



УДК 547.963.32.02

СТРУКТУРА УЧАСТКА ТЕРМИНАЦИИ ТРАНСКРИПЦИИ
РНК-ПОЛИМЕРАЗЫ *E. COLI**Васильев Г. В., Гуревич А. И., Малина З. А.,
Артемов И. В.**Институт биоорганической химии им. М. М. Шемякина
Академии наук СССР, Москва*

На основе статистического анализа 17 фрагментов ДНК с помощью ЭВМ установлена структура участка терминации транскрипции РНК-полимеразой *E. coli* (ρ -независимый терминатор). Первым признаком этой структуры является нуклеотидная последовательность $GGGG\left(\frac{C}{T}\right)TTTTTTTT$, предшествующая точке терминации. Эта нуклеотидная последовательность, имеющая вероятностный характер, использована для математического поиска расчетных точек терминации по критерию максимума функции узнавания β . Вторым признаком участка терминации — наличие слева от расчетной точки терминации на расстоянии 4–10 нуклеотидных остатков элементов осевой симметрии второго порядка. Участок симметрии (палиндром) должен, таким образом, включать левую часть нуклеотидной последовательности терминатора.

Среди факторов, вызывающих терминацию транскрипции РНК-полимеразой *E. coli*, согласно современным представлениям [1–3], определяющим является структура участка ДНК-матрицы или РНК-транскрипта вблизи точки терминации. Определенная нуклеотидная последовательность (терминатор) может служить сигналом окончания транскрипции в отсутствие дополнительных факторов (ρ -независимые терминаторы), хотя эффективность терминации вблизи той же точки увеличивается в присутствии дополнительного белкового фактора (ρ -фактора) [4]. Этот ρ -фактор способен также вызывать терминацию транскрипции на других участках ДНК-матрицы (ρ -зависимые терминаторы), однако небольшое число таких участков с установленной структурой не позволяет в настоящее время провести статистическую обработку с получением надежных результатов. Имеющиеся данные свидетельствуют о том, что ρ -фактор, по-видимому, осуществляет отделение РНК от РНК-полимеразы в терминаторном комплексе или при паузе в транскрипции [5]. Эффективность терминации транскрипции определяется не только наличием или отсутствием дополнительных факторов, но и непосредственно структурой терминатора.

В ряде случаев на терминаторе оканчивается лишь часть транскриптов (например, на одном из двух терминаторов ранних генов фага Т3 [1]). Одним из механизмов регуляции экспрессии некоторых оперонов служит регулируемая частичная терминация транскрипции в особых участках ДНК — так называемых аттенуаторах, которые могут вызывать терминацию до 95% транскриптов [1]. Поэтому для выявления характерных особенностей структуры терминатора мы использовали также известные нуклеотидные последовательности участков аттенуаторов.

