

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ  
КАФЕДРА ОБЩЕЙ ХИМИИ

**В. В. ХРУСТАЛЁВ**

**БИОХИМИЧЕСКИЕ МЕХАНИЗМЫ  
МУТАЦИОННОГО ДАВЛЕНИЯ  
В МЕТОДОЛОГИИ  
ВЫЧИСЛИТЕЛЬНОЙ БИОЛОГИИ**

Монография

Под редакцией Е. В. Барковского



Минск БГМУ 2010

УДК 577.1:575.224/.23

**Хрусталеv, В. В.** Биохимические механизмы мутационного давления в методологии вычислительной биологии : монография / В. В. Хрусталеv ; под ред. Е. В. Барковского. – Минск : БГМУ, 2010. – 212 с. – ISBN 978–985–528–110–9.

Посвящена описанию многочисленных биохимических аспектов мутационного давления, биоинформатических методов его изучения и применению знаний о данных аспектах в различных областях биологии и медицины. Актуальность издания обусловлена тем, что описанные в ней методы впервые представлены в систематизированном и расширенном виде. До момента издания монографии сведения о биохимических аспектах мутационного давления, в том числе и о тех, которые были раскрыты непосредственно автором, можно было найти исключительно в научных статьях, опубликованных в международных рецензируемых журналах на английском языке. Представляет интерес не только для исследователей, но и для врачей-инфекционистов, эпидемиологов. Последние смогут найти на страницах издания новейшие сведения о патогенезе ВИЧ, герпетической инфекции и аутоиммунных осложнений стрептококковой инфекции.

Предназначена для студентов, аспирантов, соискателей и преподавателей биологических и медицинских вузов, генетиков, биохимиков, эпидемиологов и врачей-инфекционистов.

Табл. 27. Ил. 58. Библиогр. назв. 135.

Рекомендовано Советом Бел. гос. мед. универ.

Р е ц е н з е н т ы :

академик НАН Беларуси, д-р биол. наук, проф. Института генетики и цитологии НАН Беларуси Л. В. Хотылёва; зав. каф. радиационной медицины и экологии Белорусского государственного медицинского университета, д-р биол. наук, проф. А. Н. Стожаров

---

Научное издание

**Хрусталеv Владислав Викторович**

**БИОХИМИЧЕСКИЕ МЕХАНИЗМЫ МУТАЦИОННОГО ДАВЛЕНИЯ  
В МЕТОДОЛОГИИ ВЫЧИСЛИТЕЛЬНОЙ БИОЛОГИИ**

Монография

Ответственный за выпуск Е. В. Барковский  
В авторской редакции  
Компьютерная верстка Н. М. Федорцовой  
Корректор Ю. В. Киселёва

Подписано в печать 09.12.09. Формат 60×84/16. Бумага писчая «Кюм Люкс».

Печать офсетная. Гарнитура «Times».

Усл. печ. л. 12,32. Уч.-изд. л. 12,44. Тираж 100 экз. Заказ 121.

Издатель и полиграфическое исполнение:

учреждение образования «Белорусский государственный медицинский университет».

ЛИ № 02330/0494330 от 16.03.2009.

ЛП № 02330/0150484 от 25.02.2009.

Ул. Ленинградская, 6, 220006, Минск.

ISBN 978–985–528–110–9

© Хрусталеv В. В.

© Оформление. Белорусский государственный  
медицинский университет, 2010

## Введение

Данная монография посвящена изучению биохимических процессов, вызывающих мутационное давление в генах и геномах вирусов (вирусов иммунодефицита человека 1-го и 2-го типа, вирусов простого герпеса человека 1-го и 2-го типа, варицелло-зостер-вируса и др.) и бактерий (золотистого стрептококка), и тех многочисленных последствий (биологических, иммунологических и собственно биохимических), к которым они приводят. Актуальность издания обусловлена также и тем, что методы исследования в области биоинформатики и вычислительной биологии, созданные автором, впервые представлены в систематизированном и расширенном виде. До момента издания монографии о большинстве этих оригинальных методов можно было узнать исключительно из научных статей, опубликованных в англоязычных международных рецензируемых журналах.

Исследования в области молекулярной эволюции, эволюционной биохимии и биоинформатики на кафедре общей химии Белорусского государственного медицинского университета были начаты в 2000 г. по инициативе и под научным руководством доктора биологических наук, профессора Барковского Евгения Викторовича, лауреата Государственной премии Республики Беларусь. За это время по данной теме было опубликовано свыше 200 научных работ, в том числе в таких журналах как *IUBMB Life* (Impact factor 2008: 2.9), *Computational Biology and Chemistry* (Impact factor 2008: 1.8), *Immunological Investigations* (Impact factor 2008: 1.8). Кроме этого, под редакцией проф. Е. В. Барковского были изданы 4 монографии.

На кафедре общей химии БГМУ ведется разработка оригинальных методов исследования, основанных на теориях нейтральной эволюции М. Кимуры и мутационного давления Н. Суеоки. В 2008 г. автор данной монографии совместно с Е. В. Барковским опубликовал в журнале *IUBMB Life* статью, внесшую существенный вклад в развитие теории мутационного давления: было введено новое биологическое понятие о *силе мутационного давления*. В 2009 г. В. В. Хрусталёв и Е. В. Барковский опубликовали статью в журнале *Computational Biology and Chemistry*, в которой были введены критерии определения силы мутационного давления в полном вирусном геноме, использован оригинальный метод выявления лидирующего биохимического механизма (причины) мутационного давления. Первая в истории молекулярной биологии статья о влиянии мутационного давления на иммуногенность белков (написанная автором данной монографии) вышла в 2009 г. в журнале *Immunological Investigations*.

Автор впервые применил теорию мутационного давления для изучения жизненных циклов и патогенеза возбудителей, вызывающих серьезные заболевания (СПИД, герпетическую инфекцию, опоясывающий лишай, аутоиммунные осложнения стрептококковой инфекции). Однако изложенные в дан-

ной монографии оригинальные методы могут быть с успехом применены и к изучению других возбудителей и вызываемых ими заболеваний. В кратко описанных выше статьях, так же как и в других научных публикациях автора данной монографии, содержатся основы нового направления в развитии методологии вычислительной биологии.

Монография представляет интерес не только для научных работников, но и для практикующих врачей-инфекционистов, эпидемиологов, которые смогут найти на страницах издания новейшие сведения о патогенезе ВИЧ, герпетической инфекции, о развитии аутоиммунных осложнений стрептококковой инфекции.

В первой части монографии изложены основы теорий нейтральности М. Кимуры и мутационного давления Н. Суеоки. Кроме того, в гл. 3 описаны основные направления научной критики данных теорий, приведены собственные взгляды автора на многие вопросы молекулярной эволюции.

Во второй части описаны оригинальные методы определения направления, силы и ведущего механизма мутационного давления как в отдельных генах, так и в полных геномах на примере альфа-герпесвирусов и вирусов иммунодефицита человека 1-го и 2-го типов.

В третьей части изложены основы методов определения степени мутабельности кодирующих участков в полных геномах, а также конкретных генов и отдельных участков генов, базирующиеся на результатах применения методов, описанных в первой части. В качестве примера были выбраны гены *env* ВИЧ1 и ВИЧ2, гены, кодирующие М-белок различных штаммов стрептококка группы А.

Четвертая часть посвящена методам изучения степени нейтральности и консервативности аминокислотных замен, вызываемых мутационным давлением (как в целом по протеому, так и в отдельных белках), опять же, основанным на результатах применения методов, описанных в первой части. В четвертой части описаны исследования не только вирусных белков и протеомов, но и препроинсулина хордовых животных (в том числе и человека). Это еще раз подчеркивает, что методы биоинформатического анализа, созданные автором в процессе изучения вирусных и бактериальных генов, могут быть успешно применены и для изучения генов человека.

Пятая часть посвящена методологии исследования, позволяющего подтвердить результаты применения методов, описанных в трех предыдущих частях монографии.

Шестая часть содержит краткое описание и инструкции к применению шести оригинальных компьютерных алгоритмов, которые были созданы автором для максимального упрощения применения описанных в монографии методов. Эти алгоритмы доступны на сайте: [www.barkovsky.hotmail.ru](http://www.barkovsky.hotmail.ru).

В конце издания размещен глоссарий, содержащий трактовку большого количества молекулярно-биологических терминов, и таблицы генетического кода (прил. 1–3), помогающие пониманию биологического смысла многих показателей, использованных в работе.

# Часть первая

## Теория мутационного давления

### Глава 1

#### Предмет и задачи молекулярной биологии, геномики, протеомики, биоинформатики и вычислительной биологии

Материал, изложенный на страницах данного издания, во многом относится к таким научным дисциплинам, как молекулярная биология, геномика, протеомика, биоинформатика и вычислительная биология, но не только к ним. Результаты исследований, методы проведения которых описаны в последующих главах, могут быть использованы в медицине, биологии, генетике, иммунологии, биохимии. Однако разъяснить предмет и задачи этих пяти дисциплин не представляется целесообразным. Гораздо более оправданно раскрыть содержание пока еще недостаточно известных научных направлений.

**Молекулярная биология** — это наука, занимающаяся изучением биологии на молекулярном уровне. Фактически, молекулярная биология оформилась как научная дисциплина на стыке генетики (отрасли биологии) и биохимии (отрасли химии) [23]. Предмет молекулярной биологии — «молекулы жизни»: ДНК, РНК, белки и их компоненты. Задачей молекулярной биологии является изучение взаимодействий молекул генетических биополимеров и изменений, возникающих в них (молекулярной эволюции) [23].

Внутри молекулярной биологии существует множество более узких дисциплин. Одной из них является геномика. **Геномика** занимается изучением геномов живых организмов (то есть таких биополимеров, как ДНК и РНК).

Сразу хочется уточнить, что значение термина «геном» в молекулярной биологии несколько отличается от его значения в биологии вообще. Традиционно под термином «геном» понимали совокупность генов, содержащихся в гаплоидном наборе хромосом организмов данного вида (но не особи) [1, 13]. В настоящее время в молекулярной биологии этот термин может использоваться для обозначения полного набора генов не только какого-либо вида организмов, но и отдельной особи. Как правило, в электронных банках биологической информации содержатся сведения о полном наборе генов отдельной особи данного вида (или подвида, штамма), которые переносятся на представителей всего вида (или подвида, штамма). Такой геном называется референтным.

Геномика включает в себя исследования, направленные на определение нуклеотидной последовательности полного генома данного вида (организма), картирование этого генома (поиск открытых рамок считывания,

регуляторных и повторяющихся последовательностей) и изучение таких феноменов, как *гетерозис*, *эпистаз*, *плейотропизм* и других эффектов, возникающих от взаимодействия генов (на системном уровне) [15].

Что касается первой задачи геномики (определения нуклеотидных последовательностей полных геномов), то она выполнена в гораздо большей степени, чем все остальные. На момент написания данной главы в GenBank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) находились записи о нуклеотидном составе полных геномов 2266 видов вирусов (всего 3449 записей о полных вирусных геномах), записи, относящиеся к геномам 2087 видов эукариот (всего 1187 записей о нуклеотидном составе хромосом, 2088 — о нуклеотидном составе генетического аппарата органелл и 42 — о нуклеотидном составе генетического аппарата плазмид), записи о полных геномах 1279 видов бактерий (включая записи о 2000 полностью просеквенированных бактериальных «хромосомах» и 1893 плазмидах), сведения о полных геномах 82 видов архей (83 «хромосомы» и 63 плазмиды), а также полные геномы 40 видов виридов и 38 плазмид, не относящихся к геному какого-либо конкретного вида организмов.

Вторая задача геномики — картирование геномов и поиск открытых рамок считывания — не всегда «успевает» за выполнением первой. Большинство записей об аминокислотном составе белков было «предсказано» с помощью различных компьютерных алгоритмов на основании наличия открытых рамок считывания в полностью просеквенированных геномах и положительных результатов проведения филогенетического анализа (поиска известных белков, похожих на предположительный продукт предсказанного гена). Такие данные нуждаются в экспериментальном подтверждении.

Что касается изучения феноменов, возникающих при взаимодействии генов, то тут дело обстоит еще сложнее. В данном издании мы как раз и касаемся феномена, возникающего при взаимодействии генов (точнее — при взаимодействии продуктов так называемых «генов-мутаторов» непосредственно с геномной ДНК) и влияющих на все гены в геноме (или в части генома) данного вида организмов [121]. Речь идет о **мутационном давлении**. Для большей части генов данного вида бактерии, археи или вируса характерно общее направление и сила мутационного давления [30, 31, 32, 33, 84, 85, 96, 121]. Этот факт дает нам основание причислить изучение мутационного давления к одной из целей геномики.

**Протеомика** — это научная дисциплина, изучающая все белки данного организма (или вида организма) в совокупности [15, 93]. Протеом — это совокупность всех белков (по аналогии с геномом — совокупностью всех генов) данного организма (или вида организмов) [20].

Заведующий лабораторией протеомики Института биофизики и клеточной инженерии НАН Республики Беларусь, д-р биол. наук, чл.-кор.

НАН Республики Беларусь В. М. Мажуль определял эту научную дисциплину следующим образом: «протеомика — это отрасль биологической науки, изучающая экспрессированные геномом белки — их состав, структуру, функциональные свойства, механизмы регуляции активности, взаимодействия белков друг с другом» [20].

Протеомика — гораздо менее изученная область знаний, чем геномика. Как оказалось, далеко не каждая мРНК подвергается трансляции (существует тонкая внутриклеточная регуляция не только на уровне транскрипции, но и на уровне трансляции) [23]. В то же время, с одного и того же гена могут транскрибироваться несколько вариантов мРНК (за счет альтернативного сплайсинга) [23].

После того как были обнаружены случаи несоответствия аминокислотного состава белков нуклеотидному составу кодирующих их участков, была открыта система редактирования РНК [62]. Функцией нескольких семейств клеточных ферментов является внесение нуклеотидных замен в мРНК [62, 115]. Кроме того, давно известно о системе посттрансляционной модификации белков (например, как в случае с препроинсулином), о системе их фосфорилирования, гликозилирования, расщепления в клетке [23].

В данном издании мы касаемся эффектов мутационного давления, которые распространяются на все белки в геномах бактерий, архей и вирусов, то есть на полные протеомы [30, 31, 32, 33, 84, 85].

**Биоинформатика** — это научная дисциплина, занимающаяся применением математических методов к биологическим процессам [15]. Термин «биоинформатика» появился в 70-х гг. XX в. Хотя собственно биоинформатикой занимались и до того. Достаточно вспомнить о многочисленных биологических исследованиях, в которых использовалась теория информации Клода Шеннона, созданная в 40-е гг. XX в. [135]. В последние же два десятка лет биоинформатика достигла своего расцвета. Связано это с усовершенствованием вычислительной техники (с повышением ее производительности и доступности) и с обилием биологической информации, ждущей математической обработки и систематизации (банки генетической информации, сведений о функциях и структуре белков достигли небывалых объемов).

Обычно биоинформатику объединяют с еще одной научной дисциплиной — *вычислительной биологией*. Четкие границы между ними провести невозможно, но все-таки биоинформатика скорее решает вопросы чисто математические (программирование, моделирование, составление баз данных и т. д.), а вычислительная биология — теоретические (то есть собственно биологические). Важно соблюдать субординацию в биоинформатических исследованиях. Математика, статистика и программирование должны всего лишь обслуживать интересы биологии. Когда ситуация ста-

новится обратной, результат вызывает безграничное удивление у любого человека с биологическим или медицинским образованием.

В данном издании мы затрагиваем такие направления биоинформатики, как анализ нуклеотидных последовательностей генов и аминокислотных последовательностей белков, предсказание вторичной и третичной структуры белка, а также расположения линейных и пространственных эпитопов на его поверхности, выравнивание нуклеотидных и аминокислотных последовательностей, филогенетический анализ.

*Основной задачей биоинформатики и вычислительной биологии является достижение недоступного ранее уровня понимания биологических процессов.*

Как уже было сказано выше, прежде чем приступить к описанию экспериментов, проведенных при помощи оригинальных биоинформатических методов (вторая–пятая части), необходимо изучить биологические теории, оправдывающие их применение (главы 2–4 первой части).

## Глава 2 Теория нейтральной эволюции м. Кимуры

Теория нейтральной эволюции, созданная Муто Кимурой (1924–1994), является фундаментальной для молекулярной биологии и молекулярной эволюции. Главный принцип данной теории — признание *отрицательного отбора* и *случайного генетического дрейфа* основными факторами эволюции, действующими на молекулярном уровне [16].

Приводим пять основных положений данной теории:

1. Скорость эволюции любого белка, выраженная через число аминокислотных замен на сайт в год, приблизительно постоянна и одинакова в разных филогенетических линиях, если только функция и третичная структура этого белка остаются практически неизменными.

2. Функционально менее важные молекулы и их части эволюционируют (накапливают мутационные замены) быстрее, чем более функционально важные.

3. Мутационные замены, приводящие к меньшим нарушениям структуры и функции молекулы (консервативные замены), в ходе эволюции происходят чаще тех, которые существенно нарушают структуру и функции данной молекулы (радикальных замен).

4. Появлению нового в функциональном отношении гена всегда должна предшествовать дупликация.

5. Селективная элиминация вредных мутаций (за счет отрицательного отбора) и случайная фиксация селективно нейтральных или очень слабо вредных мутаций (за счет случайного генетического дрейфа) происходят



в ходе эволюции гораздо чаще, чем положительный дарвиновский отбор благоприятных мутаций.

Может показаться, что первое и второе положение теории М. Кимуры противоречат друг другу. В первом положении М. Кимура говорит о том, что скорость эволюционных изменений всех белков в филогенетических рядах примерно одинакова, а именно, равняется  $10^{-9}$  фиксированным аминокислотным заменам на сайт в год [16]. Второе же положение вытекает из оговорки, входящей в состав первого: если белок не является функционально важным, то аминокислотные замены в нем накапливаются быстрее, чем в функционально важном и, следовательно, «законсервированном» отрицательным отбором белке [16]. Как показали многочисленные исследования, скорость эволюции различных белков действительно может варьировать, как правило, в пределах одного порядка относительно средней скорости, найденной М. Кимурой [105].

Третье положение объясняет, почему менее важные в функциональном смысле белки и их части накапливают больше аминокислотных замен, чем функционально важные, за тот же промежуток времени. Нейтральные замены фиксируются в популяции (то есть «происходят в ходе эволюции») гораздо чаще, чем радикальные (приводящие к серьезным изменениям структуры и/или функции белка). О фиксации замены можно говорить в том случае, если частота ее встречаемости среди особей в популяции приблизилась к 100 % [16]. Если белок менее функционально значим, то более серьезные структурные перестройки в нем могут оказаться нейтральными или слабо вредными для жизнеспособности организма, а, следовательно, не будут подвергаться отрицательному отбору и получают шанс зафиксироваться в популяции.

Четвертое положение является производным от существовавшей до создания теории нейтральности гипотезы (гипотезы Гудмена) о ведущей роли дубликации генов в увеличении размеров геномов и в появлении семейств структурно схожих белков. В своей же теории М. Кимура уточняет, что новый в функциональном отношении ген не может возникнуть из ранее существовавшего, не обладавшего вновь появившейся функцией. После дубликации одна из копий гена, кодирующего некий белок, остается под влиянием отрицательного отбора, поскольку продолжает быть вовлеченной в сложнейшие механизмы внутриклеточной регуляции. Вторая же копия, как правило, не подвергается тем же механизмам внутриклеточной регуляции, что и первая (исходная), в результате чего замены в этой копии гена начинают накапливаться быстрее [16].

Дальнейшую судьбу дублицированных генов (и их исходных копий) раскрыла в своей книге «Генетические механизмы прогрессивной эволюции» С. Оно, внесшая немалый вклад в создание теории нейтральности. До тех пор пока жизненно важная функция контролируется только одним

геном, естественный отбор будет препятствовать закреплению мутаций, изменяющих активные центры кодируемого этим геном полипептида [21]. По мнению С. Оно: «... только накопление запрещенных мутаций в активных центрах может привести к изменению основных свойств белка и обеспечить превращение его в новый белок, а естественный отбор игнорирует серьезные изменения только в избыточных копиях, в результате чего именно в дубликатах накапливаются ранее запрещенные мутации; именно этот процесс и может приводить (с очень низкой вероятностью — *прим. автора данной монографии*) к возникновению нового гена, кодирующего белок с несуществовавшей ранее функцией» [21].

Нельзя не упомянуть о наиболее распространенных причинах возникновения дубликаций генетического материала. Универсальным механизмом возникновения дубликации является «проскальзывание синтезируемой цепи ДНК с последующей репарацией» [23]. Если в цепи ДНК на определенном расстоянии друг от друга находятся два прямых повтора, то сразу же после своего появления на вновь синтезированной цепи второй повтор может спариться с участком на комплементарной цепи, соответствующим первому. Репликация в таком случае продолжается, и на вновь синтезированной цепи появляются три прямых повтора и дубликат участка цепи, располагавшегося между повторами. Последующая репарация достраивает участок цепи напротив дубликата. Между двумя прямыми повторами могут оказаться как несколько генов, так и участок одного гена [23].

Немалый вклад в дубликацию вносят мобильные элементы (транспозоны и ретротранспозоны). Ретротранспозоны — это мобильные элементы, кодирующие (кроме остальных белков) обратную транскриптазу [23]. Обратная транскриптаза строит копию ДНК на матрице РНК, транскрибируемой с ретротранспозона. Синтезированная ДНК затем встраивается в другой локус хромосомы при помощи интегразы. Обратная транскриптаза может построить копию ДНК и на любой другой молекуле мРНК. Если это происходит, то копия кодирующего участка попадает в далекий от родительской копии гена локус хромосомы (или в другую хромосому) [23]. Особенностью таких дубликатов (ретрогенов) у эукариот является то, что в их состав зачастую не входят интроны, но включаются полиаденилированные концевые последовательности (поскольку синтез ДНК обратной транскриптазой ретротранспозонов чаще всего происходит на матрице полностью процессированной мРНК). Образование ретрогена может происходить и на матрице частично процессированной мРНК, содержащей меньшее количество интронов, чем ген, с которого она транскрибировалась [23].

Транспозоны (мобильные элементы, не кодирующие обратную транскриптазу) содержат на своих концах инвертированные повторы [23]. Если два транспозона размещаются недалеко друг от друга, то расположенные между ними гены рискуют быть вырезанными из одной гомологичной

хромосомы и перенесенными в другую в результате «неравномерного кроссинговера» [23].

Известно, что ДНК любого организма обогащена как прямыми, так и инвертированными повторяющимися последовательностями. Взаимодействие между инвертированными повторами (то есть комплементарное спаривание) может происходить и в том случае, если они не являются абсолютно идентичными [23]. Именно на этом комплементарном взаимодействии и основываются основные механизмы дупликации. Сам собой напрашивается вывод о том, что молекуле ДНК по описанным выше биохимическим причинам присуще такое свойство, как постепенное увеличение своего размера за счет дупликаций различного объема. Недавние филогенетические исследования показали, что дупликации могут подвергаться не только сравнительно короткие, но и довольно длинные участки эукариотических хромосом [79]. Известная фраза Н. В. Тимофеева-Рессовского о том, что «неизбежность эволюции следует из основных свойств живой материи» [24] имеет вполне конкретный биохимический смысл.

Перед тем как подробно остановиться на рассмотрении пятого положения, необходимо четко отграничить друг от друга два уровня, на которых могут действовать факторы эволюции.

Существует известный всем фактор эволюции — естественный отбор, открытый Ч. Дарвином [1]. Этот фактор оперирует на фенотипическом уровне. Естественный отбор основан на борьбе за существование. Борьба за существование происходит внутри популяции (наиболее ожесточенная), с другими биологическими видами и с факторами окружающей среды. Существование популяции поддерживается за счет наследственности. Особи внутри популяции отличаются друг от друга за счет изменчивости. Сразу хочется подчеркнуть, что изменчивость бывает не только генотипической, но и фенотипической. Каждый организм обладает генетически детерминированными пределами своей индивидуальной фенотипической изменчивости (нормой реакции). В борьбе за существование чаще побеждают особи, обладающие определенными фенотипическими признаками, помогающими им выжить в данных конкретных условиях. Выжившие особи оставляют потомство (если оказываются к этому способными). Если признак, обеспечивший их выживание, является наследуемым, частота встречаемости этого признака среди особей в популяции постепенно возрастает [1]. Под термином «естественный отбор» Ч. Дарвин понимал процесс, приводящий к приспособлению *популяции в целом* к условиям среды [24].

Еще раз подчеркнем, что естественный отбор ведется по критическим фенотипическим признакам [30]. Число таких признаков, как правило, невелико. Классический пример — темный цвет крыльев бабочек, который помогал им прятаться от птиц в темном и туманном пейзаже наполненного смогом Лондона XIX в. Выживали, в основном, бабочки с темными крыль-

ями, а бабочки со светлыми крыльями, как правило, погибали. При этом на выживание темнокрылых бабочек практически не влиял ни размер этих крыльев, ни их форма, ни даже количество конечностей на теле бабочки. Фенотипический признак — темный цвет крыльев — детерминируется небольшим количеством генов (если не одним). Соответственно, если темнокрылость обусловлена мутацией в одном единственном гене, то естественный (положительный) отбор на фенотипическом уровне привел к фиксации этой генной мутации в популяции лондонских бабочек. А что же произошло со всеми остальными генами?

Сам Ч. Дарвин считал, что отбор в идеале должен проходить по всем фенотипическим признакам [24]. В его времена не знали о существовании генов и хромосом, рецессивных и доминантных признаков, совместном наследовании генов, расположенных на одной хромосоме, кроссинговере и его вероятности, в конце концов, о том, что лишь малая часть генов постоянно проявляет себя на фенотипическом уровне [1, 16, 21, 24, 42, 123, 131]. Нельзя забывать и о том, что гены, продукты которых обуславливают жизнеспособность сперматозоида и яйцеклетки (гаплоидной формы жизни *homo sapiens*), и гены, продукты которых обуславливают жизнеспособность (в широком смысле слова) человека (диплоидной формы жизни *homo sapiens*), — это абсолютно разные гены. Поэтому основная суть теории эволюции Ч. Дарвина применима исключительно для описания естественного отбора на фенотипическом уровне.

В принципе, разногласий между теориями Ч. Дарвина и М. Кимуры не существует, поскольку одна из них описывает эволюцию на фенотипическом уровне, а другая — на молекулярном [16]. Примеры воздействия естественного отбора на эволюцию, происходящую на молекулярном уровне, полностью убеждают нас в правоте М. Кимуры: «Селективная элиминация вредных мутаций (за счет отрицательного отбора) и случайная фиксация селективно нейтральных или очень слабо вредных мутаций (за счет случайного генетического дрейфа) происходят в ходе эволюции гораздо чаще, чем положительный дарвиновский отбор благоприятных мутаций (хочется добавить — происходящий на фенотипическом уровне по отдельным критическим признакам)» [16].

Интересно отметить, что даже один из критиков теории нейтральности Р. Левонтин утверждал, что «так называемая гипотеза нейтральных мутаций — это на самом деле модернизированная классическая гипотеза Дарвина о структуре популяции и эволюции» [17].

После приведенного выше вступления читатель подготовлен к пониманию того, что представляют из себя основные факторы эволюции на генотипическом уровне: *отрицательный отбор* и *случайный генетический дрейф* [16].

В синтетической теории эволюции естественный отбор на фенотипическом уровне принято разделять на три типа: *дизруптивный*, *направленный* и *стабилизирующий* [1]. Дизруптивный отбор приводит к тому, что популяция разделяется на несколько групп, обладающих различными вариантами данного фенотипического признака, а исходный вариант признака элиминируется [1]. Направленный отбор предполагает то, что «положительным» оказывается вновь возникший (менее распространенный) вариант фенотипического признака, а предшествующий вариант (наиболее распространенный) — элиминируются (при этом популяция не расщепляется, как при дизруптивном отборе) [1, 42]. Отрицательный отбор часто отождествляют с отбором стабилизирующим. Последний приводит к элиминации всех вновь возникающих вариантов фенотипического признака и к «консервации» наиболее распространенного варианта [1, 42].

Под термином же «отрицательный отбор на фенотипическом уровне» следует понимать отбор, направленный против определенных вариантов фенотипического признака (то есть обладатели определенного варианта признака часто оказываются неспособными воспроизводить потомство и элиминируются естественным отбором в результате борьбы за существование). В случае стабилизирующего отбора отбор отрицательный направлен на все варианты признака, за исключением наиболее распространенного [42]; в случае направленного и дизруптивного отбора — на ранее существовавший вариант признака [1, 42]. При направленном отборе, в отличие от дизруптивного, популяция не расщепляется на несколько групп, обладающих разными вновь возникшими вариантами признака [1]. Несмотря на сходство в названиях и в результатах воздействия, следует различать отрицательный отбор на фенотипическом и на молекулярном уровне.

Отрицательный отбор на молекулярном уровне (по М. Кимуре) воздействует на белок, а следовательно, и на кодирующий его ген, препятствуя фиксации мутаций (нуклеотидных замен, делеций и инсерций в гене), приводящих к определенным аминокислотным заменам в белке [16], а не на вариант фенотипического признака. Можно сказать, что все мутации в данном гене, приводящие к неспособности мутантной особи воспроизводить потомство или к существенному снижению этой способности, элиминируются из популяции *отрицательным отбором на молекулярном уровне* [16]. Особи с «вредными» (радикальными) мутациями в данном белке чаще всего погибают до вступления в репродуктивный возраст или же вообще не обладают способностью к размножению. Степень вреда, наносимого данной конкретной мутацией, во многом зависит от особенностей внутренней среды организма (от набора всех остальных генов, продукты которых обуславливают способность адаптации к последствиям мутации) [42].

Кроме мутаций, приводящих к гибели организма, к существенному снижению его жизнеспособности, к невозможности или снижению способности к размножению, существуют мутации, которые не влияют на жизнеспособность вообще или влияют на нее незначительно. Именно такие мутации, по мнению М. Кимуры, чаще всего и фиксируются в популяции путем случайного генетического дрейфа [16].

*Случайный генетический дрейф* — сборное понятие, включающее в себя все возможные вероятностные процессы, так или иначе изменяющие частоты встречаемости генов в популяции [16, 123]. В нашей книге большинство глав посвящены изучению генов и геномов бактерий и вирусов. Геном у бактерий и у большинства вирусов (за исключением, например, вируса иммунодефицита человека) является гаплоидным, а размножение у них бесполое (хотя половой процесс у бактерий происходит: существует возможность обмена плазмидами) — многие описанные в литературе проявления случайного генетического дрейфа к ним не применимы [30]. Тем не менее случайный генетический дрейф может проявляться у бактерий и вирусов, как минимум, в виде эффекта «горлышка бутылки» и «эффекта родоначальника» [30].

Эффект «горлышка бутылки» возникает тогда, когда большая часть популяции подвергается уничтожению в результате случайного процесса. В качестве примера можно привести любой катаклизм. В этом случае, как известно, выживают не наиболее приспособленные к ранее существовавшим условиям среды особи, а *наиболее удачливые* [123]. В результате в генофонде популяции (среди немногих выживших особей) увеличивается частота встречаемости наиболее «удачливых» аллельных генов [123]. Последний эффект имеет название «эффекта родоначальника» [1].

«Эффект родоначальника» может возникать не только как результат прохождения популяции через «горлышко бутылки». Пространственная изоляция небольшого количества особей из популяции приводит к тому же: в геномах их потомков будут присутствовать только те гены (за исключением мутантных), которые были в геномах родоначальников [21, 123].

Нельзя забывать и о том, что естественный отбор по какому-либо лимитирующему фенотипическому признаку приводит к случайной фиксации определенных вариантов менее важных в данных условиях среды признаков [30]. Возможность закрепления в популяции генетической мутации, повышающей жизнеспособность и/или плодовитость особи, то есть возможность положительного отбора на молекулярном уровне, не отрицается М. Кимурой. В пятом положении теории нейтральности констатировано, что положительный отбор на генетическом уровне происходит крайне редко [16].

Многие могут возразить утверждению, содержащемуся в этом положении, вспомнив о том, что сложность организации живых организмов возрастала с течением времени (естественно, «усложнилась» организация

лишь для меньшей части видов живых организмов), а человек стоит на вершине эволюционной лестницы. Чем же, как ни положительным отбором по Ч. Дарвину объяснить этот факт?

Теория нейтральности не противоречит этому, но указывает на то, что «положительные» генетические изменения происходят в эволюции весьма редко. Однако этой частоты оказалось достаточно, чтобы в течение четырех миллиардов лет появился человек разумный [16, 21]. Кстати, число генов, кодирующих белок, в геноме человека лишь в два-три раза превышает число кодирующих участков у некоторых современных бактерий. Это указывает на то, что «прогресс» скорее связан с качественными изменениями, чем с количественными. Известно о таком направлении «прогресса», как увеличение размеров головного мозга и площади его коры в филогенетическом ряду хордовых животных (классический пример). Несомненно, есть общий фенотипический признак, по которому идет положительный отбор в большинстве популяций высших животных, — это уровень развития центральной нервной системы (причем не в целом, а только тех особенностей нервной деятельности, которые помогают выжить, адаптироваться и оставить потомство). Виды животных, приспособившиеся к окружающей среде за счет закрепления анатомических и/или физиологических признаков, не относящихся к центральной нервной системе (например, за счет «узкоспециализированного» клюва, как это произошло у ткачиков с Галапагосских островов), а не в результате приспособительных реакций центральной нервной системы, качество которых может быть обусловлено анатомическими и физиологическими особенностями последней, оказываются обреченными на вымирание при резком изменении условий окружающей среды. Человек — поразительный пример минимума закрепленных приспособительных анатомо-физиологических признаков, не имеющих отношения к ЦНС, на фоне (и за счет) максимального (среди ныне существующих видов) развития высшей нервной деятельности.

В популяции диплоидных организмов, размножающихся половым путем, роль случая в фиксации каких-либо признаков еще выше, чем у ранее упомянутых бактерий и вирусов [123]. Необходимо напомнить, что вероятность закрепления того или иного генетически детерминированного признака в популяции находится в обратной зависимости от ее численности (в условиях свободного скрещивания между всеми особями). Чем меньше размер популяции, тем больше вероятность того, что какие-либо аллельные варианты генов будут утеряны по случайным причинам, а другие, по этим же причинам, закрепятся [16, 21].

Фиксация генетически детерминированных признаков в малой популяции диплоидных организмов, размножающихся половым путем, называется «дрейф генов» [1]. Это всего лишь один из процессов, объединенных под общим названием «случайный генетический дрейф» [123].

В 40-е гг. Р. Л. Берг доказала тот факт, что между степенью проявления мутаций в гетерозиготе и их концентрацией в генофонде популяции существует зависимость [9, 42]. Если в результате неполного доминирования мутация, снижающая приспособленность особи, частично проявляется фенотипически у гетерозиготных организмов, то концентрация ее в генофонде снижается под воздействием стабилизирующего отбора [9].

Однако чаще всего (в результате полного доминирования) рецессивные признаки фенотипически не проявляются [1]. Поэтому естественный отбор не может влиять на гетерозиготных носителей таких признаков. В больших популяциях частота встречаемости аллельных генов (как рецессивных, так и доминантных) должна быть относительно постоянной. Различные мутации в генах, которые не проявляются фенотипически, могут закрепляться или исчезать под влиянием случайных процессов.

Открытие такого свойства генов, как пенетрантность [24], привело к признанию существования еще одной разновидности процессов случайного генетического дрейфа. Пенетрантность — это вероятность того, что ген, находящийся в гомозиготном состоянии, проявится фенотипически [1]. Как оказалось, пенетрантность далеко не всегда равна 100 %. Гены с низкой пенетрантностью мало подвержены влиянию естественного отбора. Это еще раз убеждает нас в том, что лишь малая часть генетической информации проявляется фенотипически, становясь (а чаще — не становясь) «аргументом» в борьбе за существование.

Еще в 30–40-е гг. XX в. многие биологи-эволюционисты пришли к выводу о том, что, кроме «отрицательных» и «положительных» вариантов фенотипических признаков, существуют и «нейтральные отклонения — все те вариации, которые выражаются в изменениях фенотипа, не влияющих на выживание и плодовитость особи в данных условиях существования» [42]. И. И. Шмальгаузен написал в своей известной книге «Факторы эволюции» (первое издание — 1946 г.) практически то же самое, что и М. Кимура в пятом положении теории нейтральности: «оценивая значение индивидуальных отклонений в эволюции, мы должны, в первую очередь, отметить существование принципиальной разницы между *крайне редкими положительными отклонениями* и условной нормой, с одной стороны, и *многочисленными отрицательными отклонениями* от нее, с другой» [42].

И до появления теории нейтральной эволюции многие ученые, причем не только биологи, приходили к признанию роли случая в природе. Даже известный психиатр Зигмунд Фрейд, который обладал поразительной способностью предполагать наличие причинно-следственной связи в самых туманных и расплывчатых сновидениях своих пациентов, признавал, что «при этом мы охотно забываем, что, собственно, все в нашей жизни — случай, начиная с нашего возникновения путем встречи сперматозоида



и яйцеклетки; случай, который по причине своего возникновения все-таки соучаствует в закономерности и необходимости природы, обходясь, впрочем, без связи с нашими желаниями и иллюзиями» [27].

Нобору Суеока основал свою теорию мутационного давления на теории нейтральности Муто Кимуры [121]. Несмотря на то, что М. Кимура знал о теории Н. Суеоки, он ни разу не прокомментировал ее в своих научных публикациях. В книге «Теория нейтральной эволюции» [16] есть только одно предложение, в котором упоминается термин «мутационное давление» (судя по контексту, этот термин был употреблен скорее в том абстрактном смысле, который придавал ему И. И. Шмальгаузен, чем в конкретной биохимической интерпретации Н. Суеоки). Разъяснить причины игнорирования М. Кимурой теории Н. Суеоки помогли бы знания, почерпнутые из трудов процитированного выше З. Фрейда. Содержание же теории мутационного давления изложено в следующей главе данного издания.

### Глава 3

#### Теория мутационного давления Н. Суеоки

Нобору Суеока в начале 60-х гг. XX в. открыл факт наличия корреляции между частотами использования многих аминокислотных остатков в бактериальных белках и GC-насыщенностью кодирующих их генов [120]. В это время генетический код еще не был расшифрован. На основании открытых закономерностей Н. Суеока предложил свой вариант кодирования аминокислот при помощи комбинаций нуклеотидов [120]. Как выяснилось, природу генетического кода Н. Суеоке вскрыть не удалось. Однако ему удалось открыть существование такого «внутреннего» фактора эволюции, как мутационное давление [120].

Основной идеей теории Н. Суеоки, на наш взгляд, является следующее утверждение: нуклеотидные замещения в ДНК (точечные мутации) имеют определенное предпочтительное *направление*, выраженное в большей или меньшей степени. Процесс возникновения нуклеотидных замен нельзя назвать случайным. Замены каждого типа в данном гене (или геноме) происходят с определенной вероятностью. Эти вероятности практически никогда не равны друг другу. Процесс же фиксации нейтральных замен в популяции является случайным. Получается, что замены по направлению мутационного давления чаще фиксируются случайным генетическим дрейфом исключительно из-за того, что они происходят чаще, чем замены противоположного направления [87, 121].

Существуют биохимические процессы, вызывающие увеличение частоты возникновения строго определенных типов нуклеотидных замен. Данные процессы Н. Суеока и объединил под сборным термином «мутационное давление» [121]. Мутационное давление в геномах, представлен-

ных двухцепочечной ДНК, может иметь два направления: **GC-давление** (при повышенной частоте возникновения замен аденина и тимина на гуанин и цитозин, относительно обратных им замен) или **AT-давление** (при повышенной частоте возникновения замен гуанина и цитозина на аденин и тимин относительно обратных им замен). Теоретически возможна ситуация, в которой AT- и GC-давление полностью уравнивают друг друга, однако в большей части геномов абсолютного равновесия этих процессов не наблюдается [121].

Общие частоты использования гуанина и цитозина в двухцепочечной ДНК равны друг другу и изменяются вместе, так как эти нуклеотиды комплементарны друг другу: при замене аденина на гуанин в одной из цепочек ДНК во время первой же репликации на дочерней цепочке появляется комплементарный гуанину цитозин.

Большую часть своих поздних работ, написанных в соавторстве с Ж. Б. Лобри, Н. Суеока посвятил изучению причин того, почему частоты гуанина и цитозина (так же как и частоты аденина и тимина) близки, но никогда абсолютно не равны друг другу в генах бактерий [96]. Помимо симметричного, он открыл еще и асимметричное мутационное давление: частоты использования нуклеотидов в кодирующих участках многих бактерий находятся в зависимости от того, на какой из двух цепочек ДНК (на лидирующей или на отстающей) эти кодирующие участки расположены [96]. Конкретных причин неравенства G и C найти не удалось, хотя авторы были уверены в «биохимическом происхождении» локальных отклонений от второго правила Чаргаффа [96].

Результатом воздействия мутационного давления на кодирующие участки ДНК является направленное изменение частот использования нуклеотидов. Нуклеотид, как известно, может находиться в первом, во втором или в третьем положении кодона.

На рис. 1 изображен полученный автором [31] график, демонстрирующий прямую линейную зависимость между средним содержанием гуанина и цитозина в каждом из трех положений кодонов (показатели 1GC, 2GC и 3GC) и общей GC-насыщенностью полных и неполных бактериальных геномов. Наличие прямой зависимости между каждым из этих трех показателей и общей GC-насыщенностью указывает на то, что мутационное давление (то есть преимущественное направление нуклеотидных замен) оказывает огромное влияние на распределение частот использования кодонов.

Показатель **3GC** демонстрирует самый высокий наклон относительно G+C (наклон = 1,84), поскольку большинство нуклеотидных замен по третьему положению синонимичны (не приводят к изменению состава кодируемого белка). В результате отрицательный отбор далеко не всегда

может оказывать влияние на содержание гуанина и цитозина в третьем положении кодона [31, 105, 121].

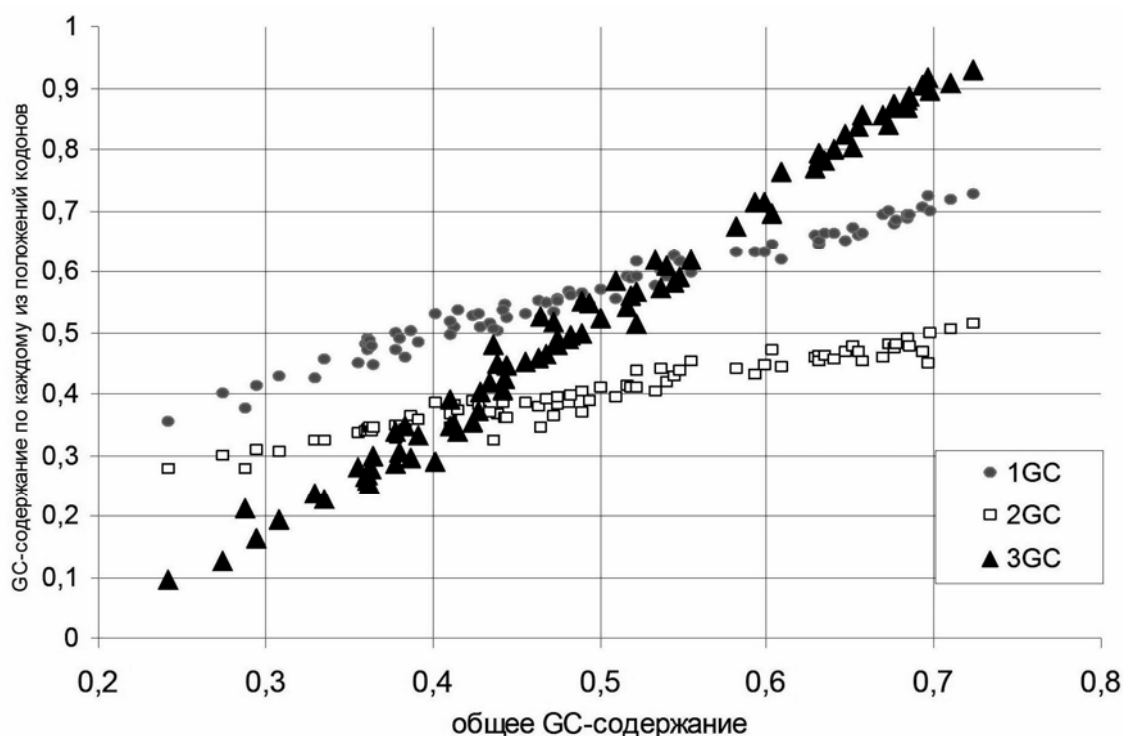


Рис. 1. Зависимость средних значений показателей 1GC, 2GC и 3GC от G+C в полных и неполных геномах 84 видов бактерий [31]

Показатель **1GC** имеет гораздо меньший наклон (наклон = 0,71), чем показатель **3GC**. Этот факт можно объяснить тем, что лишь незначительная часть замен по первому положению синонимична [31, 105, 121].

Несмотря на то, что все замены по второму положению несинонимичны, показатель **2GC** находится в линейной зависимости от общего содержания G+C (наклон = 0,45). Объяснить это можно, приняв во внимание тот факт, что не все замены аминокислотных остатков в белках подвергаются воздействию отрицательного отбора [31, 105, 121]. Мутационное давление предоставляет материал для отрицательного отбора и для случайного генетического дрейфа [31]. В том случае, если мутация нейтральна, отрицательный отбор на нее не влияет, и она получает шанс закрепиться.

В пользу существования направленного мутационного давления говорит тот факт, что далеко не все нуклеотидные замены фиксируются в популяции, причем не столько из-за их элиминации отбором, сколько из-за непредсказуемости, присущей случайному генетическому дрейфу. Вероятность закрепления одной нейтральной мутации ничтожно мала [42], но мутационное давление вызывает повышение частоты возникновения однотипных мутаций. Благодаря появлению множества нуклеотидных замен одного и того же направления, часть из них все же закрепляется в популяции, изменяя общую GC-насыщенность генома данного вида организмов:

у отдельных бактерий общая GC-насыщенность в ходе эволюции снизилась до 24 % (*Mycoplasma mycoides*), у других — повысилась до 72 % (*Streptomyces coelicolor*) [31].

Наиболее «чувствителен» к мутационному давлению показатель 3GC. Если изучить зависимость 3GC от G+C, можно заметить, что при  $G+C > 0,5$ , 3GC тоже превышает 0,5 и возрастает с вышеуказанным наклоном. При  $G+C < 0,5$  показатель 3GC убывает с тем же наклоном. Приведенные на рис. 1 данные говорят о том, что в большинстве случаев при  $G+C > 0,5$  преимущественное направление нуклеотидных замен — АТ на GC, а при  $G+C < 0,5$  преимущественное направление нуклеотидных замен — GC на АТ [31].

По теории мутационного давления, частоты использования нуклеотидов в генах и геномах изменяются **по причине** неравновероятности возникновения различных типов нуклеотидных мутаций. Согласно трактовке исследователей, называющих себя «селекционистами», частоты использования нуклеотидов в генах и геномах изменяются **для того, чтобы** стабилизировать (или дестабилизировать) вторичную структуру транскрибируемых с них молекул РНК [59, 65] или же **для того, чтобы** снизить вероятность рекомбинации между дублицировавшимися генами, а значит, увеличить вероятность успеха дубликации [50, 65].

Что касается вторичной структуры РНК, то она неизбежно образуется у любого транскрипта [23]. Одной из фаз инициации трансляции является «расплетение» многочисленных и разнообразных шпилек РНК при помощи специальных белков, обладающих РНК-хеликазной активностью (их причисляют к факторам инициации трансляции) [23].

