



УДК 577.113.5

НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ УЧАСТКА ГЕНОМА
ВИРУСА КЛЕЩЕВОГО ЭНЦЕФАЛИТА,
КОДИРУЮЩЕГО СТРУКТУРНЫЕ БЕЛКИ ВИРИОНА

Шлетнев А. Г., Ямциков В. Ф., Блинов В. М. *

Новосибирский институт биоорганической химии
Сибирского отделения Академии наук СССР;* Всесоюзный научно-исследовательский институт молекулярной биологии,
п. Кольцово Новосибирской обл.

С помощью методов генетической инженерии получены рекомбинантные плазмиды, несущие ДНК-копии участков генома флавивируса — вируса клещевого энцефалита. Такой подход позволил начать исследование структуры генома вируса и обойти затруднения, связанные с работой с инфекционным вирусным материалом. Методом Максама — Гилберта исследована нуклеотидная последовательность ДНК-вставок рекомбинантных плазмид. Гибридизацией цепей ДНК-вставок с РНК вируса определена цепь, соответствующая вирусной мРНК. Нуклеотидная последовательность этой цепи ДНК-вставок переводится в аминокислотную последовательность белков только в одной рамке. На основании данных о концевой аминокислотной последовательности белков вириона ТЕV локализован участок генома, кодирующий структурные белки С, М, Е и неструктурный белок nsI. Гены структурных белков расположены единым блоком на геномной РНК ТЕV. Проведено сопоставление аминокислотных последовательностей белков и нуклеотидных последовательностей геномов ТЕV, вируса желтой лихорадки и вируса Западного Нила. Выявлена высокая гомология в структуре белков и организации геномов флавивирусов.

Вирусный энцефалит, переносимый клещами, широко распространен в природных очагах лесных зон СССР и других государств Европы и Азии. Вирус относится к семейству *Togaviridae*, к роду *Flaviviruses*, вирионы которого представляют собой нуклеокапсид, окруженный липопротеиновой оболочкой. В состав оболочки входят два белка: белок М с молекулярной массой 7,5–8 кДа и гликопротеин Е (53–60 кДа). Нуклеокапсид представляет собой комплекс белка С (14 кДа) с информационной РНК (4·10⁶ кДа) [1]. Вирус клещевого энцефалита (ТЕV) вызывает генерализованные инфекции у людей и животных, приводящие к лихорадкам и энцефалитам.

Прямые исследования структуры генома ТЕV затруднены в силу высокой патогенности вируса, инфекционности вирусной РНК и низких титров вируса при культивировании на перевиваемых клетках. Применение современных методов молекулярной биологии позволяет непосредственно приступить к таким исследованиям и обойти затруднения, связанные с работой с опасным вирусным материалом. Мы сообщили о клонировании ДНК-копий участков генома ТЕV в клетках кишечной палочки и исследовании нуклеотидных последовательностей этих копий [2, 3]. Настоящая работа подводит итог исследования структуры участка генома ТЕV, кодирующего вирусные структурные белки М, С и Е. В работе приводится сопоставление аминокислотных последовательностей структурных белков ТЕV, вируса желтой лихорадки YF [4] и вируса Западного Нила WN [5], относящихся к более отдаленной группе флавивирусов, переносимых комарами. Выявлена высокая гомология структурных белков этих вирусов.

РНК ТЕV получали из очищенных вирионов фенольной экстракцией белков и фракционировали двумя последовательными циклами центрифугирования в градиенте сахарозы. Для последующей генно-инженерной работы использовали фракцию РНК с константой седиментации 40–45S, имеющую максимальную инфекционность в перевиваемой культуре клеток почки эмбриона свиньи. По данным электрофореза в 0,8% агарозном

геле с формамидом, РНК ТЕV представляет собой гомогенную фракцию длиной 10—12 тыс. оснований. ДНК-копии участков генома ТЕV получали обратной транскрипцией и клонировали коннекторным методом с помощью плазмиды pBR322 в клетках *E. coli* (подробно см. [12]). Отбор клонов, несущих в составе рекомбинантных плазмид ДНК-копии участков генома ТЕV, осуществляли методом молекулярной гибридизации с фрагментированной [³²P]РНК ТЕV. Было получено 187 позитивных клонов. Среди них для дальнейшей работы по секвенированию были отобраны те клоны, которые имели в составе плазмиды вставки длиной 800—2000 нуклеотидных пар, выщепляемые из состава плазмидных ДНК при гидролизе эндонуклеазой *Pst*I.

Перекрестной гибридизацией ДНК-вставок нами была построена предварительная карта взаимного расположения вставок [3]. Такое картирование позволило выбрать оптимальную стратегию секвенирования ДНК-копий генома ТЕV.

Первичную структуру фрагментов ДНК определяли модифицированным методом Максама—Гилберта [6, 7]. Гибридизацией ³²P-меченых цепей ДНК-вставок с РНК ТЕV была найдена цепь, соответствующая вирусной РНК. В этой цепи нуклеотидная последовательность переводится в аминокислотную только в одной рамке. В двух других рамках этой цепи и во всех рамках комплементарной ей цепи аминокислотная последовательность, полученная из нуклеотидной, обрывается на коротких участках кодонами терминации трансляции.

На основании литературных данных об исследовании N-концевой аминокислотной последовательности белков С и Е вируса клещевого энцефалита (штамм Найдорф) [8] и данных о структуре N-концевых последовательностей белков вирионов других представителей флавивирусов [9] среди исследованных участков генома ТЕV нам удалось локализовать тот участок, который кодирует структурные белки ТЕV [2]. Правильность локализации гена гликопротеина Е на геноме ТЕV была в дальнейшем подтверждена прямыми исследованиями структуры четырех пептидов, полученных гидролизом белка Е трипсином [10]. Дальнейшие наши усилия были направлены на расшифровку участка генома, кодирующего структурные белки вируса, который покрывается вставками плазмид из клонов 2, 4 и 10. Схема секвенирования этих вставок представлена на рис. 1. Нуклеотидная последовательность и определенная из нее аминокислотная последовательность белков С, М и Е представлены соответственно на рис. 2 и 3.

