция, как и в случае промоторов первой группы, начинается с А, которому, однако, предшествует не С, а G или Т. Наконец, в одном промоторе (2.22) стартовым нуклеотидом является Т.

В промоторах третьей и четвертой групп транскрипция начинается с самых разнообразных нуклеотидов, хотя преимущественный старт с нуклеотидов А или G наблюдается и здесь. В третьей группе (в отличие от первых двух групп) имеются лишь четыре промотора, в которых стартовый нуклеотид А входит в состав тринуклеотида САТ (в четвертой группе такие промоторы вообще отсутствуют); следует отметить, что эти промоторы (3.14, 3.18, 3.35, 3.53), по-видимому, наиболее сильные среди промоторов третьей группы, так как уровень соответствия суммарной обобщенной структуре в этой группе выше среднего показателя.

* Здесь и далее в обозначениях промоторов первая цифра указывает номер группы, а вторая — номер промотора в группе.

Таблица 2. Нуклеотидные последовательности промоторов группы 1.

Номер промотора	Название гена	TGTTCACAAATG C C C T T	TATAATR RTAT	+1	Функция
1	T5 26	<u>TTCAGTTCCTAAATCC</u> <u>ACAATCTTGA</u>	<u>TATAATA</u> <u>TTCTC</u>	<u>ATAGTTTGA</u>	8, 13
2	T5 25	<u>TTTATTTGCCT</u> <u>TCAGGAATAATTTCTG</u>	<u>TATAATA</u> <u>GATTC</u>	<u>ATGAAATTC</u>	8, 13
3	T5 P28	<u>GTAGTTGCTAA</u> <u>ATGCTTAAATACTTGC</u>	<u>TATAATA</u> <u>TTTAT</u>	<u>ATAAATGA</u>	13
4	15 P207	<u>ATTCAATTCCT</u> <u>AAACGCTTCAATTTCTCG</u>	<u>TATAATA</u> <u>TACTTC</u>	<u>ATAAATGA</u>	13
5	T7 30	<u>GGCGTTGACTT</u> <u>GATGGCTCTT7AG</u>	<u>CTGAGG</u> <u>CTTLAGGT</u>	<u>CTTGGCTTT</u>	17
6	T7 31	<u>AGTATTGACTT</u> <u>AAAGTCTAACCTATAGG</u>	<u>ATATTA</u> <u>CAGCC</u>	<u>ATCGAGAGG</u>	18, 19
7	T7 A2	<u>GGTATTGACAA</u> <u>CATGAAGTAACATGCAGTAAGATACA</u>	<u>AAATC</u>	<u>GCTAGGTAA</u>	19
8	T7 A3	<u>ACGGTTGACAA</u> <u>CATGAAGTAAACACGC</u>	<u>TACGATC</u> <u>TACAC</u>	<u>ATGAAACCA</u>	19
9	E. coli rrnD P1	<u>CTTGTGCAAAA</u> <u>AATTCGGATCCC</u>	<u>TATAATG</u> <u>CGCTCC</u>	<u>CTTGAGACG</u>	20
10	E. coli rrnX P1	<u>CGGCTTGTCTT</u> <u>CCTGAGCCGACTCC</u>	<u>TATAATG</u> <u>CGCTCC</u>	<u>ATCGACACG</u>	20
11	E. coli rrnE P1	<u>TTTTCTATTGGCGCTCCGACACCTCTTAAAG</u>	<u>CGCTCC</u>	<u>ATCGACACG</u>	21
12	E. coli rrnA P1	<u>CCTCTTGTCTAG</u> <u>CCCGAATAACTCCC</u>	<u>TATAATG</u> <u>CGCCAC</u>	<u>ACTGACACG</u>	21
13	E. coli rrnB P1	<u>CGCTTGTCTAG</u> <u>CCCGAATAACTCCC</u>	<u>TATAATG</u> <u>CGCC</u>	<u>ACCACTGAC</u>	22, 23
14	E. coli rrnD P2	<u>ATTCAGGGTTGACTCTGAAGACGAAAGCTAAATA</u>	<u>GAGC</u>	<u>ACCTCGCGA</u>	20
15	E. coli rrnX P2	<u>ATTCAGGGTTGACTCTGAAGACGAAAGCTAAATA</u>	<u>GAGC</u>	<u>ACCTCGCGA</u>	20
16	E. coli rrnE P2	<u>ATTCAGGGTTGACTCTGAAGACGAAAGCTAAATA</u>	<u>GAGC</u>	<u>ACCTCGCGA</u>	21

17	E. coli rrnA P2	<u>ATGCTTGACTC</u> <u>TGTAGCGGGAAGCGG</u>	<u>TATTATG</u> <u>CAC</u>	<u>ACCCCGCGC</u>	21
18	E. coli rrnB P2	<u>ATGCTTGACTC</u> <u>TGTAGCGGGAAGCGG</u>	<u>TATTATG</u> <u>CAC</u>	<u>ACCCCGCGC</u>	22, 23
19	pBR322 P4	<u>GTCTTGAAGT</u> <u>GGTGCCTAACTACGGC</u>	<u>TACACTA</u> <u>GAAGGAC</u>	<u>AGTATTTGG</u>	12, 24
20	pRRN 2 p1	<u>CTTCAAGAATTCGGATCCGGAATAACTCCCTATAATG</u>	<u>CGCC</u>	<u>ACCACTGAC</u>	25
21	pRRN 2 p2	<u>ATGCTTGACTC</u> <u>TGTAGCGGGAAGCGG</u>	<u>TATTATG</u> <u>CAC</u>	<u>ACCCCGATC</u>	25
22	pRRN 3 p1	<u>CCTCTTGTCTAG</u> <u>CCCGAATAACTCCC</u>	<u>TATAATG</u> <u>CGCC</u>	<u>ACCACTGAC</u>	25
23	pRRN 3 P2	<u>ATGCTTGACTC</u> <u>TGTAGCGGGAAGCGG</u>	<u>TATTATG</u> <u>CAC</u>	<u>ACCGATCCG</u>	25
24	φ 29 P1	<u>CTAGACAAACTATCCTTTAASCATGTTATATAATA</u>	<u>GAAGT</u>	<u>AAGGTATA</u>	26
25	φ 29 P2	<u>ATGCTTGACAA</u> <u>CTATACAGAGTATGC</u>	<u>TATAATG</u> <u>GTAGT</u>	<u>ATCAATGGT</u>	26
26	λ PL-360	<u>GGTGTTCAT</u> <u>AAAGACCCTAGCGCT</u>	<u>GATCATG</u> <u>AGCAC</u>	<u>ATCAGCAGG</u>	27
27	λ PR	<u>CTTGACTATTTTACCTCTGCGCGT</u>	<u>GATAATG</u> <u>GTTC</u>	<u>ATGTACTAA</u>	28
28	λ PL	<u>GGTGTTCAT</u> <u>AAATACCCTAGCGCT</u>	<u>GATACTG</u> <u>AGCAC</u>	<u>ATCAGCAGG</u>	29
29	Tn3 R	<u>TGAGTCTCCAT</u> <u>TAAATCGTCAATTTGG</u>	<u>CATAATA</u> <u>GACAC</u>	<u>ATCGTCTCT</u>	30
30	pRRN4	<u>ATGTTTGACAG</u> <u>CTTATCATCGATAAGC</u>	<u>TTTAATG</u> <u>CGGT</u>	<u>AGTTTATCA</u>	24
31	Col E1 IOO RNA	<u>GTCTTGAAGT</u> <u>GTTGCCSTAACATCCGC</u>	<u>TACACTA</u> <u>GAAGG</u>	<u>ACAGTATTT</u>	31, 32

