



# БИООРГАНИЧЕСКАЯ ХИМИЯ

том 5 \* № 11 \* 1979

## ПИСЬМА РЕДАКТОРУ

УДК 547.963.32.02

### НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ НАЧАЛА ГЕНА *rpoB* *ESCHERICHIA COLI*

Гуревич А. И., Аваков А. Э., Колосов М. Н.

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

Из трансдуцирующего фага  $\lambda$ rif<sup>d</sup>47 [1] мы выделили EcoRI-фрагмент бактериальной ДНК длиной около 1100 нуклеотидных пар (н.п.), который содержит конец гена *rplL*, начало гена *rpoB* и регуляторный участок между ними. Ранее нами было выяснено положение этого фрагмента (обозначенного EcoRI-K) на физической карте  $\lambda$ rif<sup>d</sup>47 [2] и исследована структура двух его *Hpa*II-субфрагментов [3, 4]. Продолжая изучение оперона *rpoBC*, мы определили полную нуклеотидную последовательность фрагмента EcoRI-K, включая начальную часть структурного гена *rpoB*, кодирующую первые 185 аминокислотных остатков  $\beta$ -белка РНК-полимеразы.

Выясненная нами карта расщепления фрагмента EcoRI-K рестриктазами и схема его секвенирования приведены на рис. 1. Как видно из этого рисунка, большая часть нуклеотидной последовательности была определена по обеим цепям ДНК. Анализ проводили методом Максама — Гильберта [5] в модификации Коробко и др. [6, 7]; 5'- и 3'-концевую  $^{32}$ P-метку вводили как описано ранее [5, 3], комплементарные цепи разделяли по методу [5]. При определении нуклеотидной последовательности некоторых участков ДНК, например термиатора гена *rplL*, встретились труд-

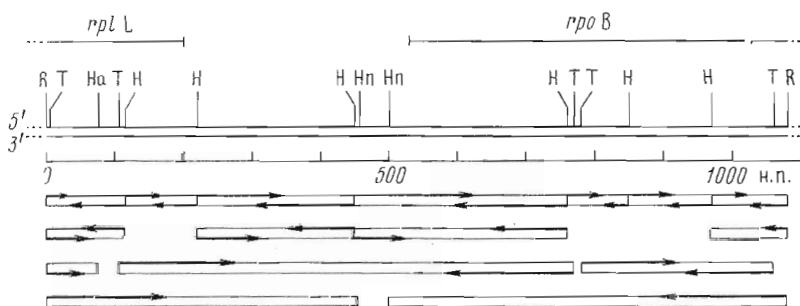


Рис. 1. Рестриктазная карта фрагмента EcoRI-K генов *rplL*-*rpoB* *E. coli* и схема определения его нуклеотидной последовательности. Указано расположение структурных генов *rplL* и *rpoB* и сайтов действия эндопукилаз *Eco*RI (R), *Hpa*II (H), *Hae*III (Ha), *Hind*II (Hn) и *Taq*I (T). Горизонтальные стрелки показывают длину установленных последовательностей и направление секвенирования в каждой из двух пытей соответствующего субфрагмента (полярность верхней цепи ДНК везде 5'→3')

...GluPheAspValIleLeuLysAlaAlaGlyAlaAsnLysValAlaValIleLysAlaValArgGly

EcoRI TaqI

50

...GAAATTCGACGTAATTCTGAAAGCTGCTGGCGCTAACAAAGTTGCTGTTATCAAAGCAGTACGTGGC  
...CTTAAGCTGCATTAAGACTTCGACGCCGATTGTTCAACGACAATAGTTCGTCATGCACCG

AlaThrGlyLeuGlyLeuLysGluAlaLysAspLeuValGluSerAlaProAlaAlaLeuLysGluGly

HaeIII

100

TaqI

HpaII

GCAACTGGCTGGGCTGAAAGAAGCTAAAGACCTGGTCGAATCTGACCAGCTGCTCTGAAAGAAGGC  
CGTTGACCCGACCCAGACTTCTCGATTCTGGACCAGCTTAGACGTGGCCGACGAGACTTCTTCG

ValSerLysAspAspAlaGluAlaLeuLysLysAlaLeuGluGluAlaGlyAlaGluValGluValLys

I50

200

GTGAGCAAAGACGACGAGCAGAACGCACTGAAAAAAAGCTCTGGAAGAAGCTGGCGCTGAAGTGAAGTTAAA  
CACTCGTTCTGCTGCGTCTCGTGACTTTTGAGCACCTTCTCGACCGCAGCTCAACTCAATT