Вторая «цель» изменения GC-насыщенности генов может объяснить психологические предпосылки селекционистской постановки научного вопроса («селекционистов», как мы уже успели подчеркнуть, интересуется *с какой целью* происходит то или иное событие, а не *по какой причине*). Согласно концепции, созданной Дж. Бернарди и развитой Д. Р. Форсдайком, дублицировавшиеся гены могут рекомбинировать, а рекомбинация может привести к потере второй копии гена. Чем больше сходство в нуклеотидном составе двух копий гена, тем выше вероятность рекомбинации [50, 65]. По мнению упомянутых выше исследователей, для того чтобы гены чаще подвергались дубликации, в геноме теплокровных животных (как выяснилось позже — не только у них) появились изохоры [50]. Изохоры — это участки хромосомной ДНК с примерно одинаковым уровнем G+C. Хромосомы теплокровных животных состоят из множества изохор с разным уровнем GC-насыщенности [50]. Если дубликат гена попадает в изохору, отличающуюся по GC-насыщенности от той, в которой расположен оригинал, — вероятность рекомбинации резко снижается [50].

Д. Р. Форсдаик, зная об отсутствии изохор у бактерий, провозгласил каждый отдельный бактериальный ген «микроизохорой» [65]. Дело в том, что гены в геномах бактерий, отличающихся средним уровнем GC-насыщенности, варьируют в большей степени по уровню GC-насыщенности в третьих, а не в первых и во вторых положениях кодонов [65, 91]. По мнению Д. Р. Форсдайка, каждый ген имеет свой «индивидуальный» уровень 3GC для того, чтобы понизить вероятность рекомбинации и увеличить вероятность успеха дупликации. Мысль о том, что уровень GC-насыщенности в третьих положениях кодонов генов таких бактерий во многом подвержен колебаниям за счет случайных процессов [103], Д. Р. Форсдаик не допускает.

В чем же положительная роль высокой частоты дупликации генов с точки зрения «селекционистов»? В том, что высокая частота дупликации генов повышает скорость эволюции и приспособления организмов, приводит к ускорению процессов дивергенции внутри вида, в общем, «способствует эволюции» [65].

После изложения точки зрения «селекционистов» читателю должно стать понятно, почему мы помещаем термин «селекционизм» в кавычки. Естественный отбор в трудах «селекционистов» предстает как некий метафизический фактор, постоянно стремящийся «способствовать эволюционированию, повышать уровень организации, создавать новые в функциональном отношении гены». Как это ни парадоксально, современный молекулярно-биологический «селекционизм» имеет гораздо больше общего с метафизической теорией эволюции Ж. Б. Ламарка, чем с теорией эволюции Ч. Дарвина. Если еще сильнее углубиться в историю, можно обнаружить корни «селекционистского» подхода в учении о *телеологии* Аристотеля [14]. Именно он развил мысль о том, что все в мире происходит с какой-либо целью, о том, что все объекты материального мира изначально содержат в себе некую цель [14]. Эта часть многогранного учения древнегреческого философа была особенно активно воспринята средневековым христианством и впоследствии вплелась в систему европейского образования и в науку [14]. В свободной энциклопедии ([www.wikipedia.com](http://www.wikipedia.com)) можно найти следующее определение: «Телеология ставит перед собой задачу ответить на вопрос «зачем, с какой целью?» без анализа причинно-следственной связи явлений». Этому учению противопоставляется натурализм. Противостояние телеологии и натурализма характерно не только для биологии.

Естественно, разница в GC-насыщенности изохор должна оказывать влияние на успех дупликации. Но успех дупликации — это не цель, а одно из многочисленных последствий возникновения изохор. Причиной же возникновения изохор должна быть разница в интенсивности мутационных и репарационных процессов в разных участках хромосом, то есть разница

в направлении и силе мутационного давления [67, 68]. Причиной разницы в интенсивности мутационных и репарационных процессов может быть наличие различных фаз жизненного цикла [85]. У человека, как известно, генетический материал проходит через длинную цепочку клеток разной степени дифференцировки (от момента оплодотворения до момента следующего оплодотворения). Дифференцировка клетки достигается путем транскрипции только определенного набора генов [23]. Если в набор генов, активных в данной клетке, входит ген-мутатор, то мутации, вызываемые продуктом последнего, будут происходить, в основном, в тех генах, с которых идет транскрипция [85]. На каждом этапе дифференцировки активным является определенный набор генов, в том числе и тех, продукты которых способны влиять на частоты возникновения мутаций. Это и приводит к расширению границ внутригеномных вариаций GC-насыщенности. Подобную ситуацию мы обнаружили [85] в геноме прокариотического организма *Methanosarcina Acetivorans*, в жизненный цикл которого входят несколько фаз (от одноклеточного состояния до многоклеточного с дифференцированными клетками) [70].

Нельзя не упомянуть о том, что термин «мутационное давление» широко использовался в эволюционной биологии и до создания Н. Суеокой теории мутационного давления. И. И. Шмальгаузен вкладывал в это понятие гораздо более широкий смысл. В приведенном ниже абзаце содержатся цитаты из его книги «Факторы эволюции» [42], в которых раскрывается содержание этого немаловажного эволюционного фактора.

«Механизм естественного отбора и его результат в различных случаях нетрудно представить себе на обычных вариационных кривых... Если *мутационное давление* (курсив автора монографии) определяет расширение вариационной кривой, то избирательная элиминация (стабилизирующий отбор) определяет ее сжатие и поднятие вершины... Всякое улучшение условий, связанное с сокращением элиминации и, следовательно, с ослаблением интенсивности отбора, ведет поэтому к расширению вариационной кривой под действием *мутационного давления* (курсив автора монографии)» [42].

Хочется подчеркнуть, что речь в приведенных выше цитатах велась о накоплении в популяции самых разнообразных мутаций, проявляющихся фенотипически. Н. Суеока же употребил термин «мутационное давление» в узком смысле — для характеристики преимущественного направления нуклеотидных замен в ДНК [120, 121, 122]. В своих ранних работах Н. Суеока чаще оперировал еще более конкретизированным термином: «направленное мутационное давление» [121]. Этим автор подчеркивал факт преимущественного возникновения однотипных замен (замен только одного из двух возможных направлений).

Следует уточнить, что мутационное давление имеет определенное направление по вполне естественным биохимическим причинам.

Процесс дезаминирования нуклеотидов является молекулярной основой транзиций [23, 72]. Дезаминирование цитозина и аденина может протекать как спонтанно, так и ферментативно.

В результате дезаминирования цитозина образуется урацил. В результате дезаминирования метилированного цитозина образуется тимин. Если такие мутации не подвергаются репарации, после репликации в одной из двух дочерних копий ДНК возникает транзиция по направлению GC на AT [72]. Урацил вырезается как из одноцепочечной, так и из двухцепочечной ДНК при помощи ферментов из семейства урацил-ДНК-гликозилаз. Тимин, будучи мажорным нуклеотидом, вырезается исключительно из неправильно образованных пар (в частности, из пары T:G, образующейся в результате дезаминирования 5-метил-цитозина) с помощью тимин-ДНК-гликозилаз [72].

Относительно недавно было выяснено, что в некоторых клетках, в частности, в иммунокомпетентных клетках человека, замены C на U (цитозина на урацил) происходят ферментативно: благодаря работе «индуцируемой активацией цитидин-ДНК-дезаминазы» (AICDA) [22, 102, 106]. AICDA находится в близком родстве с ферментами, дезаминирующими цитозин в РНК (APOBEC1) [61].

В результате дезаминирования аденина образуется гипоксантин (инозин) [23, 72]. Инозин комплементарен цитозину. Если мутация аденина в инозин не репарируется, после репликации в одной из двух дочерних копий ДНК возникает транзиция по направлению AT на GC [72]. На данный момент времени считается, что дезаминирование аденина в ДНК — процесс исключительно спонтанный (чаще всего аденин в ДНК дезаминируется под воздействием оксида азота) [72]. При этом известно о существовании целого семейства ферментов, дезаминирующих аденин в РНК (ADAR) [62].

Когда мы употребляем выражение «инозин комплементарен цитозину», имеется в виду, что инозин чаще всего образует водородные связи с цитозином. Большое количество экспериментальных исследований [63] посвящено выяснению вероятности образования водородных связей между нуклеотидами. Как выяснилось, нельзя утверждать, что аденин образует водородные связи исключительно с урацилом и тиминем. Аденин может образовывать водородные связи также и с цитозином, гуанином и аденином, но прочность таких связей, а следовательно, и вероятность их возникновения, значительно ниже, чем в паре аденин – тимин (урацил).

Транверсии чаще всего вызываются биохимическими процессами, в которые вовлечен окисленный гуанин (8-оксо-G) [72].

Процесс замены аденина на цитозин довольно сложен (рис. 2). Впервые он был изучен у *Escherichia coli*. Как оказалось впоследствии, все про-

кариотические и эукариотические организмы содержат в своих геномах гены, гомологичные «генам-мутаторам» кишечной палочки [72]. Под воздействием свободных радикалов дезоксигуанитрифосфат (ГТФ) окисляется с образованием 8-оксогуанитрифосфата (8-оксо-ГТФ). Последний комплементарен не цитозину, а аденину. Опять же, понятие комплементарности в данном случае относительно. При репликации ДНК 8-оксо-Г часто оказывается в дочерней цепи напротив аденина [72]. Фермент, продукт гена MutY, вырезает аденин, оказавшийся в паре с 8-оксо-Г (рис. 2) [72]. После этого напротив 8-оксо-Г образуется апуриновый/апиримидиновый сайт (AP-сайт). Репарация таких сайтов осуществляется с помощью AP-эндонуклеаз, которые формируют разрыв цепочки ДНК, и ДНК-полимеразы, которая достраивает недостающий нуклеотид [72]. Как оказалось, предпочтительность включения аденина или цитозина напротив 8-оксо-Г зависит, в том числе, и от свойств ДНК-полимеразы [63]. Когда напротив 8-оксо-Г оказывается цитозин, образуется пара неправильно спаренных нуклеотидов «8-оксо-Г:С». Еще один фермент, MutM, вырезает 8-оксо-Г из различных неправильно спаренных нуклеотидных пар [72]. AP-сайт, образовавшийся напротив цитозина, чаще всего заполняется гуанином. Так происходят трансверсии АТ на GC [72].

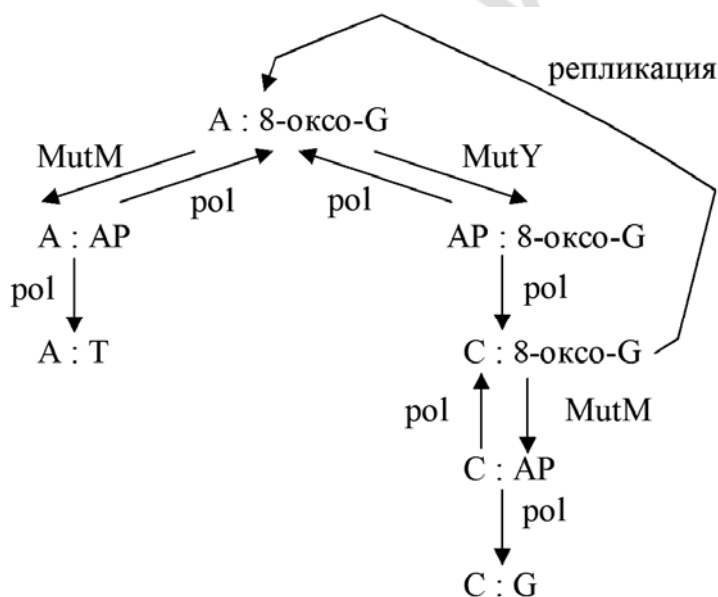


Рис. 2. Схема возможных путей репарации ДНК, содержащей 8-оксо-Г

Функцией фермента, кодируемого геном MutT, является элиминация 8-оксо-ГТФ из пула нуклеотидов [72]. Чем больше содержание 8-оксо-ГТФ в клетке, тем выше вероятность возникновения трансверсий АТ на GC.

Следует отметить, что вышеописанный процесс может происходить при совместном влиянии, по крайней мере, трех факторов: 1) окислительный стресс, 2) высокая экспрессия генов MutY и MutM, 3) низкая актив-



ность продукта гена MutT [31]. Продукты всех трех упомянутых генов (MutT, MutY и MutM) являются представителями системы репарации [72].

Получается, что ферменты из системы репарации могут вносить вклад в создание мутационного давления не только за счет снижения своей функции, но и за счет своего «нормального» функционирования [72].

Гуанин может окисляться и непосредственно в цепи ДНК. В результате образуется неправильно спаренная пара (англ. mispair) нуклеотидов — «8-оксо-G:C» (см. рис. 2) [72]. Далее, в норме фермент, обладающий способностью вырезать цитозин из пары с 8-оксо-G (как мы уже упоминали, этой способностью обладают гомологи фермента MutM), вырезает его из пары «8-оксо-G:C» [72]. После этого в образовавшийся AP-сайт напротив цитозина встраивается гуанин [72]. Последствия окисления гуанина в ДНК полностью устраняются.

Если комплекс репарационных ферментов (MutM и MutY) работает эффективно, то гуанин, окислившийся в ДНК, успешно вырезается (трансверсии GC на AT предотвращаются), но 8-оксогуанинтрифосфат, встроившийся в растущую цепь ДНК, приводит к появлению трансверсий AT на GC [72].

Фермент MutY выступает в роли «второго звена» в репарации окисленного в ДНК гуанина: если MutM не вырезал 8-оксо-G из пары «8-оксо-G:C» до репликации, а во время репликации образовалась пара «8-оксо-G:A», то MutY вырезает из этой пары не 8-оксо-G, а аденин, что приводит к последующему восстановлению пары «G:C» [72].

Если же репарационный фермент MutM меняет свою специфичность: начинает чаще связываться с парой «8-оксо-G:A», чем с «8-оксо-G:C», то равновесие (см. рис. 2) смещается в сторону учащения возникновения трансверсий GC на AT [72].

Известно, что не только окисленный гуанин, но и окисленный тимин (5-формил-урацил — часто встречающееся повреждение ДНК в результате воздействия радиации), может приводить к возникновению мутаций в результате неправильного спаривания нуклеотидов во время репликации [72]. Однако, по данным новейших исследований, 5-формил-урацил предпочтительно образует водородные связи с аденином, причем сила этих связей даже больше, чем таковая в исходной паре T:A. На втором месте по частоте образования находится пара 5-формил-урацила с гуанином, на третьем — с цитозином [72]. Это означает, что окисленный тимин, хоть и с малой вероятностью, но может вызывать транзиции и трансверсии по направлению AT на GC [72].

Присутствие в геномах подавляющего большинства организмов генов, кодирующих такие ферменты, как аденин-ДНК-гликозилаза (MutY) и тимин-ДНК-гликозилаза, само по себе создает условия для возникновения мутационного GC-давления. Известно, что во время репликации ДНК-

полимераза может допускать ошибки. Если аденин и тимин будут неправильно встроены напротив цитозина и гуанина, то ферменты аденин-ДНК-гликозилаза и тимин-ДНК-гликозилаза должны будут вырезать ошибочно встроены ДНК-полимеразой аденин и тимин. Если же ДНК-полимераза встроит гуанин и цитозин напротив аденина и тимина, то те же самые ферменты из системы репарации должны будут вырезать из этих неправильных пар «правильные» аденин и тимин, и привести к возникновению замен по направлению АТ на GC на обеих цепях ДНК.

Известно, что частота, с которой ДНК-полимераза допускает ошибки, варьирует среди организмов [72]. Каждая ДНК-полимераза имеет свой собственный спектр наиболее и наименее часто вносимых типов ошибок [72]. Судя по данным о системе репарации, большинство организмов «застраховано» от мутагенного воздействия ДНК-полимераз, часто допускающих ошибки по направлению GC на АТ, но результаты ошибок встраивания нуклеотидов по направлению АТ на GC под воздействием ферментов из системы репарации не только не устраняются, но и «усугубляются».

Подобно ферментам MutY и MutM, урацил-ДНК-гликозилазы [47] и гипоксантин-ДНК-гликозилазы [100] вырезают урацил и гипоксантин из одноцепочечной и/или из двухцепочечной ДНК. Есть ферменты, которые, как и MutY, распознают «неправильно спаренные пары», то есть работают с двухцепочечной ДНК [72]. Есть и ферменты, распознающие непосредственно *минорный* нуклеотид (например, 8-оксо-G, гипоксантин, 5-формил-урацил, урацил — в ДНК), то есть работающие с одноцепочечной ДНК [47]. Часто ферменты из системы репарации способны вырезать широкий спектр минорных нуклеотидов как из одноцепочечной, так и из двухцепочечной ДНК (при этом они проявляют предпочтительность к тому или иному субстрату) [72]. Например, как бактериальные, так и эукариотические гипоксантин-ДНК-гликозилазы способны вырезать из ДНК не только инозин, но и 5-формил-урацил [72].

Как правило, частоты возникновения транзиций и трансверсий одного и того же направления, не равны друг другу [84, 85]. Причиной этому могут быть особенности работы сложной системы репарации.

Нобору Суеока приписывал ведущую роль в создании мутационного давления двум биохимическим процессам: дезаминированию цитозина и окислению гуанина [122]. На данный момент число известных процессов, приводящих к возникновению мутационного давления, увеличилось. Особенно интенсивно изучаются ферментативные мутационные процессы в РНК [62]. Клеточные ферменты, дезаминирующие цитозин и аденин в РНК, способны взаимодействовать как с клеточной РНК, так и с РНК вирусов [51, 53, 108, 134].

## Глава 4

### Понятие о силе мутационного давления

В совместных с Е. В. Барковским исследованиях, посвященных изучению мутационного давления в геномах прокариот (бактерий и архей) [5, 6, 30, 31, 32, 37, 85] и вирусов [33, 39, 82, 83, 84], автором была обнаружена важная для понимания последствий этого явления закономерность. В первую очередь на процессы мутационного давления реагирует GC-насыщенность третьих положений кодонов. Наступает момент, когда GC-насыщенность в третьих (наиболее нейтральных) положениях кодонов достигает максимума (при GC-давлении) или минимума (при AT-давлении) [82, 84].

Как правило, при 99%-ном заполнении третьих положений кодонов гуанином и цитозином первые и, особенно, вторые положения имеют гораздо более низкие значения GC-насыщенности [31, 82, 84, 85]. Замены AT на GC в таких условиях могут происходить исключительно в первых и во вторых положениях кодонов. Вероятность того, что нуклеотидная замена AT на GC будет несинонимичной, приближается к 100 % [82, 84, 85], а того, что замена AT на GC окажется синонимичной, приближается к 0 % [82, 84, 85]. Синонимичные замены при этом могут происходить за счет мутаций обратного направления (GC на AT) и за счет трансверсий, не изменяющих GC-насыщенность (G на C и C на G), в третьих положениях кодона [81, 84]. Большое количество субстрата для обратных замен (GC на AT) в последних должно повышать вероятность их возникновения [81]. Как показали наши исследования GC-богатых геномов симплексовирусов, обратные замены (GC на AT) действительно происходят в третьих положениях кодонов, но после этого GC-насыщенность восстанавливается [84]. Причем снижается GC-насыщенность за счет транзиций, а восстанавливается, чаще всего, за счет трансверсий. В результате в заполненных гуанином и цитозином третьих положениях кодонов происходят и фиксируются, в основном, замены G на C и C на G [81]. В таких условиях GC-насыщенность начинает постепенно повышаться за счет фиксации замен AT на GC в первых и во вторых положениях кодонов, поскольку GC-насыщенность третьих положений кодонов больше повыситься не может [82, 84, 85].

Ситуацию, когда третьи положения кодонов практически полностью насытились гуанином и цитозином, мы назвали «*сильным* мутационным GC-давлением» [82, 84, 85]. Если третьи положения кодонов еще не полностью заполнились гуанином и цитозином, мутационное GC-давление является *слабым* [84, 85]. Аналогичным образом, в случае практически полного насыщения третьих положений кодонов аденином и тиминном, необходимо признать факт наличия *сильного* мутационного AT-давления [85].

Биологическим свойством сильного мутационного давления является то, что вероятность возникновения несинонимичной замены (вызванной мутационным давлением) в данном случае значительно выше, чем при слабом мутационном давлении [82, 84]. Нами был описан случай, когда сильное мутационное GC-давление вызвало ситуацию статистически достоверного превышения несинонимичной эволюционной дистанцией синонимичной дистанции [82]. Другими словами, сильное мутационное давление приводит к повышению частоты возникновения несинонимичных замен в целом, за счет увеличения вероятности того, что замена, вызванная мутационным давлением, будет несинонимичной и за счет снижения возможности возникновения синонимичной замены [82].

Необходимо уточнить, что же такое несинонимичная ( $dN$ ) и синонимичная ( $dS$ ) эволюционные дистанции. Значение  $dN$  равно отношению количества несинонимичных замен между двумя нуклеотидными последовательностями к количеству несинонимичных сайтов [10, 11, 105]. Значение  $dS$ , соответственно, равно отношению количества синонимичных замен между двумя нуклеотидными последовательностями к количеству синонимичных сайтов [10, 11, 105]. Дело в том, что абсолютно синонимичными сайтами являются только четырехкратно вырожденные сайты [105]. Двукратно вырожденные сайты являются синонимичными на одну треть, а на две трети — несинонимичными [105]. Такое свойство генетического кода само по себе вносит погрешность в вычисление синонимичных и несинонимичных дистанций. Так в чем же смысл вычисления  $dN$  и  $dS$ ?

Создатели многочисленных методик вычисления  $dN$  и  $dS$  стремились определить наличие положительного отбора на молекулярно-генетическом уровне [12, 105]. По их мнению, если  $dN$  статистически достоверно превышает  $dS$ , ген находится под воздействием положительного отбора [105]. Вдумайтесь в определение положительного отбора, которое дают авторы методики вычисления  $dN$  и  $dS$ : «Положительный отбор приводит к повышению частоты фиксации несинонимичных замен **для того, чтобы** продукт гена обрел свою функцию» [105]. Как и в гл. 3 первой части, мы снова встречаемся с учением о телеологии, скрывающимся под названием «селекционизм». В данном примере положительный отбор выступает как «трансцендентальный» фактор, который закрепляет в популяции предпочтительно несинонимичные нуклеотидные замены в данном гене «в надежде» на обретение кодируемым им белком некой полезной функции.

Статистически достоверное превышение несинонимичной эволюционной дистанцией синонимичной дистанции — явление крайне редкое. В связи с этим многие исследователи предполагают наличие положительного отбора и при статистически недостоверном превышении или даже при достижении определенного соотношения между ними (например, при  $dN/dS > 0,6$ ) [105].

Описанная выше концепция положительного отбора противоречит теории М. Кимуры. По теории нейтральности, чем чаще закрепляются несинонимичные замены в гене, тем он менее функционально значим [16]. Положительный же отбор (как по теории Ч. Дарвина, так и по теории М. Кимуры) может зафиксировать в популяции одну-единственную аминокислотную замену, приведшую к изменению функции белка, и, как следствие, вызвавшую повышение приспособленности особи и популяции в целом к условиям среды, но никак не огромное количество аминокислотных замен в белке, которые так ни к чему и не привели, но теоретически могут привести, если будут с той же интенсивностью происходить и дальше [16].

В упомянутой выше работе [82] мы впервые указали на то, что статистически достоверное превышение несинонимичной эволюционной дистанцией синонимичной дистанции может происходить из-за сильного мутационного давления. Основной вклад в ситуацию  $dN > dS$  в случае сильного мутационного давления вносит резкое снижение вероятности возникновения синонимичных замен [82].

В данном издании описаны методы определения силы мутационного давления как в отдельных генах, так и в полных геномах прокариот и вирусов. Интересен тот факт, что ситуация сильного мутационного давления может быть обнаружена и в отдельных участках генов [86]. Именно на основании понятия о сильном мутационном давлении построена большая часть приведенных в монографии оригинальных методов. Надеемся, что предложенные нами пути применения теории мутационного давления к решению многих вопросов микробиологии, вирусологии и иммунологии будут обладать достаточным практическим выходом.

Что касается теоретической значимости введения понятия о сильном мутационном давлении, то частично о ней рассказано в гл. 1 четвертой части. Методика, предложенная Н. Суеоккой для определения степени нейтральности аминокислотных замен в выборке гомологичных белков [121], не учитывала саму возможность существования сильного мутационного давления.

Влияние сильного мутационного давления на процесс развития жизни на Земле, по нашим представлениям, было огромным. Нами было выяснено, что геном общего предшественника архей был подвержен сильному мутационному АТ-давлению [8, 85], а геном общего предшественника бактерий — сильному GC-давлению [8]. Это во многом обусловило разницу в среднем аминокислотном составе протеомов архей и бактерий. Кроме того, сильное мутационное давление действительно должно способствовать увеличению интенсивности процессов естественного отбора, но об этом подробнее написано в следующей главе.

## Глава 5 Последствия мутационного давления

С целью демонстрации последствий мутационного давления предлагаем прибегнуть к наглядному примеру. На рис. 3 изображена NJ-дендрограмма [28], построенная на основании эволюционных дистанций, вычисленных с использованием двухпараметрического метода Кимуры [3, 16]. Эта дендрограмма, впрочем, как и любое другое филогенетическое древо, должна быть отражением эволюционного родства между нуклеотидными последовательностями. В данном случае — между генами, кодирующими тимидинкиназу у различных альфа-герпесвирусов. Справа на рис. 3 указана общая GC-насыщенность (G+C) и GC-насыщенность третьих положений кодонов (3GC) этих генов. Жирным шрифтом обозначены значения G+C и 3GC, превышающие 0,6. Курсивом обозначены значения G+C и 3GC ниже 0,6.

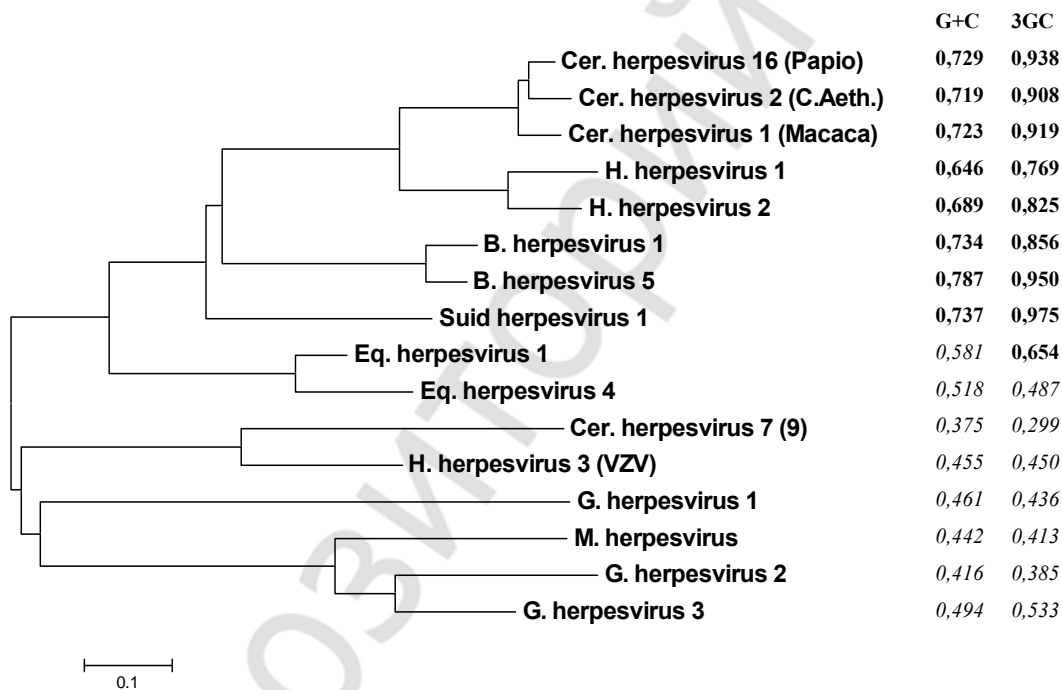


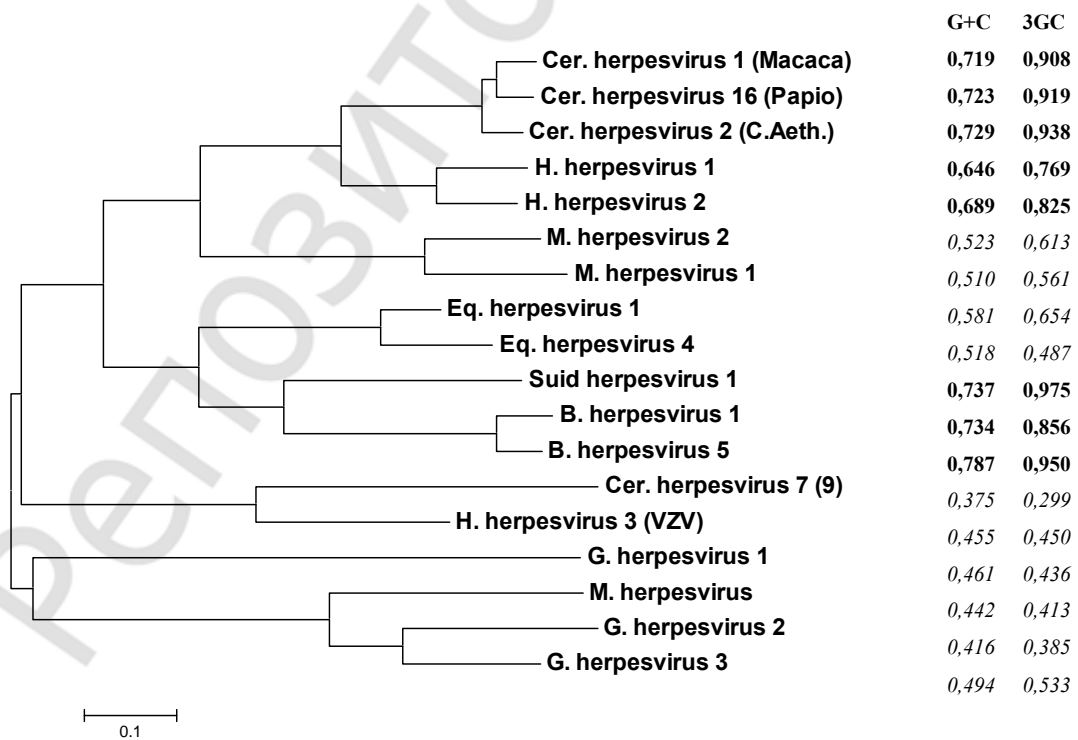
Рис. 3. NJ-дендрограмма, построенная по результатам вычисления эволюционных дистанций (Kimura-2-parameter) между генами, кодирующими тимидинкиназу у 16 альфа-герпесвирусов

Как видно на рис. 3, гены вирусов простого герпеса 1-го и 2-го типа (*H. herpesvirus 1* и *2*) расположены рядом с генами вирусов простого герпеса обезьян (*Cer. herpesvirus 1*, *2* и *16*). Эти вирусы простого герпеса объединены в род симплексвирус. Род варицелловирус представлен варицеллозостер вирусами человека и обезьяны (*VZV* и *Cer. herpesvirus 9*), варицелловирусами лошадей (*Eq. herpesvirus 1* и *4*), варицелловирусами крупного

рогатого скота (*B. herpesvirus 1* и *5*) и свиней (*Suid herpesvirus 1*). Наибольшую степень родства с симплексвирусами, судя по рис. 3, проявляют варицелловирусы крупного рогатого скота и свиней. Варицелло-зостер-вирусы человека и обезьяны группируются вместе с альфа-герпесвирусами птиц (*G. herpesvirus 1, 2* и *3, M. herpesvirus*).

На основании подобных дендрограмм была создана теория о коэволюции симплекс- и варицелловирусов с их хозяевами [97]. Верхняя большая ветвь дендрограммы с рис. 3 очень напоминает филогенетическое древо млекопитающих, на котором более мелкие ветви расположены в том же порядке (совпадает топология дендрограмм): сначала отходит ветвь, ведущая к непарнокопытным, затем — к парнокопытным, а последней разделяется ветвь, ведущая к приматам (человеку и обезьянам). По причине сходства топологии филогенетических деревьев вирусов и их хозяев долгое время считалось, что симплексвирусы возникли от варицелловирусов исключительно у приматов. Оставался открытым вопрос возникновения варицелло-зостер-вируса человека и обезьяны, так же как и альфа-герпесвирусов, поражающих птиц.

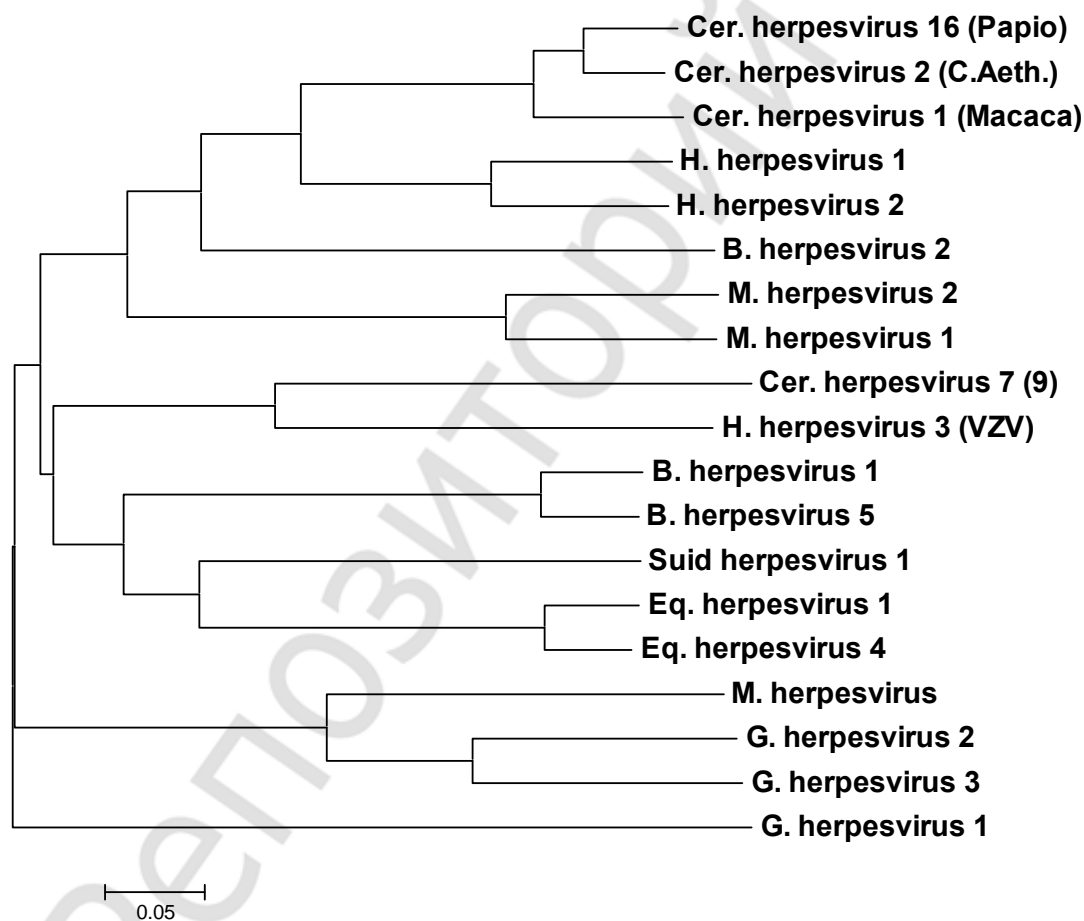
На рис. 4 представлена дендрограмма, построенная с добавлением еще двух генов, кодирующих тимидинкиназу. Эти гены были получены от вирусов герпеса кенгуру (*M. herpesvirus 1* и *2*). Справа указана общая GC-насыщенность (G+C) и GC-насыщенность третьих положений кодонов (3GC) этих генов. Жирным шрифтом обозначены значения G+C и 3GC, превышающие 0,6. Курсивом обозначены значения G+C и 3GC ниже 0,6.



*Рис. 4.* NJ-дендрограмма, построенная по результатам вычисления эволюционных дистанций (Kimura-2-parameter) между генами, кодирующими тимидинкиназу у 18 альфа-герпесвирусов

После добавления генов герпесвирусов кенгуру топология дендрограммы резко изменилась. Гены этих вирусов группируются вместе с генами симплексовирусов человека и обезьян. Варицелловирусы лошадей, свиней и крупного рогатого скота группируются на ветви, отдельной от симплексовирусов. Картина коэволюции, вроде бы, пропадает. Но пропадает ли?

На рис. 5 представлена дендрограмма, построенная на основании вычисления р-дистанции [3] между аминокислотными последовательностями тимидинкиназ все тех же видов вирусов с добавлением аминокислотной последовательности тимидинкиназы еще одного альфа-герпесвируса крупного рогатого скота (*B. herpesvirus 2*). К сожалению, нуклеотидная последовательность гена, кодирующего тимидинкиназу *B. herpesvirus 2*, отсутствует в GenBank.



*Рис. 5.* NJ-дендрограмма, построенная по результатам вычисления эволюционных дистанций (p-distance) между аминокислотными последовательностями тимидинкиназ 19 альфа-герпесвирусов



На рис. 6 четко видны четыре основные ветви. Верхняя ветвь объединяет все тимидинкиназы симплексвирусов. В их состав входят вирусы простого герпеса человека и обезьян, вирусы герпеса кенгуру 1 и 2, так же как и вирус герпеса крупного рогатого скота 2. Топология ветвей внутри рода симплексвирус соответствует топологии ветвей филогенетического дерева млекопитающих. Топология ветвей внутри рода варицелловирус тоже соответствует топологии ветвей филогенетического дерева их хозяев (ферменты варицелловирусов объединились на одной большой ветви, отдельно от ферментов вирусов, поражающих птиц). Получается, что коэволюция между альфа-герпесвирусами и их хозяевами действительно существует, но каждый представитель животных, по-видимому, поражается как «собственными» симплекс-, так и «собственными» варицелловирусами. Дивергенция же симплекс- и варицелловирусов произошла явно до дивергенции сумчатых и плацентарных.

Остается открытым вопрос, почему на дендрограмме с рис. 3 гены варицелловирусов крупного рогатого скота 1 и 5 и варицелловируса свиней так близки генам симплексвирусов? Обратите внимание на сходство в GC-насыщенности этих генов (см. рис. 3). Дело в том, что аминокислотный состав белков, которые кодируются генами, подверженными мутационному давлению одного направления, становится похожим. Возникает сходство между белками, не основанное на их филогенетической близости. Если добавить в выравнивание аминокислотные последовательности, кодируемые не столь GC-богатыми генами (как в случае с тимидинкиназами герпесвирусов кенгуру), но филогенетически более близкими лишь одной из групп белков (группе белков симплексвирусов человека и обезьян), сходство между белками, кодируемыми далекими, но одинаково GC-богатыми генами, заметно уменьшится.

Итак, можно заключить, что мутационное давление одного направления и силы в филогенетически далеких генах повышает степень сходства кодируемых ими белков. С другой стороны, разные направления и сила мутационного давления в филогенетически близких генах снижает степень сходства кодируемых ими белков (как в случае с варицелло-зостер-вирусом человека и варицелловирусами крупного рогатого скота 1 и 5).

Об этих эффектах мутационного давления необходимо помнить при выяснении филогенетических взаимоотношений между генами.

Сильное мутационное давление может оказывать влияние на скорость эволюции гена: в некоторых случаях она может значительно увеличиться (если кодируемый им белок не несет важной функции, несинонимичные нуклеотидные замены в нем фиксируются, как правило, за счет того, что они чаще происходят). В других случаях скорость эволюции гена, наоборот, может существенно снизиться (если кодируемый им белок несет важ-

ную функцию, то несинонимичные замены в нем не фиксируются, а синонимичные происходят реже, чем при слабом мутационном давлении).

Нельзя не остановиться на характере аминокислотных замен, вызываемых мутационным давлением. Существует шкала вероятности включения аминокислотных остатков в ту или иную разновидность вторичной структуры белка [58]. Мы рассортировали аминокислотные остатки в зависимости от GC-насыщенности кодонов, которыми они кодируются, и указали вероятность их включения в альфа-спираль (табл. 1).

Таблица 1

**Вероятность включения аминокислотных остатков в альфа-спираль [58]**

Аминокислотные остатки, кодируемые GC-богатыми кодоном (GARP)			Аминокислотные остатки, кодируемые триплетом средней GC-насыщенности			Аминокислотные остатки, кодируемые GC-бедными кодоном (FYMINK)		
Ala	1,45	сильный форм-ль	Glu	1,53	сильный форм-ль	Met	1,20	форм-ль
Arg	0,79	инидиф.	Leu	1,34	сильный форм-ль	Phe	1,12	форм-ль
Pro	0,59	сильный разр-ль	His	1,24	форм-ль	Lys	1,07	слабый форм-ль
Gly	0,53	сильный разр-ль	Gln	1,17	форм-ль	Ile	1,00	слабый форм-ль
			Trp	1,14	форм-ль	Asn	0,73	разр-ль
			Val	1,14	форм-ль	Tyr	0,61	разр-ль
			Asp	0,98	инидиф.			
			Thr	0,82	инидиф.			
			Ser	0,79	инидиф.			
			Cys	0,77	инидиф.			

*Примечание.* Аминокислотные остатки рассортированы в зависимости от GC-насыщенности кодирующих их кодонов.

Как видно из табл. 1, в группе аминокислотных остатков, которые кодируются GC-богатыми кодоном (см. прил. 1), присутствует как сильный формирователь альфа-спирали (аланин), так и два сильных разрушителя ее (пролин и глицин).

При GC-давлении, особенно при сильном, замены, приводящие к появлению этих аминокислотных остатков должны происходить довольно часто. Они должны естественным образом отражаться на вторичной структуре белка. Как показали наши исследования, у альфа-герпесвирусов при GC-давлении особенно резко растет частота использования аланина, а частоты использования пролина и глицина, в среднем, растут далеко не так резко [83]. Получается, что отрицательный отбор препятствует закреплению большинства мутаций, приводящих к разрушению альфа-спиральных участков в белках альфа-герпесвирусов. Замены, повышающие вероят-

ность образования альфа-спиральных участков, часто оказываются нейтральными.

Как видно из табл. 2, мутационное давление совместно с отрицательным отбором могут оказывать влияние и на образование бета-структурных участков белков (приводить к повышению или снижению вероятности их формирования). Например, при АТ-давлении у альфа-герпесвирусов особенно резко повышается частота использования изолейцина — сильного формирователя бета-структуры (см. гл. 2 четвертой части).

Таблица 2

**Вероятность включения аминокислотных остатков в бета-структуру [58]**

Аминокислотные остатки, кодируемые GC-богатыми кодонами (GARP)			Аминокислотные остатки, кодируемые триплетами средней GC-насыщенности			Аминокислотные остатки, кодируемые GC-бедными кодонами (FYMINK)		
Ala	0,97	слабый форм-ль	Val	1,65	сильный форм-ль	Met	1,67	сильный форм-ль
Arg	0,90	инидиф.	Cys	1,30	форм-ль	Ile	1,60	сильный форм-ль
Gly	0,81	инидиф.	Gln	1,23	форм-ль	Tyr	1,29	форм-ль
Pro	0,62	разр-ль	Leu	1,22	форм-ль	Phe	1,28	форм-ль
			Thr	1,20	форм-ль	Lys	0,74	разр-ль
			Trp	1,19	форм-ль	Asn	0,65	разр-ль
			Asp	0,80	инидиф.			
			Ser	0,72	разр-ль			
			His	0,71	разр-ль			
			Glu	0,26	сильный разр-ль			

*Примечание.* Аминокислотные остатки рассортированы в зависимости от GC-насыщенности кодирующих их кодонов.

В гл. 2 третьей части представлена табл. 16, похожая на табл. 1 и 2 из данной главы. В ней находятся вероятности включения аминокислотных остатков в линейный В-клеточный эпитоп. Как выяснилось, мутационное GC-давление повышает иммуногенность белков [87], в то время как АТ-давление, наоборот, снижает их иммуногенность.

Существуют и другие вероятностные шкалы для аминокислотных остатков, например, шкала вероятности включения их в активные центры определенного типа ферментов, шкала гидрофобности и гидрофильности [2]. Мутационное давление, совместно с отрицательным отбором, может приводить к изменению самых разнообразных свойств белковой молекулы. К тому же, мутационное давление за счет изменения частот использования нуклеотидов оказывает определенное влияние на вторичную структуру РНК, на вероятность успеха дубликации генов (см. гл. 3 первой части). Изучить все последствия, в сущности, простых биохимических процессов, вызывающих мутационное давление, представляется весьма сложной зада-

чей. В последующих главах данного издания приведены методы изучения отдельных его эффектов, представляющие определенный практический интерес для исследований в области микробиологии, вирусологии, фило-генетики и иммунологии.

Описанные нами последствия мутационного давления в узком значении этого термина прекрасно согласуются с тем, что уже давно было известно о мутационном процессе вообще. Как мы уже упоминали в гл. 2, 3, первой части, большое внимание «мутационному давлению» как фактору эволюции (в широком значении термина) уделял в своих трудах И. И. Шмальгаузен, который подчеркивал, что «эволюция строится, все же, только на мутациях» [42]. Для того чтобы завершить и обобщить первую часть монографии, а также предвосхитить те данные, которые будут изложены в последующих главах, лучше всего подходит отрывок из книги «Факторы эволюции», написанной им.

«Лишь при непрерывном появлении все новых мутаций они могут накапливаться в популяции. Единичные мутации обычно теряются, даже и в том случае, если они благоприятны для данного вида организмов в данных условиях. С другой стороны, повторно возникающие мутации могут при *достаточно высоком мутационном давлении* (курсив автора монографии) накапливаться в популяции до известной степени и в том случае, если они неблагоприятны. Они могут преодолеть «давление» отбора вплоть до достижения уровня равновесия, определяемого интенсивностью обоих противоположных процессов... Значение темпа мутирования для скорости распространения мутаций в популяции ясно само собой. Для нас важно, что при прочих равных условиях, то есть, прежде всего, при равной интенсивности элиминации (отрицательного отбора), повышенная частота возникновения мутаций (мутационное давление) ведет к увеличению концентрации мутаций и к установлению равновесия на более высоком ее уровне» [42].

## **Часть вторая**

### **Характеристика мутационного давления в генах и геномах**

#### **Глава 1**

**Анализ частот нуклеотидных замен между генами, кодирующими отдельные белки герпесвирусов (вирусов простого герпеса человека и обезьяны, варицелло-зостер-вирусов человека и обезьяны)**

**Цель исследования:** определить направление, силу и ведущий механизм мутационного давления на основании анализа гомологичных вирусных генов.

**Область применения:** данный тип молекулярно-биологического анализа может быть использован при необходимости определить направление, силу и ведущий механизм мутационного давления в геномах родственных вирусов или бактерий, если в распоряжении имеются нуклеотидные последовательности, кодирующие только один (или несколько) гомологичных генов.

**Используемые оригинальные алгоритмы:** «CodonChanges» (см. гл. 1 шестой части).

**Материалы.** Для исследования, на примере которого продемонстрированы возможности метода характеристики мутационного давления в геномах микроорганизмов на основании анализа отдельных кодирующих участков, были использованы гены, кодирующие 4 белка наружного капсида (UL6, UL19, UL25, UL38) и гликопротеин L (gL) из полных геномов вирусов простого герпеса человека 1-го и 2-го типа (ВПГ1 и ВПГ2 соответственно) и обезьяньего вируса герпеса 16 (ОВГ16); также был исследован большой капсидный белок (UL19), принадлежащий варицелло-зостер-вирусу человека (ВЗВ) и варицелло-зостер-вирусу обезьяны (ОВГ9).