В табл. 2-7 приведены нуклеотидные последовательности промоторов (одна нить ДНК в направлении 5'→3'), в которых участки наибольшего соответствия сайтам узнавания и инициации расположены под изображенными сверху обобщенными структурами (R — остаток пурина); обозначено также положение точки старта транскрипции (+1). Элементы симметрии второго порядка подчеркнуты сходящимися стрелками. В некоторых случаях элементы симметрии включают участки за пределами приведенных в таблице последовательностей и соответственно они обозначены лишь частично (либо даже одной стрелкой).

Структуры сайтов инициации промоторов третьей группы хорошо соответствуют обобщенным структурам (среднее значение функции соответствия — 0,76), и длина участка между сайтом инициации и точкой старта лежит в пределах 3-9 п.о. У промоторов четвертой группы наблюдается меньшее подобие в этой области (для сайтов инициации среднее значение функции соответствия — 0,69). В районе сайта узнавания в обеих группах имеется достаточно высокое соответствие рассматриваемой обобщенной структуре (среднее значение функции узнавания в третьей и четвертой группах — 0,74). Среднее значение функции соответствия суммарного

Таблица 3. Нуклеотидные последовательности промоторов группы 2.

Номер промотора	Название гена	TGGTTGACAATTC C C G T T	TATAATR RTAT	+1	Литература
1	λPO	GTTCAGTATTTTGCCTGATTTGT	CATAATG ACTCCT	GTTGATAGA	28
2	λPRM RRL	CTGTTACACAT TTATCCCTTGGCGTC	ATAGATT TAACGT	ATGAGCACA	33
3	λ434 PO	GTCAGIATTTTGCCTGATTTGT	CATAATG ACTCCT	GTCGATAGA	34, 35
4	λ434 PR	GTATTTGACAA ACAAGATACATGGTAT	GAATAA CAAGAA	GTTTGTGGA	34, 35
5	φX174 A	AGGATGGAC CCTGCCAATGTATGT	TTTCATG CCTCC	AAACTTGG	36
6	φX174 B	ATAGCTTGCAA AATACCTGGCCTTAIGGT	TACAGTA TGCCC	ATCGCAGTT	36
7	X174 D	CTTGACNTTTAAAAGAGCGCTGGAITA	CTATCTG AGTCC	GATGCTGTT	36
8	G4 A	CTTCCTAAACTCAACACCAC	TCTAATA GCCTCCC	ATCAAAACGG	37
9	G4 B	TAGGTTGCAAA ACACGGGCGCTTATGGT	TACTCTATG CCC	ATCGCAGTC	37
10	G4 D	ATGGTTGACAT ACTGAAAGAAGCTGGCC	TATTAIC CACATC	CTCAACTGA	37
11	fd X	TTTGATGCAA TCGCTTGGCTTCGAC	TATAAIA GACAG	GGTAAAGAC	38, 39
12	fd II	CTTTACAATTTAAATTTGCTTA	TACAATC ATCCT	GTTTITGGG	38, 39
13	fd VIII	GTTCACCTTGTTCGGCCITGG	TATAATC GCTGG	GGGTCAAAC	38, 39
14	M13 X	TTTGATGCAAT CCGCTTTCGCTTCGAC	TATAATA GTCAG	GGTAAAGAC	39, 40
15	M13 II	GTTCACAATTTAAATATTTGCTTA	TACAATC TTCCT	GTTTITGGG	39, 40
16	M13 VIII	GTTCACCTTGTTCGGCCITGG	TATAATC GCTGG	GGGTCAAAC	39, 40

17	M13-RNA	TTTCATGACGA AGTTCGATTAACCGGG	TAAAATA CGTAAT	GCCACTACG	8, 40
18	pBR322 P3	ATTCAAATATGTATCCGCTCATGA	GACAATA ACCCT	GATAAATGC	24
19	pBR322mpt3	GTTGTGACAT TTTAAGCTTGGCGGT	TATAATC GACCC	ATTATAGCT	41
20	Tn3 A	GAAGCCGTTTACTATGCTGATAATT	TATAATA TTTC	GAACGGTTC	50
21	SV40	GTTGTGTTAA CTGTATTGACGCT	TATAATC GITACA	AAATAAGCA	8
22	E. coli ompA	GTTCACACTTGTAGTTTTCAC	TACCTTC TAGAC	TTTACATCG	42
23	E. coli lacUV5	CTTTACACTTATGCTTCGGCTCG	TATAATC TGTGG	AATTTGAG	8, 13
24	T. coli lacPSA	CTTTACACTTATGCTTCGGCTC	GTAATTC TGTGG	AATTTGAG	43
25	E. coli lacPRIA	CTTTACACTTATGCTTCGGGATC	GTAATTC TGTGG	AATTTGAG	43, 44
26	E. coli lpp	ATTCTCAACAT AAAAACTTGTGTA	ATACTTC TAAC	GCTACATGG	45
27	E. amylovora lpp	TTTATTCCTT TAGAAAACCTTG	TGTAATA CTGTGAC	GCTACATGG	46
28	S. marcescens lpp	ATTATTCCTT TCGAAAACCTTG	TGTAATA CTGTGAC	GCTACATGG	46
29	E. coli tRNA Tyr	CGTACACTTACAGCGCGCGTCAITTCATGATG	CGCCCC	GCTTCGCGA	47
30	E. coli tRNA Leu	ACTATTGACGA AAGCTGAAAACCAC	TAGAATG CCGCTCC	GTGGTAGCA	48
31	E. coli UVRB	TTTGTGGCA ATTAAGTAGGAC	GAGTAA ATTAC	ATACCTGCC	49
32	pN0615	TTTCTTGCAA GTTGGGTGAGCTGCC	TAGATTA GCCA	GCCAACTTT	50
33	pN0613	TTTCTTGCAA GTTGGGTGAGCTGCC	TAGATTA GCCA	GCCAACTTT	50
34	f1 X	TTTGATGCAAT CCGCTTTCCTTCGAC	TATAATA GTCAG	GGTAAAGAC	39, 40
35	f1 II	CTTTACAATTTAAATATTTGCTTA	TACAATC TTCCT	GTTTITGGG	39, 40
36	f1 VIII	GTTCACCTTGTTCGGCCITGG	TATAATC GCTGG	GGGTCAAAG	39, 40
37	E. coli lacIQI	TACGTTGACAC CACCTTCGCTATGG	CAIGATA GCGCC	GGAAGAGAG	51

участка в третьей группе — 0,73, в четвертой группе — 0,70 (см. табл. 1). Длина промежуточного участка между сайтами узнавания и инициации 12—19 п.о.

