



ПИСЬМА РЕДАКТОРУ

УДК 547.963.32.02+547.96.02

ПЕРВИЧНАЯ СТРУКТУРА РНК-ПОЛИМЕРАЗЫ *E. coli*.
НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ФРАГМЕНТА
ДНК *E. coli*, СОДЕРЖАЩЕГО ЧАСТЬ ГЕНА *groC*,
И СООТВЕТСТВУЮЩАЯ С-КОНЦЕВАЯ АМИНОКИСЛОТНАЯ
ПОСЛЕДОВАТЕЛЬНОСТЬ β' -СУБЪЕДИНИЦЫ

Овчинников Ю. А., Монастырская Г. С., Губанов В. В.,
Саломатина И. С., Шувалева Т. М., Липкин В. М.,
Свердлов Е. Д.

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

При параллельном исследовании структуры *groBC* оператор *E. coli* и аминокислотной последовательности β - и β' -субъединиц РНК-полимеразы ранее мы установили полную нуклеотидную последовательность гена *groB* и аминокислотную последовательность β -субъединицы РНК-полимеразы *E. coli*, последовательность начала гена *groC* и соответствующую ему N-концевую аминокислотную последовательность β' -субъединицы, а также структуру межцистронной области *groB* — *groC* [1].

В данной работе публикуется нуклеотидная последовательность фрагмента ДНК *E. coli*, содержащего часть гена *groC*, и соответствующая ему аминокислотная последовательность С-концевой части β' -субъединицы РНК-полимеразы *E. coli*.

Источником исследуемого фрагмента ДНК служила «космида» *pJc703* [2], из которой выделяли фрагмент *EcoRI* — *SalI*, содержащий С-концевую область гена *groC* (рис. 1). Этот фрагмент рекомбинировали *in vitro* с плазмидой *pBR-322*, расщепленной эндонуклеазами *EcoRI* и *SalI*, и после трансформации клеток *E. coli* HB101 селектировали трансформанты с фенотипом *Ap^rTc^s*. Клоны, содержащие необходимый фрагмент, отбирали путем рестрикционного анализа плазмид. Фрагмент для анализа последовательности получали с помощью препаративного электрофореза расщепленной эндонуклеазами *EcoRI* и *HindIII* рекомбинантной плазмиды в горизонтальных блоках 1% агарозы. Нужный фрагмент с *M* 1,3 · 10⁶ (R—H-фрагмент) извлекали из геля путем электроолиговирования.

Общая стратегия исследования первичной структуры фрагмента R—H аналогична примененной нами в работе [3] (см. рис. 2). Фрагмент расщепляли поочередно одной из рестриктаз: *HpaII*, *Sau3A1*, *HinfI*. Последовательности комплементарных цепей выделенных субфрагментов определяли модифицированными нами [4] методом Максама — Гилберта [5] после введения ³²P-концевой фосфатной группы с помощью [γ -³²P]АТФ и полинуклеотидкиназы фага Т4. Для «перекрытия» сайта рестриктазы *HinfI* в положении 229—233 определили структуру фрагментов, получен-

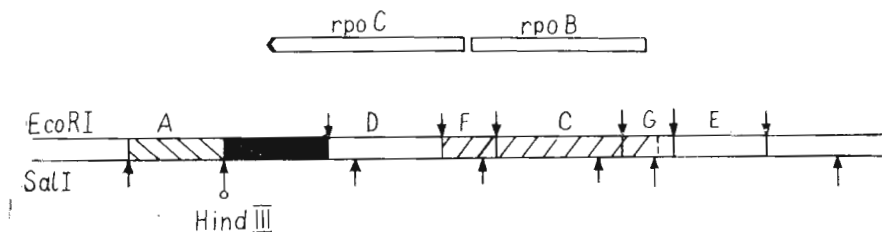


Рис. 1. Карта расщепления участка ДНК *E. coli*, содержащего структурные гены β - и β' -субъединиц РНК-полимеразы (*rpoB* и *rpoC* соответственно) рестрикционными эндонуклеазами *EcoRI* и *SalI*. Буквенные обозначения фрагментов соответствуют принятым в работе [2]. Указан также сайт расщепления *HindIII*. R—H-фрагмент зачернен