Недавно были опубликованы первичные структуры генов белков С и М вируса Западного Нила [5] и полная нуклеотидная последовательность РНК вируса желтой лихорадки [4]. Оба вируса являются представителями флавивирусов, распространенных в Африке и Южной Америке; они вызывают лихорадку у людей и переносятся комарами. Эти вирусы выделяют в отдельную группу, имеющую отдаленное родство с вирусами группы клещевого энцефалита. Мы провели поиск гомологии между аминокислотными последовательностями структурных белков и геномов этих вирусов, с одной стороны, и последовательностями, расшифрованными в данной работе,— с другой (см. рис. 3).

Гены структурных белков С, М и Е у всех флавивирусов образуют единый блок и находятся вблизи 5'-конца генома. Этот блок для РНК ТЕV может транслироваться *in vitro* в полицистронный белок с молекулярной массой 90 кДа [11] и в последующем расщепляться до вирусных белков С и Е протеиназами клеточных мембран [12]. Белок С кодируется в начале блока структурных генов. Синтез его инициируется с остатка метионина (кодон АТГ, положение 49—51, рис. 2), который отщепляется в процессе созревания белка. Это подтверждается тем, что радиоактивно меченый формилметионин в бесклеточной системе Кребс-2 при синтезе белка на РНК ТЕV включается в белок С [11]. Белок С является нуклеокапсидным белком вириона, ответственным за упаковку РНК. Согласно аминокислотной последовательности, полученной в данной работе, молекулярная масса белка С составляет 10 580, а содержание отдельных аминокис-

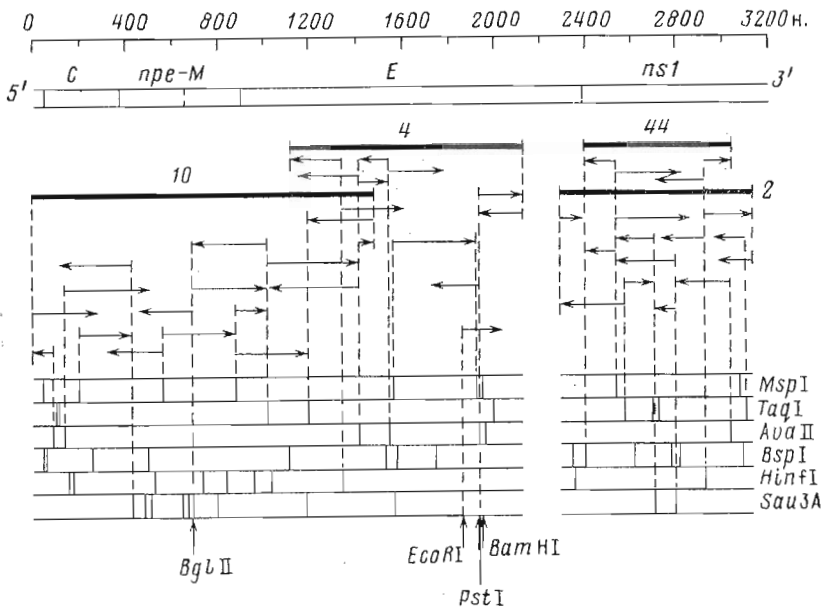


Рис. 1. Схема установления нуклеотидных последовательностей и карта расщепления эндонуклеазами рестрикции фрагментов-вставок рекомбинантных плазмид клонов 2, 4, 10. Номера клонов, содержащих рекомбинантные плазмиды, указаны над соответствующими ДНК-вставками. После гидролиза эндонуклеазой *Pst*I ДНК рекомбинантных плазмид исходные фрагменты-вставки выделяли электрофорезом в 1% агаровом геле. Субфрагменты, получающиеся при действии рестриктаз, изображены прямоугольниками. Стрелки указывают направление от кодовой 32 P-метки, а сплошная часть стрелок обозначает установленную последовательность данной цепи субфрагмента. *C*, *pre-M*, *E* и *ns1* — гены соответствующих белков в составе РНК ТЕV

лотных остатков хорошо согласуется с данными, полученными при анализе аминокислотного состава вирионного белка С [8]. Нуклеокапсидный белок обогащен остатками основных аминокислот: доля остатков лизина и аргинина составляет 25,5%. По-видимому, высокое содержание остатков основных аминокислот в белке С необходимо для электростатического взаимодействия с РНК вируса при плотной упаковке в нуклеокапсид. В аминокислотной последовательности белка С трех представителей флавивирусов присутствует гидрофобный участок (остатки 46–62, согласно нумерации, представленной на рис. 3). Функция этого участка белка С, по-видимому, связана с взаимодействием с липопротеиновой оболочкой вируса. Сопоставление аминокислотных последовательностей структурных белков флавивирусов и выявление максимальной гомологии, проведенное с помощью пакета программ [13, 14], позволило обнаружить общие закономерности строения структурных белков. Максимальная гомология для ТЕV/УЕ — 32,4%, ТЕV/УН — 26,8%, а для вирусов УF/УН она составляет 28,7%. Следует отметить высокую степень гомологии (82%) N-концов нуклеокапсидных белков С из двух штаммов ТЕV: штамма Найдорф, обнаруженного в европейском ареале [15], и штамма Софьи, выделенного из крови больного на Дальнем Востоке [16]. С-Концевая аминокислотная последовательность вирионного белка С ТЕV_N не определена. У вируса УF этот белок заканчивается последовательностью Lys-Arg-Arg, сходной с последовательностью белка С ТЕV_S в этом районе. Наличие гидролитического сайта клеточной протеиназы (Lys-Arg-Arg-Ser, положения 107–110, см. рис. 3) позволяет предположить, что именно по этой последовательности происходит процессинг белка С у ТЕV_S.