Таким образом, представляется вероятным, что все промоторы, независимо от их силы, должны обладать характерной структурой сайтов узнавания (со значением функции соответствия указанной обобщенной структуре (не ниже 0,55) и инициации (со значением функции соответствия не ниже 0,57) с ограниченным в пределах 12—19 п.о. промежуточным участком. Значение функции соответствия суммарной структуры должно быть не менее 0,61.

Таблица 4. Нуклеотидные последовательности промоторов группы 3.

Номер промотора	Название гена	TTGTTGACAATTG C C G T T	TATAATR RTAT	+1	Температура		
1	fd IV	ATTCACATATTGACTCTTCTCAGCGCTTTA	ATCTAAG	CTATCGCT	ATGTTTTC	38,59	
2	M13 IV	ATTCACATATTGACTCTTCTCAGCGCTTTA	ATCTAAG	CTATCGCT	ATGTTTTC	39,40	
3	M13 11'	CTACACATTACTCAGGCAATGGCATT	TAAAATA	TATGAG	GGTTCTAAA	39,40	
4	λ C17	TTTATTGGCAT	ACATTCAATCAATTGT	TATAAAT	GTTAT	CTAAGGAAA	52,55
5	λ PRM	GTGTTAGATAT	TTATLCCGTGCGGTG	ATAGATT	TAACGI	ATGAGCACA	54
6	λ PRM 11b	GTAAAATTTTATCCCTTGGCGGTG	ATAGATT	TAACGI	ATGAGCACA	55,54	
7	λ PR'	GATATGCACTT	ATGAATAAAATTGG	GTAATTT	TGACTCA	AGGATGGGT	8
8	λ PE	TTTGTTTGCAC	GAACCATATGTAA	GTAATTC	CTT	AGATAACAA	28,55
9	λ PE 3001	TTTGTTTGCAC	GAACCATATGTAA	GTAATTC	CTT	AGATAACAA	28,55
10	λ PE 3008	TTTGTTTGCAC	GAACCATATGTAA	GTAATTC	CTT	AGATAACAA	28,55
11	λ PE 844	TCTGTTTGCAC	GAACCATATGTAA	GTAATTC	CTT	AGATAACAA	28,55
12	λ PE 42	TTTGTTTGCAC	GAACCATATGTAA	GTAATTC	CTT	AGATAACAA	28,55
13	λ PE 3003	TTTGTTFACAC	GAACCATATGTAA	GTAATTC	CTT	AGATAACAA	28,55
14	λ PL-230	GGTGTGACAT	AAAGACCCTGGCGGT	GATCATG	AGCAC	ATCAGCAGG	27
15	λ PX	CAAAATGATAA	GCAATGCTTTT	TATAATG	CCAA	CTTAGTATA	25
16	pBR322 P1	GTGCGTGCATG	CGTTAGCAATTAAAC	TGTGATA	AAC	TACCGCATAAAGCI	12,24

17	pBR322 tet	ATGTTTGACAG	CITATCATGATAAGC	TTTAAIG	CGGI	AGTTTATCA	12,24
18	pBR322mpt2	AIGTTTGACAG	CITATCATGATAAGC	TATAAIG	GTACC	ATATATAGCT	41
19	Col T1 31 RN'	ATITGTAGGAC	CAAACAAAGTAGA	TTATATG	GCAT	AAATGGGTT	56
20	Col E1 37 RNA	ATITGTAGGAC	CAAACAAAGTAGA	TTATATG	GCAT	AAATGGGTT	56
21	pSC101 tet	ATGTTTGACAG	CITATCATGATAAGC	TTTAAIG	CGGTA	GTTTATCAC	57
22	TK905 EM R	TCTGTGATGTTACATTCGACAACATAA	AAATATA	TCATCATG	AACAATAA		58
23	E. coli lac	CTTTACACTTTATGCTTCCGGCTC	GTATGTT	GTGTGG	AAATGTGAC		15
24	E. coli rp18A	AICGTTGCACA	AGGCGTGAGATIG	GAATACA	ATTTCCG	GGCTTTTGT	59
25	E. coli rp1JL-rpoBC	TGCGTTTACGT	GGGCGGTGATTT	GTCTACA	ATCTTACC	GGCAGGAT	59
26	E. coli araBAD	CGTGACCGTTTATCGCAACTC	TCTACTG	TTCCTCATACC	CGGTTT		8,60,61
27	E. coli recA	CTTGACTCTGATGAGCATACAG	TATAAAT	GGTTCA	ACAGAACAT		62
28	E. coli spe	TTTGTACCCAT	AICGTTGAAGCCGTT	TATAATG	CCGC	GGCTGGAT	63
29	E. coli str	TTTCTTGACAC	CTTTCCGGCATCCGCC	TAAAATY	CGGC	GTCTCTAIA	63,64
30	E. coli araC	CGTGACCGTTTATCGCAACTC	TCTACTA	TTCCTCATACC	CGGTTT		60
31	E. coli rpsM-rpoA	TTTCTTGCAAA	GTGGGCTGAGCTGCC	TAGATA	GCCA	GGCAATCTT	65
32	E. coli bioA	TTCCAAAACGT	GTGTTTCTGTAAATCGGCTGAGA	CTGTAA	ACTTAAATC		8
33	E. coli araC	ATTATAGCAC	TITGCTTACCGGTTT	TGTCATG	GCCTT	GGTCCGGCT	4,61
34	E. coli sroH	CTAETGCCATTAGCTTATTTTGT	TATCAIG	CTAACC	ACCCGCTAT		66
35	E. coli lexA	TATGTTTCAA	ATGCGCTTTCCT	GTATATA	CTCAGAC	ATAACTCTA	67,68,69

То обстоятельство, что между уровнем соответствия двух сайтов обобщенным структурам и эффективностью промоторов не соблюдается строгой зависимости, указывает на наличие дополнительных факторов, влияние которых должно накладываться и изменять минимальный уровень транскрипции. К числу таких факторов, вероятно, относятся структура стартового тринуклеотида от -1 до +2 и длина участка между сайтом инициации и точкой старта транскрипции.

