



УДК 577.213.7 : 577.152.34

КЛОНИРОВАНИЕ, ЭКСПРЕССИЯ И СТРУКТУРА
ФУНКЦИОНАЛЬНО АКТИВНОГО УКРОЧЕННОГО ГЕНА
lon *ESCHERICHIA COLI*Америк А. Ю., Чистякова Л. Г., Острозумова Н. И.,
Гурезич А. И., Антонов В. К.Институт биоорганической химии им. М. М. Шемякина
Академии наук СССР, Москва

Белки с измененной или неполной структурой подвергаются в клетке быстрой деградации [1]. Доминирующую роль в этом процессе в клетках *E. coli* играет АТР-зависимая La-протеиназа, кодируемая геном *lon* [2, 3]. Этот фермент относится к группе белков теплового шока [4] и участвует в таких процессах, как чувствительность клетки к УФ-облучению, клеточное деление и образование капсулярных полисахаридов. Протеиназа La представляет значительный интерес как с энзимологической точки зрения, так и в генной инженерии, для разработки способов предотвращения быстрой деградации экспрессируемых белков и пептидов.

Ранее ген *lon* был клонирован в F-эписоме [5] и фаге λ [6]. Была получена рестриктная карта гена [6, 7] и установлена нуклеотидная последовательность фрагмента, включающего регуляторный участок и 5'-концевую часть (174 п.о.) структурного гена La-протеиназы [8]. Сама La-протеиназа была получена экспрессией клонированного гена в клетках *E. coli* и оказалась тетрамерным белком с M_r^2 субъединицы ~ 90 кДа [9].

В настоящей работе для клонирования гена *lon* мы использовали плазмидный вектор pBR327. Суммарную ДНК *E. coli* K12 дикого типа подвергали совместному расщеплению рестриктазами *EcoRI* и *SphI*, полученную смесь фракционировали в градиенте сахарозы и фрагменты размером в 3—6 т.п.о клонировали в pBR327. Полученные рекомбинантные плазмиды амплифицировали в штамме *E. coli* HB101 и выделенными плазмидами трансформировали клетки *E. coli* AB1189, несущие мутацию по гену *lon*. Первичный отбор клонов проводили по увеличению устойчивости клеток к нитрофурантоину (имитирующему действие ультрафиолета [5]) и по виду колоний (отсутствие ослизненности, вызванной гиперпродукцией капсулярных полисахаридов), выращенных на минеральной среде М9. Среди отобранных было обнаружено 7 клонов, плазмидная ДНК которых гибридизовалась с нуклеотидным зондом CTGAACGCATTGAAA-TCCCGTATTGCC, соответствующим 5'-концевой части гена [8]. Рестриктная карта полученных плазмид (рис. 1) почти полностью совпадала

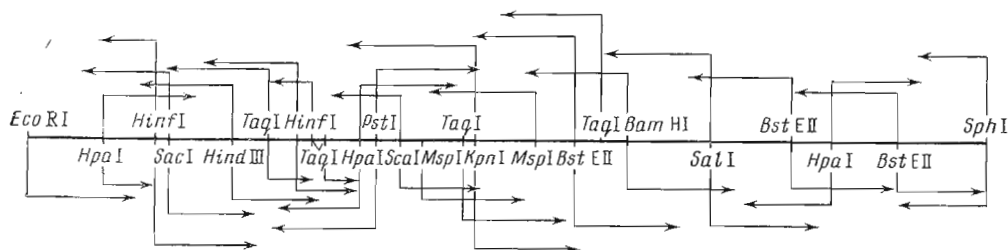


Рис. 1. Стратегия секвенирования гена *lon*. Указаны сайты рестрикции и направленные секвенирования по основной (вверху) и комплементарной (внизу) цепям