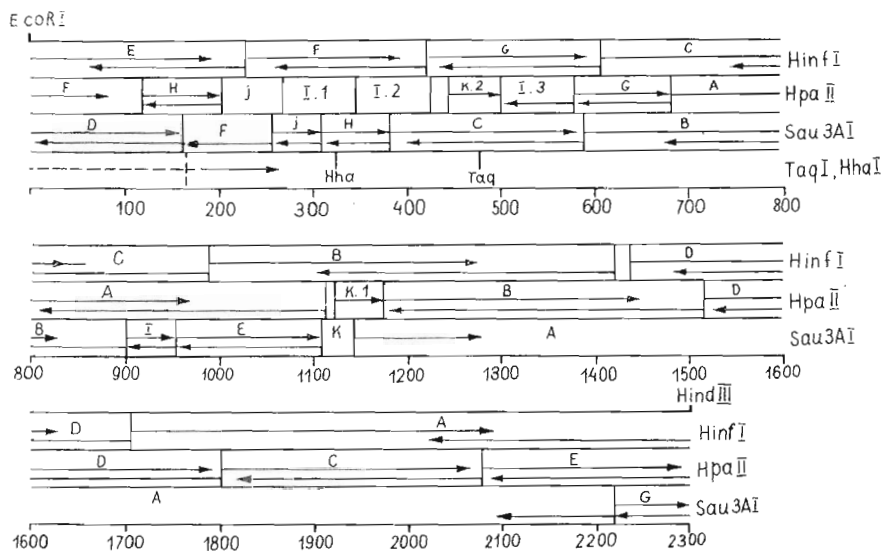


Рис. 2. Детальная карта расщепления фрагмента R—H рестрикционными эндонуклеазами и схема установления его первичной структуры. Субфрагменты, образующиеся при действии рестриктаз, изображены прямоугольниками. Стрелки обозначают длину установленной последовательности комплементарных цепей данного субфрагмента. Вертикальная пунктирная линия обозначает положение последовательности TCGA, не расщепляемой рестриктазой *TaqI*

ных путем расщепления рестриктазами *TaqI* и *HhaI* фрагмент R—H после введения в него ^{32}P -концевых фосфатных групп. Кроме того, была определена структура фрагментов, образующихся при расщеплении субфрагмента *Sau-F* рестриктазами *HpaII* и *BspI*. Практически вся структура определена по двум цепям. Установленная последовательность, содержащая 2306 пар оснований, представлена на рис. 3.

Изучение аминокислотной последовательности β' -субъединицы начато с расщепления ее полипептидной цепи бромцианом. Из полученного гидролизата путем гель-фильтрации, хроматографии на бумаге и в тонком слое, электрофореза в блоках ацетатцеллюлозы, а также экстракцией гидрофобных пептидов бутанолом выделено 17 пептидов, содержащих в сумме ~850 аминокислотных остатков. Кроме того, проведен триптический гидролиз β' -субъединицы, модифицированной по остаткам лизина цитраконовым ангидридом.

Выбор значащей цепи в исследуемом фрагменте ДНК осуществлен на основании того, что, согласно известной физико-генетической карте участка *rpoBC*, транскрипция осуществляется от сайта *EcoRI* к сайту *HindIII* [6]. Следовательно, матричной является та цепь, 3'-конец которой лежит

в сайте рестриктазы *EcoRI*. Цепь, последовательность которой адекватна мРНК, комплементарна ей. Поиск «рамки» кодирования аминокислотной последовательности в этой цепи был осуществлен с помощью компьютера путем сравнения нуклеотидной последовательности с аминокислотными последовательностями пептидов β' -субъединицы. Полная или частичная последовательность пептидов, соответствующая последовательности R—H-фрагмента, подчеркнута на рис. 3. В найденной таким образом «рамке» считывания завершающий кодон TAA находится в положении 1264—1266*.