Рассмотрение структур, лишенных промоторной активности (группа 6), показывает, что в этих случаях сильно снижен уровень соответствия идеализированным структурам возможных сайтов узнавания и инициации или (как в случае структур 6.2, 6.5) в промежуточном участке между возможными сайтами располагается вторая структура сайта узнавания с более

Таблица 4 (окончание)

36	Sh.dysent.rev.trp	GCTGTTGACAA TAAATATCGAACAT TAACTA GTACGC AAGTTCAGC	70
37	pBR322 PP	CTTCTTGAGAT CCTTTTTTCTCCCGG TAATCTG CTGCTT GCAAACAAA	12, 24
38	Col E1 imm	CCTCATAACTCGATCCTATAAAAAGAAAAGAAATATA TGGCGAG GTTTAAATT	56
39	T7 C	CAACTTCAGCC AATGTTAATGGGCT GATAGTC TTAATCT ACAGGTCAT	71
40	fd II'	CTACTCATTACTCCGGCATTCGATT TAAAATA TATGAG GGTCTAAA	58, 59
41	Tn3 h1a	CCTATTTTTAT AGGTTAATGTCATGAT AATAATG GTTCTTAGACGGTCAGGI	50
42	E.coli trp	GCTGTTGACAA TAAATCATCGAACAT TAACTA GTACGC AAGTTCAGC	71
43	S.typh.trp	GCTGTTGACAT TATTCCATCGAACAT TAACTA GTACGC AAGTTCAGC	71
44	E.coli bioK	CTTGTAAACCA AATTGAAAAGATTAG GTTTACA AGTC TACACCGAA	8
45	λC1N	ATGATTTGAAT GTATGCAATAAAT GCATACA CTATAGGT GGTCTTAA	8
46	S.melan.trp	AGGGTTGACTT TGGCTTCGGCAACCAG TTACTA GTACAC AAGTTCAGC	8
47	E.coli gal S2 P8-3	TCTCAGCTTTTCGCACTTCTTTTGC TATGTT ATTCA TACATAAG	71
48	pBR322mpt1	CTTGTGACAT TTTAAGCTTGGCAGC TTTAAGC CGGT AGTTTATCA	41
49	E.coli tufB	GTCGGGATGTTCCCTGGTTGATGT GGTCTATA TCAACG ATTTATCG	74
50	E.coli thrU	TTTACTTGCAAT GAAGTCCATGCTTCCA TAGAATG CCGG CTACTGAT	74
51	E.coli trpR	TTACTGATCCGGACGTTTATGATATGC TATGTA CTCTTT AGCGAGTAC	66
52	E.coli araC	AATGHTACTT TTCTGCCCTGATATAGACACTTTTG TTAC CGGTTTTG	60, 61, 65

53	E.coli his	GTTCTTGCTTT TTAACGTGAAAGTG GTTTAGG TTAAGAAGACATCACTGA	76
54	F.coli phoA	ACAGCTGTGAT AAAGTTGTACGGCCGAG ACTTATA GTCCGTTT GTTTTATT	77
55	T7 B	TCACTTACTT ATGAGGGAGTAAAT GTATATG CTTACTATCGGCTACTC	17
56	S.typh.araC	GTATATGACAC TTCTGTACTTAA TTTATC GCCTGAAC TGTACGCTT	78
57	S.typh.araBAD	CCTGACGGTTTTTCCCGCACTC TCTATCG TTTCTCCATACCTGTTT	78
58	E.coli ssu-1	CGTAAACACTTTACAGCGGGCCGTCATTTGATATGATG CGCCCC GCTTCCGA	79
59	E.coli ssu-2	CCTAAACACTTTACAGCGGGCCGTCATTTGATATGATG CGTTAA GCTTCCCGA	79
60	E.coli ssu-3	CCTAAACACTTTACAGCGGGCCGTCATTTGATATATG CGTTAA GCTTCCCGA	79
61	E.coli ssu-4	CCTAAAATTTACAGCGGGCCGTCATTTGATATGATG CGTTAA GCTTCCCGA	79
62	E.coli atp	ATTGTTTCAA TCAAGCGGGCCACCG TATAAT TCAAC GCTTTTTCA	80
63	E.coli rpsB P1	TTTGCCAAAC GTGCCACTGAAGBTTTTC TATAATA GAAA ATTCCAGCT	81
64	pBR325 Ptet	ATGTTTGACAG CTTATCATCGATAAGC TTTAATG CGGT AGTTTATCA	82
65	λrexB	CAGGTCGGCAA AGTTAAGGATTAAT TATCAGG AGTAAT ATCGGGAAC	83
66	E.coli rpsB P2	CGGCACATATTCGGGGTCCCTTTGGC GTGGGA ATATGGGATACGGGAGC	84
67	fd IV	ATTCASATATGACTCTTCTCAGCCTCTTA ATCTAAG CTATCGCT ATGTTTAA	59, 10
68	fd II'	CTACACATTAAGCATTCGATT TAAAATA TATGAG GGTCTAAA	59, 10
69	E.coli colicin E1 P1	CTTGCAGGGGAAAATGCAGCCCGTA GCTTTTA TGGT GTATAAAA	84
70	E.coli colicin E1 P2	AATGTTTTAA AGTCAAGAGCAATT TATAATG GAAACC CGGCTTCCG	84

высоким значением функции соответствия, чем первая. Последнее обстоятельство, вероятно, приводит к лучшему связыванию РНК-полимеразы с этим вторым участком и, как следствие, к существенному сокращению расстояния до сайта инициации, который поэтому не может функционировать.

2. Распределение А·Т-богатых участков

Распределение А·Т-пар в исследованных структурах было изучено с помощью программы DISTR. В этой программе за основу берется значение содержания А·Т-пар, вычисленное на отрезке ДНК длиной 10 п.о.

А·Т-богатыми мы считали те участки, в которых содержание А·Т-пар составляло не менее 70%. При этом оказалось, что в структуре самих промоторов А·Т-богатыми, как правило, являются лишь короткие участки в сайтах узнавания и инициации. Лишь в некоторых промоторах различ-

Таблица 5. Нуклеотидные последовательности промоторов группы 4.

Номер промото- ра	Название гена	TTGTTGACAATTG - C C C T T	TATAATR RTAT	+1	Литература
1	fd V	AGATTTTTCCTCCCAAGCTCCTGACTGG	TATAATG AGCCA	GTCTTAAA	38, 39
2	fd III	ATTACCTCGA AAGCAAGCTGATAAACCGATACAATT	AAAGGC	TCCTTTGG	38, 39
3	M13 V	AGATTTTTCCTCCCAAGCTCCTGACTGG	TATAATG AGCCA	GTCTTAAA	39, 40
4	M13 III	ATTACCTCGA AAGCAAGCTGATAAACCGATACAATT	AAAGGC	TCCTTTGG	39, 40
5	λ C17 SCS10	TTTATTGGCAT ACATTCAATCAATTGTT	ATAAATG TTAT	CTAAGGAAA	52, 53
6	λ C17 SCS20	TTTATTGGCAT ACATTCAATCAATTGTT	GTAAATG TTAT	CTAAGGAAA	52, 53
7	λ PE 3019	TTTGTTGGCAC GAACCATATATAA	GTATTTC CTT	AGATAACAA	28, 55
8	λ PRM 11	GTTAAATATTATCCCTTGGCGTG	ATAGATT TAACGT	ATGAGCACA	33, 54
9	λ PI	TTGCGTGTAAAT TGGCGAGACTTGGCGAT	GTACTTG ACACT	TCAGGAGTG	85
10	pBR313-H tet	TTTGACAGCTTATCATCGATAAGCTAGC	TTTAAATG CCGTA	GTTTATCAC	57
11	E. coli gal S1	TGTCACACTTTTGGCACTTTTGT	TATGCTA TGGTTA	TTTCATACC	86
12	E. coli gal S2	TGTCACACTTTTGGCACTTTTGTAAAGC	TATGCTT ATTTCA	TACCATAAG	86
13	E. coli lac I	AATGCGCAAA ACCTTTCGCGGTATGG	CATGATA GCGCCC	GGAAGAGAG	87
14	E. coli lac IQ	AATGCTGCAAA ACCTTTCGCGGTATGG	CATGATA GCGCCC	GGAAGAGAG	8, 87
15	E. coli LAC PIIS	CTTCCGGCTCGTATGTTGTGTGG	TATTGTG AGCGG	ATAACAATT	8

