



УДК 547.963.32.02+547.96.02

ПЕРВИЧНАЯ СТРУКТУРА РНК-ПОЛИМЕРАЗЫ *E. coli*.
НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ФРАГМЕНТА *EcoRI*-С
ГЕНА *groV* И АМИНОКИСЛОТНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ
СООТВЕТСТВУЮЩЕГО ФРАГМЕНТА β -СУБЪЕДИНИЦЫ

*Овчинников Ю. А., Свердлов Е. Д., Дипкин 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. Н.*

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

В результате параллельного изучения первичных структур β -субъединицы ДНК-зависимой РНК-полимеразы *E. coli* и ее структурного гена *groV* установлена полная нуклеотидная последовательность (2874 пары оснований) фрагмента *EcoRI*-С гена *groV*. Соответствующая ей аминокислотная последовательность средней части β -субъединицы содержит 958 аминокислотных остатков. Сочетание методов белковой и нуклеотидной химии при совместном структурном исследовании белка и ДНК резко ускоряет и значительно упрощает решение обеих задач и, главное, увеличивает надежность структурного анализа.

Для понимания механизмов транскрипции необходимо иметь детальную информацию об организации активных центров РНК-полимеразы на разных этапах синтеза РНК. Такая информация может быть получена только на основании изучения первичной и пространственной структуры фермента.

Ранее мы установили аминокислотную последовательность α -субъединицы ДНК-зависимой РНК-полимеразы *E. coli* [1]. При этом использовались исключительно методы белковой химии. Однако в процессе установления первичной структуры β - и β' -субъединиц, имеющих значительно больший молекулярный вес (~155 000 и 165 000) [2], возникает ряд методических трудностей [3]. В частности, резко затрудняется выделение и очистка фрагментов, а также реконструкция аминокислотной последовательности по «перекрывающимся» пептидам.

Развитие методов анализа последовательности ДНК сделало возможным на основе генетического кода выводить гипотетические аминокислотные последовательности, исходя из известных нуклеотидных. При реализации такого подхода следует, однако, иметь в виду целый ряд ограничений и возможные источники ошибок. Во-первых, выведенная из нуклеотидной аминокислотная последовательность может не соответствовать реальной вследствие процессинга — превращения белка-предшественника в конечный белок. Во-вторых, процессинг на уровне информационной РНК в свою

* В обозначениях дезоксирибонуклеотидов символ «d» везде опущен.

очередь может приводить к отличию истинной структуры белка от выводимой. Особенно это касается эукариот, для которых обнаружено явление «сплайсинга», т. е. превращение мРНК-предшественника в конечную мРНК путем делеции фрагментов из ее внутренних областей. Хотя у прокариот «сплайсинг» пока не обнаружен, окончательно исключить его нельзя. В-третьих, исходя из структуры ДНК, весьма часто трудно обнаружить начало структурного гена. Критерием такого поиска служит наличие иницирующего кодона в сочетании с близлежащей последовательностью, комплементарной 3'-концу 16S РНК [4, 5]. Однако часто для одного и того же белка можно обнаружить два и более подобных сочетаний. Нами и другими авторами [6, 7] были замечены такие случаи, в частности при определении последовательности гена «0» — белка фага λ . Вероятность таких совпадений, естественно, увеличивается с увеличением длины определяемой последовательности. Кроме того, следует иметь в виду, что лишь одна ошибка в определении последовательности ДНК (пропуск или вставка) приводит к совершенно неправильной аминокислотной последовательности белка. Такие ошибки возможны, в частности, вследствие влияния особенностей вторичной структуры и наличия модифицированных оснований в ДНК.

Таким образом, установление первичной структуры ДНК не может заменить непосредственное исследование аминокислотной последовательности кодируемого ею белка. В то же время параллельное изучение первичных структур белка и ДНК должно быть чрезвычайно эффективным. Ввиду этого для определения первичной структуры β - и β' -субъединиц РНК-полимеразы мы решили использовать сочетание методов белковой и нуклеотидной химии, независимо определяя последовательности структурных генов *proB* (β -субъединица) и *proC* (β' -субъединица) и соответствующих белков. Знание нуклеотидной последовательности соответствующих участков ДНК даст возможность расположить изученные пептидные фрагменты в непрерывную полипептидную цепь. Такой подход облегчит решение наиболее сложной проблемы в изучении первичной структуры белка большого молекулярного веса — проблемы реконструкция аминокислотной последовательности, требующей в классическом варианте выделения и установления структуры большого числа относительно крупных «перекрывающихся» пептидов. Основной задачей при таком подходе становится выделение и определение аминокислотной последовательности пептидов, покрывающих всю полипептидную цепь белка. Для этого можно использовать пептиды сравнительно небольшого молекулярного веса (5–20 аминокислотных остатков). Такие пептиды обычно получают при гидролизе белка трипсином или протеазой из *St. aureus*; методы их разделения хорошо разработаны, и для определения их структуры можно обойтись небольшим количеством материала (до 10–30 нмоль пептида).

В настоящее время разработаны пути полного [8] или частичного [9] введения оперона *proBC* *E. coli*, содержащего гены β - и β' -субъединиц, в состав ДНК трансдуцирующих фагов. Это обеспечило возможность его функциональных исследований и упростило проблему структурного анализа.

Рядом исследователей получена карта расщепления области ДНК *E. coli*, содержащей структурные гены β - и β' -субъединиц РНК-полимеразы, рестрикционной эндонуклеазой *EcoRI* [10–12] (см. рис. 1). Недавно Пост и др. [13] установили полную структуру фрагмента *EcoRI*-E и большую часть структуры фрагмента *EcoRI*-G, содержащих промоторы оперона *proBC* и генов некоторых рибосомных белков, а также начало структурного гена β -субъединицы. Позже А. И. Гуревичем и др. [14] структура фрагмента *EcoRI*-G была установлена полностью.

