



УДК 577.214.622 : 577.113.5 : 577.152.344'152

© 1990 г.

А. Ю. Америк, В. В. Антонов, Н. И. Остроумова,  
Т. В. Ротанова, Л. Г. Чистякова

КЛОНИРОВАНИЕ, СТРУКТУРА И ЭКСПРЕССИЯ  
ПОЛНОРАЗМЕРНОГО ГЕНА *lon* *ESCHERICHIA COLI*.  
КОДИРУЮЩЕГО АТР-ЗАВИСИМУЮ La-ПРОТЕИНАЗУ

Институт биоорганической химии им. М. М. Шемякина АН СССР, Москва

Осуществлена сборка полноразмерного гена *lon* *Escherichia coli* K-12 из *EcoRI/SphI*-фрагмента («модифицированный» ген) геномной ДНК, клонированного и секвенированного нами ранее, и *PstI/PstI*-фрагмента той же ДНК, содержащего 3'-концевую область гена *lon*. Как «модифицированный», так и полноразмерный гены проявляют все фенотипические свойства гена *lon*. Определена полная нуклеотидная последовательность гена (2770 п. о.) и выведена последовательность аминокислот кодируемой этим геном АТР-зависимой La-протеиназы. В структуре протеиназы обнаружен АТР-связывающий участок. Высказаны предположения о локализации каталитически активных остатков серина, гистидина и аспарагиновой кислоты. Сравнение структуры гена *lon* и La-протеиназы со структурами, опубликованными в работе D. T. Chin и соотр., выявило различия в двух областях гена. Анализируются причины обнаруженных различий.

Доминирующую роль в деградации аномальных и чужеродных белков в клетках *Escherichia coli* играет АТР-зависимая протеиназа La, кодируемая геном *lon* (*capR*, *deg*) [1—3]. Протеиназа La относится к группе белков теплового шока [4] и участвует в таких процессах, как клеточное деление и образование капсулярных полисахаридов [5]. Клетки, мутантные по гену *lon*, обладают повышенной чувствительностью к УФ-облучению и к имитаторам этого процесса (нитрофурантоин и др.) [6].

Ген *lon* был клонирован в F-эписоме [7] и фаге  $\lambda$  [8]. Была получена рестриктная карта гена [7, 8] и выведена нуклеотидная последовательность фрагмента, включающего в себя регуляторный участок и 5'-концевую часть (174 п. о.) структурного гена La-протеиназы [9]. Сам фермент, полученный экспрессией клонированного гена в клетках *E. coli*, оказался тетрамерным белком с молекулярной массой субъединицы около 90 кДа [10]. Имеющиеся данные [11] позволяют отнести его к группе сериновых протеиназ.

Ранее [12] мы опубликовали в виде краткого сообщения результаты опытов по клонированию гена *lon* по рестриктным сайтам *EcoRI* и *SphI* в плазмидном векторе pBR327 и по определению его нуклеотидной последовательности. Полученный в этих экспериментах ген оказался неполным («модифицированный» ген, *lon<sub>m</sub>*). Структура полноразмерного гена была доложена нами [13] и одновременно А. Гольдбергом [14] на XIV биохимическом конгрессе. В дальнейшем появилось подробное сообщение группы американских авторов на эту тему [15].

В настоящей работе мы приводим полное описание данных по клонированию «модифицированного» гена *lon* и по конструированию полноразмерного гена, его структуре и экспрессии. Анализируются структура La-протеиназы, а также отличия полученной нами последовательности гена *lon* от данных, приведенных в работе [15].

Геномную ДНК *E. coli* K-12 сначала клонировали по рестриктным сайтам *EcoRI* и *SphI* в плазмидном векторе pBR327 [12]. Полученные рекомбинантные плазмиды амплифицировали в штамме *E. coli* HB101 и выделен-

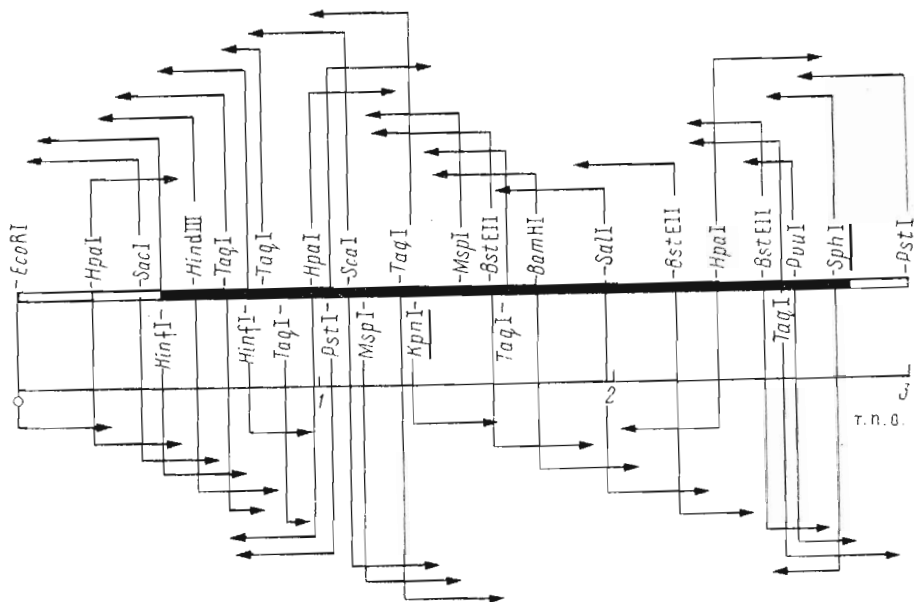


Рис. 1. Рестриктивная карта и стратегия определения нуклеотидной последовательности гена *lon* (кодирующая область зачернена). Стрелками указано направление секвенирования и длина субфрагментов кодирующей (под картой) и комплементарной (над картой) цепей ДНК. Выделены сайты, отсутствующие на рестриктивной карте гена *lon*, полученной авторами работы [15]

