



УДК 547.963.32.02+547.96.02

ПЕРВИЧНАЯ СТРУКТУРА ФРАГМЕНТА *EcoRI*-D ГЕНА *groC*
И СООТВЕТСТВУЮЩЕЙ ОБЛАСТИ β' -СУБЪЕДИНИЦЫ
РНК-ПОЛИМЕРАЗЫ *E. COLI*Монастырская Г. С., Гурьев С. О., Калинин Н. Ф.,
Сорокин А. В., Саломатина И. С., Шубаева Т. М.,
Липкин В. М., Свердлов Е. Д., Овчинников Ю. А.Институт биоорганической химии им. М. М. Шемякина
Академии наук СССР, Москва

В процессе параллельного изучения структуры *groB,C*-оперона и аминокислотной последовательности β - и β' -субъединиц РНК-полимеразы *E. coli* ранее мы установили полную нуклеотидную последовательность гена *groB* и аминокислотную последовательность β -субъединицы РНК-полимеразы *E. coli* [1–4], а также нуклеотидную последовательность двух фрагментов гена *groC*, кодирующих N- и C-концевые последовательности β' -субъединицы [4, 5].

В настоящей работе приводятся результаты определения первичной структуры последнего фрагмента из этой области — *EcoRI*-D-фрагмента

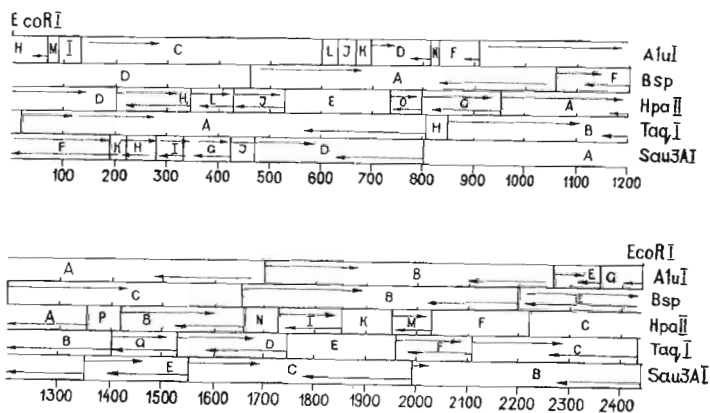


Рис. 1. Детальная карта расщепления фрагмента *EcoRI*-D рестрикционными эндонуклеазами *TaqI*, *BspI*, *AluI*, *HpaII*, *Sau3AI* и схема установления его последовательности. Субфрагменты, образующиеся при действии рестриктаз, изображены прямоугольниками. Стрелки обозначают длину установленной последовательности комплементарных цепей данного субфрагмента

ДНК *E. coli* (см. в [1] рис. 1), содержащего центральную и большую часть гена *groC*. Общая стратегия исследования аналогична примененной нами в работах [1, 2].

Схема установления нуклеотидной последовательности *EcoRI*-D-фрагмента дана на рис. 1. Большая часть структуры определена по двум цепям. Установленная последовательность, содержащая 2442 пары оснований, представлена на рис. 2.

Правильность определенной структуры подтверждается совпадением аминокислотной последовательности β' -субъединицы, выведенной нами