В данной работе определена нуклеотидная последовательность *EcoRI*-C-фрагмента гена *proB*, составляющего его среднюю, большую часть

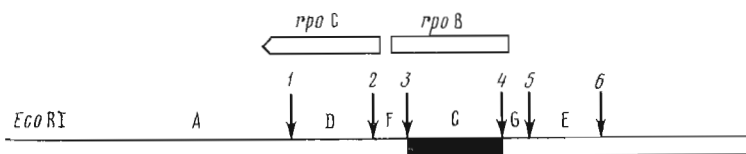


Рис. 1. Карта расщепления области ДНК *E. coli*, содержащей структурные гены β - и β' -субъединиц РНК-полимеразы (*rpoB* и *rpoC* соответственно), рестрикционной эндонуклеазой *EcoRI*. Фрагмент *EcoRI-C* зачернен. Буквенные обозначения фрагментов соответствуют принятым в работе [12]

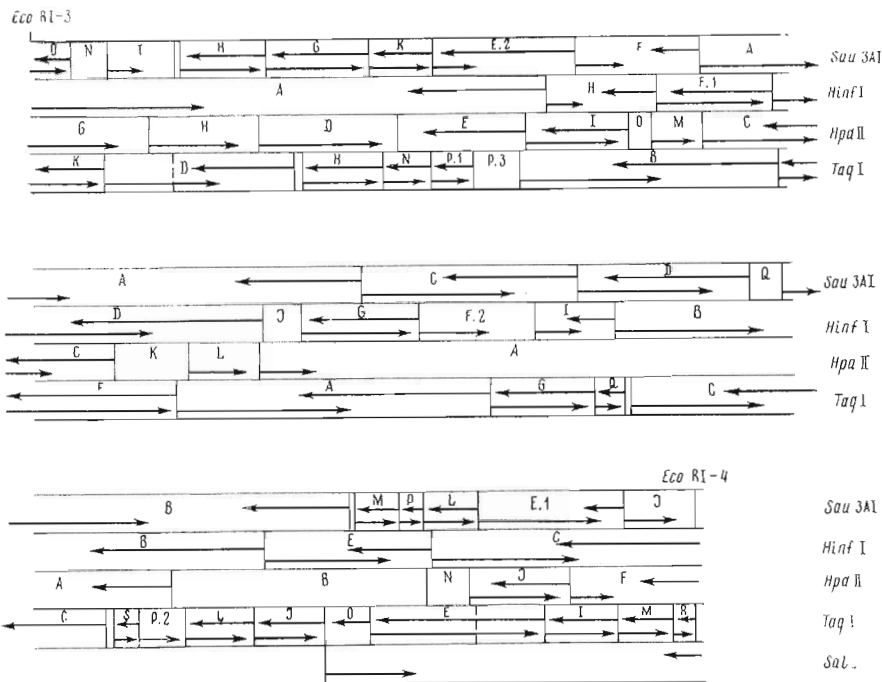


Рис. 2. Детальная карта расщепления фрагмента *EcoRI-C* рестрикционными эндонуклеазами и схема установления его последовательности. Ориентировка фрагмента соответствует ориентировке оперона *rpoBC* на рис. 1. Субфрагменты, получающиеся при действии рестриктаз, изображены прямоугольниками. Стрелки обозначают установленные части последовательностей комплементарных цепей данного субфрагмента. В субфрагментах *Taq-D* и *Taq-E* вертикальные пунктирные линии обозначают положение последовательности ТСГА, не расщепляемой *TaqI*

(2874 п.о. *). Из сопоставления полученной структуры с аминокислотными последовательностями пептидов, выделенных при расщеплении β -субъединицы, выведена последовательность 958 аминокислотных остатков β -субъединицы РНК-полимеразы *E. coli*.

Стратегия установления структуры фрагмента *EcoRI-C*

Общая стратегия определения нуклеотидной последовательности фрагмента *EcoRI-C*, представленная на рис. 2, аналогична стратегии, примененной Сатклиффом [15] при изучении плазмиды pBR-322 (карта участков расщепления рестрикционными эндонуклеазами предварительно не составлялась). Фрагмент поочередно гидролизовали одной из рестриктаз, расщепляющих ДНК на сравнительно короткие фрагменты (*Sau3A1*, *HinfI*, *HpaII*, *TaqI*), смесь фрагментов фосфорилировали с помощью $[\gamma\text{-}^{32}\text{P}]\text{гАТФ}$ и полинуклеотидкиназы фага T4 и разделяли электрофорезом в полиакриламидном геле. Как правило, определяли структуру обеих ком-

* п. о. — пар оснований.