**Номера доступа (идентификаторы) в GenBank.** Исследованные последовательности можно найти в GenBank в записях о нуклеотидном составе полных геномов референтных штаммов альфа-герпесвирусов. Human herpesvirus 1 (ВПГ1) [NC\_001806], Human herpesvirus 2 (ВПГ2) [NC\_001798], Cercopithecine herpesvirus 16 (обезьяний герпесвирус 16) [NC\_007653], Human herpesvirus 3 (человеческий герпесвирус 3, или варицелло-зостер-вирус) [NC\_001348], Cercopithecine herpesvirus 9 (обезьяний герпесвирус 9) [NC\_002686].

Полные геномы альфа-герпесвирусов, гены из которых были исследованы, доступны в GenBank. Характеристика мутационного давления на основании анализа этих полных геномов будет приведена в следующей главе. В данной же главе мы продемонстрировали методику исследования отдельных генов. Аналогичным образом могут быть предварительно исследованы гены микроорганизмов, чьи геномы еще полностью не просеквенированы.

#### **Описание метода.**

**Предварительные операции.** Алгоритм «CodonChanges» работает с записями о порядке расположения нуклеотидов в гене [81, 83], далее именуемыми нуклеотидные последовательности (этот термин, естественно, относится к записи в текстовом формате, а не к собственно биополимеру).

Перед анализом при помощи «CodonChanges» две нуклеотидные последовательности, кодирующие гомологичные белки, необходимо выров-

нять. Выравнивание — это процесс поиска и последующего совмещения схожих участков нуклеотидных или аминокислотных последовательностей друг с другом [13]. Выравнивание может производиться с помощью множества компьютерных алгоритмов [105]. Главное, о чем стоит помнить при работе с кодирующими белок участками, это о сохранении открытой рамки считывания. Зачастую, в результате выравнивания нуклеотидных последовательностей, рамка считывания сбивается. Поэтому мы рекомендуем переводить (транслировать) нуклеотидные последовательности в аминокислотные, производить выравнивание аминокислотных последовательностей, а затем переводить их обратно (ретранслировать) в нуклеотидные. Наилучшим образом с этой задачей справляется программа MEGA4 [124].

Алгоритм «CodonChanges» распознает следующие однобуквенные символы, обозначающие нуклеотиды: а (аденин), t (тимин), g (гуанин), c (цитозин) и символ «n», обозначающий гэп. Гэпы — это разрывы (пустые места), образующиеся в нуклеотидных и аминокислотных последовательностях после выравнивания [13]. Биологической причиной таких разрывов являются делеции и инсерции. В программе MEGA4 гэп обозначается символом «-». К сожалению, программа MS Excel, с помощью которой работает наш алгоритм «CodonChanges», иногда допускает ошибки при работе с этим символом, который может ей прочитываться и как дефис, и как минус. Чтобы избежать появления ошибок, мы стали обозначать гэп в наших электронных таблицах другим символом — латинской буквой «n».

Прежде чем скопировать текст с нуклеотидной последовательностью в соответствующую ячейку на листе «sequence1» или «sequence2» алгоритма «CodonChanges», необходимо заменить в нем символы, обозначающие гэп, на символ «n». Прodelать эту процедуру можно с помощью MS Word или любого другого текстового процессора.

**Производимые вычисления.** После того, как в соответствующую ячейку на листе «sequence1» алгоритма «CodonChanges» (эта ячейка подписана и выделена цветом) копируется нуклеотидная последовательность, на листе «gene1» рассчитываются все показатели, необходимые для характеристики мутационного давления в исследуемом гене. Аналогичные показатели рассчитываются и на листе «gene2» после копирования второй нуклеотидной последовательности в соответствующую ячейку на листе «sequence2». На листе «gene1» находится таблица с частотами нуклеотидных замен от нуклеотидной последовательности, введенной в ячейку на листе «gene1», к нуклеотидной последовательности, введенной в ячейку на листе «gene2». Согласно нашему методу, расчет частот нуклеотидных замен следует производить от нуклеотидной последовательности, менее подверженной мутационному давлению, к нуклеотидной последовательности, подверженной мутационному давлению в большей степени [81].

**Определение направления мутационного давления.** Для того чтобы определить направление мутационного давления в каждой из исследуемых последовательностей, необходимо изучить показатели, характеризующие частоту использования гуанина и цитозина в них. Если общая (суммарная) частота использования гуанина и цитозина (G+C) превышает 0,5 (то есть 50 %), то исследуемый ген, по всей вероятности, находится под влиянием мутационного GC-давления [30, 121]. Если же уровень G+C ниже 0,5, то ген, скорее всего, подвержен AT-давлению [30, 121].

Более веским доказательством GC-давления является повышенная частота использования гуанина и цитозина в третьих положениях кодонов (3GC). Если 3GC превышает как 2GC, так и 1GC, то в данном гене частоты замен AT на GC значительно превышают частоты замен GC на AT [82]. Если 3GC меньше 1GC и 2GC, то в данном гене существует AT-давление [82].

Приведенные выше критерии, при всей их убедительности, нельзя рассматривать как абсолютные. Повышение или понижение 3GC может быть вызвано внедрением tandemных повторов или сдвигом рамки считывания, а не постепенным процессом накопления единичных нуклеотидных замен преимущественного направления [82]. Чаще всего о внедрении tandemных повторов свидетельствует нарушение правила соотношения 1GC, 2GC и 3GC в кодирующих участках [82]. Во всех кодирующих участках, возникших от далеких предшественников из прогеноты (первого живого организма), выполняются следующие правила. При GC-давлении:  $2GC \ll 1GC \ll 3GC$ . При AT-давлении:  $3GC \ll 2GC \ll 1GC$ . В целом, как это ни удивительно, в подавляющем большинстве кодирующих участков, в том числе и в кодирующих участках вирусов, выполняется неравенство:  $2GC \ll 1GC$  [82].

Все изложенные выше критерии справедливы, если направление мутационного давления определяется в гене, который кодируется геномом микроорганизма, представленным двухцепочечной ДНК, в жизненном цикле которого отсутствуют фазы с одноцепочечным состоянием генома. Только в этом случае можно использовать суммарную частоту использования гуанина и цитозина, поскольку, по правилам Чаргаффа,  $G=C$  как в двухцепочечном геноме, так и внутри одной цепи ДНК [23]. Известно, что в вирусных и бактериальных генах и кодирующих геномах частота использования гуанина не всегда равна частоте цитозина, однако отклонения от равенства обычно не настолько велики, чтобы отказаться от использования показателя G+C [96].

Усомниться в том, что в жизненном цикле исследуемого микроорганизма отсутствуют фазы с одноцепочечным геномом, можно глядя на частоты использования каждого нуклеотида в каждом из трех положений кодонов (такие таблицы есть на листах «gene1» и «gene2»). Если общие частоты G и C далеки от равенства, то наиболее подходящим методом изу-

чения мутационного давления для данного микроорганизма является анализ выборки нуклеотидных последовательностей от разных штаммов и клонов (см. гл. 3 второй части). Теоретически, к неравенству G и C внутри кодирующего участка гена может приводить асимметричное мутационное давление, связанное как с транскрипцией, так и с репликацией [96].

Как видно из табл. 3, в гене, кодирующем большой капсидный белок (БКБ) вируса простого герпеса 1-го типа, существует GC-давление ( $G+C > 0,5$ ;  $3GC \gg 1GC \gg 2GC$ ). Общие частоты использования гуанина и цитозина не равны друг другу, но разница между ними не велика.

Таблица 3

**Частоты использования нуклеотидов в гене, кодирующем большой капсидный белок вируса простого герпеса 1-го типа. Лист «gene1» алгоритма «CodonChanges»**

Показатель	Значение	Показатель	Значение
G+C	0,685	1GC	0,709
GC4f	0,923	2GC	0,461
GC2f3p	0,813	3GC	0,886
1A	0,163	1T	0,128
2A	0,260	2T	0,279
3A	0,042	3T	0,072
A	0,155	T	0,160
1C	0,311	1G	0,398
2C	0,282	2G	0,179
3C	0,485	3G	0,401
C	0,360	G	0,326

**Определение ведущего механизма мутационного давления.** Путем сравнения таких показателей, как GC4f и GC2f3p (табл. 3), можно предположить ведущий механизм мутационного давления в гене [84]. Показатель GC4f равен доле гуанина и цитозина в четырехкратно вырожденных сайтах [84]. В четырехкратно вырожденных сайтах любая нуклеотидная замена является синонимичной (см. прил. 2). Показатель GC2f3p равен доле гуанина и цитозина в двукратно вырожденных сайтах, расположенных в третьих положениях кодонов [84]. В двукратно вырожденных сайтах транзиция синонимична, а трансверсии — нет (см. прил. 2). В подавляющем большинстве кодирующих участков показатель GC4f не равен показателю GC2f3p [84, 85].

Если GC4f превышает GC2f3p при GC-давлении, как в случае с геном, кодирующим большой капсидный белок ВПГ1 (табл. 3), то ведущим механизмом мутационного GC-давления следует считать повышенную частоту **трансверсий** АТ на GC. Содержание гуанина и цитозина повышается за счет трансверсий АТ на GC, в основном, в четырехкратно вырожденных сайтах, поскольку в них эти замены являются синонимичными. В то же время, содержание гуанина и цитозина может снижаться за счет транзиций



GC на AT, в основном, в двукратно вырожденных сайтах, поскольку в них эти транзиции являются синонимичными, а трансверсии обратного направления — несинонимичными.

Итак, если при GC-давлении GC4f превышает GC2f3p, то ведущим механизмом GC-давления являются трансверсии AT на GC. Если при GC-давлении GC2f3p превышает GC4f, то ведущим механизмом GC-давления являются транзиции AT на GC. Если при AT-давлении GC2f3p превышает GC4f, то ведущий механизм AT-давления — трансверсии AT на GC.

Если при AT-давлении GC4f превышает GC2f3p, то ведущим механизмом AT-давления являются *транзиции* GC на AT. Именно такая ситуация была обнаружена нами в участке, кодирующем большой капсидный белок варицелло-зостер-вируса человека (табл. 4). Как видно из табл. 4, G+C меньше 0,5, а 3GC ниже, чем 2GC и 1GC в участке, кодирующем большой капсидный протеин ВЗВ.

Таблица 4

**Частоты использования нуклеотидов в гене, кодирующем большой капсидный белок варицелло-зостер-вируса человека. Лист «gene1» алгоритма «CodonChanges»**

Показатель	Значение	Показатель	Значение
G+C	0,473	1GC	0,573
GC4f	0,414	2GC	0,440
GC2f3p	0,382	3GC	0,407
1A	0,246	1T	0,181
2A	0,268	2T	0,293
3A	0,278	3T	0,315
A	0,264	T	0,263
1C	0,256	1G	0,317
2C	0,273	2G	0,166
3C	0,218	3G	0,188
C	0,249	G	0,224

Подтвердить результаты, полученные при анализе частот использования нуклеотидов, можно путем расчета частот нуклеотидных замен между гомологичными генами (от гена, менее подверженного мутационному давлению, к гену, более подверженному ему) [81]. В нашем случае мы рассчитали частоты нуклеотидных замен от гена, кодирующего большой капсидный белок ВПГ1, к гену, кодирующему большой капсидный белок ВПГ2 (табл. 5).

Как видно из табл. 5, наибольшей является общая частота трансверсий G на C. Частота обратных трансверсий, C на G, несколько ниже. На втором месте — частота транзиций T на C, значительно превышающая частоту обратных замен. Частота транзиций A на G также превышает частоту обратных транзиций (G на A).

Таблица 5

**Частоты нуклеотидных замен от гена, кодирующего большой капсидный белок ВПГ1 (G+C = 0,68), к гомологичному гену ВПГ2 (G+C = 0,70).  
Лист «gene1» алгоритма «CodonChanges»**

Вид замены	Частота нуклеотидных замен, нуклеотидных замен на нуклеотидный сайт			
	Положения кодонов			
	1	2	3	Всего
<i>Транзиции:</i>				
С на Т	0,000727	0,001212	0,009212	0,011151
G на А	0,002424	0,001212	0,004364	0,008000
Т на С	0,001939	0,001455	0,014061	0,017455
А на G	0,002182	0,000970	0,007273	0,010425
<i>Трансверсии:</i>				
С на А	0,001212	0,000242	0,002424	0,003878
С на G	0,001939	0,001212	0,013818	0,016969
G на Т	0,000485	0	0,002667	0,003152
G на С	0,001697	0,000485	0,016000	0,018182
Т на А	0	0	0,000242	0,000242
Т на G	0,001455	0,000242	0,002424	0,004121
А на Т	0	0	0,000485	0,000485
А на С	0,000485	0,000485	0,003636	0,004606
<i>Гэпы (АК):</i>				
Ген 1	0			
Ген 2	0			

Частоты трансверсий по направлению GC-давления превышают частоты обратных трансверсий, но сами по себе они ниже, чем частоты транзиций по направлению GC-давления. Как же этот факт согласуется с тем, что ведущим механизмом GC-давления в вирусах простого герпеса являются трансверсии?

По нашему мнению, замены G на C и C на G наблюдаются не из-за того, что гуанин меняется непосредственно на цитозин (и наоборот). Судя по всему, в четырехкратно вырожденных сайтах относительно часто происходят транзиции GC на AT (табл. 5). Например, на месте гуанина образуется аденин. Далее GC-насыщенность четырехкратно вырожденных сайтов восстанавливается путем трансверсий AT на GC: на месте аденина образуется цитозин. То есть в результате двух замен на месте гуанина образуется цитозин. Наша гипотеза исчерпывающе объясняет, почему наблюдаются (табл. 5) высокие частоты трансверсий G на C и C на G при более низких наблюдаемых частотах трансверсий AT на GC.

Для изучения частот замен при AT-давлении мы рассчитали их от гена, кодирующего большой капсидный белок ВЗВ, к гомологичному и менее GC-насыщенному гену обезьяньего герпес вируса 9 (варицеллозостер-вируса обезьян, ОВГ9) [81].

Как видно из табл. 6, между гомологичными генами ВЗВ и ОВГ9, кодирующими большой капсидный белок, наибольшей является частота транзиций по направлению АТ-давления. Это согласуется с данными, полученными на основании изучения нуклеотидного состава гена, кодирующего большой капсидный белок ВЗВ. Среди трансверсий наиболее часто встречаются замены С на А. Замены Т на А и А на Т уступают им по частоте. Руководствуясь этими данными можно подтвердить наш вывод о том, что ведущим механизмом АТ-давления в генах ВЗВ являются транзиции GC на AT.

Таблица 6

**Частоты нуклеотидных замен от гена, кодирующего большой капсидный белок ВЗВ (G+C = 0,47), к гомологичному гену ОВГ9 (G+C = 0,40). Лист «gene1» алгоритма «CodonChanges»**

Вид замены	Частота нуклеотидных замен, нуклеотидных замен на нуклеотидный сайт			
	Положения кодонов			
	1	2	3	Всего
<i>Транзиции:</i>				
С на Т	0,011182	0,005948	0,034261	0,051391
G на А	0,015227	0,005472	0,023793	0,044492
Т на С	0,005234	0,005472	0,017131	0,027837
А на G	0,007614	0,002617	0,013324	0,023555
<i>Трансверсии:</i>				
С на А	0,010945	0,005234	0,016655	0,032834
С на G	0,002379	0,003093	0,004996	0,010468
G на Т	0,007138	0,000952	0,014751	0,022841
G на С	0,003093	0,002855	0,005710	0,011658
Т на А	0,004045	0,001428	0,021651	0,027124
Т на G	0,002379	0,000714	0,010707	0,013800
А на Т	0,007138	0,001428	0,022365	0,030931
А на С	0,007138	0,003093	0,008327	0,018558
<i>Гэпы (AK):</i>				
Ген 1	4			
Ген 2	8			

**Определение силы мутационного давления.** О сильном мутационном давлении можно говорить в том случае, если вероятность возникновения синонимичной замены значительно ниже, чем вероятность возникновения несинонимичной [82]. Такая ситуация имеет место после того, как третьи положения кодонов практически полностью насыщаются гуанином и цитозином при GC-давлении или, наоборот, теряют почти все данные нуклеотиды при АТ-давлении [82].

Вероятность возникновения синонимичной замены можно оценить путем расчета вероятности того, что замена, вызванная мутационным давлением, произойдет в третьем положении кодона. Дело в том, что боль-

шинство замен в третьих положениях кодонов (но не все) являются синонимичными.

Для GC-давления расчет выглядит следующим образом. Необходимо сложить разности между 1,0 и значениями 1GC, 2GC и 3GC. Данная величина представляет собой некий потенциал накопления мутаций по направлению GC-давления. Затем следует разделить разность по каждому положению на их общую сумму. В результате мы определили, что для гена, кодирующего большой капсидный белок ВПГ1, вероятность возникновения замены по направлению АТ на GC в первом положении кодона — 30,8 %, во втором положении — 57,1 %, а в третьем — только 12,1 %. Для гомологичного гена ВПГ2 вероятности возникновения мутаций по направлению GC-давления следующие: в первом положении кодона — 31,5 %, во втором — 58,9 %, в третьем — 9,6 % [81].

Для гена, кодирующего большой капсидный белок ВЗВ, потенциал накопления мутаций был рассчитан относительно АТ-давления. В данном случае эта величина равняется сумме показателей 1GC, 2GC и 3GC. Для гена, кодирующего большой капсидный белок ВЗВ, вероятность возникновения замены по направлению GC на АТ в первом положении кодона — 40,4 %, во втором — 31,0 %, в третьем — 28,6 % [81].

Вероятность возникновения замены, вызванной мутационным давлением, в третьем положении кодона гена, кодирующего большой капсидный белок ВЗВ, в 2,36 раз выше, чем в гомологичном гене ВПГ1, и в 3 раза выше, чем в гомологичном гене ВПГ2 [81].

Эти данные свидетельствуют о большей устойчивости аминокислотного состава большого капсидного белка ВЗВ к мутационному давлению, нежели чем у БКБ ВПГ. То есть мутационное давление в генах ВПГ1 и ВПГ2 сильнее, чем в гене ВЗВ [81].

Для того чтобы доказать, что при сильном мутационном давлении замены чаще накапливаются в первых и во вторых положениях кодонов, чем в третьих, можно использовать следующий тест. Для его проведения необходимы нуклеотидные последовательности, кодирующие гомологичные белки трех вирусов. Нами были выбраны ВПГ1, ВПГ2 и их ближайший родственник, поражающий обезьян, — ОВГ16. По GC-насыщенности гены этих вирусов можно расположить в следующем порядке:  $G+C$  ВПГ1 <  $G+C$  ВПГ2 <  $G+C$  ОВГ16 [81].

Первым этапом теста являются расчеты частот нуклеотидных замен от исследуемого белка ВПГ1 к исследуемому белку ВПГ2 и нуклеотидных замен от исследуемого белка ВПГ1 к гомологу из ОВГ16. После этого частоты замен в каждом из трех положений кодонов по направлению GC-давления между белками ВПГ1 и ОВГ16 необходимо разделить на соответствующие частоты замен между белками ВПГ1 и ВПГ2 [81].

Мы выполнили данный тест для генов, кодирующих четыре капсидных белка, включая БКБ, и для гена, кодирующего гликопротеин L (gL) [39]. Последний был использован по той причине, что уровень GC-насыщенности в нем ниже, чем в генах, кодирующих капсидные белки, особенно в третьих положениях кодонов [39].

Таблица 7

**Соотношение частот транзиций и трансверсий по направлению АТ на GC, рассчитанных между генами двух пар вирусов**

Ген	Положения кодонов	Частоты во 2-й паре / частоты в 1-й паре	
		транзиции АТ на GC	трансверсии АТ на GC
UL19	1	2,18	3,12
	2	3,10	2,33
	3	1,27	1,76
UL6	1	2,81	2,26
	2	2,87	2,01
	3	1,24	1,44
UL25	1	1,92	2,17
	2	1,69	3,04
	3	1,54	2,16
UL38	1	1,82	1,36
	2	2,04	3,57
	3	1,59	0,97
gL	1	1,33	1,24
	2	1,52	2,30
	3	1,62	2,39

*Примечание.* В первой паре частоты были рассчитаны от генов ВПГ1 к генам ВПГ2; во второй — от генов ВПГ1 к генам ОВГ16.

Как видно из табл. 7, для генов, кодирующих капсидные белки, прирост частоты замен по направлению АТ на GC в третьих положениях кодонов всегда ниже, чем прирост в первых или во вторых положениях. В гене, кодирующем gL, все наоборот: наибольший прирост частоты отмечен для третьих положений кодонов [39]. Получается, что тест на наличие сильного GC-давления для генов, кодирующих капсидные белки, положителен, а для гена, кодирующего гликопротеин L, — отрицателен.

Подтвердить, можно ли переносить сведения о мутационном давлении в отдельных генах на полный геном микроорганизма, представляется возможным с использованием метода, описанного в следующей главе. Если же вернуться к генам, кодирующим капсид и гликопротеин L, то, глядя на табл. 8, можно убедиться в том, что такая категория, как сила мутационного давления, несет в себе большой биологический смысл.

Таблица 8

**Уровень 3GC в четырех генах, кодирующих капсидные белки (UL19, UL6, UL25, UL38), и в гене, кодирующем гликопротеин L, исследованных вирусов (ВПГ1, ВПГ2 и ОВГ16)**

Ген	Уровень 3GC в генах		
	ВПГ1	ВПГ2	ВОГ16
UL19	0,89	0,91	0,99
UL6	0,85	0,89	0,97
UL25	0,86	0,88	0,98
UL38	0,85	0,86	0,96
gL	0,63	0,66	0,79

Действительно, ген, кодирующий гликопротеин L, сохранил гораздо больший «буфер» для замен АТ на GC (то есть в третьих положениях кодонов этого гена осталось много аденина и тимина), чем гены, кодирующие капсидные белки (табл. 8).

**Биологические проблемы, которые можно решить с использованием данного метода.** После первого контакта человека с ВПГ1, ВПГ2 и ВЗВ, эти вирусы, как правило, переходят в латентное состояние [19, 25]. Это значит, что геномы ВПГ1, ВПГ2 и ВЗВ пожизненно сохраняются в нейронах (чаще всего — в ганглиях тройничных и межпозвоночных нервов) [19, 25]. Далеко не каждая реактивация (выход из латенции — синтез полного набора белков и репликация генома) этих вирусов приводит к развитию клинически выраженного рецидива [41, 78]. По нашему мнению, рецидив инфекции, вызываемой ВПГ1, ВПГ2 и ВЗВ, связан с образованием новых эпитопов в белках этих вирусов [81]. Иммунный ответ на эти новые эпитопы через некоторое время приводит к подавлению инфекции (в отсутствие иммунодефицита), но геномы вирусов продолжают находиться в латенции.

С помощью описанного в данной главе метода можно определить направление, ведущий механизм и силу мутационного давления в генах микроорганизмов. Знание такого рода может быть использовано для исследования патогенеза вызываемых ими инфекций. В частности, сильное GC-давление, вызванное трансверсиями АТ на GC, должно способствовать частому рецидивированию герпетической инфекции (термин «герпетическая инфекция» относится к инфекции, вызванной ВПГ1 и ВПГ2) [39, 81]. Слабое АТ-давление, вызванное транзициями GC на АТ, должно приводить к более редкому рецидивированию опоясывающего лишая (вызывается ВЗВ) по сравнению с герпетической инфекцией [39, 81].

Кроме того, с помощью описанного метода можно производить отбор генов, наименее подверженных несинонимичной мутации, для дальнейшего использования в качестве антигенов при вакцинации. Возвращаясь к табл. 7 и 8, можно заключить, что гликопротеин L будет реже подвергаться несинонимичным мутациям, чем четыре исследованных капсидных белка. Аналогичным образом можно производить отбор наиболее стабильных эпитопов внутри белка (см. гл. 2 третьей части).

## Глава 2

### Анализ полных геномов альфа-герпесвирусов

**Цель исследования:** определить направление, силу и ведущий механизм мутационного давления на основании анализа полных вирусных геномов.

**Область применения:** данный тип молекулярно-биологического анализа может быть применен в случае необходимости определить направление, силу и ведущий механизм мутационного давления в геномах вирусов или бактерий, если в распоряжении имеются записи о частотах использования кодонов в каждом кодирующем участке данного микроорганизма.

**Используемые оригинальные алгоритмы:** «Coding Genome Scanner» (см. гл. 4 шестой части).

**Материалы.** Для описания метода исследования полных геномов с целью характеристики мутационного давления в них были использованы записи о частотах использования кодонов в каждом кодирующем участке десяти геномов альфа-герпесвирусов. Такие записи хранятся в Codon Usage Database ([www.kazusa.or.jp/codon](http://www.kazusa.or.jp/codon)) [104]. Периодически файлы с информацией о полных геномах из GenBank автоматически перерабатываются в таблицы использования кодонов для каждого вида. Последние хранятся в Codon Usage Database [104]. Запись с частотами использования кодонов в каждом кодирующем участке является своеобразным промежуточным продуктом между нуклеотидной последовательностью полного генома и таблицей частот использования кодонов в целом по кодирующему геному. Создатели Codon Usage Database сделали подобные записи общедоступными [104].

**Номера доступа (идентификаторы) в GenBank.** Полностью просеквенированные геномы, сведения о которых мы использовали, можно найти в GenBank под следующими номерами доступа: вирусы из рода симплекс-вирус: Human herpesvirus 1 (HSV1) [NC\_001806], Human herpesvirus 2 (HSV2) [NC\_001798], Cercopithecine herpesvirus 1 (CeHV1) [NC\_004812], Cercopithecine herpesvirus 2 (CeHV2) [NC\_006560], Cercopithecine herpesvirus 16 (CeHV16) [NC\_007653]; вирусы из рода варицелловирус: Bovine herpesvirus 5 (BoHV5) [NC\_005261], Equid herpesvirus 1 (EqHV1) [NC\_001491], Equid herpesvirus 4 (EqHV4) [NC\_001844], Cercopithecine herpesvirus 9 (CeHV9) [NC\_002686], Human herpesvirus 3 (VZV) [NC\_001348].

#### **Описание метода.**

**Предварительные операции.** Алгоритм «Coding Genome Scanner» работает с записями о частотах использования кодонов в каждом кодирующем участке исследуемого генома [40, 83, 84, 85]. Такая запись из Codon Usage Database представляет собой текстовый файл в формате «txt». Создаются эти файлы автоматически (с помощью специального компью-

терного алгоритма) [104]. По этой причине перед использованием следует просматривать содержание каждого из них. Довольно часто в запись о полном геноме включаются не только кодирующие участки из референтного штамма, но и другие записи. Иногда в один файл могут быть включены несколько записей о полных геномах разных штаммов вируса. Для работы согласно описанному в данной главе методу достаточно использовать записи о частотах использования кодонов исключительно в полном геноме референтного штамма.

Текстовый файл из Codon Usage Database записан следующим образом. Информация о кодирующем участке занимает две строки. В первой строке содержатся сведения о кодирующем участке (номер белка в GenBank, длина и источник поступления), во второй строке — количество каждого из шестидесяти четырех кодонов в кодирующем участке. Если кодирующий участок относится к полностью просеквенированному геному, то в первой строке имеется запись «complete genome» [104]. В задачу исследователя входит удаление записей, которые не относятся к полностью просеквенированному геному. Такие записи всегда располагаются перед последовательным изложением сведений о кодирующих участках из полного генома или же после них.

Удалять строки со сведениями о «лишних» кодирующих участках можно непосредственно в MS Excel.

**Производимые вычисления.** Для того чтобы алгоритм «Coding Genome Scanner» произвел все необходимые вычисления, необходимо скопировать текстовый файл со сведениями о частотах использования кодонов в каждом кодирующем участке на лист «All CDSs». После этого на листе «List of indexes» рассчитываются все показатели, необходимые для характеристики мутационного давления в исследуемом кодирующем геноме [40, 83].

**Определение направления мутационного давления.** Повышенный уровень 3GC во всех кодирующих участках генома говорит о том, что в процессе мутагенеза замен АТ на GC в среднем происходит больше, чем замен GC на АТ. Другими словами, высокий уровень 3GC — свидетельство GC-давления, а низкий — свидетельство АТ-давления. То есть если уровень 3GC выше 0,5 у большинства генов, то весь геном подвержен GC-давлению. Если уровень 3GC ниже 0,5, то в геноме существует АТ-давление [84].

В геноме вируса, состоящего из двухцепочечной ДНК, так же как и в геноме бактерии, могут быть участки с относительно повышенным или пониженным содержанием G+C [60]. Для того чтобы их обнаружить, мы предлагаем проводить в начале исследования следующий тест. Поскольку кодирующие участки в файле из Codon Usage Database расположены последовательно [104], появляется возможность сканировать содержание G+C, 1GC, 2GC и 3GC в генах на протяжении генома. Для этого достаточно построить диаграмму с перечисленными показателями, расположенны-



ми по оси Y. При этом ось X будет отражать порядок расположения кодирующих участков.

На рис. 6 изображено распределение G+C и 3GC в кодирующих участках варицелло-зостер-вируса на протяжении его генома. Четко видно, что в большинстве кодирующих участков существует АТ-давление (G+C и 3GC меньше 0,5). Тем не менее на рис. 6 выделяются шесть кодирующих участков с повышенным содержанием G+C и 3GC. Эти участки располагаются в больших инвертированных повторах (LTR). Последние характерны для всех альфа-герпесвирусов. Участки LTR полностью идентичны друг другу, но расположены в разных направлениях (обратно комплементарны) [33, 84, 126]. В связи с описанной особенностью геномов альфа-герпесвирусов мы изучали распределение частот использования нуклеотидов в генах, расположенных в LTR, отдельно [33, 84].

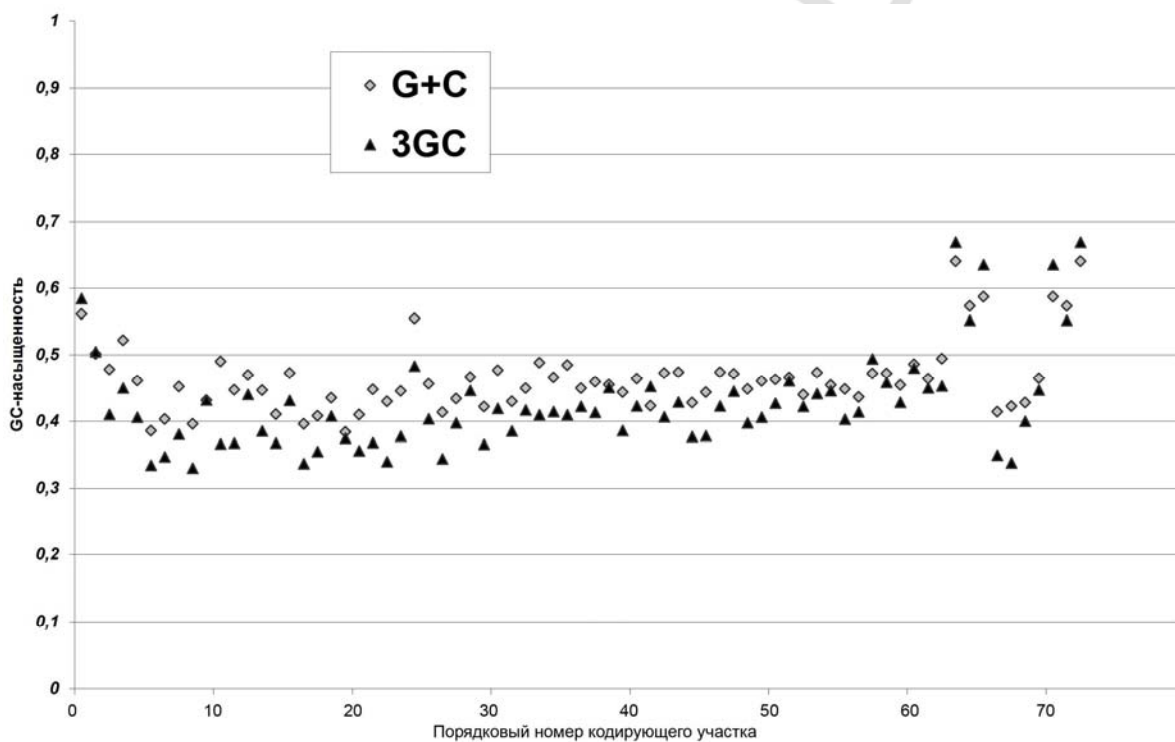


Рис. 6. Общая GC-насыщенность (G+C) и GC-насыщенность в третьих положениях кодонов (3GC) генов из полного генома варицелло-зостер-вируса

Встречаются случаи, когда в одном геноме существуют области с разными направлениями мутационного давления. Действительно, в большинстве генов ВЗВ существует АТ-давление, в то время как гены из LTR подвержены GC-давлению [33, 84].

Исследование полных геномов позволяет применять средства описательной статистики для изучения частот использования нуклеотидов: открывается возможность получить средние по кодирующим геномам значения показателей и сравнить различные показатели между собой.

В табл. 9 приведены средние значения показателей GC-насыщенности трех положений кодонов для генов из уникальных областей (UL и US) и инвертированных повторов (LTR) десяти альфа-герпесвирусов.

**Определение ведущего механизма мутационного давления.** Третьи положения кодонов неоднородны: они представлены, в основном, двукратно и четырехкратно вырожденными сайтами. В четырехкратно вырожденных сайтах любая замена является синонимичной (нуклеотид может мутировать путем одной транзиции и двух трансверсий), в двукратно вырожденных сайтах третьих положений кодонов синонимична только транзиция (см. прил. 2) [84]. При наличии разницы между GC-насыщенностью четырехкратно (GC4f) и двукратно вырожденных сайтов в третьих положениях кодонов (GC2f3p) можно определить лидирующий механизм мутационного давления [84]. Теоретически GC-насыщенность может повышаться (или снижаться) за счет неравенства частот возникновения как разнонаправленных транзиций, так и разнонаправленных трансверсий.

Если при GC-давлении GC4f превышает GC2f3p, то ведущим механизмом GC-давления являются трансверсии AT на GC. Если при GC-давлении GC2f3p превышает GC4f, то ведущий механизм GC-давления — транзиции AT на GC. Если при AT-давлении GC2f3p превышает GC4f, то ведущий механизм AT-давления — трансверсии AT на GC. Если при AT-давлении GC4f превышает GC2f3p, то ведущий механизм AT-давления — транзиции GC на AT [84, 85].

Средние значения показателей GC4f и GC2f3p для генов из уникальных областей (UL и US) и инвертированных повторов (LTR) десяти альфа-герпесвирусов приведены в табл. 9.

На рис. 7 размещен график зависимости GC4f и GC2f3p от G+C для генов из полного генома ВПГ1. Поразительно, но факт: уровень GC2f3p практически у всех генов ВПГ1 значительно ниже уровня GC4f [84].

На рис. 8 размещен аналогичный график для генов из полного генома ВЗВ. Разность между средними значениями GC4f и GC2f3p не является результатом наличия таковой лишь у отдельных генов. Графики на рис. 7 и 8 позволяют заключить, что речь в данном случае идет о существовании общегеномных закономерностей.

Таблица 9

**Уровни GC-насыщенности в трех положениях кодонов (1GC, 2GC, 3GC), в четырехкратно (GC4f) и в двукратно вырожденных сайтах, расположенных в третьих положениях кодонов (GC2f3p), рассчитанные для всех кодирующих участков, находящихся в уникальных областях (UL+US) и в больших инвертированных повторах (LTR) симплекс- и варицелловирусов [35, 84]**

Имя вируса, регион	Уровни GC-насыщенности в положениях кодонов и сайтах				
	1GC	2GC	3GC	GC4f	GC2f3p
HSV1	0,670±0,011	0,530±0,012	0,796±0,014	0,831±0,015	0,725±0,018

<i>UL+US</i>					
<b>HSV1</b> <i>LTR</i>	0,787±0,079	0,688±0,043	0,922±0,045	0,930±0,043	0,886±0,052
<b>HSV2</b> <i>UL+US</i>	0,685±0,011	0,539±0,013	0,831±0,015	0,865±0,015	0,757±0,022
<b>HSV2</b> <i>LTR</i>	0,793±0,073	0,689±0,031	0,928±0,039	0,928±0,046	0,929±0,049
<b>CeHV1</b> <i>UL+US</i>	0,720±0,011	0,556±0,015	0,913±0,012	0,924±0,012	0,888±0,016
<b>CeHV1</b> <i>LTR</i>	0,825±0,040	0,680±0,017	0,952±0,080	0,949±0,089	0,960±0,055
<b>CeHV2</b> <i>UL+US</i>	0,738±0,012	0,565±0,015	0,936±0,011	0,947±0,011	0,910±0,015
<b>CeHV2</b> <i>LTR</i>	0,817±0,085	0,660±0,011	0,971±0,054	0,967±0,061	0,978±0,044
<b>CeHV16</b> <i>UL+US</i>	0,738±0,012	0,566±0,015	0,943±0,012	0,950±0,013	0,924±0,014
<b>CeHV16</b> <i>LTR</i>	0,826±0,070	0,654±0,022	0,950±0,089	0,940±0,111	0,975±0,040
<b>BoHV5</b> <i>UL+US</i>	0,748±0,011	0,578±0,016	0,916±0,013	0,923±0,012	0,901±0,018
<b>BoHV5</b> <i>LTR</i>	0,761±0,185	0,676±0,025	0,923±0,047	0,929±0,044	0,899±0,059
<b>CeHV9</b> <i>UL+US</i>	0,461±0,011	0,401±0,012	0,308±0,011	0,312±0,015	0,281±0,013
<b>CeHV9</b> <i>LTR</i>	0,601±0,040	0,513±0,067	0,534±0,069	0,575±0,078	0,478±0,090
<b>VZV</b> <i>UL+US</i>	0,515±0,009	0,435±0,011	0,410±0,011	0,444±0,013	0,355±0,015
<b>VZV</b> <i>LTR</i>	0,633±0,054	0,549±0,023	0,619±0,069	0,658±0,042	0,561±0,100
<b>EqHV1</b> <i>UL+US</i>	0,570±0,012	0,473±0,014	0,614±0,012	0,614±0,014	0,610±0,014
<b>EqHV1</b> <i>LTR</i>	0,676±0,044	0,567±0,053	0,882±0,032	0,893±0,009	0,871±0,012
<b>EqHV4</b> <i>UL+US</i>	0,541±0,010	0,456±0,012	0,481±0,010	0,456±0,014	0,500±0,012
<b>EqHV4</b> <i>LTR</i>	0,665±0,053	0,587±0,002	0,751±0,062	0,745±0,051	0,760±0,079

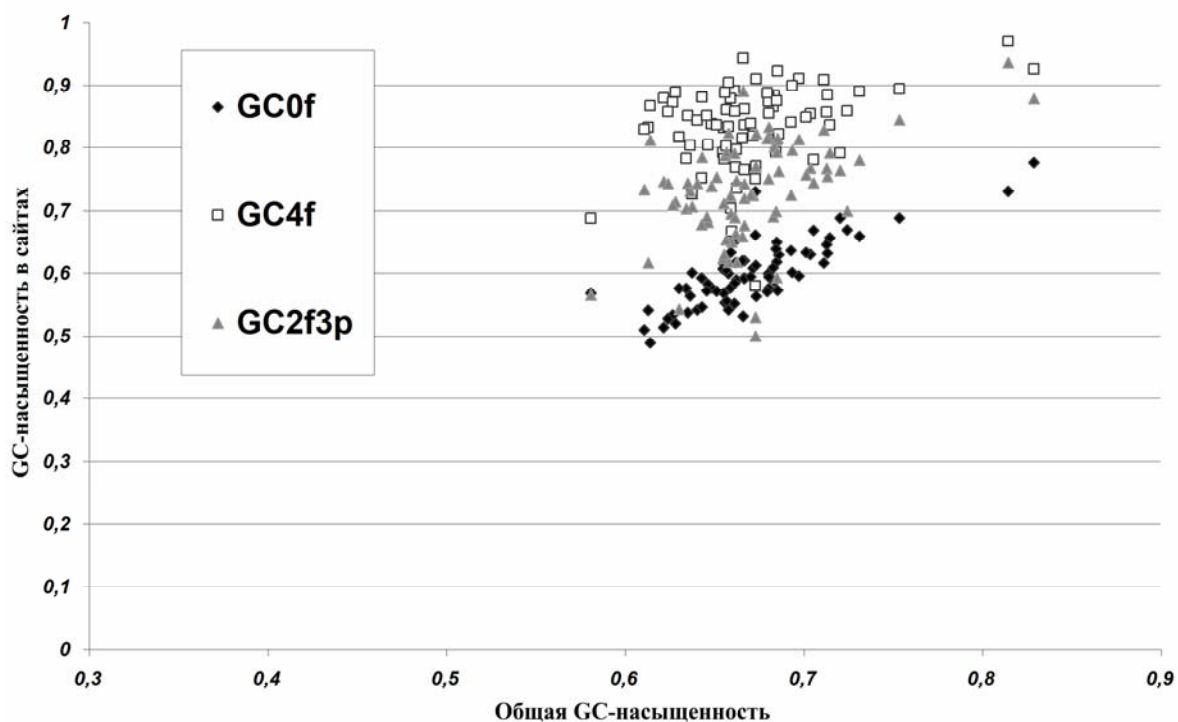


Рис. 7. Зависимость GC-насыщенности в четырехкратно вырожденных сайтах (GC4f), двукратно вырожденных сайтах, расположенных в третьих положениях кодонов, (GC2f3p) и в нолькратно вырожденных сайтах (GC0f) генов из полного генома ВПГ1 от их общей GC-насыщенности (G+C)

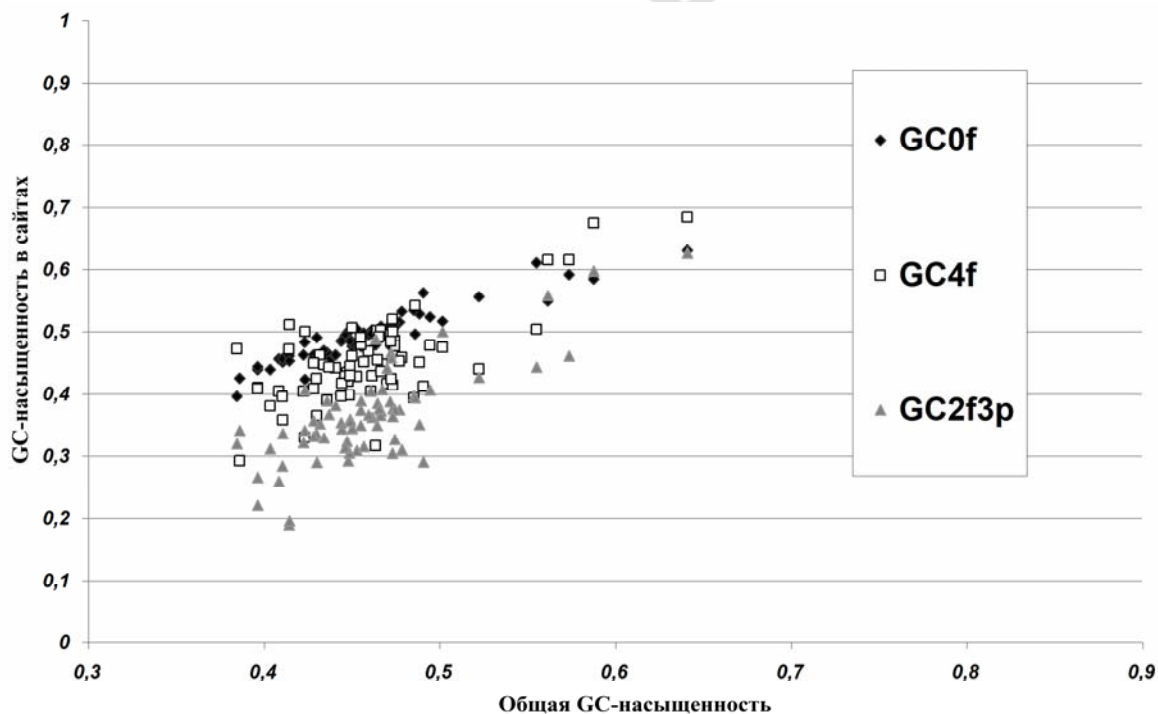


Рис. 8. Зависимость GC-насыщенности в четырехкратно вырожденных сайтах (GC4f), двукратно вырожденных сайтах, расположенных в третьих положениях кодонов, (GC2f3p) и в нолькратно вырожденных сайтах (GC0f) генов из полного генома ВЗВ от их общей GC-насыщенности (G+C)

**Определение силы мутационного давления.** Нами предложен новый критерий для определения силы мутационного давления в крупных геномах, представленных двухцепочечной ДНК [84, 85]. По мере увеличения GC-насыщенности третьих положений кодонов, в них сокращается количество субстрата для замен по направлению GC-давления (снижается содержание аденина и тимина). Чем больше количество гуанина и цитозина в третьих положениях кодонов, тем выше вероятность обратных замен GC на AT. Исходя из этих представлений, рост уровня 3GC (так же, как и его снижение) со временем должен остановиться на определенном уровне. Свидетельством того, что уровень 3GC стабилизировался во всех кодирующих участках данного генома, является низкий коэффициент корреляции ( $R < 0,5$ ) между 3GC и G+C [84, 85]. Если уровни 3GC у всех генов в геноме выровнялись на одном (высоком) уровне, можно констатировать наличие сильного мутационного GC-давления [84, 85]. В такой ситуации мутационное давление должно часто вызывать замены в первых и во вторых положениях кодонов, GC-насыщенность в которых значительно ниже, чем в третьих [84, 85]. При сильном AT-давлении, наоборот, замены чаще будут происходить в GC-богатых вторых и первых положениях кодонов, а не в потерявших большую часть гуанина и цитозина третьих положениях [84, 85]. Все замены во вторых и большинство замен в первых положениях кодонов несинонимичны [105]. Из этого следует, что биологический смысл сильного мутационного давления заключается в том, что оно повышает вероятность возникновения несинонимичной замены относительно синонимичной [82].

В табл. 10 приведены коэффициенты корреляции 1GC, 2GC и 3GC с общей GC-насыщенностью для участков, расположенных в уникальных областях десяти геномов альфа-герпесвирусов.

Таблица 10

**Коэффициенты корреляции 1GC, 2GC и 3GC с общей GC-насыщенностью для всех кодирующих участков, расположенных в уникальных регионах (UL+US) симплекс- и варицелловирусов**

Имя вируса	R (1GC от G+C)	R (2GC от G+C)	R (3GC от G+C)
HSV1	0,817	0,537	0,350
HSV2	0,872	0,499	0,474
CeHV1	0,791	0,668	0,195
CeHV2	0,854	0,719	0,226
CeHV16	0,798	0,694	0,148
BoHV5	0,813	0,642	0,352
CeHV9*	0,676	0,787	0,470
VZV	0,744	0,790	0,786
EqHV1	0,676	0,600	0,593
EqHV4	0,694	0,594	0,636

\* При расчете коэффициентов корреляции из числа кодирующих участков UL и US регионов CeHV9 был исключен недавно переместившийся из LTR в UL ген ICP0.

Коэффициент корреляции, при всей его демонстративности, нельзя абсолютизировать. Одна выдающаяся точка может привести к увеличению  $R$ , хотя для остальных точек всякая зависимость между показателями отсутствует. Вместе с тем зависимость между двумя показателями, имеющаяся для большей части точек, может быть скрыта (значение  $R$  станет меньше 0,5) при наличии одной точки, которая не ложится на общую прямую. Чтобы избежать неправильного определения силы мутационного давления, мы рекомендуем строить графики зависимости 1GC, 2GC и 3GC от G+C для всех генов исследуемого генома. Главное в данном методе — раскрыть биологический смысл сильного мутационного давления, что не всегда удается сделать, руководствуясь исключительно значением коэффициента корреляции.