Следующим за геном белка С на геноме флавивирусов УF и УН расположен ген белка оболочки М. Белок М не обнаруживается в зараженных флавивирусами клетках, а присутствует в вирионах. Этот белок формируется через гликопротеиновый предшественник пре-М, обозначаемый в различных работах как GP23, GP19, NV2 или р36 [12, 17, 18]. В ряде случаев в составе вирионов вместо белка М обнаруживают белок пре-М [9, 19].

a

50
GTC GTG AAC GTG TTG AGA AAA GAC AGC TTA GGA GAA CAA GAG CTG GGG ATG GCC GGG AAG GCC ATT CTG AAA GGA 150
AAG GGG GGC GGT CCC CCT CGA CGA GTG TCG AAA GAG ACC GCG AAG AAG ACG CGT CAA TCT AGG GTC CAA ATG CCA
200
AAT GGA CTC GTG TTG ATG CGC ATG ATG GGG ATT CTA TGG CAC GCC GTA GCC GGA CCT GCT AGA AGT CCC GTG TTG 300
AAG TCT TTC TGG AAA TCA GTT CCA CTG AAA CAG GCC ACG GCA GCA CTT CGG AAA ATT AAG AAG GCA GTG AGC ACC
350
CTG ATG GTA GGT CTG CAA AGA CGT GGC AAA AGA AGG TCG GCA GTA GAC TGG GCA GGT TGG TTG CTG GTT GTT GTC 450
CTG TTG GGA GTG ACA CTT GCA GCC ACA GTG CGG AAG GAA AGA GAT GGC ACC
500
GAT GCG GCA ACC CAG GTG CGT GTG GAA AAT GGC ACC TGT GTG ATC CTG GCC ACC GAC ATG GGA TCA TGG TGT GAT 600
GAT TCA CTA ACC TAT GAG TGT GAG ACC ATA GAC CAG GGG GAG GAA CCG GTT GAC GTG GAT TGC TCT TGC AGG ATT
650
GTT GAT GGA GTT TAC CTG GAG TAT GGG CGG TGT GGA AAA CAA GAA GGA TCA AGA ACA AGG CCT TCA GTC CTG ATC

Рис. 2. Нуклеотидная последовательность ДНК-вставок рекомбинантных плазмид. *a* — клоны 4 и 10, *b* — клон 2. Показана структура цепи, гомологичная последовательности нуклеотидов вирусной РНК TEV. В рамку взят кодон метионина, с которого иницируется синтез полипептидной цепи белков TEV

700
 CCA TCC CAC GCT CAG GGA GAT CTC ACA GGA AGG GGA CAC AAA TGG TTA GAA GGG GAT TCA TTA CGG ACG CAC CTC 750
 800
 ACT AGA GTT GAG GGA TGG GTC TGG AAG AAT AAA GTG CTC ACC CTC GCG GTG ATC GCC GTT GTG TGG CTG ACC GTG 900
 850
 GAA AGT GTG GTG ACT CGG GTC GCC GTA GTG GTG CTC TTG TGC CTG GCT CCG GTT TAT GCC TCA CGG TGC ACA
 950
 CAT TTG GAA AAC AGG GAT TTT GTT ACT GGC ACT CAG GGG ACC ACT CGT GTG ACT CTG GTG TTG GAA CTG GGA GGA 1050
 1000
 TGC GTC ACC ATA ACA GCT GAG GGG AAG CCC TCG ATG GAT GTG TGG CTT GAC TCC ATC TAC CAG GAG AAC CCT GCC 1050
 1100
 AAG ACA CGT GAG TAC TGC CTT CAC GCA AAA CTA TCG GAT ACC AAA GTC GCG GCC AGG TGC CCA ACA ATG GGA CCT 1200
 1150
 GCC ACT TTG GCT GAA GAG CAC CAG AGC GGC ACA GTG TGT AAG AGA GAC CAG AGT GAT CGA GGC TGG GGC AAC CAT 1250
 1250
 TGT GGA TTA TTT GGA AAA GGC AGC ATT GTG ACC TGT GTC AAG GCG TCT TGT GAG GCA AAA AAG AAA GCC ACA GGA 1350
 1300
 CAC GTG TAT GAC GCT AAC AAA ATT GTG TAC ACA GTC AAA GTA GAG CCG CAC ACG GGG GAT TAC GTC GCT GCT AAT 1350

Рис. 2

1400
 GAG ACT CAC AGT GGA AGA AAA ACC GCG TCC TTC ACG GTT TCC TCG GAG AGG ACC ATC TTG ACC ATG GGA GAC TAC
 1500
 GGA GAC GTG TCC TTG TTA TGC AGA GTA GCC AGC GGT GTT GAC CTT GCT CAG ACC GTC ATC CTG GAG CTT GAC AAG
 1550
 ACC TCA GAA CAC CTA CCG ACG GCC TGG CAG GTC CAC CGG GAC TGG TTC AAT GAT CTG GCC CTA CCG TGG AAA CAT
 1600
 GAA GGG GCA CAG AAT TGG AAC AAC GCG GAA CCG CTG GTT GAG TTT GGA GCT CCA CAT GCT GTG AAA ATG GAC GTG
 1700
 TAC AAC CTT GGA GAC CAG ACT GGA GTG TTG CTC AAA TCA CTT GCT GGT GTT CCT GTG GCG CAC ATT GAT GGA ACC
 1750
 AAG TAC CAC CTG AAA AGT GGC CAC GTC ACA TGC GAG GTA GGA CTA GAA AAA CTT AAG ATG AAA GGT CTT ACA TAC
 1850
 ACA ATG TGT GAC AAG ACG AAA TTC ACG TGG AAA AGA ATT CCA ACA GAC AGT GGA CAT GAC ACA GTG GTC ATG GAA
 1900
 GTT GCG TTC TCT GGG ACC AAA CCC CCC TGC AGG ATC CCG GTG AGG GCC GTG GCA CAC GGC TCC CCG GAT GTG AAC
 2000
 GTG GCC ATG TTG ATA ACA CCC AAC CCC ACA ATC GAA AAC AAT GGC GGT GGC TTC ATA GAA ATG CAG TAA CCT CCA
 2050
 GGA GAT AAT ATC ATC TAT GTT GGG GAA CTG AGT CAC CAA TGG TTC CAA AAA GGG AGT AGC ATT GGA AGG GTT TTT

