



УДК 577.113.5

НУКЛЕОТИДНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ГЕНОМА И ПОЛНАЯ
АМИНОКИСЛОТНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ПОЛИПРОТЕИНА
ВИРУСА КЛЕЩЕВОГО ЭНЦЕФАЛИТА

Плетнев А. Г., Ямщиков В. Ф., Блинов В. М.*

Новосибирский институт биоорганической химии СО АН СССР, Новосибирск;

* Всесоюзный научно-исследовательский институт молекулярной биологии,
пос. Кольцово Новосибирской обл.

Клонирована и установлена полная структура РНК вируса клещевого энцефалита (ТВЕ), кодирующая вирусные и неструктурные белки вируса. Длина исследованного генома составляет 10 477 оснований. Нуклеотидная последовательность содержит единственную открытую рамку считывания белка (127–10 363 н.о.), кодирующую 3412 аминокислотных остатков. 5'- и 3'-Нетранслируемые концы РНК представлены шпильчато-петлевыми структурами. Протеолиз полипротеина-предшественника протекает, по-видимому, по механизму, предложенному для полипротеинов вирусов желтой лихорадки (YF), Западного Нила (WN), Капжи (Kun). Определенный нами порядок генов на геноме вируса клещевого энцефалита следующий: 5' С-преМ (М)-Е-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS53'. Геномы и полипротеины флавивирусов и вируса клещевого энцефалита сходны по строению, хотя эти вирусы передаются в природных очагах от хозяина к хозяину различными переносчиками — комарами и клещами. Анализ гомологий последовательностей полипротеинов флавивирусов показывает, что вирус клещевого энцефалита более близок к вирусу желтой лихорадки, чем к другим представителям серологических групп флавивирусов (к группе вирусов Западного Нила или Денге). Профили гидрофобности полипротеинов флавивирусов высококонсервативны. Неструктурные белки NS2A, NS2B, NS4A, NS4B крайне гидрофобны. Предполагается, что эти белки, вероятно, связаны с клеточными мембранами. Белки Е, NS1, NS3, NS5 наиболее консервативны и могут быть вовлечены в выполнение общих основных ферментативных процессов при репликации РНК или сборке вирионов.

Семейство *Flaviviridae* насчитывает более 60 представителей, среди которых по реакции нейтрализации выделяют несколько групп. Подгруппа клещевого энцефалита объединяет серологически родственные вирусы, переносимые клещами: Лангат, Негиши, шотландского энцефаломиелита овец, Омской геморрагической лихорадки, Киасаурской лесной болезни и клещевого энцефалита (ТВЕ). Среди флавивирусов наибольшее эпидемиологическое значение имеют вирусы желтой лихорадки (YF), Денге (Den), японского энцефалита (JE), а для СССР — вирус клещевого энцефалита. Вирионы флавивирусов имеют сферическую форму диаметром 40–50 нм и включают в себя нуклеокапсид, окруженный липопротеиновой оболочкой. Геном флавивирусов представлен одноцепочечной РНК положительной полярности длиной около 11 000 оснований, кодирующий единый белок-предшественник, из которого формируются вирусные белки и белки, необходимые для развития вируса в клетках [1, 2].

Весенне-летний клещевой энцефалит, возбудителем которого является ТВЕ, остается серьезной проблемой здравоохранения страны, особенно в регионах Сибири и Дальнего Востока. В природных очагах этих районов распространены высокопатогенные штаммы вируса. Несмотря на ежегодную систематическую вакцинацию населения, уровень заболеваемости не снижается. Решение задачи практического здравоохранения требует разработки более эффективных средств профилактики, диагностики и лече-

Принятые сокращения: ТВЕ — вирус клещевого энцефалита (штамм Софбин), YF — вирус желтой лихорадки, Den — вирус Денге, JE — вирус японского энцефалита, SLE — вирус энцефалита Сент-Луис, MVE — вирус долины Мюррей, ТВЕ_n — вирус клещевого энцефалита (штамм Найдорф).

ния заболевания «клещевой энцефалит». Успешное решение поставленных проблем требует понимания молекулярных основ строения и механизмов развития вируса в клетках.

Ранее [3–5] мы сообщали о клонировании участков генома ТВЕ в бактериальных клетках и исследовании нуклеотидной последовательности генов, кодирующих белки вириона. Настоящая работа подводит итог исследований структуры кодирующей зоны генома ТВЕ. Впервые определена структура нетранслируемых зон генома и первичная структура вирусного полипротеина; на основании данных по исследованию N-концевых аминокислотных последовательностей белков вирусов YF, SLE, Kun [2, 6, 7] и высокого сходства мест процессинга полипротеинов флавивирусов определена последовательность генов на геноме ТВЕ и аминокислотная последовательность всех белков ТВЕ. Показано, что ТВЕ — типичный представитель семейства *Flaviviridae*.

В качестве объекта исследований нами был взят штамм Софьи ТВЕ, выделенный на Дальнем Востоке от больного в 1937 г. и прошедший несколько десятков пассажей на лабораторных животных и клетках [8]. В библиотеке клонов, полученной нами ранее [3], ДНК рекомбинантных плазмид содержали вставки, покрывающие 75% генома ТВЕ. Нуклеотидная последовательность генома ТВЕ, кодирующая структурные белки вириона — С, М, Е и неструктурный белок NS1, опубликована в работах [3–5, 9]. Для заполнения брешей в структуре генома ТВЕ мы предприняли клонирование ДНК-копий генома ТВЕ, полученных при помощи олигонуклеотидов-затравок в реакции, катализируемой обратной транскриптазой. Такой подход позволил нам получить и исследовать структуру генома ТВЕ на участках, кодирующих неструктурные белки и 3'-концевую нетранслируемую зону (см. рис. 1 и «Экспериментальную часть»). Строение 5'-концевой нетранслируемой зоны, представленной 127 нуклеотидами, определяли прямым секвенированием кДНК, синтезированной при помощи праймера d(CAGACCTACCATCAGGGT), комплементарного последовательности РНК ТВЕ на участке 376–393 (рис. 2). 5'-Концевой участок РНК ТВЕ отличается от всех исследованных 5'-концевых структур геномов флавивирусов по длине и имеет частичное сходство по первичной структуре, однако он не обладает характерной консервативной последовательностью, наблюдаемой у представителей семейства флавивирусов, переносимых комарами [2, 10, 11]. В то же время возможная шпильочно-петлевая структура 5'-концевого участка РНК ТВЕ имеет сходную конфигурацию, характерную для этого участка РНК всех флавивирусов.

ДНК-копия, покрывающая 3'-концевую зону генома, была получена при использовании в качестве затравки смеси олигонуклеотидов — d(AGT(G/T) (G/T)T(G/T)TATGTGT), комплементарных 3'-концевым участкам РНК флавивирусов YF, WN, JE [2, 12, 13]. При клонировании кДНК получили три идентичных рекомбинантных плазмидных ДНК (p9), вставки которых включают участок терминации трансляции полипротеина (кодон UAA, положение 10 363–10 366, рис. 2) и 115 нуклеотидных остатков нетранслируемой зоны. Для более детального исследования 3'-концевой части генома ТВЕ дополнительно мы предприняли несколько вариантов клонирования этого участка: 1) использовали вирусную РНК с синтезированным на 3'-конце poly(A)-трактом для получения кДНК с затравки (dT)_{15–18}; вторую цепь кДНК получали достройкой цепи *HindIII*-*SalGI*-фрагмента из ДНК плазмиды p23 (цепь фрагмента идентична по последовательности, приведенной на рис. 2, в положениях 9 923–10 009); 2) при синтезе кДНК в качестве затравки использовали цепь *HindIII*-*SalGI*-фрагмента ДНК плазмиды p23, комплементарную минус-цепи РНК ТВЕ; с этой целью выделяли репликативную форму РНК ТВЕ из мозга мыши, зараженной ТВЕ (штамм Софьи); на кДНК достраивали poly(A)-тракт и вторую цепь синтезировали, используя олиготимидилатную затравку. В первом случае получено 6 рекомбинантных плазмид, во втором — 2. Эти плазмидные ДНК по структуре вставок были аналогичны ДНК плазмиды p9, т. е. они не удлиняли нетранслируемую зону генома ТВЕ.

РНК флавивирусов обладают высокой структурированностью концевых

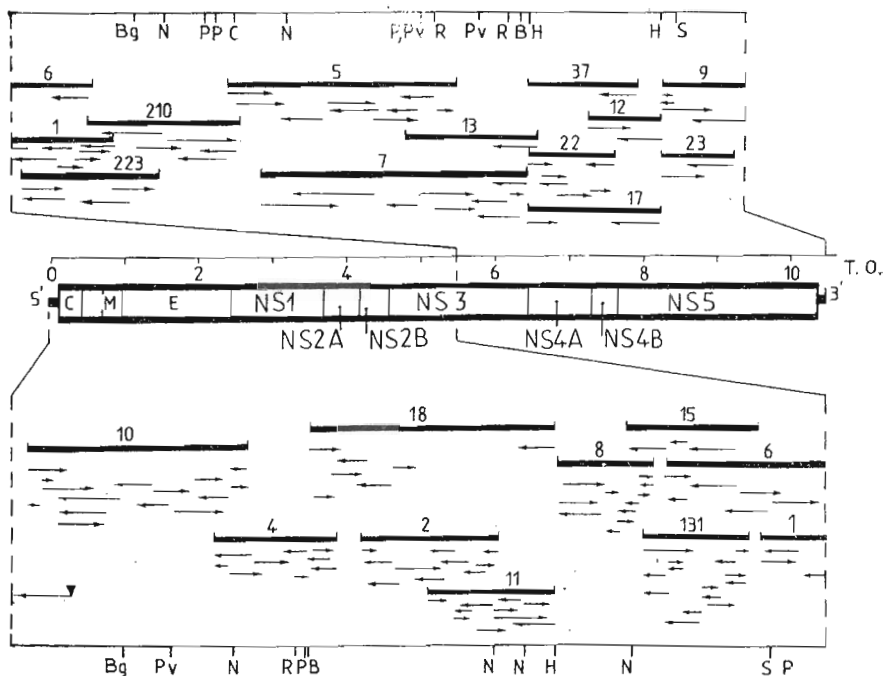


Рис. 1. Карта расположения клонированных ДНК-копий (указаны номера соответствующих плазмид) и генов вдоль генома ТВЕ. Представлена стратегия определения нуклеотидной последовательности генома. ДНК-вставки рекомбинантных плазмид выделены жирными линиями и отмечены цифрой. Стрелками указано направление и длина секвенированных последовательностей субфрагмента. Знаком (▼) обозначено положение олигонуклеотидной затравки, с помощью которой была получена кДНК и исследована структура 5'-концевой части генома. С, М, Е, NS1 – NS5 – гены соответствующих белков ТВЕ. В нижней и верхней частях рисунка указаны места гидролиза ДНК-вставок эндонуклеазами рестрикции: *Bgl*II (Bg), *Pvu*II (Pv), *Nco*I (N), *Eco*RI (R), *Pst*I (P), *Bam*HI (B), *Cla*I (C), *Hind*III (H)