Мы приводим графики, изображающие зависимости между 1GC, 2GC, 3GC и общей GC-насыщенностью для отдельных геномов альфа-герпесвирусов (рис. 9, 10). На этих рисунках обозначены некоторые гомологичные гены, которые могут смещаться из области LTR в уникальные области и наоборот, а также ген, находящийся в LTR у всех альфа-герпесвирусов (ICP4) [84]. Глядя на такие графики можно не только оценивать характер общей зависимости 3GC от G+C, но и находить гены с нестандартным распределением GC-насыщенности по положениям кодонов. Как правило, такие гены включают в себя тандемные повторы или другие «макрмутации».

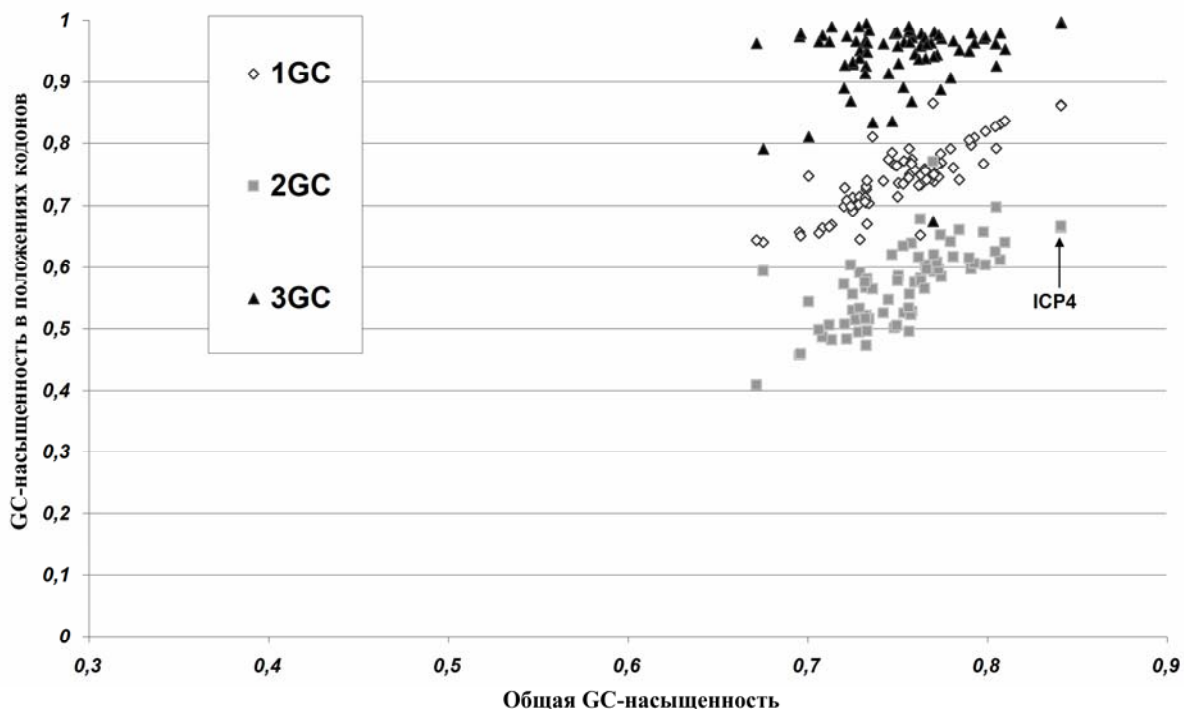


Рис. 9. Зависимость GC-насыщенности в первых (1GC), вторых (2GC) и в третьих (3GC) положениях кодонов в генах из полного генома ВГО16 от их общей GC-насыщенности (G+C)

На рис. 9 можно наблюдать характерную для сильного GC-давления «крышу», состоящую из значений 3GC [84]. На рис. 10 показана картина, характерная для сильного AT-давления: значения 3GC «лежат на полу» [84].

График на рис. 10 демонстрирует наличие сильного AT-давления в уникальных областях ВГО9, но коэффициент корреляции 3GC и G+C превышает 0,5. Причиной этому является ген ICP0, который по своей GC-насыщенности гораздо ближе к генам из LTR (ICP22, US10 и ICP4), чем к генам из уникальных областей. По нашему мнению, этот кодирующий участок переместился из LTR в уникальную область генома относительно недавно, вследствие чего GC-насыщенность в нем не успела упасть до уровня, характерного для генов из уникальных областей [84].

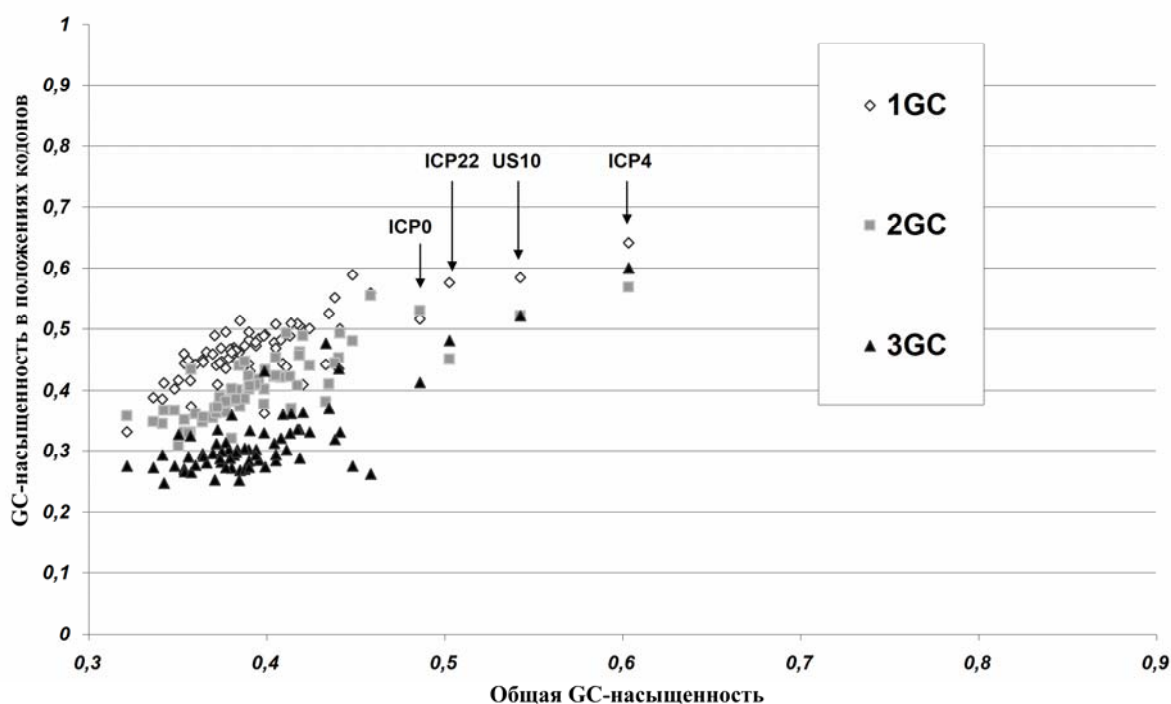


Рис. 10. Зависимость GC-насыщенности в первых (1GC), вторых (2GC) и в третьих (3GC) положениях кодонов в генах из полного генома ВГО9 от их общей GC-насыщенности (G+C)

**Общее заключение о направлении, ведущем механизме и силе мутационного давления.** Как видно из табл. 9, в геномах всех пяти симплексовирусов и в геноме Bovine herpesvirus 5 существует сильное (см. табл. 10) мутационное GC-давление, вызванное преимущественно повышенной частотой возникновения трансверсий AT на GC. Под этим следует понимать то, что отношение частоты трансверсий AT на GC к частоте трансверсий GC на AT выше, чем отношение частоты транзиций AT на GC к частоте противоположных по направлению транзиций [84].

В геноме EqHV1 мутационное GC-давление слабое. Судя по близким значениям GC4f и GC2f3p (см. табл. 9), отношение частот возникновения

разнонаправленных транзиций примерно равно отношению частот разнонаправленных трансверсий [84].

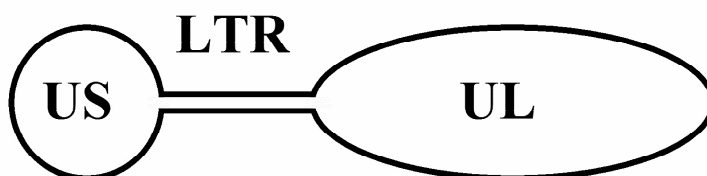
В геноме CeHV9 существует сильное мутационное АТ-давление (см. табл. 9, 10), вызванное, преимущественно, транзициями GC на АТ [84]. В геноме VZV слабое мутационное давление также вызвано, в основном, повышенной частотой транзиций GC на АТ [84].

В областях UL и US EqHV4 существует слабое АТ-давление, вызванное трансверсиями GC на АТ [84].

GC-насыщенность генов, находящихся в LTR, у всех симплекси и варицелловирусов гораздо выше, чем G+C генов из уникальных областей [84]. 3GC в генах из LTR всегда выше 50 % (см. табл. 9). Вирусные гены в процессе эволюции могут перемещаться из области LTR в область UL или в область US, могут дублироваться, попадая в LTR из US [84]. При этом у генов, сместившихся в уникальную область, 3GC снижается, а у генов, попавших в LTR, 3GC возрастает [84]. Это наводит на мысль о том, что причина повышения 3GC кроется в неких свойствах инвертированных повторов, а не в особенностях генов, находящихся в них.

Как показал наш эксперимент *in-silico*, отношение транзиций к трансверсиям достоверно ниже у генов, находящихся в LTR, чем у генов из UL и US [84]. Исходя из этих данных можно сделать вывод о том, что GC-насыщенность в генах, находящихся в LTR, растет из-за повышенной частоты трансверсий АТ на GC. Такие трансверсии могут происходить только в двухцепочечной ДНК, при участии ферментов системы репарации. Следовательно, большие инвертированные повторы альфа-герпесвирусов существуют в двухцепочечной форме дольше, чем их уникальные регионы [84]. Такой вывод правомочен, поскольку две копии инвертированных повторов, находящиеся на одной цепи ДНК, комплементарны друг другу.

На основании результатов *in-silico* анализа геномов альфа-герпесвирусов представляется возможным доказать факт образования вторичной структуры из их одноцепочечной геномной ДНК [84]. Такая структура (две петли на одной ножке) может образовываться непосредственно после репликации генома по типу катящегося кольца (рис. 11) [92].



*Рис. 11.* Схема вторичной структуры, которая может образовываться на одном из этапов жизненного цикла альфа-герпесвирусов. Уникальные области (UL и US) представлены петлями из одноцепочечной ДНК, инвертированные повторы (LTR) комплементарно соединены друг с другом, они образуют двухцепочечный участок ДНК



Причиной мутационного давления может быть не только высокая частота возникновения биохимических модификаций нуклеотидов (их окисления и дезаминирования), но и недостаточность процессов репарации. В геномах всех альфа-герпесвирусов находится ген, кодирующий урацил-ДНК-гликозилазу [47]. Этот фермент вырезает урацил (продукт дезаминирования цитозина) из ДНК, предотвращая возникновение транзиции С на U (G на A — на комплементарной цепи) [47]. Наши данные позволяют предположить снижение активности урацил-ДНК-гликозилазы в геномах CeHV9 и VZV [33].

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью описанного метода можно определить направление, ведущий механизм и силу мутационного давления как в полных геномах, так и в отдельных генах микроорганизмов. Открываются возможности для изучения мутационного давления как в целом по геному, так и в отдельных его участках. Знание такого рода может быть использовано для исследования патогенеза вирусных и бактериальных инфекций. Например, как в случае с альфа-герпесвирусами, можно предполагать наличие ранее неизвестных фаз в жизненном цикле [84].

С помощью описанного метода гораздо легче производить отбор генов, наименее подверженных несинонимичной мутации, чем при помощи метода, описанного в предыдущей главе.

Изучение полных геномов микроорганизмов позволяет делать более обоснованные выводы о причинах, вызывающих мутационное давление. Зная причину мутационного давления, можно в терапевтических целях пытаться влиять на его интенсивность. Некоторые исследователи уже рассматривают подобные стратегии борьбы с вирусными заболеваниями [77, 109] (подробнее см. гл. 3 второй части).

### Глава 3

#### **Анализ выборки нуклеотидных последовательностей гена *env*, полученных от разных штаммов и клонов вируса иммунодефицита человека 1-го типа**

**Цель исследования:** определить направление, силу и ведущий механизм мутационного давления на основании анализа нуклеотидных последовательностей гена, полученных от разных штаммов и клонов вируса.

**Область применения:** данный тип молекулярно-биологического анализа может быть применен в случае необходимости определить направление, силу и ведущий механизм мутационного давления в геноме вируса или бактерии, если в распоряжении имеются нуклеотидные последова-

тельности одного или нескольких генов, полученных от разных штаммов и клонов данного микроорганизма.

**Используемые оригинальные алгоритмы:** «VVK Consensus» (см. гл. 5 шестой части).

**Материалы.** Для описания метода исследования выборки нуклеотидных последовательностей с целью характеристики мутационного давления в геноме вируса были использованы 667 нуклеотидных последовательностей, полученных от 33 пациентов. Эти последовательности были добавлены в GenBank в ходе пяти независимых исследований.

**Номера доступа (идентификаторы) в GenBank.** Пять групп последовательностей (от пяти пациентов) были получены в ходе исследования изменчивости вируса под воздействием гуморального иммунитета [57], их идентификаторы в GenBank: EU744175–EU743973. Две группы — в ходе работы по изучению тропизма вируса [66], их идентификаторы: EU604642–EU604549. Еще девять групп были получены теми же авторами и с той же целью, но при передаче данных в GenBank произошла ошибка: имя гена во всех этих последовательностях (EF575486–EF575363) было указано неправильно (вместо *env* их обозначили как *pol*). Шесть групп последовательностей (EU852934–EU853141) были получены в работе, посвященной изучению передачи вируса половым путем [113]. Одиннадцать групп генов (DQ410040 to DQ410649) были получены в ходе исследования мутагенеза ВИЧ1 у так называемых «элитных контроллеров», то есть у ВИЧ-инфицированных пациентов, способных «контролировать» интенсивность репликации вируса [44]. Мы выбрали для анализа группы последовательностей, в которых количество трансверсий превышало 20. В результате только четыре группы, принадлежащие «элитным контроллерам», попали в наше исследование. Остальные семь групп принадлежат «быстрым прогрессорам» — они использовались в упомянутой работе [44] в качестве контрольных.

#### **Методика.**

**Предварительные операции.** Перед подсчетом количества нуклеотидных замен в каждой выборке последовательности необходимо выровнять. Мы использовали с этой целью программу MEGA 4 [124]. Методом выравнивания был избран PAM [105].

Алгоритм «VVK Consensus» способен обработать до 100 нуклеотидных последовательностей [86]. «VVK Consensus» распознает следующие однобуквенные символы, обозначающие нуклеотиды: а (аденин), т (тимин), г (гуанин), с (цитозин), — и символ «n», обозначающий гэп. Прежде чем скопировать текст с выборкой нуклеотидных последовательностей в соответствующие ячейки на листе «sequences», необходимо заменить в них символы, обозначающие гэп, на символ «n». Прodelать эту процедуру можно с помощью MS Word или любого другого текстового редактора.

Как уже было сказано выше, качество расчета процентного соотношения нуклеотидных замен зависит от их количества: чем больше их число, тем достоверней их соотношения. В связи с этим мы рекомендуем использовать выборки с достаточным количеством замен. Общее число замен на сайт (*per site*) и на нуклеотид (*per nucleotide*) указывается на листе «results». На этом листе располагаются две таблицы с количеством замен всех типов *per site* и *per nucleotide* соответственно.

**Производимые вычисления.** Алгоритм «VVK Consensus» производит расчет количества нуклеотидных замен (на листе «results»), вычисляет их соотношения (на листе «calculations») и рассчитывает нуклеотидный состав каждой введенной последовательности (на листе «content») сразу же после того как все выровненные кодирующие участки попадают в соответствующие ячейки на листе «sequences».

Вопрос о том, как именно надо подсчитывать количество и определять направление нуклеотидной замены в выборке последовательностей, до сих пор остается открытым. Здесь мы приводим наш метод подсчета замен от консенсусной последовательности. В каждом сайте консенсусной последовательности находится наиболее часто используемый в нем (среди всех последовательностей из данной группы) нуклеотид [13, 23].

С помощью «VVK Consensus» можно подсчитывать количество нуклеотидных замен двумя способами. При подсчете первым способом (при подсчете «*per site*») алгоритм суммирует количество сайтов, содержащих данную мутацию [35]. Процесс подсчета мутаций «*per site*» можно выразить следующей формулой:

$$N_{XY} = \sum_{i=X} S_Y, \quad (3.1)$$

где  $N_{XY}$  — количество замен нуклеотида X на нуклеотид Y;  $S_Y$  — сайт, в котором содержится нуклеотид Y;  $i$  — сайт консенсусной последовательности, содержащий нуклеотид X.

Расчет «*per site*» основан на гипотезе о том, что аналогичные нуклеотидные замены в одном и том же сайте разных последовательностей являются результатом одной единственной замены, имевшей место в их общей предковой последовательности [35].

Второй способ (подсчет «*per nucleotide*») рассматривает по отдельности все нуклеотидные замены, произошедшие в одном и том же сайте разных последовательностей. Для подсчета замен «*per nucleotide*» используется следующая формула:

$$N_{XY} = \sum_{i=X} nS_Y, \quad (3.2)$$

где  $N_{XY}$  — количество замен нуклеотида X на нуклеотид Y;  $nS_Y$  — количество последовательностей, содержащих нуклеотид Y в данном сайте;  $i$  — сайт консенсусной последовательности, содержащий нуклеотид X.

Способ подсчета «*per site*» лучше подходит для нуклеотидных последовательностей, в которых мутации происходят редко [35], а «*per nucleotide*» — для нуклеотидных последовательностей часто мутирующих генов [35]. Поскольку невозможно четко определить, является ли данный ген редко или часто мутирующим, алгоритм «VVK Consensus» производит подсчет обоими способами. В данной главе приведены результаты подсчета нуклеотидных замен как «*per site*», так и «*per nucleotide*».

При подсчете нуклеотидных замен от консенсусной последовательности велика вероятность неправильного определения направления мутации. Причем чем чаще происходят мутации, тем выше вероятность неправильного определения их направления данным способом [35]. Если, например, мутация G на A, часто происходящая в генах ВИЧ [35, 48, 49], зафиксировалась в более чем 50 % последовательностей, ее направление будет ошибочно определено как замена A на G. Поэтому в данной работе мы оперировали суммами замен противоположных направлений (то есть, возвращаясь к приведенному примеру, мы вычисляли сумму замен G на A и A на G). После этого мы определяли их направление по методу, близкому к таковому, рассмотренному в гл. 1 второй части.

**Определение ведущего механизма мутационного давления.** Ведущий механизм нуклеотидных замен по данному методу определяется путем непосредственного сравнения процентных соотношений замен разных типов.

**Результаты подсчета нуклеотидных замен «*per site*».** В табл. 11 приведено процентное соотношение замен в спектре мутагенеза ВИЧ1 для каждой из проанализированных 33 групп последовательностей. Для наглядности мы приводим рис. 12, где отражены средние значения для процентного соотношения замен в спектре мутагенеза ВИЧ1.

Таблица 11

**Процентное соотношение замен в спектре мутагенеза ВИЧ1 при подсчете «*per site*» в 33 группах генов *env***

Имя группы	Число посл-ей	Число замен	G↔A, %	C↔T, %	C↔A, %	G↔T, %	A↔T, %	G↔C, %
S33	50	300	53,1	24,8	11,2	2,7	6,5	1,7
S35	40	402	48,9	25,2	12,0	4,5	7,2	2,2
H5	30	487	40,3	21,6	13,8	10,5	10,7	3,1
H4	49	542	50,6	18,1	13,1	7,6	6,5	4,1
H3	42	499	43,1	21,0	16,1	8,1	8,7	3,0
H2	41	536	44,8	22,0	13,9	8,1	9,6	1,7
H1	41	685	41,1	18,3	15,5	7,8	12,1	5,3
DM9	16	293	46,0	22,7	12,7	6,0	9,0	3,7
DM8	16	388	45,6	18,0	15,6	5,5	10,9	4,4
DM7	11	124	45,8	33,3	6,9	6,9	4,2	2,8
DM6	15	185	47,5	25,7	12,8	8,4	4,5	1,1
DM5	15	260	39,7	19,8	19,1	8,9	8,2	4,3

Имя группы	Число посл-ей	Число замен	G↔A, %	C↔T, %	C↔A, %	G↔T, %	A↔T, %	G↔C, %
DM4	14	109	53,7	21,1	12,6	3,2	7,4	2,1
DM3	16	244	44,5	21,8	13,0	6,3	10,1	4,2
DM2	14	183	49,2	24,9	12,2	3,3	5,5	5,0
DM1	7	187	47,9	23,6	10,0	6,4	10,0	2,1
C98	17	225	46,0	15,6	13,4	10,7	8,5	5,8
C109	11	159	44,8	18,2	20,1	6,5	7,1	3,2
C96	9	176	30,7	15,7	15,7	14,5	13,9	9,6
C94	13	375	45,7	15,7	14,9	9,1	10,5	4,1
C93	15	304	36,6	19,3	19,7	10,5	8,1	5,8
C62	15	164	47,6	16,1	12,5	8,9	13,1	1,8
C61	15	203	37,6	23,4	14,1	11,2	10,2	3,4
ES4	18	118	40,7	22,0	18,6	8,5	8,5	1,7
ES7	16	225	42,5	21,7	12,8	10,6	6,2	6,2
ES8	47	138	41,3	20,3	14,5	12,3	6,5	5,1
ES9	27	221	46,8	21,8	15,5	9,5	4,5	1,8
9F	8	240	47,2	19,3	13,2	7,1	8,6	4,6
605F	8	226	45,6	23,5	8,8	7,4	10,3	4,4
605M	8	345	41,7	21,4	13,6	6,8	12,5	4,1
32M	7	191	39,8	20,4	12,0	8,9	14,7	4,2
183M	8	221	41,5	17,1	15,2	5,1	16,6	4,6
120F	8	223	42,7	20,7	14,5	4,8	11,0	6,2

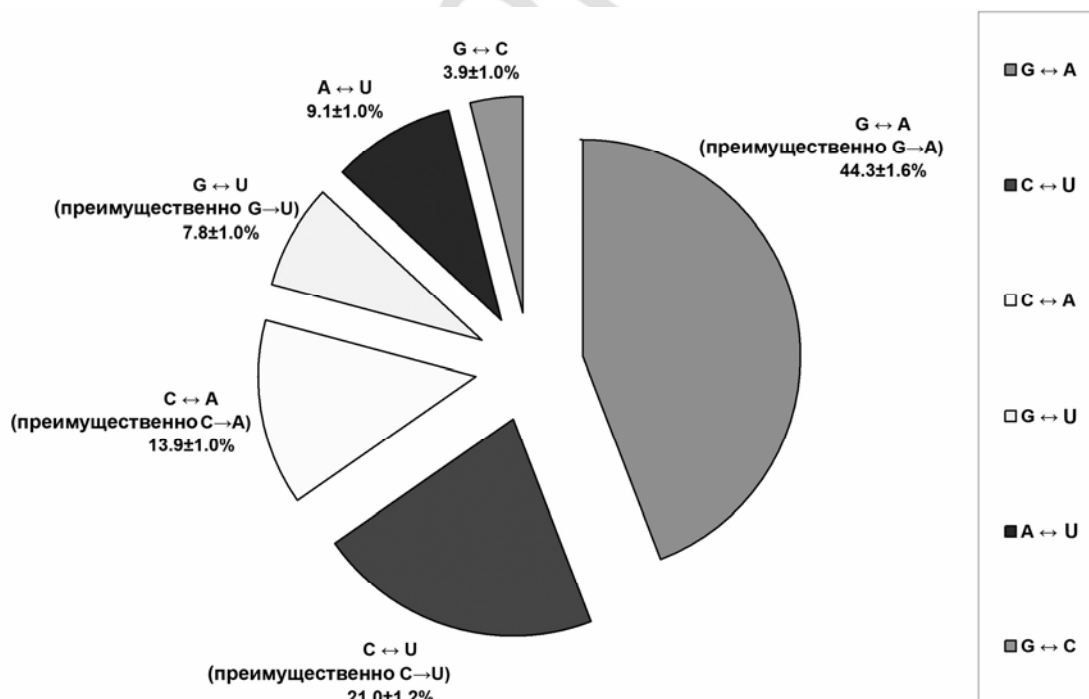


Рис. 12. Процентное соотношение замен в спектре мутагенеза ВИЧ1 при подсчете «per site»

Соотношение транзиций (G↔A/C↔U) равно  $2,15 \pm 0,13$  (граница доверительного интервала определена при  $P < 0,05$ ) [35].

Соотношение изменяющих GC-насыщенность трансверсий (C↔A/G↔U) равно  $2,02 \pm 0,29$  ( $P < 0,05$ ) [35].

Соотношение неизменяющих GC-насыщенность трансверсий (A↔U/G↔C) равно  $2,79 \pm 0,48$  ( $P < 0,05$ ) [35].

Различия в процентном соотношении между всеми типами замен (рис. 12) достоверны. Для разности процентов замен G↔U и A↔U получен уровень достоверности  $P < 0,05$ , для всех остальных замен различия еще более достоверны ( $P < 0,001$ ) [35].

**Результаты подсчета нуклеотидных замен «per nucleotide».** В табл. 12 приведено процентное соотношение замен в спектре мутагенеза ВИЧ1 для каждой из проанализированных 33 групп последовательностей, на рис. 13 показаны средние значения для процентного соотношения замен в спектре мутагенеза ВИЧ1.

Таблица 12

**Процентное соотношение замен в спектре мутагенеза ВИЧ1 при подсчете «per nucleotide» в 33 группах генов env**

Имя группы	Число посл-ей	Число замен	G↔A, %	C↔T, %	C↔A, %	G↔T, %	A↔T, %	G↔C, %
S33	50	1096	48,8	25,4	16,4	5,6	2,7	1,1
S35	40	1795	49,4	22,7	13,6	6,1	6,3	1,9
H5	30	1921	43,3	18,8	17,5	9,7	8,3	2,5
H4	49	2845	47,9	15,5	16,8	5,7	5,6	8,4
H3	42	2484	42,5	17,2	20,4	9,4	8,8	1,7
H2	41	2295	45,9	21,9	10,5	8,4	11,4	1,8
H1	41	3381	46,0	15,0	14,3	8,0	11,5	5,1
DM9	16	780	43,7	20,0	14,2	7,8	10,8	3,5
DM8	16	1115	42,7	15,2	19,1	5,3	12,7	5,0
DM7	11	206	44,8	34,3	8,2	8,2	2,2	2,2
DM6	15	380	48,0	23,0	12,3	10,2	5,2	1,3
DM5	15	1078	36,7	20,3	20,9	8,2	9,1	4,8
DM4	14	351	58,2	17,1	12,2	4,5	7,3	0,7
DM3	16	899	39,5	19,6	18,3	7,1	11,5	3,9
DM2	14	292	52,9	23,0	12,0	2,4	6,2	3,4
DM1	7	380	48,5	22,3	10,2	6,2	10,9	1,8
C98	17	615	50,4	13,6	16,6	7,7	6,6	5,1
C109	11	302	46,8	15,3	17,6	8,8	8,1	3,4
C96	9	334	31,2	17,7	19,0	12,5	11,0	8,6
C94	13	700	48,3	13,6	15,1	8,4	9,9	4,7
C93	15	834	35,0	19,7	18,7	10,1	11,1	5,4
C62	15	600	51,0	14,0	15,6	4,7	12,5	2,3
C61	15	542	38,1	21,5	17,1	11,8	9,4	2,2

Имя группы	Число посл-ей	Число замен	G↔A, %	C↔T, %	C↔A, %	G↔T, %	A↔T, %	G↔C, %
ES4	18	425	37,3	24,8	17,9	7,8	10,1	2,1
ES7	16	751	43,1	20,6	13,2	12,0	5,0	6,2
ES8	47	580	46,7	25,7	15,3	6,6	4,3	1,4
ES9	27	1149	47,6	18,7	16,5	11,8	3,7	1,7
9F	8	359	49,0	19,6	12,8	6,1	9,0	3,5
605F	8	352	45,3	21,4	8,8	8,3	11,7	4,6
605M	8	476	42,7	21,9	12,4	7,0	12,0	4,1
32M	7	289	37,0	22,5	12,5	8,0	16,6	3,5
183M	8	303	43,3	16,5	15,1	4,6	16,2	4,2
120F	8	341	42,9	19,7	15,4	5,2	11,3	5,5

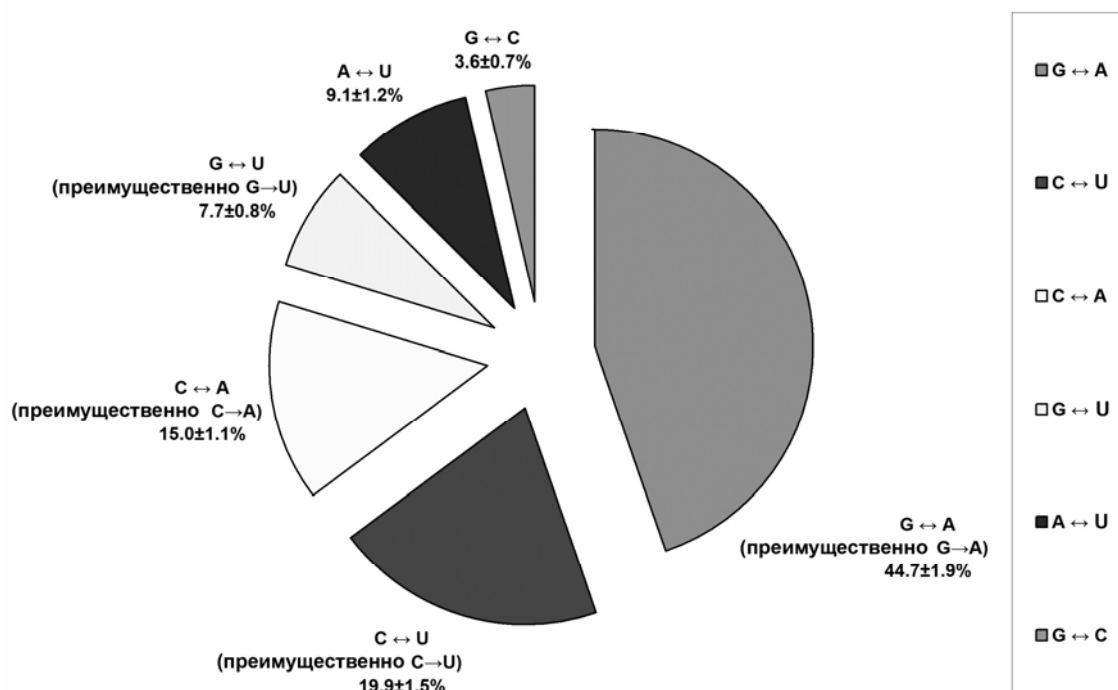


Рис. 13. Процентное соотношение замен в спектре мутагенеза ВИЧ1 при подсчете «per nucleotide»

Соотношение транзиций (G↔A/C↔U) равно  $2,35 \pm 0,22$  ( $P < 0,05$ ) [35].

Соотношение изменяющих GC-насыщенность трансверсий (C↔A/G↔U) равно  $2,16 \pm 0,29$  ( $P < 0,05$ ) [35].

Соотношение неизменяющих GC-насыщенность трансверсий (A↔U/G↔C) равно  $3,20 \pm 0,67$  ( $P < 0,05$ ) [35].

Различия в процентном соотношении всех замен, за исключением такого для G↔U и A↔U (рис. 13), достоверны ( $P < 0,001$ ) [35].

**Определение направления мутационного давления.** Поскольку в жизненном цикле ВИЧ присутствует фаза, в которой его геном представлен

одноцепочечной РНК плюс цепью, так же как и фаза, в которой он представлен частично одноцепочечной ДНК минус цепью [23], ожидать выполнения правил Чаргаффа в его генах не приходится. Получается, что мутационное давление в генах ВИЧ имеет не одно, а несколько направлений.

Чтобы определить предпочтительное направление мутации, мы использовали теорию мутационного давления [121]. Чем больше процент содержания данного нуклеотида, особенно в третьих положениях кодонов, тем чаще должны происходить замены, приводящие к увеличению частоты его использования [122]. Таким образом были определены направления наиболее частых мутаций в генах *env* ВИЧ1.

Дополнительной функцией алгоритма «VVK Consensus» является расчет частот использования нуклеотидов в трех положениях кодонов анализируемых последовательностей [86].

В изученных нуклеотидных последовательностях гена *env* ВИЧ1 частота использования аденина в третьих положениях кодонов (3А) равна  $(36,84 \pm 0,31) \%$ , в то время как частота использования гуанина в них (3G) равна  $(21,16 \pm 0,22) \%$  [35]. Согласно теории мутационного давления [121], большинство замен, определенных нами как замены  $G \leftrightarrow A$ , должны иметь направление G на A.

Аналогично, большинство замен  $C \leftrightarrow U$  должно иметь направление C на U, так как  $3C = (16,49 \pm 0,21) \%$ , а  $3U = (25,50 \pm 0,31) \%$  [35].

Большинство замен  $C \leftrightarrow A$  имело направление C на A, так как  $3A > 3C$  [35].

Большинство замен  $G \leftrightarrow U$  имело направление G на U, поскольку  $3U > 3G$  [35].

**Определение силы мутационного давления.** Сила мутационного давления должна определяться отдельно для каждого из его направлений. В качестве примера можно привести наиболее часто встречающиеся в процессе мутагенеза ВИЧ замены G на A [48, 49]. По нашей теории, чем ниже содержание G в третьих положениях кодонов и чем выше оно в первых и во вторых, тем больше вероятность того, что замена G на A будет несинонимичной [86]. Этому вопросу практически полностью посвящены гл. 1, 2 третьей части.

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью описанного метода можно определить направление, ведущий механизм и силу мутационного давления в генах микроорганизмов на основании анализа выборки (или нескольких выборок) нуклеотидных последовательностей гена (или нескольких генов) от разных штаммов и клонов. Преимущественное направление и тип нуклеотидных замен во многом определяются особенностями жизненного цикла вируса. Располагая информацией о мутационном давлении можно предположить наличие определенных особенностей жизненного цикла. С целью



подтверждения биологической важности результатов, которые можно получить с использованием данного метода, необходимо подробнее рассмотреть особенности жизненного цикла и мутагенеза ВИЧ.

В конце 90-х гг. было открыто семейство белков, обладающих цитидин дезаминазной активностью по отношению к минус цепи ДНК ВИЧ1 [134]. О существовании фермента, редактирующего мРНК аполипопротеина В (сокращенно АРОВЕС), было известно и до этого открытия [62]. Поскольку открытые ферменты, дезаминирующие цитидин в одноцепочечной ДНК, обладают высокой степенью гомологии с ферментом, дезаминирующим цитидин в мРНК аполипопротеина В, они были объединены в общее суперсемейство под названием АРОВЕС [62]. Причем собственно АРОВЕС (редактирующий мРНК) отнесли к семейству АРОВЕС1, а цитидин-ДНК-дезаминазы — к семейству АРОВЕС3 [62].

Подавляющее большинство современных статей о мутагенезе ВИЧ1 начинаются общей фразой: «Человеческие ферменты из семейства АРОВЕС3 дезаминируют цитидин в ДНК минус цепях ВИЧ1 во время обратной транскрипции, вызывая G на A гипермутагенез» [77]. В нашем исследовании был подтвержден тот известный факт, что наиболее частым типом мутаций в генах ВИЧ1 является транзиция G на A [48, 49, 77].

Замены C на U, являющиеся вторыми по частоте возникновения, судя по всему, вызываются дезаминированием цитозина в РНК ВИЧ1 [35]. Считается, что ферменты из семейства АРОВЕС3 специфичны к одноцепочечной ДНК и не вызывают дезаминирование цитидина в одноцепочечной РНК [110]. Что же касается АРОВЕС1, то в недавних экспериментах на крысиной модели было доказано, что этот фермент способен дезаминировать цитозин в РНК ВИЧ1 [51]. Мышиный АРОВЕС1 является мощным «мутатором» как для клеточной, так и для вирусной РНК [108]. Судя по нашим данным [35], АРОВЕС1 человека так же способен дезаминировать цитозин в РНК ВИЧ1.

Возникновение трансверсий C на A и G на U связано с окислением гуанина [72, 81]. Если гуанин окисляется в ДНК минус цепи, то при синтезе на ней РНК плюс цепи возникает замена C на A, так как окисленный гуанин (8-оксо-G) комплементарен не цитозину, а аденину. Если же гуанин окислится в РНК плюс цепи, то во время обратной транскрипции может возникнуть замена C на A в минус цепи, что впоследствии должно привести к замене G на U в новом поколении РНК плюс цепей.

Согласно экспериментальным исследованиям, обратная транскриптаза ВИЧ1 чаще встраивает напротив 8-оксо-G цитозин, чем аденин [63]. Однако при транскрипции напротив 8-оксо-G чаще должен оказываться аденин, чем цитозин [63, 72]. Получается, что при синтезе ДНК-минус цепи обратной транскриптазой 8-оксо-G часто оказывается напротив цитозина, а во время транскрипции с этой ДНК-минус цепи напротив встроенного 8-оксо-

G часто появляется аденин — замены C на A происходят часто. Напротив же гуанина, окисленного в РНК, аденин во время обратной транскрипции оказывается редко — замены G на U происходят реже. Но неправильно спаренная нуклеотидная пара 8-оксо-G:C репарируется клеточными ферментами: 8-оксо-G заменяется на G [72]. Неужели система репарации не работает с интегрированными в хромосомы провирусами ВИЧ?

Как оказалось, далеко не каждый двухцепочечный геном ВИЧ может интегрироваться в человеческую хромосому [71]. Неудача в процессе встраивания в хромосому может постигнуть вирус в случае нарушения функции интегразы, так же как и в результате других причин. На один интегрированный приходится около 100 неинтегрированных геномов [71]. Такой неинтегрированный геном ВИЧ, как показали недавние экспериментальные исследования, не является «тупиковой ветвью». С него может осуществляться транскрипция ранних генов [71]. В случае же наличия в клетке одного интегрированного вируса, транскрипция со всех неинтегрированных геномов осуществляется в полном объеме. Происходит это за счет экспрессии с интегрированного провируса генов, в частности гена *vpr*, необходимых для транскрипции поздних вирусных генов с неинтегрированных геномов [110]. Процесс «помощи» интегрированного вирусного генома в осуществлении транскрипции поздних генов и, что самое главное, геномной РНК с неинтегрированных геномов называется комплементацией [71, 133].

Результаты наших исследований [35] указывают на то, что комплементация действительно является широко распространенным процессом. Геномные РНК, синтезированные на матрицах ДНК минус цепей неинтегрированных геномов, вносят значительный вклад в общую вирусную популяцию. По нашему мнению, ДНК минус цепи неинтегрированных геномов — своеобразный накопитель мутаций, в том числе окисленного гуанина (8-оксо-гуанина) и дезаминированного цитозина (урацила). Это должно быть связано с тем, что *долгоживущие ДНК минус цепи неинтегрированных геномов не подвергаются репарации* [35].

Некоторые исследователи предлагают усилить процесс G на A гипермутационного давления в геномах ВИЧ путем ингибирования гена *vif* [77]. Продукт гена *vif* способствует ускорению протеолиза, в том числе и протеолиза цитидин-дезаминаз, таких как АРОВЕС1 и АРОВЕС3 [77]. Но можно ли считать усиление мутационного давления панацеей? Будет ли происходить обещанное исчезновение вирусной популяции?

По нашему мнению, усиление мутационного давления будет иметь гораздо больший положительный эффект для вируса, чем для человека. Чем чаще будут возникать мутации, тем быстрее будет происходить уход от иммунного ответа и выработка устойчивости к лекарственным препаратам [86].

## Часть третья Уровень мутабельности кодирующих участков

### Глава 1

#### Поиск участков с высокой и низкой мутабельностью в генах *env* вирусов иммунодефицита человека 1-го и 2-го типа

**Цель исследования:** найти наиболее и наименее подверженные несинонимичным мутациям участки гена *env* вируса иммунодефицита человека 1-го типа; сравнить уровень мутабельности участка гена *env*, кодирующего V3-петлю вируса иммунодефицита человека 1-го типа (ВИЧ1), с уровнем мутабельности гомологичного участка вируса иммунодефицита человека 2-го типа (ВИЧ2).

**Область применения:** данный тип молекулярно-биологического анализа применяется для поиска наименее мутабельных участков в вирусных генах, что может быть использовано при отборе наиболее устойчивых эпитопов для создания вакцин.

**Используемые оригинальные алгоритмы:** «VVK in length», «VVK Consensus» и «VVK Dinucleotides» (см. гл. 2, 5, 6 шестой части).

**Материалы.** В первой части данной главы мы изучали распределение гуанина по трем положениям кодонов на протяжении гена *env* референтного штамма ВИЧ1. Во второй части мы искали ответ на вопрос о том, можно ли переносить знания об участках, кодирующих V3-петлю gp120, полученные на основании изучения референтных штаммов, на популяции ВИЧ1 и ВИЧ2. Для этого были использованы выборки нуклеотидных последовательностей, кодирующих V3-петлю, из разных штаммов и клонов ВИЧ1 и ВИЧ2, по 100 штук для каждого вируса. В третьей части главы мы изучали мутабельность участков, кодирующих V3-петлю gp120, полученных от двух субсов, один из которых был медленным прогрессором ВИЧ-инфекции, а другой — быстрым. В конце главы описан метод подсчета частот использования динуклеотидов на примере участков, кодирующих консервативные и вариабельные регионы gp120 из референтного штамма ВИЧ1.

**Номера доступа (идентификаторы) в GenBank.** Геномы референтных штаммов ВИЧ1 и ВИЧ2 доступны в GenBank под следующими номерами: NC\_001802 и NC\_001722 соответственно. Выборка нуклеотидных последовательностей *env* ВИЧ1 состоит из 100 перечисленных ниже: EU664612–EU664683; U70809–U70821; U70597–U70614. Выборка из 100 нуклеотидных последовательностей *env* ВИЧ2: U24287–U24388. К сожалению

нию, на момент написания данной главы было выполнено только одно исследование ВИЧ-инфицированных сибсов, один из которых был быстрым прогрессором, а второй — медленным [112]. Материал этого исследования представляет собой 44 нуклеотидные последовательности, полученные во время «мониторинга» изменчивости ВИЧ1 у быстрого прогрессора EF657933–EF657976, и 42 последовательности, полученные от медленного прогрессора: EF657890–EF657932 [112].

#### **Описание метода.**

**Предварительные операции.** Перед тем как заняться поиском наиболее и наименее подверженных несинонимичным мутациям участков вирусного или бактериального гена, необходимо определить преимущественное направление нуклеотидных замен в нем, то есть направление мутационного давления. Методы определения направления мутационного давления описаны во второй части. Конкретно для гена *env* ВИЧ1 основные направления нуклеотидных замен были определены в гл. 3 второй части.

Алгоритмы «VVK in length» и «VVK Dinucleotide», так же как и «VVK Consensus», распознают следующие однобуквенные символы, обозначающие нуклеотиды: а (аденин), т (тимин), г (гуанин), с (цитозин), и символ «n», обозначающий гэп. Если алгоритм «VVK Consensus» используется исключительно для расчета частот использования нуклеотидов, как в данной главе, предварительное выравнивание изучаемых последовательностей можно не производить.

**Производимые вычисления.** Алгоритм «VVK in length» производит расчет частот использования нуклеотидов в участках одинаковой длины, нарезаемых из нуклеотидной последовательности, введенной в соответствующую ячейку на листе «full sequence». Длина каждого участка задается на листе «length». Результаты подсчетов находятся на листе под названием «results». На листе «parts» находятся «нарезанные» участки нуклеотидной последовательности [86]. Во избежание неправильной трактовки результатов необходимо обращать внимание на длину последнего участка. Зачастую последний участок оказывается слишком коротким.

Алгоритм «VVK Dinucleotides», так же как и «VVK in length», способен обработать только одну нуклеотидную последовательность, в то время как «VVK Consensus» способен производить расчеты в 100 последовательностях одновременно [86]. С помощью «VVK Dinucleotides» можно рассчитать частоту использования каждого динуклеотида в гене [86]. Динуклеотид — это два последовательно расположенных нуклеотида. Существует мнение, что некоторые ферменты, вносящие замены в нуклеотидные последовательности, имеют специфичность к определенным динуклеотидам. Например считается, что цитозин-ДНК-метилтрансфераза метилирует цитозин, находящийся исключительно в CG динуклеотидах [23, 67]. После дезаминирования метилированный цитозин превращается в тимин [72].

По разности в частотах использования CG динуклеотидов и GC динуклеотидов многие исследователи делали заключение о том, подвержен ли ген или геном метилированию [67].

Различные ферменты из семейства АРОВЕС3 предпочтительно дезаминируют цитозин, находящийся в определенных динуклеотидах. Поскольку процесс дезаминирования происходит на ДНК минус цепи, в кодирующем участке «целевые» динуклеотиды будут обратно комплементарны реально существующим. Например, АРОВЕС3G предпочтительно дезаминирует цитозин в динуклеотиде CC, производя замены CC на CU [109]. Это значит, что в кодирующем участке из-за активности АРОВЕС3G будут происходить замены GG на AG. Упомянутый фермент также способен вызывать замены в динуклеотидах (относительно кодирующего участка) GA и GC [109]. Ферменты АРОВЕС3С, АРОВЕС3F и АРОВЕС3D3 вызывают замены GA на AA [53, 61, 95].

Из этих данных следует, что горячими точками для мутаций, вызываемых ферментами из семейства АРОВЕС3, являются динуклеотиды GA, GG и GC [86]. В специфичности АРОВЕС3 можно усомниться по той причине, что ген *env* обогащен аденином, в результате чего гуанин, как правило, оказывается окруженным им. Частота использования динуклеотида GA высока: замены G на A должны происходить чаще в наиболее часто встречающемся динуклеотиде. Именно из-за этого может создаваться впечатление о специфичности АРОВЕС3.

Генетический код, как известно, состоит из тринуклеотидов, а не из динуклеотидов. Динуклеотид может занимать три позиции относительно положений кодона. А именно: есть динуклеотиды, находящиеся в первых и во вторых положениях кодона, во вторых и в третьих, в третьих и в первых. Важно рассчитывать частоты их использования по отдельности [80]. Именно такую возможность и предоставляет алгоритм «VVK Dinucleotides».

**Изучение частоты использования нуклеотида на протяжении гена.** Ведущий механизм нуклеотидных замен в геноме ВИЧ и в гене *env*, в частности — это транзиции G на A. Такой вывод можно сделать не только на основе результатов расчетов, описанных в гл. 3 второй части, но и по результатам многочисленных экспериментальных исследований [48, 49, 53, 61, 95].

