



УДК 577.113.5

ПЕРВИЧНАЯ СТРУКТУРА ГЕНА ЭСТЕРАЗЫ
*S DROSOPHILA VIRILIS*Сергеев П. В., Кастильо Х. Э., Шеунова Н. И.,
Евниколопов Г. Н.Институт молекулярной биологии им. В. А. Энгельгардта Академии наук
СССР, Москва

Эстераза S (КФ 3.1.1.1.) *Drosophila virilis*, относящаяся к классу β -карбоксилэстераз, синтезируется в семявыносящей луковиче зрелого самца в строго детерминированный период индивидуального развития и влияет на половое поведение и фертильность дрозофил [1, 2]. Ген эстеразы S входит в состав сложной регуляторной цепи, так что интенсивность синтеза фермента, время начала его синтеза и другие характеристики находятся под генетическим контролем [1–3]. Ранее нами была клонирована хромосомная копия гена эстеразы S и проведено его начальное описание — локализован промотор гена и определена экзоинтронная структура [4–6]. Настоящее сообщение посвящено установлению первичной структуры гена *EstS*, выведению аминокислотной последовательности белка и сравнению ее с последовательностями других эстераз животных.

Для определения первичной структуры мы использовали клонированные субфрагменты хромосомной копии гена эстеразы S [4, 5]. Последовательность ДНК определяли методом химической модификации по Максиму — Гилберту в обычной [7] и твердофазной модификациях [8], а также ферментативным методом Сепгера в модификации для двуспиральной плазмидной ДНК [9]. Были использованы отдельные рестриктные фрагменты гена и наборы делеционных мутантов, полученных после переклонирования отдельных фрагментов исходной плазмиды pVE9 с последующей обработкой ДНКазой I, лигированием и трансформацией [10]. Фрагменты и субклоны секвенировали так, чтобы каждый нуклеотид был прочитан несколько раз и в обоих направлениях.

Длина мРНК составляет 1808 нуклеотидов (без учета 3'-концевой poly(A)-последовательности, добавляемой посттранскрипционно), длина белка — 542 аминокислоты (рисунок). Мы предполагаем, что указанная структура соответствует проферменту и что первые 16 триплетов гена кодируют сигнальный пептид, отщепляющийся при прохождении белка через мембрану (большая часть молекул эстеразы S секретируется в просвет семявыносящей луковичи). Этот участок белка несет отличительные особенности сигнальных пептидов [11]: насыщенность гидрофобными аминокислотами (12 из 16), наличие нескольких полярных аминокислот на N-конце (Met-Thr-Gln) и гидрофобной аминокислоты с короткой боковой цепью (Ala) в последнем положении сигнального пептида. На границу между сигнальным пептидом и зрелым ферментом указывают также данные определения N-конца родственного белка *Est6 D. melanogaster* [12]. Зрелый белок, таким образом, должен иметь длину 526 аминокислот и мол. массу 59 220. Результаты биохимического изучения говорят о том, что белок является гликопротеином, и нами было обнаружено несколько потенциальных участков N-гликозилирования по остаткам аспарагина [13] (отмечены стрелками на рисунке).

Сравнение эстеразы *S D. virilis* с другими исследованными эстеразами (генами и белками) демонстрирует значительное сходство с карбоксилэстеразами млекопитающих и ацетилхолинэстеразами различных классов

GAACAA TTCAAGTTTG AGCTCCGTCC TGTGCACACC

ATG ACT CAA ATA CTG TTG CCG ATC GCG TTG CTC TGT CTG TTT GCA
Met Thr Gln Ile Leu Leu Pro Ile Ala Leu Leu Cys Leu Phe Ala 15

GCA TCA ACC CTC AGC AAT CCC CTG CTC GTG GAG TTG CCA AAT GGA
Ala Ser Thr Leu Ser Asn Pro Leu Leu Val Glu Leu Pro Asn Gly 30

GAA TTG CCG GGA CGC GAC AAT GGA TTC TAT TAC AGC TAC GAG TCA
Glu Leu Arg Gly Arg Asp Asn Gly Phe Tyr Tyr Ser Tyr Glu Ser 45