участков [2, 10–16]. Предполагается, что эти концевые шпильчно-петлевые структуры играют существенную роль в формировании комплексов РНК с белковыми компонентами, например в процессе трансляции РНК рибосомами, при формировании иницирующих комплексов в репликации РНК или при упаковке РНК в нуклеокапсид. РНК ТВЕ также высокоструктурирована в 3'-концевой части и особенно в зоне, кодирующей С-концевую часть белка NS5. Одна из предполагаемых вторичных структур может быть сформирована за счет инвертированных повторов на участке 9 251–10 204 н.о. в последовательности, представленной на рис. 2. Образование шпильчно-петлевой структуры РНК ТВЕ на этом участке находит подтверждение в экспериментах по синтезу кДНК на вирусной РНК. Наблюдаются «остановки» синтеза кДНК перед структурированными участками и накопление продуктов кДНК соответствующей длины при разделении их в ПААГ. В библиотеке клонов, в двух ДНК рекомбинантных плазмид нами найдены вставки, не содержащие последовательностей, вовлеченных в образование шпилек (последовательности 9 251–10 204 н.о. или 9 531–10 052 н.о., рис. 2), т. е. копирование РНК происходит с «про скакиванием» структурированного участка. Подобные участки нами обнаружены на геноме ТВЕ перед геном NS1 и перед началом гена С.

Полная нуклеотидная последовательность РНК ТВЕ и выведенная из нее аминокислотная последовательность полипротеина представлены на рис. 2. Геном ТВЕ имеет длину 10 477 оснований, из которых 127 нуклеотидных остатков предшествуют иницирующему кодону AUG; затем следует открытая рамка считывания белка, представленная 10 236 основаниями и кодирующая 3 412 аминокислотных остатков полипротеина, и 115 оснований 3'-нетранслируемой зоны генома. Геном имеет следующий нуклеотидный состав (%): А – 24,8; С – 22,6; G – 31,5; U – 21,1. Наименее

встречаемый динуклеотид в РНК ТВЕ — UA. Частота использования кодонов существенно не отличается от таковой для других представителей семейства флавивирусов [2, 17] и не составляет среднестатистическую величину, поскольку наименее используемыми являются кодоны типа XUA для Leu (CUA), Val (GUA), Ile (AUA).

Положение генов белков ТВЕ на геноме основано на сопоставлении данных прямого анализа N-концевых последовательностей белков флавивирусов со структурой полипротеина ТВЕ (табл. 1). Наличие значительного сходства в участках расщепления полипротеинов флавивирусов позволяет считать, что в РНК ТВЕ гены белков расположены в следующем порядке: С-преМ(М)-Е-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5. Правильность локализации генов белков М и Е ТВЕ подтверждается прямым исследованием структуры триптических пептидов этих белков [24], а также данными по трансляции вирусной РНК в бесклеточных системах синтеза белка [25, 26].

Нуклеокапсидный белок С кодируется в N-концевой части полипротеина и состоит из 111 остатков. Молекулярная масса (М) белка С равна 12 108 Да, он обогащен остатками основных аминокислот: доля остатков Arg и Lys составляет 21,4%, по-видимому, необходимых для нейтрализации зарядов на РНК при ее упаковке в нуклеокапсид. Иницирующий остаток Met удаляется клеточной аминопептидазой [2]. С-концевая часть белка, представленная 12 гидрофобными аминокислотными остатками, вероятно, формирует трансмембранный лидер, необходимый для правильного гидролиза клеточной сигнализацией полипротеина между белками С и преМ.

Следующим от N-конца полипротеина выщелачиваемым белком является гликопротеин преМ (или NV2) — полипептид длиной в 168 аминокислотных остатков (М 18 385 Да). Единственный потенциальный сайт гликозилирования Asn-Gly-Thr расположен в 32 остатках от N-конца. По-видимому, сахарный остов имеет N-тип связи с остатком Asn, поскольку подобное гликозилирование белка преМ вируса *Kip* блокируется туникамицином при внутриклеточном развитии вируса [27, 28]. ПреМ-белок, в свою очередь, является предшественником вирионного белка М, который не удается обнаружить в зараженных флавивирусами клетках. Считается, что процессинг белка преМ осуществляет клеточная протеиназа со специфичностью к последовательностям типа Arg-X-Arg(Lys)-Arg↓, найденным также в местах процессинга полипротеинов у вирусов ряда других семейств [2].

Белок оболочки вируса М представлен 75 остатками аминокислот (М 8 209 Да). В С-концевой части белка находятся два гидрофобных участка, разделенных остатком Arg. Подобные последовательности имеются в белках М всех представителей флавивирусов, переносимых комарами. По-видимому, они обеспечивают этим белкам трансмембранную локализацию и важны для посттрансляционной модификации [2, 14, 17, 20, 29, 30].

Имеются веские основания считать, что процессинг полипротеина между белком Е и пеструктурным белком NS1 осуществляет связанная с мембранами аппарата Гольджи протеиназа [2, 16, 19]. Структура сайта протеолиза между этими белками высококонсервативна у флавивирусов (табл. 1). Полипептидная цепь белка Е, включающая 496 аминокислотных остатков (М 53 680 Да), заканчивается двумя гидрофобными участками, вероятно, служащими трансмембранным «якорем» белка. Белок Е слабоосновный, так как на 47 остатков Asp и Glu приходится 50 остатков основных аминокислот. С учетом вклада гликозилирования (2% от массы белка) мол. масса гликопротеина Е составляет 55 000 Да, что хорошо согласуется с данными по определению белка М электрофорезом в ПААГ [33]. В цепи гликопротеина Е присутствуют три потенциальных места гликозилирования (положения 434-436, 641-643 и 753-755 а.о., рис. 2). В вирионном белке Е гликозилирование происходит только по первому месту (434-436), так как при исчерпывающем гидролизе белка цианбромидом гликозилированным оказывается только пептид (358-456 а.о., М 14,5 Да), способный окрашиваться конканавалином и связываться с моноклональными антителами, полученными к вирусному белку Е. Два

CCUGCAUGCGUUGCUCGGAUAGCAACAGCAGCGACAGGUUUGAGAGAGAGACAAUCUU	66
UCGCUUGAUCAGUCGUGAACGUGUUGAGAAAAAGACAGCUUAGGAGAACAAGAGCUGGGG	126
M $\xrightarrow{C}$ A G K A I L K G K G G G P P R R V S K	20
AUGGCCGGGAAGGCCAUUCUGAAAGAAAGGGGGCGGUCGCCCCUCGACGAGUGCGAAA	186
E T A K K T R Q S R V Q M P N G L V L M	40
GAGACCGGAAGAAGACGCGUCAAUCUAGGGUCCAAUGCCAAUUGGACUCGUGUUGAUG	246
R M M G I L W H A V A G T A R S P V L K	60
CGCAUGAUGGGGAUUCUUAUGGCACGCCGUAGCCGGCACUGCUAGAAGUCCCGUGUUGAAG	306
S F W K S V P L K Q A T A A L R K I K K	80
UCUUUCUGGAAAUACAGUUCACUGAAACAGGCCACGGCAGCACUUCGGAAAAUUAAGAAG	366
A V S T L M V G L Q R R G K R R S A V D	100
GCAGUGAGCACCCUGAUGGUAGGUCUGCAAAGACUGGGCAAAGAAGGUCGGCAGUAGAC	426
W T G W L L V V V L L G V T L A A T V R	120
UGGACAGGUUGGUGUGUGUUGUUGUCCUGUUGGGAGUGACACUUGCAGCCACAGUGCGG	486
K E R D G T T V I R A E G K D A A T Q V	140
AAGGAAAGAGAUGGCACCACCGUGAUCACAGCUGAAGGAAAAGAUGCGGCAACCCAGGUG	546
R V E N G T C V I L A T D M G S W C D D	160
CGUGUGGAAAAUUGGCACCUGUGUGAUCCUGGCCACGGACAUGGGAUCAUGGUGUGAUGAU	606
S L T Y E C V T I D Q G E E P V D V D C	180
UCACUAAACCUAUGAGUGUGUGACCAUAGACCAGGGGAGGAACCGGUUGACGUGGAUUGC	666
S C R N V D G V Y L E Y G R C G K Q E G	200
UCUUGCAGGAUUGUUGAUGGACUUUACCUGGAGUUGGGCGGUGUGGAAAACAAGAAGGA	726
S R T R R S V L I P S H A Q G D L T G R	220
UCAAGAACAAGGCGUUCAGUGCUGAUCCCAUCCACGCUCAGGGAGAUUCACAGGAAGG	786
G H K W L E G D S L R T H L T R V E G W	240
GGACACAAUUGGUUAGAAGGGGAUUCAUUACGGACGCACCUCACUAGAGUUGAGGGGAUGG	846
V W K N K V L T L A V I A V V W L T V E	260
GUCUGGAAGAAUAAAGUGCUCACCCUGGCGGUGAUCGCCGUUGUGUGGUGACCGUGGAA	906
S V V T R V A V V V V L L C L A P V Y A	280
AGUGUGGUGACUCGGGUGCGCGUAGUGGUGGUGCUCUUGUGCCUGGCUCGCGUUUAGGCC	966
S R C T H L E N R D F V T G T Q G T T R	300
UCACGGUGCACACAUUUGGAAAACAGGGAUUUUGUUAUGGCACUCAGGGGACCACUCGU	1026
V T L V L E L G G C V T I T A E G K P S	320
GUGACUCUGGUGUUGGAACUGGGAGGAUGCGUCACCAUAACAGCUGAGGGGAAGCCUCG	1086
M D V W L D S I Y Q E N P A K T R E Y C	340
AUGGAUGUGUGGCUUGACUCCAUCUACCAGGAGAACCUGCCAAGACACGUGAGUACUGC	1146
L H A K L S D T K V A A R C P T M G P A	360
CUUCACGCAAAACUUAUCGGAUACCAAAGUCGCGGCCAGGUGCCCAACAAUUGGACCUGGC	1206
T L A E E H Q S G T V C K R D Q S D R G	380
ACUUUGGCGUAGAGCACCAGAGCGGCACAGUGUGUAAGAGAGACCAGAGUGAUCGAGGC	1266
W G N H C G L F G K G S I V T C V K A S	400
UGGGGCAACCAUUGUGGAUUAUUUGGAAAAGGCAGCAUUGUGACCUGUGUCAAGGCUGUC	1326
C E A K K K A T G H V Y D A N K I V Y T	420
UGUGAGGCAAAAAAGAAAGCCACAGGACACGUGUAUGACGCUAACAAAAUUGUGUACACA	1386
V K V E P H T G D Y V A A N E T H S G R	440
GUCAAAGUAGAGCCGCACACGGGGGAUUACGUCGUGCUAAUGAGACUCACAGUGGAAGA	1446
K T A S F T V S S E R T I L T M G D Y G	460
AAAACCGCGUCCUUCACGGUUCUGCGGAGAGGACCAUCUUGACCAUGGGAGACUACGGA	1506