Наиболее часто в генах ВИЧ мутациям подвергается гуанин. Именно поэтому мы сконцентрировали свое внимание на распределение гуанина по положениям кодонов на протяжении гена *env* ВИЧ1. Как видно на рис. 14, частота использования G подвергается заметным колебаниям на протяжении гена *env* ВИЧ1. Мы отметили на этом рисунке части гена, в которых частота использования гуанина в третьих положениях (3G) значительно ниже 1G и 2G. Именно в этих участках вероятность того, что замена G на A будет несинонимичной (произойдет в первом или во втором

положении кодона, а не в третьем), должна быть высокой. Мы разбили ген *env* на участки по 40 кодонов. При этом последний участок оказался слишком коротким (3 кодона) для того, чтобы говорить о высоком уровне его мутабельности. Второй участок (рис. 14) кажется наименее мутабельным (уровень 3G в нем достигает наибольшего значения, в то время как 2G — наименьшего среди участков по 40 кодонов).

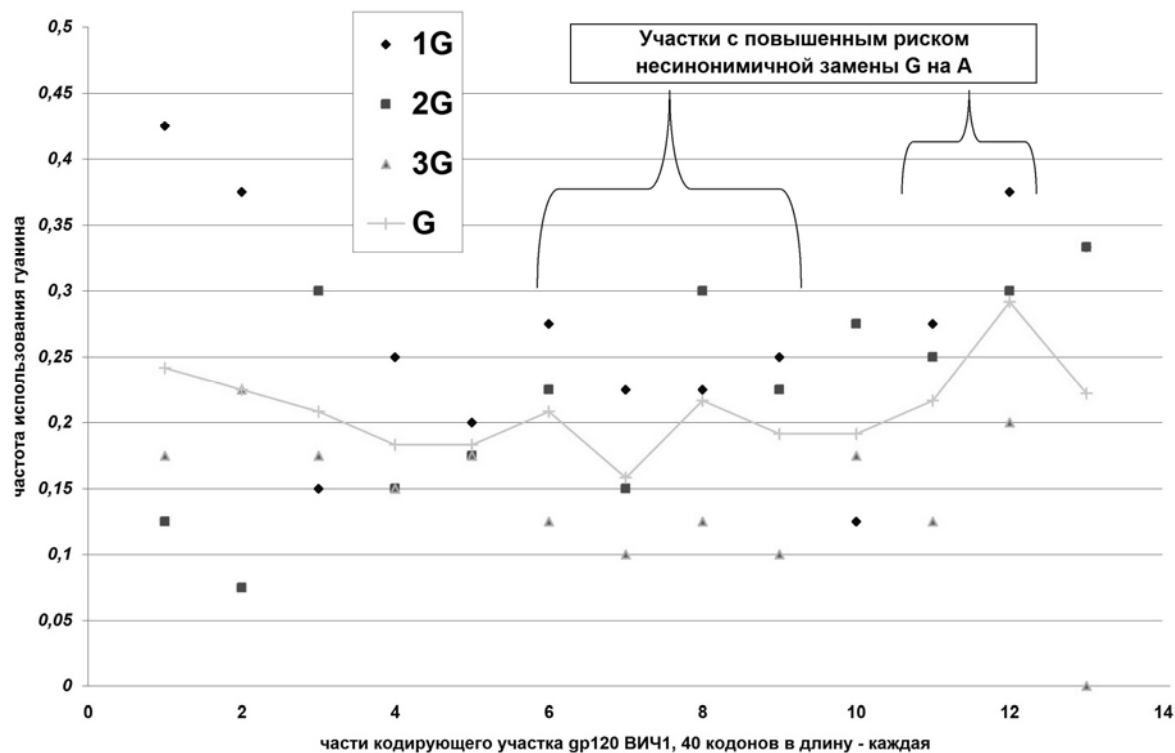


Рис. 14. Распределение гуанина по положениям кодона на протяжении гена *env* ВИЧ1

Большое значение имеет не только отношение 3G к 1G и 2G, но и общий уровень частоты использования G. При одинаковых вероятностях возникновения несинонимичной замены наиболее мутабельным необходимо признавать тот участок, в котором выше общее содержание часто мутирующего нуклеотида [86].

**Сравнение уровней мутабельности гомологичных кодирующих участков.** Согласно большинству публикаций [114, 119, 132], наиболее изменчивым участком поверхностного гликопротеина ВИЧ1 является его третий варибельный участок, так называемая V3-петля. На рис. 14 участок, кодирующий V3-петлю, расположен (частично) в части 7 и 8. Действительно, глядя на рис. 14, можно предполагать, что изменчивость V3-петли во многом обусловлена высокой вероятностью несинонимичной замены в участке, который ее кодирует.

Еще один интереснейший факт известен о V3-петле: к ней вырабатывается наибольшая часть антител при ВИЧ-инфекции (нельзя не упомянуть

о данных, противоречащих этому утверждению: не в каждом случае V3-петля является наиболее иммуногенной и вариабельной) [19, 57, 119].

Известно, что V3-регион ВИЧ2 гораздо менее изменчив, чем V3-регион ВИЧ1 [114]. В данной части главы мы продемонстрировали тот факт, что участок, кодирующий V3-регион ВИЧ2, в гораздо большей степени защищен от несинонимичных замен, чем гомологичный участок ВИЧ1 [86].

По нашим представлениям, гипервариабельность какого-либо участка белка может быть вызвана двумя основными причинами. Если аминокислотные замены происходят с одинаковой вероятностью на протяжении всего белка, то чаще всего они будут фиксироваться на наименее функционально важных его участках. То есть первая причина гипервариабельности участка белка — его низкая функциональная значимость [16, 86]. Второй причиной может быть его гипермутабельность, вызванная особенностями нуклеотидной последовательности, в которой он закодирован [86]. Под термином «гипермутабельность» мы понимаем высокую вероятность того, что нуклеотидная замена будет несинонимичной [86].

При сравнении рис. 15 и 16 можно прийти к выводу, что уровень 3G в участке, кодирующем V3-регион ВИЧ1, значительно ниже, чем уровень 3G в участке, кодирующем V3-регион ВИЧ2. Более того, уровень 3G в участке, кодирующем V3-регион ВИЧ1, стремится к нулю, в то время как уровень 2G в нем колеблется от 30 до 40 %. В участке, кодирующем V3-регион ВИЧ2, уровень 3G, наоборот, превышает уровень 2G.

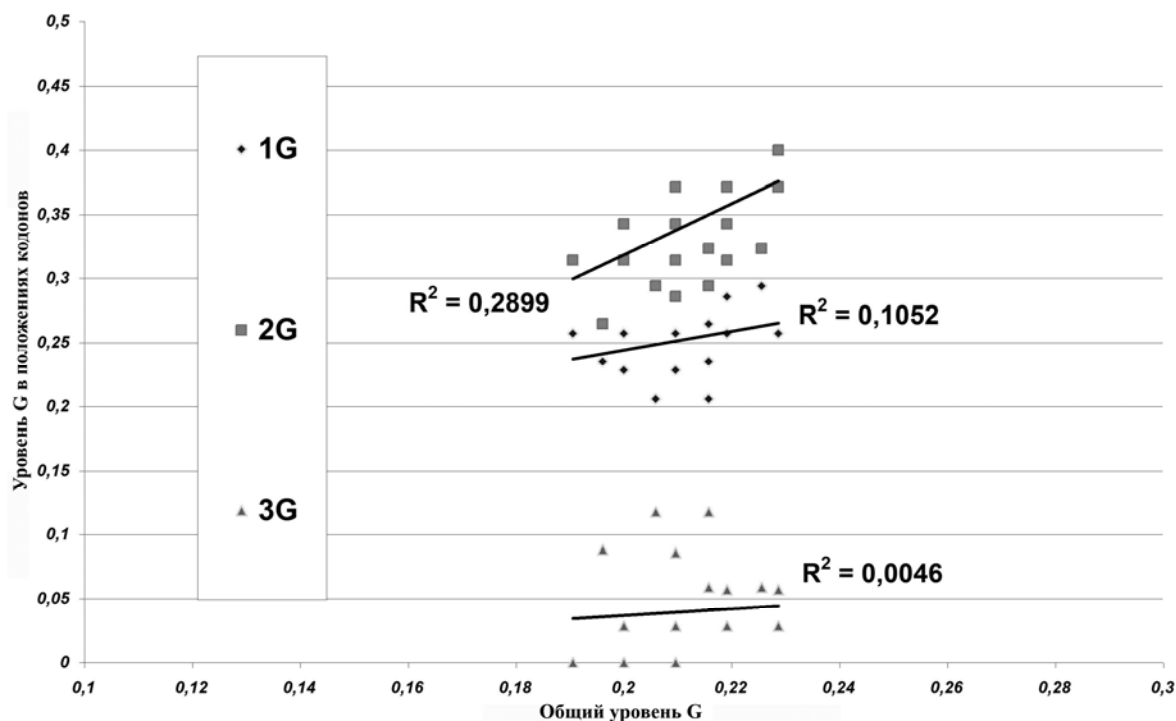


Рис. 15. Зависимость между частотами использования гуанина в трех положениях кодонов 100 нуклеотидных последовательностей участка, кодирующего V3-регион, от разных штаммов и клонов ВИЧ1 [86]

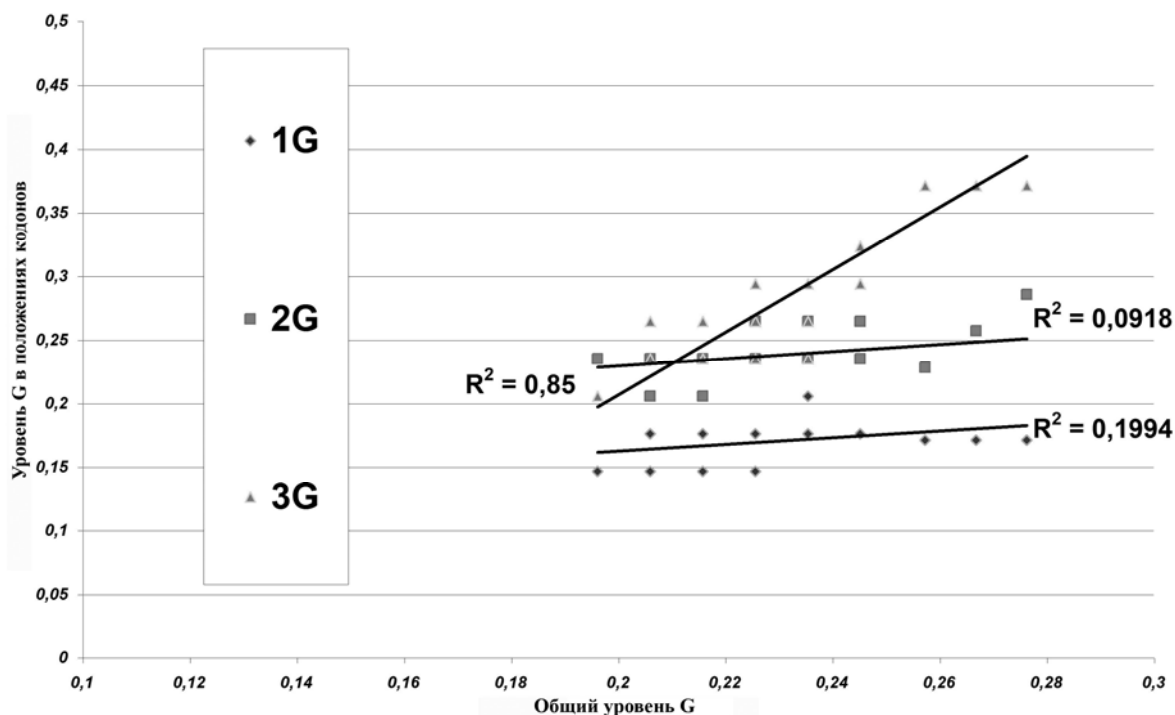


Рис. 16. Зависимость между частотами использования гуанина в трех положениях кодонов 100 нуклеотидных последовательностей участка, кодирующего V3 регион, от разных штаммов и клонов ВИЧ2 [86]

Низкий уровень 3G и высокий уровень 1G и 2G характерны для участков, кодирующих V3-регион у всех 100 исследованных разнородных штаммов ВИЧ1. Кроме того, как видно на рис. 15, линейная зависимость между 3G и общей частотой использования G для участков, кодирующих V3-регион ВИЧ1, отсутствует. У многих штаммов ВИЧ1 3G в данном участке вообще равен нулю. При этом уровень G находится в прямой зависимости средней силы от уровня 2G. Это подтверждает наше предположение о том, что вероятность несинонимичной замены G на A в участке, кодирующем V3-регион ВИЧ1, велика по причине высокого уровня использования гуанина во вторых положениях кодонов и низкой частоты его использования — в третьих. В результате уровень использования гуанина снижается не за счет синонимичных замен в третьих положениях кодонов, а за счет несинонимичных замен во вторых положениях [86].

Известно, что V3-регион выполняет важную функцию в жизненном цикле ВИЧ. Именно V3-регион gp120 связывается с клеточными ко-рецепторами [66], благодаря чему происходит проникновение вируса в клетку [66]. Логично предположить, что существуют селекционные ограничения аминокислотного состава V3-региона: большая часть аминокислотных замен в нем будет приводить к ухудшению или полной потере возможности взаимодействия с ко-рецепторами. Как же тогда объяснить наблюдаемые вариации в его аминокислотном составе?



Дело в том, что популяция вирусов в организме ВИЧ-инфицированного гетерогенна. Большинство вирионов действительно нежизнеспособны, в том числе и в результате неспособности к проникновению в клетку. Исходная нуклеотидная последовательность, кодирующая V3-регион, долгое время может сохраняться в неактивированных провирусах. Активация таких провирусов в поздний период ВИЧ-инфекции наряду с вирусными геномами, накопившими большое количество мутаций, может приводить к появлению «квазивидов» вирусов (нескольких наиболее часто встречающихся вариантов). Несмотря на то, что большинство мутантных V3-регионов не позволяют вирусу проникнуть в клетку, именно на них и может переключаться иммунный ответ, создавая картину «ухода от иммунного ответа», тем более что из-за определенного направления мутационного давления (А-давления) одинаковые варианты мутантных нежизнеспособных вирусов должны производиться постоянно и в больших количествах [86].

Следующим этапом нашего анализа явился расчет частот использования гуанина в участках, кодирующих V3-регион у быстрого и медленного прогрессора ВИЧ1. Согласно работе, в которой были получены эти нуклеотидные последовательности, «скорость эволюции» V3-региона быстрого прогрессора была гораздо выше «скорости эволюции» медленного прогрессора [112].

Как медленный, так и быстрый прогрессор были заражены внутриутробно их общей матерью. Между рождением первого ребенка (медленного прогрессора) и второго (быстрого прогрессора) прошло три года [112]. Как показал наш анализ, в популяции вирусов из организма матери за эти три года заметно истощился «защитный буфер» в виде гуанина в третьих положениях кодонов. Те вирусы, которые проникли в организм второго ребенка, имели более мутабельные участки, кодирующие V3-регион gp120, чем вирусы, проникшие в организм первого [86].

На рис. 17 видно, что уровень 3G в участках, кодирующих V3-регион медленного прогрессора, коррелирует с общим уровнем G, в то время как для 1G и 2G такая корреляция отсутствует. Это говорит о том, что 3G действительно «предохранял» участок, кодирующий V3-регион, от несинонимичных замен [86].

На рис. 18 все наоборот: есть корреляция между 1G и G, так же как и между 2G и G, но нет никакой зависимости между 3G и G. Объяснить это можно тем, что исходный уровень 3G у вирусов, поразивших второго ребенка (быстрого прогрессора) был ниже (рис. 18), чем исходный уровень 3G у вирусов, поразивших первого (рис. 17) [86].

Результаты наших исследований заставляют сделать предположение о том, что уровень 3G в участках, кодирующих V3-регион ВИЧ1, можно

использовать в качестве прогностического критерия [86]. Чем ниже 3G, тем выше вероятность того, что ВИЧ-инфекция будет прогрессировать быстро. Естественно, на данном этапе можно делать лишь предположения. Существует много других факторов, влияющих на скорость прогрессии ВИЧ-инфекции, как со стороны вируса, так и со стороны иммунной системы инфицированного. Тем не менее предложенный нами тест кажется удобным и обоснованным с теоретической стороны, если основной мишенью для гуморального и клеточного иммунного ответа у данного пациента является именно V3-петля [86].

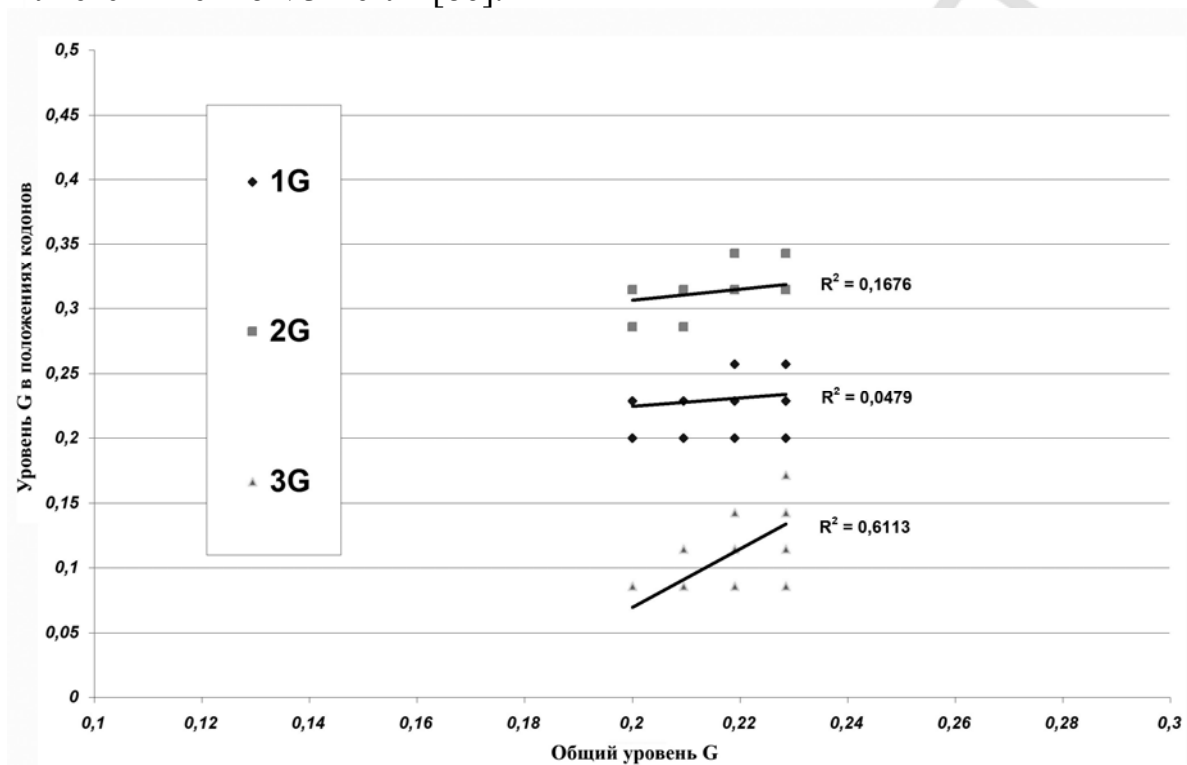


Рис. 17. Зависимость между частотами использования гуанина в трех положениях кодонов 42 нуклеотидных последовательностей участка, кодирующего V3-регион, от разных клонов ВИЧ1 медленного прогрессора [86]

**Расчет частот использования динуклеотидов и изучение картины их распределения по положениям кодонов.** Поскольку в генах *env* ВИЧ1 и ВИЧ2 можно предположить наличие более конкретных «горячих точек для мутаций», чем просто гуанин, а именно, трех динуклеотидов (GA, GG и GC), мы углубили наши исследования.

В табл. 13 можно видеть частоты использования всех динуклеотидов, а также процентное соотношение динуклеотидов, расположенных в первых и во вторых, во вторых и в третьих, в третьих и в первых положениях кодонов для участка, кодирующего V3-регион *gp120* из референтного штамма ВИЧ1. Поскольку мутации во всех трех «горячих» динуклеотидах подвергается первый нуклеотид, то динуклеотиды, расположенные

в третьих и в первых положениях кодонов, можно считать «безопасными»: мутация в них, как правило, будет синонимичной. Таких безопасных динуклеотидов в участке, кодирующем V3-петлю ВИЧ1, довольно мало — только 16 % [86].

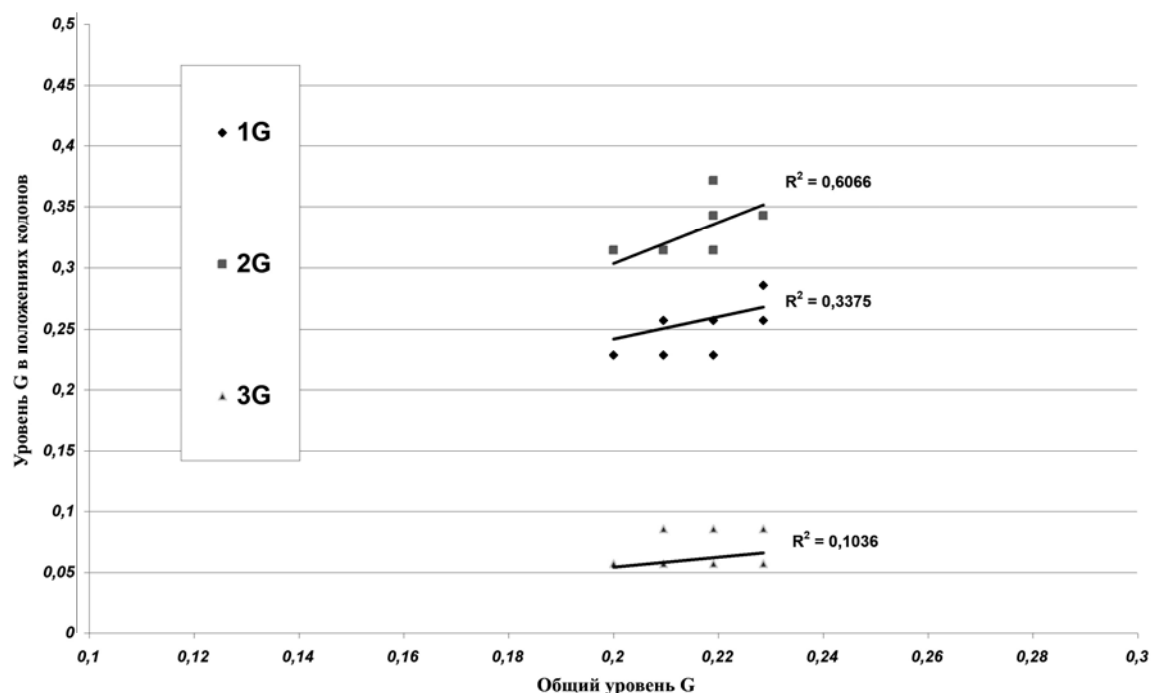


Рис. 18. Зависимость между частотами использования гуанина в трех положениях кодонов 44 нуклеотидных последовательностей участка, кодирующего V3-регион, от разных клонов ВИЧ1 быстрого прогрессора [86]

Таблица 13

**Частоты использования динуклеотидов и их распределение по положениям кодонов в участке, кодирующем V3-регион gp120 ВИЧ1 (VVK Dinucleotide) [86]**

Динуклеотид	Общая частота, %	Положение кодона, %		
		1–2-е	2–3-е	3–1-е
CC	4,67	40	20	40
CG	0,93	100	0	0
CU	0,00	—	—	—
CA	11,22	25	50	25
GC	1,87	100	0	0
GG	4,67	80	20	0
GU	3,74	25	75	0
GA	11,22	0	75	25
UC	1,87	0	100	0
UG	3,74	50	25	25
UU	3,74	25	50	25
UA	6,54	0	29	71
AC	8,41	33	22	45
AG	12,15	46	7	46
AU	8,41	56	33	11

AA	16,82	33	17	50
----	-------	----	----	----

В табл. 14 мы поместили частоты использования динуклеотидов в участке, кодирующем V3-регион ВИЧ2. Сразу хочется отметить, что содержание «безопасных» нуклеотидов в этом участке (38 %) гораздо выше, чем в гомологичной нуклеотидной последовательности ВИЧ1 [86].

Общая частота использования трех мутабельных динуклеотидов (GA, GG и GC) в участке, кодирующем V3-регион ВИЧ1, выше, чем в гомологичном участке ВИЧ2 (14,96 и 9,88 % соответственно) [86].

Невозможно не усомниться в том, насколько V3-петли gp120 ВИЧ1 и gp105 ВИЧ2 гомологичны друг другу. По результатам выравнивания gp120 ВИЧ1 и gp105 ВИЧ2 из референтных штаммов нами была получена приведенная на рис. 19 картина [86].

Таблица 14

**Частоты использования динуклеотидов и их распределение по положениям кодонов в участке, кодирующем V3-регион gp120 ВИЧ1 (VVK Dinucleotide) [86]**

Динуклеотид	Общая частота, %	Положение кодона, %		
		1-2-е	2-3-е	3-1-е
CC	3,96	100	0	0
CG	0,00	–	–	–
CU	3,96	25	50	25
CA	12,87	23	62	15
GC	4,95	20	20	60
GG	5,94	33	50	17
GU	5,94	33	17	50
GA	4,95	0	60	40
UC	4,95	40	20	40
UG	4,95	60	40	0
UU	6,93	43	43	14
UA	6,93	0	43	57
AC	7,92	38	24	38
AG	10,89	28	36	36
AU	5,94	50	17	33
AA	8,91	44	0	56



Рис. 19. Выровненные аминокислотные последовательности V3-регионов ВИЧ1 и ВИЧ2 [86]

Различия в частотах использования нуклеотидов, динуклеотидов и аминокислот в V3-петлях ВИЧ1 и ВИЧ2 можно объяснить тем, что в их эволюции существенную роль сыграли инсерции. Инсерция, обогащенная гуанином во вторых и в первых положениях кодонов, произошла в гене *env* общего предшественника ВИЧ1. Другая инсерция, обогащенная гуанином в третьих положениях кодонов и обедненная в первых и во вторых, произошла в гене *env* общего предшественника ВИЧ2 [86].

**Расчет уровня мутабельности в различных регионах гена.** Для того чтобы выяснить, что чаще вызывает повышение варибельности участков гена *env*: высокая мутабельность или низкая функциональная значимость — мы рассчитали частоты использования трех часто мутирующих динуклеотидов в канонических консервативных и варибельных регионах кодирующего участка *gp120* референтного штамма ВИЧ1.

Как видно из табл. 15, варибельность региона V1–V2 (первый и второй варибельные регионы разделены одним консервативным аминокислотным остатком, и потому обычно изучаются совместно) можно связать с повышенной мутабельностью: общая частота «опасных» динуклеотидов в нем выше, чем в соседних консервативных регионах [86].

Таблица 15

**Частоты использования динуклеотидов GA, GG и GC в различных регионах кодирующего участка *gp120* референтного штамма ВИЧ1 [86]**

Регионы кодирующего участка <i>gp120</i> ВИЧ1	Длина, нт.	Общая частота использования GA, GG и GC	Частота использования GA, GG и GC в первых и вторых и во вторых и третьих положениях кодонов	Частота использования GA, GG и GC в третьих и первых положениях кодонов	Процент динуклеотидов GA, GG и GC, способных мутировать несинонимично, %
C1	306	0,141	<b>0,092</b>	0,049	<b>65,25</b>
V1-V2	198	0,168	<b>0,112</b>	0,056	<b>66,67</b>
C2	297	0,122	<b>0,081</b>	0,047	<b>66,39</b>
V3	108	0,178	<b>0,150</b>	0,028	<b>84,27</b>
C3	159	0,165	<b>0,127</b>	0,038	<b>76,97</b>
V4	87	0,116	<b>0,105</b>	0,011	<b>90,52</b>
C4	138	0,146	<b>0,117</b>	0,029	<b>80,14</b>
V5	24	0,174	<b>0,131</b>	0,043	<b>75,29</b>
C5	132	0,237	<b>0,176</b>	0,061	<b>74,26</b>

Варибельность V3-региона, судя по табл. 15, тоже должна быть связана с его повышенной мутабельностью. Как общая частота использования «опасных» динуклеотидов, так и вероятность того, что мутация в них будет несинонимичной, выше в V3-регионе, чем в соседних C2- и C3-регионах [86].

Что касается региона V4, то его варибельность может быть усилена снижением частоты использования «безопасных» мутабельных динуклео-

тидов по сравнению с таковой для C3 и C4 [86]. Между тем общее содержание GA, GG и GC гораздо выше в C3 и C4, чем в V4. Как показали наши дальнейшие исследования (см. гл. 1 пятой части), вариабельность V4 в большей степени связана с нейтральностью происходящих в нем мутаций, чем с риском несинонимичной замены: в этом регионе часто происходят делеции и инсерции различной длины.

Пятый вариабельный регион кажется слишком коротким для заключения об уровне его мутабельности. В то же время, пятый консервативный регион отличается самым высоким содержанием трех динуклеотидов, являющихся «горячими точками для мутаций». Приходится ожидать от него высокого уровня мутабельности [86]. Тот факт, что данный регион относят к консервативным, может быть объяснен тем, что он несет большую функциональную нагрузку, а потому аминокислотные замены в нем фиксируются реже. С другой стороны, границы вариабельных и консервативных регионов gp120 были определены на ограниченном количестве материала. Возможно, эти границы не соответствуют реальному положению вещей.

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью описанного метода можно обнаружить наиболее и наименее мутабельные регионы в генах вирусов и бактерий. Такого рода анализ должен существенно облегчить поиск устойчивых эпитопов для создания более эффективных вакцин. В данной главе речь шла о «сканировании» генов на наличие устойчивых участков, в то время как в следующей главе будет описана методика изучения мутабельности предсказанных пространственных эпитопов gp120 ВИЧ1.

Идея об использовании уровня мутабельности наиболее иммуногенного эпитопа (V3-региона в случае с ВИЧ) в качестве прогностического критерия тяжести и длительности заболевания [86] может быть перенесена и на другие вирусные и бактериальные инфекции.

## Глава 2

### **Предсказание уровня мутабельности пространственных эпитопов поверхностного белка gp120 вируса иммунодефицита человека 1-го типа**

**Цель исследования:** определить уровень мутабельности предсказанных пространственных эпитопов поверхностного гликопротеина gp120 ВИЧ1 на основании анализа участков гена *env*, кодирующих их.

**Область применения:** данный тип молекулярно-биологического анализа может быть применен для отбора наиболее устойчивых эпитопов в вирусных или бактериальных белках в том случае, если 3D-структура этих белков уже определена.

**Используемые оригинальные алгоритмы:** «CodonChanges» (см. гл. 1 шестой части).

**Материалы.** Для расчета вероятности возникновения несинонимичных замен G на A и C на U мы использовали участки, кодирующие предсказанные 3D эпитопы, вырезанные из нуклеотидной последовательности гена *env* референтного штамма ВИЧ1. Границы эпитопов были предсказаны на основании данных о 3D-структуре gp120 из базы данных PDB.

**Номера доступа (идентификаторы) в GenBank.** Геном референтного штамма ВИЧ1 доступен в GenBank под следующим номером — NC\_001802. Записи о третичной структуре gp120 ВИЧ1 из базы данных PDB: 2NY0–2NY7, 2NZY и 2NZX («ядро» gp120 без V3-региона), 2B4C («ядро» gp120 с V3 регионом).

**Описание метода.**

**Предварительные операции.** Эпитоп — это участок белка, на который вырабатывается иммунный ответ [26]. Существуют В-клеточные эпитопы, на которые вырабатывается гуморальный иммунитет в виде специфических антител, и Т-клеточные эпитопы, на которые вырабатывается клеточный иммунитет, а именно, специфические рецепторы Т-лимфоцитов [26]. В данной главе речь идет о В-клеточных эпитопах. В иммунный ответ на ВИЧ, естественно, вовлечены как гуморальное, так и клеточное звено. Роль их в спонтанном контроле ВИЧ-инфекции дискутируется. Метод, описанный ниже, может быть легко перенесен с В-клеточных эпитопов на Т-клеточные.

В-клеточные эпитопы бывают двух видов: линейные (2D) и пространственные (3D) [26, 43, 90]. Линейный эпитоп представляет собой участок аминокислотной последовательности. Такая относительно короткая последовательность аминокислот становится мишенью для антител в том случае, если она находится на поверхности белковой глобулы [75, 76, 90]. Хочется подчеркнуть, что естественные антитела вырабатываются на пространственные эпитопы, то есть на участки, находящиеся на поверхности белка [90]. Пространственный эпитоп может состоять из аминокислотных остатков, располагающихся рядом друг с другом исключительно за счет формирования третичной структуры белка, но находящихся далеко друг от друга в первичной структуре [43]. Лишь меньшая часть пространственных эпитопов включает в себя участок аминокислотной последовательности, являющийся линейным эпитопом [90].

Известна шкала акрофильности — вероятности включения аминокислотного остатка в линейный В-клеточный эпитоп [75, 76]. Основана эта шкала на обобщении эмпирических результатов, но причины разницы в вероятности включения того или иного аминокислотного остатка в линейный эпитоп — чисто биохимические. На поверхности белка чаще всего располагаются пролин и глицин, аминокислотные остатки с небольшим

полярным радикалом (аспарагиновая и глутаминовая кислоты, аспарагин, глутамин, серин, треонин), так же как и аминокислотные остатки с длинным положительно заряженным радикалом (лизин и аргинин) [75, 76]. Внутри белковой глобулы чаще располагаются аминокислотные остатки с гидрофобными радикалами [75, 76] (табл. 16).

Таблица 16

**Шкала акрофильности аминокислотных остатков [87]**

Аминокислотные остатки, кодируемые GC-богатыми кодонами (GARP)		Аминокислотные остатки, кодируемые триплетами средней GC-насыщенности		Аминокислотные остатки, кодируемые GC-бедными кодонами (FYMINK)	
Gly	3,0	Asp	2,1	Asn	2,3
Pro	2,6	Ser	1,8	Lys	1,4
Arg	0,3	Glu	0,5	Met	-1,8
Ala	-0,5	Thr	-0,1	Tyr	-2,0
		Gln	-0,2	Ile	-2,5
		His	-0,4	Phe	-2,7
		Val	-0,7		
		Leu	-2,5		
		Cys	-2,6		
		Trp	-3,0		
«Средняя акрофильность»	1,35	«Средняя акрофильность»	-0,51	«Средняя акрофильность»	-0,88

Шкалу акрофильности/акрофобности можно переносить и на вероятность включения аминокислотных остатков в 3D-эпитопы. Итак, как видно из табл. 16, белки, кодируемые GC-богатыми генами, должны включать больше В-клеточных эпитопов, чем белки, кодируемые GC-бедными генами. Нами было доказано на примере гликопротеина В ВПГ1, что аминокислотные замены, вызванные GC-давлением, приводят к удлинению ранее существовавших или возникновению новых В-клеточных эпитопов с вероятностью 25 % [87], то есть каждая четвертая несинонимичная замена АТ на GC способна вызвать появление нового или удлинение ранее существовавшего линейного В-клеточного эпитопа [87].

А что же тогда происходит с линейными В-клеточными эпитопами при АТ-давлении? Многие аминокислотные замены, вызванные АТ-давлением, действительно приводят к смене акрофильных аминокислотных остатков на акрофобные, в частности замены пролина на лейцин, глицина на валин. Такие замены должны влиять (с определенной долей вероятности) на третичную структуру белка, вызывая исчезновение соответствующих участков белка с его поверхности. Получается, что пространственный эпитоп может попросту «скрыться» в глубине белковой глобулы. Антигены возбудителей, гены которых подвержены АТ-давлению, и в частности антигены ВИЧ1, могут уходить от иммунного ответа таким радикальным способом.



Из приведенных выше теоретических выкладок следует вывод о том, что наиболее сильные эпитопы gp120, содержащие много таких акрофильных аминокислотных остатков, как глицин и, особенно, пролин, являются наименее устойчивыми. А существует ли такой 3D-эпитоп, в котором отсутствуют пролин и глицин и содержание мишеней для АТ-давления минимально? Чтобы ответить на этот вопрос, необходимо предсказать расположение пространственных эпитопов на поверхности gp120.

На данный момент разработано довольно много компьютерных алгоритмов, позволяющих предсказать расположение пространственных эпитопов на поверхности белка, но все они работают с записями о его третичной структуре. В данной главе мы оперировали границами пространственных эпитопов, определенными с помощью программы DiscoTore ([www.cbs.dtu.dk/services/DiscoTore/](http://www.cbs.dtu.dk/services/DiscoTore/)) [43]. Эта программа предсказывает расположение 3D-эпитопов в белке, третичная структура которого была определена при помощи рентгеноструктурного анализа или методом ядерно-магнитного резонанса [43]. Данные о третичной структуре белков можно найти в базе данных PDB (<http://www.rcsb.org/pdb/home/home.do>). Программа DiscoTore работает исключительно с записями из базы данных PDB [43].

Дело в том, что количество записей о третичной структуре различных фрагментов gp120 в базе данных PDB велико. Мы выбрали из нее 11 наиболее длинных участков: 2NY0–2NY7, 2NZY и 2NZX представляют собой «ядро» gp120 без V3-региона, 2B4C — «ядро» gp120 с V3-регионом. Хотелось отметить, что существуют некоторые различия в 3D-структуре этих белков даже при полной идентичности их аминокислотных последовательностей. В связи с этим непостоянны и границы 3D-эпитопов. Аминокислотные последовательности 2NY0–2NY7, 2NZY и 2NZX идентичны друг другу, но третичная структура изменчива: дело в том, что рентгеноструктурный анализ производился для белков gp120 во взаимодействии с антителами или рецепторами CD4. Белок-белковые взаимодействия вызывают конформационные изменения. В результате таких изменений аминокислотные остатки, входящие в состав 3D-эпитопов, становятся доступными для антител в большей или меньшей степени.

Руководствуясь результатами предсказаний DiscoTore, мы определили границы четырех наиболее длинных и постоянных пространственных эпитопов gp120 и сравнили нуклеотидный состав участков, кодирующих их, с таковым для V3-региона. Для работы с DiscoTore требуется знать номер записи о третичной структуре белка в базе данных PDB. После введения этого номера в специальную ячейку, DiscoTore выдает результат — аминокислотную последовательность с отмеченными на ней остатками, которые с наибольшей вероятностью будут формировать пространственные эпитопы [43].

Ниже расположены рисунки с выровненными аминокислотными последовательностями пяти исследованных эпитопов.

Как видно на рис. 20, границы 3D-эпитопа 1 действительно варьируют в зависимости от конформации белка и от его аминокислотной последовательности. Например, новый участок включился в данный эпитоп у белка под номером 2В4С за счет замены акрофобного валина на более акрофильный глутамин. Вместе с тем и обратная замена (глутамин на валин) могла вызвать исчезновение близлежащих аминокислотных остатков с поверхности белка.

	3D Эпитоп 1
<b>Референтный штамм</b>	VVLVNVTFENFMWKNDMVEQMHEDIISLWDQ
2В4С	VVLQNVT-H-NMWKNDMVEQMQED--S---Q
2NY7	-----KND-VEQMHE-----
2NY6	-----KND--EQ--E-----
2NY5	-----KND--EQ-HE-----
2NY4	-----WKNDMVEQMHE- IS---Q
2NY3	-----WKNDMVEQMHE- IS---Q
2NY2	-----N-WKNDMVEQMHE--S---Q
2NY1	-----N-WKND-VEQ-HE-----
2NY0	-----KND--EQ-HE- IS---Q
2NXZ	-----N-WKNDMVEQMHE- IS---Q
2NXY	-----N-WKNDMVEQMHE- IS---Q

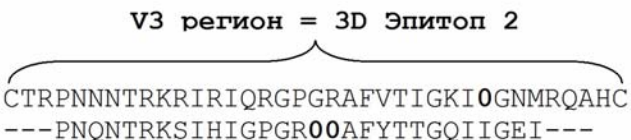
Рис. 20. Аминокислотные остатки, включенные в 3D-эпитоп 1 белка gp120, выровненные с аминокислотной последовательностью gp120 референтного штамма ВИЧ1. Знаком «-» обозначены аминокислотные остатки, не входящие в состав 3D-эпитопов

Пространственный эпитоп 1 образован такими акрофильными аминокислотными остатками, как аспарагин и лизин (кодируются GC-бедными кодонами), глутаминовая и аспарагиновая кислоты, глутамин (кодируются средними по GC-насыщенности кодонами). В эпитопе отсутствуют такие «горячие точки для мутаций», как глицин, пролин, аланин и аргинин.

Как уже было сказано, V3-регион не был вырезан из ядра белка gp120 только в записи под номером 2В4С. Как видно из рис. 21, V3-регион 2В4С значительно отличается от такового у референтного штамма. Пространственный эпитоп на рис. 21 богат глицином, аргинином и пролином.


Если внимательно изучить рис. 22 23, 24, можно заметить, что пространственные эпитопы 3, 4 и 5 содержат GC-насыщенный глицин, в эпитопах 3 и 5 присутствует аргинин, а в эпитопе 3 — еще и аланин. Судя по аминокислотному составу, наименее мутабельным среди пяти эпитопов gp120, должен быть первый из них (см. рис. 20).

**V3 регион = 3D Эпитоп 2**

<b>Референтный штамм</b> 2B4C	
	CTRPNNNTRKRIRIQRGPGRAFVTIGKIOGNMRQAHC ---PNQNTRKSIHIGPGR00AFYTTGQIIGEI---

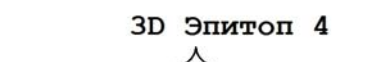
*Рис. 21.* Аминокислотные остатки, включенные в 3D-эпитоп 2 белка gp120, выровненные с аминокислотной последовательностью gp120 референтного штамма ВИЧ1. Знаком «-» обозначены аминокислотные остатки, не входящие в состав 3D-эпитопов; знаком «0» обозначены гэпы

**3D Эпитоп 3**

<b>Референтный штамм</b>	
2B4C	AKWNNTLKQIASKLREQFGNNKTIIFKQSSGGD
2NY7	A--ND--K-----REQFEN?KT-----SGGD
2NY6	---NN-----REQFGNNK-----SGGD
2NY5	---NN-----REQFGNNK-----SGGD
2NY4	---NN--K---S--REQFGNNK-----SGGD
2NY3	---NN-----REQFGNNK-----SGGD
2NY2	A--NN-----REQFGNNK-----SGGD
2NY1	---NN-----REQFGNNK-----SGGD
2NY0	A--NN--K-----REQFGNNK-----SGGD
2NXZ	---NN-----S--REQFGNNK-----SGGD
2NXY	A--NN-----S--REQFGNNK-----SGGD

*Рис. 22.* Аминокислотные остатки, включенные в 3D-эпитоп 3 белка gp120, выровненные с аминокислотной последовательностью gp120 референтного штамма ВИЧ1. Знаком «-» обозначены аминокислотные остатки, не входящие в состав 3D-эпитопов; знаком «?» обозначены аминокислотные остатки, не включенные в запись о третичной структуре из базы данных PDB

**3D Эпитоп 4**

<b>Референтный штамм</b>	
2B4C	STWFNSTWSTEGSNNTEGSD
2NY7	STWNN??NTEGSNNTEGN-
2NY6	S---NSTWS????????GSD
2NY5	---FN????????????GSD
2NY4	S-WFN????????????GSD
2NY3	---FNST????????????GSD
2NY2	---FNS????????????GSD
2NY1	---FN????????????GSD
2NY0	--WFN????????????GSD
2NXZ	---FNST????????????GSD
2NXY	---FN????????????GSD

*Рис. 23.* Аминокислотные остатки, включенные в 3D-эпитоп 4 белка gp120, выровненные с аминокислотной последовательностью gp120 референтного штамма ВИЧ1. Знаком «-» обозначены аминокислотные остатки, не входящие в состав 3D-эпитопов; знаком «?» обозначены аминокислотные остатки, не включенные в запись о третичной структуре из базы данных PDB

### 3D Эпитоп 5

Референтный штамм	DGGNSNNESEIFRPGGGDMRDNWR
2B4C	DGGINENGTE-----DMRD-WR
2NY7	DGGNSNNESEI-----GDMR---R
2NY6	DGGNSNNESEI-----R--W-
2NY5	DGGNSNNESEI-----D-RD-WR
2NY4	DGGNSNNESE-----DMRD--R
2NY3	DGGNSNNESE-----DMRD--R
2NY2	DGGNSNNESEI-----DMRD--R
2NY1	DGGNSNNESEI-----D-R---R
2NY0	DGGNSNNESEI-----DMR--W-
2NXZ	DGGNSNNESEI-----DMRD--R
2NXY	DGGNSNNESEI-----DMRD--R

Рис. 24. Аминокислотные остатки, включенные в 3D-эпитоп 5 белка gp120, выровненные с аминокислотной последовательностью gp120 референтного штамма ВИЧ1. Знаком «-» обозначены аминокислотные остатки, не входящие в состав 3D-эпитопов

**Производимые вычисления.** Алгоритм «CodonChanges» производит расчет частот использования нуклеотидов в последовательностях, введенных в соответствующую ячейку на листе «sequence1» (или «sequence2») [83]. Вероятность того, что замена нуклеотида произойдет в третьем положении кодона, рассчитывается путем деления уровня часто мутирующего нуклеотида в третьих положениях кодонов на сумму уровней его использования в первых, вторых и третьих положениях [81]. Например, вероятность замены G на A в третьем положении кодона равна частному от деления 3G на 1G+2G+3G. В табл. 17 такие вероятности для улучшения восприятия приведены в процентах. Возможности возникновения замены в третьем положении кодона были рассчитаны в каждом участке, кодирующем пространственный эпитоп, как для гуанина, так и для цитозина, поскольку гуанин и цитозин являются наиболее часто мутирующими нуклеотидами в генах ВИЧ [35] (см. гл. 3 второй части).

**Предсказание уровня мутабельности пространственных эпитопов.** Как видно из табл. 17, наиболее защищенным от процессов мутационного давления является первый 3D-эпитоп, расположенный в пределах первого консервативного участка (С1) gp120. По нашему мнению, использование этого эпитопа для создания вакцины смогло бы существенно повысить ее эффективность.

Примечательно, что в участке, кодирующем первый эпитоп, не только высока вероятность замены С на U в третьем положении кодона (50 %), но и общая частота использования С является наименьшей среди всех остальных участков, представленных в табл. 17. Вероятность того, что замена G

на А произойдет в третьем положении кодона, также имеет наибольшее значение в участке, кодирующем первый 3D-эпитоп.

Таблица 17

**Использование гуанина и цитозина в трех положениях кодонов в участках, кодирующих пять 3D-эпитопов gp120 ВИЧ1. В скобках указана область gp120, в которой находится эпитоп**

3D-эпитоп, его границы	G	1G	2G	3G	Вер-ть замены в 3 пол., %	C	1C	2C	3C	Вер-ть замены в 3 пол., %
66–86 (C1)	0,254	0,286	0,143	0,333	43,7	0,095	0,143	0	0,143	50,0
268–303 (V3)	0,213	0,194	0,333	0,111	17,4	0,167	0,167	0,194	0,139	27,8
308–340 (C3)	0,172	0,212	0,182	0,121	23,5	0,141	0,091	0,182	0,152	35,8
365–384 (V4)	0,267	0,250	0,400	0,150	18,8	0,117	0	0,250	0,100	28,6
429–452 (C4-V5-C5)	0,347	0,417	0,417	0,208	20,0	0,125	0,042	0,083	0,250	66,7

Можно ли переносить данные о частотах использования нуклеотидов в трех положениях кодонов участков, кодирующих 3D-эпитопы, из референтного штамма на все остальные штаммы ВИЧ1?