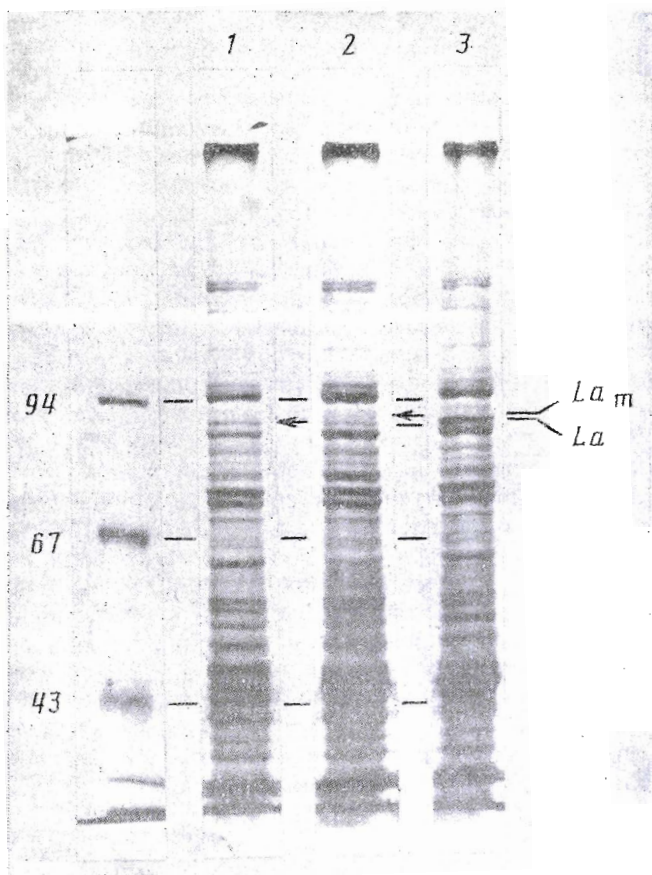
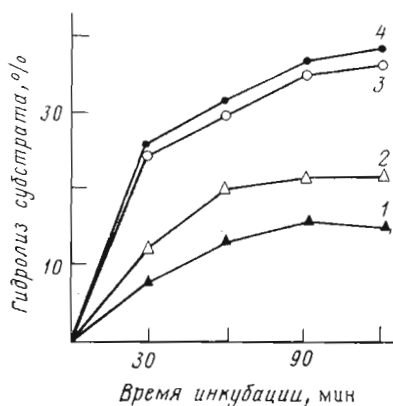


Рис. 2. Гель-электрофорез в присутствии SDS клеточных экстрактов *E. coli*: штаммы AB1899/pBR327 (1), AB1899/pBR $lon_m$  (2) и AB1899/pBR $lon$  (3). Положения *La*-протеиназы и продукта модифицированного гена ( $La_m$ ) отмечено стрелками. Цифры соответствуют значениям молекулярных масс белков-маркеров (в кДа). Препараты для электрофореза готовили обработкой суспензии клеток ( $A_{600}$  0,8) фенолом, насыщенным 50 мМ трис-НСI-буфером, рН 8,0, с последующим центрифугированием (5 мин, 10 000g) и осаждением белков из фенольной фракции добавлением 5 объемов этанола

Рис. 3. Внутриклеточная деградация пуромидинсодержащих пептидов в штаммах *E. coli* АВ1899/pBR327 (1), АВ1157/pBR327 (2), АВ1899/pBRlon (3) и АВ1899/pBRlon<sub>m</sub> (4)



ными плазмидами трансформировали клетки *E. coli* АВ1899, несущие мутацию по гену *lon*. Первичный отбор клонов проводили по увеличению устойчивости клеток к нитрофурантоину и по виду колоний (отсутствие осозленности, вызванной гиперпродукцией капсулярных полисахаридов), выращенных на минеральной среде М9. Среди отобранных клонов было обнаружено 7 клонов, плазмидная ДНК которых гибридизовалась с нуклеотидным зондом 5'CTGAACGCATTGAAATCCCCGATTGCC, соответствующим 5'-концевой части гена [9]. Рестриктивная карта полученных плазмид (рис. 1) почти полностью совпадала с известной [7, 8] картой гена *lon*, однако в нашем случае был обнаружен новый *SphI*-сайт (выделен на рис. 1).

Введение рекомбинантных плазмид (pBR327 *lon<sub>m</sub>*) вызывало исчезновение всех характерных признаков *lon*-мутации у штаммов *E. coli* АВ1899 и VL901. Электрофорез белков клеток *E. coli* VL901, содержащих плазмиду с клонированным геном, обнаруживал появление полосы с  $M \sim 90$  кДа, отсутствующей у контрольных клеток, содержащих плазмиду pBR327 (рис. 2). Была измерена скорость внутриклеточного протеолиза пуромидиновых пептидов [16] в клетках *E. coli* АВ1899 и в изогенном штамме АВ1157. Показано, что при введении в клетки АВ1899 плазмиды, содержащей клонированный ген, скорость протеолиза значительно возрастала и существенно превышала скорость протеолиза в клетках штамма АВ1157, не несущих мутации в области *lon*-гена (рис. 3). Полученный эффект можно объяснить увеличением числа копий гена в клетках с плазмидой. Тем не менее выделить соответствующий белок не удалось из-за его крайней неустойчивости.

Стратегия определения нуклеотидной последовательности клонированного гена методом Максама — Гилберта показана на рис. 1. Структура гена и соответствующая ей аминокислотная последовательность Ла-протеиназы представлены на рис. 4.

Выяснилось, что клонированный нами ген не имеет терминирующего кодона. Терминация происходит уже после *SphI*-сайта, по которому проводилось расщепление геномной ДНК, на участке вектора. Таким образом, клонированный нами ген *lon<sub>m</sub>* имел искаженную структуру в 3'-концевой области, а образующийся при его экспрессии белок обладал измененной по сравнению с протеиназой Ла С-концевой последовательностью аминокислот.

Для установления истинной 3'-концевой последовательности гена *lon* геномную ДНК подвергали расщеплению рестриктазой *PstI* и фрагменты размером 1,5—3,0 т. п. о. клонировали в векторе pSP64. Полученной плазмидой трансформировали клетки *E. coli* МН1 и отбирали клоны, дающие положительный сигнал с зондом для *HpaI/SphI*-фрагмента «модифицированного» гена (около 400 п. о.). Структурный анализ полученного фрагмента показал, что его 3'-концевая последовательность совпадает с последовательностью 3'-участка клонированного *EcoRI/SphI*-фрагмента (см. выше) и содержит в рамке считывания после *SphI*-сайта дополнительный участок из 18 нуклеотидов, после которого следуют два тер-

419 ATG AAT CGT GAG CGT TCT GAA CGC ATT GAA ATC CCG GTA TTG CCC  
Met Asn Pro Glu Arg Ser Glu Arg Ile Glu Ile Pro Val Leu Pro 15

464 CTG CGC GAT GTG GTG GTT TAT CCG CAC ATG GTC ATC CCC TTA TTT  
Leu Arg Asp Val Val Val Tyr Pro His Met Val Ile Pro Leu Phe 30