Рис. 2

d

50
TGG TCA GGC AAG GTT GGG CCT GAA CAT GAG GAA TCC GAC CAT GCC ATG AGC TTC CTT CTG GCT GGA GGA CTG GTT
100
CTG GCC ATG ACA CTC GGA GTG GGA GCT GAT GTT GGC TGT GCT GTG GAC ACT GAA CGG ATG GAG CTC CGC TGT GGT
200
GAG GGT CTG GTT GTG TGG AGA GAG GTA TCC GAA TGG TAT GAC AAT TAT GCA TAC TAC CCG GAG ACA CTA GGA GCT
250
CTT GCT TCG GCC ATA AAG GAG ACC TTC GAG GAG GGA ACT TGT GGC ATA GTG CCC CAA AAC AGA CTT GAG ATG GCC
300
ATG TGG AGG AGT TCG GCG ACA GAA CTG AAC TTG GCT TTG GTG GAG GGA GAC GCA AAT CTC ACA GTG GTG GTG GAC
350
AAA CTC GAT CCC ACA GAT TAT CGA GGT GGC ATT CCT AGC TTG CTA AAA AAG GGG AAA GAC ATA AAG GTT TCT TGG
400
AAG AGT TGG GGC CAC TCA ATG ATC TGG AGC GTC CCC GAG GCC CCC CGT CTG TTC ATG GTG GGA ACA GAG GGA AGC
500
AGT GAG TGC CCA CTA GAG AGA AGG AAA ACA GGT GTC TTC ACA GTG GCA GAG TTT GGA GTT GGC TTG AGA ACA AAA
600
GTA TTT TTG GAC TTC AGA CAG GAA TCA ACA CAC GAG TGT GAC ACA GGA GTG ATG GGA GCT GCT GTC AAG AAT GGC
700
ATG GCA GTC CAC ACA GAC CAG AGC CTC TGG ATG AAA TCC GTG AGA AAT GAC ACA GGG ACC TAC ATA GTG GAA CTT
800
CTG GTC ACT GAC CTG AGA AAC TGC TCA TGG CCG GCT AGC CAC ACT ATC GA