Подвержен ли первый эпитоп вариабельности, связанной с низкой функциональной значимостью данной части gp120?

Ответы на эти вопросы можно найти в гл. 1 пятой части.

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью метода, описанного в данной главе, можно производить отбор наименее мутабельных пространственных эпитопов в бактериальных и вирусных белках для дальнейшей работы по созданию вакцин.

Такой предварительный этап, как предсказание расположения 3D-эпитопов на поверхности белка с помощью специальных компьютерных программ, работающих с записями о третичной структуре, может быть заменен на непосредственное картирование эпитопов при помощи иммунологических методов и рентгеноструктурного анализа. Такая замена будет целесообразной в случае отсутствия сведений о третичной структуре изучаемого белка в электронных базах данных. Необходимо отметить, что объем таких общедоступных баз данных заметно увеличивается с течением времени.

### Глава 3

#### Расчет риска нонсенс-мутации в гене, кодирующем м-протеин *Streptococcus pyogenes*

**Цель исследования:** рассчитать частоту использования претерминальных кодонов в генах, кодирующих М-протеин различных штаммов *Streptococcus pyogenes*, и сравнить ее с частотой использования претерминальных кодонов в полных кодирующих геномах различных штаммов *Streptococcus pyogenes*.

**Область применения:** расчет частоты использования претерминальных кодонов с определением направления мутационного давления дает представление о риске возникновения нонсенс-мутации в гене, что может быть использовано в различных отраслях биотехнологии и иммунологии.

**Используемые оригинальные алгоритмы:** «VVK in group» и «Coding Genome Scanner» (см. гл. 3, 4 шестой части).

**Материалы.** Для расчета G+C, 1GC, 2GC и 3GC в полных геномах одиннадцати штаммов *Streptococcus pyogenes* и в генах, кодирующих их М-белки, мы использовали записи с частотами использования кодонов в каждом кодирующем участке этих штаммов из Codon Usage Database [104]. На основании тех же таблиц были рассчитаны частоты использования аминокислотных остатков в полных протеомах и в М-белках этих штаммов *Streptococcus pyogenes*. Алгоритм «Coding Genome Scanner» рассчитывает интересующие нас показатели по каждому гену и белку [84]; на основании последних легко рассчитать средние по геному (и протеому) значения, используя статистический аппарат MS Excel. Названия штаммов перечислены ниже: M1 GAS; MGAS315; str. Manfredo; MGAS5005; MGAS6180; MGAS9429; MGAS10270; MGAS10394; MGAS10750; MGAS8232 и MGAS2096.

Из GenBank были выбраны нуклеотидные последовательности генов, кодирующих различные изоформы тяжелых цепей миозина человека [74]. Миозины класса II: MYH1, MYH2, MYH3, MYH4, MYH6, MYH7, MYH8, MYH9, MYH10, MYH11, MYH13, MYH14, MYH15; миозины класса V: MYH12, MYO5C. Для расчета G+C, 1GC, 2GC и 3GC в 15 генах, кодирующих изоформы миозина человека, а также аминокислотного состава соответствующих белков, использовался алгоритм «VVK in group».

#### **Описание метода.**

**Предварительные операции.** Кодон считается претерминальным в том случае, если в результате одной нуклеотидной замены он может стать терминальным [2, 5, 6, 32, 45, 99] (см. прил. 3). Следствием такой нуклеотидной замены (нонсенс-мутации) является преждевременное завершение синтеза соответствующего белка [1, 23].

Ранее нами было доказано наличие обратной зависимости между частотой использования претерминальных кодонов (ПТК) и GC-насыщенностью бактериальных генов [45]. В этом исследовании мы обращались к таблицам со средними частотами использования кодонов в полных и неполных геномах 84 бактерий, в том числе и в геноме одного из штаммов *Streptococcus pyogenes* [84]. На рис. 25 показана полученная нами обратная зависимость между ПТК и G+C для бактериальных геномов и для активных центров аденилатциклаз I и III классов [29, 84]. Зависимость имеется во всех трех случаях, хотя в активных центрах аденилатциклаз I класса уровень ПТК несколько выше, чем в среднем по кодирующим геномам, а в активных центрах аденилатциклаз III класса — несколько ниже.

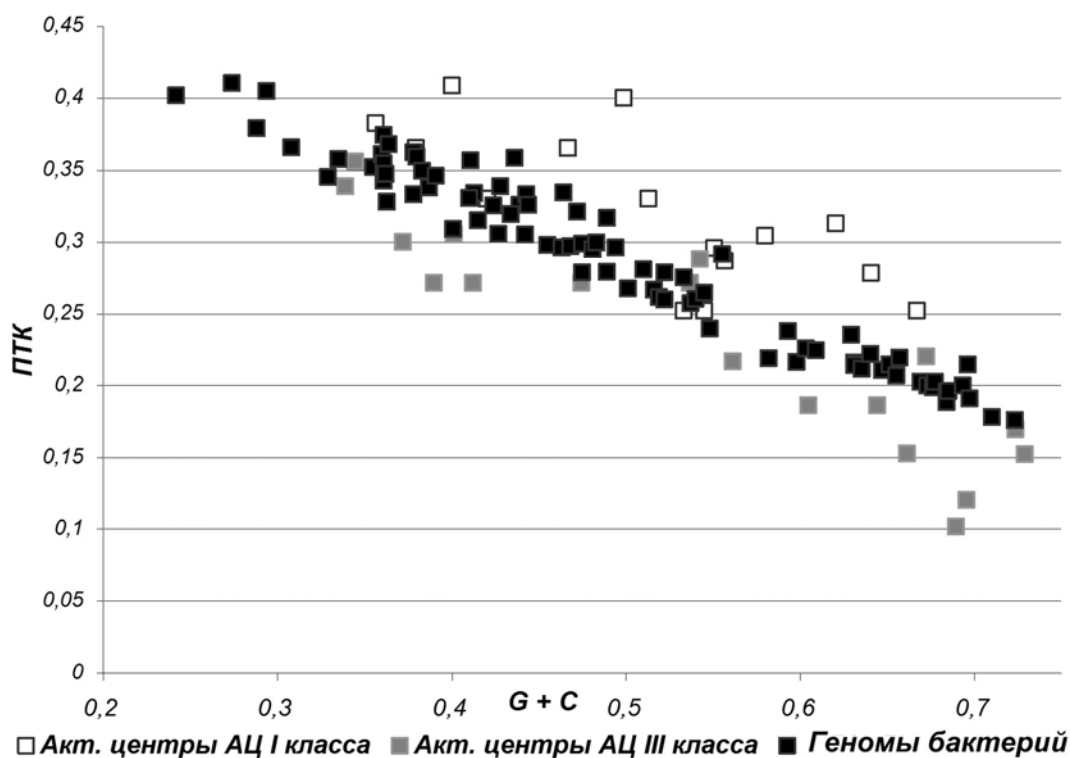


Рис. 25. Зависимость между GC-насыщенностью (G+C) и общей частотой использования претерминальных кодонов (ПТК) в геномах бактерий и в нуклеотидных последовательностях, кодирующих активные центры бактериальных аденилатциклаз [84]

Все три терминальных кодона (стоп-кодона) являются GC-бедными, как и большинство претерминальных кодонов (абсолютно GC-бедными, или GC-бедными относительно синонимичных им кодонов) (см. прил. 3) [84]. Именно из-за этого частота использования ПТК закономерно возрастает под воздействием АТ-давления [29, 84]. То же самое АТ-давление способно вызывать большинство нонсенс-мутаций, то есть замен претерминальных кодонов на терминальные [32, 84]. У GC-бедных бактерий такие замены периодически фиксируются, что приводит к относительному укорочению длины их кодирующих участков и к уменьшению размеров их

кодирующих геномов (то есть к уменьшению количества генов в геномах) [32, 84].

Если обобщить эти данные, можно заключить, что претерминальные кодоны становятся субстратом для нонсенс-мутаций, в основном, под воздействием АТ-давления. Чем больше претерминальных кодонов в гене, подверженном АТ-давлению, тем выше вероятность нонсенс-мутации в нем [84].

Инфекция, вызванная стрептококком группы А (*Streptococcus pyogenes* group A), опасна своим осложнением — острой ревматической лихорадкой (ОРЛ) [18, 25]. Острая ревматическая лихорадка возникает из-за выработки антител и Т-клеточных рецепторов к М-протеину стрептококка, которые перекрестно реагируют с миозином и тропомиозином человека [18, 25]. Согласно новейшим данным, причиной перекрестного реагирования является структурное сходство М-протеина, миозина и тропомиозина [98]. Как показали иммунологические исследования, титр антител против М-протеина, перекрестно реагирующих с миозином, одинаков у больных ОРЛ, хронической ревматической болезнью и у инфицированных стрептококком пациентов без развившихся аутоиммунных осложнений [54]. Значит ли это, что одного структурного сходства бактериальных и человеческих антигенов недостаточно для развития ОРЛ?

В результате анализа частот использования кодонов в гене, кодирующем М-протеин референтного штамма *Streptococcus pyogenes*, нами было обнаружено, что частота использования ПТК в нем повышена.

Для дальнейших исследований были выбраны все записи о полных геномах различных штаммов стрептококка из Codon Usage Database (их на данный момент оказалось одиннадцать) [104] и нуклеотидные последовательности участков, кодирующих различные изоформы миозина человека, из GenBank (пятнадцать генов).

По результатам вычисления эволюционной дистанции (по М. Кимуре) [3, 16] между нуклеотидными последовательностями, кодирующими М-протеины различных штаммов *Streptococcus pyogenes*, была построена NJ-дендрограмма [28, 105] с помощью программы MEGA 4 [124].

**Производимые вычисления.** Алгоритм «VVK in group» производит расчет всех интересующих нас показателей (частот использования кодонов, нуклеотидов, G+C, 1GC, 2GC, 3GC, GC4f, GC2f3p и ПТК) в последовательностях, введенных в соответствующие ячейки на листе «sequences». Алгоритм способен обработать до 50 последовательностей одновременно. Результаты вычислений, которые находятся на листе «results», могут служить для расчета средних значений показателей и для поиска зависимостей между ними.

О работе алгоритма «Coding Genome Scanner» подробно написано в гл. 2 второй части. Существенной деталью для данного исследования яв-



ляется то, что в нем были использованы данные об одном (или нескольких) отдельных кодирующих участках из полных геномов. Чтобы найти данные об интересующем вас кодирующем участке, необходимо осуществить поиск на листе «All CDSs». Поиск можно вести вручную или с помощью специальной опции MS Excel. Номер строки на листе «All CDSs» соответствует номеру на листе «results».

**Определение направления, силы и ведущего механизма мутационного давления в геноме *Streptococcus pyogenes*.** Прежде чем говорить о риске нонсенс-мутации, необходимо узнать направление, силу и ведущий механизм мутационного давления в исследуемом геноме. Для этого мы использовали методику, описанную в гл. 2 второй части. В кодирующем геноме *Streptococcus pyogenes* MGAS5005 общая GC-насыщенность (38,5 %) ниже 50 %, уровень 3GC (32,0 %) ниже уровня 1GC и 2GC (49,2 и 34,4 %). Это свидетельствует о том, что гены *Streptococcus pyogenes* подвержены мутационному АТ-давлению (частота нуклеотидных замен GC на АТ значительно превышает частоту замен АТ на GC).

Тем не менее уровень 3GC в кодирующих участках *Streptococcus pyogenes* MGAS5005 снизился еще не настолько, чтобы корреляция между 3GC и G+C была утрачена ( $R = 0,544$ ). Это свидетельствует о том, что некоторый «защитный буфер» в виде 3GC, предохраняющий гены *Streptococcus pyogenes* от несинонимичных замен в первых и вторых положениях кодонов, сохранился. Соответственно, мутационное АТ-давление в геноме *Streptococcus pyogenes* можно считать слабым.

При сравнении GC-насыщенности в четырехкратно (GC4f) и двукратно вырожденных сайтах (GC2f3p), находящихся в третьих положениях кодонов генов *Streptococcus pyogenes*, оказалось, что их уровни примерно одинаковы (30,0 и 30,1 % соответственно). Это свидетельствует о том, что АТ-давление в геноме *Streptococcus pyogenes* вызвано как транзициями, так и трансверсиями по направлению GC на АТ.

**Сравнение уровней использования ПТК в участках, кодирующих М-белок, и в полных геномах *Streptococcus pyogenes*.** Принимая во внимание направление и механизмы мутационного давления в геноме *Streptococcus pyogenes*, отдельно было рассчитано содержание претерминальных кодонов, способных стать терминальными в результате замены GC на АТ (ПТК GC на АТ) (см. прил. 3), в полных кодирующих геномах и в участках, кодирующих М-протеины (табл. 18). Действительно, «мишеней» для образования стоп-кодонов в участках, кодирующих М-протеин и М-подобные белки значительно больше, чем в остальных кодирующих участках *Streptococcus pyogenes*.

Интересно отметить, что статистически достоверная разница между G+C участков, кодирующих М-белки, и G+C полных кодирующих геномов разных штаммов *Streptococcus pyogenes* отсутствует. При этом уровень

ПТК и ПТК GC на АТ значительно (и статистически достоверно) выше в участках, кодирующих М-протеин и М-подобные белки. Найти причину этого повышения помог анализ аминокислотного состава М-белков. На одну и ту же диаграмму были помещены средние уровни использования аминокислотных остатков в М-протеинах и М-подобных белках, в полных протеомах *Streptococcus pyogenes* и в различных изоформах тяжелых цепей человеческого миозина (рис. 26).

Таблица 18

**Использование претерминальных кодонов (ПТК) и насыщенность гуанином и цитозином (G+C) генов, кодирующих М-протеин и М-подобные белки, а также полных кодирующих геномов различных штаммов *Streptococcus pyogenes***

Штамм	Полные геномы и протеомы				М-протеин и М-подобные белки			
	ПТК	ПТК GC на АТ	G + C	3GC	ПТК	ПТК GC на АТ	G + C	3GC
M1 GAS	0,337	0,171	0,386	0,317	0,462	0,239	0,393	0,245
MGAS315	0,339	0,171	0,386	0,320	0,460	0,237	0,376	0,208
MGAS5005	0,338	0,170	0,385	0,320	0,460	0,237	0,395	0,247
MGAS6180	0,341	0,170	0,383	0,318	0,434	0,234	0,400	0,231
MGAS6180	0,341	0,170	0,383	0,318	0,475	0,261	0,393	0,260
MGAS8232	0,339	0,171	0,385	0,320	0,476	0,229	0,369	0,222
MGAS9429	0,341	0,169	0,383	0,321	0,455	0,243	0,386	0,208
MGAS10270	0,341	0,170	0,383	0,321	0,464	0,236	0,379	0,209
MGAS10270	0,341	0,170	0,383	0,321	0,491	0,275	0,385	0,228
MGAS10270	0,341	0,170	0,383	0,321	0,437	0,225	0,399	0,258
MGAS10394	0,339	0,170	0,386	0,323	0,442	0,212	0,389	0,252
MGAS10750	0,341	0,170	0,383	0,320	0,461	0,242	0,391	0,215
MGAS10750	0,341	0,170	0,383	0,320	0,494	0,270	0,396	0,239
MGAS10750	0,341	0,170	0,383	0,320	0,437	0,228	0,404	0,262
Str. Manfredo	0,338	0,171	0,386	0,319	0,479	0,237	0,371	0,243

*Примечание.* Разница в ПТК между генами, кодирующими М-протеин, и полными кодирующими геномами статистически достоверна ( $P < 0,001$ ), в отличие от разницы в G+C ( $P > 0,1$ ).

Лейцин, глутаминовая кислота и лизин являются наиболее часто используемыми аминокислотными остатками в тяжелых цепях миозина, частоты использования лейцина и глутаминовой кислоты в них, как это ни странно, примерно равны таковым в М-протеинах и М-подобных белках. Глутаминовая кислота кодируется двумя триплетами (GAA и GAG), которые могут стать стоп-кодонами в результате замены G на T в первом положении кодона. Лизин тоже кодируется претерминальными кодонами (AAA и AAG), но они могут стать терминальными в результате замены A на T в первом положении.

Известно, что молекула тяжелой цепи миозина состоит из глобулярной головки и стержневидного хвоста [74]. Две тяжелые цепи миозина соединяются друг с другом своими С-концевыми хвостами, образуя скру-

ченную двойную альфа-спираль (coiled-coil, в русской биохимической литературе встречается следующая версия перевода: «двухцепочечная альфа-спиральная суперспираль») [74]. Такую же структуру имеет спираль, образующаяся при соединении двух молекул М-протеина *Streptococcus pyogenes* [98].

Скрученная спираль образуется двумя альфа-спиральными белками. Как известно, три аминокислотных остатка являются сильными формирователями альфа-спирали (а именно, глутаминовая кислота, аланин и лейцин) [75, 76]. В скрученных спиральных, как правило, присутствуют тандемные повторы, обогащенные такими аминокислотными остатками. По данным, приведенным на рис. 26, аминокислотный состав как М-протеинов *Streptococcus pyogenes*, так и человеческих миозинов класса II и V действительно обогащен альфа-спиральными аминокислотными остатками, особенно глутаминовой кислотой и лейцином.

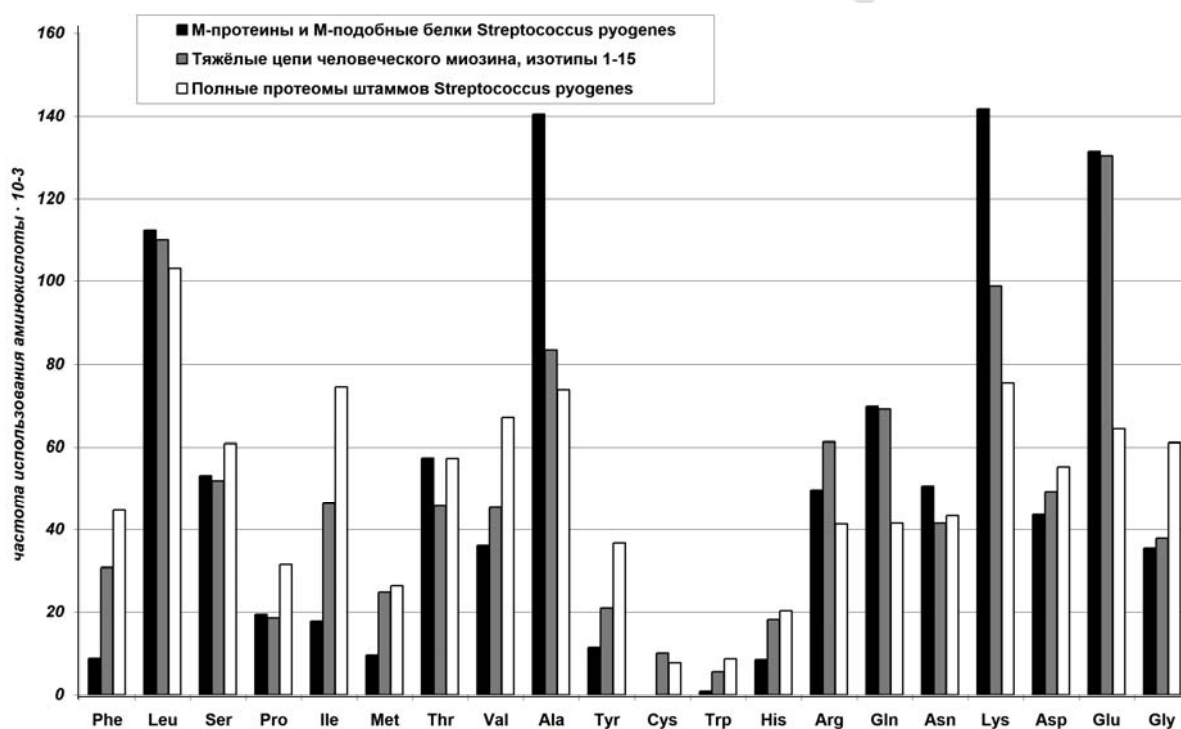


Рис. 26. Аминокислотный состав исследованных М-протеинов и М-подобных белков *Streptococcus pyogenes*, изоотипов человеческого миозина и полных протеомов *Streptococcus pyogenes*

Если высокая частота использования лейцина характерна для всех белков *Streptococcus pyogenes*, то уровни лизина, глутаминовой кислоты и аланина в М-протеинах и М-подобных белках примерно в два раза превышают значения, характерные для протеома *Streptococcus pyogenes* (рис. 26).

Частоты использования фенилаланина, изолейцина, метионина, валина, тирозина, глицина и пролина (последние два являются сильными раз-

рушителями альфа-спирали) в М-протеинах и М-подобных белках *Streptococcus pyogenes*, наоборот, значительно ниже, чем в среднем по протеому.

Тандемные повторы, обогащенные глутаминовой кислотой, не только способствуют образованию альфа-спирали и скрученной двойной спирали, но и обогащают кодирующие участки претерминальными кодонами (ПТК). Из-за высокой частоты использования ПТК при мутационном АТ-давлении резко повышается вероятность нонсенс-мутации (образования терминального кодона) в генах, кодирующих М-протеины и М-подобные протеины *Streptococcus pyogenes*.

По нашей гипотезе, высокая вероятность утраты М-протеина путем закрытия рамки считывания в кодирующем его участке ДНК может приводить к быстрому снижению антигенной нагрузки в течение заболевания. Стрептококки, утратившие ген М-протеина, должны уходить от иммунного ответа и активно размножаться в организме больного. Снижение количества бактериальных мишеней (М-протеинов) должно способствовать переключению иммунного ответа на схожие по структуре белки собственного организма (миозин и тропомиозин).

В пользу нашей гипотезы о том, что высокая частота использования ПТК при АТ-давлении учащает возникновение нонсенс-мутаций, свидетельствуют данные филогенетического исследования изученных генов, кодирующих М-протеин и М-подобные белки. На рис. 27 изображена дендрограмма, отражающая степень родства этих генов из 10 штаммов *Streptococcus pyogenes*. Сразу хочется отметить, что в геноме штамма MGAS2096 отсутствует ген, кодирующий М-протеин. Как показали недавние экспериментальные работы, утрата гена, кодирующего М-белок, не является летальной мутацией для стрептококка, несмотря на то, что продукт этого гена проявляет антифагоцитарную активность [52]. У штамма MGAS6180 есть два гена, кодирующих М-протеин и М-подобный белок. У штаммов MGAS10270 и MGAS10750 таких генов три [64].

На дендрограмме (рис. 27) четко видно, что гены, кодирующие М-подобные белки, присутствующие в геноме отдельных штаммов наряду с геном, кодирующим собственно М-протеин, не являются продуктами недавней дубликации. Все три варианта гена М-протеина возникли у общего предшественника проанализированных штаммов в относительно далеком прошлом. Только у двух из 10 штаммов сохранились все три копии гена М-протеина. У штамма MGAS6180 сохранились две копии, причем, судя по дендрограмме, эти копии кодируют не М-протеин, а М-подобные белки.

У большинства (у 7 из 11) штаммов сохранилась только одна копия гена М-протеина (кодирующая собственно М-протеин). Видимо, высокая частота использования ПТК сыграла свою роль в исчезновении открытых рамок считывания в копиях кодирующего участка М-протеина. Еще раз хочется напомнить, что у штамма MGAS2096 исчезли все три копии белка,

антитела к которому перекрестно реагируют с человеческим миозином. Инфекция, вызванная таким штаммом, должна протекать без аутоиммунных осложнений. По нашей гипотезе, наиболее ревматогенными штаммами должны быть те, в геномах которых присутствует только одна копия кодирующего участка М-протеина или М-подобного белка. Действительно, *Streptococcus pyogenes* серотипа MGAS8232, в геноме которого есть только одна копия гена М-протеина, обладает повышенной ревматогенностью [117].

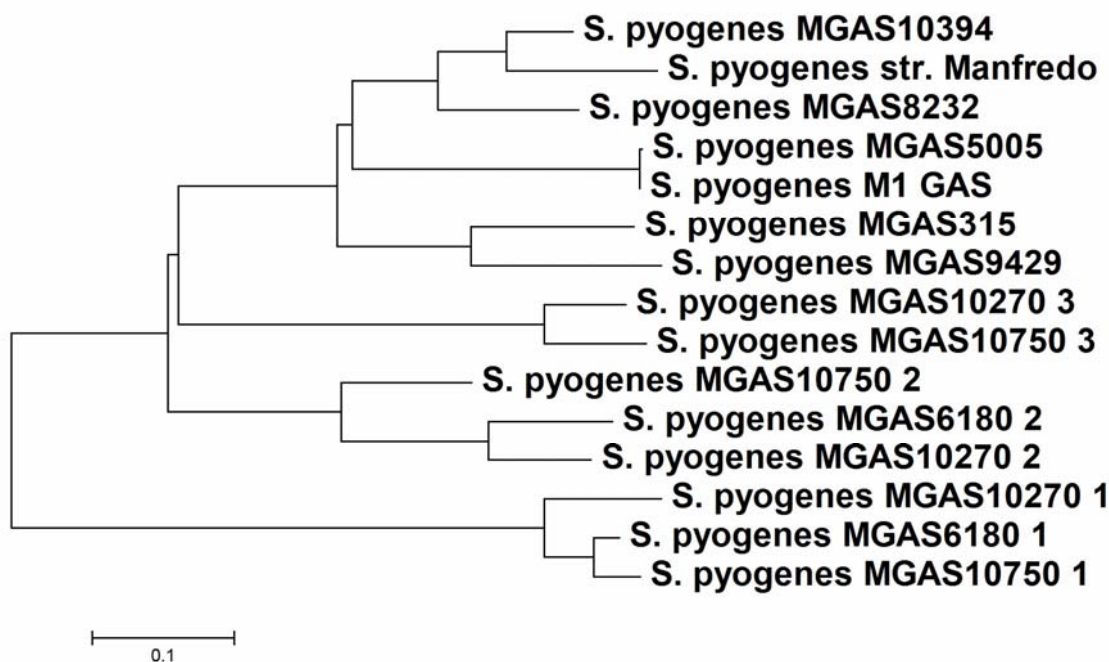


Рис. 27. Дендрограмма, отражающая филогенетические взаимоотношения М-протеинов и М-подобных белков исследованных штаммов *Streptococcus pyogenes*

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью метода, описанного в данной главе, можно выяснить, насколько интересующий вас ген или геном подвержен нонсенс-мутациям, которые происходят в результате единичных нуклеотидных замен. Знание такого рода может быть использовано в различных отраслях биотехнологии.

Изложенная в этой главе гипотеза о том, что развитию острой ревматической лихорадки, вызываемой перекрестно реагирующими антителами, способствует резкое снижение титра бактериального антигена (М-белка), может быть перенесена и на другие аутоиммунные болезни. Высокий риск нонсенс-мутации в кодирующем участке является одним из наиболее значимых факторов, способствующих прекращению выработки того или иного бактериального или вирусного антигена.

## Часть четвертая

### Степень селекционных ограничений, накладываемых на аминокислотный состав белка

#### Глава 1

#### Анализ механизмов упрощения аминокислотного состава препроинсулина хордовых животных под воздействием GC-давления

**Цель исследования:** определить степень и характер селекционных ограничений, накладываемых отрицательным отбором на аминокислотный состав различных частей препроинсулина: инсулин, С-пептид и сигнальный пептид.

**Область применения:** данный тип молекулярно-биологического анализа может быть применен в случае необходимости определить степень и характер селекционных ограничений, накладываемых на определенный белок отрицательным отбором в условиях мутационного давления, что может быть использовано в молекулярной филогенетике и в различных областях биотехнологии.

**Используемые оригинальные алгоритмы:** «VVK in group» (см. гл. 3 шестой части).

**Материалы.** Для описания метода изучения того, как часто и какие именно аминокислотные замены, вызванные мутационным давлением, фиксируются в белках, мы выбрали наиболее яркий пример. Молекула препроинсулина состоит из нескольких частей. Первый (N-концевой) участок препроинсулина представляет собой сигнальный пептид. Функция последнего заключается в том, чтобы вновь синтезированная молекула препроинсулина попала в эндоплазматический ретикулум [23]. Для выполнения этого сигнальный пептид должен содержать гидрофобные аминокислотные остатки, такие как лейцин (рис. 28). Необходимо упомянуть о том, что все белки, которые транспортируются через эндоплазматический ретикулум, имеют короткие N-концевые участки (около 20 аминокислотных остатков) с преимущественно гидрофобными аминокислотными остатками [23].

После того как препроинсулин попадает в эндоплазматический ретикулум, сигнальный пептид отщепляется протеолитическими ферментами. Молекула препроинсулина, лишенная сигнального пептида, называется «проинсулин» [23]. Проинсулин транспортируется в аппарат Гольджи, а затем — в секреторные везикулы. В везикулах происходит следующий этап посттрансляционной модификации: из проинсулина вырезается средняя часть, которая называется С-пептид [23]. Такое название является сокращением от английского «connecting peptide», то есть «соединяющий пептид». Действительно, С-пептид, в каком-то смысле соединяет В-

и А-цепь инсулина (именно в таком порядке они расположены в молекуле проинсулина). Считается, что С-пептид способствует правильной пространственной ориентации А- и В-цепи друг относительно друга, необходимой для образования дипептидных связей между ними [23]. Кроме С-пептида и инсулина от проинсулина отщепляются еще два дипептида. Молекула С-пептида секретируется бета-клетками островков Лангерганса поджелудочной железы вместе с инсулином [23].

В работе, описанной в данной главе, были использованы гены пре-проинсулина из геномов следующих животных (в скобках указаны их идентификаторы в GenBank): *Homo sapiens* (NM\_000207); *Pan troglodytes* (NM\_001008996); *Gorilla gorilla* (AH011815); *Pongo pygmaeus* (AH011816); *Cercopithecus aethiops* (X61092); *Aotus trivirgatus* (J02989); *Macaca fascicularis* (J00336); *Bos taurus* (NM\_173926); *Felis catus* (NM\_001009272); *Canis lupus familiaris* (NM\_001130093); *Oryctolagus cuniculus* (U03610); *Mus musculus* (NM\_008386 и X04724); *Rattus norvegicus* (NM\_019129 и V01243); *Gallus gallus* (NM\_205222); *Xenopus laevis* (M24443 и M24442); *Danio rerio* (NM\_131056); *Lophius americanus* (V00634); *Cyprinus carpio* (X00989); *Myxine glutinosa* (V00649); *Sus scrofa* (AF064555); *Ovis aries* (U00659); *Mesocricetus auratus* (M26328); *Spermophilus tridecemlineatus* (AY038604); *Cricetulus longicaudatus* (M26328); *Psammomys obesus* (X98241); *Octodon degus* (M57671); *Cavia porcellus* (K02233); *Selasfolus rufus* (S66612); *Oreochromis niloticus* (AF038123); *Verasper moseri* (AB029318).

В геномах *Mus musculus*, *Rattus norvegicus* и *Xenopus laevis* (мыши, крысы и лягушки) содержится по две копии гена препроинсулина. Мы использовали в наших исследованиях обе копии. Отдельно были изучены (i)-участки, кодирующие сигнальный пептид, (ii)-участки, кодирующие С-пептид и (iii)-участки, кодирующие собственно инсулин (то есть мы объединили участки, кодирующие А-цепь инсулина с участками, кодирующими В-цепь инсулина). Кроме того, к выборке участков, кодирующих С-пептид, мы добавили соответствующие последовательности из геномов *Oncorhynchus keta* (J00936) и *Callorhynchus Milii* (U82395); к выборке участков, кодирующих сигнальный пептид, — последовательности из геномов *Anas platyrhynchos* (AY206835); *Anas crecca* (AY206805); *Tadorna tadorna* (AY206858) и *Cairina moschata* (AY206809); к выборке участков, кодирующих инсулин, — последовательность из генома *Oncorhynchus keta*. На момент проведения исследования в GenBank отсутствовали полные кодирующие участки гена препроинсулина для отдельных видов животных, поэтому мы использовали те части гена, которые были доступны.

NJ-дендрограммы [28] были построены с помощью программы MEGA 4 [124] на основании вычисления р-дистанции [3] при попарном сравнении аминокислотных последовательностей инсулина, С-пептида и сигнального пептида, выровненных методом identity (более сложные методы не могли



быть применены для изучения филогенетических связей между большим количеством относительно коротких, но значительно отличающихся друг от друга последовательностей) [105].

Аминокислотные последовательности препроинсулина гистрикоморфных грызунов (*Cavia porcellus* и *Octodon degus* — морской свинки и дегу соответственно) не были исследованы на предмет консервативности аминокислотных замен в них, поскольку еще М. Кимура констатировал тот факт, что инсулин морской свинки утратил свою функциональную активность и эволюционирует в режиме полной нейтральности [16], но эти последовательности были тщательно изучены при помощи анализа построенных дендрограмм.

#### Описание метода.

**Предварительные операции.** Перед тем как использовать алгоритм «VVK in group», мы разрезали нуклеотидные последовательности генов препроинсулина на части, согласно границам, обозначенным в записях из GenBank. На рис. 28 приведена аминокислотная последовательность препроинсулина человека. Мы исследовали три части (за исключением дипептидов), на которые разрезается препроинсулин: сигнальный пептид, С-пептид и собственно инсулин (В-цепь плюс А-цепь).

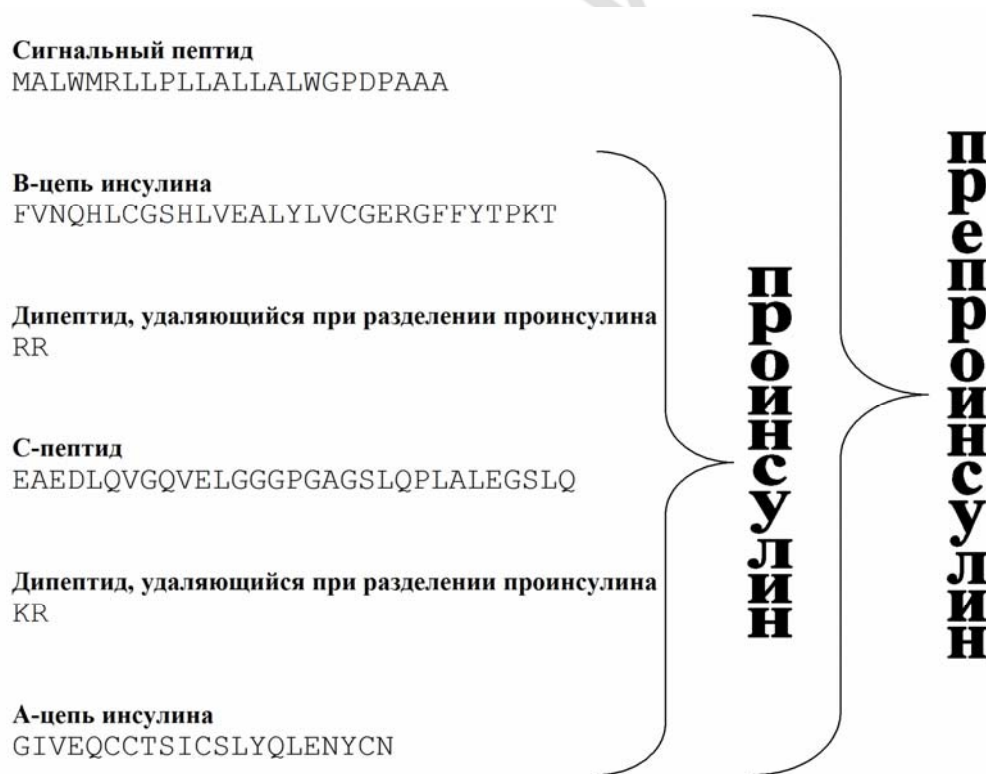


Рис. 28. Структура препроинсулина (на примере аминокислотной последовательности препроинсулина человека)

Алгоритм «VVK in group» способен обработать до 50 нуклеотидных последовательностей одновременно. Предварительного выравнивания по-



следовательностей не требуется, поскольку «VVK in group» рассчитывает все показатели «внутри» каждой из введенных последовательностей, а не «между» ними.

**Производимые вычисления.** Для того чтобы алгоритм «VVK in group» произвел все необходимые вычисления, необходимо скопировать набор нуклеотидных последовательностей в соответствующие ячейки на листе «sequences». В столбцах на листе «results» появятся все вычисляемые показатели. Такие показатели, как G+C, 1GC, 2GC и 3GC (общая GC-насыщенность и GC-насыщенность в каждом из трех положений кодонов) должны быть знакомы читателям данной главы. Подробно об их биологической значимости написано в гл. 1 второй части. Кроме частот использования нуклеотидов в гене и частот использования аминокислот в кодируемом белке, алгоритм «VVK in group» рассчитывает энтропию распределения аминокислотных остатков в белке [83, 135].

Согласно теории информации Клода Шеннона, максимальное значение энтропии (H) в системе достигается в том случае, если частоты всех ее компонентов равны друг другу. Поскольку в формуле логарифм имеет основание 2, количество информации в белке измеряется в битах. Максимальным значением энтропии для аминокислотного состава белка (как и в любой системе из 20 компонентов) является 4,322 бит [83].

Расчет энтропии для каждого кодирующего участка производился по приведенной ниже формуле:

$$H = -\sum f_{aa} \cdot \log_2 f_{aa}, \quad (1.1)$$

где  $f_{aa}$  — частота использования каждой из 20 аминокислот.

По формуле (1.1) получается, что общее количество информации равно сумме (с обратным знаком) значений выражения « $f_{aa} \cdot \log_2 f_{aa}$ » для каждого вида аминокислотных остатков. Если какой-то вид аминокислотного остатка вообще не встречается в белке, значение выражения « $f_{aa} \cdot \log_2 f_{aa}$ » признается равным нулю. Отсюда следует вывод, что применение этой формулы для филогенетического ряда коротких кодирующих участков является биологически оправданным, только если длина кодирующих участков примерно одинакова у всех видов. Длина инсулина, С-пептида и сигнального пептида среди хордовых животных не подвергается значительным вариациям.

**Определение направления мутационного давления в генах из филогенетического ряда.** Метод определения направления мутационного давления в генах из филогенетического ряда (гомологичных генов, принадлежащих разным видам организмов) идентичен таковому, описанному в гл. 1 второй части. Одна часть генов из филогенетического ряда могут быть подвержены GC-давлению, другая — AT-давлению.

Нами были описаны случаи, когда GC-насыщенность существенно варьировала и внутри одного гена — в гене US9 варицелловируса лошади

4-го типа [84]. Такая ситуация возникла, когда одна часть гена расположилась в LTR (большом инвертированном повторе), а вторая — в US (в уникальной короткой области генома). Частота возникновения трансверсий АТ на GC существенно выше в LTR, чем в US. Соответственно, 3GC в части US9 из LTR равняется 86,3 %, а в части US9 из US — 68,6 % [84].

Геномы эукариотических организмов содержат участки, существенно различающиеся друг с другом по GC-насыщенности. Такие участки называются изохорами [50]. Гены, находящиеся в GC-богатых изохорах, должны быть подвержены GC-давлению, которое, собственно, и должно приводить к появлению GC-богатых изохор [68]. Гены, находящиеся в АТ-богатых изохорах, должны быть подвержены АТ-давлению [68]. После транслокации гена из GC-богатой изохоры в GC-бедную GC-насыщенность в большей степени снижается в интронах, чем в экзонах [36, 38].

С помощью «VVK in group» можно легко получить график, объединяющий значения 1GC, 2GC и 3GC всех исследуемых гомологичных кодирующих участков. По характеру зависимости 1GC, 2GC и 3GC от G+C в генах из филогенетического ряда можно сделать первые предварительные выводы о характере селекционных ограничений, накладываемых на аминокислотную последовательность кодируемого ими белка [41].

Как видно на рис. 29, значения 3GC во всех исследованных участках, кодирующих инсулин, превышают 60 %. Существует сильная прямая зависимость между 3GC и G+C, у некоторых животных уровень 3GC превышает 90 %. Эти факты заставляют признать, что в участках, кодирующих инсулин, существует мутационное GC-давление. Тем не менее интенсивность биохимических процессов, вызывающих нуклеотидные замены АТ на GC в кодирующем участке инсулина, различается у разных видов животных. Нельзя не заметить полного отсутствия зависимости между 2GC и G+C (рис. 29). Известно, что все замены во вторых положениях кодонов являются несинонимичными. Они должны происходить довольно часто в участках, кодирующих инсулин. Однако, как показал наш анализ, такие замены практически не фиксируются. Зависимость между 1GC и G+C существует, но угол наклона линии тренда этой зависимости значительно меньше, чем угол наклона зависимости 3GC от G+C. Только незначительная часть замен в первых положениях кодонов синонимична. По данным, изображенным на рис. 29, можно сделать вывод о том, что аминокислотная последовательность инсулина «устойчива» к GC-давлению. Причина такой устойчивости — отрицательный отбор.

Особи с несинонимичными заменами в участке, кодирующем инсулин, в подавляющем большинстве случаев оказываются нежизнеспособными, то есть подвергаются элиминации из популяции путем отрицательного отбора. В данном случае отрицательный отбор является «стабилизирующим».

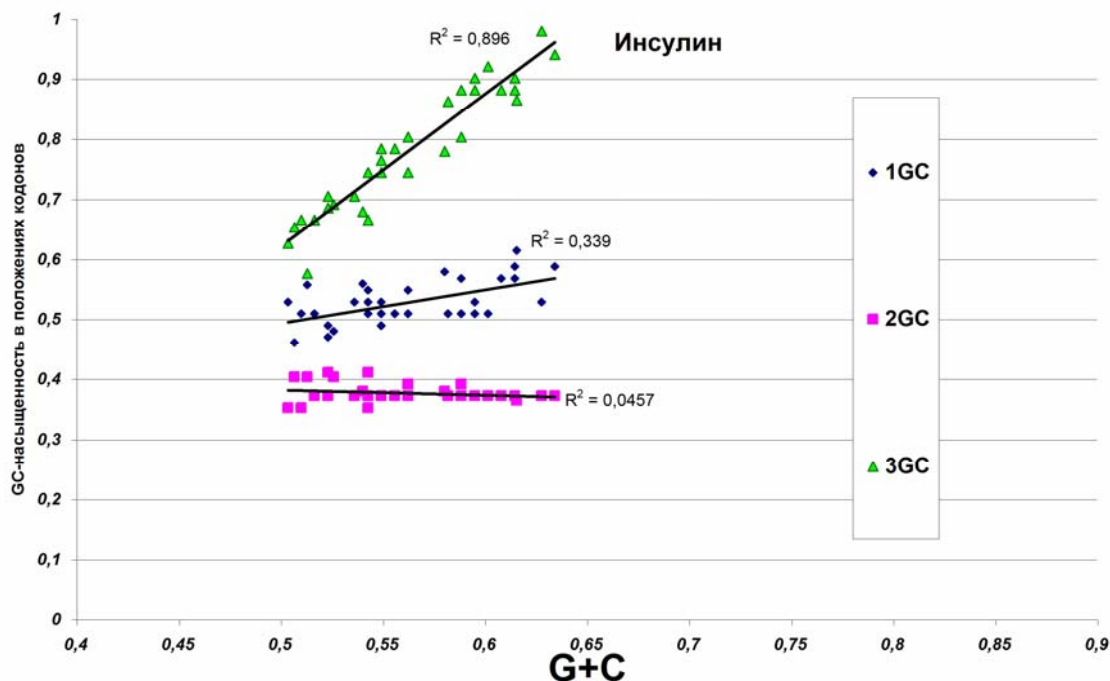


Рис. 29. Зависимость 1GC, 2GC и 3GC от G+C в участках, кодирующих инсулин хордовых животных

На рис. 30 изображен график зависимости 1GC, 2GC и 3GC от G+C в участках, кодирующих С-пептид хордовых животных. Различия рис. 29 и 30 бросаются в глаза. GC-насыщенность в каждом из трех положений кодонов участков, кодирующих С-пептид, находится в прямой линейной зависимости от общей GC-насыщенности. В связи с этим диапазон колебаний G+C в участках, кодирующих С-пептид, примерно в два раза шире, чем в участках, кодирующих инсулин. Уровень 1GC в целом выше в участках, кодирующих С-пептид, чем в участках, кодирующих инсулин.

Наличие прямой зависимости между 2GC и G+C в участках, кодирующих С-пептид, указывает на то, что функциональные ограничения «позволяют» определенным аминокислотным заменам, вызванным мутационным давлением, фиксироваться. Некоторые замены аминокислот оказываются нейтральными. Отрицательный отбор не элиминирует организмы с такими аминокислотными заменами в С-пептиде из популяции. Фиксация нейтральных замен в популяции происходит за счет случайного генетического дрейфа. К тому же, большую роль в повышении частоты встречаемости мутантных особей в популяции должно играть мутационное давление одного направления (повышается вероятность того, что у разных особей возникнут одинаковые мутации).

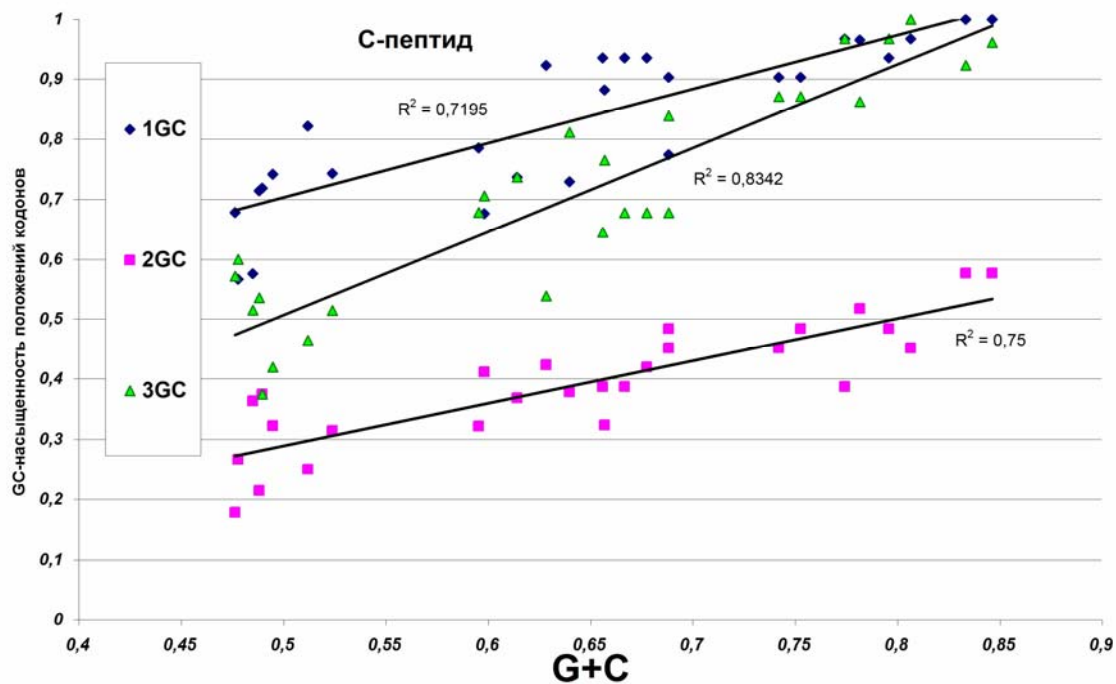


Рис. 30. Зависимость 1GC, 2GC и 3GC от G+C в участках, кодирующих С-пептид хордовых животных

Графики зависимости 1GC, 2GC и 3GC от G+C в участках, кодирующих сигнальный пептид хордовых животных, являются двухфазными (рис. 31).