	10	20	30	40	50	60
I-60	GAATTCGATCCGAAGGACAACCTGTTTCGTACGTATCGACCGTCGCCGTA	AACTGCCTGGC				
	GluPheAspProLysAspAsnLeuPheValArgIleAspArgArgArgLysLeuProAla					
61-120	ACCATCATTCTGCGCCCTGAACTACACCACAGAGCAGATCCTCGACCTGTTCTTTGAA					
	ThrIleIleLeuArgAlaLeuAsnTyrThrThrGluGlnIleLeuAspLeuPhePheGlu					
121-180	AAAGTTATCTTTGAAATCCCTGATAACAAGCTGCAGATCGAACTGGTCCCGAACCGCTG					
	<u>LysValIlePheGluIleArgAspAsnLysLeuGlnMetGluLeuValProGluArgLeu</u>					
181-240	CGTGGTGAACCCGATCTTTTGACATCGAAGCTAACCGTAAAGTGTACCTAGAAAAAGGC					
	<u>ArgGlyGluThrAlaSerPheAspIleGluAlaAsnGlyLysValTyrValGluLysGly</u>					
241-300	CGCCGTACTCTGCGCCACATTCCGCCAGCTGGAAAAAGACGACGCTCAA	ACTGATCGAA				
	ArgArgIleThrAlaArgHisIleArgGlnLeuGluLysAspAspValLysLeuIleGlu					
301-360	GTCCCGTTGAGTACATCGCAGGTAAGTGGTTGCCTAAAGACTATATTGATGAGTCTACC					
	ValProValGluTyrIleAlaGlyLysValValAlaLysAspTyrIleAspGluSerThr					
361-420	GCGGAGCTGATCTGCCGACGGAACATGGAGCTGAGCC*GGATCTGCTGGCTAAGCTGAGC					
	<u>GlyGluLeuIleCysAlaAlaAsnMetGluLeuSerLeuAspLeuLeuAlaLysLeuSer</u>					
421-480	CAGTCTGGTCAACCGTATCGAAACCGCTGTTACCAACGATCTGGATCAGGCCCATAT					
	<u>GlnSerGlyHisLysArgIleGluThrLeuPheThrAsnAspLeuAspHisGlyProTyr</u>					
481-540	ATCTCTGAAACCTTACGTGTGACCCAACTAACGACCGTCTGAGCCACTCGTAGAAATC					
	<u>IleSerGluThrLeuArgValAspProThrAsnAspArgLeuSerAlaLeuValGluIle</u>					
541-600	TACCGCATGATGCCCT*GGCGAGCCCGGACTCGTGAACGAGCTGAAAGCCTGTTTCGAG					
	<u>TyrArgMetMetArgProGlyGluProProThrArgGluAlaAlaGluSerLeuPheGlu</u>					
601-660	AACCTGTTCTTCTCCGAAGACCGTTATGACTTGTCTGCGGTTGGTCTGATGAAGTCCAAC					
	<u>AsnLeuPhePheSerGluAspArgTyrAspLeuSerAlaValGlyArgMetLysPheAsn</u>					
661-720	CGTTCTCTGCTGCCGAAGAATCGAAGGTTCCGGTATCCTGAGCAAAGACGACATCATT					
	<u>ArgSerLeuLeuArgGluGluIleGluGlySerGlyIleLeuSerLysAspAspIleIle</u>					
721-780	GATGTTATGAAAAAGCTCATCGATATCCGTAACGGTAAAGCGGAAGTCGATGATATCGAC					
	<u>AspValMetLysLysLeuIleAspIleArgAsnGlyLysGlyGluValAspAspIleAsp</u>					
781-840	CACCTCGGCAACCGTCTGATCCGTTCCGTTGGCGAAATGGCGAAAACCACTCCCGGTT					
	<u>HisLeuGlyAsnArgArgIleArgSerValGlyGluMetAlaGluAsnGlnPheArgVal</u>					
841-900	GGCCTGGTACGTGTAGAGCGTGGCGTGAAGAGCGTCTGTCTCTGGCGGATCTGGATACC					
	<u>GlyLeuValArgValGluArgAlaValLysGluArgLeuSerLeuGlyAspLeuAspThr</u>					
901-960	CTGATGCCACAGGATATGATCAACGCCAACCCGATTTCCGCAGCAGTGAAGAGTCTCTC					
	<u>LeuMetProGlnAspMetIleAsnAlaLysProIleSerAlaAlaValLysGluPhePhe</u>					
961-1020	GGTTCACGCCAGCTGTCTCAGTTTATGGTCCAGAACAACCCGCTGTCTGAGATTACGCAC					
	<u>GlySerSerGlnLeuSerGlnPheMetValGlnAsnAsnProLeuSerGluIleThrHis</u>					
1021-1080	AAACGTCGTATCTCCGCACTCGGCC*AGCGGCTCTGACCCGTGAACCTGCAGGCTTCGAA					
	<u>LysArgArgIleSerAlaLeuGlyProGlyGlyLeuThrArgGluArgAlaGlyPheGlu</u>					
1081-1140	GTTCCGAGACGTACACCCGACTCACTACGGTCCGGTATGTCCAATCGAAACCCCTGAAGGT					
	<u>ValArgAspValHisProThrHisTyrGlyArgValCysProIleGluThrProGluGly</u>					
1141-1200	CGAAACATCGGTCTGATCAACTCTCTGTCGGTGTACGCCAGACTAACGAATACGGCTTC					
	<u>ProAsnIleGlyLeuIleAsnSerLeuSerValTyrAlaGlnThrAsnGluTyrGlyPhe</u>					
1201-1260	CTTGAGACTCCGTATCGTAAAGTGACCGACGGTGTGTAACTGACGAAATTCACTACCTG					
	<u>LeuGluThrProTyrArgLysValThrAspGlyValValThrAspGluIleHisTyrLeu</u>					
1261-1320	TCTGCTATCGAAGAAGGCAACTACGTTATCGCC*AGGCGAATCCAACCTGGATGAAGAA					
	<u>SerAlaIleGluGluGlyAsnTyrValIleAlaGlnAlaAsnSerAsnLeuAspGluGlu</u>					
1321-1380	GGCCACTTCGTAGAAGACCTGGTAACTTCCCGTAGCAAAGGCGAATCCAGCTTGTTCCAGC					
	<u>GlyHisPheValGluAspLeuValThrCysArgSerLysGlyGluSerSerLeuPheSer</u>					
1381-1440	CGCGAC*AGGTTGACTACATGGACGTATCCACCCAGCAGGTGCTATCCGTCGGTCCGCTCC					
	<u>ArgAspGlnValAspTyrMetAspValSerThrGlnGlnValValSerValGlyAlaSer</u>					
1441-1500	CTGATCCCGTTCCTGGAAACACGATGACGCCAACCGTGCATTGATGGGTGCCAAACATGCAA					
	<u>LeuIleProPheLeuGluHisAspAspAlaAsnArgAlaLeuMetGlyAlaAsnMetGln</u>					