509 GTC GGG CGG GAA AAA TCT ATC CGT TGT CTG GAA GCG GCG ATG GAC  
Val Gly Arg Glu Lys Ser Ile Arg Cys Leu Glu Ala Ala Met Asp 45

554 CAT GAT AAA AAA ATT ATG CTG GTC GCG CAG AAA GAA GCT TCA ACG  
His Asp Lys Lys Ile Met Leu Val Ala Gln Lys Glu Ala Ser Thr 60

599 GAT GAG CCG GGT GTA AAC GAT CTT TTC ACC GTC GGG ACC GTG GCC  
Asp Glu Pro Gly Val Asn Asp Leu Phe Thr Val Gly Thr Val Ala 75

644 TCT ATA TTG CAG ATG CTG AAA CTG CCT GAC GGC ACC GTC AAA GTG  
Ser Ile Leu Gln Met Leu Lys Leu Pro Asp Gly Thr Val Lys Val 90

689 CTG GTC GAG GGG TTA CAG CGC GCG CGT ATT TCT GCG CTC TCT GAC  
Leu Val Glu Gly Leu Gln Arg Ala Arg Ile Ser Ala Leu Ser Asp 105

734 AAT GGC GAA CAC TTT TCT GCG AAG GCG GAG TAT CTG GAG TCG CCG  
Asn Gly Glu His Phe Ser Ala Lys Ala Glu Tyr Leu Glu Ser Pro 120

779 ACC ATT GAT GAG CCG GAA CAG GAA GTG CTG CTG CGT ACT GCA ATC  
Thr Ile Asp Glu Arg Glu Gln Glu Val Leu Val Arg Thr Ala Ile 135

824 AGC CAG TTC GAA GGC TAC ATC AAG CTG AAC AAA AAA ATC CCA CCA  
Ser Gln Phe Glu Gly Tyr Ile Lys Leu Asn Lys Lys Ile Pro Pro 150

869 GAA GTG CTG ACG TCG CTG AAT AGC ATC GAC GAT CCG GCG CGT CTG  
Glu Val Leu Thr Ser Leu Asn Ser Ile Asp Asp Pro Ala Arg Leu 165

914 GCG GAT ACC ATT GCT GCA CAT ATG CCG CTG AAA CTG GCT GAC AAA  
Ala Asp Thr Ile Ala Ala His Met Pro Leu Lys Leu Ala Asp Lys 180

959 CAG TCC GTT CTG GAG ATG TCC GAG GTT AAC GAA CCG CTG GAA TAT  
Gln Ser Val Leu Glu Met Ser Asp Val Asn Glu Arg Leu Glu Tyr 195

1004 CTG ATG GCA ATG ATG GAA TCG GAA ATC GAT CTG CTG CAG GTT GAG  
Leu Met Ala Met Met Glu Ser Glu Ile Asp Leu Leu Gln Val Glu 210

1049 AAA CGC ATT CGC AAC CGC GTT AAA AAG CAG ATG GAG AAA TCC CAG  
Lys Arg Ile Arg Asn Arg Val Lys Lys Gln Met Glu Lys Ser Gln 225

1094 CGT GAG TAC TAT CTG AAC GAG CAA ATG AAA GCT ATT CAG AAA GAA  
Arg Glu Tyr Tyr Leu Asn Glu Gln Met Lys Ala Ile Gln Lys Glu 240

1139 CTC GGT GAA ATG GAC GAC GCG CCG GAC GAA AAC GAA GCC CTG AAG  
Leu Gly Glu Met Asp Asp Ala Pro Asp Glu Asn Glu Ala Leu Lys 255

1184 CGC AAA ATC GAC GCG GCG AAG ATG CCG AAA GAG GCA AAA GAG AAA  
Arg Lys Ile Asp Ala Ala Lys Met Pro Lys Glu Ala Lys Glu Lys 270  
\*CGA AAG AGG CAA AAG AGA AAG

1229 CCG GAA GCA GAG TTG CAG AAG CTG AAA ATG ATG TCT CCG ATG TCG  
Ala Glu Ala Glu Leu Gln Lys Leu Lys Met Met Ser Pro Met Ser 285  
CGG ACA GGA GTT GCA GAA GCT GAA AAT GAT GTC TCC GAT GTC GGC

1274 GCA GAA CCG ACC GTA GTG CGT GCT TAT ATC GAC TGG ATG GTA CAG  
Ala Glu Ala Thr Val Val Arg Gly Tyr Ile Asp Trp Met Val Gln 300  
AGA AGC GAC CGT ACT CCG TGG TTA TAT CCA CTG GAT GGT ACA GGT

1319 GTA CCG TGC AAT CCG CCG AGC AAG GTC AAA AAA GAC CTG CGT CAG  
Val Pro Trp Asn Ala Arg Ser Lys Val Lys Lys Asp Leu Arg Gln 315  
GCC GTG GAA TGC CCG TAC GAA GGT CAA AAA AGA CCT GCG TCA GGC

1364 GCG CAG GAA ATC CTT GAT ACC GAC CAT TAT GGT CTG GAG CCG GTG  
Ala Gln Glu Ile Leu Asp Thr Asp His Tyr Gly Leu Glu Arg Val 330  
GCA ----\*

1409 AAA GAT CGC ATC CTT GAG TAC CTT GCG GTT CAA AGC CGT GTC AAC  
Lys Asp Arg Ile Leu Glu Tyr Leu Ala Val Gln Ser Arg Val Asn 345

1454 AAA ATC AAG GGA CCA ATC CTC TGC CTG GTA GGG CCG CCG GGG GTA  
Lys Ile Lys Gly Pro Ile Leu Cys Leu Val Gly Pro Pro Gly Val 360

1499 GGT AAA ACC TCT CTT GGT CAG TCC ATT GCC AAA GCC ACC GGG CGT  
Gly Lys Thr Ser Leu Gly Gln Ser Ile Ala Lys Ala Thr Gly Arg 375

1544 AAA TAT GTC CGT ATG GCG CTG GGC GGC GTG CGT GAT GAA GCG GAA  
Lys Tyr Val Arg Met Ala Leu Gly Gly Val Arg Asp Glu Ala Glu 390

1589 ATC CGT GGT CAC CGC CGT ACT TAC ATC GGT TCT ATG CCG GGT AAA  
Ile Arg Gly His Arg Arg Thr Tyr Ile Gly Ser Met Pro Gly Lys 405

1634 CTG ATC CAG AAA ATG GCG AAA GTG GGC GTG AAA AAC CCG CTG TTC  
Leu Ile Gln Lys Met Ala Lys Val Gly Val Lys Asn Pro Leu Phe 420