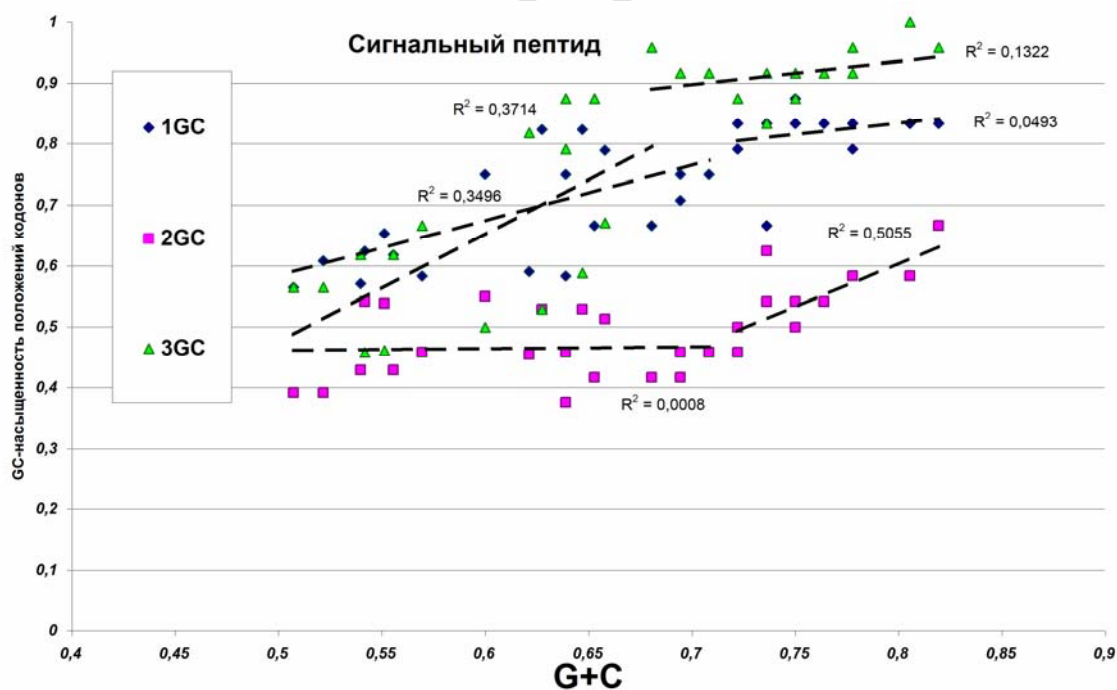


Рис. 31. Зависимость 1GC, 2GC и 3GC от G+C в участках, кодирующих сигнальный пептид хордовых животных

Интересно, что в генах с  $G+C < 0,71$ , GC-давление приводит к росту 1GC, а в генах с  $G+C > 0,71$  — к росту 2GC. GC-насыщенность вторых положений кодонов (2GC) стала повышаться только после того, как уровни 3GC и 1GC достигли своего предела, образовав «крышу» на рис. 31. Причем сначала своего предела достигла GC-насыщенность третьих положений кодонов, а потом — первых. Данный график напоминает комбинацию графиков из гл. 2 второй части. В целом, можно сказать, что аминокислотный состав сигнального пептида подвержен влиянию GC-давления. Это значит, что многие аминокислотные замены в сигнальном пептиде, вызываемые GC-давлением, являются нейтральными.

Если обобщить полученные данные, можно сказать, что селекционные ограничения на аминокислотный состав инсулина являются сильными, в отличие от селекционных ограничений состава сигнального пептида и С-пептида.

**Определение показателя нейтральности эволюции белка по методу Н. Суеоки.** Создатель теории мутационного давления, Н. Суеока, предложил метод, позволяющий определить степень нейтральности эволюции белка количественно [121]. Для получения показателя нейтральности эволюции белка по Н. Суеоке необходимо выполнить следующие действия. Первым этапом является построение графика зависимости среднего значения показателей 1GC и 2GC « $(1GC + 2GC) / 2$ » от уровня 3 GC в филогенетическом ряду. Вторым этапом — получение тренда линейной зависимости между этими показателями. Третьим этапом — продление этой линии до пересечения с осью Y и с перпендикуляром к точке 1,0 на оси X. Четвертым этапом — построение параллельной ей линии, проходящей через начало координат. Пятым этапом — измерение длины отсекаемого этой линией отрезка на перпендикуляре к точке 1,0 оси X. Финальным этапом — вычисление отношения длины этого отрезка к длине оси Y (от 0 до 1,0). В результате получается величина, отражающая, по мнению Н. Суеоки, степени нейтральности эволюции белка [121]. Чем выше эта величина, тем круче наклон « $(1GC + 2GC) / 2$ » к 3GC, тем больше несинонимичных замен, вызванных мутационным давлением, фиксируется по отношению к синонимичным заменам [121].

На рис. 32 видно, что наклон « $(1GC + 2GC) / 2$ » к 3GC минимален в участках, кодирующих инсулин. Нейтральность по Суеоке равна 3 %. Этот показатель, вычисляемый с привлечением геометрических расчетов, примерно равен множителю в формуле, отражающей характер линейной зависимости. Формулу линейной зависимости можно легко получить с использованием MS Excel. Значение  $R^2$  ( $R^2$  — это квадрат коэффициента корреляции, сам R в данном случае равен 0,17) на рис. 32 является минимальным, что указывает на полное отсутствие всякой зависимости между « $(1GC + 2GC) / 2$ » и 3GC в участках, кодирующих инсулин. Можно ли

доверять углу наклона тренда недостоверной зависимости? Автор метода ответа на этот вопрос не дает.

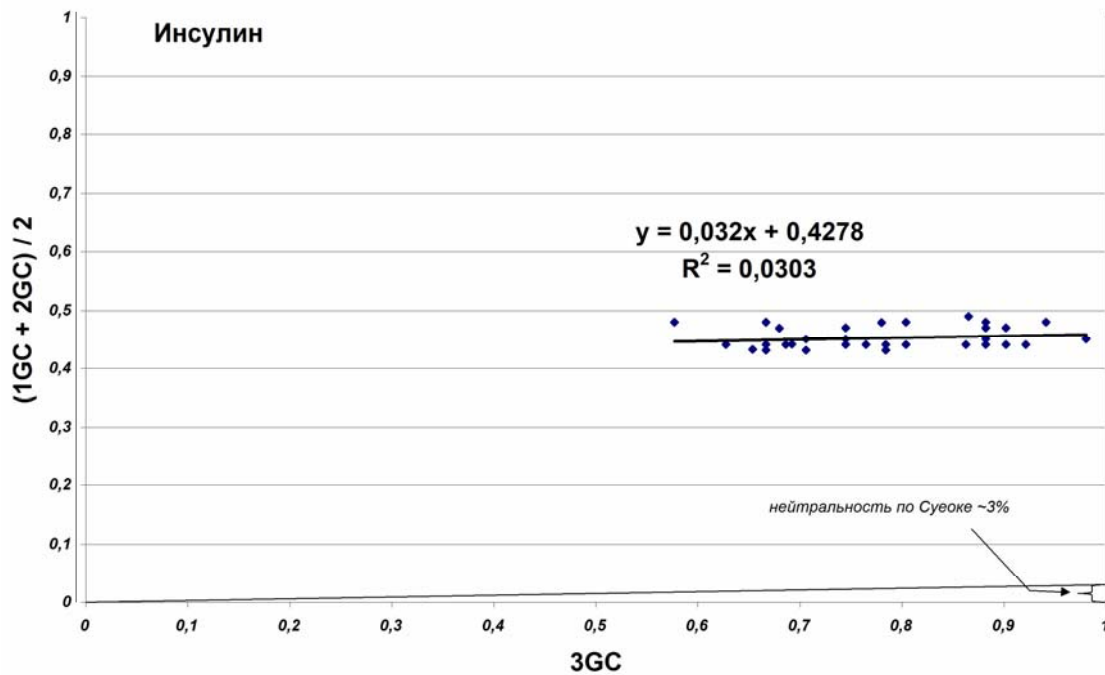


Рис. 32. Зависимость « $(1GC + 2GC) / 2$ » от 3GC в участках, кодирующих инсулин хордовых животных

Уровень нейтральности эволюции С-пептида более чем в 10 раз превышает таковой для инсулина (рис. 33). Зависимость между « $(1GC + 2GC) / 2$ » и 3GC в участках, кодирующих С-пептид, достоверна.

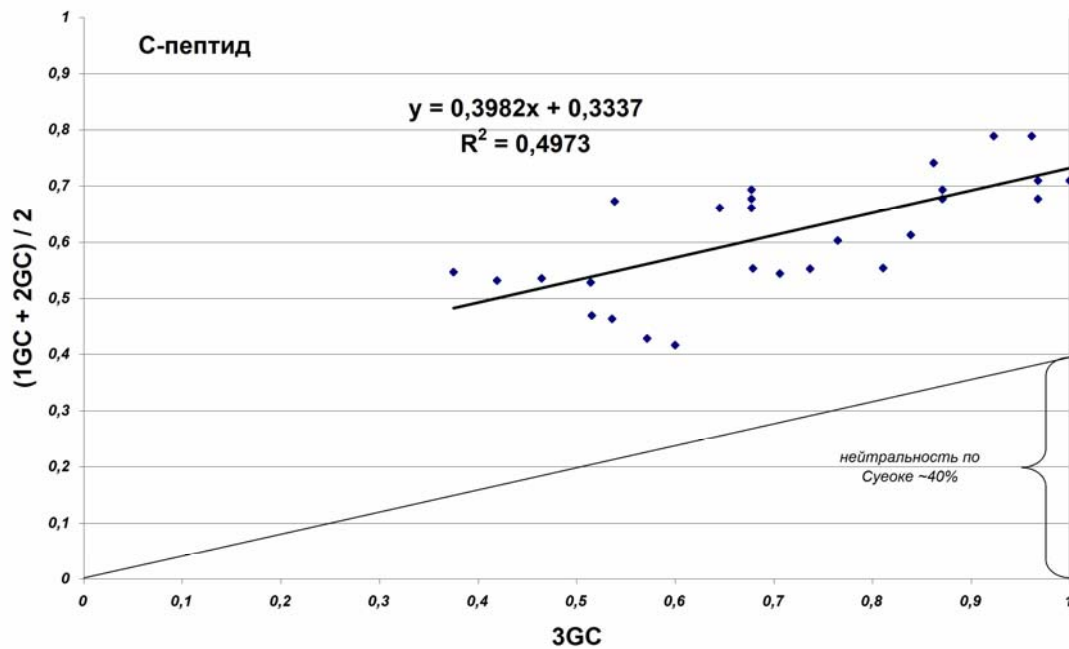


Рис. 33. Зависимость « $(1GC + 2GC) / 2$ » от 3GC в участках, кодирующих С-пептид хордовых животных

Показатель нейтральности эволюции сигнального пептида (рис. 34) более чем в три раза ниже, чем таковой для С-пептида. Судя по результатам определения нейтральности по методу Н. Суеоки, сигнальный пептид должен быть гораздо более консервативным, чем С-пептид. Но так ли это?

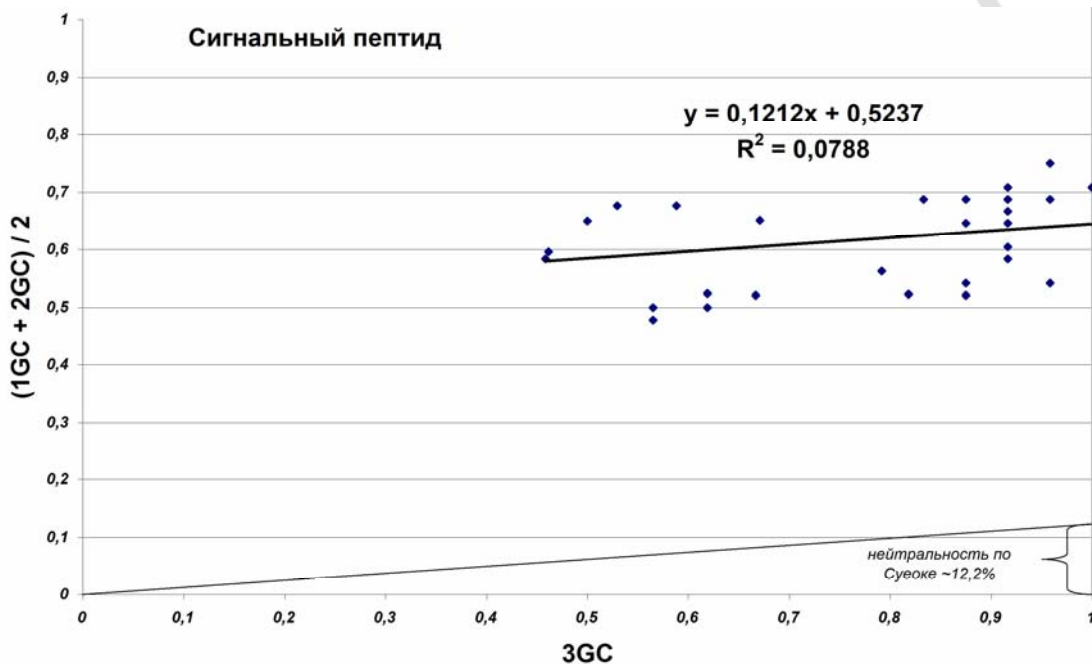


Рис. 34. Зависимость « $(1GC + 2GC) / 2$ » от 3GC в участках, кодирующих сигнальный пептид хордовых животных

**Упрощение аминокислотного состава белка под воздействием мутационного давления.** Судить о том, насколько аминокислотный состав белка подвержен упрощению под влиянием мутационного GC-давления можно путем расчета энтропии распределения аминокислотных остатков в белке [83]. Наибольшее значение энтропии будет в белке, в котором частоты использования всех 20 аминокислотных остатков равны друг другу. Чем больше неравномерность в распределении частот использования аминокислотных остатков, тем ниже энтропия [83]. GC-богатыми (в первых и вторых положениях) кодонами в универсальном генетическом коде кодируются только четыре аминокислотных остатка (см. прил. 1). Частоты использования этих аминокислотных остатков (GARP: Gly, Ala, Arg и Pro) должны расти за счет замен AT на GC. Рост частот использования этих четырех аминокислотных остатков должен происходить за счет закономерного снижения частот использования остальных шестнадцати [83]. Можно предполагать, что в белках, кодируемых более GC-насыщенными генами, энтропия распределения аминокислотных остатков будет ниже, чем в белках, кодируемых генами со средней GC-насыщенностью. На рис. 35 приведены графики зависимости энтропии распределения ами-



нокислотных остатков в инсулине, С-пептиде и сигнальном пептиде различных животных от GC-насыщенности участков, кодирующих их.

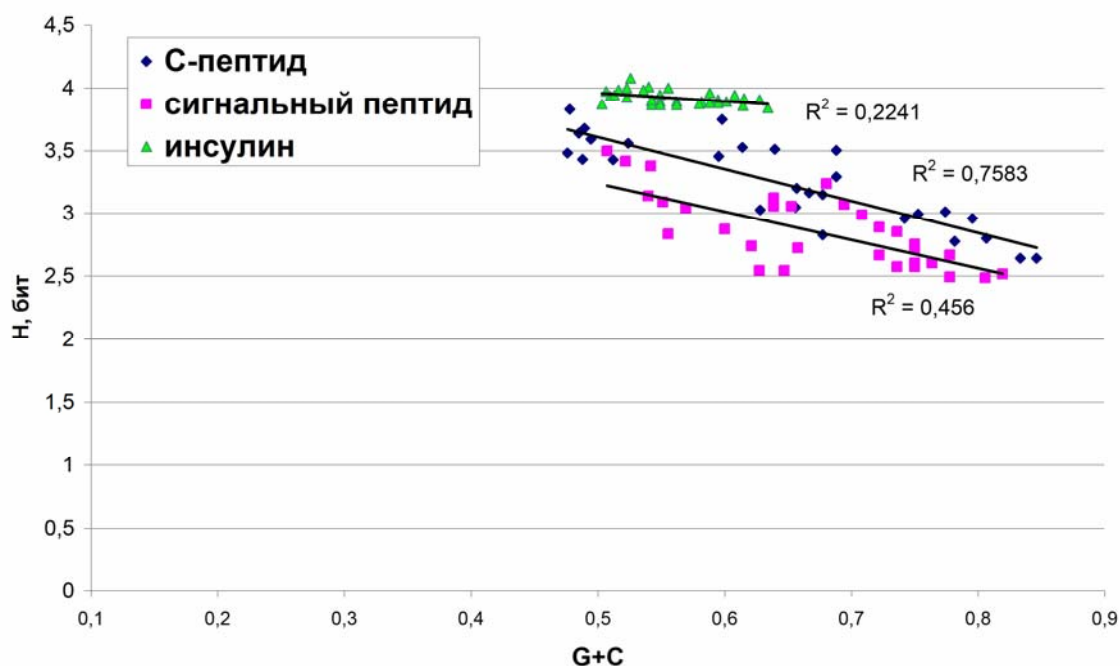


Рис. 35. Зависимость энтропии распределения аминокислотных остатков (H) от GC-насыщенности участков, кодирующих их, в С-пептиде, сигнальном пептиде и инсулине хордовых животных

Как видно на рис. 35, энтропия распределения аминокислотных остатков в С-пептиде и в сигнальном пептиде действительно закономерно снижается по мере увеличения GC-насыщенности кодирующих их участков. Зависимость H от G+C для инсулина слаба. Отсутствие сильной или средней силы зависимости H от G+C является доказательством наличия отрицательного отбора, стабилизирующего аминокислотный состав белка, не дающего ему упроститься под влиянием мутационного давления.

Энтропия распределения аминокислотных остатков в целом выше для С-пептида, чем для сигнального пептида. Действительно, если вернуться к рис. 28, можно заметить большое количество аминокислотных остатков лейцина в сигнальном пептиде. Обогащенность лейцином повышает гидрофобность сигнального пептида, приводя к снижению энтропии.

С-пептид, кодируемый наиболее GC-богатым участком гена (из генома *Bos taurus*, G+C = 84,5 %), содержит только 7 видов аминокислотных остатков (глицин, аланин, аргинин, пролин, а так же серин, глутаминовую кислоту и триптофан).

Человеческий С-пептид, кодируемый несколько менее GC-насыщенным участком гена, содержит 11 видов аминокислотных остатков (глицин, аланин, аргинин, пролин, а также лейцин, треонин, валин, цистеин, серин, глутамин и триптофан).



Два С-пептида лягушки, кодируемые участками со средним уровнем GC-насыщенности ( $G+C = 48,8$  и  $47,6\%$ ), отличаются наибольшим разнообразием аминокислотного состава (14 видов аминокислотных остатков как в первом, так и во втором С-пептиде).

Получается, что под воздействием GC-давления происходит постепенное «выжигание» всех аминокислотных остатков в белке, за исключением четырех аминокислотных остатков из группы GARP, кодируемых GC-богатыми кодонами, частоты использования которых растут.

Судя по рис. 35, аминокислотный состав С-пептида и сигнального пептида «упрощается под одним и тем же углом». Чем же тогда объяснить более чем трехкратную разницу в степени их эволюционной нейтральности? Для этого необходимо провести следующий этап анализа.

Как мы уже упоминали, GC-насыщенными в первых и вторых положениях кодонами кодируются четыре аминокислоты (GARP), GC-бедными — шесть (FYMINK: Phe, Tyr, Met, Ile, Asn и Lys). Еще десять аминокислот (10AA) кодируются средними по GC-насыщенности аминокислотными остатками (см. прил. 1).

На рис. 36, 37 и 38 представлены графики с зависимостями уровней использования этих трех групп аминокислотных остатков от GC-насыщенности кодирующих их участков в инсулине, С-пептиде и сигнальном пептиде соответственно.

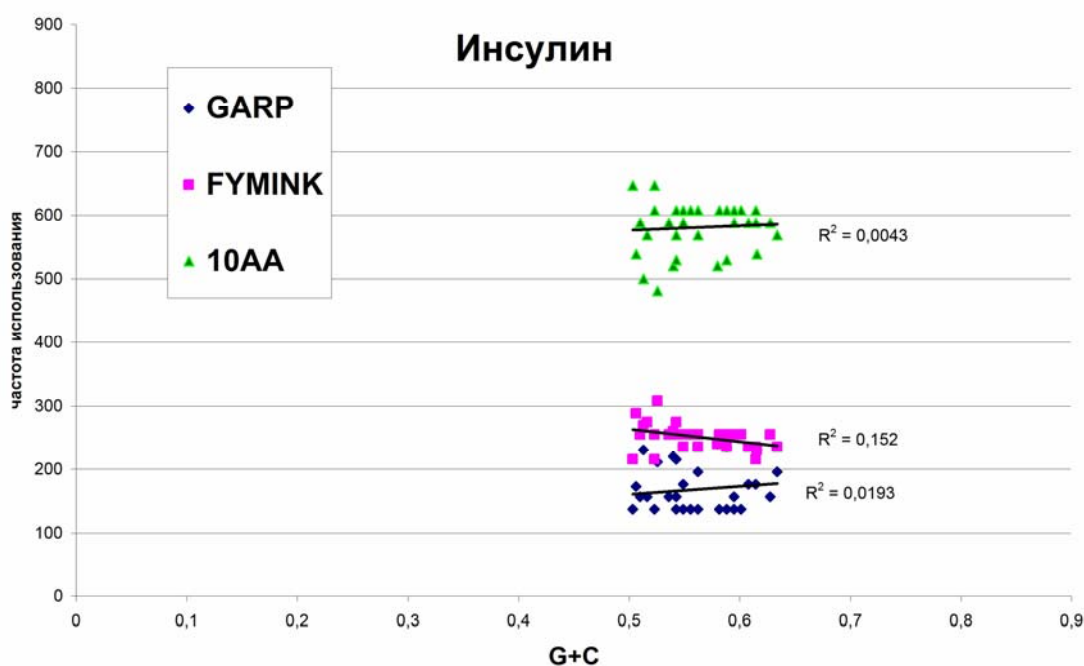


Рис. 36. Зависимость частот использования трех групп аминокислотных остатков (GARP, FYMINK и 10AA) от GC-насыщенности участков, кодирующих его, в инсулине хордовых животных

Сильной или средней силы зависимости между частотами использования трех групп аминокислотных остатков в инсулине и GC-насыщен-

ностью соответствующих кодирующих остатков нет. Тем не менее нельзя сказать, что аминокислотный состав инсулина неизменен среди исследованных хордовых животных. Определенные вариации хорошо заметны на рис. 36.

Частота использования GARP в С-пептиде закономерно растет по мере увеличения G+C в соответствующих кодирующих участках, в то время как частота использования FYMINK падает практически до нуля (рис. 37). При этом общая частота использования средних по GC-насыщенности аминокислотных остатков остается независимой от G+C кодирующих участков.

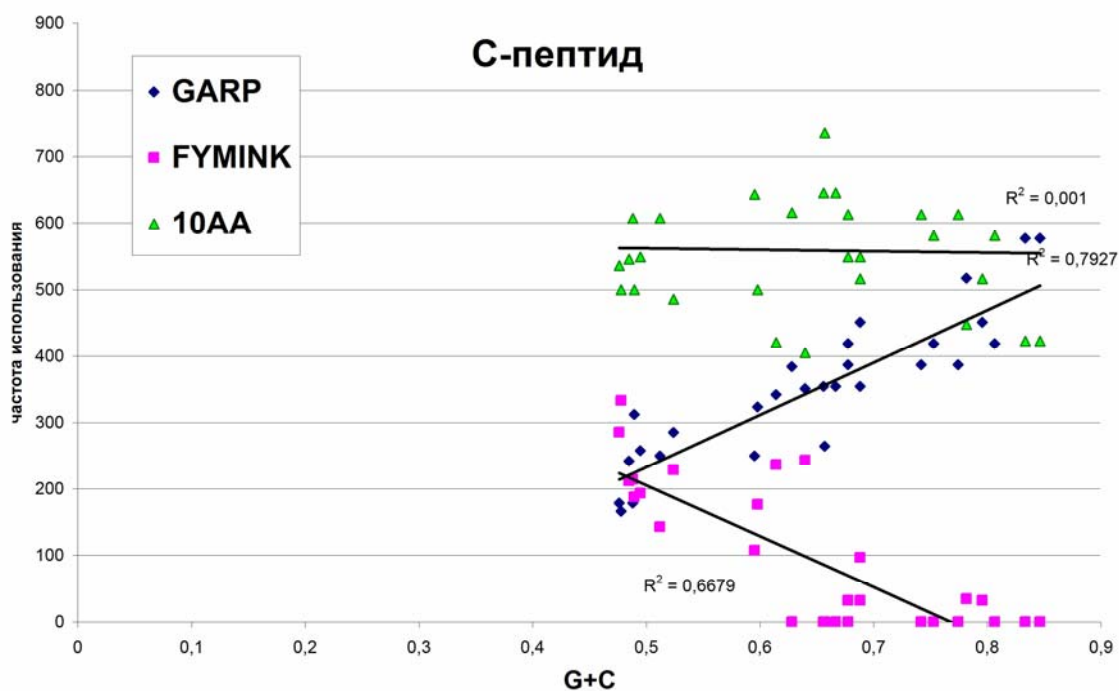


Рис. 37. Зависимость частот использования трех групп аминокислотных остатков (GARP, FYMINK и 10AA) в С-пептиде хордовых животных от GC-насыщенности участков, кодирующих его

В сигнальном пептиде, наоборот, уровень 10AA падает по мере увеличения GC-насыщенности кодирующих его участков, а коэффициент корреляции для зависимости FYMINK от G+C ниже 0,5. При этом, как и в С-пептиде, уровень GARP демонстрирует прямую зависимость от GC-насыщенности кодирующих участков (рис. 38).

Единственным способом выяснить, за счет чего графики на рис. 37 и 38 отличаются друг от друга, является расчет коэффициентов корреляции частот использования каждого аминокислотного остатка в белке с G+C в кодирующих его участках. Для начала приводим табл. 19, в которой показаны коэффициенты корреляции с G+C частот использования аминокислот в инсулине.

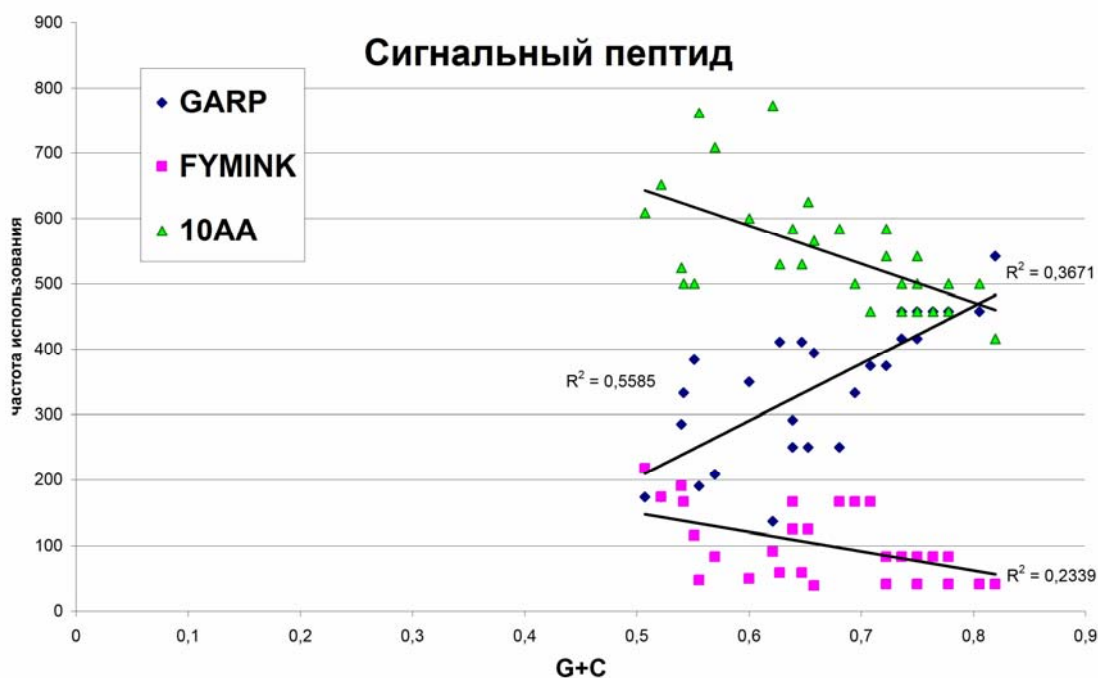


Рис. 38. Зависимость частот использования трех групп аминокислотных остатков (GARP, FYMINK и 10AA) от GC-насыщенности участков, кодирующих его, в сигнальном пептиде хордовых животных

Таблица 19

**Коэффициенты корреляции (R) между частотами использования аминокислотных остатков в инсулине хордовых животных и GC-насыщенностью кодирующих их участков**

FYMINK		10AA		GARP	
Phe	0,181	Leu	0,222	Gly	0,020
Tyr	0,069	Ser	-0,245	Ala	0,322
Met	-0,331	Thr	-0,180	Arg	-0,074
Ile	-0,330	Val	0,532	Pro	0,053
Asn	-0,059	His	-0,094		
Lys	-0,291	Gln	0,086		
		Asp	-0,323		
		Glu	0,398		
		Cys	0,155		
		Trp	–		

*Примечание.* Курсивом выделен коэффициент корреляции, модуль которого превышает 0,5.

Единственным аминокислотным остатком в инсулине хордовых животных, который демонстрирует зависимость от G+C, является валин. Частота использования валина, по-видимому, увеличивается в результате замен АТ на GC, происходящих в кодонах изолейцина или лейцина. Эти аминокислоты обладают относительно близкими физико-химическими свойствами с валином [2].

Весьма неожиданно выглядят данные из табл. 20. Как оказалось, высокая общая частота использования GARP в С-пептидах, кодируемых GC-богатыми участками гена препроинсулина, обусловлена исключительно приростом частоты использования глицина. Последняя же растет исключительно за счет снижения частоты использования аспарагиновой кислоты. Замена Asp на Gly вызывается транзицией А на G. Получается, что все остальные аминокислотные замены от 10AA к GARP в С-пептиде фиксируются очень редко. Снижение частоты использования FYMINK, в частности Phe и Ile, приводит к росту частоты использования Leu. Замены же последнего на GARP практически не фиксируются. Прирост Gly, по видимому, во многом обусловлен еще и снижением частот использования Lys и Asn из группы FYMINK. Получается, что аминокислотный состав С-пептида относительно стабилен. Уровень FYMINK снизился почти до нуля еще на отметке G+C = 0,65 (см. рис. 37). Дальнейшее же упрощение аминокислотного состава происходит только за счет замен Asp на Gly.

Таблица 20

**Коэффициенты корреляции (R) между частотами использования аминокислотных остатков в С-пептиде хордовых животных и GC-насыщенностью кодирующих их участков**

FYMINK		10AA		GARP	
Phe	-0,691	Leu	0,674	Gly	0,879
Tyr	-0,581	Ser	-0,071	Ala	0,405
Met	-0,500	Thr	-0,461	Arg	-0,065
Ile	-0,579	Val	-0,121	Pro	0,179
Asn	-0,591	His	-0,401		
Lys	-0,733	Gln	0,141		
		Asp	-0,789		
		Glu	0,341		
		Cys	-		
		Trp	-		

*Примечание.* Курсивом выделены коэффициенты корреляции, модуль которых превышает 0,5.

Совершенно по-другому выглядит картина для сигнального пептида (табл. 21). Уровень GARP в сигнальных белках, кодируемых GC-богатыми участками гена препроинсулина, растет за счет прироста частоты использования аланина, аргинина и пролина. Происходит это за счет фиксации замен Ser и Val (из группы 10AA) на эти три аминокислоты. При этом, как и в С-пептиде, отмечается прирост частоты использования Leu, судя по табл. 21, за счет замен Phe на Leu. Уровень FYMINK в сигнальном белке ниже при средних значениях G+C, чем в С-пептиде. Из-за этого (из-за изначально низкого содержания) обратная зависимость между FYMINK и G+C является слабой для сигнального белка.

**Коэффициенты корреляции (R) между частотами использования  
аминокислотных остатков в сигнальном пептиде хордовых животных  
и GC-насыщенностью кодирующих их участков**

FYMINK		10AA		GARP	
Phe	<i>-0,528</i>	Leu	<i>0,615</i>	Gly	<i>-0,253</i>
Tyr	<i>-0,261</i>	Ser	<i>-0,666</i>	Ala	<i>0,577</i>
Met	<i>0,155</i>	Thr	<i>-0,232</i>	Arg	<i>0,650</i>
Ile	<i>-0,280</i>	Val	<i>-0,627</i>	Pro	<i>0,620</i>
Asn	<i>-0,471</i>	His	<i>-0,061</i>		
Lys	<i>-0,015</i>	Gln	<i>-0,380</i>		
		Asp	<i>0,332</i>		
		Glu	<i>-0,169</i>		
		Cys	<i>-0,322</i>		
		Trp	<i>0,485</i>		

*Примечание.* Курсивом выделены коэффициенты корреляции, модуль которых превышает 0,5.

Наши данные убедительно показывают, что для С-пептида только одна аминокислотная замена 10AA на GARP (Asp на Gly) нейтральная, в то время как для сигнального пептида таких замен гораздо больше (замены Ser и Val на Ala, Arg и Pro). Как же объяснить тот факт, что нейтральность эволюции по Н. Суеоке для сигнального белка гораздо ниже, чем для С-пептида?

Дело в том, что на графике с рис. 34 значения « $(1GC + 2GC) / 2$ », что называется, «разбросаны в высоту». И действительно, одному и тому же значению 3GC соответствует много различных значений « $(1GC + 2GC) / 2$ ». Особенно хорошо это заметно для предельно высоких значений 3GC. Если вернуться к рис. 31, то можно вспомнить о существовании «крыши» для значений 3GC. После заполнения третьих положений гуанином и цитозинном, начинается заполнение первых и вторых. Этот процесс не может быть правильно отражен в методике Н. Суеоки.

«Разброс в высоту» на рис. 34 приводит к тому, что линия тренда проводится по середине этого диапазона значений « $(1GC + 2GC) / 2$ », а не по верхней его границе, что было бы правильным с точки зрения биологии, но не с точки зрения математики. Получается, что метод Н. Суеоки не адаптирован к определению нейтральности эволюции белков, которые кодируются генами, подверженными сильному мутационному давлению.

Как показал наш детальный анализ, вычисление единого уровня нейтральности эволюции иногда может быть дезориентирующим. Гораздо более обоснованным кажется определение уровня нейтральности замен, вызванных мутационным давлением, для каждого отдельного аминокислотного остатка.

Об общем уровне нейтральности, по нашему мнению, нужно судить, используя «качественный подход». Чем больше типов аминокислотных замен являются нейтральными, тем больше должна быть общая нейтральность эволюции белка, подверженного мутационному давлению. С этой точки зрения, С-пептид можно считать гораздо более консервативным белком, чем сигнальный пептид. Действительно, аминокислотные последовательности С-пептида практически идентичны у большинства приматов [16]. Хочется подчеркнуть, что в С-пептиде присутствуют серин, лейцин, валин, глутамин, которые могли бы замениться на аланин, пролин и аргинин в результате замен АТ на GC. Судя по запредельной GC-насыщенности третьих положений кодонов в участках, кодирующих С-пептид приматов, такие замены должны периодически происходить. Но они практически не закрепляются в популяции. Значит ли это, что С-пептид выполняет некую важную функцию? Экспериментальные данные свидетельствуют о том, что это действительно так.

Полученные нами данные филогенетического анализа указывают на то, что С-пептид выполняет некие важные функции не только у человека, но и у некоторых гистрикоморфных грызунов, в частности у морской свинки и дегу. На рис. 39 приведена дендрограмма, отражающая степень сходства аминокислотных последовательностей инсулина среди хордовых животных. Инсулины морской свинки и дегу группируются на отдельной ветви, расположенной между общими ветвями, на которых расположены инсулины рыб и всех остальных позвоночных. По данным, приведенным в книге М. Кимуры, инсулин морской свинки действительно накопил большое количество аминокислотных замен, не встречающихся в молекулах инсулина всех остальных животных. Вследствие одной критической замены (замены гистидина в 10 положении В-цепи) инсулин морской свинки потерял свою функцию, после чего все селекционные ограничения для него исчезли [16]. Организм морской свинки (*Cavia porcellus*) каким-то образом сумел адаптироваться к отсутствию функционального инсулина. По нашим данным, аналогичная ситуация характерна и для другого представителя гистрикоморфных грызунов — для дегу (*Octodon degus*).

На дендрограмме с рис. 40 С-пептиды морской свинки и дегу занимают положение, характерное для других генов гистрикоморфных грызунов: рядом с ветвями, отходящими к С-пептидам кошки, собаки, кролика и суслика.

Топология дендрограммы, отражающей степень сходства сигнальных пептидов хордовых животных (рис. 41), несколько отличается от таковой для дендрограммы с рис. 40. Тем не менее сигнальные пептиды морской свинки и дегу на ней группируются среди сигнальных пептидов животных (а именно — приматов), а не на отдельной ветви, как на рис. 39.

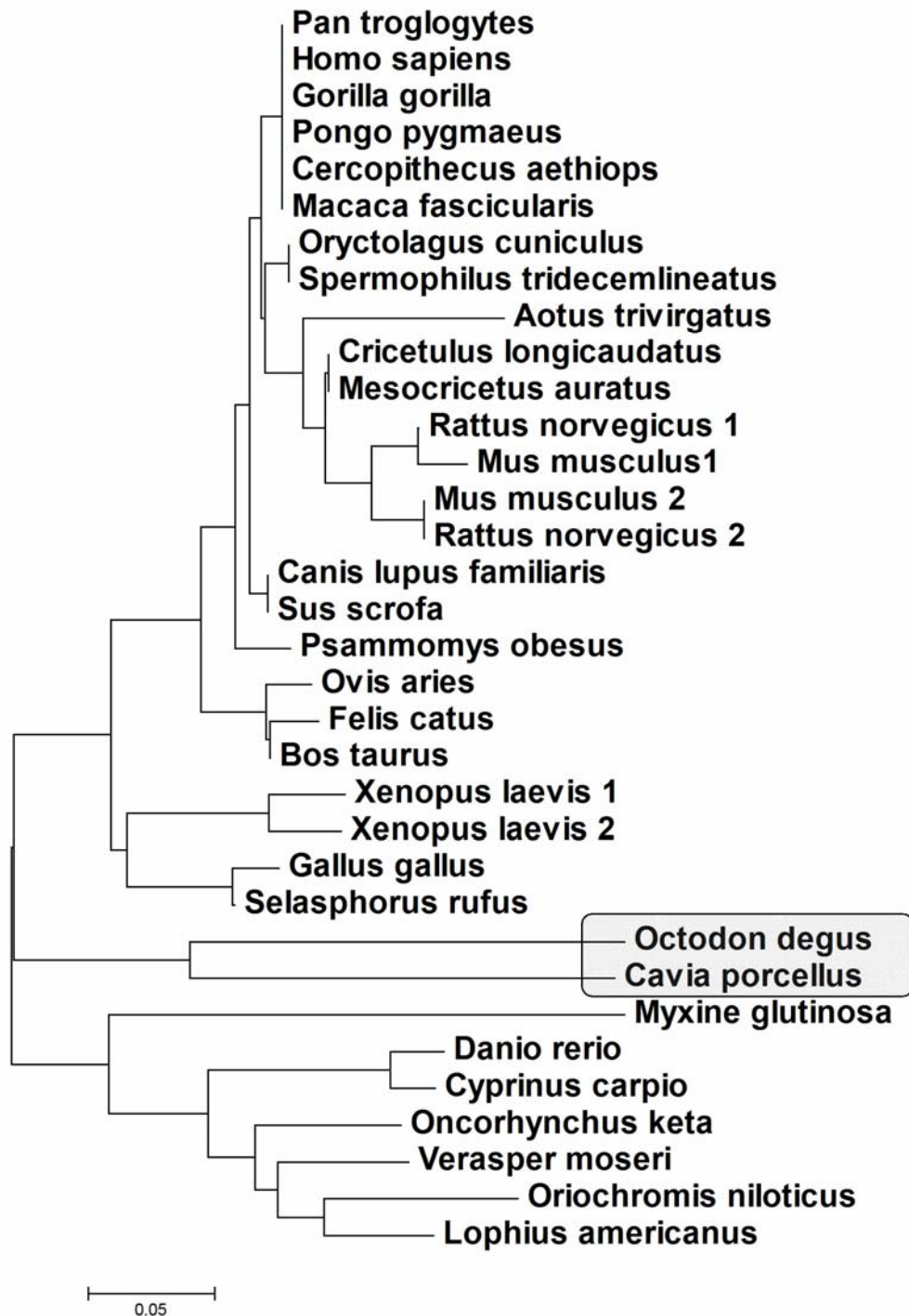


Рис. 39. NJ-дендрограмма, отражающая сходство (р-дистанция) аминокислотных последовательностей инсулина хордовых животных

Получается, что, несмотря на утерю функционального инсулина, С-пептид морской свинки и дегу продолжает выполнять свои функции (селекционные ограничения с него не сняты), а сигнальный пептид по-прежнему обеспечивает транспорт их препроинсулина в эндоплазматический ретикулум.



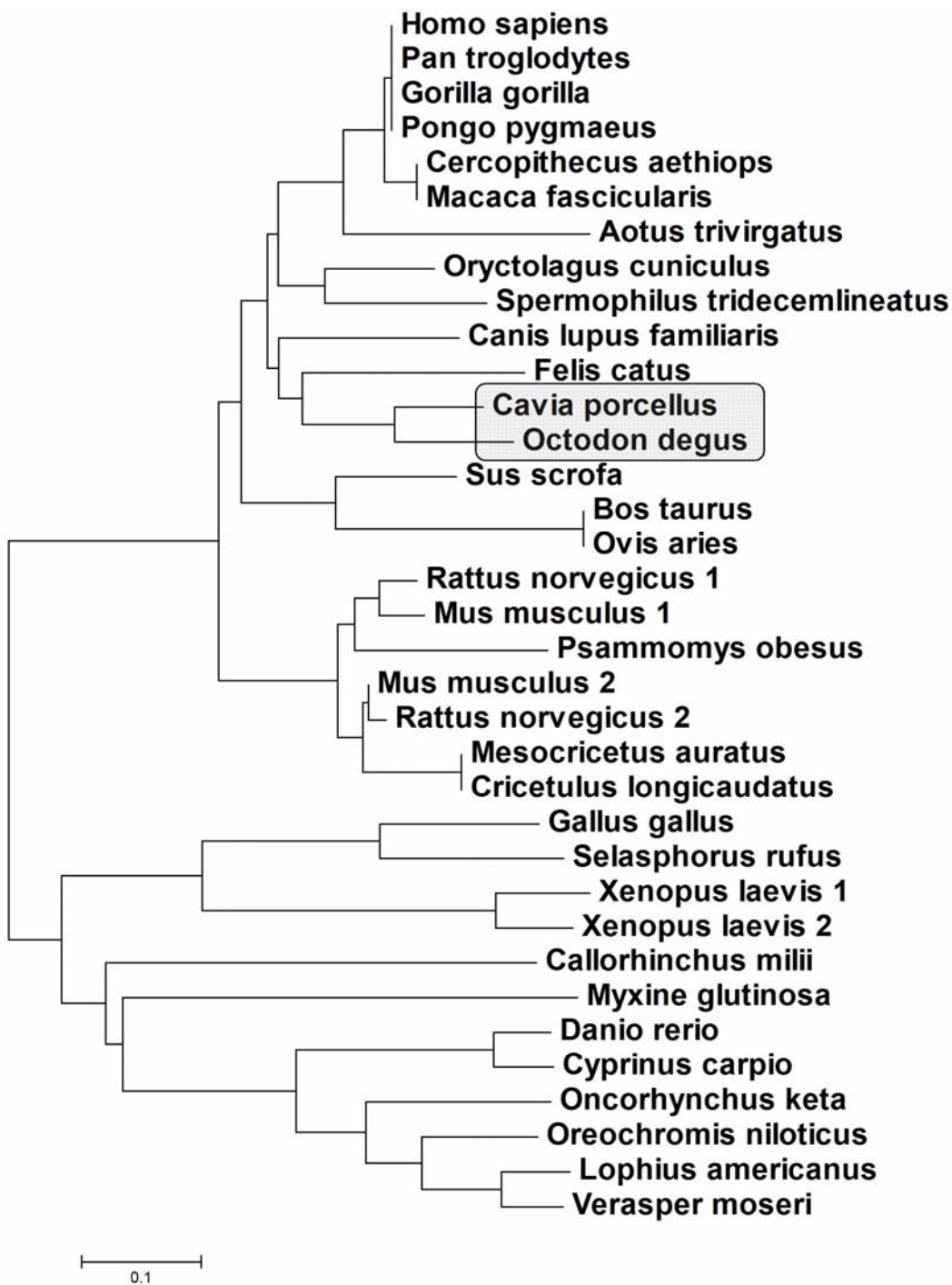


Рис. 40. NJ-дендрограмма, отражающая сходство (р-дистанция) аминокислотных последовательностей С-пептида хордовых животных



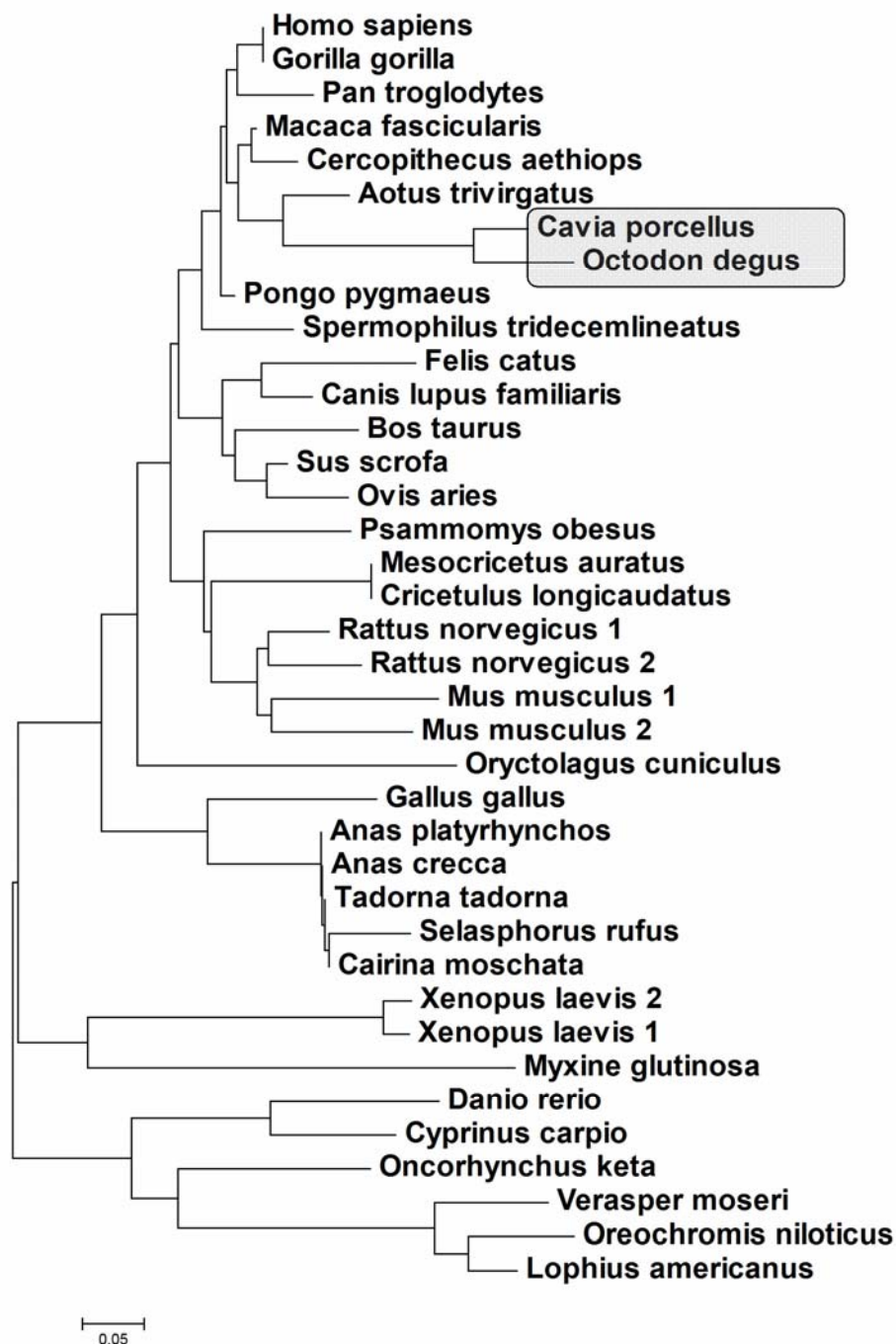


Рис. 41. NJ-дендрограмма, отражающая сходство (p-дистанция) аминокислотных последовательностей сигнального пептида хордовых животных

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью описанного метода можно определить степень нейтральности каждой отдельной аминокислотной замены, происходящей под воздействием мутационного давления в конкретном белке или в его участке. С помощью нашего метода была четко продемонстрирована относительная консервативность аминокислотного состава С-пептида, что противоречит результатам, полученным с помощью методики Н. Суеки.