С х е м а

	60	50	40	30	20	10	1	-10	
1	AAAGAGTTGGTAGCTCTTGATCCGGCAAACAACCCACCGCTGGTAGCGGTGGTTTTTTTGT	TTGCAAGCA					t		(840)
2	TGTGAACAGAATGCGAAGACGAACAATAAAGGCCCTCCAAATCGGGGGCCTTTTTTATTGATAACAAA						t		(157)
3	TACTTGCCGTAACACAGTGTGATAAGGGCCTTTGAAGTTACCGCTTCAAGGGCTTTTTAA						t		(62)
4	AATCCCGAAACAGTTCGCGAGGTAATAGTTAGAGCCTGCATAACGGTTTCGGGATTTTTATATCTGCACA						t		(203)
5	TGACCTCAGAACTCCATCTGGATTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTTGGTGAGAA						t		(87)
6	ACCTCGAAAGCAAGCTGATAAACCGGATACAATTAAGGCTCCTTTGGAGCCTTTTTTTGGAGATTTT						t		(274)
7	CCGATGAAAGCGGGCACGCGCAGTTAATCCACAGCCGCCAGTTCCGCTGGCGGCATTTAACTTTCTTT						t		(146)
8	CAGTGGTGCATGAACGCATGAGAAAGCCCCGGAAGATCACCTCCGGGGGCTTTATAATAATTAGCGCGG						t		(160)
9	CCAGTGGTGCATGAACGCATGAGAAAGCCCCGGAAGATCATCTCCGGGGGCTTTTTTTTGGCGCGCG						t		(216)
10	GTATTCACCATGCGTAAAGCAATCAGATACCCAGCCGCCAATGAGCGGGCTTTTTTTTGAACAAAATT						t		(151)
11	GTATGAACAGCTGTAATCAGCCAAACGATACCCGGCCCGCCTGTTAAGCGTGGCTTTTTTGAACAAAATA						t		(152)
12	CAGGATTGACACCCTCCCAATTGTATGTTTTCATGCCTCCAATCTTGGAGGCTTTTTTATGGTTGCTTC						t		(113)
13	GTAATACAGTACACTCTCTCAATACGAATAAACGGCTCAGAAATGAGCCGTTATTTTTTCTACCCATAT						t		(142)
14	TGACACCTTTTCGGCATCGCCCTAAAATTCGGCGTCTCATATTGTGTGAGGACGTTTTATTACGTGTTT						t		(270)
15	GGCTGACGCGTACAGGAAACACAGAAAAAGCCCGCACCTGCACAGTGGCGGGCTTTTTTTTTCGACCAAAG						t		(161)
16	CGGTTGCAGCCTGAGAAATCAGGCTGATGGCTGGTGACTTTTTAGTCACCAGCTTTTTTTCGCGCTGAAG						t		(288)
17	GATTGGAATACAATTTTCGGCCTTTTGTTTTATGGGCTTGCCCGTAAACGATTTTTTATATCACGGGG						t		(130)

Фрагменты нуклеотидных последовательностей с участками терминции. 1 – терминатор гена лидерной РНК плазмиды *Col EI* [10], 2 – аттенуатор *phe*-оперона *E. coli* [9], 3 – терминатор гена M_3 фага $\varnothing 80$ [11], 4 – терминатор 6S РНК фага λ [12, 13], 5 – терминатор 4S «*оор*» РНК фага λ [14, 15], 6 – терминатор фагов $M13$ и *fd* [16, 17], 7 – терминатор *trp*-оперона *E. coli* [18], 8 – аттенуатор *his*-оперона *E. coli* [19], 9 – аттенуатор *his*-оперона *S. tiphimurium* [20], 10 – аттенуатор *trp*-оперона *E. coli* [21], 11 – аттенуатор *trp*-оперона *S. tiphimurium* [21], 12 – терминатор фага $\varnothing X 174$ [22], 13 – терминатор оперона рибосомных белков S10 *E. coli* [23], 14 – терминатор в районе, предшествующем *str*-оперону рибосомных белков *E. coli* [23], 15 – аттенуатор *thr*-оперона *E. coli* [24], 16 – аттенуатор *rpo*-оперона *E. coli* [25], 17 – терминатор перед геном *rpK* *E. coli* [26]. За точку отсчета (1) приняты экспериментальные (t_e) или предположительные (t_p) точки терминции. *t* – расчетная точка терминции. Цифры в скобках указывают номера крайних правых нуклеотидов в соответствии с нумерацией в цитированных работах. Подчеркнуты найденные участки симметрии

При анализе нуклеотидных последовательностей в участках терминции мы принимали за точку терминции остаток дезокси-нуклеотида, соответствующий 3'-концевому нуклеотиду в транскрипте, пренебрегая известной неоднородностью структуры 3'-концов РНК [6,7].