1-75 GAA TTC GAC GCG AAG ATG GGG GCG GAA GCA ATC CAG GCT CTG CTG AAG AGC ATG GAT CTG GAG CAA GAG TGC GAA
 1-25 Glu-Phe-Asp-Ala-Lys-Met-Gly-Ala-Glu-Ala-Ile-Gln-Ala-Leu-Leu-Lys-Ser-Met-Asp-Leu-Glu-Gln-Cys-Glu-
 76-150 CAG CTG CGT GAA GAG CTG AAC GAA ACC AAC TCC GAA ACC AAG CGT AAA AAG CTG ACC AAG CGT ATC AAA CTG CTG
 26-50 Gln-Leu-Arg-Glu-Glu-Leu-Asn-Glu-Thr-Asn-Ser-Glu-Thr-Lys-Arg-Lys-Leu-Thr-Lys-Arg-Ile-Lys-Leu-Leu-
 151-225 GAA GCG TTC GTT CAG TCT GGT AAC AAA CCA GAG TGG ATG ATC CTG ACC GTT CTG CCG GTA CTG CCG CCA GAT CTG
 51-75 Glu-Ala-Phe-Val-Gln-Ser-Gly-Asn-Lys-Pro-Glu-Trp-Met-Ile-Leu-Thr-Val-Leu-Pro-Val-Leu-Pro-Asp-Leu-
 226-300 CGT CCG CTG GTT CCG CTG GAT GGT GGT TTC GCG ACT TCT GAC CTG AAC GAT CTG TAT CGT CGC GTC ATT AAC
 76-100 Arg-Pro-Leu-Val-Pro-Leu-Asp-Gly-Gly-Arg-Phe-Ala-Thr-Ser-Asp-Leu-Asn-Asp-Leu-Tyr-Arg-Arg-Val-Ile-Asn-
 301-375 CGT AAC AAC CGT CTG AAA CGT CTG GAT CTG GCT GCG CCG GAC ATC ATC GTA CTG AAC GAA AAA CGT ATG CTG
 101-125 Arg-Asn-Asn-Arg-Leu-Lys-Arg-Leu-Leu-Asp-Leu-Ala-Ala-Pro-Asp-Ile-Ile-Val-Arg-Asn-Glu-Lys-Arg-Met-Leu-
 376-450 CAG GAA GCG GTA GAC GCC CTG GAT AAC GGT CGT GCG GGT CGT GCG ATC ACC GGT TCT AAC AAG CGT CCT CTG
 126-150 Gln-Glu-Ala-Val-Asp-Ala-Leu-Leu-Asp-Asn-Gly-Arg-Gly-Arg-Ala-Ile-Thr-Gly-Ser-Asn-Lys-Arg-Pro-Leu-
 451-525 AAA TCT TTG GCC GAC ATG ATC AAA GGT AAA CAG GGT TTC CGT CAG AAC CTG CTC GGT AAG CGT GTT GAC TAC
 151-175 Lys-Ser-Leu-Ala-Asp-Met-Ile-Lys-Gly-Lys-Gln-Gly-Arg-Phe-Arg-Gln-Asn-leu-Leu-Gly-Lys-Arg-Val-Asp-Tyr-
 526-600 TCC GGT CGT TCT GTA ATC ACC GTA GGT CCA TAC CTG CGT CTG CAT CAG TGC GGT CTG CCG AAG AAA ATG GCA CTG
 176-200 Ser-Gly-Arg-Ser-Val-Ile-Thr-Val-Gly-Pro-Tyr-Leu-Arg-Leu-His-Gln-Cys-Gly-Leu-Pro-Lys-Lys-Met-Ala-Leu-
 601-675 GAG CTG TTC AAA CCG TTC ATC TAC GGC AAG CTG GAA CTG CGT GGT CTT GCT ACC ACC ATT AAA GCT GCG AAG AAA
 201-225 Glu-Leu-Phe-Lys-Pro-Phe-Ile-Tyr-Gly-Lys-Leu-Glu-Lou-Arg-Gly-Leu-Ala-Thr-Ile-Lys-Ala-Ala-Lys-Lys-
 676-750 ATG GTT GAG CGC GAA GAA GCT GTT TGG GAT ATC CTG GAC GAA GTT ATC CGC GAA CAC CCG GTA CTG CTG AAC
 226-250 Met-Val-Glu-Arg-Glu-Glu-Ala-Val-Ile-Leu-Asp-Glu-Val-Ile-Arg-Glu-His-Pro-Val-Leu-Leu-Asn-
 751-825 CGT GCA CCG AET CTG CAC CGT CTG GGT ATC CAG GCA TTT GAA CCG GTA CTG ATC GAA GGT AAA GCT ATC CAG CTG
 251-275 Arg-Ala-Pro-Thr-Leu-His-Arg-Leu-Gly-Ile-Gln-Ala-Phe-Glu-Pro-Val-Leu-Ile-Glu-Gly-Lys-Ala-Ile-Gln-Leu-
 826-900 CAC CCG CTG GTT TGT GCG GCA TAT AAC GCC GAC TTC GAT GGT GAC CAG ATG GCT GTT CAC GTA CCG CTG ACG CTG
 276-300 His-Pro-Leu-Val-Cys-Ala-Ala-Tyr-Asn-Ala-Asp-Phe-Asp-Gly-Asp-Gln-Met-Ala-Val-His-Val-Pro-Leu-Thr-Leu-