	10	20	30	40	50	60
150I-1560	CCTCAGGCCGTTCCGACTCTGCCGCTGATAAGCCGCTGGTTGGTACTGGTATGGAACCT	ArgGlnAlaValProThrLeuArgAlaAspLysProLeuValGlyThrGlyMetGluArg				
156I-1620	GCTGTTGCCGTTGACTCCGCTGTAACCTGCCGTAGCTAACCGTGGTGGTCTCCTCAGTAC	AlaValAlaValAspSerGlyValThrAlaValAlaLysArgGlyGlyValValGlnTyr				
162I-1680	GTGGATGCTTCCCGTATCGTTATCAAAGTTAACGAAGACGAGATCTATCCGGGTGAAGCA	ValAspAlaSerArgIleValIleLysValAsnGluAspGluMetTyrProGlyGluAla				
168I-1740	GGTATCGACATCTACAACCTGACCAAAATACACCCCTTCTAACCCAGAACACCTGTATCAAC	GlyIleAspIleTyrAsnLeuThrLysTyrThrArgSerAsnGlnAsnThrCysIleAsn				
174I-1800	CAGATGCCGTGTGTCTCTGGGTGAACCGCTGAAACGTGGCGACCTGCTGGCAGACCGT	GlnMetProCysValSerLeuGlyGluProValGluArgGlyAspValLeuAlaAspGly				
180I-1860	CCGTCACCCGACCTCGGTGAACTGCCGCTTGGTCAGAACATGCCGCTACCGTTCATCCCG	ProSerThrAspLeuGlyLeuAlaLeuGlyGlnAsnMetArgValAlaAlaMetPro				
186I-1920	TGGAATGGTTACAACCTCGAAGACTCCATCCTCGTATCCGACCGTGTGTTGTTGAGGAAGC	TrpAsnGlyTyrAsnPheGluAspSerIleLeuValSerGluArgValValGlnGluAsp				
192I-1980	CGTTTCACCACATCCACATTCAGGAACCTGGCGTGTGTCTCCCGTGACACCAAGCTGGGT	ArgPheThrThrIleHisIleGlnGluLeuAlaCysValSerArgAspThrLysLeuGly				
198I-2040	CCGGAAGAGATCACCGCTGACATCCCGAACCTGGGTGAAGCTGCGCTCTCCAACTGGAT	ProGluGluIleThrAlaAspIleProAsnValGlyGluAlaAlaLeuSerLysLeuAsp				
204I-2100	GAATCCGGTATCGTTTACATTGCTGCCGAAGTGACCGGTGGCGACATTCTGGTTGTTAAG	GluSerGlyIleValTyrIleGlyAlaGluValThrGlyGlyAspIleLeuValGlyLys				
210I-2160	GTAACGCCGAAAGGTGAAACTCAGCTGACCCGAGAAGAAAACTGCTGCGTGGCATCTTC	ValThrProLysGlyGluThrGlnLeuThrProGluGluLysLeuLeuArgAlaIlePhe				
216I-2220	GGTGAGAAAGCCTCTGACGTTAAAGACTCTTCTCTGCCGCTACCAAACCGTGTATCCGCT	GlyGluLysAlaSerAspValLysAspSerSerLeuArgValProAsnGlyValSerGly				
222I-2280	ACGGTTATCGACGTTACGCTCTTACTCGCGATGGCGTAGAAAAAGACAAACCTGCCGCTG	ThrValIleAspValGlnValPheThrArgAspGlyValGluLysAspLysArgAlaLeu				
228I-2340	GAAATCGAAGAAATGCAGCTCAAACAGCGAAGAAAGACCTGTCTGAAGAACTGCAGATC	GluIleGluGluMetGlnLeuLysClnAlaLysLysAspLeuSerGluGluLeuGlnIle				
234I-2400	CTCGAAGCGGTCTGTTTCAAGCGTATCCGTGCTGTCTGCTGAGCGGTGGCGTTGAAGCT	LeuGluAlaGlyLeuPheSerArgIleArgAlaValLeuValAlaGlyGlyValGluAla				
240I-2460	GAGAAGCTCGACAAACTGCCGCGGATCGCTGGCTGGAGCTGGCGCTGACAGACGAAGAG	GluLysLeuAspLysLeuProArgAspArgTrpLeuGluLeuGlyLeuThrAspGluGlu				
246I-2520	AAACAAAATCAGCTGGAACAGCTGGCTGAGCAGTATGACGAACTGAAACACGAGTTCGAG	LysGlnAsnGlnLeuGluGlnLeuAlaGluGlnTyrAspGluLeuLysHisGluPheGlu				
252I-2580	AAGAACTCGAAGCGAAACGCCGCAAAATCACC [*] AGGGCGCAGCATGTCGACCGGGCGTG	LysLysLeuGluAlaLysArgArgLysIleThrGlnGlyAspAspLeuAlaProGlyVal				
258I-2640	CTGAAGATTGTTAAGTATATCTGGCGGTTAAACGCCGATCCAGC [*] TGGTGACAAGATG	LeuLysIleValLysValTyrLeuAlaValLysArgArgIleGlnProGlyAspLysMet				
264I-2700	GCAGTCTGTCACGGTAAACAAGGTGTAATTTCTAACATCAACCCGATCGAAGATATGCCT	AlaGlyArgHisGlyAsnLysGlyValIleSerLysIleAsnProIleGluAspMetPro				
270I-2760	TACGATGAAAACGCTACGCCGTAGACATCGTACTGAACCCGCTGGCGTACCGTCTCGT	TyrAspGluAsnGlyThrProValAspIleValLeuAsnProLeuGlyValProSerArg				
276I-2820	ATGAACATCGGTGAGTCTCGAAACCCAC [*] TGGGTATGGCTGCGAAAGTATCGCGCAC	MetAsnIleGlyGlnIleLeuGluThrHisLeuGlyMetAlaAlaLysGlyIleGlyAsp				
282I-2874	AAGATCAACGCCATGCTGAAACAGCAGCAAGAAGTCCGGAACCTGCGCGAATTC	LysIleAsnAlaMetLeuLysGlnGlnGlnGluValAlaLysLeuArgGluPhe				

Рис. 3. Нуклеотидная последовательность фрагмента *EcoRI*-С и аминокислотная последовательность соответствующей части 3-субъединицы РНК-полимеразы. Приведена последовательность комплементарной цепи ДНК, адекватная последовательности мРНК. Ориентировка фрагмента *EcoRI*-С противоположна приведенной на рис. 1 и 2. Подчеркнуты аминокислотные последовательности, структура которых установлена анализом соответствующих пептидов. С* — остатки 5-метилцитидина

плементарных цепей, получаемых после денатурации каждого субфрагмента и разделения в полиакриламидном геле. Всего по обоим комплементарным цепям было определено 96% структуры. В случае субфрагмента *Sau*-F (рис. 2), цепи которого разделить не удавалось, было применено вторичное расщепление рестриктазой *Hpa*II. Реконструкция структуры фрагмента *Eco*RI-C была осуществлена по «перекрывающимся» последовательностям субфрагментов.