Элементы сходства в структурах различных терминаторов неоднократно отмечались ранее. Так, Гилберт обратил внимание на то, что 3'-концевые последовательности ряда РНК оканчиваются U_{5-7} или $U_{5-7}A$ и в структуре ДНК соответствующей последовательности T_{5-7} или $T_{5-7}A$ предшествует G-C-богатый участок [2]. Это много раз подтверждалось в последующих работах, причем в районе терминатора всегда наблюдались элементы осевой симметрии [1,3]. Эта качественная картина не могла, однако, служить достаточным основанием для компьютерного поиска структур терминаторов в последовательностях ДНК. Используя ранее суммированные данные, а также структуры районов терминаторов и аттенуаторов, опубликованные в последнее время [1, 9–26], мы провели их статистический анализ на ЭВМ с целью выявления характерных структурных особенностей. Для этого были составлены три программы

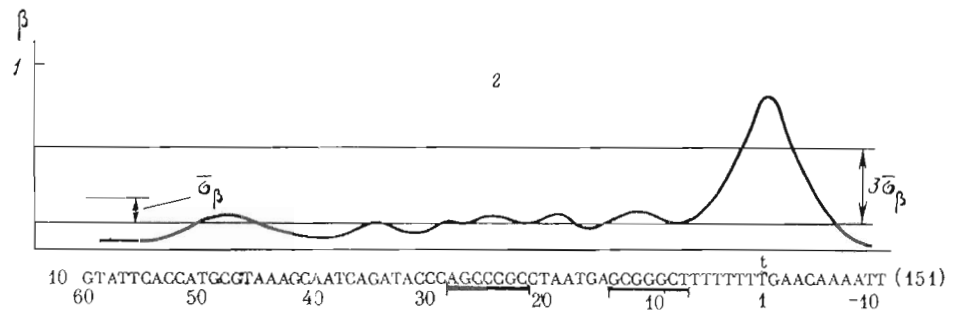
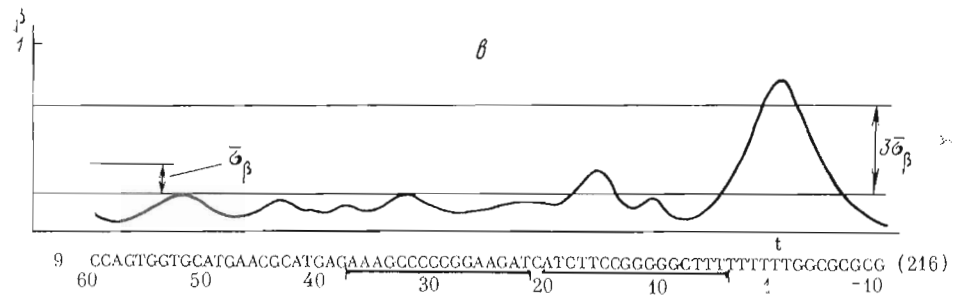
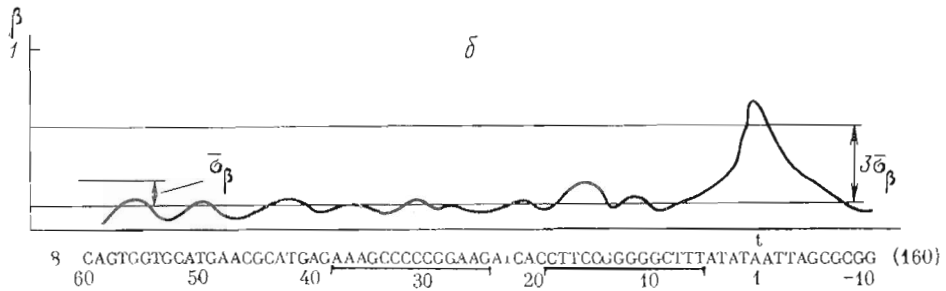
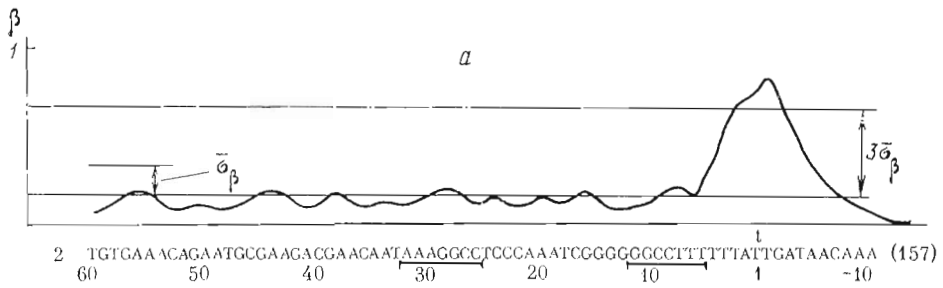
Распределение вероятностей (p) расположения нуклеотидов на различных расстояниях (l) относительно экспериментальной (t_e) или предположительной (t_p) точки терминации ($l=1$)
 Подчеркнуты значения $p > 0,51$ в компактной группе, прилегающей к точке терминации