1 GAATTCCTGACGAGGCGCTGGATGCTATCGCTAAGAAAGCGATGGCGCTAAAACCG
 59 GTGCCCGTGGCCTGCGTTCCATCGTAGAAGCCGCACTGCTCGATACCATGTACGATCTGCCGTCCATGGAAG
 131 ATGTCGAAAAAGTGGTTATCGACGAGTCGGTAATTGATGGTCAAAGCAAACCGTTGCTGATTTATGGTAAGC
 203 CGGAAGCGCAACAGGCATCTGGTGAATAATTAACCAATCCCATACAATTAGTTAACCAAAAAGGGGGGATTT
 275 TATCTCCCTTTAATTTTTCTCTATTCTCGGCGTTGAATGTGGGGAAACATCCCATATACTGACGTACA
 347 TGTTAATAGATGGCGTGAAGCACAGTCGTGTCATCTGATTACCTGGCGAAATTAACCTAAGAGAGAGCTCT
 419 ATG AAT CCT GAG CGT TCT GAA CGC ATT GAA ATC CCC GTA TTG CCG CTG CGC GAT
 Met Asn Pro Glu Arg Ser Glu Arg Ile Glu Ile Pro Val Leu Pro Arg Asp
 473 GTG GTG GTT TAT CCG CAC ATG GTC ATC CCC TTA TTT GTC GGG CGG GAA AAA TCT
 Val Val Val Tyr Pro His Met Val Ile Pro Leu Phe Val Gly Arg Glu Lys Ser
 527 ATC CGT TGT CTG GAA GCG GCG ATG GAC CAT GAT AAA AAA ATT ATG CTG GTC GCG
 Ile Arg Cys Leu Glu Ala Ala Met Asp His Asp Lys Lys Ile Met Leu Val Ala
 581 CAG AAA GAA GCT TCA ACG GAT GAG CCG GGT GTA AAC GAT CTT TTC ACC GTC GGG
 Gln Lys Glu Ala Ser Thr Asp Glu Pro Gly Val Asn Asp Leu Phe Thr Val Gly
 635 ACC GTG GCC TCT ATA TTG CAG ATG CTG AAA CTG CCT GAC GGC ACC GTC AAA GTG
 Thr Val Ala Ser Ile Leu Gln Met Leu Lys Leu Pro Asp Gly Thr Val Lys Val
 689 CTG GTC GAG GGG TTA CAG CGC GCG CGT ATT TCT GCG CTC TCT GAC AAT GGC GAA
 Leu Val Glu Gly Leu Gln Arg Ala Arg Ile Ser Ala Leu Ser Asp Asn Gly Glu
 743 CAC TTT TCT GCG AAG GCG GAG TAT CTG GAG TCG CCG ACC ATT GAT GAG CGG GAA
 His Phe Ser Ala Lys Ala Glu Tyr Leu Glu Ser Pro Thr Ile Asp Glu Arg Glu
 797 CAG GAA GTG CTG GTG CGT ACT GCA ATC AGC CAG TTC GAA GGC TAC ATC AAG CTG
 Gln Glu Val Leu Val Arg Thr Ala Ile Ser Gln Phe Glu Gly Tyr Ile Lys Leu
 851 AAC AAA AAA ATC CCA CCA GAA GTG CTG ACG TCG CTG AAT AGC ATC GAC GAT CCG
 Asn Lys Lys Ile Pro Pro Glu Val Leu Thr Ser Leu Asn Ser Ile Asp Asp Pro
 905 GCG CGT CTG GCG GAT ACC ATT GCT GCA CAT ATG CCG CTG AAA CTG GCT GAC AAA
 Ala Arg Leu Ala Asp Thr Ile Ala Ala His Met Pro Leu Lys Leu Ala Asp Lys
 959 CAG TCC GTT CTG GAG ATG TCC GAC GTT AAC GAA CGT CTG GAA TAT CTG ATG GCA
 Gln Ser Val Leu Glu Met Ser Asp Val Asn Glu Arg Leu Glu Tyr Leu Met Ala
 1013 ATG ATG GAA TCG GAA ATC GAT CTG CTG CAG GTT GAG AAA CGC ATT CGC AAC CGC
 Met Met Glu Ser Glu Ile Asp Leu Leu Gln Val Glu Lys Arg Ile Arg Asn Arg
 1067 GTT AAA AAG CAG ATG GAG AAA TCC CAG CGT GAG TAC TAT CTG AAC GAG CAA ATG
 Val Lys Lys Gln Met Glu Lys Ser Gln Arg Glu Tyr Tyr Leu Asn Glu Gln Met