Определение последовательности субфрагментов

Определение последовательности комплементарных нитей было выполнено по методу Максама — Гилберта [16]. Однако вместо стандартной реакции $A > G$ мы использовали частичную апуринизацию муравьиной кислотой, предложенную нами ранее [17, 18] для определения положения пуриновых звеньев в олиго- и полинуклеотидах. Эта реакция затем в разных вариантах была использована нами [6] и другими авторами [19]. Она проста в исполнении и всегда дает надежные результаты.

При определении положения пиримидиновых звеньев с помощью гидразина во многих случаях затруднительна дифференцировка тимидина и цитидина. Это происходит вследствие малой специфичности реакции по цитидину, достигаемой в стандартной методике. Ввиду этого в ряде случаев для определения положения тимидиновых звеньев мы использовали реакцию с четырехокисью осмия в присутствии аминов [20]. Ранее мы детально исследовали применение этой модификации для установления положения тимидиновых звеньев в олиго- и полинуклеотидах [21]. Несколько позже эту реакцию использовал Браун [22]. В большинстве случаев она дает хорошие результаты, однако модификация тимидина четырехокисью осмия, по-видимому, весьма чувствительна к особенностям вторичной структуры. Вследствие этого в некоторых случаях наблюдаются различия в степени модификации остатков тимидина в зависимости от их положения в полинуклеотидной цепи.

При определении структуры *Eco*RI-C в 11 положениях нами был обнаружен 5-метилцитидин (С*) (рис. 3). Во всех случаях это модифицированное основание находится в последовательности $5'-CC^*_{\text{T}}GG-3'$, что согласуется с данными Максама [23]. После гидразинолиза в структурном геле полосы, соответствующие 5-метилцитидину, проявляются слабо либо вообще отсутствуют. При определении же последовательности с помощью четырехокиси осмия 5-метилцитидин модифицируется аналогично тимидину. Возможные ошибки в этих случаях исключались благодаря анализу комплементарных цепей. Кроме того, во многих случаях правильность выведенной последовательности подтверждалась сопоставлением с аминокислотной последовательностью соответствующих пептидов.

Другим примером отклонения от регулярного расположения полос в структурном геле является компрессия. Яркий ее пример встретился нам при определении структуры субфрагмента *Taq*-J (область 562—564 н.о.). Три нуклеотида — GAG — в этой области давали только две полосы. Это, по-видимому, связано с возможностью образования «шпильки» с участием этих нуклеотидов. Подтверждением правильности публикуемой структуры в этом участке является последовательность комплементарной цепи и аминокислотная последовательность пептида, перекрывающего этот участок (см. рис. 3).

Определение значащей цепи и установление «рамки» трансляции

Выбор значащей цепи в случае фрагмента *Eco*RI-C легко осуществляется на основании асимметричного расположения единственного *Sal*I-сайта. Согласно физико-генетическим картам области оперона *gro*BC [10—12],

этот сайт занимает положение, проксимальное по отношению к промотору оперона. Следовательно, значащей является цепь ДНК, 3'-конец которой расположен ближе к участку расщепления *SalI*. Последовательностью, адекватной последовательности мРНК, обладает комплементарная цепь.

Поиск «рамки» кодирования аминокислотной последовательности в этой цепи был осуществлен с помощью компьютера. Только в одной из трех возможных рамок не было получено терминирующих кодонов. Эта рамка и была взята за основу для выведения аминокислотной последовательности, приведенной на рис. 3.

В процессе установления нуклеотидной последовательности *EcoRI*-С производилось постоянное сравнение с аминокислотными последовательностями пептидов β -субъединицы, что позволяло устанавливать значащую цепь и рамку считывания в отдельных субфрагментах и помогало ориентировать их друг относительно друга.

Выделение и определение аминокислотной последовательности пептидов

Исследование первичной структуры β -субъединицы было начато с ограниченного триптического гидролиза, при котором β -субъединица, согласно данным аналитических экспериментов, расщеплялась на пять крупных фрагментов (*M* 62 000, 52 000, 37 000, 24 000 и 10 000). Однако в препаративном опыте наряду с этими фрагментами, выход которых был сравнительно невысок, образовалось большое число более мелких пептидов [3]. Из полученного гидролизата было выделено 53 низкомолекулярных пептида, содержащих 380 аминокислотных остатков. Для 46 пептидов установлена полная и для 7 — частичная аминокислотные последовательности [24]. Выделение высокомолекулярных пептидов ограниченного триптического гидролиза осложнилось из-за малой специфичности гидролиза и низкого выхода большинства пептидов.

В качестве другого метода расщепления полипептидной цепи β -субъединицы был выбран гидролиз протеиназой из *St. aureus*, которая с успехом использовалась нами при изучении первичной структуры α -субъединицы [4]. Фермент гидролизует в основном связи, образованные α -карбоксыльными группами глутаминовой кислоты. Гидролиз проводили в 0,1 М аммоний-бикарбонатном буфере (рН 8,0) и фермент-субстратном соотношении 1 : 60. Для улучшения растворимости β -субъединицы в этом буфере перед гидролизом карбоксиметилированный белок модифицировали цитраконовым ангидридом.

Первоначальное фракционирование гидролизата, содержащего ~150 пептидов, проводилось гель-фильтрацией на «биогеле Р-4». В результате были получены четыре объединенные фракции. Для последующего разделения фракций II, III и IV, содержащих сравнительно небольшие пептиды (2—20 аминокислотных остатков), использовали ионообменную хроматографию на катионите AG-50W \times 4 и хроматографию на бумаге. Фракция I состояла из смеси наиболее высокомолекулярных пептидов. С целью облегчения разделения и последующего анализа пептидного материала этой фракции смесь пептидов подвергалась дополнительному гидролизу химотрипсином. Гидролизат обессоливали на сефадексе G-10 и полученные пептидные фрагменты разделяли аналогично пептидам фракций II—IV.