l	40	39	38	37	36	35	34	33	32	31
$p(A)$	0,353	0,412	0,294	0,471	0,471	0,412	0,471	0,294	0,235	0,412
$p(C)$	0,294	0,235	0,118	0,176	0,118	0,118	0,059	0,118	0,353	0,412
$p(G)$	0,235	0,176	0,235	0,118	0,118	0,412	0,294	0,059	0,059	0,059
$p(T)$	0,118	0,176	0,353	0,235	0,294	0,059	0,176	0,529	0,353	0,118
l	30	29	28	27	26	25	24	23	22	21
$p(A)$	0,235	0,353	0,235	0,176	0,235	0,235	0,176	0,059	0,118	0,235
$p(C)$	0,353	0,294	0,294	0,294	0,235	0,471	0,471	0,529	0,471	0,353
$p(G)$	0,235	0,176	0,294	0,471	0,412	0,176	0,235	0,176	0,176	0,118
$p(T)$	0,176	0,176	0,176	0,059	0,118	0,118	0,118	0,235	0,235	0,294
l	20	19	18	17	16	15	14	13	12	11
$p(A)$	0,353	0,412	0,176	0,118	0,059	0,176	0,118	0,176	0,176	0,176
$p(C)$	0,235	0,235	0,235	0,235	0,235	0,176	0,294	0,294	0,235	0,059
$p(G)$	0,118	0,118	0,235	0,059	0,471	0,176	0,471	0,412	0,529	0,706
$p(T)$	0,294	0,235	0,353	0,588	0,235	0,471	0,118	0,118	0,059	0,059
l	10	9	8	7	6	5	4	3	2	1
$p(A)$	0,059	0,000	0,176	0,059	0,000	0,118	0,000	0,059	0,118	0,412
$p(C)$	0,176	0,412	0,294	0,059	0,059	0,000	0,000	0,000	0,000	0,000
$p(G)$	0,647	0,529	0,118	0,118	0,059	0,000	0,000	0,000	0,000	0,059
$p(T)$	0,118	0,059	0,412	0,765	0,882	0,882	1,000	0,941	0,882	0,529
l	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10
$p(A)$	0,125	0,313	0,313	0,188	0,250	0,313	0,375	0,333	0,375	0,250
$p(C)$	0,063	0,125	0,063	0,438	0,250	0,313	0,250	0,067	0,125	0,125
$p(G)$	0,188	0,188	0,375	0,188	0,188	0,250	0,250	0,267	0,125	0,313
$p(T)$	0,625	0,375	0,250	0,188	0,313	0,125	0,125	0,333	0,375	0,313

на языке «Фортран», по которым проводилась обработка на ЭВМ HP-3000.

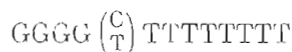
В основу алгоритма поиска характерной нуклеотидной последовательности терминатора мы положили определение статистической оценки вероятности нахождения остатков одного из четырех дезоксирибонуклеотидов N(A, T, C или G) на определенных местах относительно экспериментально найденной (t_e) или предположительной (t_p) точки терминации.