1679 CTG CTC GAT GAG ATC GAC AAA ATG TCT TCT GAC ATG CGA GGC GAT  
Leu Leu Asp Glu Ile Asp Lys Met Ser Ser Asp Met Arg Gly Asp 435

1724 CCG GCC TCT GCA CTG CTT GAA GTG CTG GAT CCA CAG CAG AAC GTA  
Pro Ala Ser Ala Leu Leu Glu Val Leu Asp Pro Glu Gln Asn Val 450

1769 GCG TTC AGC GAC CAC TAC CTG GAA GTG GAT TAC GAT CTC AGC GAC  
Ala Phe Ser Asp His Tyr Leu Glu Val Asp Tyr Asp Leu Ser Asp 465

1814 GTG ATG TTT GTC GCG ACG TCG AAC TCC ATG AAC ATT CCG GCA CCG  
Val Met Phe Val Ala Thr Ser Asn Ser Met Asn Ile Pro Ala Pro 480

1859 CTG CTC GAT CGT ATG GAA GTG ATT CGC CTC TCG GCT TAT ACC GAA  
Leu Leu Asp Arg Met Glu Val Ile Arg Leu Ser Gly Tyr Thr Glu 495

1904 GAT GAA AAA CTG AAC ATC GCC AAA CGT CAC CTG CTG CCG AAG CAG  
Asp Glu Lys Leu Asn Ile Ala Lys Arg His Leu Leu Pro Lys Gln 510

1949 ATT GAA CGT AAT GCA CTG AAA AAA GGT GAG CTG ACC GTC GAC GAT  
Ile Glu Arg Asn Ala Leu Lys Lys Gly Glu Leu Thr Val Asp Asp 525

1994 AGC CCC ATT ATC GGC ATT ATT CGT TAC TAC ACC CGT GAC CCG GGC  
Ser Ala Ile Ile Gly Ile Ile Arg Tyr Tyr Thr Arg Glu Ala Gly 540  
\*CGG GCG

2039 GTG CGT GGT CTG GAG CGT GAA ATC TCC AAA CTG TGC CCG AAA GCG  
Val Arg Gly Leu Glu Arg Glu Ile Ser Lys Leu Cys Arg Lys Ala 555  
TGC CTC GTC TGC AGC CTG AAA TCT CCA AAC TGT GTC GCA AAG CCG

2084 GTT AAG CAG TTA CTG CTC GAT AAG TCA TTA AAA CAT ATC GAA ATT  
Val Lys Gln Leu Leu Leu Asp Lys Ser Leu Lys His Ile Glu Ile 570  
TTA AGC AGT TAC TCC TCG ATA ACG\*

2129 AAC GCG GAT AAC CTG CAT GAC TAC CTC GGT GTT CAG CGT TTC GAC  
Asn Gly Asp Asn Leu His Asp Tyr Leu Gly Val Gln Arg Phe Asp 585

2174 TAT GGT CGC GCT GAT AAC GAA AAC CGT GTC GGT CAG GTA ACC GGT  
Tyr Gly Arg Ala Asp Asn Glu Asn Arg Val Gly Gln Val Thr Gly 600

2219 CTG GCC TGC ACG GAA GTG GGC GGT GAC TTG CTC ACC ATT GAA ACC  
Leu Ala Trp Thr Glu Val Gly Gly Asp Leu Leu Thr Ile Glu Thr 615

2264 GCA TGT GTT CCC GGT AAA GGC AAA CTC ACC TAT ACC GGT TCG CTC  
Ala Cys Val Pro Gly Lys Gly Lys Leu Thr Tyr Thr Gly Ser Leu 630

2309 GGC GAA CTG ATG CAG GAG TCT ATT CAG GCG GCG TTA ACG GTG GTT  
Gly Glu Val Met Gln Glu Ser Ile Gln Ala Ala Leu Thr Val Val 645

2354 CGT GCG CGT GCC GAA AAA CTG GGG ATC AAC CCT GAT TTT TAC GAA  
Arg Ala Arg Ala Glu Lys Leu Gly Ile Asn Pro Asp Phe Tyr Glu 660

2399 AAA CGT GAC ATC CAC GTC CAC GTA CCG GAA GGT GCG ACG CCG AAA  
Lys Arg Asp Ile His Val His Val Pro Glu Gly Ala Thr Pro Lys 675

2444 GAT GGT CCG AGT GCC GGT ATT GCT ATC TCC ACC GCG CTG GTT TCT  
 Asp Gly Pro Ser Ala Gly Ile Ala Met Cys Thr Ala Leu Val Ser 690

2489 TGC CTG ACC GGT AAC CCG GTT CGT GCC GAT GTG GCA ATG ACC GGT  
 Cys Leu Thr Gly Asn Pro Val Arg Ala Asp Val Ala Met Thr Gly 705

2534 GAG ATC ACT CTG CGT GGT CAG GTA CTC CCG ATC GCT GCT TTG AAA  
 Glu Ile Thr Leu Arg Gly Gln Val Leu Pro Ile Gly Gly Leu Lys 720

2579 GAA AAA CTC CTG GCA GCG CAT CGC GGC GCG ATT AAA ACA GTG CTA  
 Glu Lys Leu Leu Ala Ala His Arg Gly Gly Ile Lys Thr Val Leu 735

2624 ATT CCG TTC GAA AAT AAA CGC CAT CTC GAA GAG ATT CCT GAC AAC  
 Ile Pro Phe Glu Asn Lys Arg Asp Leu Glu Glu Ile Pro Asp Asn 750

2669 GTA ATT GCC GAT CTG GAC ATT CAT CCT GTG AAG CGC ATT GAG GAA  
 Val Ile Ala Asp Leu Asp Ile His Pro Val Lys Arg Ile Glu Glu 765

2714 GTT CTG ACT CTG GCC CTG CAA AAT GAA CCG TCT GCC ATG CAG GTT  
 Val leu Thr Leu Ala Leu Gln Asn Glu Pro Ser Gly Met Gln Val 780  
 ..CAC CAT  
 His His

2759 GTG ACT GCA AAA TAG TGA  
 Val Thr Ala Lys  
 TCC TTG CCG CCG CCG TCC TCA ACC GCC TCA ACC TAC TAC TCG GCT  
 Ser Leu Arg Arg Arg Cys Ser Thr Ala Ser Thr Tyr Tyr Trp Ala 795

2804 GCT TCC TAA  
 Ala Ser \*\*

Рис. 4. Нуклеотидная последовательность структурной части гена *lon* и соответствующая ей первичная структура протеиназы La. Звездочкой указаны фрагменты последовательности из работы [15], двумя звездочками — последовательность модифицированного гена *lon* (*lon<sub>m</sub>*) [12]

минирующих кодона (рис. 4). Таким образом была определена полная структура гена *lon*.