D V S L L C R V A S G V D L A Q T V I L	480
GACGUGUCCUUGUUAUGCAGAGUAGCCAGCGGUGUUGACCUUGCUCAGACCGUCAUCCUG	1566
E L D K T S E H L P T A W Q V H R D W F	500
GAGCUUGACAAGACCUCAGAACACCUACCGACGGCCUGGCAGGUCCACCGGGACUGGUUC	1626
N D L A L P W K H E G A Q N W N N A E R	520
AAUGAUCUGGCCCUACCGUGGAAACAUGAAGGGGCACAGAAUUGGAACAACGCGGAACGG	1686
L V E F G A P H A V K M D V Y N L G D Q	540
CUGGUUGAGUUUGGAGCUCCACAUGCUGUGAAAAUGGACGUGUACAACCUUGGAGACCAG	1746
T G V L L K S L A G V P V A H I D G T K	560
ACUGGAGUGUUGCUCAAAUCACUUGCUGGUUCCUGUGGCGCACAUUGAUGGAACCAAG	1806
Y H L K S G H V T C E V G L E K L K M K	580
UACCACCUGAAAAUGGGCCACGUGACAUGCGAGGUAGGACUAGAAAAACUUAAGAUGAAA	1866
G L T Y T M C D K T K F T W K R I P T D	600
GGUCUUAUCAACACAAUGUGUGACAAGACGAAAUUCACGUGGAAAGAAUCCAACAGAC	1926
S G H D T V V M E V A F S G T K P C R I	620
AGUGGACAUGACACAGUGGUCAUGGAAGUUGCGUUCUCUGGGACCAAACCCUGCAGGAUC	1986
P V R A V A H G S P D V N V A M L M T P	640
CCGGUGAGGGCCGUGGCACACGGCUCGCCGAUGUGAACGUGGCAUGUUGAUGACACCC	2046
* N P T I E N N G G G F I E M Q L P P G D	660
AACCCCAACAUCGAAACAAUGGCGGUGGCUUCAUGAAUUGCAGUUACCUCAGGAGAU	2106
N I J Y V G E L S H Q W F Q K G S S I G	680
AAUAUCAUCUUGUUGGGAAACUGAGUCACCAUUGGUUCCAAAAAGGGAGUAGCAUUGGA	2166
R V F Q K T R K G I E R L T V I G E H A	700
AGGGUUUUUCAAACCAAGGCAUAGAAAGGCUGACAGUGAUCGGAGAACAUGCC	2226
W D F G S T G G F L T S V G K A L H T V	720
UGGGAUUUUGGCUCUACUGGUGGUUCCUGACCUUGGUUGUAAGGCGCUGCACACAGUU	2286
L G G A F N S L F G G V G F L P K I L V	740
CUUGGCGGUGCCUUUAACAGCCUUUUUGGAGGAGUGGGGUUCUUGCCCAAGAUCCUAGUG	2346
G V V L A W L G L N M R N P T M S M S F	760
GGAGUGGUCCUGGCCUGGUUGGGCCUGAACAUAGGAAUCCGACCAUGUCCAUAGAGCUUC	2406
L L A G G L V L A M T L G V G A D V G C	780
CUUCUGGCUGGAGGACUGGUUCUGGCCAUGACACUCGGAGUGGGAGCUGAUGUUGGCUGU	2466
A V D T E R M E L R C G E G L V V W R E	800
GCUGUGGACACUGAACGGAUGGAGUCCGUGUGGUGAGGGUCUGGUUGUGUGGAGAGAG	2526
V S E W Y D N Y A Y Y P E T L G A L A S	820
GUAUCCGAAUGGUAUGACAAUUAUGCAUACUACCGGAGACACUAGGAGCUCUUGCUUCG	2586
A I K E T F E E G T C G I V P Q N R L E	840
GCCAUAAGGAGACCUUCGAGGAGGGAACUUGUGGCAUAGUGCCCCAAAACAGACUUGAG	2646
M A M W R S S A T E L N L A L V E G D A	860
AUGGCCAUGUGGAGGAGUUCGGCGACAGAACUGAACUUGGCUUUGGUGGAGGGAGACGCA	2706
* N L T V V V D K L D P T D Y R G G I P S	880
AAUCUCACAGUGGUGUGGACAAACUCGAUCCACAGAUUAUCGAGGUGGCAUUCUAGC	2766
L L K K G K D I K V S W K S W G H S M I	900
UUGCUAAGAAAGGGAAAGACAUAAAGGUUUCUUGGAAGAGUUGGGGCCACUCAUGAUC	2826
W S V P E A P R L F M V G T E G S S E C	920
UGGAGCGUCCCCGAGCCCCCGUCUGUUAUGGUGGGAACAGAGGGAAGCAGUGAGUGC	2886
P L E R R K T G V F T V A E F G V G L R	940
*CACUAGAGAGAAGGAAAACAGGUGUCUUCACAGUGGCAGAGUUUGGAGUUGGCUUGAGA	2946
T K V F L D F R Q E S T H E C D T G V M	960
ACAAAAGUAUUUUGGACUUCAGACAGGAUCAACACACGAGUGUGACACAGGAGUGAUG	3006

G A A V K N G M A V H T D Q S L W M K S 980
 GGAGCUGCUGUCAAGAAUGGCAUGGCAGUCCACAGACCAGAGCCUCUGGAUGAAAUCC 3066
 * V R N D T G T Y I V E L L V T D L R N C 1000
 GUGAGAAAUGACACAGGGACCUACAUAUGUGAACUUCUGGUCACUGACCUGAGAAACUGC 3126
 S W P A S H T I D N A E V V D S E L F L 1020
 UCAUGGC 'GGCUAGCCACACUAUCGACAAUGCUGAGGUGGUGGACUCAGAGCUCUUCUU 3186
 P A S L A G P R S W Y N R I P G Y S E Q 1040
 CCGGCCAGUCUGGCGGGGCCAGAUCCUGGUAUAACAGGAUACCUGGGUACUCAGAACAA 3246
 V K G P W K Y S P I R V T R E E C P G T 1060
 GUGAAAGGACCAUGGAAGUACUCACCCAUCGAGUAACAAGAGAAGAGUGCCUGGCACG 3306
 R V T I N A D C D K R G A S V R S T T E 1080
 AGGGUCACCAUAAUUGCUGACUGUGAUAAAGGGGGGCUUCUGAGGAGUACCACAGAG 3366
 S G K V I P E W C C R T C T L P P V T F 1100
 AGUGGCAAGGUGAUUCCAGAGUGGUGCUGCCGAACGUGCACACUACCUCAGUGACGUUC 3426
 R T G T D C W Y A M E I R P V H D Q G G 1120
 CGCACGGGGACAGACUGUUGGUAUGCCAUGGAAUACGUCCUGUUAUGACCAGGGAGGG 3486
 L V R S M V V A D N G E L L S E G G G I P 1140
 CUUGUCCGCUCUAAUGGUGGUGGCAGACAAUGGAGAGCUGCUCAGCGAGGGGGGCAUCCCC 3546
 G I V A L F V V L E Y V I R R R P A T G 1160
 GGAAUAGUGGCCUUGUUUGUGGUCCUGGAAUACGUCAUCCGGAGGAGGCCAGCCACUGGA 3606
 T T A M W G G I V V L A L L V T G L V K 1180
 ACAACGGCCAUGUGGGGAGGCAUUGUUGUCCUUGCACUGCUGACUGGUCUGGUAAG 3666
 I E S L V R Y V V A V G I T F H L E L G 1200
 AUCGAAAGCUUAGUGCGUUAUGUCGUGGCAGUUGGGAUCACAUUCCAUCUUGAGCUGGGG 3726
 P E I V A L T L L Q A V F E L R V G L L 1220
 CCAGAGAUUGUGGCUCUGACACUGUUAACAGGCUGUGUUUGAGUUGAGGGUGGGCCUGCUC 3786
 S A F A L R S N L T V R E M V T I Y F L 1240
 AGCGCUUUUGCGCUGCGCAGCAACCCUACCGUCAGAGAGAUUGGUAACCAUCUACUCCUU 3846
 L L V L E L G L P S E G L G A L W K W G 1260
 CUGCUGGUUCUGGAGCUGGGAUUGCCAAGCGAGGGCCUUGGGGCCUCUGCAAUUGGGGA 3906
 D A L A M G A L I F R A C T A E E K T G 1280
 GAUGCACUGGCUAUGGGGGCAUUGAUUUUCAGAGCUUGCACAGCAGAGGAAAAGACCGGU 3966
 V G L L L M A L M T Q Q D L A T V H Y G 1300
 GUUGGACUUCUGCUCUAGGCUCUCAUGACACAGCAAGACCUGGCAACUGUACAUAUUGG 4026
 L M L F L G V A S C C S I W K L I R G H 1320
 CUCAUGCUCUUCUUGGGCGUGGCUCUACUUGCUCAAUCUGGAAACUGAUUCGAGGACAC 4086
 R E Q K G L T W I V P L A G L L G G E G 1340
 AGAGAACAGAAGGGACUGACUUGGAUUGUCCUUGGCCGGGUUACUGGAGGAGAGGGG 4146
 S G V R L V A F W E L T V H G R R R S F 1360
 UCUGGAGUUAUGAUUGGUGGCCUUUUGGAGCUAACCGUCCAUGGAAGGAGACGUUCAUUC 4206
 S E P L T V V G V M L T L A S G M I R H 1380
 AGUGAACCCACUGACUGUCGUGGAGUCAUGCUAACCCUGGCCAGCGGCAUGAUCCGGCAC 4266
 T S Q E A L C A L A V A S F L L L M L V 1400
 ACCUCUCAGGAGGCUCUUUGUGCGCUCGCCGUAGCUUCGUUCCUCCUGCUAUGCUGGUG 4326
 L G T R K M Q L V A E W S G C V E W H P 1420
 CUAGGGACAAGGAAGAUAGUGGCUGAAUGGAGUGGCUGUGUGGAGUGGCACCCA 4386
 E L M N E G G E V S L R V R Q D S M G N 1440
 GAACUAAUGAAUGAAGGUGGAGAGGUGAGCCUGCGGGUCCGGCAGGAUUCAAUUGGGGAAC 4446
 F H L T E L E K E E R V M A F W L L A G 1460
 UUCCACCUGACAGAGCUUGAGAAAGAGGAAAGAGUGAUGGCUUUUUGGUGCUGGCAGGA 4506