HpaII

250

TAAGCCAACCCCTCCGGTTCAGCCTGAGAAATCAGGCTGATGGCTGGTGACTTTTAGTCACCAGGTT  
CTTCGGTTGGGAAGGCCAAGCTCGGACTCTTAGTCCGACTACCGACCACTGAAAAATCAGTGGTCGAAA

300

TTTTTGCGCTGTAAGGCGCCAGTAGCGTTTCAACTGTTGACTACTGCTGTGCCTTCAATGCTTGTT  
AAAAACCGCACATTCCCGCGTCATCGAAAGTGTGACAAACTGATGACGACACGGAAAGTTACGAACAA

350

400

TCTATCGACGACTTAATATACTGCGACAGGACGTCGTTCTGTAAAATCGCAATGAAATGGTTAAGC  
AGATAGCTGTAATTATGACGCTGCCTGCAGGAAGACACATTAGCGTTACTTACCAAATTCG

HpaII HindII

GTGATAGCAACAGGCATTCGGAAAGTGTCCATTTTCCCAACAAAATAGTGTGACAAACTGTC  
CACTATCGTTGTCGTAACGCCTTCAAGGTAAAGGCCAGTTGTTATACACGTGTTGACAG

MetValTyrSerTyrThrGlu

500 HindII

CGCTCAATGGACAGATGGGTCGACTTGTCAGCGAGCTGAGGAACCCTATGGTTACTCCTATACCGAG  
GCGAGTTACCTGTTACCCAGCTGAACAGTCGCTGACTCCTTGGATACCAATGAGGATATGGCTC

LysLysArgIleArgLysAspPheGlyLysArgProGlnValLeuAspValProTyrLeuLeuSerIle

550

600

AAAAAACGTATTCGTAAGGATTTGGTAAACGCTCCACAGTCTGGATGTACCTTATCTCCTTCTATC  
TTTTTGCATAAGCATTCTAAACCATTGCAGGGTGTTCAAGACCTACATGGAATAGAGGAAAGATAG

GlnLeuAspSerPheGlnLysPheIleGluGluAspProGluGlyGlnTyrGlyLeuGluAlaAlaPhe

650

CAGCTTGACTCGTTCAGAAATTATCGAGCAAGATCTGAGCTGAGGCAGTATGGTCTGGAAGTCTGTTTC  
GTCGAACTGAGCAAGTCTTAAATAGCTCGTTAGGACTCCCGTCATACCAGACCTTCGACGAAAG

ArgSerValPheProIleGlnAlaThrAlaValIleProSerCysAsnAsnValThrThrProTrpArg

700

750

CGTTCGTATTCCGATTCGGCTACAGCGGTAATTCCGAGCTGCAATAACGTCACTACCCCTGGCGA  
GCAAGGCATAAGGCTAAGTCCGATGTCGCCATTAGGCTCGACGTTATTGCAGTGTCGGACCGCG

ThrGlyPhePheAspValGluGluProTrpArgAspLeuPheArgThrAlaArgValLysLeuArgLeu  
 HpaII TaqI TaqI 800  
 ACCGGCTTTTCGACGTCAGGAACCGTGGCGTGACCTATTCCGACCGCGCGCGTAAACTGCGTC  
 TGGCGAAAAGCTGCAGCTCCTGGCACCGCACTGGATAAGGCGTGGCGCGCAATTGACGCAGAC

ValIleTyrGluArgGluAlaProGluGlyThrValLysAspIleLysGluGlnGluValTyrMetGly  
 HpaII 850  
 GTGATTATGAGCGCGAAGCGCCGAAGGACCGTAAAGACATTAAAGAACAGAAAGTCTACATGGC  
 CACTAGATACTCGCGCTTCGCGGCCCCGTGGCATTTCTGTAATTCTGTTCTCAGATGTACCCG

GluIleProLeuMetAlaAspAsnGlyThrPheValIleAsnGlyThrGluArgValIleValProSer  
 900 950  
 GAAATTCCGCTCATGGCAACACGGTACCTTGTTATCAACGGTACTGAGGCGTGTTATCGTCCAGC  
 CTTAAGGCGAGTACCGTCTGGCCATGGAAACAATAGTGGCCATGACTCGCACAATAGCAAGGGTCG

CysThrValValArgAlaSerSerPheThrProThrLysValLysProThrLeuArgValLysCysCys  
 HpaII 1000  
 TGCACCGTGGTCGCGCGTCTTCTTGACTCCGACAAAGGTAAACCCACTCTTCGGTAAAGTGCTGT  
 ACGTGGCACCAGGCCCGAGAAGAAACTGGCTGTTCCATTTGGGTGAGAAGCCCATTACGACA

ThrThrArgValSerSerLeuThrValAspSerTrpLeuAspPheGluPhe...