Для конструирования плазмиды, содержащей полноразмерный ген *lon*, вектор рSP64, несущий *Pst*I-фрагмент гена *lon*, расщепляли рестриктазой *Sal*I (рис. 5) и полученный *Sal*I-фрагмент клонировали в обработанной рестриктазой *Sal*I плазмиде рBR327, имеющей *Eco*RI/*Sph*I-участок гена *lon*. Клоны, содержащие плазмиды с нужной ориентацией *Sal*I-фрагмента, отбирали для трансформации клеток *E. coli* AB1899. Такие клетки, как и клетки, трансформированные «модифицированным» геном, проявляют все фенотипические свойства *lon*<sup>+</sup>-клеток: они устойчивы к нитрофурантоину, не ослизняются при выращивании на минеральной среде и способны быстро гидролизовать белки с неполной структурой, образующиеся при действии на клетки пуромицина (рис. 3). Белок, образующийся при экспрессии полноразмерного гена *lon*, электрофоретически отличается от белка, полученного экспрессией «модифицированного» гена (рис. 2). Клетки, содержащие плазмиду с полноразмерным геном *lon*, растут заметно медленнее, чем клетки дикого типа или содержащие «модифицированный» ген, вследствие токсичности La-протеиназы для *E. coli* (ср. [17]). При длительном хранении они трансформируются в ослизненные колонии, по-видимому, за счет спонтанных мутаций в гене *lon*.

Экспрессия La-протеиназы клетками, содержащими полноразмерный ген *lon*, была доказана определением АТФ-зависимой протеолитической активности в лизатах с использованием в качестве субстрата радиоактивно меченного [18] ацетил- $\alpha$ -казеина. Следует отметить, что обнаружить такую активность удается лишь после сорбции белков лизата на фосфоцеллюлозе и последующего элюирования фермента. Полученные данные приведены в табл. 1.

Анализ выведенной из структуры гена *lon* последовательности аминокислот протеиназы La показывает, что ее субъединица содержит 784 аминокислотных остатка и имеет молекулярную массу 87 304 Да. Белок характеризуется почти одинаковым содержанием полярных и неполярных



Рис. 5. Схема конструирования плазмиды, несущей полноразмерный ген *lon*

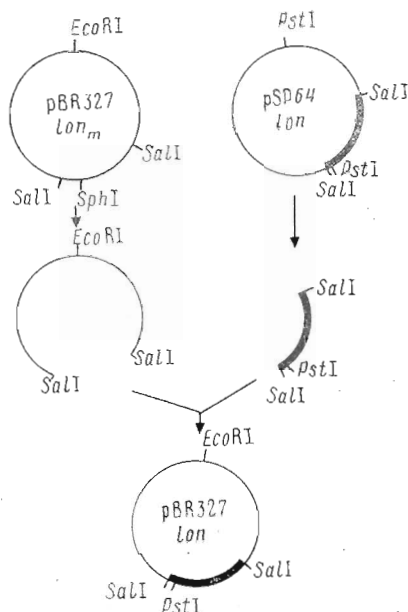


Рис. 5

Рис. 6. Предполагаемая вторичная структура La-протеиназы. 1 —  $\alpha$ -спиральные участки, 2 —  $\beta$ -складки, 3 —  $\beta$ -шипилки. Цифровая шкала соответствует нумерации остатков аминокислот (по рис. 4)

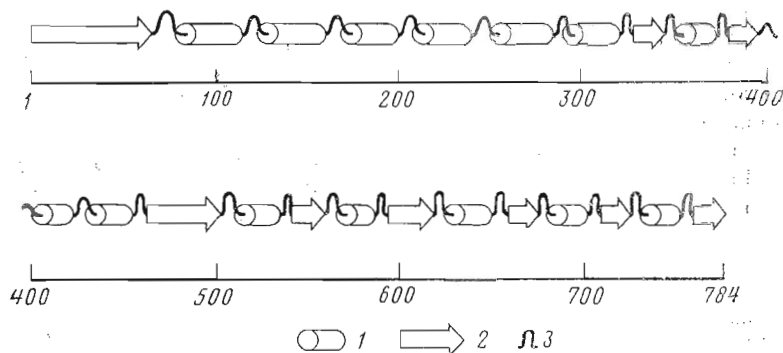


Рис. 6

остатков (табл. 2). Его вычисленная изоэлектрическая точка 6,0. Расчет вторичной структуры (по методу [19]) показывает, что белок относится, по-видимому, к типу  $\alpha/\beta$  с чередующимися  $\alpha$ -спиральными и  $\beta$ -складчатыми структурами (рис. 6). Из шести остатков цистеина четыре локализованы в С-концевой части белка.

Сравнение последовательности аминокислот La-протеиназы со структурой других АТФ-связывающих белков (рис. 7) выявило наличие АТФ-связывающего центра в области остатков 351—365. Этот фрагмент расположен на стыке коротких  $\alpha$ -спирального и  $\beta$ -складчатого участков, соединенных  $\beta$ -шипилкой (рис. 6).

Имеются веские основания [11] отнести La-протеиназу к группе сериновых протеиназ. Вопрос о локализации каталитически активных остатков серина, гистидина и аспарагиновой кислоты остается открытым. Однако сопоставление последовательностей аминокислот ряда сериновых протеиназ и La-протеиназы позволяет сделать некоторые предположения. Почти для всех известных сериновых протеиназ характерно наличие остатка глицина, предшествующего серину и отделенного от него одной аминокислотой (Gly-X-Ser). В структуре La-протеиназы обнаруживаются два таких остатка серина: Ser-368 в последовательности LGQZIA и Ser-679 в последовательности DGPSAG. Каталитическая функция Ser-368 была недавно постулирована в работе [24]. Однако, поскольку известно [25], что АТФ-азная активность фермента увеличивается в присутствии белковых субстратов, этот остаток, расположенный вблизи АТФ-связывающего центра (Phe<sup>251</sup>-Leu<sup>365</sup>), вряд ли может быть каталитически активным.

Влияние АТФ на протеолитическую активность в клетках *E. coli*, содержащих полноразмерный или модифицированный ген *lon*\*

Штамм	А	Б	Коэффициент ускорения
AV1899/pBR327 (контроль)	3,88	1,31	0,34
AV1157/pBR327 (контроль)	1,89	4,30	2,28
AV1899/pBR <i>lon</i>	6,33	39,80	6,29
AV1899/pBR <i>lon<sub>m</sub></i>	5,66	23,70	4,19

\* Приведена степень гидролиза [<sup>14</sup>C]ацетил-α-казеина (%) за 1 ч без АТФ (А) и в присутствии АТФ (Б) в расчете на 1 мг препарата фермента. Концентрация препарата фермента в пробе 0,26—0,42 мг/мл. Степень превращения субстрата в каждом опыте не превышала 15%.