При помощи первой программы была проведена обработка 17 нуклеотидных последовательностей фрагментов ДНК; части этих последовательностей вблизи точек терминации приведены на схеме. Статистической обработке подвергались опубликованные последовательности целиком; результаты обработки вблизи точек терминации приведены в таблице. Полученные числа могут рассматриваться только как статистические оценки вероятности нахождения определенных нуклеотидов на конкретном расстоянии от t_e (t_p), так как они являются случайными числами, стремящимися к истинным вероятностям только при бесконечном увеличении числа исследуемых фрагментов. С учетом этого критерий, по которому может быть установлена нуклеотидная последовательность участков терминации, должен иметь вероятностно-статистический характер. Это означает, что выбор одного из четырех нуклеотидов производился не только по принципу максимальной вероятности (p_{\max}), но и по сравнению ее с некоторым заданным числовым уровнем. Принимая этот числовой уровень $p_{\max} \geq 0,51$, что, согласно теории статистических оце-



Функции узнавания β -участков терминации для фрагментов ДНК, обозначенных на схеме номерами: 2 (а), 8 (б), 9 (в), 10 (г).

нок [8], соответствует вероятности правильного выбора нуклеотида $P = 0,95$, мы пришли к изображенной ниже нуклеотидной последовательности участка терминации:



При этом мы не учитывали отдельные нуклеотиды, удовлетворяющие вероятностному критерию, но отстоящие на значительном расстоянии от основной компактной группы, прилегающей к точке терминации.

Для поиска участков терминации на исследуемых фрагментах ДНК мы использовали последовательное сопоставление с найденной нуклео-

тидной последовательностью и в качестве количественной меры близости структур различных участков ДНК ввели коэффициент узнавания β , который может изменяться от 1, когда сравниваемый участок полностью соответствует структуре терминатора, до 0, когда сравниваемый участок во всех элементах отличается от терминатора.

При помощи второй из составленных нами программ были рассчитаны функции узнавания β для всех 17 исследуемых фрагментов. Анализ этих функций показал, что расчетные точки терминации (t), соответствующие максимуму функции узнавания β , находятся вблизи или совпадают с экспериментальными или предположительными точками терминации (см. схему). При этом, согласно статистическим оценкам, в расчетных точках терминации превышение коэффициента β над среднестатистическим уровнем, выраженное в единицах стандартного отклонения $\bar{\sigma}_\beta$, в большинстве случаев превосходит величину 3 и нигде не менее 2,5. Согласно теории статистических оценок [8], это соответствует вероятности 0,98 того, что данная точка является точкой терминации. На рисунке в качестве примера приведены графики функций узнавания для некоторых из исследованных фрагментов ДНК.

Другим формальным признаком, характеризующим структурные особенности участка ДНК в районе терминации, согласно современным представлениям, являются элементы осевой симметрии второго порядка (комплементарность). Поиск этих структурных признаков в 17 изученных фрагментах ДНК был произведен при помощи третьей программы. Найденные участки симметрии в нуклеотидных последовательностях, приведенных на схеме и рисунке, подчеркнуты.

Из полученных данных следует, что от расчетных точек терминации (t) на расстоянии 4—10 нуклеотидных остатков всегда находится структура с элементами симметрии второго порядка в виде пары комплементарных отрезков длиной не менее 5 нуклеотидных остатков каждый, находящихся на переменном расстоянии один от другого.

Таким образом, вторым существенным признаком участка терминации, который следует использовать для его математического поиска, является наличие палиндрома, охватывающего левую часть нуклеотидной последовательности терминатора. Следует отметить, что структуры, соответствующие критерию математического поиска участка терминации, иногда встречаются вблизи старта транскрипции (например, в приведенных на схеме последовательностях 13 [23] или 16 [25]). Возможно, терминация в этих участках в действительности не происходит.