L A A S A F H W S G I L G V M G L W T L	1480
CUGGCGGCUUCAGCCUUCACUGGUCGCGCAUCCUUGGUGUGAUGGGAUUGUGGACGCUU	4566
NS3	
S E M L R T A R R S G L V F S G Q G G R	1500
UCAGAAAUGCUGAGAACGGCUCGAAGAUCAGGCUUGGUCUUCUCUGGACAAGGAGGACGU	4626
E R G D R P F E V K D G V Y R I F S P G	1520
GAGCGUGGUGACAGGCCCUUUGAGGUCAAGGAUGGCGUGUACAGAAUCUUCAGCCAGGA	4686
L L W G Q R Q V G V G Y G S K G V L H T	1540
CUGCUUUGGGGGCAGCUCUAGUGGGAGUUGGCUAUGGUUCCAAAGGUGUCCUACACACG	4746
M W H V T R G A A L S I D D A V A G P Y	1560
AUGUGGCAUGUGACGAGAGGGGCGCGUUGUCCAUGAUGACGCCGUGCGAGGGCCCUAU	4806
W A D V K E D V V C Y G G A W S L E E K	1580
UGGGCUGAUGUCAAGAGGACGUUGUAUGCUACGGCGGAGCCUGGAGUCUUGAGGAGAAG	4866
W K G E T V Q V H A F P P G R A H E V H	1600
UGGAAAGGUGAGACAGUGCAGGUCCAUGCCUUCACCGGGGAGAGCUCUAGAGGUGCAU	4926
Q C Q P G E L L L D T G R R I G A V P I	1620
CAAUGUCAGCCCGGAGAACUGCUCCUGGACACAGGUAGGAGGAUAGGGGACUGCCAAUU	4986
D L A K G T S G S P I L N S Q G V V V G	1640
GAUUUGGCAAAGGGGACAUUGGCAGUCCAUCCUACUCCAGGGAGUGGUUGUGGGG	5046
*	
L Y G N G L K T N E T Y V S S I A Q G E	1660
CUGUAUGGGAACGGACUGAAGACCAAUGAAACCUACGUCAGCAGCAUUGCUCAGGGUGAG	5106
A E K S R P N L P P A V T G T G W T A K	1680
GCGGAAAAGAGUCGACCCAACCUCCGCCGCGGUCACUGGCACAGGCUGGACAGCAAAA	5166
G Q I T V L D M H P G S G K T H R V L P	1700
GGGCAGAUACAGUGCUGGACAUUGCACCCAGGCUCUGGGAAGACCCACAGAGUCCUCCG	5226
E L I R Q C I D R R L R T L V L A P T R	1720
GAGCUAUUCGCCAAUGCAUUGACAGACGCCUAAGGACAUUGGUGUUGGCCCAACCCGU	5286
V V L K E M E R A L N K R V R F H S P	1740
GUGGUGCUUAAGGAAAUGGAGCGUGCCUUGAAUGGGAAGAGGGUCAGGUUCCAUCUCCU	5346
A V G D Q Q V G G S I V D V M C H A T Y	1760
GCAGUUGGAGAUACGCAAGUGGAGGGUUAUCGUCGAUGUGAUGUGCCAUGCAACCUAU	5406
V N R R L L P Q G R Q N W E V A I M D E	1780
GUCAACAGACGCCUGCUCCGCCAGGGGAGACAGAACUGGGAAGUAGCAAUCAUGGAUGAA	5466
A H W T D P H S J A A R G H L Y T L A K	1800
GCUCACUGGACAGAUCCACACAGCAUAGCCGCUCCGGGUCACUUGUACACCCUGGCUAAA	5526
E N K C A L V L M T A T P P G K S E P F	1820
GAAAACAAAUGUGCCUUGGUUCUUGAUGACGGCAACGCCACCUGGGAAGAGCGAGCCUUC	5586
P E S N G A I S S E E K Q I P D G E W R	1840
CCAGAGUCCAACGGAGCAAUUCAGUGAAGAGAAGCAGAUCCCCGAUGGGGAGUGCGGU	5646
D G F D W I T E Y E G R T A W F V P S S	1860
GAUGGAUUCGACUGGAUACCGAGUAUGAGGGGCGUACCGCAUGGUUCGUUCCUCCGAGU	5706
A K G G I I A R T L I Q K G K S V I C L	1880
GCAAAAGGUGGCAUCAUAGCACGUACCCUGAUACAAAAGGAAAAAGUGUGAUCUGUCUG	5766
N S K T F E K D Y S R V R D E K P D F V	1900
AACAGCAAGACCUUUGAAAAGGACUACUCCAGAGUGAGAGAUGGAAACAGACUUCGUG	5826
V T T D I S E M G A N L D V S R V I D G	1920
GUCACAACCGACAUAUCUGAAAUGGGGGCCAACCUCGAUGUGAGCCGUGUCAUAGAUGGA	5886
R T N I K P E E V D G R V E L T G T R R	1940
CGAACAAACAUCAAACCGGAAGAGGUUGAUGGGCGAGUUGAGCUCACAGGGACCAGACGU	5946
V T T A S A A Q R R G R V G R Q E G R T	1960
GUGACCACGGCCUCUGCGGCCCAACGCCGUGGGAGAGUCGGAAGACAGGAGGGAAGAACA	6006

D E Y I Y S G Q C D D D S G L V Q W K GAUGAAUACAUUACUCUGGACAGUGUGAUGAUGAUGAGUGGACUUGUGCAGUGGAAG	1980 6066
E A Q I L L D N I T T L R G P V A T F Y GAAGCGCAGAUACUUCUUGACAACAUACAACACUGCGGGGCCUGUGGCUACCUUCUUAU	2000 6126
G P E Q D K M P E V A G H F R L T E E K GGACCAGAGCAGGACAAGAUGCCAGAGGUGGCGGGUCAUUCGCCUCACUGAAGAGAAA	2020 6186
R K H F R H L L T H C D F T P W L A W H AGAAAGCAUUUUGGACAUUCUCACCCACUGUGACUUCACGCCUGGCUAGCAUGGCAC	2040 6246
V A A N V S S V T S R N W T W E G P E E GUCGCAGCAAACGUGUCUAGUGUGACAAGUCGGAACUGGACUGUGGAAGGUCCUGAGGAG	2060 6306
N T V D E A N G D L V T F R S P N G A E AACACCGUGGACGAGGCCAAUGGAGAUUCUGGUCACCUUCAGGAGCCCGAAUGGGGCGUAA	2080 6366
R T L R P V W R D A R M F R E G R D I R AGAAUUGAGGCCAGUGUGGAGGGAUGCGCGCAUGUUCAGAGAGGGACGUGACAUCAGA	2100 6426
E F V A Y A S G R R S F G D V L S G M S GAGUUUGUGCGUAUGCCUCAGGGAGACGCGAGCUUCGGAGACGUGUGAGCGGAUUGUCC	2120 6486
G V P E L L R H R C V S A M D V F Y T L GGGUUCCUGAGCUUCUGCGCCACAGAUUGUUAUGGCCAUGGAUGUCUUCUACACACUG	2140 6546
M H E E P G S R A M K M A E R D A P E A AUGCAUGAGGAGCCUGGCAGUAGAGCAAUGAAGAUGCCGAGAGAGAUUGCUCCAGAGGCC	2160 6606
F L T V V E M M V L G L A T L G V V W C UUCUUGACGGUGUAGAGAUGAUGGUGCUCGGCCUGGCUACUCUUGGCGUCUUUGGUGC	2180 6666
F V V R T S I S R M M L G T L V L L A S UUUGUUGUUGCAGUUCUAAUCAGCCGAUGAUGCUUGGCACGUGGUACUGCUGGCCUCU	2200 6726
L A L L W A G G V S Y G N M A G V A L I UUGGCGCUCUUAUGGGCUGGGGUGUAAGCUACGGGAACAUGGCAGGAGUGGCUCUCAU	2220 6786
F Y T L L T V L Q P E A G K Q R S S D D UUUACACGUGGUGACGGUGCUGCAGCCUGAGCGGGAAAACAGAGAAGCAGCGAUGAC	2240 6846
N K L A Y F L L T L C S L A G L V A A N AACAAGUUGGCUUACUUCUGUUGAGCGCUCUGCAGUCUGGUGGACUGGUGGCGUCCAAU NS4B	2260 6906
E M G F L E K T K A D L S T V L W S E H GAGAUGGGAUUUCUGGAGAAGACCAAGGCGGACCUGUCCACGGUGUUGUGGAGUGAACAU	2280 6966
E E L R S W E E W T N I D I Q P A R S W GAGGAGUUGCGUGUGGGAAGAGUGGACCAUAUUCGAUAUCCAGCCUGCAGGUUCUUGG	2300 7026
G T Y V L V V S L F T P Y I I H Q L Q T GGAACUUCGUGUUGGUGGUCUCCUGUUCACACCAUACAUAUCCACCAACUUCAGACC	2320 7086
K I Q Q L V N S A V A T G A Q A M R D L AAGAUCACAGCUCGUCAACAGUGCUGUUGCAACUGGAGCCCAGGCCAUGGCAGACCUC	2340 7146
G G G A P F F G V A G H V M A L G V V S GGAGGAGGGGCUCAUUUUUGGGUAGCAGGGCAUGUAAUGGCUUUGGAGUGGUAUCU	2360 7206
L V G A T P T S L V V G V G L A A F H L CUGGUUGGUGCAACGCCAACUCCUUGGUGGUUGGUGUUGGUCUGGCGGCGUCCACCUG	2380 7266
A I V V S G L E A E L T Q R A H K V F F GCCAUUGUGGUGUCCGGACUAGAGGCUGAAUUGACACAAGAGCCCAAAAGUCUUCUUC	2400 7326
S A M V R N P M V D G D V I N P F G E G UCGGCAAUGGUACGCAAUCCCAUGGUGGAUGGAGACGUCAUCAAUCCAUUUGGAGAGGGG	2420 7386
E A K P A L Y E R K M S L V L A I V L C GAGGCAAAACCGUCUUAUUGAGAGAAAAUAGGCCUGGUCUUGGCAUAGUGCUUUGC	2440 7446
L M S V V M N R T V P S T P R L L L W D UUGAUGUCAGUGGUGAUGAACAGAACAGUGCCUUCUACACCGAGGCUUCUCCUGUGGGAU	2460 7506