Всего из фракции IV выделено 3, из фракции III — 73, из фракции II — 48 и из фракции I — 60 пептидных фрагментов. N-Концевую аминокислотную последовательность пептидов определяли методом Эдмана с идентификацией Pth- и Dns-производных аминокислот, а C-концевую — с помощью карбоксипептидазы Y. К настоящему времени установлена полная аминокислотная последовательность 83 и частичная 35 пептидов, содержащих 539 и 435 аминокислотных остатков соответственно.

Очевидно, что из такой сложной смеси, которая образовалась при гидролизе β -субъединицы протеиназой из *St. aureus*, невозможно выделить все пептиды. Поэтому для получения недостающих фрагментов был предпринят исчерпывающий гидролиз β -субъединицы трипсином по остаткам аргинина после модификации остатков лизина цитраконовым ангидридом. Разделение продуктов триптического гидролиза проводится в настоящее время.

Сочетание исследований первичной структуры ДНК со структурными исследованиями пептидов

На рис. 3 подчеркнуты аминокислотные последовательности, установленные в процессе изучения пептидов, полученных при расщеплении трипсином и протеиназой из *St. aureus* β -субъединицы РНК-полимеразы. Поиск соответствия между аминокислотной последовательностью пептидов и нуклеотидной последовательностью фрагментов ДНК проводился с помощью компьютера. Обычно проводили сравнение последовательности пептида с тремя возможными аминокислотными последовательностями, соответствующими данной нуклеотидной. В тех случаях, когда не было известно, какая цепь субфрагмента является значащей, исследовались шесть возможных аминокислотных последовательностей. Программа поиска была составлена таким образом, чтобы в процессе сравнения можно было выявлять возможные ошибки как в структуре ДНК, так и в структуре пептидов. Как показали результаты настоящего исследования, совместное изучение аминокислотной последовательности белка и нуклеотидной последовательности его структурного гена позволяет с помощью компьютера легко идентифицировать положение каждого пептидного фрагмента в полипептидной цепи β -субъединицы (рис. 3) и обеспечивает их «стыковку» между собой.

Данные по нуклеотидной последовательности помогают также выявлять возможные ошибки, возникающие в процессе изучения структуры пептидов. Так, для пептида Ser-Gly-Ile-Val-Tyr-Ile-Gly-Ala-Glu (см. координаты 2044—2070, рис. 3) первоначально была найдена последовательность Ser-Gly-Ile-Tyr-Tyr-Ile-Gly-Ala-Glu. Структура соответствующих субфрагментов ДНК указывала на наличие в четвертом положении остатка валина, что и было подтверждено при повторном анализе пептида. Ошибка была вызвана тем, что при кислотном гидролизе плохо расщепляется связь Val—Tyr, а хроматографическая подвижность образующегося в результате неполного гидролиза Dns-валил-тирозина аналогична подвижности Dns-тирозина.

С другой стороны, известные аминокислотные последовательности пептидов упрощали анализ нуклеотидной последовательности. Так, структуры некоторых субфрагментов (например, *HinfD* и *HinfJ*) могли быть объединены в непрерывные последовательности уже в начале анализа благодаря тому, что их концевые последовательности перекрывались с пептидами. Известная структура пептидов, как уже отмечалось выше, также помогала определять значащую цепь и «рамку» считывания в субфрагментах ДНК. Наиболее значительный результат совместного изучения — возможность на ранних стадиях исследования выявлять неточности и ошибки, возникающие в процессе исследования в отдельных участках нуклеотидной последовательности, с тем чтобы при последующем анализе уделять этим участкам особое внимание. Так, например, в структуре *EcoRI-C* (рис. 3) в положения 1596, 2049, 2051 и 2088 были вставлены пропущенные остатки тимидина, в положениях 1358 и 2761 гуанозин заменен на аденозин, после остатков тимидина в положениях 158 и 1650 выброшены ошибочно вставленные ранее тимидин и цитидин соответственно. Во всех случаях ошибки, выявленные первоначально путем сравнения со структурой соответствующих пептидов, были исправлены при последующем ана-

лизе субфрагментов ДНК. Некоторые причины таких ошибок (компрессия, метилирование цитозина) обсуждались выше.

Определенная к настоящему времени аминокислотная последовательность пептидов составляет ~60% от последовательности β -субъединицы, выведенной из структуры *EcoRI*-С фрагмента. При сопоставлении этих двух последовательностей следует иметь в виду, что аминокислотная последовательность β -субъединицы определяется нами для РНК-полимеразы, выделяемой из *E. coli* дикого типа, тогда как ДНК, используемая для определения нуклеотидной последовательности, содержит в гене *rpoB* *rif^r*-мутацию, обуславливающую доминантную устойчивость к рифампицину. По данным Номуры [11] и Коллинза [12], *rif^r*-мутация локализована в *EcoRI*-С-фрагменте. Имеющиеся данные пока не позволяют локализовать эту мутацию. Очевидно, она находится в одном из участков фрагмента, для которых структура пептидов еще не установлена.

Особенности нуклеотидной последовательности фрагмента *EcoRI*-С

Определенная нами структура содержит два участка, в которых тетра-нуклеотид TCGA, обычно узнаваемый и расщепляемый рестрикционной эндонуклеазой *TaqI*, не расщепляется (или расщепляется крайне медленно). Это участки с координатами 296—299 и 2687—2690. В обоих случаях последовательности, включающие этот тетра-нуклеотид, одинаковые: GATCGA. Можно видеть, что в этой структуре участок узнавания *TaqI* перекрывается с участком узнавания другой рестрикционной нуклеазы — *Sau3AI* (GATC), которая легко расщепляет специфичный для нее тетра-нуклеотид. Причины, по которым данная последовательность не расщепляется *TaqI*, неясны. Возможно, близлежащая последовательность GATC играет роль внутреннего ингибитора, с которым связывается активный центр *TaqI*. В этой связи интересно было бы проверить, образует ли рестриктаза *TaqI* комплексы с данной последовательностью. До настоящего времени мы не встречали в литературе подобных случаев.