16	E. coli lac L157	CTTAAACCTTTATGCTTCGGGCTC	GTATGTT GTGTGG	AATGTGAG	43
17	E. coli lac L241	CTTAAACCTTTATGCTTCGGGCTC	GTATGTT GTGTGG	AATGTGAG	43, 73
18	E. coli lac L305	CTTAAACCTTTATGCTTCGGGCTC	GTATGTT GTGTGG	AATGTGAG	43, 14
19	E. coli str P105	TTTCTTGACAC CTTCTGGCAATGG	CTTAAA CTCGGC	GTCTTATA	88
20	E. coli bio P98	GGTGTAGACTT GTAAACCTAAATCTTTT	AAATTG GTTT	ACAAGTCGA	8
21	E. coli protein A	CTAATTAATAA ATAGTAAATTAACGCTC	ATCATTG TACAATG	AATGTATA	76
22	pD01 trp M1	GGTGTGACAT TATTCCATCGAACTAGT	TAACTCA GCGTC	CGCCGCGGT	89
23	pD01 trp MII	GGTGTGACAT TATTCCATCGAACTAGT	TAACTCA GCGTC	CGCCGCGGT	89
24	Sh. dysent. trp	CCTGTGACAA TTAATATCGAACTAC	TAACTCA GTACGC	AAGTTCACG	70
25	E. coli hca1C	CCTGGCGCTTTTATCGCAAC1C	TCTACTG TTTCTCCATACCCGTTT		22
26	pBRH3B tet	TTTGACAGCTTATCATCGATAAGCTAGC	TTTAAATG CCGTA	GTTTATCAC	51
27	λ PE 2001	TTTGTTTGAC GAACCATATGTAC	GTATTTC CTT	AGATAACAA	28, 55
28	λ PE 3048	TTTGTTTGAC GAACCATATGTAA	GTACTTC CTT	AGATAACAA	28, 55
29	f1 V	AGATTTTTCCTCCCAAGCTCCTGACTGG	TATAATG AGCCA	GTCTTAAA	39, 40
30	f1 III	ATTACCTCGA AAGCAAGCTGATAAACCGATACAATT	AAAGGC	TCCTTTGG	39, 40

ной эффективности (1.2, 1.3, 1.4, 1.29; 2.1, 2.12, 2.15, 2.28, 2.27; 3.4, 3.41; 4.5, 4.6, 4.7; 5.1) А·Т-богатые участки более протяженны и охватывают почти всю структуру промотора. В то же время почти у всех промоторов первой группы непосредственно перед сайтом узнавания имеется протяженный (15–25 п.о.) А·Т-богатый участок. Лишь в промоторах 1.7, 1.8, 1.19 такого участка не оказалось, а в промоторах 1.29, 1.30 А·Т-богатый участок отстоит от сайта узнавания больше чем на 20 п.о. Во второй группе промоторов аналогичный А·Т-богатый участок в большинстве случаев имеется у тех промоторов, в которых стартовым нуклеотидом является не А, а G, С или Т; в противном случае А·Т-богатый участок либо невелик (2.2), либо отстоит от сайта узнавания больше чем на 20 п.о. (промоторы 2.24, 2.25, 2.23). Лишь в шести промоторах этой группы (2.11, 2.13, 2.14, 2.16, 2.20, 2.30) со стартовым нуклеотидом G и в одном (2.22) со стартовым нуклеотидом T перед сайтом узнавания имеются недостаточно протяженные А·Т-богатые участки. У подавляющего большинства промоторов третьей и четвертой групп в предпромоторной области отсутствуют совсем либо имеются лишь небольшие А·Т-богатые участки.

Таблица 6. Нуклеотидные последовательности промоторов группы 5.

Номер промотора	Название гена	TTGTTGACAATG Г С Г Г Т	TATAATR RTAT	Литература
1	fd VI	CATATGAATTTCTATTGATTCGACACA	AAATAAA CTATTCGGTGG	38, 39
2	fd IV'	GTTCGGGCAAA GGATTAAATAAGGCTTG	TAGAAAT GTTGTAAATC	38, 39
3	fd I	CGTCTAAIGCGCTGCCGTCTTTTAAIGI	TATTCTC TCTGTAAAGGCT	38, 39
4	fd I'	GCTCTGGAAAG ACGCTGCTTAGCGTTGG	TAAGATT CAGGATAAATGT	38, 39
5	M13 VI	CATATGAATTTCTATIGATTCGACACA	AAATAAA CTATTCGGTGG	39, 40
6	M13 I	CTTCCCCTCTTTTATGTTAATCTCTCT	GTAAAGG CTGCIATTTTCA	39, 40
7	M13 IV	GTTCGGGCAAA GCAATTAAACGAGGTIG	TCGAATT GTTGTAAAGTC	38, 39
8	M13 I'	GCTCTGGAAAG ACGCTGCTTAGCGTIGG	TAAGATT CAGGATAAATGT	38, 39
9	f1 VI	CATATGAATTTCTATIGATTCGACACA	AAATAAA CTATTCGGTGG	38, 39
10	f1 IV'	GTTCGGGCAAA GCAATTAAACGAGGTIG	TAGAAAT GTTGTAAATC	38, 39
11	f1 I	CGTCTAAIGCGCTGCCGTCTTTTAAIGI	TATTCTC TCTGTAAAGGCT	38, 39
12	f1 I'	GCTCTGGAAAG ACGCTGCTTAGCGTIGG	TAAGATT CAGGATAAATGT	38, 39
13	E. coli Gly I	TCTCTCGATAI TCAGTGCAGAAATGAA	AATCAGG TAGCCGAGTTCC	73
14	E. coli P ORF II	CTACAAGCCAT CCCCCACAGA	TACGGTA AACTAGCCTCGT	25

Таблица 7. Нуклеотидные последовательности промоторов группы 6.