 1121 AAA GCT ATT CAG AAA GAA CTC GGT GAA ATG GAC GAC CGC CCG GAG GAA AAC GAA
 Lys Ala Ile Gln Lys Leu Glu Gly Glu Met Asp Asp Ala Pro Asp Glu Asn Glu
 1175 GCC CTG AAG CGC AAA ATC GAC GCG GCG AAG ATG CCG AAA GAG GCA AAA GAG AAA
 Ala Leu Lys Arg Lys Ile Asp Ala Ala Lys Met Pro Lys Glu Ala Lys Glu Lys
 1229 GCG GAA GCA GAG TTG CAG AAG CTG AAA ATG ATG TCT CCG ATG TCG GCA GAA GCG
 Ala Glu Ala Glu Leu Gln Lys Leu Lys Met Met Ser Pro Met Ser Ala Glu Ala
 1283 ACC GTA GTG CGT GGT TAT ATC GAC TGG ATG GTA CAG GTA CCG TGG AAT GCG CGC
 Thr Val Val Arg Gly Tyr Ile Asp Trp Met Val Gln Val Pro Trp Asn Ala Arg
 1337 AGC AAG GTC AAA AAA GAC CTG CGT CAG GCG CAG GAA ATC CTT GAT ACC GAC CAT
 Ser Lys Val Lys Lys Asp Leu Arg Gln Ala Gln Glu Ile Leu Asp Thr Asp His
 1391 TAT GGT CTG GAG CGC GTG AAA GAT CGC ATC CTT GAG TAC CTT GCG GTT CAA AGC
 Tyr Gly Leu Glu Arg Val Lys Asp Arg Ile Leu Glu Tyr Leu Ala Val Gln Ser
 1445 CGT GTC AAC AAA ATC AAG GGA CCA ATC CTC TGC CTG GTA GGG CCG CCG GGG GTA
 Arg Val Asn Lys Ile Lys Gly Pro Ile Leu Cys Leu Val Gly Pro Pro Gly Val
 1499 GGT AAA ACC TCT CTT GGT CAG TCC ATT GCC AAA GCC ACC GGG CGT AAA TAT GTC
 Gly Lys Thr Ser Leu Gly Gln Ser Ile Ala Lys Ala Thr Gly Arg Lys Tyr Val
 1553 CGT ATG GCG CTG GGC GGC GTG CGT GAT GAA GCG GAA ATC CGT GGT CAC CGC CGT
 Arg Met Ala Leu Gly Gly Val Arg Asp Glu Ala Glu Ile Arg Gly His Arg Arg
 1607 ACT TAC ATC GGT TCT ATG CCG GGT AAA CTG ATC CAG AAA ATG GCG AAA GTG GGC
 Thr Tyr Ile Gly Ser Met Pro Gly Lys Leu Ile Gln Lys Met Ala Lys Val Gly
 1661 GTG AAA AAC CCG CTG TTC CTG CTC GAT GAG ATC GAC AAA ATG TCT TCT GAC ATG
 Val Lys Asn Pro Leu Phe Leu Leu Asp Glu Ile Asp Lys Met Ser Ser Asp Met
 1715 CGA GGC GAT CCG GCC TCT GCA CTG CTT GAA GTG CTG GAT CCA GAG CAG AAC GTA
 Arg Gly Asp Pro Ala Ser Ala Leu Leu Glu Val Leu Asp Pro Glu Gln Asn Val
 1769 GCG TTC AGC GAC CAC TAC CTG GAA GTG GAT TAC GAT CTC AGC GAC GTG ATG TTT
 Ala Phe Ser Asp His Tyr Leu Glu Val Asp Tyr Asp Leu Ser Asp Val Met Phe
 1823 GTC GCG ACG TCG AAC TCC ATG AAC ATT CCG GCA CCG CTG CTG GAT CGT ATG GAA
 Val Ala Thr Ser Asn Ser Met Asn Ile Pro Ala Pro Leu Leu Asp Arg Met Glu
 1877 GTG ATT CGC CTC TCC GGT TAT ACC GAA GAT GAA AAA CTG AAC ATC GCC AAA CGT
 Val Ile Arg Leu Ser Gly Tyr Thr Glu Asp Glu Lys Leu Asn Ile Ala Lys Arg
 1931 CAC CTG CTG CCG AAG CAG ATT GAA CGT AAT GCA CTG AAA AAA GGT GAG CTG ACC
 His Leu Leu Pro Lys Gln Ile Glu Arg Asn Ala Leu Lys Lys Gly Glu Leu Thr