Еще одной важной особенностью изученного фрагмента является то, что, согласно данным [12], в области 4—500 п.о. он содержит участок прочного связывания с РНК-полимеразой *E. coli*. Структура этого участка и обсуждение его возможной роли в экспрессии оперона *rpoBC* опубликованы нами ранее [25].

Таким образом, в результате проведенного исследования установлена полная нуклеотидная последовательность фрагмента *EcoRI*-С гена *rpoB* и аминокислотная последовательность средней части β -субъединицы РНК-полимеразы *E. coli*, содержащей 958 аминокислотных остатков. Совместное изучение первичной структуры белка и соответствующего ему структурного гена, во-первых, резко ускорило и значительно упростило решение обеих поставленных задач и, во-вторых, увеличило надежность структурного анализа.

Экспериментальная часть

Фрагмент *EcoRI*-С получали из ДНК плазмиды рОД 162 [26] путем гидролиза рестриктазой *EcoRI* и последующего разделения полученных фрагментов в градиенте сахарозы аналогично тому, как описано в работе [27].

Рестриктазы *EcoRI* [28], *SalI* [29] и *TaqI* [30] выделяли как описано в соответствующих работах. Рестриктазы *HinfI*, *Sau3AI* и *HpaII* были любезно предоставлены проф. Цахау и проф. Мюллер-Хилом (ФРГ).

Полинуклеотидкиназа фага Т4 выделялась согласно методике [31]. Ферменты не проявляли фосфатазной и нуклеазных активностей. [γ -³²P]rATP, удельной активности 2000 Ки/ммоль, — препарат фирмы Amersham (Англия).

Гидролиз *EcoRI*-С-фрагмента рестриктазами *HinfI*, *Sau3AI* и *SalI* выполняли в буфере (рН 7,5), содержащем 0,01 М трис-НСl, 0,01 М $MgCl_2$, 0,01 М меркаптоэтанол, при 37° С в течение 2 ч, используя 1 ед. акт. на 1 мкг фрагмента. Расщепление рестриктазой *TaqI* проводили аналогично, но при 70° С. Для расщепления рестриктазой *HpaII* был использован буфер (рН 7,4), содержащий 0,01 М трис-НСl, 0,01 М $MgCl_2$, 0,006 М КCl, 0,001 М дитиотреит, и те же условия гидролиза (37° С).

Определение нуклеотидной последовательности субфрагментов выполнялось по методу Максама — Гилберта [16]. Комплементарные нити получали по методу [32]. Для определения положения пуриновых оснований проводили апуринизацию ДНК 60% $HCOOH$ [17, 18]. При определении положения тимидиновых звеньев с помощью модификации с OsO_4 [20—22] к 50 мкл раствора ^{32}P -меченого фрагмента в 0,1 М NH_4OH добавляли 2 мкл 0,2 М раствора OsO_4 . После инкубации в течение 5 мин при 0° С добавляли 200 мкл охлажденного до 0° С 0,3 М ацетата натрия, рН 5,5, и 750 мкл спирта. Образцы после высаживания обрабатывали пиперидином, как описано в работе [16]. Для модификации использовали выделенные комплементарные цепи.

Гидролиз β -субъединицы протеазы из *St. aureus* и трипсином. β -Субъединицу (2 мкмоль) растворяли в 50 мл буфера, содержащего 6 М гуанидин-НСl, 1,0 М трис-НСl (рН 8,6) и 0,34 мл меркаптоэтанола, и карбоксиметилировали иодуксусной кислотой по методу [33]. Раствор белка обессоливали на колонке (2,5×100) с сефадексом G-50, уравновешенным буфером, содержащим 6 М мочевины, 0,01 М трис-НСl (рН 8,5). К раствору, термостатированному при 2—4° С, добавляли 1 мл свежеперегнанного цитраконового ангидрида, поддерживая рН раствора 8,5 с помощью 5 н. NaOH на автотитраторе ТТ1/ВР2 (Radiometer, Дания). Белок обессоливали на колонке с сефадексом G-50, уравновешенным водой, содержащей аммиак (рН 9—10), и лиофилизовали. Лиофилизированный препарат растворяли в 50 мл 0,1 М аммоний-бикарбонатного буфера (рН 8,0) и добавляли 5 мг протеиназы из *St. aureus* (Miles, Англия) или трипсина (Worthington, США). В случае протеиназы из *St. aureus* гидролиз вели 20 ч, а в случае трипсина — 4 ч при 37° С. После гидролиза смесь лиофилизовали.

Разделение, очистку и определение аминокислотной последовательности пептидов осуществляли методами, описанными ранее [24, 34—36].

Авторы выражают благодарность проф. Цахау и проф. Мюллер-Хилу (ФРГ) за рестрикционные эндонуклеазы, проф. Р. Б. Хесину и О. Н. Данилевской (Институт молекулярной генетики АН СССР) за предоставленную плазмиду рОД 162, Э. И. Царьковой и Г. М. Аринушкиной за помощь в работе, П. В. Костецкому за помощь в математической обработке данных.

ЛИТЕРАТУРА

1. Ovchinnikov Yu. A., Lipkin V. M., Modyanov N. N., Chertov O. Yu., Smirnov Yu. V. (1977) FEBS Letters, **76**, 108—111.
2. Burgess R. R. (1976) in: RNA Polymerase (Chamberlin M., Losick R., eds); pp. 68—100, Cold Spring Harbor Lab., Cold Spring Harbor.
3. Марчечко Т. В., Модянов Н. Н., Липкин В. М., Овчинников Ю. А. (1980). Биорган. химия, **6**, 325—331.
4. Shine J., Dalgarno L. (1974) Proc. Nat. Acad. Sci. USA, **71**, 1342—1346.
5. Steitz J. A., Jakes K. (1975) Proc. Nat. Acad. Sci. USA, **72**, 4734—4738.
6. Ovchinnikov Yu. A., Guryev S. O., Kravtsov A. S., Monastyrskaya G. S., Skryabin K. G., Sverdlov E. D., Zakharyev B. M., Bayev A. A. (1979) Gene, **6**, 235—249.
7. Schwarz E., Scherer G., Hobom G., Kössel H. (1978) Nature, **272**, 410—414.
8. Kirschbaum J. B., Konrad B. E. (1973) J. Bacteriology, **116**, 517—526.
9. Мивдлин С. З., Ильина Т. С., Горленко Х. М., Хачикян Н. А., Ковалев Ю. Н. (1976) Генетика, **12**, 116—130.
10. Lindahl L., Yamamoto M., Nomura M. (1977) J. Mol. Biol., **109**, 23—47.
11. Yamamoto M., Nomura M. (1978) Proc. Nat. Acad. Sci. USA, **75**, 3891—3895.