Номер промотора	Название гена	TTGTTGACAATG С С Г Г Т	TATAATR RTAT	Литература
1	PBRH tet	CAGCTTATCAI CGATAAGCTGAATTCAGC	TATAATG CGGTAGTTFATCAC	57
2	PBRH tet	CGAGGCCCTTTCGCTTCAAGAAATC	ATAATG CGGAGTTFATCAC	57
3	E. coli gal SI. P8-5	TGTCACACTTTTCGCACTTIGT	TITGCTA TGGTATTTCAATAC	73
4	PBO1 trp	GGTGTGACAI TATTCATCGAAXTAGT	TAAATCA GCTGGGCCCGGCT	89
5	PBRH2 tet	CCTTTGGCTT CAAGAAATCAGC	TATAATG CGGTAGTTFATCAC	57
6	PRO617	ATTGGTCCGCC CAGAAGTACGGAGA	GTAAAAA TGAAGATTCGAGCT	50

Таким образом, представляется вероятным, что А·Т-богатая нуклеотидная последовательность существенно увеличивает эффективность промоторов лишь в том случае, если она расположена перед промотором, в непосредственной близости к сайту узнавания, и ее протяженность достаточно велика (порядка 15 п.о. и более).

3. Предшествующие промотору участки связывания РНК-полимеразы и целые структуры промоторов

С использованием программ PROM1, PROM2, PROM3 нами было исследовано наличие в предшествующей промотору области дополнительных сайтов узнавания РНК-полимеразы или целых структур промоторов.

За целую структуру промотора мы принимали совокупность структур сайтов узнавания и инициации, найденных в разделе 1, паходящихся на расстоянии 12—19 п.о. друг от друга, при условии, что значение функций соответствия обобщенным структурам должно быть не ниже средних значений функций соответствия по четвертой группе промоторов. Таким образом, при исследовании использовались значения функции соответствия 0,71 (район узнавания), 0,69 (район инициации) и 0,70 (целая структура промотора). Исследованию подвергался район ДНК от положения -45 до -120 ÷ -220. При этом статистической обработке были подвергнуты нуклеотидные последовательности с 23 промоторами первой группы, 26 —

второй, 56 — третьей и 25 — четвертой группы. Для этих районов ДНК были опубликованы достаточно протяженные для исследования структуры; в остальных случаях отсутствовали данные о протяженных предпромоторных структурах.

Анализ показал, что в участках, предшествующих промоторам первой и второй групп, как правило (у 97% структур), имеются целые промоторные структуры или множественные (два и более) сайты узнавания (кроме структур 1.9, где не обнаружено ни целых структур, ни сайтов узнавания с принятым уровнем значения функции соответствия, а также 1.19 и 1.31, где имеется только по одному сайту узнавания).

Среди промоторов третьей группы 25% последовательностей не имеют предшествующих целых структур промотора либо множественных сайтов узнавания, а в четвертой группе такие последовательности составляют 40%.

Таким образом, можно сказать, что предшествующие реально действующему промотору сайты узнавания РНК-полимеразы или целые промоторные структуры могут служить фактором, усиливающим эффективность промотора. Эти результаты соответствуют представлению о латеральном перемещении РНК-полимеразы при поиске промотора и о влиянии, которое оказывает локальная концентрация фермента вблизи промотора на эффективность его работы.

4. Распределение элементов симметрии второго порядка

Распределение в структуре промоторов обращенных повторов (палиндромов), в том числе несовершенных, было изучено по программе COMPL. При составлении этой программы мы учитывали обращенные повторы не короче тринуклеотидных. Обнаруженные при этом элементы симметрии второго порядка отмечены в табл. 2—5.

Лишь в редких случаях единый обращенный повтор охватывает всю структуру промотора (промоторы 2.24, 2.25, 2.23, 3.24, 3.25, 3.26, 3.27, 3.31, 3.69, 4.5, 4.6, 4.8, 4.20). Большинство же промоторов характеризуется тремя группами обращенных повторов. Первая группа таких элементов симметрии располагается в районе старта транскрипции, с центром вблизи точки старта, между положениями -12 и $+3$ (70% промоторов группы 1, 73% — группы 2, 72% — 3, 66% — 4). Второй обращенный повтор располагается так, что хотя бы одна из его ветвей составляет часть сайта узнавания (90% промоторов группы 1, 75% — группы 2, 85% — 3, 66% — 4). Наконец, третий обращенный повтор захватывает промежуточный участок между сайтами узнавания и инициации. При этом промоторы, в структуре которых не имеется такого элемента симметрии, как правило, содержат вместо него А·Т-богатый участок (исключения составляют промоторы 1.23, 2.6, 2.9, 3.14, 3.52, 4.9, 4.15).

Такое распределение элементов симметрии говорит о том, что их возможная роль в структуре промоторов заключается в снижении энергетического барьера разъединения нитей ДНК. Это согласуется с экспериментально обнаруженным участком с разделенными нитями ДНК в районе старта транскрипции [8], но также указывает на возможное расплетание ДНК и в других участках промотора, в особенности вблизи сайта узнавания. Следует отметить, что по всей длине промотора, помимо уже отмеченных, наблюдаются также многочисленные перекрывающиеся короткие (трехнуклеотидные) элементы симметрии. Эта особенность структуры позволяет предположить образование «скользящей шпильки» при перемещении РНК-полимеразы вдоль промотора.

Еще один обращенный повтор, как правило, содержится непосредственно за точкой старта. Центр его располагается в районе между положениями $+5$ и $+20$. Можно предположить, что указанный обращенный повтор (и/или вторичная структура 5'-концевого участка мРНК) является сигналом для диссоциации σ -субъединицы РНК-полимеразы.

Таким образом, между структурой промотора и его эффективностью нет однозначной зависимости. Промотор должен обладать минимальным

соответствием найденной обобщенной структуре (ундекануклеотид в сайте узнавания и гептануклеотид в сайте инициации, разделенные участком 12—19 п.о.), что обеспечивает минимальный уровень его эффективности. Усиливают промотор дополнительные структурные факторы: нахождение на оптимальном расстоянии за сайтом инициации и предпочтительное строение стартового триплексида (от -1 до +2), протяженный А·Т-богатый участок непосредственно перед промотором, а также предшествующие промотору сайты узнавания РНК-полимеразы и целые промоторные структуры.