W R Q R D N C S N Q R R T P F G R C Q A	2480
UGGCGGCAGCGGACAACUGCUGAACCCAGAGGCGGACCCUUUGGACGAUGCCAGGCC	7566
C G L S G V V R G S L W G F C P L G H R	2500
UGUGGCCUGAGUGGCGUGGUCAGGGGAGUCUCUGGGGAUUCUGUCCCCUUGGGCACAGA	7626
L W L R A S G S R R G ^{NS5} G S E G D T L G D	2520
CUCUGGCUCAGGGCUUCUGGGAGCAGGCGUGGCGGCUCUGAAGGGGACACUCUUGGUGAC	7686
L W K R K L N G C T K E E F F A Y R R T	2540
UUGUGGAAACGGAAGCUCAAUGGCUGUACCAAGGAAGAGUUCUUCGCCUAUAGGCGCACU	7746
G I L E T E R D K A R E L L K R G E T N	2560
GGCAUCCUGGAGACGGAAAGGGACAAGGCGCGGGAACUCCUCAAGAGAGGGGAGACCAAC	7806
M G L A V S R G T A K L A W L E E R G Y	2580
AUGGGGCGUGGUGUCACGGGGCAGGGCAAAACUUGCCUGGCUUGAGGAACGAGGUUAC	7866
A T L K G E V V D L G C G R G G W S Y Y	2600
GCAACUCUCAAGGGUGAGGUAGUGGACCUUGGAUGUGGAAGAGGCGGCGUGUCAUACUUA	7926
A A S R P A V M S V K A C A I A G K G H	2620
GCAGCCUCUCGGCCGGCAGUCAUGAGUGUCAAGCCUGUGCAAUUGCUGGAAAGGGACAU	7986
E T P K M V T S L G W N L I K F R A G M	2640
GAGACUCCAAAGAUGUGGACAAAGCCUGGGUUGGAACCGAUCAAGUUCAGAGCGGGAUUG	8046
D V F S M Q P H R A D T I M C D I G E S	2660
GAUGUGUUCAGUAUGCAACCCACCCGAGCUGAUACCAUCAUGUGUGACAUUGGAGAAAGC	8106
N P D A V V E G E R T R K V I L L M E Q	2680
AACCCAGAUGCCGUGGUGAGAGGACACGGAAAGUGAUUCUACUCAUGGAACAG	8166
W K N R N P T A T C V F K A L A P Y R P	2700
UGGAAAAACCGUAACCCACGGCCACCUGUGUCUUUAAAGCGUGGCCCCCGUACCGCCCA	8226
E V T E A L H R F Q L W G G G L V R T	2720
GAGGUCACAGAAGCACUGCAGACAUUCCAGCUGCAGUGGGGGCGGAGGACUGGUGAGGACC	8286
P F S R N S T H E M Y Y S T A I T G N I	2740
CCUUUUUCGAGGAACUCAACCCAUGAGAUGAUUACUCGACUGCUAUCACUGGAAAAUUA	8346
V N S V N I Q S R K L L A R F G D Q R G	2760
GUGAAUUCUGUCAACAUCCAGUCAAGGAAACUCCUGGCCCGGUUGGGGACCAGAGGGGA	8406
P T R V P E L D L G V G T R C V V L A E	2780
CCCACCAGGUGCCUGAGCUGGACCUUGGAGUUGGGACCCGGUGUGUUGUCUUGGCUGAG	8466
D K V K E K D V Q E R I S A L R E Q Y G	2800
GACAAGGUGAAAGAAAAAGAUUGUCAGGAGAGGAUCAGUGCGUUGCGGGAACAGUAUGGU	8526
E T W H M D R E H P Y R T W Q Y W A A T	2820
GAGACCUGGCACAUGGACAGAGAGCACCCGUACAGGACCUGGCAGUACUGGGCAGCUACC	8586
A C A N R V G G A L I N G V V K L L S W	2840
GCCUGCGCCAACCGGGUCGGCGGCACUGAUUAAUGGAGUCUGUAAACUCCUCAGCUGG	8646
P W N A R E D V V R M A M T D T T A F G	2860
CCAUGGAACGCGCGGGAGGAUGUCGUGCGAAUGGCCAUGACUACACCACAGCCUUUGGA	8706
Q Q R V F K E K V D T K A Q E P Q P G T	2880
CAGCAGCGAGUGUUAAGGAGAAGGUUGACACUAAGGCUCAGGAACCUCAGCCUGGCACA	8766
K V I M R A V N D W I L E R L A R K S K	2900
AAGGUCAUCAUGAGAGCAGUGAAUGACUGGAUUCUGGAACGAUUGGCACGAAAAAGCAAA	8826
P R M C S R E E F I A K V K S N A A L G	2920
CCACGAUUGUCAGCAGAGAGGAAUUAUAGCGAAAGUGAAAUCCAAUGCAGCUCUGGGG	8886
A W S D E Q N R W S S A K E A V E D P A	2940
GCUUGGUCUGAUGAGCAGAACAGGUGGUCAAGUGCAAAAGAGGCUGUGGAGGAUCCCGCA	8946
F W Q L V D E E R E R H L A G R C A H C	2960
UUCUGGCAGCUCUGGACGAAGAGAGAGAGACACCUCGCGGGGAGAUUGCGCCACUGU	9006

V Y N M M G K R E K K L G E F G V A K G GUGUAUACAUGAUGGGCAAGAGAGAAAAAGCUUGGAGAGUUUGGAGUGGCCAAAGGA	2980 9066
S R A I W Y M W L G S R F L E F E A L G AGCCGGGCCAUUAUGGUACAUGUGGCUGGGGAGCCGCUUUCUGGAGUUCGAAGCUCUUGGC	3000 9126
F L N E D H W A S R G S S G S G V E G I UUUUUGAAUGAGGACCACUGGGCCUCUAGGGGUCCAGUGGAUCUGGAGUGGAAGGAUA	3020 9186
S L N Y L G W H L K G L S T L E G G L F AGCCUGAACUACCUGGGCUGGCAUCUCAAGGGUUGUCAACAUGGAAGGAGGCCUUUUU	3040 9246
Y A D D T A G W D T K V T N A D L E D E UAUGCAGAUGACACAGCCGGCUGGGAUACCAAGGUCACUAACGCAGACCUAGAGGAUGAA	3060 9306
E Q L L R Y M E G E H K Q L A A T I M Q GAGCAGCUCUUACGCUACAUGGAGGGUGAACACAAGCAACUGGCGGCUACAAUUAUGCAG	3080 9366
K A Y H A K V V K V A R P S R D G G C I AAGGCAUACCACGCCAAAGUGGUAAAAAGUUGCCCGGCCUCCCGAGAUGGAGGCUGCAUC	3100 9426
M D V I T R R D Q R G S G Q V V T Y A L AUGGAUGUCAUCACAAGAAGAGACCAAGAGGUUCAGGACAGGUUGUGACUUUAGCCCCUC	3120 9486
N T L T N I K V Q L I R M M E G E G V I AACACCCUCACCAACAUAAAGGUUCAACUAAUCCGAAUGAUGGAGGGGGAGGGGUGCAUU	3140 9546
E A S D A H N P R L L R V E R W L R D H GAGGCCUCGGACGCACAUAAACCCAAGACUACUUCGAGUGGAACGAUGGCUGAGAGAUCAU	3160 9606
G E E R L G R M L V S G D D C V V R P V GGAGAAGAAGCUCUGGAAGAAUGCUCUGAGCGCGCAUGACUGUGUGGUGAGGCCAGUG	3180 9666
D D R F S G A L Y F L N D M A K T R K D GAUGAUAGGUUCAGUGGGGGCUCUACUUUCUGAUGACAUGGCAAAAACAGGAAAGAC	3200 9726
I G E W D H S V G F S N W E E V P F C S AUCGGGGAUUGGGACCAUUCGGUUGGCUUCUCCGAACUGGGAGGAGGUUCCUUUCUGCUA	3220 9786
H H F H E L V M K D G R T L I V P C R D CACCAUUUCCACGAGCUGGUGAUGAAAGAUGGGCGCACCCUCAUAGUGCCAUGCCGAGAC	3240 9846
Q D E L V G R A R V S P G C G R S V R E CAAGAUGAACUGGUUGGAAGAGCCCCGUCUCACCAGGCUGCGCGCGGAGCGUCCGUGAA	3260 9906
T A C L S K A Y G Q M W L L S Y F H R R ACUGCCUGCCUUUCAAAGCUUAUGGGCAGAUGUGGCUGCUGAGCUACUCCACCGGCGC	3280 9966
D L R T L G L A I C S A V P V D W V P A GACUUGCGAAGCGCUUGGACUUGCCAUUCUGUGCGGCGGUGCCCGUCACUGGGUCCUGCU	3300 10026
G R T T W S I H A S G A W M T T E D M L CGUGCCACGACCUGGAGCAUUCACGCUAGCGGAGCCUGGAUGACCACAGAGGACAUGCUG	3320 10086
D V W N R V W I L D N P F M H S K E K I GAUGUCUGGAACAGGGUGUGGAUUCUGGACAACCCCUCAUGCACAGCAAAGAAAAGAUU	3340 10146
A E W R D V P Y L P K S H D M L C S S L GCGGAAUGGAGGGGAUGUCCCGUACCUCUCCCAAUCCCAUGACAUUGCUGUUCUCCCUU	3360 10206
V G R K E R A E W A K N I W G A V E K V GUUGGAGGAAAGAGAGGGCAGAGUGGGCUAAGAACAUCUGGGGAGCAGUAGAGAAAGUC	3380 10266
R K M I G Q E K F K D Y L S C M D R H D AGGAAGAUGAUCGGACAAGAGAAGUUAAGGACUACCUUCCUGCAUGGACCGGAUGAC	3400 10326
L H W E S K L E S S I I UUGCACUGGGAGUCCAAACUGGAGAGCUAAUAAUC <u>UAAA</u> ACCAGACUGUGACUGAGCAA	3413 10386
AACCUGGAGUGCUGGUAAAACAUUGUCCAGAACCAAAAACACAGCAAGCAAUUCACAGA	10446
AACAGAGCUCGAACUGAAGAGCUCUUUAAAC	10477