ATA CCT TAT GCC GAA CCC CCA ATC GAT GAT CTG TGC TTG GAA GAA
Ile Pro Tyr Ala Glu Pro Pro Ile Asp Asp Leu Cys Leu Glu Glu 60

CCT CGT CCC TAT ACC GAA AGA TGG GAA AAT ACC TTT GAC GCG ACT
Pro Arg Pro Tyr Thr Gln Arg Trp Glu Asn Thr Phe Asp Ala Thr 75

CGG CCC CCA GTT GAC TGC CTG CAG TGG AGT CAA CTC ATT TCA CAG
Arg Pro Pro Val Glu Cys Leu Gln Trp Ser Gln Leu Ile Ser Gln 90

CCT AAT AAG CTG ACA GGG AGC GAG GAC TGT CTA ACC GTC AGC ATC
Pro Asn Lys Leu Thr Gly Ser Glu Asp Cys Leu Thr Val Ser Ile 105

TAC AAG CCA AAG AAT CTG ACT CGC ATC TCT TTT CCG GTG GTG GCG
Tyr Lys Pro Lys Asn Leu Thr Arg Ile Ser Phe Pro Val Val Ala 120

CAT ATA TTT GGC GGC GGC TGG TCA TTT GGT GCT GCG ATC GAT GAC
His Ile Phe Gly Gly Gly Trp Ser Phe Gly Ala Ala Ile Asp Asp 135

GGA GTG AGG CCC TTC AGC AGC AGC GGC AAT GTG ATA GTG GTG AAG
Gly Val Arg Pro Phe Ser Ser Ser Gly Asn Val Ile Val Val Lys 150

ACA ACC ACA GAG TGG GAG CGC TTG GGT TTT ATG AGC ACT GGT GAT
Thr Thr Thr Glu Trp Glu Arg Leu Gly Phe Met Ser Thr Gly Asp 165

TCT GTG ATT CCG GGC AAC TTC GGA CTA AAG GAT CAG CGT CTG GCA
Ser Val Ile Pro Gly Asn Phe Gly Leu Lys Asp Gln Arg Leu Ala 180

ATC AAA TGG ATT AGG AAC AAC ATT GCA CGC TTT GGC GGA GAT CCA
Ile Lys Trp Ile Arg Asn Asn Ile Ala Arg Phe Gly Gly Asp Pro 195

CAT AAT ATA ATT CTT CTC GGT TTC AGT ACA GGC GGC TCC TCG GTG
His Asn Ile Ile Leu Leu Gly Phe Ser Thr Gly Gly Ser Ser Val 210

CAC TTG CAG CTT ATG CAC AAG GAA TAT GGA CAG CTG GTG AAG GGG
His Leu Gln Leu Met His Lys Glu Tyr Gly Gln Leu Val Lys Gly 225

GCC ATA TCC ATT AGT GGA ACT GCA ACG GTT CCC TGG GCT GTA CAG
Ala Ile Ser Ile Ser Gly Thr Ala Thr Val Pro Trp Ala Val Gln 240

GCC AAT GCA CGT GAT CTC GCA TTC CGA TAT GGC AAA CTC TTG GGT
Ala Asn Ala Arg Asp Leu Ala Phe Arg Tyr Gly Lys Leu Leu Gly 255

TGT AAT AAC CCT AAA AAT TCA CGC GAG CTG AAA GAT TGC CTG AAA
Cys Asn Asn Pro Lys Asn Ser Arg Glu Leu Lys Asp Cys Leu Lys 270

AAA ACG GAT GCG GAA GAA TTC GTC AGC ACC TTA AGG CAC CTT CAG
Lys Thr Asp Ala Glu Glu Phe Val Ser Thr Leu Arg His Leu Gln 285

GTG TTT GAC TAT GTG CCT TTT GGT CCG TTT GGC CCA GTC GTA GAG
Val Phe Asp Tyr Val Pro Phe Gly Pro Phe Gly Pro Val Val Glu 300

TCC CCC GAA GTG GAA AGC CCC TTC CTC ACC GAG CTG CCC CTC GAC
Ser Pro Glu Val Glu Ser Pro Phe Leu Thr Glu Leu Pro Leu Asp 315