1985 GTC GAC GAT AGC GCC ATT ATC GGC ATT ATT CGT TAC TAC ACC CGT GAG GCG GGC
Val Asp Asp Ser Ala Ile Ile Gly Ile Ile Arg Tyr Tyr Thr Arg Glu Ala Gly

2039 GTG CGT GGT CTG GAG CGT GAA ATC TCC AAA CTG TGC CGC AAA GCG GTT AAG CAG
Val Arg Gly Leu Glu Arg Glu Ile Ser Lys Leu Cys Arg Lys Ala Val Lys Gln

2093 TTA CTG CTC GAT AAG TCA TTA AAA CAT ATC GAA ATT AAC GGC GAT AAC CTG CAT
Leu Leu Leu Asp Lys Ser Leu Lys His Ile Glu Ile Asn Gly Asp Asn Leu His

2147 GAC TAC CTC GGT GTT CAG CGT TTC GAC TAT GGT CGC GCT GAT AAC GAA AAC CGT
Asp Tyr Leu Gly Val Gln Arg Phe Asp Tyr Gly Arg Ala Asp Asn Glu Asn Arg

2201 GTC GGT CAG GTA ACC GGT CTG GCG TGG ACG GAA GTG GGC GGT GAC TTG CTG ACC
Val Gly Gln Val Thr Gly Leu Ala Trp Thr Glu Val Gly Gly Asp Leu Leu Thr

2255 ATT GAA ACC GCA TGT GTT CCG GGT AAA GGC AAA CTG ACC TAT ACC GGT TCG CTC
Ile Glu Thr Ala Cys Val Pro Gly Lys Gly Lys Leu Thr Tyr Thr Gly Ser Leu

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

2363 GCG GAA AAA CTG GGG ATC AAC CCT GAT TTT TAC GAA AAA CGT GAC ATC CAC GTC
Ala Glu Lys Leu Gly Ile Asn Pro Asp Phe Tyr Glu Lys Arg Asp Ile His Val

2417 CAC GTA CCG GAA GGT GCG ACG CCG AAA GAT GGT CCG AGT GCC GGT ATT GCT ATG
His Val Pro Glu Gly Ala Thr Pro Lys Asp Gly Pro Ser Ala Gly Ile Ala Met

2471 TGC ACC GCG CTG GTT TCT TGC CTG ACC GGT AAC CCG GTT CGT GCC GAT GTG GCA
Cys Thr Ala Leu Val Ser Cys Leu Thr Gly Asn Pro Val Arg Ala Asp Val Ala

2525 ATG ACC GGT GAG ATC ACT CTG CGT GGT CAG GTA CTG CCG ATC GGT GGT TTG AAA
Met Thr Gly Glu Ile Thr Leu Arg Gly Gln Val Leu Pro Ile Gly Gly Leu Lys

2579 GAA AAA CTC CTG GCA GCG CAT CCG GGC GGG ATT AAA ACA GTG CTA ATT CCG TTC
Glu Lys Leu Leu Ala Ala His Arg Gly Gly Ile Lys Thr Val Leu Ile Pro Phe

2633 GAA AAT AAA CGC GAT CTG GAA GAG ATT CCT GAC AAC GTA ATT GCC GAT CTG GAC
Glu Asn Lys Arg Asp Leu Glu Ile Pro Asp Asn Val Ile Ala Asp Leu Asp

2687 ATT CAT CCT GTG AAG CGC ATT GAG GAA GTT CTG ACT CTG GCG CTG CAA AAT GAA
Ile His Pro Val Lys Arg Ile Glu Glu Val Leu Thr Leu Ala Leu Gln Asn Glu

2741 CCG TCC GGC ATG CAC CAT TCC TTG CGG CGG CGG TGC TCA ACG GCC TCA ACC TAC
Pro Ser Gly Met His His Ser Leu Arg Arg Arg Cys Ser Thr Ala Ser Thr Tyr

2795 TAC TGG GCT GCT TCC TAA
Tyr Trp Ala Ala Ser ***

Рис. 2. Нуклеотидная последовательность ДНК гена *lon* и кодируемая ею аминокислотная последовательность. Рамкой выделен нуклеотид 464 (см. текст). Подчеркнуты сайты рестрикции *Hind* III (одной чертой) и *Sph* I (двумя чертами). Стрелкой указан конец клонированного гена.

с известной картой *lon*-гена, однако в нашем случае был обнаружен новый *Sph*I-сайт.

Рекомбинантные плазмиды вызывали исчезновение всех характерных признаков *lon*-мутации у штаммов *E. coli* AB1189, VL 901 и *pVI deg*. Электрофорез белков клеток *E. coli* VL901, содержащих плазмиду с клонированным геном, обнаруживал появление полосы с $M_r \sim 90$ кДа, отсутствующей среди белков контрольных клеток, содержащих плазмиду *pBR327*. Была измерена скорость внутриклеточного протеолиза пуромидиновых пептидов в клетках *E. coli* AB1189 и в изогенном штамме AB1157 [6]. Было показано, что при введении в клетки AB1189 плазмиды, содержащей клонированный ген, скорость протеолиза значительно возрастала и существенно превышала скорость протеолиза в клетках штамма AB1157, не несущих мутации в области *lon*-гена.