ЛИТЕРАТУРА

1. Никуфоров В. Г., Зограф Ю. Н. Итоги науки и техники. Молекулярная биология, 1977, т. 13.
2. Gilbert W. In: RNA polymerase/Eds Losick R., Chamberlin M. N. Y.: Acad. Press, 1976, p. 193—205.
3. Pribnow D. Proc. Nat. Acad. Sci. USA, 1975, v. 72, № 3, p. 784—788.
4. Seeburg P. H., Nusslein C., Schaller H. Eur. J. Biochem., 1977, v. 74, № 1, p. 107—113.
5. Scherer G. E. F., Walkinshaw M. D., Arnott S. Nucl. Acids Res., 1978, v. 5, № 10, p. 3759—3773.
6. Siebenlist U. Nucl. Acids Res., 1979, v. 6, № 5, p. 1895—1907.
7. Rosenberg M., Court D. Ann. Rev. Genet., 1979, v. 13, p. 319—353.
8. Siebenlist U., Simpson R. B., Gilbert W. Cell, 1980, v. 20, № 2, p. 268—281.
9. Johnsrud L. Proc. Nat. Acad. Sci. USA, 1978, v. 75, № 11, p. 5314—5318.
10. Simpson R. B. Cell, 1979, v. 18, № 2, p. 277—285.
11. Von Gabain A., Bujard H. Proc. Nat. Acad. Sci. USA, 1979, v. 76, № 1, p. 189—193.
12. Stuber D., Bujard H. Proc. Nat. Acad. Sci. USA, 1981, v. 78, № 1, p. 167—171.
13. Bujard H., Niemann A., Breunig K., Roisch U., Dresel A., von Cabain A., Gentz R., Stuber D., Weiher H. In: Promoter Structure and Function/Eds Rodriguez R., Chamberlin M. N. Y.: Praeger Press, 1982.
14. Шарский И. И. В кн.: Итоги науки и техники. Сер. Биоорганическая химия, т. 15, Структура и функции рибосом. М., ВИНТИ, 1981, с. 3—37.
15. Vollenweider H. J., Fiantl M., Szybalski W. Science, 1979, v. 205, № 4405, p. 508—511.
16. Lund E., Dalberg J. E. Proc. Nat. Acad. Sci. USA, 1979, v. 76, № 11, p. 5480—5484.
17. Dunn J. J., Studier F. W. J. Mol. Biol., 1981, v. 148, № 4, p. 303—330.
18. McConnell D. J. Nucl. Acids Res., 1979, v. 6, № 11, p. 3491—3503.
19. Siebenlist U. Nucl. Acids Res., 1979, v. 6, № 5, p. 1895—1907.
20. Young R. A., Steitz J. A. Cell, 1979, v. 18, № 1, p. 225—234.
21. De Boer H. A., Gilbert S. F., Nomura M. Cell, 1979, v. 17, № 1, p. 201—209.
22. Csordas-Toth E., Boros I., Venetianer P. Nucl. Acids Res., 1979, v. 7, № 8, p. 2189—2197.
23. Brosius J., Dull T. J., Sleeter D. D., Noller H. F. J. Mol. Biol., 1981, v. 148, № 2, p. 107—127.
24. Sutcliffe J. G. S. H. Symp. Quant. Biol., 1978, v. 43, p. 77—90.
25. Коробко В. Г., Добрынин В. Н., Чувило С. А., Северцова И. В., Шингарова Л. Н., Колосов М. И. Биоорганическая химия, 1981, т. 7, № 2, с. 309—312.
26. Yoshikawa H., Friedmann T., Ito J. Proc. Nat. Acad. Sci. USA, 1981, v. 78, № 3, p. 1336—1340.
27. Horn G. T., Wells R. D. J. Biol. Chem., 1981, v. 256, № 4, p. 1998—2002.
28. Schwarz E., Scherer G., Hobom G., Kossel H. Nature, 1978, v. 272, № 5644, p. 410—428.
29. Maniatis T., Ptashne M., Barrell B. G., Donelson J. Nature, 1974, v. 250, № 5465, p. 394—397.
30. Heffron F., McCarthy B. J., Ohtsubo H., Ohtsubo E. Cell, 1979, v. 18, № 4, p. 1153—1163.
31. Backman K., Bellach M., Boyer H. W., Yanofsky S. S. H. Symp. Quant. Biol., 1978, v. 43, p. 69—76.
32. Hashimoto-Gotoh T., Timmis K. N. Cell, 1981, v. 23, № 1, p. 229—238.
33. Meyer B. J., Kleid D. G., Ptashne M. Proc. Nat. Acad. Sci. USA, 1975, v. 72, № 12, p. 4785—4789.
34. Grosschedl R., Schwarz E. Nucl. Acids Res., 1979, v. 6, № 3, p. 867—881.
35. Ovchinnikov Yu. A., Guryev S. O., Krayev A. S., Monastyrskaya G. S., Skryabin K. G., Sverdlov E. D., Zakharjev V. M., Bayev A. A. Gene, 1979, v. 6, № 3, p. 235—249.
36. Sanger F., Coulson A. R., Friedmann T., Air G. M., Barrell B. G., Brown N. L., Fiddes J. C., Hutchison III C. A., Slocombe P. M., Smith M. J. Mol. Biol., 1978, v. 125, № 2, p. 225—246.
37. Godson G. N., Barrell B. G., Staden R., Eiddes J. C. Nature, 1978, v. 276, № 5685, p. 236—247.
38. Schaller H., Beck E., Takamami M. In: The Single-Stranded DNA Phages. Cold Spring Harbor, 1978, p. 139—163.
39. Beck E., Zink B. Gene, 1981, v. 16, № 1, p. 35—58.