Кроме того, известно, что расстояния между каталитически активными остатками у многих сериновых протеиназ варьируют в относительно узких пределах — от 80 до 90 остатков между Asp и Ser и от 30 до 40 между His и Asp. Ни одного остатка гистидина на указанном удалении от Ser-368 в структуре La-протеиназы не обнаруживается. В то же время, если за каталитически активный остаток принять Ser-679, то, с нашей точки зрения, наиболее вероятными кандидатами на роль каталитически активных остатков His и Asp у La-протеиназы будут остатки His-576 и Asp-609. Конечно, эти соображения не более чем гипотеза, нуждающаяся в проверке, которую мы сейчас и осуществляем методом направленного мутагенеза. Однако если принять это предположение, то каталитически активные остатки оказываются сосредоточенными в С-концевой трети молекулы. Как уже указывалось, АТФ-связывающий центр расположен в области остатков 350—360, т. е. во второй трети. Возможно, что La-протеиназа состоит из трех доменов, причем два последних выполняют соответственно роль АТФ-азы и протеиназы.

Таблица 2

Аминокислотный состав La-протеиназы, рассчитанный по предполагаемой (рис. 4) структуре фермента

Аминокислота	Число остатков в молекуле
Ala	62
Arg	49
Asn	26
Asp	51
Cys	6
Gln	28
Glu	68
Gly	51
His	14
Ile	51
Leu	84
Lys	61
Met	30
Phe	10
Pro	37
Ser	38
Thr	32
Trp	3
Tyr	20
Val	63
Общее число остатков	784
Молекулярная масса	87 304

Наконец, можно предположить, что причиной неустойчивости «модифицированной» La-протеиназы является наличие в ее структуре измененной С-концевой аминокислотной последовательности.

Как уже указывалось выше, в работе [12] мы привели последовательность 778 аминокислот (из 784) La-протеиназы. В настоящей работе представлена полная структура фермента. Сопоставление полученных нами данных с результатами Гольдберга и сотр. [15] выявило различия в отношении строения двух участков в молекуле La-протеиназы: в области остатков аминокислот 264—317 и 537—563 (нуклеотиды 1208—1369 и 2033—2107 по нашей нумерации и 1202—1360 и 2024—2098 по нумерации [15]). Вся остальная последовательность аминокислот полностью совпадает.

Причины указанных различий в настоящее время окончательно не выяснены. Они могут обуславливаться, например, различиями использованных штаммов *E. coli*. Спорные участки нуклеотидной последовательности были



1	Ile Leu Cys Leu Val <u>Gly Pro Pro Gly Val</u> <u>Gly Lys Thr</u> Ser Leu	
	145	159
2	Lys Val Gly Leu Phe <u>Gly Gly Ala Gly Val</u> <u>Gly Lys Thr</u> Val Asn	
	151	165
3	Lys Ile Gly Leu Phe <u>Gly Gly Ala Gly Val</u> <u>Gly Lys Thr</u> Val Phe	
	64	78
4	Ile Val Glu Ile Tyr <u>Gly Pro Glu Ser Ser</u> <u>Gly Lys Thr</u> Thr Leu	
	172	186
5	Ser Ile Leu Ile Thr <u>Gly Glu Ser Gly Ala</u> <u>Gly Lys Thr</u> Val Asn	
	123	137
6	Cys Leu Leu Val His <u>Gly Ser Pro Gly Thr</u> <u>Gly Lys Ser</u> Val Ala	

Рис. 7. Сопоставление последовательностей АТФ-связывающих доменов  $\lambda$ -протеиназы (1) и  $\beta$ -цепей АТФ-азы *E. coli* [20] (2) и АТФ-азы митохондрий быка [21] (3), RecA-белка [22] (4), миозина кролика [21] (5) и 2С-белка полиовируса [23] (6)

нами пройдены по ДНК трех независимых клонов, при этом полученные результаты получили подтверждение. Важно отметить, что на одном из участков (нуклеотиды 1318—1323) замена обнаруженной нами последовательности на последовательность, опубликованную в работе [15], привела бы к отсутствию *KpnI*-сайта (рис. 1), который мы успешно использовали при секвенировании. Вместе с тем некоторые различия в использованных штаммах могут иметь место. Так, уже отмечалось, что мы обнаружили дополнительный *SphI*-сайт (рис. 1), отсутствующий в ДНК штамма, использованного в работе [8]. В настоящее время структура спорных участков проверяется методами белковой химии.

### Экспериментальная часть

В работе использовались следующие штаммы *E. coli*: K-12, АВ1157, АВ1899, VL901, НВ101 и МН1 (табл. 3). Использовались рестрикционные эндонуклеазы и ферменты модификации ДНК следующих фирм: Pharmacia (Швеция), Amersham (США) и Boehringer mannheim (ФРГ). Радиоактивные [ $\alpha$ - $^{32}$ P]dNTP — фирмы Amersham (3000 Ки/ммоль). [ $^{14}$ C]Ацетил- $\alpha$ -казеин (30—50 мКи/ммоль) получали по модифицированной методике [17]. Олигонуклеотидный зонд был синтезирован С. А. Филипповым (ИБХ им. М. М. Шемякина АН СССР).

**Выделение ДНК *E. coli* K-12.** Клетки *E. coli* K-12 выращивали в течение ночи в 500 мл среды LB при 30° С 50 мМ центрифугировали (120 мин, 3000g) и суспендировали в 50 мл 50 мМ трис-НСl-буфера (рН 8,0) с, EDTA. К суспензии, замороженной охлаждением до -70° С, добавляли 5 мл раствора лизоцима (10 мг/мл) и инкубировали в водяной бане при 20° С до полного оттаивания клеток, после чего продолжали инкубацию при 0° С в течение 1 ч. К полученной суспензии добавляли 10 мл 50 мМ трис-НСl-буфера (рН 7,5), содержащего 0,5% SDS и 0,4 М EDTA, смесь выдержива-