других потенциальных сайта гликозилирования Asn-Pro-Thr (641-643 и 753-755 а.о.) не реализуются.

Расщепление полипротеинов флавириусов протеиназой со специфичностью к последовательности Val-X-Ala, расположенной на расстоянии (-3)-(-1) от места гидролиза, обеспечивает формирование N-концов белков NS2A и NS4B [7, 16]. Такие места процессинга для полипротеина TBE находятся в положениях 1126-1129 и 2257-2260 а.о. (последовательности Val-Val-Ala¹-Asp и Val-Ala-Ala¹-Asn, соответственно). Кроме такого типа протеолиза полипротеинов флавириусов формирование N-концевых последовательностей белков NS3, NS4A и NS5 осуществляет, по-видимому, вирус-специфическая протеиназа по последовательности, обогащенной остатками основных аминокислот, которым предшествуют протяженные гидрофобные участки полипротеинов. Функцию вирус-специфической протеиназы могут, вероятно, выполнять белки преМ или NS1, так как в их составе присутствуют последовательности Cys-Trp, часто встречающиеся в активных центрах тиролевых протеиназ. Учет этих двух типов гидролиза полипротеинов флавириусов приводит к локализации семи белковых продуктов (NS1, NS2A, NS2B, NS3, NS4A, NS4B и NS5), наблюдаемых в инфицированных флавириусами клетках, в том числе и при развитии TBE [2, 25, 34]. Низкомолекулярные белки NS2A, NS2B, NS4A и NS4B с *M* 24 873, 14 531, 16 063 и 27 585 Да — гидрофобные белки с протяженными участками остатков незаряженных аминокислот (рис. 2). Функции этих белков в развитии вируса в клетках не выяснены.

По сравнению с низкомолекулярными продуктами белки NS1, NS3 и NS5 легко обнаруживаются в инфицированных флавириусами клетках при использовании радиоактивных предшественников. Белок NS1 служит сильным иммуногеном; антитела к этому белку способны вызывать комплемент-зависимый лизис инфицированных вирусом клеток [35]. Предварительная иммунизация животных белком NS1 вирусов Den-2 или YF приводит к защите животных от высоких доз заражения этими вирусами [36, 37], однако в сыворотке крови животных, иммунизированных белком NS1, не обнаруживают вирус-нейтрализующих антител. Белок NS1, состоящий из 352 а.о. (*M* 39 162 Да), является гликопротеином. В структуре белка присутствуют три потенциальных места гликозилирования. Гликозилирование NS1 не всегда сопутствует инфекционному процессу; так, радиоактивно меченная манноза или глюкозамин в клетках, инфицированных вирусом Kun, не способны включаться в белок [38]. В соответствии с локализацией белка NS1 на полипротеине его молекулярная масса оказывается ниже той, которую определяют при электрофорезе внутриклеточного белка в ПААГ. Белок NS1 обнаруживают в димерной форме [39] либо во фракции белка с *M* 49 000 Да. Последняя, по-видимому, представляет собой слитую форму NS1-NS2A (*M* 45 710 Да), которая за счет гидрофобности цепи NS2A приобретает способность удерживаться на клеточных мембранах [2, 39].

Белки NS3 и NS5 (*M* 68 949 и 102 639 Да) необходимы для репликации вирусной РНК, так как в клетках их обнаруживают ассоциированными с репликативной формой вирусной РНК [40]. В структурах белков NS5 всех флавириусов присутствует последовательность, характерная для РНК-зависимых РНК-полимераз вирусов животных и растений [2]: Gly-Asp-Asp. В ДНК одной из рекомбинантных плазмид, кодирующих белок

Рис. 2. Нуклеотидная последовательность генома и аминокислотная последовательность полипротеина TBE. Положения мест расщеплений полипротеина и локализация генов на геноме находятся в соответствии с данными, представленными в табл. 1, и указаны стрелками на рисунке. Цифрами справа дана нумерация аминокислотных остатков полипротеина (верхняя цифра) и нуклеотидных остатков генома (нижняя цифра). Потенциальные места гликозилирования белков по последовательности Asn-X-Ser/Thr обозначены звездочками. Протяженные участки остатков незаряженных аминокислот полипротеина отмечены чертой сверху. Двойной линией указан кодон терминации трансляции полипротеина. Названия аминокислот даны в однобуквенном коде в соответствии с рекомендацией IUPAC-IUB

Аминокислотные последовательности полипротеинов
флавивирусов в местах протеолитического гидролиза

Белок	Вирус	Место процессинга полипротеина*	Литературный источник
C	TBE	M AG KAILKGKGGGP P	
	TBE _h	VKKAILKGKKGXGP P	23
	YF	M SGRKA QGKTLGVNMV	2
	WN	M SK KPGGPGKNRAVNML	18, 17
	Kun	M SK KPGGPGKSRVNML	16
	SLE	M SK KPGKPGRNRVVNML	21
npe-M	TBE	VVVLGG VTAAATVRKERDGTTVI	
	WN	IACAGA VTLSNFQGV MMTVN	18, 17
	Kun	IAGVGA VTLSNFQGV MMTVN	16
M	TBE	SR TRR SVLIPSHAQGDLTGRGH	
	YE	SRRSRR AIDLPTHEHGLKTRQE	2
	WN	SRRSRR SLTVQTHGESTLANKKG	18, 17
E	TBE	LAPVYA SRCTHLENRDFVTGTQG	
	YF	VGPAYS AHCIGITDRDFIEGVHG	2, 21
	WN	VAPAYS FNCLGMSNRDFLEGVSG	17, 19
	Kun	VAPAYS FNCLGMSNRDFLEGVSG	16
	JE	VAPAYS FNCLGMGNRDFIEGASG	22
	Den-2	VARSMT MRCIGISNRDFVEGVSG	2, 21
NS1	TBE	TLGVGA DVGCAVDTERMELRCGE	
	YF	SLGVGA DQGCAINFGKRELKCGD	2, 6
	SLE	ATSVQA DSGCAISLQRRELKCGG	6, 20
NS2A	TBE	RSMVVA DNGELLSEGGIPGIVAL	
	Kun	QSQVNA YNAD MIDPFQLGLLVV	7, 16
NS2B	TBE	VHGRRR SR SERLTVGVMLTL	
	Kun	DPNRKR GWPATEVMTAVGLMFAI	7, 16
NS3	TBE	LRTARR SGLVFSGQGGREGRCDRP	
	YF	VRGARR SGDVLWDIPTPKIIEEC	2, 6
	SLE	GKHSKR GG ALWDVPSPKVYPKC	6, 20
	Kun	LQYTKR GG VLWDTPSPKEYKRG	7, 16
NS4B	TBE	AGLVAA NEMGFLEKTKADLSTVL	
	Kun	VGAVAA NEMGWLDKTKSDISGLF	7, 16
NS5	TBE	ASGSRR GGSEGDTLGDLWKRKLN	
	YF	MRTGRR GSANGKTLGEVWKRELN	2, 6
	SLE	XXXXXATLXXTXXXXXX	6
	Kun	KPGLKR GGAKGRTLGEVWKERLN	7, 16

* Подчеркнуты остатки аминокислот N-концевых участков белков, определенных прямым секвенированием белка. Аминокислотные последовательности TBE приведены на основании исследования нуклеотидной последовательности генома.

NS5, в этом участке белка мы обнаружили иную последовательность — Gly-Leu-Asn-Asp (3 171—3 175 а.о.). Такая последовательность возникла в результате вставки триплета CUC и мутации G→A в первом положении кодона GAU. Маловероятно, что такая последовательность могла возникнуть в процессе клонирования; скорее всего, данный клон представляет копию минорной популяции вирусной РНК.

По нашим предварительным данным, синтез РНК in vitro репликативным комплексом, выделенным из инфицированных TBE клеток почки эмбриона свиньи, способен блокироваться мышечными антителами, полученными к вирусным белкам NS3 и NS5.