ACC ATC AGA AGT GGA AAC TTT GCT CAA GTG CCT TGG TTG GCC AGC
Thr Ile Arg Ser Gly Asn Phe Ala Gln Val Pro Trp Leu Ala Ser 330

TAC ACA CCC GAG AAT GGT ATC TAT AAC GCC GCT CTT CTT TTA GCT
Tyr Thr Pro Glu Asn Gly Ile Tyr Asn Ala Ala Leu Leu Leu Ala 345

AAG GAC GCC AAT GGT AAA GAG AGG ATT GAA GAG TTA AAC ACT CGC
Lys Asp Ala Asn Gly Lys Glu Arg Ile Glu Glu Leu Asn Thr Arg 360

TGG AAC GAA CTG GCT CCA TAC TTT TTG GCT TAC CCA TAC ACA TTG
Trp Asn Glu Leu Ala Pro Tyr Phe Leu Ala Tyr Pro Tyr Thr Leu 375

AAG AGG TCT GAA ATG AAT GCT CAT TCT CAG AAA CTG AAA TAT CAA
Lys Arg Ser Glu Met Asn Ala His Ser Gln Lys Leu Lys Tyr Gln 390

TAT CTA GGA TAT AAG AAC TTC AGC GTC GTA AAC TAT TTC GAT GTT
Tyr Leu Gly Tyr Lys Asn Phe Ser Val Val Asn Tyr Phe Asp Val 405

CAG CGC TTG TTT ACA AAC GAG TTG TAC AAA AAA GGC ATC GAA CTG
Gln Arg Leu Phe Thr Asn Glu Leu Tyr Lys Lys Gly Ile Glu Leu 420

TCA TTA GAT TCA CAT CGC AAG CAC GGA GCC AGC CCC GTC TAT GCG
Ser Leu Asp Ala His Arg Lys His Gly Ala Ser Pro Val Tyr Ala 435

TAC GTC TAC GAC AAT CCC GCG GAT AAG TCG CTG GCA CAA TTC CTG
Tyr Val Tyr Asp Asn Pro Ala Asp Lys Ser Leu Ala Gln Phe Leu 450

GCC AAG AGA TCT GAT ATA TCC TTG GGC ACC GGA ATG GGC GAC GAT
Ala Lys Arg Ser Asp Ile Ser Leu Gly Thr Gly Met Gly Asp Asp 465

TAC TAT CTC TTG ATG AAC AAT CCA CTG CGT GAA CCG CTG CGA GCT
Tyr Tyr Leu Leu Met Asn Asn Pro Leu Arg Glu Pro Leu Arg Ala 480

GAC GAG AAA ATC GTT TCA TGG AAG CTA GTC AAG ATG GTG GAA GAT
Asp Glu Lys Ile Val Ser Trp Lys Leu Val Lys Met Val Glu Asp 49

TTC GCC GCG CAC GAG ACC TTG GTC TAT GAC GAC TGT GTG TTC CCA
Phe Ala Ala His Glu Thr Leu Val Tyr Asp Asp Cys Val Phe Pro 510

AAC AAC TTG GGC AAA AAG AAA TTC CAG TTG GTG GTC ATA GGA CGT
Asn Asn Leu Gly Lys Lys Lys Phe Gln Leu Val Val Ile Gly Arg 525

AAC TAT TGC AAG CAA TTG GAA GTG GAA TCG TTT GCC CGA CAC GGT
Asn Tyr Cys Lys Gln Leu Glu Val Glu Ser Phe Ala Arg His Gly 540

GTC CAA TAA TCTACCACCA AATCCAATTG TAATTTACTG AGCCGATCTC
Val Gln

TATGAAACCA TCAAATCAAT AGTCGATGCT TGTAATTTTC TATAAATGTA TATG

TGGGTA TGTAATCTT CTGCAAATAA ATCATGGCAT TATTTAATAA CTCAAAAA

Первичная структура гена эстеразы *S. D. virilis* и кодируемая им аминокислотная последовательность. Стрелками отмечены потенциальные участки N-гликозилирования

животных. Особенно велико сходство с нуклеотидной и аминокислотными последовательностями эстеразы 6 *D. melanogaster* [12] — ферментом, по своей физиологической роли соответствующим эстеразе *S. D. virilis* и находящимся под сходным генетическим контролем. Особенного внимания заслуживает область белка с координатами 197—213, сходная с областями активных центров известных эстераз.