В интервале 1288—1313 в непосредственной близости от завершающего кодона находится блок тимидинов и перед ним последовательности с осью симметрии второго порядка (см. рис. 3), т. е. участок, структура которого близка структурам р-независимых терминаторов [7]. Недавно установлено [8], что терминация транскрипции *rhoBC*-оперона действительно осуществляется в непосредственной близости от конца структурного гена *rhoC*. Указанная последовательность — наиболее вероятный кандидат на роль терминатора транскрипции *rhoBC*-оперона.

Интересной особенностью исследованной последовательности является наличие участка узнавания рестриктазы *TaqI* в положении 164—167, не расщепляемого ею. Два таких же примера мы описали в работе [3]. Во всех трех случаях наблюдается одна и та же последовательность GATCGA, в которой нерасщепляемый сайт рестриктазы *TaqI* (TCGA) перекрывается с сайтом (расщепляемым) рестриктазы *Sau3AI* (GATC).

Возможно, что в комплементарной последовательности TCGATC отмеченный звездочкой аденозин модифицирован.

Соответствующая R—H-фрагменту С-концевая аминокислотная последовательность β' -субъединицы (рис. 3) содержит 421 аминокислотный остаток.

Авторы выражают благодарность проф. Г. Чахау и проф. Б. Мюллер-Хилу (ФРГ) за предоставление рестрикционных эндонуклеаз, д-ру Дж. Коллинзу (ФРГ) за предоставленную плазмиду *pJc703*, Ю. И. Вишневному, Э. И. Царьковой и Г. М. Аринушкиной за помощь в работе, Г. В. Васильеву и И. В. Артемьеву за помощь в математической обработке данных.

* Авторы считают необходимым отметить, что на одном из структурных гелей, соответствующих матричной цепи фрагмента, проявилась дополнительная полоса, читаемая как «А», в положении между 1215 и 1216 на рис. 3. В результате такой вставки выводимая последовательность сократилась бы на 17 аминокислотных остатков ввиду образования терминатора TAG. Последовательность данного участка определялась дважды по этой цепи и трижды по комплементарной, однако в других гелях этого дуплекета не обнаружено. На основании регулярности структуры гелей мы считаем эту дополнительную полосу артефактной. Тем не менее, поскольку данный участок входит в последовательность, способную к образованию «шпильки», что является потенциальной основой для компрессии, окончательный вывод о положении терминатора трансляции будет сделан при анализе структуры С-концевого пептида β' -субъединицы.