Рис. 2

901-975 GAA GCC CAG CTG GAA GCG CGT GCG CTG ATG ATG TCT ACC AAC AAC ATC CTG TCC CCG GCG AAC GGC GAA CCA ATC
 301-325 Glu-Ala-Gln-Leu-Glu-Ala-Arg-Ala-Leu-Met-Ser-Thr-Asn-Asn-Ile-Leu-Ser-Pro-Ala-Asn-Gly-Glu-Pro-Ile-
 976-1050 ATC GTT CCG TCT CAG GAC GTT GTA CTG GGT CTG TAC TAC ATG ACC CGT GAC TGT GTT AAC GCC AAA GGC GAA GGC
 326-350 Ile-Val-Pro-Ser-Gln-Asp-Val-Val-Leu-Gly-Leu-Tyr-Met-Thr-Arg-Asp-Cys-Val-Asn-Ala-Lys-Gly-Glu-Gly-
 1051-1125 ATG GTG CTG ACT GGC CCG AAA GAA GCA GAA CGT CTG TAT CGC TCT GGT CTG GCT TCT CTG CAT GCG CGC GTT AAA
 351-375 Met-Val-Leu-Thr-Gly-Pro-Lis-Glu-Ala-Glu-Arg-Leu-Tyr-Arg-Ser-Gly-Leu-Ala-Ser-Leu-His-Ala-Arg-Val-Lis-
 1126-1200 GTG CGT ATC ACC GAG TAT GAA AAA GAT GCT AAC GGT GAA TTA GTA GCG AAA ACC AGC CTG AAA GAC ACG ACT GTT
 376-400 Val-Arg-Ile-Thr-Glu-Tyr-Glu-Lys-Asp-Ala-Asn-Gly-Glu-Leu-Val-Ala-Lys-Thr-Ser-Leu-Lys-Asp-Thr-Thr-Val-
 1201-1275 GGC CGT GCC ATT CTG TGG ATG ATT GTA CCG AAA GGT CTG CCT TAC TCC ATC GTC AAC CAG GCG CTG GGT AAA AAA
 401-425 Gly-Arg-Ala-Ile-Leu-Trp-Met-Ile-Val-Pro-Lys-Gly-Leu-Pro-Tyr-Ser-Ile-Val-Asn-Gln-Ala-Leu-Gly-Lys-Lys-
 1276-1350 GCA ATC TCC AAA ATG CTG AAC ACC TGC TAC CGC ATT CTC GGT CTG AAA CCG ACC GTT ATT TTT GCG GAC CAG ATC
 426-450 Ala-Ile-Ser-Lys-Met-Leu-Asn-Thr-Cys-Tyr-Arg-Ile-Leu-Gly-Leu-Lys-Pro-Thr-Val-Ile-Phe-Ala-Asp-Gln-Ile-
 1351-1425 ATG TAC ACC GGC TTC GCC TAT GCA GCG CGT TCT GGT GCA TCT GGT ATC GAT GAC ATG GTC ATC CCG GAG AAG
 451-475 Met-Tyr-Thr-Gly-Phe-Ala-Tyr-Ala-Ala-Arg-Ser-Gly-Ala-Ser-Val-Gly-Ile-Asp-Asp-Met-Val-Ile-Pro-Glu-Lys-
 1426-1500 AAA CAC GAA ATC ATC TCC GAG GCA GAA GCA GAA GTT GCT GAA ATT CAG GAG CAG TTC CAG TCT GGT CTG GTA ACT
 476-500 Lys-His-Glu-Ile-Ile-Ser-Glu-Ala-Glu-Ala-Glu-Val-Ala-Glu-Ile-Gln-Glu-Phe-Gln-Ser-Gly-Leu-Val-Thr-
 1501-1575 GCG GGC GAA CCG TAC AAC AAA GTT ATC GAT ATC TGG GCT GCG GCG AAC GAT CGT GTA TCC AAA GCG ATG ATG GAT
 501-525 Ala-Gly-Glu-Arg-Tyr-Asn-Lys-Val-Ile-Asp-Ile-Trp-Ala-Ala-Ala-Asn-Asp-Arg-Val-Ser-Lys-Ala-Met-Met-Asp-
 1576-1650 AAC CTG CAA ACT GAA ACC GTG ATT AAC CGT GAC GGT CAG GAA GAG AAG CAG GTT TCC TTC AAC AGC ATC TAC ATG
 526-550 Asn-Leu-Gln-Thr-Glu-Thr-Val-Ile-Asn-Arg-Asp-Gly-Gln-Glu-Lys-Gln-Val-Ser-Phe-Asn-Ser-Ile-Tyr-Met-
 1651-1725 ATG GCC GAC TCC GGT GCG CGT GGT TCT GCG GCA CAG ATT CGT CAG CTT GCT GGT ATG CGT GGT CTG ATG GCG AAG
 551-575 Met-Ala-Asp-Ser-Gly-Ala-Arg-Gly-Ser-Ala-Ala-Gln-Ile-Arg-Gln-Leu-Ala-Gly-Met-Arg-Gly-Leu-Met-Ala-Lys-
 1726-1800 CCG GAT GGC TCC ATC ATC GAA ACG CCA ATC ACC GCG AAC TTC CGT GAA GGT CTG AAC GTA CTC CAG TAC TTC ATC
 576-600 Pro-Asp-Gly-Ser-Ile-Ile-Glu-Thr-Pro-Ile-Thr-Ala-Asn-Phe-Arg-Glu-Gly-Leu-Asn-Val-Leu-Gln-Tyr-Phe-Ile-