Сравнение аминокислотных последовательностей предполагаемого активного центра эстеразы S и активных центров других эстераз

Ests	AAT Asn	ATA Ile	ATT Ile	CTT Leu	CTC Leu	GGT Gly	TTC Phe	AGT Ser	ACA Thr	GGC Gly	GGC Gly	TCC Ser	TCG Ser	GTG Val	CAC His	TTG Leu	CAG Gln						
197																	213						
<i>EstS D. virilis</i>							N	I	I	L	L	G	F	S	T	G	G	S	S	V	H	L	Q
<i>Est6 D. melanogaster</i>							N	V	L	L	V	G	H	S	A	G	G	A	S	V	H	L	Q
Ацетилхолинэстераза ската							T	V	T	I	F	G	E	S	A	G	G	A	S	V	G	M	H
» утря											G	E	S	S	E	G	A	A	G				
Алиэстераза лошади											F	G	E	S	A	G	A	A	S				
Холинэстераза человека							S	V	T	L	F	G	E	S	A	G	A	A	S	V	S	L	H
Карбоксилэстераза курицы											G	E	S	A	G	G	I	S					
» свиньи											G	E	S	A	G	G	E	S					
» овцы											G	E	S	A	G	G	E	S					
» быка											G	E	S	A	G	G	E	S					
Консенсус											G	E	S	A	G	G	A	S	V				

Из сравнения структур этих участков может быть выведена усредненная последовательность Gly-Glu-Ser-Ala-Gly-Gly-Ala-Ser-Val, общая для эстераз таких эволюционно далеких объектов, как человек, скат, дрозофила и др. (рис. 2). Обращает на себя внимание неизменность расположения остатков Gly в положениях 202, 206, 207 и Ser в положениях 204 и 309. Сходство особенно велико у эстеразы *S. D. virilis* и эстеразы 6 *D. melanogaster* — совпадают 11 аминокислот из 17. Отметим, что сходство велико и в других областях, входящих в формирование активных центров ферментов.

Сравнение аминокислотной последовательности эстеразы *S. D. virilis* с известными последовательностями карбоксилэстераз, ацетилхолинэстераз и сериновых протеиназ позволит глубже понять взаимоотношения членов суперсемейства сериновых гидролаз, к которому относятся эти классы ферментов.

Авторы благодарны А. Б. Жаданову, О. А. Малеванчук (ИМБ АН СССР) за оказанную помощь в работе, а также Г. П. Георгиеву за поддержку работы.

Экспериментальная часть

Использовали рестрикционные эндонуклеазы (НПО «Фермент», Вильнюс), фрагмент Кленова ДНК-полимеразы *E. coli*, полинуклеотидкиназу фага T4 (P-L Pharmacia, США), материалы для электрофореза в полиакриламидном геле (Serva, ФРГ).

Плазмидную ДНК выделяли щелочным методом [14]. Для определения первичной структуры методом Максама — Гилберта фрагменты ДНК после обработки рестрикционными эндонуклеазами, введения метки в 5'-или 3'-концевое положение и вторичной обработки рестриктазами разделяли электрофорезом в агарозном или полиакриламидном геле и подвергали стандартным химическим модификациям [7]. В части экспериментов меченый материал после разделения в агарозном геле электрофоретически переносили на DEAE-бумагу и после отмывки подвергали процедурам Максама — Гилберта в модификации Чувпило — Кравченко для твердофазных носителей [8]. Часть последовательности была определена методом Сенгера в модификации для двуспиральной плазмидной ДНК [9].