Рис. 2

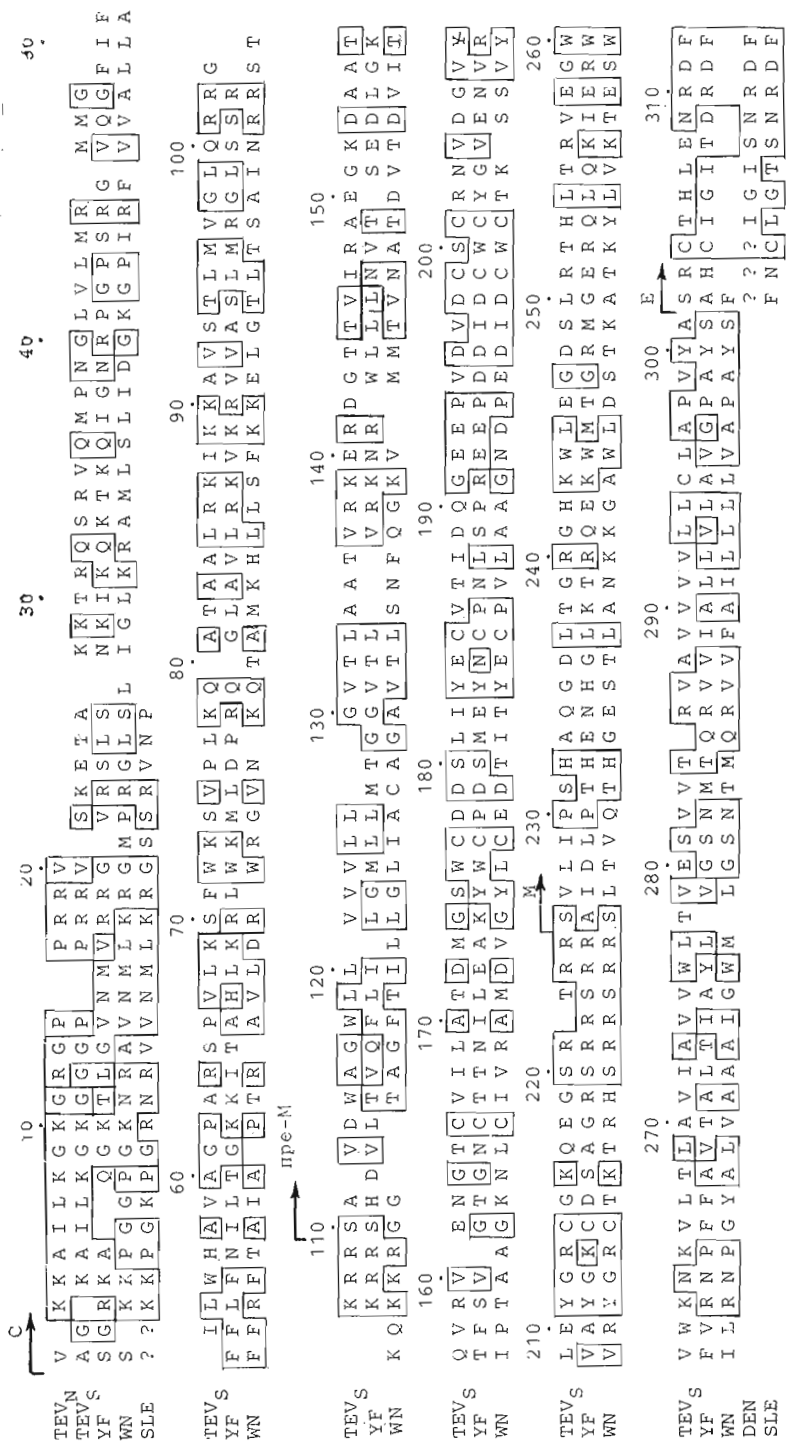


Рис. 3. Аминокислотная последовательность структурных белков С. пре-M, Е и неструктурного белка ns1 флавивируса N-Концевая последовательность белка С вируса TEV^N (штам Найдорф) приводится по работе [8], белок С SLE (вирус энцефалита Сент-Луи) — по работе [9]; N-концевые последовательности белков Е вируса SLE и DEN (вирус Деңге-2) приводятся согласно данным работы [9]. Аминокислотные последовательности белков С и пре-M вируса WN даны в соответствии с работой [5], аминокислотные последовательности белков вируса YF — по работе [4]. TEV^S — аминокислотные последовательности белков вируса клеточного энцефалита (штам Софьян), выведенные из нуклеотидных последовательностей ДНК-вставок плазмид клонов 4, 10 и 2 (согласно данным, представленным на рис. 2). Стрелками указаны начала соответствующих белков. В рамки взяты участки гомологий аминокислотных последовательностей. Названия аминокислот даны в однобуквенном коде в соответствии с рекомендацией IUPAC — IUB

| | | | | | |
|------|---|-----|-----|-----|-----|
| | 320 | 330 | 340 | 350 | 360 |
| TEVS | V T G T Q G T T R V T L V L E L G G C V T I T A E G K P S M D V W L D S I Y Q E N P A K T R E Y C L H | | | | |
| YF | I E G V H G C T W V S A T L E Q D K C V T V M A P D D K P S L D I S L E T V A I D R R P A E V R K V C Y N | | | | |
| DEN | V E G V S G G S W V D I V L E H G R ? V P T M A K N T ? L D F K V M K M E A S E ? A S | | | | |
| SLE | V E G A S G A T W I D L V L E G G S ? V T V M A P E K P ? | | | | |
| | 370 | 380 | 390 | 400 | 410 |
| TEVS | A K L S D T K V A A R C P T M G P A T L A E E H Q S G T V C K R D Q S D R G W G N H C G L F G K G S I | | | | |
| YF | A V L T H V K I N D K C P S T G E A H L A E E N E G D N A C K R T Y S D R G W G N G C G L F G K G S I | | | | |
| | 420 | 430 | 440 | 450 | 460 |
| TEVS | V T C V K A S C E A K K K A T G H V Y D A N K I V Y T V K V E P H T G D Y V A A N E T H S G R K T A | | | | |
| YF | V A C A K F T C A K S M S L F E V D Q T K I Q Y V I R A Q L H V G A K Q E N W N T D I K R L K F D | | | | |
| | 470 | 480 | 490 | 500 | 510 |
| TEVS | S F T V S S E R T I L T M G D Y G D V S L L C R V A S G V D L A Q T V I L E L D K T S E H L P T A W Q | | | | |
| YF | A L S G S Q E V E F I G Y G K A T L E C Q V Q T A V D F G N S Y I A E M E T E S | | | | |
| | 520 | 530 | 540 | 550 | 560 |
| TEVS | V H R D W F N D L A L P W K T E G A Q N W N N A E R L V E F G A P H A V K M D V Y N L G D Q T G V L L | | | | |
| YF | V D R Q W A Q D L T L P W Q S G S G V W R E M H H L V E F E P P H A T I R V L A L G N Q E G S L K | | | | |
| | 570 | 580 | 590 | 600 | 610 |
| TEVS | K S L A G V L V A H I D G T K Y H L K S G H V T C E V G L E K K M K G L T Y T M C D K T K F T | | | | |
| YF | T A L T G A M R V T K D T N D N N L Y K L H G G H V S C R V K L S A L T L K G T S Y K I C T D K M F | | | | |
| | 620 | 630 | 640 | 650 | 660 |
| TEVS | W K R I P T D S G H D T V V M E V A F S G T K P P S R I P V R A V A H G S P D V N V A M L I T P N P T | | | | |
| YF | F V K N P T D T G H G T V V M Q V K V S K G A P C R I P V I V A D D L T A A I N K G I L L V T V N P I | | | | |

FIG. 3

670
 TEVS I E N N G G G F I E M Q L P P G D N I I Y V G 690
 YF A S T N D D E V L I E V N P P F G D S Y I I V G R G D S R L L S H Q W F Q K G S S I G R V F 710 720
 730
 TEVS V E P L A V M G D T A W D F S A G G F F T S V G K G I H T V F G S A F Q G L F G G L N W I T K V I M 750 770
 YF
 780
 TEVS G A V L T W V G I N T R N M T M S D H A M S F L L A G G L V L A M T L G V G A D V G C A V D T E R M E 810 820
 YF W S G K V G P E H E E S W Y D N Y A Y Y P E T L G A L A S A I K E T F E E G T C G I V P O N R L 830 870
 830
 TEVS L R C G E G L V V W R R E V S E W Y D N Y A Y Y P E T L G A L A S A I K E T F E E G T C G I V P O N R L 860 870
 YF L K C G D G I F I F R D S D D W L N K Y S Y Y P E D P V K L A S I V K A S F E E G K C G L N S V D S L
 880
 TEVS E M A M W R S S A T D L N L A L V E G D A N L T V V V D K L D P T D Y R G G I P S L L K K G K D G 910 920
 YF E H E M W R S R A D E I N A I F E E N E V D I S V V V Q 900
 930
 TEVS I K V S W K S W G H S M I W S V P E A P R L F M V G T E G S S E C P L E R R K T G V F T V A E F G 960 970
 YF L Q Y G W K T W G K N L V F S P G R K N G S F I I D G K S R K E C P F S N R V W N S F Q I E E F G
 980
 TEVS V G L R T K V F L D F R Q E S T H E C D T C V M C A A V K N C M A V H T D Q S L W M K S V R N D 1000 1020
 YF T G V F T T R V Y M D A V F E Y T I D C D G S I L L G A A V N G K K S A H G S P T F W M G S H E V N D
 1030
 TEVS T G T Y I V E L L V T D L R N C S W P A S H T I I 1040
 YF T W M I H T L E A L D Y K E C F W P L T H T I