12. Collins J. (1979) *Mol. Gen. Genet.*, **173**, 217-220.
13. Post L. E., Strycharz G. D., Nomura M., Lewis H., Dennis P. P. (1979) *Proc. Nat. Acad. Sci. USA*, **76**, 1697-1701.
14. Гуревич А. И., Аваков А. Э., Колосов М. Н. (1979) *Биооргани. химия*, **5**, 1735-1739.
15. Sutchliffe J. G. (1978) *Nucl. Acids Res.*, **5**, 2721-2728.
16. Maxam A., Gilbert W. (1977) *Proc. Nat. Acad. Sci. USA*, **74**, 560-564.
17. Sverdlov E. D., Monastyrskaya G. S., Chestukhin A. V., Budowsky E. I. (1973) *FERS Letters*, **33**, 15-17.
18. Свeрдлов Е. Д., Левитан Т. Л. (1976) *Биооргани. химия*, **2**, 370-375.
19. Коробко В. Г., Грачев С. А. (1977) *Биооргани. химия*, **3**, 1419-1422.
20. Burton K., Riley W. T. (1966) *Biochem. J.*, **98**, 70-77.
21. Свeрдлов Е. Д., Монастырская Г. С., Будовский Э. И. (1977) *Мол. биол.*, **11**, 116-123.
22. Brown D. M. (1978) *Nucl. Acids Res.*, **5**, 615-621.
23. Ohmori H., Tomizawa J., Maxam A. M. (1978) *Nucl. Acids Res.*, **5**, 1479-1485.
24. Липкин В. М., Марченко Т. В., Хохлаков В. С., Половникова И. Н., Потапенко Н. А., Модянов Н. Н., Овчинников Ю. А. (1980) *Биооргани. химия*, **6**, 332-347.
25. Свeрдлов Е. Д., Липкин В. М., Монастырская Г. С., Губанов В. В., Гурьев С. О., Чертов О. Ю. (1980) *Биооргани. химия*, **6**, 309-312.
26. Bass I. A., Danilevskaya O. N., Mekhedov S. L., Fedoseeva V. V., Gorlenko Zh. M. (1979) *Mol. Gen. Genet.*, **173**, 101-107.
27. Свeрдлов Е. Д., Монастырская Г. С., Росташов В. М. (1978) *Биооргани. химия*, **4**, 894-900.
28. Sümegi J., Breedveld D., Hossenlopp P., Chambon P. (1977) *Biochem. and Biophys. Res. Commun.*, **76**, 78-85.
29. Greene P. J., Heyneker H. L., Bolivar F., Rodriguez R. L., Betlach M. C., Covarrubias A. A., Backman K., Russel D., Tait R., Boyer H. W. (1978) *Nucl. Acids Res.*, **5**, 2373-2380.
30. Bickle T. A., Pirrotta V., Imber R. (1977) *Nucl. Acids Res.*, **4**, 2561-2572.
31. Richardson C. C. (1971) in: *Procedures in Nucleic Acids Research* (Cantony G. L., Davies D. R., eds), vol. 2, pp. 815-828, Harper and Row, N. Y.—London.
32. Szalay A. A., Grohmann K., Siusheimer R. L. (1977) *Nucl. Acids Res.*, **4**, 1569-1578.
33. Hirs C., Moore S., Stein W. H. (1956) *J. Biol. Chem.*, **219**, 623-642.
34. Овчинников Ю. А., Киришкин А. А., Егоров Ц. А., Абдулаев Н. Г., Киселев А. П., Модянов Н. П. (1972) *Биохимия*, **37**, 451-460.
35. Виноградова Е. И., Фейгина М. Ю., Алданова Н. А., Липкин В. М., Смирнов Ю. В., Потапенко Н. А., Абдулаев Н. Г., Киселев А. П., Егоров Ц. А., Овчинников Ю. А. (1973) *Биохимия*, **38**, 3-21.
36. Модянов Н. Н., Липкин В. М., Смирнов Ю. В., Чертов О. Ю., Потапенко Н. А., Шуваева Т. М. (1978) *Биооргани. химия*, **4**, 158-179.

Поступила в редакцию
1.II.1980.

**PRIMARY STRUCTURE OF RNA POLYMERASE FROM *E. COLI*.
NUCLEOTIDE SEQUENCE OF *EcoRI*-C FRAGMENT OF GENE *rpoB* AND AMINO
ACID SEQUENCE OF THE CORRESPONDING FRAGMENT OF β -SUBUNIT**

OVCHINNIKOV Yu. A., SVERDLOV E. D., LIPKIN V. M.,
MONASTYRSKAYA G. S., CHERTOV O. Yu., GUBANOV V. V.,
GURIEV S. O., MODYANOV N. N., GRINKEVICH V. A.,
MAKAROVA I. A., MARCHENKO T. V., POLOVNIKOVA I. N.

M. M. Shemyakin Institute of Bioorganic Chemistry

Academy of Sciences of the USSR, Moscow

In the course of the parallel study of primary structures of β -subunit of DNA — dependent RNA polymerase from *E. coli* and its structural gene *rpoB* the complete nucleotide sequence (2874 base pairs) of *EcoRI*-C fragment of gene *rpoB* was established. The corresponding amino acid sequence of the middle part of β -subunit contains 958 amino acid residues. A conjunction of the methods of protein and nucleotide chemistry for the combined structural investigation of a protein and DNA sharply accelerates and considerably simplifies the solution of both problems and enhances the reliability of the structural analysis.