Генотипы штаммов *E. coli*

Штамм	Генотип
AB1157	F <sup>-</sup> , <i>recA99, thr-1, leu-6, thi-1, lacY1, galK2, ara-14, xyl-5, mtl-1, proA2, his-4, argE3, str-31, tsx-33</i>
AB1899	F <sup>-</sup> , <i>recA99, thr-1, leu-6, thi-1, lacY1, galK2, ara-12, xyl-5, mtl-1, proA2, his-4, argE3, str-31, tsx-33, lon100</i>
HB101	F <sup>-</sup> , <i>hsdS20 (r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>), recA13, ara-14, proA2, lacY1, galK2, rpsL20 (Sm<sup>r</sup>), xyl-5, mtl-1, supE44, λ<sup>-</sup></i>
K-12	<i>thi-1, galK2, strA</i>
MH1	<i>araD139, lacX74, galU, galK, hsr<sup>-</sup>, hms<sup>+</sup>, strA</i>
VL901	<i>hsdR<sup>-</sup>, met<sup>-</sup>, supE, thi-1, lonΔ11</i>

ли 1 ч при 60° С с осторожным перемешиванием, добавляли 60 мл фенола, насыщенного 50 мМ трис-НСI-буфером, рН 8,0, тщательно перемешивали до образования эмульсии, центрифугировали (30 мин, 4000g), собирали водную фазу, содержащую высокомолекулярную ДНК, и добавляли 0,1 объема 3 М ацетата натрия (рН 5,2) и 2 объема этанола. Выпавшую в осадок ДНК выдерживали в течение ночи в 50 мл 50 мМ трис-НСI-буфера, рН 7,5, с 1 мМ EDTA и 200 мкг/мл РНКазы (4° С, слабое перемешивание). Полученный вязкий раствор дважды экстрагировали хлороформом, осаждали этанолом, осадок растворяли в 10 мл 50 мМ трис-НСI-буфера, рН 7,5, с 1 мМ EDTA при 4° С и слабом перемешивании. Высокомолекулярную ДНК окончательно очищали центрифугированием в градиенте концентрации CsCl [26]. Фракции, содержащие ДНК, идентифицировали по увеличению вязкости раствора. Препарат ДНК, полученный после интенсивного диализа против 20 мМ трис-НСI-буфера, рН 7,5, с 1 мМ EDTA, эффективно гидролизовался рестрикционными эндонуклеазами и использовался для проведения опытов по клонированию и блот-гибридизации.

**Гибридизация ДНК.** Для опытов по блот-гибридизации использовали мембраны Gen Screen Plus (DuPont, США). Гибридизацию колоний проводили на нитроцеллюлозных фильтрах BA-85 (Schleicher und Schüll, ФРГ) [27]. Олигонуклеотидный зонд метили с помощью T4-полинуклеотидкиназы [27], а фрагменты ДНК — ник-трансферазой [27].

**Клонирование.** Очищенную ДНК *E. coli* обрабатывали (2 ч, 37° С) рестрикционными эндонуклеазами (10 ед. акт. на 1 мг ДНК), фрагменты фракционировали центрифугированием в градиенте концентрации NaCl (5—30%) [26] или сахарозы (10—40%) [27]. Фрагменты размером 1—4 т. п. о. использовали для клонирования в плазмидных векторах pBR327 и pSP64. Рестрикты плазмид были обработаны щелочной фосфатазой из кишечника тельца (КФ 3.1.3.1) [27]. Лигирование [27] проводили при 14° С в течение ночи. Полученный материал был использован для трансформации штаммов *E. coli* HB101 и MH1.

**Трансформация.** Компетентные клетки были приготовлены с использованием хлористого кальция и хранились при -70° С в присутствии 15% глицерина. Трансформацию проводили по стандартной методике [27]. Тепловой шок клеток AB1899 и VL901 проводили в течение 3 мин при 37° С. Трансформационную смесь рассеивали на чашки с LB-агаром, содержащим 50 мкг/мл ампициллина, и инкубировали в течение ночи при 37° С.

**Выделение плазмидной ДНК** осуществляли методом щелочного лизиса с последующим центрифугированием в градиенте концентрации CsCl [27].

**Нуклеотидную последовательность** определяли по методу Максама — Гилберта [28] в модификации Чувпило — Кравченко [29]. Для введения радиоактивной метки в ДНК использовали [ $\gamma$ -<sup>32</sup>P]АТФ и T4-полинуклео-

тидкиназу (5'-концевая метка), а также [ $\alpha$ - $^{32}$ P]dNTP и фрагмент Кленова ДНК полимеразы Т (3'-концевая метка). При секвенировании фрагментов ДНК, полученных в результате действия рестриктаз, образующих 3'-выступающие концы (*Pst*I, *Sph*I, *Sac* I, *Kpn*I), были использованы терминальная трансфераза и [ $\alpha$ - $^{32}$ P]ddATP (3'-labelling kit). Химическую модификацию оснований [26] проводили на DEAE-бумаге DE-81 (Whatman, Англия). При реконструкции нуклеотидной последовательности использовали программу Contig [30], любезно предоставленную П. В. Костецким (ИБХ им. М. М. Шемякина АН СССР).

*Внутриклеточный протеолиз* пуромидиновых и канаваниновых пептидов измеряли по методу [16].

*Определение АТР-зависимой протеолитической активности.* Клетки *E. coli* из различных клонов выращивали на среде LB при 30° С до оптического поглощения  $A_{600}$  1,0—1,2 и затем проводили термоиндукцию в течение 1 ч при 42° С. Культуру клеток (250 мл) центрифугировали (15 мин, 6000 об/мин), осадок промывали 0,1 М калий-фосфатным буфером, рН 7,0, и суспендировали в 2 мл того же буфера. Суспензию обрабатывали (20 мин, 0° С) 200 мкл раствора лизоцима (10 мг/мл), клетки разрушали ультразвуком (22 кГц, 5 × 5 с), суспензию разбавляли холодной водой до 10 мл и центрифугировали (45 мин, 40 000g). 4 мл супернатанта обрабатывали 2 мл фосфоцеллюлозы Р-11, уравновешенной 0,1 М калий-фосфатным буфером, рН 7,0, встряхивали 12 ч (0° С), отделяли смолу центрифугированием и промывали ее тем же буфером до величины  $A_{280} \sim 0,01$ . Связавшиеся со смолой белки элюировали 2 мл 0,7 М калий-фосфатного буфера, рН 7,0, в элюате определяли протеолитическую активность с использованием [ $^{14}$ C]ацетил- $\alpha$ -казеина. Реакционная смесь объемом 500 мкл содержала 100 мкл раствора субстрата (20—50 мкг, (3—5) · 10<sup>4</sup> имп/мин) в 50 мМ трис-НСI-буфере (рН 7,5), 20 мМ MgCl<sub>2</sub>, 3 мМ АТР и 400 мкл раствора фермента. Контрольные пробы не содержали АТР. Смесь инкубировали 1,5 ч (37° С), приливали 50 мкл 10% раствора бычьего сывороточного альбумина и 100 мкл 50% трихлоруксусной кислоты. Через 30 мин осадок отделяли центрифугированием и определяли радиоактивность в аликвоте супернатанта. Степень гидролиза субстрата рассчитывали как отношение радиоактивности в супернатанте к общей радиоактивности в пробе, каждый раз вычитая фон. Проводили по 3—5 определений.