Такие вирионы обладают пониженной инфекционностью. Предполагается, что в таких вирионах при культивировании зараженных клеток нарушен процессинг белка пре-М. Прямое определение N-концевой аминокислотной последовательности пре-М белка осуществлено только для вируса WN. Пре-М-белок начинается последовательностью Val-Thr-Leu-, которая в районе 131—133 присутствует у белков всех трех флавивирусов. Возможно, что все они имеют общую аминокислотную последовательность на N-конце белка пре-М. Сопоставление аминокислотных последовательностей белков пре-М трех флавивирусов выявляет высокую гомологию (31—37%); особенно следует отметить, что в полипептидных цепях белков пре-М положения остатков Cys полностью совпадают. Поскольку остатки Cys в белках, как правило, формируют дисульфидные мостики, можно предположить, что белки пре-М флавивирусов имеют сходное строение на более высоких уровнях организации. Вблизи N- и C-концов белков пре-М расположены кластеры незаряженных аминокислотных остатков. Белок М формируется на клеточных мембранах из гликозилированного предшественника, гидрофобные сигналы на концах полипептидной цепи белка пре-М могут выступать как «якоря», удерживающие предшественник на мембране. Потенциальным местом гликозилирования белка пре-М TEV является участок Asn-Gly-Thr (положения 163—165, рис. 3). Процессинг белков пре-М флавивирусов YF и WN осуществляет связанная с мембраной клеточная протеиназа по последовательности Arg-Arg-Ser или Arg-Arg-Ala. Такая же последовательность (Arg-Arg-Ser, остатки 224—226, рис. 3) присутствует в полипептиде пре-М TEV; возможно, этот участок служит местом гидролиза белка пре-М при созревании белка М. Однако последовательность белка М TEV (штамм Найдорф) начинается с Lys-Arg-Arg-Thr [8]. Этой последовательности мы не обнаружили. Причиной могут быть штаммовые различия. Аминокислотная последовательность от Ser-226 до остатка Ser-302, соответствующая полипептиду М у флавивирусов YF и WN и имеющая высокую гомологию с ними, может соответствовать белку М у TEV (молекулярная масса 8344 и процентное содержание аминокислотных остатков у этого полипептида хорошо коррелируют с соответствующими параметрами для белка М [8]). В работе Хайнца и Кунца [8], а также в нашей предыдущей работе [10] для N-конца белка Е была определена последовательность Ser-Val-Leu-Ile-Pro-, расположенная в полипептидной цепи после сайта процессинга Arg-Arg-Ser (остатки 224—226). Возможно, гликопротеин Е TEV формируется через предшественник М-Е, процессинг которого не всегда осуществляется при выращивании вируса на культуре клеток. В пользу этого предположения свидетельствуют данные о гетерогенности белка Е по молекулярной массе. По одним данным эта величина составляет 51—53 кДа [11, 20], по другим — 56—60 кДа [1, 15, 21]. Альтернативно этому предположению можно говорить и о том, что место процессинга между белками М и Е было утрачено у флавивирусов, переносимых клещами (TEV), в отличие от вирусов, переносимых комарами (YF, WN, SLE, DEN). Вывод о правильности одного из этих предположений требует дополнительных экспериментальных данных.

Белок Е является гликопротеином и основным белком вириона, к которому в организме вырабатываются антитела. В исследованной нами структуре белка Е TEV обнаружено два потенциальных места гликозилирования типа Asn-Glu-Thr (положения 455—457) и Asn-Pro-Thr (положения 667—669). Гомология белков Е вирусов YF и TEV составляет 35%. Максимальная непрерывная гомологичная последовательность, равная 10 аминокислотным остаткам, расположена вблизи N-концов (положения 406—415). Остатки Cys в белке Е TEV полностью соответствуют по положениям аналогичным остаткам в белке Е вируса YF. Это еще раз указывает на идентичность организации структур этих белков и на возможную общность их происхождения.

В полипептидной цепи белка Е TEV нам не удалось определить аминокислотную последовательность 716—778, так как в библиотеке клонов мы не обнаружили плазмидную вставку, перекрывающую последовательность между вставками клонов 4 и 2.

Белок E вируса YF заканчивается кластером незаряженных аминокислотных остатков с концевой последовательностью Leu-Gly-Val-Gly-Ala (положения 806—810). Эта последовательность, характерная также и для полипептида E TEV, по-видимому, заканчивает белок E. Гидрофобные сигналы вблизи концов полипептидной цепи белка E могут удерживать его на цитоплазматических мембранах клетки, где происходит сборка вируса.

Несомненно, в структуре белка E наиболее интересны иммунологические аспекты, прежде всего поиск поверхностных антигенных детерминант. Поскольку вирусы TEV и YF имеют отдаленное родство и слабый иммунологический перекрест, логично было связать локализацию поверхностных антигенных детерминант с участками белков E, имеющих максимальные различия. Эти участки не должны быть обогащены остатками Cys и должны содержать остатки заряженных аминокислот. Мы провели поиск таких участков, исходя из этих критериев и учитывая данные по частоте встречаемости аминокислотных остатков в поверхностных антигенных детерминантах других белков [22]. Были выявлены два участка в цепи белка E, которые могли бы выполнять роль поверхностных антигенных детерминант: первый — от 362-го до 371-го остатка, второй — от 424-го до 431-го.

Вставка в плазмиде клона 2, структура которого представлена на рис. 2б, покрывает участок генома TEV, кодирующий первый неструктурный белок ns1 или NV3 [4]. В ходе развития флавивирусов в зараженных клетках кроме уже упомянутого белка пре-M обнаруживают еще 12 неструктурных белков [4, 12, 18, 21, 23]. Большинство из них связано с процессами репликации РНК и процессингом белков. Функция белка ns1 неясна. Известно, что добавление к зараженному вирусом YF клеткам моноклональных антител к белку ns1 приводит к комплементзависимому лизису этих клеток [4]. Белок ns1 TEV является гликопротеином. В определенной нами N-концевой аминокислотной последовательности белка ns1 выявляются два потенциальных места гликозилирования — Asn-Asp-Thr (положения 1025—1027) и Asn-Cys-Ser (положения 1041—1043). Гомология N-концевой части белков ns1 TEV и YF несколько выше, чем для структурных белков этих вирусов, и составляет 37,9%.