1801-1875 TCC ACC CAC GGT GCT CGT AAA GGT CTG GCG GAT ACC GCA CTG AAA ACT GCG AAC TCC GGT TAC CTG ACT CCG CCG Ser-Thr-His-Gly-Ala-Arg-Lys-Gly-Leu-Ala-Asp-Thr-Ala-Lys-Thr-Ala-Asn-Ser-Gly-Tyr-Leu-Thr-Arg-Arg-Val-601-625

1876-1950 CTG GTT GAC GTG GCG CAG GAC CTG GTG GTT ACC GAA GAC GAT TGT GGT ACC CAT GAA GGT ATC ATG ATG ACT CCG Leu-Val-Asp-Val-Ala-Gln-Asp-Leu-Val-Thr-Glu-Asp-Cys-Gly-Thr-His-Glu-Gly-Ile-Met-Met-Thr-Pro-626-650

1951-2025 GTT ATC GAG GGT GGT GAC GTT AAA GAG CCG CTG CCG GAT CCG GTA CTG GGT GTA ACT GCT GAA GAC GTT CTG Val-Ile-Glu-Gly-Gly-Asp-Val-Lys-Glu-Pro-Leu-Arg-Asp-Arg-Val-Leu-Gly-Arg-Val-Thr-Ala-Glu-Asp-Val-Leu-651-675

2026-2100 AAG CCG GGT ACT GCT GAT ATC CTC GTT CCG CCG AAC ACG CTG CTG CAG GAA CAG TGG TGT GAC CTG CTG GAA GAG Lys-Pro-Gly-Thr-Ala-Asp-Ile-Leu-Val-Pro-Arg-Asn-Thr-Leu-His-Glu-Gln-Trp-Cys-Asp-Leu-Leu-Glu-Glu-676-700

2101-2175 AAC TCT GTC GAC GCG GTT AAA GTA CGT TCT GTT GAT ACC GAC TTT GGT GTA TGT GCG CAC TGC TAC Asn-Ser-Val-Asp-Ala-Val-Lys-Val-Arg-Ser-Val-Val-Ser-Cys-Asp-Thr-Asp-Phe-Gly-Val-Cys-Ala-His-Cys-Tyr-701-725

2176-2250 GGT CGT GAC CTG GCG CGT GGC CAC ATC ATC AAC AAG GGT GAA GCA ATC GGT GTT ATC GCG GCA CAG TCC ATC GGT Gly-Arg-Asp-Leu-Ala-Arg-Gly-His-Ile-Ile-Asn-Lys-Gly-Glu-Ala-Ile-Gly-Val-Ile-Ala-Ala-Gln-Ser-Ile-Gly-726-750

2251-2325 GAA CCG GGT ACA CAG CTG ACC ATG CGT ACC TTC CAC ATC GGT GGT GCG GCA TCT CGT GCG GCT GCT GAA TCC AGC Glu-Pro-Gly-Thr-Gln-Leu-Thr-Met-Arg-Thr-Phe-His-Ile-Gly-Gly-Ala-Ala-Ser-Arg-Ala-Ala-Glu-Ser-Ser-751-775