1050 TaqI EcoRI

ACAACCGCGGTTATCCCTACCGTCGATTCTGGCTGGACTTCGAAATTC...  
 TGTTGCGCGCATAGTAGGGGAATGGCAGCTAAGGACCGACCTGAAGCTTAAG...

Рис. 2. Нуклеотидная последовательность фрагмента *EcoRI-K* генов *rplL-groB* *E. coli* и соответствующие аминокислотные последовательности рибосомного белка L7/L12 и β-белка РНК-полимеразы. Подчеркнуты экспериментально найденные сайты расщепления рестриктаазами. Двойной чертой отмечен предполагаемый участок связывания рибосомы. В рамки заключены элементы вращательной симметрии второго порядка

ности, возможно, из-за невыгодной (для химической модификации и/или разделения в геле) вторичной структуры анализируемых одноцепочечных фрагментов. Поэтому, чтобы уменьшить вероятность ошибки, каждую последовательность анализировали многократно и, где было возможно, использовали разные фрагменты, полученные с помощью других рестриктиаз. В результате были уточнены и исправлены прежние данные о структуре участков 1–118 [3] и 186–452 [4] и определено строение всего фрагмента *EcoRI-K*, изображенное на рис. 2.

Установленная нами последовательность насчитывает 1082 н.п. Из них 204 н.п. занимает дистальная часть гена *rplL*, структура которой полностью согласуется с известной аминокислотной последовательностью белка L7/L12 [8]. Ген *groB*, очевидно, начинается в положении 528. Этот вывод основан на следующих соображениях: 1) инициирующему триплету ATG (528–530, здесь и далее указаны нуклеотиды верхней цепи) предшествует тетрануклеотид GAGG (518–521), который комплементарен 3'-концевому участку 16S рРНК и может служить сайтом связывания рибосомы (так называемая последовательность Шайн–Дальгарно [9]); 2) первые три кодона предполагаемого гена *groB* – ATG.GTT.TAC (528–539) – соответствуют опубликованной для β-белка РНК-полимеразы N-концевой структуре Met-Val-Tyr [10]; 3) в тринадцатой последовательности 528–530, 531–533 и т. д., в отличие от двух других рамок считывания генетической информации, нет терминирующих кодонов до конца исследованного фрагмента ДНК. Это позволяет постулировать для

$\beta$ -белка РНК-полимеразы 185-членную N-концевую аминокислотную последовательность, изображенную на рис. 2. Спайсер между генами *rplL* и *groB* занимает почти треть фрагмента EcoRI-K (около 320 н.п.). В нем недалеко от конца гена *rplL* имеются два обращенных повтора: 8-членный (225–232, 237–244) и обнаруженный ранее [4] 10-членный (248–257, 262–271), сразу за которым находится последовательность TTTTTTTG; по нашему мнению, именно этот участок является аттенюатором оперона *rplJL-groB*C. В средней части спайсера есть еще один (правда, несовершенный) обращенный повтор: CAATGCPyTGTT (332–342, 420–430); его функциональная роль и значение для вторичной структуры ДНК (и соответствующей мРНК) не известны, но интересно, что находящиеся в этой области сайты TCGA (347–350, 500–503) не расщепляются рестриктазой *TaqI*.