Выявленная нами гомология флавивирусов на уровне аминокислотных последовательностей их белков может свидетельствовать об общности их происхождения. Она еще более наглядна при сравнении нуклеотидных последовательностей геномов. С учетом делеций и вставок при сравнении расшифрованных участков геномов TEV и YF (гомология составляет 45%, причем в геномах обоих вирусов имеется несколько следующих друг за другом гомологичных блоков длиной 10—12 нуклеотидов (например, в составе участков 1184—1196, 1208—1230). 40% нуклеотидных замен между TEV и YF приходится на третье положение в кодонах, что не изменяет аминокислотных остатков в белках. Для достижения максимальной гомологии между вирусными геномами TEV и YF введено 12 делеций длиной от 3 до 12 нуклеотидных звеньев (всего 66 нуклеотидных остатков) и 13 вставок (общая длина — 69 нуклеотидов), при этом общая длина полипептидной цепи структурных белков одинакова. Сохранение общей длины белка мы наблюдаем и при сравнении областей геномов, кодирующих неструктурные белки. Подобная картина общности геномов наблюдается при сравнении геномов TEV с WN и WN с YF (гомология TEV/WN составляет 39%, а для WN/YF — 40,3%).

Итак, в организации флавивирусов выявляется консервативность как на нуклеотидном, так и на аминокислотном уровнях. Геномы имеют общую последовательность генов на РНК и, по-видимому, общую стратегию экспрессии генов. Следует отметить, что по сравнению с представителями рода *Alphaviruses* [24, 25], также относящимися к семейству *Togaviridae*, структурная организация геномов *Flaviviruses* принципиально иная: во-первых, флавивирусы не имеют субгеномной РНК; во-вторых, гены структурных белков на геноме предшествуют генам неструктурных белков (у альфа-вирусов наоборот); в-третьих, при экспрессии геномов флавиви-

русов в полицистронный белок области структурных и неструктурных генов не разобцены кодонами терминации трансляции. В этом отношении флавивирусы ближе по организации и экспрессии геномов к семейству пикорновирусов.

Экспериментальная часть

В работе использованы гидразингидрат, диметилсульфат, пиперидин, формамид (Merck, ФРГ); ампициллин, тетрациклин, акриламид, агароза, сахароза (Sigma, США). Гибридизацию нуклеиновых кислот осуществляли с использованием фильтров BA-85 (Schleicher und Schüll, ФРГ). [α - 32 P]dNTP (более 3000 Ки/ммоль) и [γ - 32 P]ATP — препараты фирмы «Изотоп» (СССР), эндонуклеазы рестрикции *Bgl*II, *Bsp*I, *Bam*HI, *Pst*I, *Eco*RI — производства НИКИ БАН (Бердск). Рестриктазы *Ava*II, *Msp*I, *Nco*I, *Taq*I, *Sau*3A любезно предоставлены С. Х. Дегтяревым (ВНИИ МБ, пос. Кольцово Новосибирской обл.), *Hin*I — И. М. Чумаковым (ИМБ АН СССР), ДНК-полимераза *E. coli* и ДНК-полимераза I (фрагмент Кленова) — В. В. Потаповым (ИЦиГ СО АН СССР). Вирус клещевого энцефалита, штамм Софийн, был выделен и очищен по методу [26] С. Г. Рубиным (ИГиВЭ АМН СССР). РНК TEV выделяли методом фенольной депротеинизации и очищали центрифугированием в градиенте сахарозы (5–20%). Получение ДНК-копий участков генома TEV, их клонирование в *E. coli*, отбор рекомбинантных плазмид были опубликованы ранее [2]. Выращивание клеток *E. coli* K802, несущих рекомбинантные плазмиды, осуществляли в LB-среде, содержащей тетрациклин (20 мкг/мл). ДНК плазмид выделяли по методу [27] с модификациями. Полисахариды из препаратов плазмид удаляли двукратным осаждением из 2,5 М ацетата натрия, pH 7,0. ДНК осаждали 2,5 объема этанола, обрабатывали РНКазой А (20 мкг/мл) 1 ч при 37°С в буфере (10 мМ трис-HCl, pH 8,0; 1 мМ EDTA) и фильтровали на колонке с сефарозой С1-4В.

Обработку ДНК плазмид эндонуклеазами рестрикции, разделение продуктов гидролиза ДНК в ПААГ, их выделение и введение 32 P-метки с помощью ДНК-полимеразы I (фрагмент Кленова) проводили как описано в работе [28]. Радиоактивно меченные комплементарные цепи фрагментов после тепловой денатурации в 75% формамиде разделяли электрофорезом в ПААГ. Определение нуклеотидной последовательности фрагментов ДНК проводили по методу Максама — Гилберта [6] в модификации [7].

Введение статистической 32 P-метки во фрагменты ДНК (цик-трансляция) осуществляли с помощью ДНК-полимеразы I в смеси, содержащей 20 мМ трис-HCl (pH 7,5), 10 мМ MgCl₂, 10 мМ β -меркаптоэтанол, 2 ммоль фрагмента ДНК, 0,5 мМ dGTP, dATP, dCTP, 1 мМ [α - 32 P]TTP и 2 сд. акт. фермента. Реакцию проводили 1 ч при 37°С, фрагменты осаждали спиртом. Перед гибридизацией [32 P]ДНК-фрагменты денатурировали прогревом в течение 5 мин при 90°С в 50% формамиде. Молекулярную гибридизацию проводили по методу [29] 12–18 ч при 68°С. Концентрация радиоактивного зонда в смеси составляла 50 нг/мл (уд. акт. зонда 0,1–0,5 мКи/мкг).