*Электрофорез белков* осуществляли по методу [31].

Авторы выражают благодарность П. В. Костецкому и А. Н. Некрасову за помощь в компьютерной обработке полученных результатов и А. Е. Горбалене за полезные дискуссии.

#### СПИСОК ЛИТЕРАТУРЫ

1. Shineberg B., Zipser D. // J. Bacteriol. 1973. V. 116. № 3 (2). P. 1469—1471.
2. Goff S. A., Goldberg A. L. // Cell. 1985. V. 41. № 2. P. 587—595.
3. Mizusawa S., Gottesman S. // Proc. Nat. Acad. Sci. USA. 1983. V. 80. № 2. P. 358—362.
4. Phillips T. A., Van Bogelen R. A., Neidhardt F. C. // J. Bacteriol. 1984. V. 159. № 1. P. 283—287.
5. Trisler P., Gottesman S. // J. Bacteriol. 1984. V. 160. № 1. P. 184—191.
6. Waksman G., Thomas G., Favre A. // Photochem. Photobiol. 1984. V. 39. P. 337—342.
7. Zehnbaauer B. A., Markovitz A. // J. Bacteriol. 1980. V. 143. № 2. P. 852—863.
8. Maurizi M. R., Trisler P., Gottesman S. // J. Bacteriol. 1985. V. 164. № 3. P. 1124—1135.
9. Gayda R. C., Stephens P. E., Hewick R., Shoemaker J. M., Dreyer W. J., Markovitz A. // J. Bacteriol. 1985. V. 162. № 1. P. 271—275.
10. Waxman L., Goldberg A. L. // Proc. Nat. Acad. Sci. USA. 1982. V. 79. № 16. P. 4883—4887.
11. Goldberg A. L., Swamy K. H. S., Chung C. H., Larimore F. S. // Meth. Enzymol. 1981. V. 80. P. 680—702.
12. Америк А. Ю., Чистякова Л. Г., Остроумова Н. И., Гуревич А. П., Антонов В. К. // Биоорганич. химия. 1988. Т. 14. № 3. С. 408—411.
13. Amerik A. Yu., Chistyakova L. G., Ostroumova N. I., Antonov V. K. // 14th International Congress of Biochemistry (Prague, Czechoslovakia), Abstracts of Papers. 1988. V. 1. № 579. P. 204.

14. Goldberg A. L., Driscoll J., Chin D. T., Chung C. H., Mennon A. S. // 14th International Congress of Biochemistry (Prague, Czechoslovakia), Abstracts of Papers. 1988. V. 3. № S6—6. P. 72—73.
15. Chin D. T., Goff S. A., Webster T., Smith T., Goldberg A. L. // J. Biol. Chem. 1988. V. 263. № 24. P. 41718—41728.
16. Goldberg A. L. // Proc. Nat. Acad. Sci. USA. 1972. V. 69. № 9. P. 2640—2644.
17. Goff S. A., Goldberg A. L. // J. Biol. Chem. 1987. V. 262. № 10. P. 4508—4515.
18. Hatcher V. B., Wertheim M. S., Rhee Ch. Y., Tsien G., Burk P. G. // Biochim. et biophys. acta. 1976. V. 451. № 2. P. 499—510.
19. Ptitsyn O. B., Finkelstein A. V. // Biopolymers. 1983. V. 22. № 1. P. 15—25.
20. Saraste M., Eberle A., Gay N. J., Runswick M. J., Walker J. E. // Nucl. Acids Res. 1981. V. 9. № 8. P. 5287—5296.
21. Walker J. E., Saraste M., Runswick M. J., Gay N. J. // EMBO. J. 1982. V. 1. P. 945—951.
22. Sancar A., Stachelek C., Konigsberg W., Rupp W. D. // Proc. Nat. Acad. Sci. USA. 1980. V. 77. № 5. P. 2611—2615.
23. Racaniell V. R., Baltimore D. // Proc. Nat. Acad. Sci. USA. 1981. V. 78. № 16. P. 4887—4891.
24. Baker M. E. // FEBS Lett. 1989. V. 244. № 1. P. 31—33.
25. Mennon A. S., Goldberg A. L. // J. Biol. Chem. 1987. V. 262. № 31. P. 14929—14934.
26. Кайзер К., Мюррей Н. Клонирование ДНК. Методы / Ред. Гловер Д. М. М.: Мир, 1988. С. 60—63.
27. Мануатис Т., Фрич Э., Сэмбрук Д. Молекулярное клонирование: Пер. с англ. М.: Мир, 1984. 479 с.
28. Maxam A. M., Gilbert W. // Meth. Enzymol. 1980. V. 65. P. 499—559.
29. Chuvpilo S. A., Kravchenko V. V. // FEBS Lett. 1985. V. 179. № 1. P. 34—36.
30. Костецкий П. В., Доброва И. Е. // Биоорган. химия. 1988. Т. 14. № 4. С. 515—521.
31. Laemmli U. K. // Nature. 1970. V. 227. № 5259. P. 680—685.

Поступила в редакцию  
26.X.1989

A. Yu., AMERIK, V. K. ANTONOV, N. I., OSTROUMOVA, T. V. ROTANOVA,  
L. G. CHISTYAKOVA

**CLONING, STRUCTURE AND EXPRESSION OF THE FULL *lon* GENE  
OF *ESCHERICHIA COLI* CODING FOR ATP-DEPENDENT PROTEASE La**

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

Assembly of the full *Escherichia coli* K-12 *lon* gene from the *Eco*RI — *Sph*I fragment of the bacterial DNA («modified» gene) cloned and sequenced earlier [12]) and the *Pst*I fragment of the same DNA containing 3'-terminal region of the *lon* gene has been performed. Both «modified» and full genes showed all phenotype properties of *lon* gene. The complete nucleotide sequence of the gene (2770 bp) coding for the 784 amino acid sequence of protease La was determined. Location of catalytically active serine, histidine and aspartic acid residues was suggested, and ATP-binding site found. The *lon* gene and protease La structures we found are compared with those described independently [15] and differences observed are discussed.