ЛИТЕРАТУРА

1. Adnya S., Gottesman M. (1978) *Ann. Rev. Biochem.*, **47**, 967—996.
2. Gilbert W. (1976) in: *RNA Polymerase* (Losick R., Chamberlin M., eds), pp. 247—271, Cold Spring Harbor.
3. Roberts J. W. (1976) in: *RNA Polymerase* (Losick R., Chamberlin M., eds), pp. 263—302, The Operon Cold Spring Harbor.
4. Howard B., de Crombrughe B., Rosenberg M. (1977) *Nucl. Acids Res.*, **4**, 827—842.
5. Platt T. (1978) in: *The Operon* (Miller J. H., Reznikoff W. S., eds), pp. 263—302, Cold Spring Harbor.
6. Bertrand K., Korn L. J., Lee F., Yanofsky C. (1977) *J. Mol. Biol.*, **117**, 227—247.
7. Rosenberg M., Court D., Shimatake H., Brady C., Wulff D. L. (1978) *Nature*, **272**, 414—423.
8. Крамер Г. (1975) Математические методы статистики, «Мир», М.
9. Zurawski G., Brown K., Killingly D., Yanofsky C. (1978) *Proc. Nat. Acad. Sci. USA*, **75**, 4271—4275.
10. Backman K., Betlack M., Boyer H. W., Yanofsky C. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 69—76.
11. Pieczenik G., Barrell B., Gefter M. (1972) *Arch. Biochem. and Biophys.*, **152**, 152—165.
12. Lebowitz P., Weismann S., Radding C. (1971) *J. Biol. Chem.*, **246**, 5120—5139.
13. Sklar J., Yot P., Weissmann S. M. (1975) *Proc. Nat. Acad. Sci. USA*, **72**, 1817—1821.
14. Rosenberg M., de Crombrughe B., Musso R. (1976) *Proc. Nat. Acad. Sci. USA*, **73**, 717—721.
15. Schwarz E., Scherer G., Hobom G., Kossel H. (1978) *Nature*, **272**, 410—414.

16. Rivera M., Konings R., Quint W., Shoemakers J. (1978) Nucl. Acids Res., 5, 2895–2913.
17. Sugimoto K., Sugisaki H., Okamoto T., Takanami M. (1977) J. Mol. Biol., 111, 487–507.
18. Wu A. M., Platt T. (1978) Proc. Nat. Acad. Sci. USA, 75, 5422–5466.
19. Nocera P. P., Blasi F., Lauro R. D., Frunzio R., Bruni C. (1978) Proc. Nat. Acad. Sci. USA, 75, 4276–4280.
20. Barnes W. M. (1978) Proc. Nat. Acad. Sci. USA, 75, 4281–4285.
21. Lee F., Bertrand K., Bennett G., Yanofsky C. (1978) J. Mol. Biol., 121, 193–217.
22. Sanger F., Air G., Barrel B., Brown N. (1977) Nature, 265, 687–695.
23. Post L. E., Arfsten A. F., Reusser F., Nomura M. (1978) Cell, 15, 215–229.
24. Gardner J. F. (1979) Proc. Nat. Acad. Sci. USA, 76, 1706–1710.
25. Гуревич А. И., Аваков А. Э., Колосов М. Н. (1979) Биооргани. химия, 5, 1735–1739.
26. Post L. E., Strycharz G. D., Nomura M., Lewis H., Dennis P. P. (1979) Proc. Nat. Acad. Sci. USA, 76, 1697–1701.

Поступила в редакцию
28.III.1980

THE STRUCTURE OF *E. COLI* RNA POLYMERASE TRANSCRIPTION TERMINATION SITE

VASILYEV G. V., GUREVICH A. I., MALINA Z. A., ARTEMYEV I. V.

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

The structure of *E. coli* RNA polymerase transcription termination site (ρ -independent terminator) is found from computer statistic analysis data of 17 DNA fragments. The first indication of this structure is the nucleotide sequence $\text{GGGG} \begin{pmatrix} \text{C} \\ \text{T} \end{pmatrix} \text{TTTTTTTT}$ preceding the termination point. This sequence, being of probabilistic character, was used for computer screening of the termination sites by the criterion of the maximal recognizing function β . The second indication of the termination site is the dyadic axial symmetry elements situated 4–10 base pairs to the left of the calculated termination point. The symmetrical element (palindrome) must therefore include the left part of the nucleotide sequence of the terminator.