I-75 GAA TTC GGT CGT ACT AAA GAA AGC TAC AAA GTA CCT TAC GGT GCG GTA CTG GCG AAA GGC GAT GGC GAA CAG GTT
 I-25 Glu-Phe-Gly-Arg-Thr-Lys-Glu-Ser-Tyr-Lys-Val-Pro-Tyr-Gly-Ala-Val-Leu-Ala-Lys-Gly-Asp-Gly-Glu-Gln-Val-
 76-150 GCT GGC GGC GAA ACC GTT GCA AAC TGG GAC CCG CAC ACC ATG CCG GTT ATC ACC GAA GTA AGC GGT TTT GTA CGC
 26-50 Ala-Gly-Gly-Glu-Thr-Val-Ala-Asn-Trp-Asp-Pro-His-Thr-Met-Pro-Val-Ile-Thr-Glu-Val-Ser-Gly-Phe-Val-Arg-
 151-225 TTT ACT GAC ATG ATC GAC GGC CAG ACC ATT ACG CGT CAG ACC GAC GAA CTG ACC GGT CTG TCT TCG CTG GTG GTT
 51-75 Phe-Thr-Asp-Met-Ile-Asp-Gly-Gln-Thr-Ile-Thr-Arg-Gln-Thr-Asp-Glu-Leu-Thr-Gly-Leu-Ser-Leu-Val-Val-
 226-300 CTG GAT TCC GCA GAA CGT ACC GCA GGT GGT AAA GAT CTG CGT CCG GCA CTG AAA ATC GTT GAT GCT CAG GGT AAC
 76-100 Leu-Asp-Ser-Ala-Glu-Arg-Thr-Ala-Gly-Gly-Lys-Asp-Leu-Arg-Pro-Ala-Leu-Lys-Ile-Val-Asp-Ala-Gln-Gly-Asn-
 301-375 GAC GTT CTG ATC CCA GGT ACC GAT ATG CCA GCG CAG TAC TTC CTG CCG GGT AAA GCG ATT GTT CAG CTG GAA GAT
 101-125 Asp-Val-Leu-Ile-Pro-Gly-Thr-Asp-Met-Pro-Ala-Gln-Tyr-Phe-Leu-Pro-Gly-Lys-Ala-Ile-Val-Gln-Leu-Glu-Asp-
 376-450 GCG GTA CAG ATC AGC TCT GGT GAC ACC CTG GCG CGT ATT CCG CAG GAA TCC GCG GGT ACC AAG GAC ATC ACC GGT
 126-150 Gly-Val-Gln-Ile-Ser-Ser-Gly-Asp-Thr-Leu-Ala-Arg-Ile-Pro-Gln-Glu-Ser-Gly-Gly-Thr-Lys-Asp-Ile-Thr-Gly-
 451-525 GGT CTG CCG GCG GTT GCG GAC CTG TTC GAA GCA GGT CGT CCG AAA GAG CCG GCA ATC CTG GCT GAA ATC AGC GGT
 151-175 Gly-Leu-Pro-Arg-Val-Ala-Asp-Leu-Phe-Glu-Ala-Arg-Arg-Pro-Lys-Glu-Pro-Ala-Ile-Leu-Ala-Glu-Ile-Ser-Gly-
 526-600 ATC GTT TCC TTC GGT AAA GAA ACC AAA GGT AAA CGT CGT CTG GTT ATC ACC CCG GTA GAC GGT AGC GAT CCG TAC
 176-200 Ile-Val-Ser-Phe-Gly-Lys-Glu-Thr-Lys-Gly-Lys-Arg-Arg-Leu-Val-Ile-Thr-Pro-Val-Asp-Gly-Ser-Asp-Pro-Tyr-
 601-675 GAA GAG ATG ATT CCG AAA TGG CGT CAG CTC AAC GTG TTC GAA GGT GAA CGT GTA GAA CGT GGT GAC GTA ATT TCC
 201-225 Glu-Glu-Met-Ile-Pro-Lys-Trp-Arg-Gln-Leu-Asn-Val-Phe-Glu-Gly-Glu-Arg-Val-Glu-Arg-Gly-Asp-Val-Ile-Ser-
 676-750 GAC GGT CCG GAA GCG CCG CAC GAC ATT CTG CGT CTG CGT GGT GAT CAT GGT ATT CCG TAC ATC GTT AAC GAA
 226-250 Asp-Gly-Pro-Glu-Ala-Pro-His-Asp-Ile-Leu-Arg-Leu-Arg-Gly-Val-His-Ala-Val-Thr-Arg-Tyr-Ile-Val-Asn-Glu-
 751-825 GTA CAG GAC GTA TAC CGT CTG CAG GGC GTT AAG ATT AAC GAT AAA CAC ATC GAA GTT ATC GTT CGT CAG ATG CTG
 251-275 Val-Gln-Asp-Val-Tyr-Arg-Leu-Gln-Gly-Val-Lys-Ile-Asn-Asp-Lys-His-Ile-Glu-Val-Ile-Val-Arg-Gln-Met-Leu-
 826-900 CGT AAA GCT ACC ATC GTT AAC GCG GGT AGC TCC GAC TTC CTG GAA GGC GAA CAG GTT GAA TAC TCT CGC GTC AAG
 276-300 Arg-Lys-Ala-Thr-Ile-Val-Asn-Ala-Gly-Ser-Ser-Asp-Phe-Leu-Glu-Gly-Gln-Val-Glu-Tyr-Ser-Arg-Val-Lys-