Расшифрованная нами структура генома TBE — пока единственная информация о молекулярных основах строения представителя флавивирусов, переносимого клещами. С помощью компьютерного анализа мы сравнили аминокислотные последовательности белков TBE и других пред-

Гомология полипротенинов ТВЕ и других представителей семейства
флавивирусов (%) *

Белок	YF	WN	Kun	JE	MVE	SLE	Den-2	Den-1	Den-4
C	30,4	21,9	22,7	17,4	20,8	20,0	18,5	20,2	15,4
преМ	34,3	24,5	25,5	20,9	23,6	21,3	19,4	20,4	15,9
M	26,3	25,0	25,0	25,0	26,3	30,3	21,1	18,4	23,7
E	42,0	40,2	39,7	39,3	39,3	39,7	36,7	37,9	36,5
NS1	42,3	43,2	41,8	44,1	42,7	39,1	38,4	37,6	39,1
NS2A	24,8	19,7	20,1	22,2	19,7	20,1	20,0	—	19,1
NS2B	24,1	20,3	20,3	18,0	21,1	22,6	19,5	—	18,0
NS3	45,5	45,8	47,2	45,6	—	—	44,4	—	43,9
NS4A	36,0	34,0	34,0	34,0	—	—	30,5	—	25,2
NS4B	29,0	24,6	24,6	23,0	—	—	24,6	—	25,8
NS5	55,8	55,9	56,0	54,1	—	—	53,4	—	54,7

* Процент прямой гомологии аминокислотных последовательностей белков вычисляли на основании данных, представленных на рис. 2, для ТВЕ и данных исследований структур гепатомов и белков вирусов YF [2], WN [17], Kun [16], JE [14], MVE [29], SLE [20], Den-2 [15, 42], Den-1 [51], Den-4 [30, 32].

ставителей семейства флавивирусов (табл. 2). Наименьшее сходство наблюдается среди низкомолекулярных неструктурных белков NS2A, NS2B, NS4B и вирионных белков M и C. Несмотря на более низкую (15–29%) прямую гомологию аминокислотных остатков, в последовательностях гидрофобных неструктурных белков флавивирусов обнаруживается закономерность в положениях протяженных участков остатков незаряженных аминокислот (рис. 2), а в структурах вирионных белков C флавивирусов гомологичность отражает консервативность в положениях остатков основных аминокислот, по-видимому, существенных в образовании нуклеокапсидного комплекса этих белков с РНК. Максимальная гомология характерна для структур белков E, NS1, NS3 и NS5, причем внутри одной серологической подгруппы вирусов эта гомология очень высока. Например, у вирусов WN и Kun гомология этих белков составляет 92–95%. Среди представителей флавивирусов различных подгруппы она существенно меньше.

Сходная картина характерна для положений потенциальных мест гликозилирования белков среди представителей различных подгрупп флавивирусов. Полипротенины ТВЕ и YF наиболее близки. Эти вирусы имеют различный тип переносчиков в природе, но обладают близкими структурами: гомология по белкам E, NS1, NS3 и NS5 составляет 42; 42,3; 45,5 и 55,8%, соответственно. Следует отметить, что наиболее консервативно положение всех остатков Cys у белка E и большинства Cys в белках преМ, NS1, NS3 и NS5. Это указывает на общность структур белков на более высоких уровнях организации и, по-видимому, на идентичность выполняемых ими функций. Исследование профиля гидрофобности полипротенина ТВЕ, согласно [41], выявляет удивительную идентичность положений гидрофобных и гидрофильных участков на полипротенинах флавивирусов [2, 15–17, 29, 32, 42], несмотря на более разнообразную их прямую гомологию. Эти предопределенные структурой особенности полипротенинов флавивирусов, возможно, играют ключевую роль в развитии вирусов в клетках, например в созревании белков при мембрано-ассоциированном сплитинге и процессинге полипротенинов-предшественников.

Таким образом, несмотря на различные способы передачи вирусов и разобщенность их циркуляции в природных очагах, флавивирусы, переносимые комарами и клещами, имеют однотипную организацию гепатомов и полипротенинов и, по-видимому, сходные механизмы их функционирования в клетках животных, а также высокую гомологию белков, необходимых для репродукции вирусов в клетках.

Виррус клещевого энцефалита, штамм Софийн, был выделен и очищен по методу [43] С. Г. Рубиным (ИПВиЭ АМН СССР). РНК ТВЕ выделяли методом фенольной депротеинизации и очищали центрифугированием в градиенте сахарозы (5–20%). Репликативную форму РНК ТВЕ выделяли из мозга мышей, инфицированных ТВЕ (штамм Софийн).

В работе использованы препараты $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$ и $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ («Изотон»), dNTP и ATP (Sigma, США), нитроцеллюлозные фильтры BA-85 (Schleicher und Schüll, ФРГ), эндонуклеазы рестрикции *Bgl*I, *Bam*HI, *Pvu*II, *Nco*I, *Eco*RI, *Pst*I, *Sal*GI, *Cla*I, *Hind*III, *Bsp*I, *Ava*II, *Msp*I, *Nin*I, *Taq*I, *Sau*3A, *Cfr*I, ДНК-полимераза I *E. coli*, ДНК-полимераза А (фрагмент Клепова), терминальная дезоксирибонуклеотидилтрансфераза из тимуса теленка, полирибонуклеотидкиназа и ДНК-лигаза фага T4 (НИКТИ БАН, Бердск, или НПО «Фермент», Вильнюс). Полиаденилаттрансфераза *E. coli* любезно предоставлена С. Х. Дегтяревым (НИКТИ БАН, Бердск). Ферментативную обработку ДНК, выделение ДНК плазмид и фрагментов ДНК, трансформацию клеток *E. coli* проводили в соответствии с ранее опубликованными данными [3–5, 9] или как описано в работе [44].

Рекомбинантные плазмиды *p10*, *p4*, *p2*, *p15*, *p131*, *p1*, *p6*, *p5*, *p13*, *p37* (рис. 1) были получены нами ранее кошекторным методом [3]. ДНК копий участков генома ТВЕ, не представленных в библиотеке предыдущего клонирования, получали с помощью синтетических олигонуклеотидов-праймеров. При синтезе первой цепи кДНК в качестве затравок использовали олигонуклеотиды d(CCTTGTCCTAGCACCAGC), d(TCCCCAAGAAC·GTGCAGG), комплементарные РНК ТВЕ на участках в положениях 4320–4338, 7012–7029 (рис. 2) и смесь олигонуклеотидов d(AGT(G/T)(G/T)T(G/T)TATGTGT), комплементарных 3'-концевым последовательностям РНК вирусом YF, WN, JE, SLE [2, 12, 13]. Вторую цепь кДНК синтезировали с помощью ДНК-полимеразы I (фрагмент Клепова) и олигонуклеотидов-затравок: d(ACGAAATTACGCTG), d(AGCAAACCACGAAT) и d(TTCGTTCCCTCG), комплементарных минус-цепи РНК ТВЕ в положениях 1894–1907, 8824–8834 и 5692–5703 соответственно (рис. 2). Олигонуклеотиды, комплементарные вирусной РНК, были синтезированы А. В. Силяковым (ВНИИ, Новосибирск); олигонуклеотиды, комплементарные минус-цепи РНК, — В. В. Горном. Двухцепочечные кДНК лигировали с *Hind*III-линкером d(GAAGCTTC), продукты реакции гидролизировали рестриктазой *Hind*III и от низкомолекулярных продуктов гидролиза освобождались гель-фильтрацией на колонке с сефарозой CL-4B. Двухцепочечные кДНК лигировали с ДНК плазмиды *pUC19-Hind*III, продуктами лигазной сшивки трансформировали клетки *E. coli* Δ M15. Отбор рекомбинантных клонов осуществляли последовательной гибридизацией колоний с ^{32}P ДНК фрагментами плазмид *p10*, *p4*, *p2*, *p15*, *p1*, *p5*, *p7*, *p37* и гибридизацией с ^{32}P РНК ТВЕ. Были получены клоны, несущие рекомбинантные плазмиды *p8*, *p11*, *p18*, *p210*, *p223*, *p22*, *p12*, *p17*, *p23*, *p9*.

Для уточнения структуры 3'-концевой части генома ТВЕ мы предприняли клонирование кДНК, полученной с вирусной РНК, несущей на 3'-конце poly(A)-тракт, который синтезировали с помощью полиаденилаттрансферазы *E. coli*. Синтез кДНК осуществляли с олиготимидилатной затравкой длиной 15–18 звеньев. Вторую цепь кДНК получали удлинением цепи *Hind*III-*Sal*GI-фрагмента ДНК-полимеразой А (цепь этого фрагмента комплементарна минус-цепи РНК ТВЕ в положениях 9923–10009 на последовательности, представленной на рис. 2, и была получена из ДНК плазмиды *p23*). Клонирование кДНК осуществляли в составе вектора *pUC19* после присоединения к кДНК *Hind*III-линкера.

Второй путь клонирования 3'-концевой части генома был по сути обратной вариацией предыдущего клонирования. Репликативную форму РНК ТВЕ после тепловой денатурации отжигали с цепью *Hind*III-*Sal*GI-фрагмента, комплементарной на участке 9923–10009 минус-цепи РНК, и осуществляли синтез кДНК с помощью обратной транскриптазы. К концам кДНК присоединяли oligo(dA)-последовательности при

помощи терминальной дезоксирибонуклеотидилтрансферазы из тимуса теленка. Вторую цепь кДНК синтезировали, используя (dT)₁₅₋₁₈ в качестве затравки в реакции, катализируемой ДНК-полимеразой А. К двухцепочечной кДНК присоединяли *Hind*III-линкер и клонировали ее в составе вектора pUC19.

После трансформации клеток *E. coli* Δ M15 отбор клонов осуществляли гибридизацией с [³²P]РНК ТВЕ и [³²P]ДНК-фрагментом плазмиды р9. В первом клонировании 3'-концевой части генома ТВЕ было отобрано 6 позитивных клонов, во втором — 2 клон. Анализ концевых структур ДНК-вставок этих рекомбинантных плазмид отличался от структуры вставки плазмиды р9 только наличием poly(A)-последовательности на концах вставок.

Первичную структуру фрагментов ДНК определяли модифицированным методом Максама — Гилберта [45, 46]. В качестве реакции специфической дегградации ДНК по цитидиновому звену использовали 0,1М раствор NaOH в гидразин-гидрате вместо солевого раствора, как рекомендовано в работе [45].