После завершения настоящей работы в печати появилась статья Номуры и сотр. [11], в которой опубликована нуклеотидная последовательность большого сегмента (3072 н.п.) ДНК *E. coli*, выделенного из трансдуцирующего фага  $\lambda rif^a 18$  и содержащего участок 1–628 исследованного нами фрагмента EcoRI-K. Наши результаты совпадают с полученными группой Номуры, за исключением следующих четырех. В гене *rplL* американскими учеными найдены нуклеотиды 100-T и 105-A, а нами – 100-C и 105-C; эти различия относятся к третьей букве триплета и не меняют его смысла в отношении кодируемой аминокислоты. Следует отметить, что и мы на основании химической деградации вначале идентифицировали нуклеотид 105 как А, но затем обнаружили, что непосредственно перед ним ДНК расщепляется рестриктазой *TaqI* и, следовательно, этот нуклеотид должен быть С. В межгенной области, по данным Поста и соавт. [11], вместо 272-T<sub>7</sub> содержится CT<sub>4</sub>, а между нуклеотидами 500 и 501 находится тринуклеотид CGT, отсутствующий в формуле на рис. 2. Мы повторно секвенировали соответствующие участки ДНК, но не смогли получить подтверждения этих данных.

#### ЛИТЕРАТУРА

- Миндлин С. З., Ильина Т. С., Горленко Ж. М., Хачикян Н. А., Ковалев Ю. Н. (1976) Генетика, 12, № 12, 116–130.
- Гуревич А. И., Аваков А. Э., Кисслева О. А., Колесов М. Н. (1978) Биоорганическая химия, 4, 628–638.
- Гуревич А. И., Аваков А. Э. (1979) Биоорганическая химия, 5, 301–304.
- Гуревич А. И., Аваков А. Э., Колесов М. Н. (1979) Биоорганическая химия, 5, 779–781.
- Махам А. М., Gilbert W. (1977) Proc. Nat. Acad. Sci. USA, 74, 560–564.
- Коробко В. Г., Грачев С. А. (1977) Биоорганическая химия, 3, 1420–1422.
- Коробко В. Г., Грачев С. А., Колесов М. Н. (1978) Биоорганическая химия, 4, 1281–1283.
- Terhorst C., Müller M., Laursen R., Wittmann-Liebold B. (1973) Eur. J. Biochem., 34, 138–152.
- Shine J., Dalgarno L. (1974) Proc. Nat. Acad. Sci. USA, 71, 1342–1346.
- Fujiki H., Zurek G. (1975) FEBS Lett., 55, 242–244.
- Post L. E., Strycharz G. D., Nomura M., Lewis H., Dennis P. P. (1979) Proc. Nat. Acad. Sci. USA, 76, 1697–1701.

Поступило в редакцию  
2.VII.1979

THE NUCLEOTIDE SEQUENCE AT THE PROXIMAL  
END OF *rpoB* GENE OF *ESCHERICHIA COLI*

GUREVICH A. I., AVAKOV A. E., KOLOSOV M. N.

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

A *EcoRI* fragment of *E. coli* DNA isolated from transducing phage  $\lambda$  rif<sup>d</sup>47 has been sequenced by a modified Maxam – Gilbert method. The fragment is 1080 nucleotide long and contains the distal and proximal ends of genes *rplL* and *rpoB*, respectively, separated by a spacer of ca. 320 b.p. The terminal 204 nucleotide sequence determined in *rplL* was found to completely agree with the published structure of the ribosomal protein L7/L12. The start of *rpoB* is assumed to be at the ATG in position 528–530, because it is preceded by a putative ribosome binding site GAGG and followed by no termination codons till the end of the DNA fragment studied while several nonsense codons are present in the two other reading frames. The starting 555 nucleotide sequence thus determined in *rpoB* codes for 185 amino acids of the  $\beta$ -protein of RNA polymerase, the first three amino acids being identical to those previously reported as the N-terminal. In the *rplL*-*rpoB* spacer, two inverted repeats are present near the end of *rplL*, an 8-membered and a 10-membered followed by T<sub>7</sub>G, which supposedly constitute an attenuator of the *rplL*-*rpoB* operon. There is also a nonperfect 11-membered inverted repeat at a longer distance whose function, if any, remains to be known.

---

Технический редактор *E. C. Кузьмишина*

---

Сдано в набор 20.08.79      Подписано к печати 02.10.79      Т-13673      Формат бумаги 70×108<sup>1/16</sup>  
Высокая печать      Усл. печ. л. 13,3+1 вкл.      Уч.-изд. л. 13,9      Бум. л. 4<sup>3/4</sup>      Тираж 870 экз. Зак. 2173

---

Издательство «Наука», 103717 ГСП, Москва, К-12, Подсосенский пер., 21  
2-я типография издательства «Наука», 121099, Москва, Шубинский пер., 10