Стратегия определения нуклеотидной последовательности клонированного гена методом Максама — Гилберта [10] показана на рис. 1. Структура гена и соответствующая ей аминокислотная последовательность белка представлены на рис. 2. Сопоставление полученной нами 5'-концевой последовательности с последовательностью фрагмента *Eco*RI — *Hind*III гена *lon*, опубликованной ранее [8], выявило отсутствие в опубликованной структуре одного нуклеотида в положении 464. По нашему мнению, этот нуклеотид был пропущен при определении структуры фрагмента в работе [8], где использовался метод Сенгера [11].

Оказалось, что клонированный нами ген не имеет терминирующего кодона. Терминация происходит уже после *Sph*I-сайта, по которому проводилась рестрикция геномной ДНК, на участке вектора. Таким образом, клонированный нами ген, по-видимому, имеет укороченную структуру,

а образующийся при его экспрессии белок имеет по сравнению с La-протеиназой измененную С-концевую последовательность аминокислот. Полученный нами укороченный ген *lon*, очевидно, мало отличается от полного гена, так как вычисленная на основе его структуры молекулярная масса белка (88913 Да) близка к массе La-протеиназы. Интересно, что эти изменения в гене *lon* не сказываются на его фенотипических проявлениях и на способности образующегося белка к АТР-зависимому протеолизу. Попытки выделения этого белка приводят к получению препаратов, обладающих выраженной АТР-зависимой протеолитической активностью, но стабильность ее крайне низка. По-видимому, С-концевой фрагмент La-протеиназы играет важную роль в стабилизации белковой молекулы.

В настоящее время нами проводятся опыты по получению полноразмерного гена *lon*.

Авторы благодарны В. Н. Добрынину и С. А. Филиппову за синтез нуклеотидного зонда, а также Т. В. Ротановой и И. А. Дывак за помощь в тестировании протеолитической активности.

ЛИТЕРАТУРА

1. Goldberg A. L., St. John A. C. // Annu. Rev. Biochem. 1976. V. 45. P. 747—803.
2. Hershko A., Ciehanover A., // Annu. Rev. Biochem. 1982. V. 51. P. 335—364.
3. Chung C. H., Goldberg A. L. // Proc. Nat. Acad. Sci. USA. 1981. V. 78. № 8. P. 4931—4935.
4. Phillips T. A., VanBogelen R. A., Neidhart F. C. // J. Bacteriol. 1984. V. 159. № 1. P. 283—287.
5. Zehnbauser B. A., Markovitz A. // J. Bacteriol. 1980. V. 143. № 3. P. 852—863.
6. Maurizi M. R., Trisler P., Gottesman S. // J. Bacteriol. 1985. V. 164. № 3. P. 1124—1135.
7. Schoemaker J. M., Markovitz A. // J. Bacteriol. 1981. V. 147. № 1. P. 46—56.
8. Gayda R. C., Stephens P. E., Newik R., Schoemaker J. M., Dreyer W. J., Markovitz A. // J. Bacteriol. 1985. V. 162. № 1. P. 271—275.
9. Charette M. F., Henderson G. W., Kezdy F. J., Markovitz A. // J. Mol. Biol. 1982. V. 162. № 2. P. 503—510.
10. Mazam A. M., Gilbert W. // Proc. Nat. Acad. Sci. USA. 1977. V. 74. № 2. P. 560—564.
11. Sanger F., Nicklen S., Coulson A. R. // Proc. Nat. Acad. Sci. USA. 1977. V. 74. № 12. P. 5463—5467.

Поступило в редакцию

23.VI.1987

После доработки

2.XI.1987

CLONING, EXPRESSION AND STRUCTURE OF THE SHORTENED, FUNCTIONALLY ACTIVE *Lon* GENE OF *ESCHERICHIA COLI*

AMERIK A. Yu., CHISTYAKOVA I. G., OSTROUMOVA N. I., GUREVICH A. I.,
ANTONOV V. K.

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

Lon gene of *E. coli* has been cloned into the plasmid pBR327. Full nucleotide sequence of the gene has been established. It was shown that the cloned gene does not possess the terminal codon and is somewhat shortened. Nevertheless it retains full phenotypic activity and expresses the C-end modified La proteinase which retains ATP-dependent proteolytic activity.