Нуклеотидную последовательность 5'-концевого участка генома ТВЕ определяли непосредственным секвенированием кДНК, синтезированной при использовании олигонуклеотидной затравки: d([³²P]CAGACCTACCA·TCAGGGT), комплементарной последовательности РНК ТВЕ на участке в положениях 376—393 (рис. 2).

Компьютерный анализ нуклеотидных и аминокислотных последовательностей был выполнен с использованием программ, описанных в работах [47, 48]. Расчет профиля гидрофобности полипротеина ТВЕ был выполнен по методу [41].

Авторы благодарны акад. Д. Г. Кнорре, акад. М. П. Чумакову (ИПВЭ АМН СССР), акад. Г. П. Георгиеву и И. М. Чумакову (ИМБ АН СССР) за интерес и поддержку работы, С. Г. Рубину, И. В. Семашко (ИПВЭ АМН СССР) и Е. К. Прессману за выполнение работ по типированию, выделению и очистке вируса клещевого энцефалита.

СПИСОК ЛИТЕРАТУРЫ

1. Westaway E. G. // The Togaviruses./Ed. Schlesinger R. W. N. Y.: Acad. Press, 1980. P. 531—581.
2. Rice C. M., Strauss E. G., Strauss J. H. // The Togaviridae and Flaviviridae./Eds Schleisinger S., Schleisinger M. J. N. Y.: Plenum Press, 1986. P. 279—326.
3. Чумаков М. П., Кусов Ю. Ю., Рубин С. Г., Сальников Я. А., Семашко И. В., Георгиев Г. П., Чумаков П. М., Грачев М. А., Шаманин В. А., Плетнев А. Г. // Биоорган. химия. 1983. Т. 9. № 2. С. 276—279.
4. Pletnev A. G., Yamshchikov V. F., Blinov V. M. // FEBS Lett. 1986. V. 200. № 2. P. 317—321.
5. Yamshchikov V. F., Pletnev A. G. // Nucl. Acids Res. 1988. V. 16. № 15. P. 7750.
6. Rice C. M., Aebersold R., Teplov D. B., Pata J., Bell J. R., Vorndam A. V., Trent D. W., Brandriss M. W., Schlesinger J. J., Strauss J. H. // Virology. 1986. V. 151. № 1. P. 1—9.
7. Speight G., Coia G., Parker M. D., Westaway E. G. // J. Gen. Virol. 1988. V. 69. № 1. P. 23—34.
8. RSSE-TBE. Prototype strain. // International Catalogue of Arboviruses. Sofjin. 1975. P. 607.
9. Плетнев А. Г., Ямщиков В. Ф., Блинов В. М. // Биоорган. химия. 1986. Т. 10. № 9. С. 1189—1202.
10. Brinton M. A., Dispoto J. H. // Virology. 1988. V. 162. № 2. P. 290—299.
11. Castle E., Wengler G. // Arch. Virol. 1987. V. 92. № 2. P. 309—313.
12. Takegami T., Washizu M., Yasui K. // Virology. 1986. V. 152. № 2. P. 483—486.
13. Brinton M. A., Fernandez A. A., Dispoto J. H. // Virology. 1986. V. 153. № 1. P. 113—121.
14. Sumiyoshi H., Mori C., Fuke I., Morita K., Kuhara S., Kondou J., Kukushi Y., Nagamatsu H., Igarashi A. // Virology. 1987. V. 161. № 2. P. 497—510.
15. Deubel V., Kinney R. M., Trent D. W. // Virology. 1988. V. 165. № 1. P. 234—244.
16. Coia G., Parker M. D., Speight G., Byrne M. E., Westaway E. G. // J. Gen. Virol. 1988. V. 69. № 1. P. 1—21.
17. Castle E., Leidner U., Nowak T., Wengler G., Wengler G. // Virology. 1986. V. 149. № 1. P. 10—26.
18. Castle E., Nowak T., Leidner U., Wengler G., Wengler G. // Virology. 1985. V. 145. № 1. P. 227—236.

19. Wengler G., Castle E., Leidner U., Nowak T., Wengler G. // *Virology*. 1985. V. 147. № 1. P. 264-284.
20. Trent D. W., Kinney R. M., Johnson B. J. B., Vorndam A. V., Grant J. A., Deubel V., Rice C. M., Hahn C. // *Virology*. 1987. V. 156. № 2. P. 293-304.
21. 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.
22. McAda P. C., Mason P. W., Schmaljahn C. S., Dalrymple J. M., Mason T. L., Fournier M. J. // *Virology*. 1987. V. 158. № 2. P. 348-360.
23. Boege U., Heinz F. X., Wengler G., Kunz C. // *Virology*. 1983. V. 126. № 2. P. 651-657.
24. Барам Г. И., Грачев М. А., Назимов И. В., Плетнев А. Г., Прессман Е. К., Рубин С. Г., Сальников Я. А., Семашко И. В., Чумаков М. П., Шемякин В. В., Ямщиков В. Ф. // *Биоорганическая химия*. 1985. Т. 11. № 12. С. 1677-1680.
25. 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.
26. Svitkin Y. V., Lyapustin V. N., Lashkevich V. A., Agol V. I. // *Virology*. 1984. V. 135. № 2. P. 536-541.
27. Wright P. J., Warr H. M., Westaway E. G. // *Virology*. 1980. V. 104. № 2. P. 482-486.
28. Wright P. J., Warr H. M., Westaway E. G. // *Virology*. 1981. V. 109. № 2. P. 418-427.
29. Dalgarno L., Trent D. W., Strauss J. H., Rice C. M. // *J. Mol. Biol.* 1986. V. 187. № 2. P. 309-323.
30. Zhao B., Mackow E., Buckler-White A., Markoff L., Chanock R. M., Lai Ch.-J., Makino Y. // *Virology*. 1986. V. 155. № 1. P. 77-88.
31. Mason P. W., McAda P. C., Mason T., Fournier M. J. // *Virology*. 1987. V. 161. № 1. P. 262-267.
32. Mackow E., Makino Y., Zhao B., Zhang Y.-M., Markoff L., Buckler-White A., Guiler M., Chanock R., Lai Ch.-J. // *Virology*. 1987. V. 159. № 2. P. 217-228.
33. Heinz F. X., Kunz C. // *J. Gen. Virol.* 1981. V. 57. № 1. P. 263-274.
34. Heinz F. X., Kunz C. // *J. Gen. Virol.* 1982. V. 62. № 1. P. 271-285.
35. Smith G. W., Wright P. J. // *J. Gen. Virol.* 1985. V. 66. № 2. P. 559-571.
36. Schlesinger J. J., Brandriss M. W., Walsh E. E. // *J. Immunol.* 1985. V. 135. № 4. P. 2805-2809.
37. Schlesinger J. J., Brandriss M. W., Walsh E. E. // *J. Gen. Virol.* 1987. V. 68. № 2. P. 853-857.
38. Wright P. J., Warr H. M. // *J. Gen. Virol.* 1985. V. 66. № 2. P. 597-601.
39. Winkler G., Randolph V. B., Cleaves G. R., Ryan T. E., Stollar V. // *Virology*. 1988. V. 162. № 1. P. 187-196.
40. Chu P. W. G., Westaway E. G. // *Virology*. 1985. V. 140. № 1. P. 68-79.
41. Kyte J., Doolittle R. F. // *J. Mol. Biol.* 1982. V. 157. № 1. P. 105-132.
42. Hahn Y. S., Galler R., Hunkapiller T., Dalrymple J. M., Strauss J. H., Strauss E. G. // *Virology*. 1988. V. 162. № 1. P. 167-180.
43. Чумаков М. П., Кусов Ю. Ю., Рубин С. Г., Семашко И. В., Сальников Я. А., Рейнгольд В. Н., Прессман Е. К., Цехановская Н. А. // *Вопросы вирусологии*. 1984. Т. 29. № 6. С. 694-701.
44. Maniatis T., Fritsch E., Sambrook J. // *Молекулярное клонирование: Пер. с англ.* М.: Мир, 1984. С. 107-240.
45. Maxam A., Gilbert W. // *Proc. Nat. Acad. Sci. USA*. 1977. V. 74. № 2. P. 560-564.
46. Чуепило С. А., Кравченко В. В. // *Биоорганическая химия*. 1983. Т. 9. № 12. С. 1634-1637.
47. McLachlan A. D. // *J. Mol. Biol.* 1971. V. 61. № 2. P. 409-424.
48. Stalen R. // *Nucl. Acids Res.* 1980. V. 8. № 10. P. 3673-3694.

Поступила в редакцию
30.I.1988

После доработки
13.IV.1989

NUCLEOTIDE SEQUENCE OF THE GENOME AND COMPLETE AMINO ACID SEQUENCE OF THE POLYPROTEIN OF THE TICK-BORNE ENCEPHALITIS VIRUS

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

*Novosibirsk Institute of Bioorganic Chemistry, Siberian
Division of the Academy of the USSR, Novosibirsk;*

* *All-Union Research Institute of Molecular Biology,
Kol'tsovo, Novosibirsk Region*

We have cloned and sequenced RNA encoding all virion and nonstructural proteins of tick-borne encephalitis virus (TBEV). Its length is 10 477 bases with a single open

reading frame (nucleotides 127--10 363) encoding 3412 amino acids. The 5'- and 3'-non-coding regions have stem- and- loop structure. The polyprotein precursor is proteolytically cleaved, apparently, by a mechanism resembling that proposed for the expression of polyproteins of other flaviviruses, such as yellow fever, West Nile and Kunjin viruses. The deduced TBEV gene order is 5'-C-preM (M)-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'. The genome and the polyprotein of TBEV and other flaviviruses appears to be structurally similar, although these flaviviruses are transmitted to and from their vertebrate hosts by different carriers, such as ticks or mosquitoes. Analysis of sequence homologies of polyproteins of flaviviruses suggests that TBEV is more closely related to yellow fever virus than to other serological subgroups of flaviviruses (West Nile or Dengue viruses). The hydrophobic profiles of the flaviviruses are highly conservative. Nonstructural proteins NS2A, NS2B, NS4A and NS4B are extremely hydrophobic, suggesting that they are likely to be associated with cellular membranes. Proteins E, NS1, NS3 and NS5 are the most conservative and may be involved in general enzymatic activities related to viral replication and virion assembly.