2326-2400 ATC CAA GTG AAA AAC AAA GGT AGC ATC AAG CTC AGC AAC GTG AAG TCG GTT GTG AAC TCC AGC GGT AAA CTG GTT Ile-Gln-Val-Lys-Asn-Lys-Gly-Ser-Ile-Lys-Leu-Ser-Asn-Val-Lys-Ser-Val-Val-Asn-Ser-Ser-Gly-Lys-Leu-Val-776-800

2401-2442 ATC ACT TCC CGT AAT ACT GAA CTG AAA CTG ATC GAC GAA TTC Ile-Thr-Ser-Arg-Asn-Thr-Glu-Leu-Lys-Leu-Ile-Asp-Glu-Phe-801-814

Рис. 2. Нуклеотидная последовательность фрагмента *Eco* RI-D и аминокислотная последовательность центральной области β'-субъединицы РНК-полимеразы. Приведена последовательность комплементарной цепи ДНК, адекватная последовательности мРНК. Ориентировка фрагмента *Eco* RI-D соответствует приведенной на рис. 1. Подчеркнуты аминокислотные последовательности, структура которых установлена анализом соответствующих пептидов. С* - нуклеотид 5-метилтиозина

из нуклеотидной, с аминокислотными последовательностями пептидов, полученных при расщеплении белка трипсином и бромцианом.

Установление структуры фрагмента *EcoRI*-D завершает определение первичной структуры оперона *rpoC* *E.coli*, кодирующего β - и β' -субъединицы РНК-полимеразы.

ЛИТЕРАТУРА

1. Ovchinnikov Yu. A., Monastyrskaya G. S., Gubanov V. V., Guryev S. O., Chertov O. Yu., Modyanov N. N., Grinkevich V. A., Makarova I. A., Marchenko T. V., Polovnikova I. N., Lipkin V. M., Sverdlov E. D. *Бур. Ж. Биохем.*, 1984, т. 116, № 3, с. 621—629.
2. Овчинников Ю. А., Свердлов Е. Д., Липкин В. М., Монастырская Г. С., Чертов О. Ю., Губанов В. В., Гурьев С. О., Модянов Н. Н., Гринкевич В. А., Макарова И. А., Марченко Т. В., Половникова И. И. *Биоорг. химия*, 1980, т. 6, № 5, с. 655—665.
3. Монастырская Г. С., Губанов В. В., Гурьев С. О., Липкин В. М., Свердлов Е. Д. *Биоорг. химия*, 1980, т. 6, № 7, с. 1106—1109.
4. Монастырская Г. С., Губанов В. В., Гурьев С. О., Липкин В. М., Свердлов Е. Д. *Биоорг. химия*, 1980, т. 6, № 9, с. 1423—1426.
5. Овчинников Ю. А., Монастырская Г. С., Губанов В. В., Саломатина И. С., Шувалева Т. М., Липкин В. М., Свердлов Е. Д. *Биоорг. химия*, 1981, т. 7, № 7, с. 1107—1112.

Поступило в редакцию
8.IX.1981

PRIMARY STRUCTURE OF *EcoRI*-D FRAGMENT OF *rpoC* GENE AND CORRESPONDING FRAGMENT OF β' -SUBUNIT OF RNA POLYMERASE FROM *E. COLI*

MONASTYRSKAYA G. S., GURJEV S. O., KALININA N. F.,
SOROKIN A. V., SALOMATINA I. S., SHUVAEVA T. M.,
LIPKIN V. M., SVERDLOV E. D., OVCHINNIKOV Yu. A.

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

In the course of the parallel study of the primary structures of β - and β' -subunits of DNA-dependent RNA polymerase from *E.coli* and their structural genes, the nucleotide sequence (2442 base pairs) of the last fragment — *EcoRI*-D fragment of gene *rpoC* was established. This allowed to complete the determination of primary structure of the operon *rpoC* *E.coli*, coding β - and β' -subunits of DNA-dependent RNA polymerase.

Технический редактор *Е. С. Кузьмишкина*

Сдано в набор 20.10.81. Подписано к печати 14.12.81 Т-28955 Формат бумаги 70×108¹/₁₆
Высокая печать Усл. печ. л. 12,6 Усл. кр.-отт. 10,9 тыс. Уч.-изд. л. 15.1 Бум. л. 4,5
Тираж 845 экз. Зак. 938

Издательство «Наука», 103062, Москва, К-62, Подсосенский пер., 21
2-я типография издательства «Наука», 121099 Москва, Г-99, Шубинский пер., 10