ЛИТЕРАТУРА

1. Westaway E. G., Schlesienger R. W., Dalrymple J. M., Trent D. W. Intervirology, 1980, v. 14, № 1, p. 114–117.
2. Чумаков М. П., Кусов Ю. Ю., Рубин С. Г., Сальников Я. А., Семашко Н. В., Георгиев Г. П., Чумаков П. М., Грачев М. А., Шаманин В. А., Плетнев А. Г. Биоорганическая химия, 1983, т. 9, № 2, с. 276–279.
3. Плетнев А. Г., Ямщиков В. Ф. Биоорганическая химия, 1985, т. 11, № 12, с. 1681–1684.
4. Rice C. M., Lenches E. M., Eddy S. R., Shin S. J., Sheets R. L., Strauss J. H. Science, 1985, v. 229, № 4715, p. 726–733.
5. Castle E., Nowak T., Leidner U., Wengler G. Virology, 1985, v. 145, № 2, p. 227–236.
6. Mazam A., Gilbert W. Proc. Nat. Acad. Sci. USA, 1977, v. 74, № 2, p. 560–564.
7. Чувицко С. А., Кравченко В. В. Биоорганическая химия, 1983, т. 9, № 12, с. 1634–1637.
8. Voegelé U., Heinz F. X., Wengler G., Kunz C. Virology, 1983, v. 126, № 2, p. 651–657.
9. Bell J. R., Kinney R. M., Trent D. W., Lenches E. M., Dalgarno L., Strauss J. H. Virology, 1985, v. 143, № 1, p. 224–229.
10. Барам Г. И., Грачев М. А., Назимов И. В., Плетнев А. Г., Прессман Е. К., Рубин С. Г., Сальников Я. А., Семашко Н. В., Чумаков М. П., Шемякин В. В., Ямщиков В. Ф. Биоорганическая химия, 1985, т. 11, № 12, с. 1677–1680.
11. Svitkin Y. V., Ugarova T. Y., Chernovskaya T. V., Lyapustin V. N., Lashkevich V. A., Agol V. I. Virology, 1981, v. 110, № 1, p. 26–34.
12. Svitkin Y. V., Lyapustin V. N., Lashkevich V. A., Agol V. I. Virology, 1984, v. 135, № 2, p. 536–541.
13. McLachlan A. D. J. Mol. Biol., 1971, v. 61, № 2, p. 409–424.
14. Stalen R. Nucl. Acids Res., 1980, v. 8, № 10, p. 3673–3694.
15. Heinz F. X., Kunz C. J. gen. Virol., 1981, v. 57, № 1, p. 263–274.
16. RSSE-TBE. Prototype strain.— In: International Catalogue of Arboviruses. Sofjin, 1975, p. 607.
17. Shope R. E. In: The Togaviruses/Ed. Schlesinger R. W. N. Y.: Acad. Press, 1980, p. 47–82.
18. Westaway E. G., Speight G., Endo L. Virus Research., 1984, v. 1, № 4, p. 333–350.

19. Shapiro D., Brandt W. E., Russell P. K. *Virology*, 1972, v. 50, № 3, p. 906–911.
20. Heinz F. X., Berger R., Tuma W., Kunz C. *Virology*, 1983, v. 130, № 2, p. 485–501.
21. Russell P. K., Brandt W. E., Dalrymple J. M. In: *The Togaviruses*/Ed. Schlesinger R. W. N. Y.: Acad. Press, 1980, p. 503–529.
22. Welling G. W., Weijer W. J., Van der Zee R., Welling-Wester S. *FEBS Lett.*, 1985, v. 188, № 2, p. 215–218.
23. Westaway E. G. In: *The Togaviruses*/Ed. Schlesinger R. W. N. Y.: Acad. Press, 1980, p. 531–581.
24. Dalgarno L., Rice C. M., Strauss J. H. *Virology*, 1983, v. 129, № 1, p. 170–187.
25. Strauss E. G., Rice C. M., Strauss J. H. *Virology*, 1984, v. 133, № 1, p. 92–110.
26. Чумаков М. И., Кусов Ю. Ю., Рубин С. Г., Семашко П. В., Сальников Я. А., Рейнгольд В. Н., Прессман Е. К., Цехановская Н. А. *Вопр. вирусологии*, 1984, т. 29, № 6, с. 694–701.
27. Birnboim H. C., Doly J. *Nucl. Acids Res.*, 1979, v. 7, № 6, p. 1513–1523.
28. Грачев М. А., Плетнев А. Г. *Биоорг. химия*, 1984, т. 10, № 6, с. 824–843.
29. Southern E. M. *J. Mol. Biol.*, 1975, v. 98, № 3, p. 503–517.

Поступила в редакцию
6.II.1986

NUCLEOTIDE SEQUENCE OF THE GENOME REGION OF THE TICK-BORNE ENCEPHALITIS VIRUS CODING FOR STRUCTURAL PROTEINS OF VIRION

PLETNEV A. G., YAMSHCHIKOV V. F., BLINOV V. M. *

*Novosibirsk Institute of Bioorganic Chemistry, Siberian Branch of the Academy
of Sciences of the USSR, Novosibirsk; *All-Union Research Institute
of Molecular Biology, Kol'tsovo, Novosibirsk Region*

RNA of a flavivirus – tick-borne encephalitis virus (strain Sofjin) was subjected to reverse transcription and the DNA copy was transformed into double-stranded DNA by action of *E. coli* DNA-polymerase I (Klenow's fragment). This DNA was annealed with pBR322 plasmid. The recombinant plasmids were cloned in *E. coli* K802. The nucleotide sequence of the inserts of the clones coding for region of structural proteins C, pre-M, E and nonstructural protein ns1 was determined by the Maxam – Gilbert method. The nucleotide sequence of these regions is translatable into an amino acid sequence of proteins without interruption. The amino acid sequences of proteins and nucleotide sequence of genome of the tick-born encephalitis virus are extensively homologous to that found for the flaviviruses Yellow Fever and West Nile.