ЛИТЕРАТУРА

1. Овчинников Ю. А., Монастырская Г. С., Губанов В. В., Гурьев С. О., Чертов О. Ю., Модянов П. Н., Гришкевич В. А., Макарова И. А., Марченко Т. В., Половникова И. Н., Липкин В. М., Свердлов Е. Д. Первичная структура РНК-полимеразы *Escherichia coli*. Нуклеотидная последовательность гена *rpoB* и аминокислотная последовательность β -субъединицы.— Докл. АН СССР, 1980, т. 253, № 4, с. 994–998.
2. Collins J. Deletions, insertions and rearrangements affecting *rpoB* gene expression.— Mol. Gen. Genet., 1979, v. 173, № 2, p. 217–220.
3. Овчинников Ю. А., Свердлов Е. Д., Липкин В. М., Монастырская Г. С., Чертов О. Ю., Губанов В. В., Гурьев С. О., Модянов П. Н., Гришкевич В. А., Макарова И. А., Марченко Т. В., Половникова И. Н. Первичная структура РНК-полимеразы *E. coli*. Нуклеотидная последовательность фрагмента *EcoRI*-С гена *rpoB* и аминокислотная последовательность соответствующего фрагмента β -субъединицы.— Биоорг. химия, 1980, т. 6, № 5, с. 655–665.
4. Ovchinnikov Yu. A., Guryev S. O., Kravov A. S., Monastyrskaya G. S., Skryabin K. G., Sverdlov E. D., Zakharyev V. M., Bayev A. A. Primary structure of an *EcoRI* fragment of λ imm 434 DNA containing regions *ci-cro* of phage 434 and *sII-O* of phage lambda.— Gene, 1979, v. 6, № 3, p. 235–249.
5. Maxam A., Gilbert W. A new method for sequencing DNA.— Proc. Nat. Acad. Sci. USA, 1977, v. 74, № 2, p. 560–564.
6. Lindahl L., Yamamoto M., Nomura M., Kirschbaum J. B., Allet B., Rochaix J. D. Mapping of a cluster of genes for components of the transcriptional and translational machineries of *Escherichia coli*.— J. Mol. Biol., 1977, v. 109, № 1, p. 23–47.
7. Rosenberg M., Court D. Regulatory sequences involved in the promotion and termination of RNA transcription.— Ann. Rev. Genet., 1979, v. 13, p. 319–353.
8. An G., Frisen J. D. Characterization of promoter-cloning plasmids: analysis of operon structure in the *rif* region of *Escherichia coli* and isolation an enhanced internal promoter mutant.— J. Bacteriol., 1980, v. 144, № 3, p. 904–916.

Поступило в редакцию
13.IV.1981

PRIMARY STRUCTURE OF RNA POLYMERASE FROM *E. coli*. NUCLEOTIDE SEQUENCE OF *E. coli* DNA FRAGMENT CONTAINING A PART OF THE *rpoC* GENE AND THE CORRESPONDING C-TERMINAL AMINO ACID SEQUENCE OF β' -SUBUNIT

OVCHINNIKOV Yu. A., MONASTYRSKAYA G. S., GUBANOV V. V.,
SALOMATINA I. S., SHUVAEVA T. M., LIPKIN V. M., SVERDLOV E. D.

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

In the course of parallel studies on the primary structures of β - and β' -subunits of DNA-dependent RNA polymerase from *E. coli* and their structural genes *rpoBC*, the complete nucleotide sequence (2306 base pairs) of *EcoRI-HindIII* fragment coding for the β' -subunit C-terminus has been established. These data in conjunction with the results obtained by the methods of protein chemistry allow to determine 421 residues in the amino acid sequence of the β' -subunit C-terminal region.