Отдельные рестрикционные фрагменты исходной плазмиды pVE9 вводили в *Sma*-участок плазмиды pUC19 после устранения лишних концов при помощи фрагмента Кленова и далее получали наборы делеционных субклонов по модифицированному методу Лина и др. [10]. Для этого ДНК субклонов обрабатывали ДНКазой I в различных концентрациях в присутствии

ионов марганца, способствующих возникновению двуцепочечных разрывов в ДНК. Число разрывов оценивали при помощи гель-электрофореза. Необходимые препараты депротенизировали, переосаждали спиртом и обрабатывали рестриктазой *Xba*I. ДНК концентрировали полиэтиленгликолем (ПЭГ-6000 добавляли до концентрации 12% в присутствии 1,6 М NaCl), осаждали спиртом и после растворения обрабатывали фрагментом Клепова (15 мин при 15° С). Далее к препарату добавляли ДНК-лигазу фага T4 и инкубировали в течение ночи при 10° С. Раствор прогревали при 68° С 10 мин и обрабатывали рестриктазой *Bam*HI. Полученный материал непосредственно использовали для трансформации компетентных клеток *E. coli* HB101. До 80% полученных трансформантов содержали делеции гена *Est*S. Положение делеций определяли на физической карте и выбрали необходимый набор субфрагментов для определения первичной структуры.

СПИСОК ЛИТЕРАТУРЫ

1. *Korochkin L. I.* // Isozymes: Current topics in biological and medical research. V. 4/ Eds Rattazzi M. C., Scandalios J. G., Whitt G. S. N. Y.: Alan R. Liss, inc., 1980. P. 159–202.
2. *Korochkin L. I.* Gene regulaton in development. Berlin – Heidelberg – New York: Springer-Verlag, 1981.
3. *Korochkin L. I., Matveeva N. M., Kuzin B. A., Karasik G. I., Maximovsky L. F.* // Biochem. Genet. 1978. V. 16. № 5. P. 709–726.
4. *Ениколопов Г. Н., Кузин Б. А., Евгеньев М. Б., Корочкин Л. И., Георгиев Г. П.* // Докл. АН СССР. 1982. Т. 267. № 6. С. 1475–1477.
5. *Yenicolopov G. N., Kuzin B. A., Evgen'ev M. B., Ludvig M. Z., Korochkin L. I., Georgiev G. P.* // EMBO J. 1983. V. 2. № 11. P. 1–7.
6. *Ениколопов Г. Н., Хечумян Р. К., Кузин Б. А., Корочкин Л. И., Георгиев Г. П.* // Генетика. 1989.
7. *Maxam A. M., Gilbert W.* // Proc. Nat. Acad. Sci. USA. 1977. V. 74. № 2. P. 560–564.
8. *Chuvpilo S. A., Kravchenko V. V.* // FEBS Lett. 1985. V. 179. № 1. P. 34–36.
9. *Sanger F., Nicklen S., Coulson A. R.* // Proc. Nat. Acad. Sci. USA. 1977. V. 74. № 12. P. 5463–5467.
10. *Lin H.-C., Lei S.-P., Wilcox G.* // Anal. Biochem. 1985. V. 147. № 1. P. 114–119.
11. *Carne T., Scheele G.* // J. Biol. Chem. 1982. V. 257. № 8. P. 4133–4140.
12. *Oakeshott J. G., Collet C., Phillis R. W., Nielsen K. M., Russell R. J., Chambers G. K., Ross V., Richmond R. C.* // Proc. Nat. Acad. Sci. USA. 1987. V. 84. № 10. P. 3359–3363.
13. *Bause E.* // Biochem. J. 1983. V. 209. № 2. P. 331–336.
14. *Dayhoff M. O.* (ed.) Atlas of protein sequence and structure. National Biochemical Research Foundation.: Washington, D. C. 1972. V. 5.

Поступило в редакцию
20.XII.1988

NUCLEOTIDE SEQUENCE OF THE *DROSOPHILA VIRILIS* ESTERASE S GENE

SERGEEV P. V., CASTILLO J. E., PEUNOVA N. I., YENIKOLOPOV G. N.

*V. A. Engelhardt Institute of Molecular Biology, Academy of Sciences
of the USSR, Moscow*

Complete nucleotide sequence of the esterase S gene of *Drosophila virilis* has been determined. The length of the gene from the transcription initiation site down to the polyadenylation signal is about 1850 bp. Conceptual translation of the DNA sequence reveals a protein 549 amino acid residues long, its presumptive active site being highly similar to active sites of the known esterases of insects and mammals.