40. *Wezenbeek P. M. G. F., Hulsebos T. J. M., Schoenmakers J. G. G.* Gene, 1980, v. 11, № 1/2, p. 129-148.
41. *Коробко В. Г., Добрынин В. Н., Северцова И. В., Чуевило С. А., Шингарова Л. Н., Колосов М. Н.* Биоорган. химия, 1980, т. 6, № 11, с. 1743-1745.
42. *Mouva N. R., Nakamura K., Inouye M. J.* Mol. Biol., 1980, v. 143, № 3, p. 317-328.
43. *Resnikoff W. S., Abelson J. N.* In: The Operon / Eds Miller J. H., Resnikoff W. S. Cold Spring Harbor, 1978, p. 221-243.
44. *Dickson R. S., Abelson J., Barnes W. M., Resnikoff W. S.* Science, 1975, v. 187, № 4171, p. 27-35.
45. *Nakamura K., Inouye M.* Cell, 1979, v. 18, № 4, p. 1109-1117.
46. *Yamagata H., Nakamura K., Inouye M. J.* Biol. Chem., 1981, v. 256, № 5, p. 2194-2198.
47. *Berman M. L., Lundy A.* Proc. Nat. Acad. Sci. USA, 1979, v. 76, № 9, p. 4303-4307.
48. *Duester G., Campen R. K., Holmes W. M.* Nucl. Acids Res., 1981, v. 9, № 9, p. 2121-2139.
49. *Van den Berg E., Zwetslott J., Noordermeer I., Pannekoek H., Dekker B., Dijkema R., van Ormondt H.* Nucl. Acids Res., 1981, v. 9, № 21, p. 5623-5643.
50. *Miura A., Klueger J. H., Itoh S., de Boer H. A., Nomura M.* Cell, 1981, v. 25, № 3, p. 773-782.
51. *Calos M. P., Miller J. H.* Mol. Gen. Genet., 1981, v. 183, № 3, p. 559-560.
52. *Rosenberg M., Court D., Shimatake H., Brady C., Wulff D.* In: The Operon / Eds Miller J. H., Resnikoff W. S. Cold Spring Harbor, 1978, p. 345-354.
53. *Mozola M. A., Friedman D. I., Crawford C. L., Wulff D. L., Shimatake H., Rosenberg M.* Proc. Nat. Acad. Sci. USA, 1979, v. 76, № 6, p. 1122-1125.
54. *Smith G. R., Eisen H., Reichardt L., Hedgpeth J.* Proc. Nat. Acad. Sci. USA, 1976, v. 73, № 3, p. 712-716.
55. *Schmeissner U., Court D., Shimatake H., Rosenberg M.* Proc. Nat. Acad. Sci. USA, 1980, v. 77, № 6, p. 3191-3195.
56. *Patil R. K.* Nucl. Acids Res., 1979, v. 6, № 8, p. 2647-2665.
57. *Rodriguez R. L., West R. W., Hegneker H. L., Bolivar P., Boger H. W.* Nucl. Acids Res., 1979, v. 6, № 10, p. 3267-3287.
58. *Grindley N. D. F., Joyce C. M.* Proc. Nat. Acad. Sci. USA, 1980, v. 77, № 12, p. 7176-7180.
59. *Post L. E., Strycharz G. D., Nomura M., Lewis H., Dennis P. P.* Proc. Nat. Acad. Sci. USA, 1979, v. 76, № 4, p. 1697-1701.
60. *Horwitz A. H., Morandi C., Wilcox G. J.* Bacteriol., 1980, v. 142, № 2, p. 659-667.
61. *Ogden S., Haggerty D., Stoner C. M., Kolodrubetz D., Schleif R.* Proc. Nat. Acad. Sci. USA, 1980, v. 77, № 6, p. 3346-3350.
62. *Sancar A., Stachelek C., Konigsberg W., Rupp W. D.* Proc. Nat. Acad. Sci. USA, 1980, v. 77, № 5, p. 2611-2615.
63. *Post L. E., Arfsten A. E., Reusser F., Nomura M.* Cell, 1978, v. 15, № 2, p. 215-229.
64. *Post L. E., Nomura M. J.* Biol. Chem., 1980, v. 255, № 10, p. 4660-4666.
65. *Post L. E., Arfsten A. E., Davis G. R., Nomura M. J.* Biol. Chem., 1980, v. 255, № 10, p. 4653-4659.
66. *Gonsalus R. P., Yanofsky Ch.* Proc. Nat. Acad. Sci. USA, 1980, v. 77, № 12, p. 7117-7121.
67. *Horii I., Ogawa I., Ogawa H.* Cell, 1981, v. 23, № 3, p. 689-697.
68. *Miki T., Ebina Y., Kishi F., Nakazawa A.* Nucl. Acids Res., 1981, v. 9, № 3, p. 529-543.
69. *Markham B. E., Little J. W., Mount D. W.* Nucl. Acids Res., 1981, v. 9, № 16, p. 4149-4161.
70. *Miozzari G. F., Yanofsky C.* Proc. Nat. Acad. Sci. USA, 1978, v. 75, № 11, p. 5580-5584.
71. *McConnell D. J.* Nucl. Acids Res., 1979, v. 6, № 2, p. 525-544.
72. *Lee F., Bertrand K., Bennett G., Yanofsky C. J.* Mol. Biol., 1978, v. 121, № 2, p. 193-217.
73. *Musso R. E., di Lauro R., Adhya S., Crombrughe R.* Cell, 1977, v. 12, № 3, p. 847-854.
74. *Ann G., Friesen J. D.* Gene, 1980, v. 12, № 1/2, p. 33-39.
75. *Wallace R. G., Lee N., Fowler A. V.* Gene, 1980, v. 12, № 3/4, p. 179-190.
76. *Verde P., Frunzio R., di Nocera P. P., Blast F., Bduni C. B.* Nucl. Acids Res., 1981, v. 9, № 9, p. 2075-2086.
77. *Kikuchi Y., Yoda K., Yamasaki M., Tamura G.* Nucl. Acid. Res., 1981, v. 9, № 21, p. 5673-5678.
78. *Horwitz A. H., Heffernan L., Morandi C., Lee J.-H., Timko J., Wilcox G.* Gene, 1981, v. 14, № 4, p. 309-319.
79. *Dunn R. J., Belagaje R., Brown E. L., Khorana H. G. J.* Biol. Chem., 1981, v. 256, № 12, p. 6109-6118.
80. *Gay N. J., Walker J. E.* Nucl. Acids Res., 1981, v. 9, № 16, p. 3919-3926.
81. *An G., Bendiak D. S., Mamelak L. A., Friesen J. D.* Nucl. Acids Res., 1981, v. 9, № 16, p. 4163-4172.
82. *Prentki P., Karch E., Iida S., Meyer J.* Gene, 1981, v. 14, № 4, p. 289-299.
83. *Pirrotta V., Ineichen K., Walz A.* Mol. Gen. Genet., 1980, v. 180, p. 369-376.
84. *Ebina Y., Kishi F., Miki T., Kagamiyama H., Nakazawa T., Nakazawa A.* Gene, 1981, v. 15, № 2, p. 119-126.

85. *Abraham J., Mascarenhas D., Fischer R., Benedik M., Campbell A., Echols H.* Proc. Nat. Acad. Sci. USA, 1980, v. 77, № 5, p. 2477-2481.
86. *Musso R. E., di Lauro R., Rosenberg M., Crombrughe B.* Proc. Nat. Acad. Sci. USA, 1976, v. 74, № 1, p. 106-110.
87. *Calos M.* Nature, 1978, v. 274, № 5673, p. 762-765.
88. *Post L. E., Arjsten A. E., Momura M., Joskunus S. R.* Cell, 1978, v. 15, № 1, p. 231-236.
89. *Oppenheim D. S., Yanofsky Ch.* J. Mol. Biol., 1980, v. 144, № 2, p. 143-161.

Поступила в редакцию
21.I.1983

STATISTIC ANALYSIS OF A PROCARYOTIC PROMOTER NUCLEOTIDE SEQUENCE
STRUCTURAL ELEMENTS UNDERLYING THE EFFICIENCY OF THE
TRANSCRIPTION
INITIATION STAGES

ARTEMYEV I. V., VASILYEV G. V., GUREVIRH A. I.

*M. M. Shemyakin Institute of Bioorganic Chemistry, Academy of Sciences
of the USSR, Moscow*

Nucleotide sequences of 188 promoter-containing DNA regions have been studied by the computer statistic analysis. Undecanucleotide NTT(G/C)TTGACA(A/T) or (G/C)·TT(G/C)A(G/C)A(A/T)TT(G/T) (recognition site) and heptanucleotide RTATATR or TATAATR (initiation site) separated by 12-19 base pairs are characteristic of a «generalized» promoter structure. Promoters can function if a minimal level of correspondence for their recognition and initiation sites to a generalized structure is attained (the correspondence function value for the whole structure is not lower than 0.61; for the most effective promoters it may be equal to 1). The transcription start is situated 3-9 base pairs after initiation site, 4-7 pairs distance being the most effective. Transcription can start from any nucleotide, preferably with A or G. The start from A is the most effective if it is contained within the CAC or CAT trinucleotides. The promoter efficiency is enhanced by some additional structural factors: the presence of an extended A-T rich region directly before the recognition site; availability of integral promoter structures or several RNA polymerase binding sites in the preceding nucleotide sequence. A characteristic feature of the promoter is the presence of either the dyadic axial symmetry elements in the initiation and recognition sites as well as in the intermediate region, or the A-T rich area in the latter.