На данный момент времени выяснено, что С-пептид связывается с клеточным трансмембранным рецептором, активируя каскад реакций, приводящих к стимуляции  $\text{Na}^+$ ,  $\text{K}^+$  АТФ-азы и эндотелиальной NO-синтазы. С-пептид способствует репарации мышечной стенки артерий. Это означает, что С-пептид должен оказывать терапевтический эффект при таких осложнениях сахарного диабета, как диабетическая невропатия, ангиопатия и нефропатия. Несмотря на то, что об этих свойствах С-пептида известно (вплоть до того, что данная информация доступна в свободной энциклопедии: [www.wikipedia.com](http://www.wikipedia.com)), он до сих пор не включен в схему терапии сахарного диабета.

С помощью описанного в данной главе метода можно предположить наличие функции у ранее мало исследованного белка. Знания об относительно нейтральных и радикальных типах аминокислотных замен можно использовать при создании синтетических белков и пептидов.

## Глава 2

### Анализ механизмов упрощения аминокислотного состава протеомов альфа-герпесвирусов под воздействием GC-давления

**Цель исследования:** определить степень и характер селекционных ограничений, накладываемых отрицательным отбором на аминокислотный состав полных протеомов альфа-герпесвирусов.

**Область применения:** данный тип молекулярно-биологического анализа может быть применен в случае необходимости определить степень и характер селекционных ограничений, накладываемых на аминокислотный состав белков родственных вирусов или бактерий отрицательным отбором в условиях мутационного давления; определение степени нейтральности отдельных аминокислотных замен у разных видов организмов имеет большое значение для теоретических исследований по созданию универсальных шкал степени радикальности аминокислотных замен; данные, полученные путем описанного ниже анализа, могут быть использованы в молекулярной филогенетике, в различных областях биотехнологии и в иммунологии.

**Используемые оригинальные алгоритмы:** «CodonChanges», «Coding Genome Scanner» (см. гл. 1 и 4 шестой части).

**Материалы.** В данной главе мы описали метод исследования полных протеомов на примере десяти альфа-герпесвирусов, полные геномы которых были исследованы нами в гл. 2 второй части. Номера доступа полных геномов этих альфа-герпесвирусов в GenBank следующие:

1. Вирусы из рода симплексвирус: Human herpesvirus 1 (HSV1) [NC\_001806], Human herpesvirus 2 (HSV2) [NC\_001798], Cercopithecine

herpesvirus 1 (CeHV1) [NC\_004812], Cercopithecine herpesvirus 2 (CeHV2) [NC\_006560], Cercopithecine herpesvirus 16 (CeHV16) [NC\_007653].

2. Вирусы из рода варицелловирус: Bovine herpesvirus 5 (BoHV5) [NC\_005261], Equid herpesvirus 1 (EqHV1) [NC\_001491], Equid herpesvirus 4 (EqHV4) [NC\_001844], Cercopithecine herpesvirus 9 (CeHV9) [NC\_002686], Human herpesvirus 3 (VZV) [NC\_001348].

Кроме записей о частотах использования кодонов в каждом гене данного генома из Codon Usage Database [104], мы исследовали нуклеотидные последовательности, кодирующие шесть капсидных протеинов и двенадцать гликопротеинов референтных штаммов вируса простого герпеса 1-го типа (ВПГ1) и обезьяньего вируса герпеса 16 (ОВГ16) из GenBank. Названия шести капсидных белков: capsid portal protein, capsid triplex subunit 1, capsid triplex subunit 2, capsid scaffold protein, small capsid protein и major capsid protein. Ген, кодирующий последний белок, был детально изучен в гл. 1 второй части. Названия двенадцати гликопротеинов: L, M, H, B, C, N, K, G, J, D, I, E. Общая длина капсидных белков (вместе с гэпами, образовавшимися в результате выравнивания) — 3297 аминокислотных остатков [83]. Общая длина выровненных гликопротеинов — 5563 аминокислотных остатка [83].

#### **Описание метода.**

**Предварительные операции.** Предварительной операцией, которую необходимо выполнить перед использованием «Coding Genome Scanner», является удаление из текстового файла, хранящегося в Codon Usage Database, записей о частотах использования кодонов в кодирующих участках, не принадлежащих полному геному референтного штамма.

Перед тем как использовать алгоритм «CodonChanges», необходимо выравнивать исследуемые последовательности и заменить в них значки, обозначающие гэпы, на символ «N» (латинская буква). Данные операции важны, если цель исследования — подсчет частот замен между гомологичными генами и поиск наиболее частых замен в кодонах.

В данном исследовании мы изучали частоты замен в кодонах по направлению мутационного GC-давления: от генов, кодирующих гликопротеины и капсидные белки ВПГ1, к гомологичным генам ОВГ16. Мы объединили выровненные последовательности, кодирующие все шесть исследованных капсидных белков ВПГ1 в одну последовательность (3297 кодонов длиной), и скопировали ее в ячейку «sequence1» алгоритма «CodonChanges». Объединенные в том же порядке выровненные последовательности шести капсидных белков ОВГ16 мы поместили в ячейку «sequence2» того же алгоритма. Это объединение позволило нам вычислить более достоверные процентные соотношения частот замен в кодонах (количество замен в каждом кодоне гораздо выше для всех шести кодирующих участков, чем для каждого в отдельности, соответственно, досто-

верней и процентное содержание каждого типа замены). Аналогичным образом мы поступили с участками, кодирующими двенадцать гликопротеинов [83].

**Производимые вычисления.** На листе «results» алгоритма «Coding Genome Scanner» находятся столбцы с формулами для расчета частот использования аминокислотных остатков в каждом белке. Кроме того, «Coding Genome Scanner» вычисляет суммы частот использования аминокислотных остатков, кодируемых GC-богатыми кодонами (GARP: Gly, Ala, Arg и Pro), GC-бедными кодонами (FYMINK: Phe, Tyr, Met, Ile, Asn и Lys) и кодонами средней GC-насыщенности (10AA) (см. прил. 1) [83]. Рассчитывает «Coding Genome Scanner» и энтропию распределения аминокислотных остатков в белке [83].

Согласно теории информации Клода Шеннона [135], максимальное значение энтропии (H) в системе достигается в том случае, если частоты всех ее компонентов равны друг другу. Максимальным значением энтропии для системы из 20 компонентов, в том числе и для белка, является 4,322 бит [83].

Расчет энтропии для каждого кодирующего участка производился по следующей формуле:

$$H = -\sum f_{aa} \cdot \log_2 f_{aa}, \quad (2.1)$$

где  $f_{aa}$  — частота использования каждой из 20 аминокислот.

Чем выше энтропия, тем более «разнообразен» аминокислотный состав белка [83], а чем ниже уровень энтропии, тем он более «однороден» [83]. Величину, обозначаемую в формуле 2.1 латинской буквой «H», называют не только «информационной энтропией», но и «количеством информации» [135]. Чем равномерней распределены частоты использования компонентов в системе, тем больше число возможных комбинаций этих компонентов, тем выше «количество информации», которое может содержаться в данной системе [135].

После вычисления всех интересующих нас показателей для каждого гена и белка из десяти геномов и протеомов, мы объединили эти данные в одну выборку [83]. Затем мы рассортировали гены альфа-герпесвирусов по уровню их GC-насыщенности и разбили их на следующие группы:  $0,3 < G+C < 0,4$ ;  $0,4 < G+C < 0,5$ ;  $0,5 < G+C < 0,6$ ;  $0,6 < G+C < 0,7$ ;  $0,7 < G+C < 0,8$ ;  $0,8 < G+C < 0,85$  [83]. Число генов в каждой группе превышает 30 [83].

Такое разделение генов на группы позволило нам сравнивать содержание аминокислотных остатков в кодируемых ими белках при помощи методов сравнительной статистики.

Алгоритм «CodonChanges» позволяет подсчитывать количество (а значит и процентное соотношение) типов замен, которые произошли в каждом кодоне. Для подсчета необходимо ввести интересующий кодон в ячейку на листе «codon» и открыть лист под названием «changes».

На листе «changes» изображена таблица универсального генетического кода, в которой указано количество замен введенного кодона на все остальные кодоны и на гэпы (по направлению от нуклеотидной последовательности, введенной в ячейку на листе «sequence1», к таковой, введенной в ячейку на листе «sequence2») [83].

**Упрощение аминокислотного состава протеомов под воздействием мутационного давления.** Судить о том, насколько аминокислотный состав протеомов родственных вирусов или бактерий подвержен упрощению под влиянием мутационного давления можно путем построения графика зависимости энтропии распределения аминокислотных остатков в белках от GC-насыщенности кодирующих их генов [83, 85]. Нами был доказан тот факт, что в геномах ВПГ1, ВПГ2, ОБГ1, ОБГ2, ОБГ16 и вируса герпеса крупного рогатого скота 5 существует сильное мутационное GC-давление (см. гл. 2 второй части). Частоты использования аминокислотных остатков, кодируемых GC-богатыми кодонами (GARP: Gly, Ala, Arg и Pro), должны расти в белках этих вирусов за счет замен АТ на GC в кодирующих их участках. Рост частот использования этих четырех аминокислотных остатков должен происходить за счет закономерного снижения частот использования остальных шестнадцати. Можно предполагать, что в белках, кодируемых более GC-насыщенными генами, энтропия распределения аминокислотных остатков будет ниже, чем в белках, кодируемых генами со средней GC-насыщенностью. К последним относится большинство белков лошадиных вирусов герпеса 1 и 4 и варицелло-зостер-вируса человека. К сожалению, судить об упрощении аминокислотного состава белков под влиянием АТ-давления на примере данной выборки нельзя. Несмотря на то, что геном ОБГ9 подвержен сильному АТ-давлению, GC-насыщенность его генов еще не успела опуститься до того уровня, когда рост частоты использования FYMINK в кодируемых ими белках вызовет снижение энтропии. На рис. 42 приведен график зависимости энтропии распределения аминокислотных остатков во всех белках десяти референтных штаммов альфа-герпесвирусов от GC-насыщенности участков, кодирующих их.

Как видно на рис. 42, энтропия распределения аминокислотных остатков в белках альфа-герпесвирусов снижается по мере увеличения GC-насыщенности кодирующих их генов. Однако снижение это происходит неравномерно: в генах с  $0,3 < G+C < 0,6$  зависимость  $H$  от  $G+C$  вообще отсутствует, а в генах с  $G+C > 0,6$  энтропия находится в обратной зависимости от  $G+C$ , причем чем выше GC-насыщенность гена, тем «круче» снижается энтропия [83].

Необходимо подчеркнуть, что снижение  $H$  по мере роста  $G+C$  — общая закономерность, характерная для всех белков, кодируемых генами с уровнем GC-насыщенности больше 0,6 [83]. Почему же значения энтропии примерно одинаковы для белков, кодируемых генами с  $G+C$  от 0,3 до 0,6?

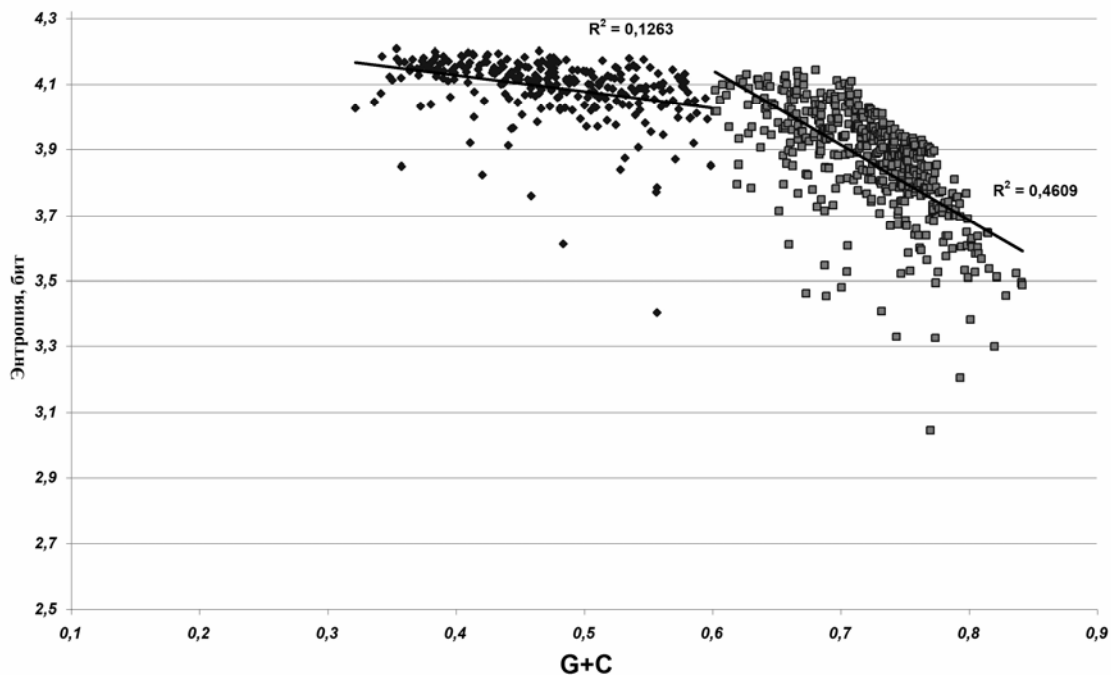


Рис. 42. Зависимость энтропии распределения аминокислотных остатков (H) в белках десяти альфа-герпесвирусов от GC-насыщенности участков, кодирующих их [83]

Чтобы ответить на этот вопрос мы построили еще один объединенный график, на оси Y которого отложены частоты использования GARP, FYMINK и 10AA для каждого белка десяти альфа-герпесвирусов (рис. 44).

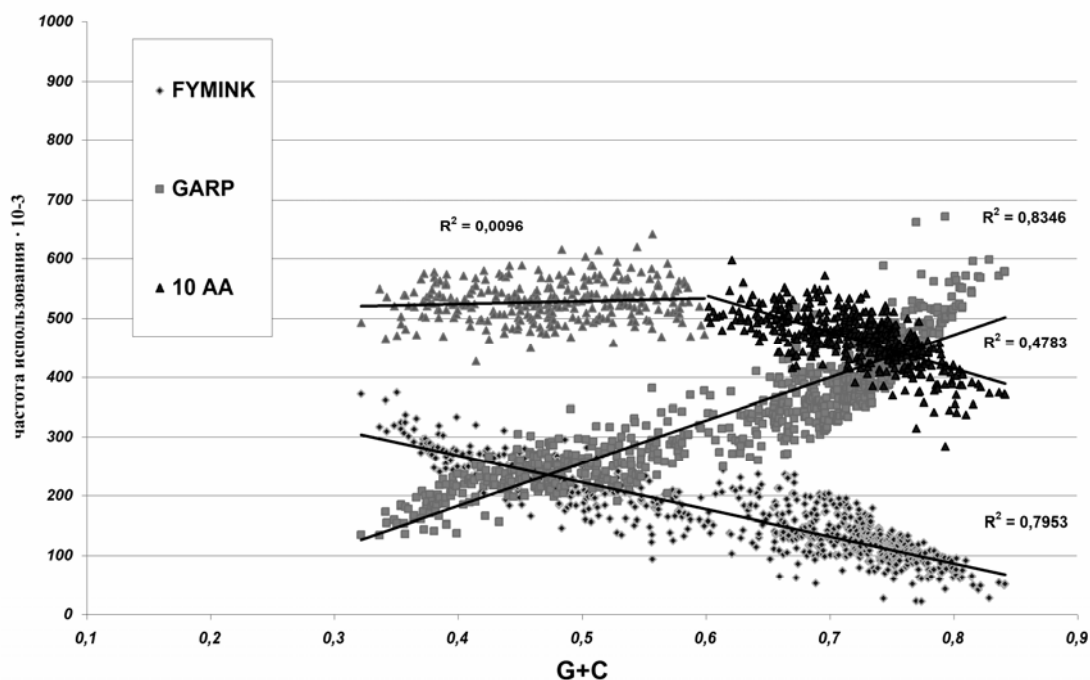


Рис. 43. Зависимость частот использования четырех аминокислотных остатков, кодируемых GC-богатыми триплетами (GARP), шести аминокислотных остатков, кодируемых GC-бедными триплетами (FYMINK), и десяти аминокислотных остатков, кодируемых средними по GC-насыщенности триплетами (10AA), в белках альфа-герпесвирусов от уровня GC-насыщенности соответствующих генов (G+C) [83]

На рис. 43 видно, что прямая зависимость между G+C и GARP присуща всем генам и кодируемым ими белкам, так же как и отрицательная зависимость между G+C и FYMINK. Для белков, которые кодируются генами с GC-насыщенностью в пределах от 0,3 до 0,6, уровень использования 10AA остается стабильным. Только для генов с  $G+C > 0,6$  характерна отрицательная зависимость между G+C и уровнем 10AA в кодируемых ими белках [83]. Этот эффект связан с тем, что сильное мутационное GC-давление вызывает частые мутации в средних по GC-насыщенности триплетах: постепенно фиксируются замены аминокислот из группы 10AA на аминокислоты из группы GARP. В результате, частота использования четырех аминокислот растет за счет падения частот использования остальных шестнадцати. Энтропия начинает снижаться только в данном случае, но остается на том же уровне, если частота использования четырех аминокислот растет за счет снижения частот использования только шести аминокислот из группы FYMINK [83].

У варицелловируса обезьяны (ОВГ9), находящегося под влиянием сильного АТ-давления, частоты использования шести аминокислот (FYMINK) растут за счет снижения частот использования четырех (GARP), а уровни использования остальных десяти (10AA) в среднем остаются на том же уровне. С этой особенностью можно связать факт стабильности энтропии распределения аминокислотных остатков в белках этого вируса. Снижение H, по нашим данным, происходит и под действием сильного АТ-давления, но в белках, кодируемых еще более GC-бедными генами ( $G+C < 0,3$ ) [85].

Чтобы оценить вклад каждого аминокислотного остатка в динамику общей частоты использования FYMINK, мы рассчитали средние уровни фенилаланина, тирозина, метионина, изолейцина, аспарагина и лизина в белках, кодируемых генами альфа-герпесвирусов разной GC-насыщенности. Результат приведен на рис. 44.

Как видно на рис. 44, снижение частот использования всех пяти аминокислот из группы FYMINK вносит вклад в снижение общей частоты использования FYMINK под воздействием GC-давления. Тем не менее можно заметить, что частоты использования фенилаланина и тирозина снижаются под воздействием GC-давления не так резко, как частоты использования лизина, аспарагина и изолейцина. Это свидетельствует о том, что замены фенилаланина и тирозина по направлению GC-давления более радикальны, чем замены лизина, аспарагина и изолейцина. В столбце с частотами использования FYMINK в генах с  $0,5 < G+C < 0,6$  наблюдается картина примерного равенства частот использования этих аминокислот. Это убеждает нас в том, что причина более резкого снижения частот использования лизина, аспарагина и изолейцина не в том, что их уровень был выше изначально, а в том, что их замены чаще оказываются нейтральными.

ми. С другой стороны, при АТ-давлении частоты использования лизина, аспарагина и изолейцина повышаются гораздо более резко, чем частоты использования фенилаланина и тирозина. Получается, что общая частота FYMINK демонстрирует корреляцию с G+C в большей степени за счет относительно нейтральных замен лизина, аспарагина и изолейцина на 10АА при GC-давлении и обратных замен — при АТ-давлении, чем за счет относительно радикальных замен тирозина и фенилаланина. Особенно широко варьирует (в зависимости от G+C) частота использования изолейцина [85].

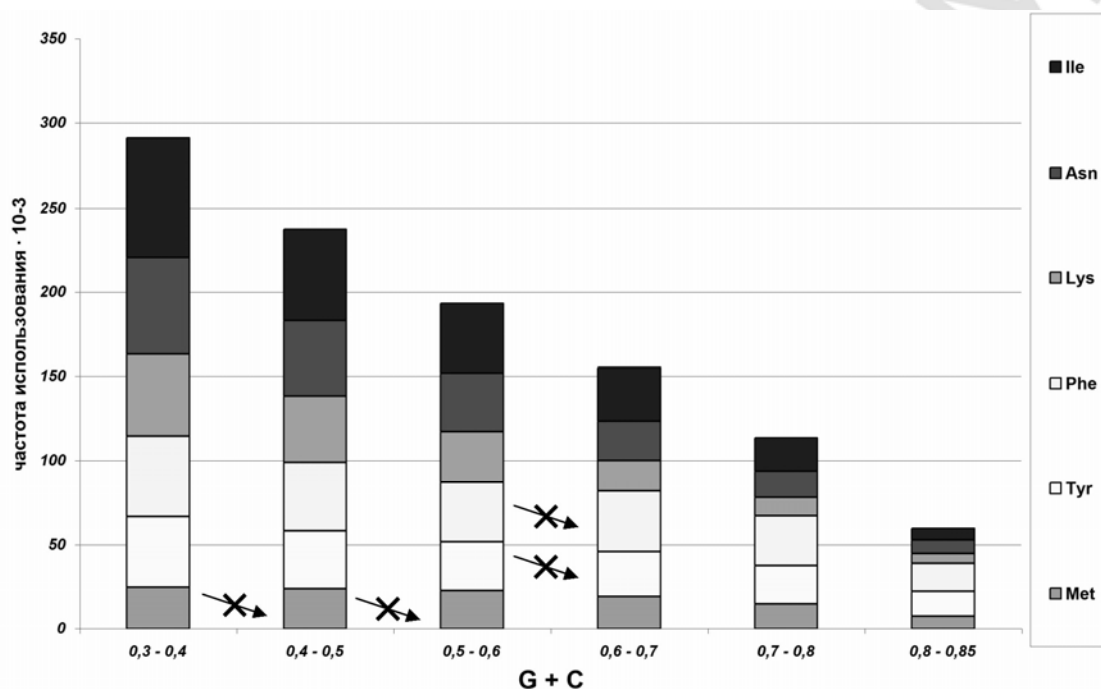


Рис. 44. Средний уровень использования аминокислот из группы FYMINK в белках альфа-герпесвирусов, кодируемых генами, разбитыми в группы по GC-насыщенности. Зачеркнутыми стрелками обозначено отсутствие статистически достоверной разности в средних уровнях использования аминокислоты в белках, кодируемых генами с различной GC-насыщенностью

На рис. 45 изображена аналогичная столбчатая диаграмма для частот использования аминокислот из группы GARP. В столбце для белков, кодируемых генами с  $0,4 < G+C < 0,5$ , частоты использования четырех аминокислот из группы GARP примерно равны друг другу. Однако, как видно на рис. 45, по мере увеличения GC-насыщенности генов неравенство частот использования глицина, аланина, аргинина и пролина усиливается.

По мере увеличения G+C в генах наиболее резко возрастает частота использования аланина в кодируемых ими белках [83, 85]. Прирост частот использования пролина, глицина и аргинина уступает приросту частоты использования аланина. Можно сказать, что колебания последней вносят наибольший вклад в зависимость GARP от G+C. Частота использования пролина заметно превышает частоты использования глицина и аргинина,



занимая второе место среди GARP, только в белках, кодируемых генами с очень высокой GC-насыщенностью (более 80 %). Глядя на диаграмму, приведенную на рис. 45, можно сделать вывод о том, что замены 10AA на аланин являются гораздо более нейтральными, чем замены 10AA на пролин, глицин или аргинин.

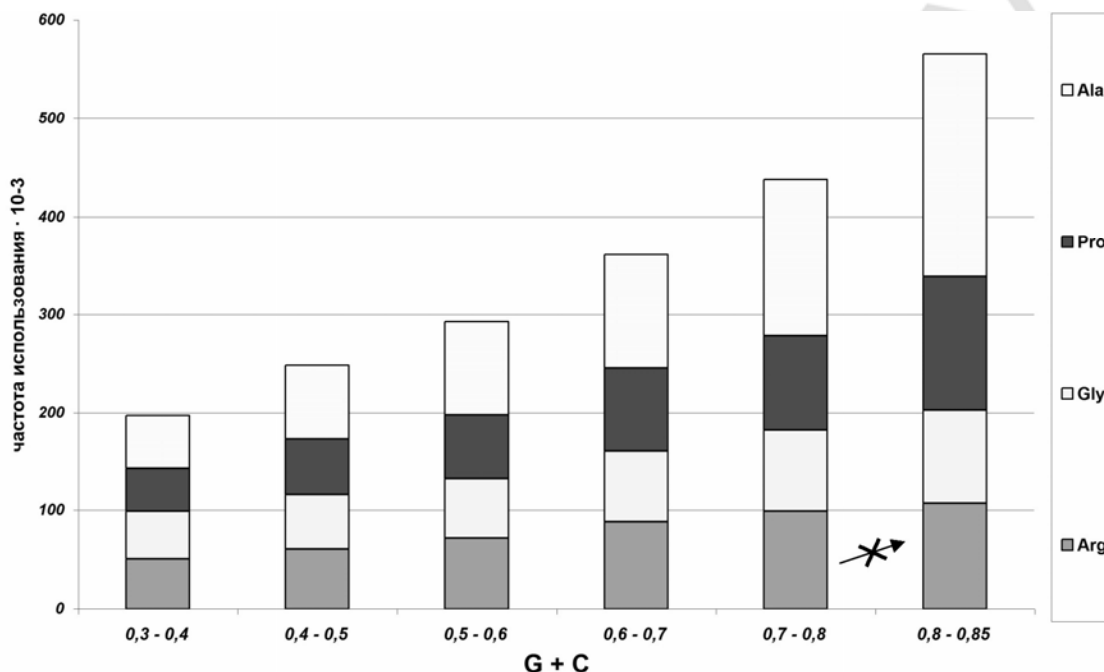


Рис. 45. Средний уровень использования аминокислот из группы GARP в белках альфа-герпесвирусов, кодируемых генами, разбитыми в группы по GC-насыщенности. Зачеркнутыми стрелками обозначено отсутствие статистически достоверной разности в средних уровнях использования аминокислоты в белках, кодируемых генами с различной GC-насыщенностью

Рис. 46 изображает динамику изменения частот использования десяти аминокислотных остатков, кодируемых средними по GC-насыщенности кодонами, в белках альфа-герпесвирусов от G+C соответствующих кодирующих участков. Частоты использования этих десяти аминокислотных остатков демонстрируют разные типы зависимости от G+C. Их можно условно разделить на три типа.

Частоты использования серина, треонина, глутамина и гистидина начинают снижаться под воздействием мутационного GC-давления уже в белках, кодируемых генами с  $G+C > 0,6$  [83]. «Раннее» снижение частот их использования заставляет предположить, что определенные типы замен этих аминокислот на GARP являются более нейтральными, чем замены того же направления остальных шести аминокислотных остатков из группы 10AA.

Частоты использования цистеина, лейцина и валина начинают снижаться только в белках, кодируемых генами с  $G+C > 0,8$  [83]. Это заставляет признать, что замены данных аминокислот на GARP в белках альфа-

герпесвирусов являются более радикальными, чем замены на GARP серина, треонина, глутамина и гистидина.

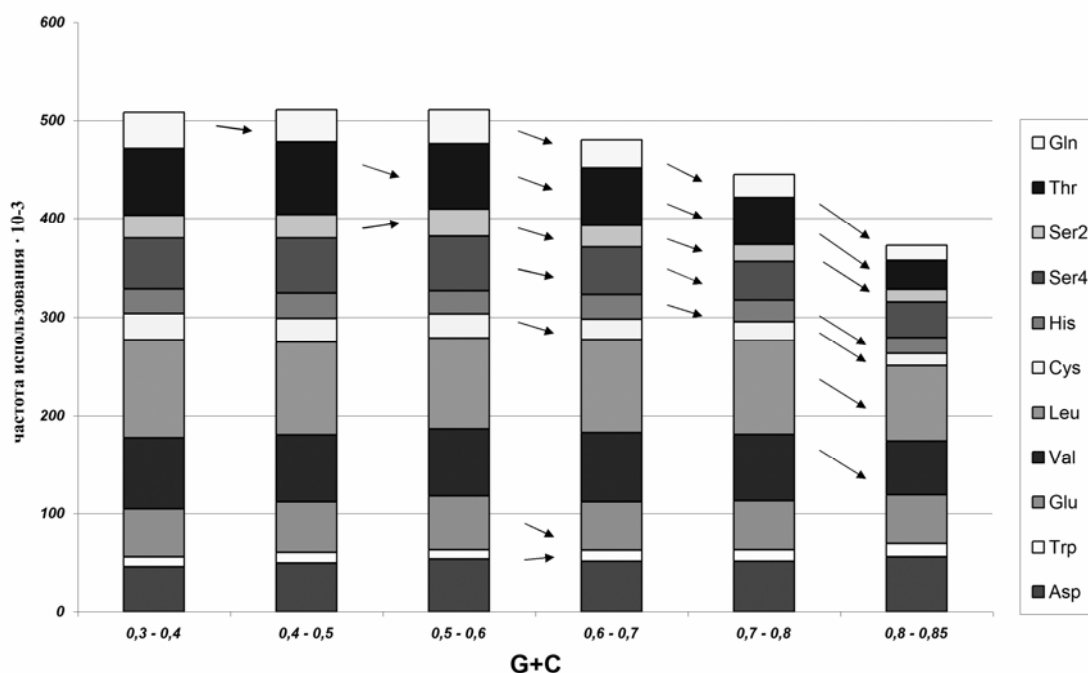


Рис. 46. Средний уровень использования аминокислот из группы 10AA в белках альфа-герпесвирусов, кодируемых генами, разбитыми в группы по GC-насыщенности. Стрелками обозначено наличие статистически достоверной разности в средних уровнях использования аминокислоты в белках, кодируемых генами с различной GC-насыщенностью [83]

Частоты использования триптофана, глутаминовой и аспарагиновой кислот в белках альфа-герпесвирусов вообще не изменяются с ростом G+C в кодирующих их генах [83]. Эти данные свидетельствуют о том, что замены триптофана, глутаминовой и аспарагиновой кислоты на GARP являются наиболее радикальными, в связи с чем они практически не фиксируются.

Хочется подчеркнуть, что приведенные на рис. 44, 45 и 46 данные — обобщение зависимостей частот использования аминокислотных остатков от G+C для всех белков альфа-герпесвирусов. В каждом конкретном белке или даже в отдельной его части могут существовать «локальные» особенности степеней нейтральности и радикальности аминокислотных замен, происходящих под воздействием GC-давления. О том, как мы предлагаем изучать эти особенности, подробно написано в гл. 1 и 3 четвертой части.

Чтобы выяснить, за счет каких именно аминокислотных замен снижается содержание серина, треонина, глутамина и гистидина в белках альфа-герпесвирусов по мере роста G+C в кодирующих их генах, мы провели следующее исследование. С помощью алгоритма «CodonChanges» мы подсчитали процентное соотношение замен в кодонах серина, треонина, глутамина и гистидина по направлению от генов ВПГ1 к генам ОВГ16. Мы

использовали две выборки кодирующих участков: (i) шесть участков, кодирующих капсидные белки, и (ii) двенадцать участков, кодирующих гликопротеины этих вирусов [83].

Капсидные белки и гликопротеины были выбраны по той причине, что именно к этим белкам синтезируется большая часть антител [89]. Антитела, как известно, вырабатываются к участкам белков, находящимся на поверхности белковой глобулы, то есть к пространственным эпитопам [90]. Примерно 10 % пространственных эпитопов содержат линейный эпитоп (участок последовательности аминокислотных остатков), остальные 90 % состоят из аминокислотных остатков, которые оказываются рядом друг с другом только благодаря формированию третичной структуры белка [90]. Тем не менее причины того, что аминокислотный остаток или последовательность аминокислотных остатков оказываются на поверхности белковой глобулы, должны быть общими как для линейных, так и для пространственных эпитопов [74, 75]. На основании данных по предсказанию появления новых линейных эпитопов или удлинения ранее существовавших в двух длинных безэпитопных участках гликопротеина В ВПГ1, мы можем привести следующие выкладки [87].

Наиболее часто возникает новый или удлиняется ранее существовавший В-клеточный линейный эпитоп в результате замен 10AA на пролин: в 92 % — при замене Ser4 на Pro, в 86 % — при замене His на Pro, в 78 % — при замене Leu4 на Pro, в 70 % — при замене Thr на Pro и в 60 % — при замене Gln на Pro [87].

Среди остальных замен 10AA на GARP с наибольшей вероятностью приводят к образованию нового линейного эпитопа или к удлинению существовавшего ранее замены Cys на Arg4 (100 %), Cys на Gly (66,6 %), Val на Gly (55 %), Leu4 на Arg4 (39 %) и Val на Ala (23 %) [87].

Возвращаясь к рис. 46, хочется отметить, что замены валина, лейцина и цистеина на GARP начинают фиксироваться только в белках, кодируемых генами с  $G+C > 0,8$ . Радикальность этих замен должна быть обусловлена разницей в физико-химических свойствах боковых цепей этих аминокислотных остатков [2, 74, 75]. Чаще всего функция белка страдает от таких замен, в том числе и из-за изменений его третичной структуры. Последнее как раз и ведет к появлению новых линейных и пространственных эпитопов. То есть замены валина, лейцина и цистеина на GARP, приводящие к появлению новых эпитопов, должны происходить довольно часто, но фиксироваться редко.

О фиксации остальных «эпитопогенных» аминокислотных замен в капсидных белках и гликопротеинах можно судить по данным из табл. 22 и 23 соответственно.

Как видно из табл. 22 и 23, «эпитопогенные» замены серина, треонина, глутамина и гистидина на пролин действительно фиксируются, но

с меньшей вероятностью, чем замены серина и треонина на аланин и глутамина и гистидина на аргинин.

Таблица 22

**Аминокислотные замены в шести капсидных белках по направлению от белков ВПГ1 к белкам ОБГ16 [83]**

АК, серия кодонов	Число АК и кодонов данной серии в белках HSV1	Неизменные и мутировавшие синонимично кодоны, %	Ala, %	Pro, %	Arg, %	Gly, %	Другие, %
Ser4 (TCX)	97	47,4	22,7 (T to G)	10,3 (T to C)	3,1	0,0	15,5 (6 AA + гэпы)
Thr (ACX)	185	60,5	17,3 (A to G)	4,3 (A to C)	1,6	2,2	14,1 (10 AA + гэпы)
Gln (CAA/G)	128	63,3	3,9	2,3 (A to C)	12,5 (A to G)	1,6	16,4 (8 AA + гэпы)
His (CAT/C)	98	73,5	2,0	5,1 (A to C)	6,1 (A to G)	2,0	11,3 (6 AA + гэпы)
Ser2 (AGT/C)	51	54,9	13,7	0,0	2,0 (A to C)	9,8 (A to G)	19,6 (4 AA + гэпы)

Таблица 23

**Аминокислотные замены в двенадцати гликопротеинах по направлению от белков ВПГ1 к белкам ОБГ16 [83]**

АК, серия кодонов	Число АК и кодонов данной серии в белках HSV1	Неизменные и мутировавшие синонимично кодоны, %	Ala, %	Pro, %	Arg, %	Gly, %	Другие, %
Ser4 (TCX)	128	37,5	17,2 (T to G)	10,9 (T to C)	4,7	5,5	24,2 (10 AA + гэпы)
Thr (ACX)	223	41,3	10,8 (A to G)	5,4 (A to C)	4,9	6,7	30,9 (8 AA + гэпы)
Gln (CAA/G)	86	36,0	7,0	9,3 (A to C)	20,9 (A to G)	3,5	23,3 (6 AA + гэпы)
His (CAT/C)	69	47,8	5,8	7,2 (A to C)	17,4 (A to G)	2,9	18,9 (6 AA)
Ser2 (AGT/C)	57	43,9	10,5	10,5	5,3 (A to C)	8,8 (A to G)	21,0 (8 AA + гэпы)

Если какая-то часть «эпитопогенных» замен была зафиксирована в популяции ОВГ16, значит, они не привели к существенному ухудшению функции белков. Это говорит о том, что есть вероятность возникновения нового эпитопа в белке ВПГ1 путем аминокислотной замены (или нескольких замен), не являющейся летальной для вируса. Эти данные позволяют пролить свет на патогенез возникновения рецидивов герпетической инфекции (см. ниже).

Сразу необходимо отметить, что процент мутировавших аминокислотных остатков выше в гликопротеинах (табл. 23), чем в капсидных белках (табл. 21). Действительно, протективными свойствами в большей степени обладают антитела к относительно более консервативным капсидным белкам ВПГ, чем к его более изменчивым гликопротеинам [89].

Как в гликопротеинах, так и в капсидных белках серин и треонин чаще всего заменяются на аланин. Объяснить относительную нейтральность замен Ser и Thr на Ala можно руководствуясь знаниями о структуре их боковых цепей (радикалов) [2]. Боковая цепь аланина представляет собой группу  $-CH_3$ . Благодаря такому радикалу аланин может находиться как в гидрофобных, так и в гидрофильных участках белка, внутри и снаружи белковой глобулы [2]. Пролин обладает весьма специфическим радикалом, глицин же вообще лишен боковой цепи. Как пролин, так и глицин являются сильными разрушителями альфа-спирали, в отличие от аланина, который является сильным ее формирователем [58]. Аргинин обладает длинным положительно заряженным радикалом. С точки зрения структурного сходства, Ser и Thr, обладающие короткими полярными радикалами с  $-OH$  группой на конце, чаще могут «находить себе эквивалентную замену» в виде аланина, чем в виде глицина, пролина и аргинина.

Как в гликопротеинах, так и в капсидных белках глутамин и гистидин чаще всего заменяются на аргинин. Хочется подчеркнуть, что глутамин и гистидин могут замениться путем единичной нуклеотидной замены по направлению AT на GC то ли на аргинин, то ли на пролин. Чаще нейтральными оказываются замены на аргинин. Боковая цепь гистидина, так же как и боковая цепь аргинина, несет положительный заряд.

Если обобщить приведенные данные, можно заключить, что наиболее нейтральными среди замен 10AA на GARP в белках альфа-герпесвирусов являются замены серина и треонина на аланин, глутамина и гистидина на аргинин [83].

Приведенные данные о степенях нейтральности аминокислотных замен (о «разрешенности» отрицательным отбором аминокислотных замен) в белках альфа-герпесвирусов претендуют на универсальность (на то, что нами были вскрыты определенные биохимические факторы, характерные и для белков других организмов) [83]. Однако не следует забывать о том, что в белках других организмов, кодируемых средними

по GC-насыщенности генами, аминокислотный состав может быть несколько иным [85]. Различия в количестве субстрата для замен по направлению мутационного давления могут оказывать существенное влияние на характер зависимости частот использования аминокислотных остатков от G+C [85].

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью описанного метода можно определить среднюю степень нейтральности каждой отдельной аминокислотной замены, происходящей под воздействием мутационного давления, для полного набора белков родственных вирусов или бактерий. С помощью нашего метода можно выяснить, фиксируются ли аминокислотные замены, приводящие к появлению новых эпитопов (при GC-давлении). Данные о нейтральности определенных типов аминокислотных замен можно использовать в биотехнологии и в иммунологии. Разумным кажется сравнивать нейтральность замен в каждом отдельном белке с их нейтральностью в среднем по протеому, что мы и сделали в гл. 3 четвертой части.

Данные, приведенные нами в этой главе, расширяют представления о патогенезе герпетической инфекции. В экспериментальных исследованиях было установлено, что у 98 % здоровых взрослых людей без всяких признаков заболеваний, причина которых — вирус простого герпеса (он способен вызывать не только привычную везикулярную сыпь на губах, но и кератит, стоматит, энцефалит), в слезах и/или слюне обнаруживается ДНК ВПГ1 [78]. Как известно, геном ВПГ1 пожизненно находится в латентном состоянии в нейронах ганглиев черепно-мозговых нервов после первого контакта организма человека с этим вирусом [19]. Периодически происходит реактивация ВПГ1: начинается синтез белков и репликация, вирусные частицы транспортируются по аксонам и попадают в иннервируемые ими ткани [19]. Традиционно считается, что реактивации ВПГ1 способствует стресс, вызывающий снижение иммунитета [19]. Как показали патологоанатомические исследования, вокруг нейронов, в которых ВПГ1 находится в латенции, существуют скопления макрофагов и Т-лимфоцитов [81].

После того как реактивировавшийся вирус попал в ткань, на него начинает воздействовать не только клеточный, но и гуморальный иммунитет. Если имеющиеся в крови у 90 % человеческой популяции антитела к ВПГ1 распознают эпитопы вируса, инфекционный процесс должен будет протекать abortивно, без клинических проявлений [78, 81]. При таком асимптомном процессе, как показали описанные выше исследования, вирус все-таки может выделяться из организма, но в малых количествах [78]. Именно такое асимптомное выделение вируса, по мнению специалистов, является причиной заражения не болевших ранее герпетической инфекцией людей (чаще всего — детей) [78].

Инфекционный процесс с клиническими проявлениями, по нашему мнению, развивается в ответ на появление новых эпитопов в белках реактивировавшегося ВПГ [81]. По данным экспериментальных исследований, во время рецидива герпетической инфекции появляется титр иммуноглобулина М против вирусных белков [89]. Как известно, иммуноглобулин М синтезируется при первой встрече В-лимфоцитов с антигеном, но не синтезируется при активации клеток памяти с уже готовыми генами иммуноглобулина G [26].

Важным обстоятельством является то, что появление нового эпитопа в вирусном белке может приводить к ухудшению доступности старых эпитопов (за счет перестроек третичной структуры). В старые эпитопы могут включаться новые аминокислотные остатки, увеличивая их длину, и тем самым нарушая аффинность ранее синтезированных к ним антител. Последние два механизма приводят к уходу ВПГ1 от иммунного ответа [81].

Частота появления новых или изменения старых эпитопов под влиянием мутационного давления должна быть гораздо более высокой в белках ВПГ1 и ВПГ2, чем в белках ВЗВ [81]. Рецидив инфекции, вызываемой ВЗВ, проявляется в виде опоясывающего лишая. Опоясывающим лишаем преимущественно заболевают лица пожилого возраста. Частота заболевания варьирует от 5 до 10 на 1000 человек в возрасте 60–80 лет. Лишь у немногих больных (около 2 % среди больных с нормальным иммунитетом и у 10 % больных с иммунодефицитами) опоясывающий лишай появляется повторно [19, 25]. Первичное же инфицирование варицелло-зостер-вирусом (ВЗВ) проявляется возникновением заболевания, известного как *ветряная оспа (варицелла)*, заболевают которым, как правило, еще в детском возрасте [25]. После стихания острых проявлений первичной инфекции этот вирус накапливается в задних корешках спинного мозга, спинномозговых или черепно-мозговых ганглиях, где сохраняется в латентном состоянии [19, 25]. Клинически выраженные рецидивы герпетической инфекции (вызванной ВПГ1) в среднем происходят 1–2 раза в год [19].

С помощью описанного в данной главе метода можно не только решать теоретические вопросы молекулярной биологии и иммунологии, но и углубляться в изучение патогенеза многих вирусных и бактериальных заболеваний.

### Глава 3

#### Анализ механизмов упрощения аминокислотного состава отдельных белков альфа-герпесвирусов под воздействием GC-давления

**Цель исследования:** определить степень и характер селекционных ограничений, накладываемых отрицательным отбором на аминокислотный состав отдельных белков альфа-герпесвирусов: на большой капсидный белок (UL19); белок, связывающий одноцепочечную ДНК (UL29); гликопротеин I (US7); регулятор транскрипции (ICP4) и на дезоксиуридинтрифосфатазу (UL50).

**Область применения:** данный тип молекулярно-биологического анализа может быть применен в случае необходимости определить степень и характер селекционных ограничений, накладываемых отрицательным отбором на различные типы аминокислотных замен в отдельном белке, в сравнении с ограничениями, накладываемыми в среднем на весь протеом.

**Используемые оригинальные алгоритмы:** «VVK in group» (см. гл. 3 шестой части).

**Материалы.** В данной главе была определена степень ограничений, накладываемых отрицательным отбором на фиксацию замен, приводящих к появлению аминокислот из группы GARP, в пяти белках альфа-герпесвирусов: в большом капсидном белке, в белке, связывающем одноцепочечную ДНК, в гликопротеине I, в дезоксиуридинтрифосфатазе и в регуляторе транскрипции ICP4. Исследованию генов, кодирующих большой капсидный протеин, посвящена гл. 1 второй части. Как этот ген, так и тот, который кодирует гликопротеин I, входят в выборки капсидных белков и гликопротеинов (соответственно), использованных в предыдущей главе.

Отдельного описания требует ген ICP4. Продукт этого гена обычно называют «положительным и отрицательным регулятором транскрипции». Действительно, многие экспериментальные данные подтверждают участие ICP4 в регуляции транскрипции с геномов альфа-герпесвирусов, что особенно важно в период латенции, во время перехода в это состояние и выхода из него (реактивации) [92]. Некоторые исследования указывают на то, что функциональный продукт гена ICP4 необходим для циркуляризации генома альфа-герпесвирусов (в вирионе геном альфа-герпесвирусов представлен линейной ДНК, в период латенции — замкнутой) [92]. ICP4 уникален тем, что является единственным геном, который присутствует в области LTR (больших инвертированных повторах) у всех альфа-герпесвирусов [84]. Как было продемонстрировано в гл. 2 четвертой части, гены, находящиеся в LTR, подвержены GC-давлению, преимущественно вызываемому трансверсиями AT на GC. Уровень GC-насыщенности генов из LTR всегда выше, чем таковой в генах из уникальных областей генома (UL и US) [84].



Дезоксиуридинтрифосфатаза — это фермент, который отщепляет три остатка фосфорной кислоты от дезоксиуридинтрифосфата (дУТФ) [23]. В результате работы фермента субстрат этой реакции — дезоксиуридинтрифосфат — лишается возможности быть встроенным в ДНК, а продукт реакции — дезоксиуридин — включается в цикл реакций внутриклеточного синтеза нуклеотидов. Ген, кодирующий дезоксиуридинтрифосфатазу, некогда попал в геном общего предшественника герпесвирусов путем ретропозиции [23]. В вирусный геном встроилась ДНК-копия процессированной клеточной мРНК, кодирующей этот фермент. В каждой клетке присутствует собственная дезоксиуридинтрифосфатаза, так что вирусная копия фермента во многом дублирует ее функции. Точно таким же путем в геном общего предшественника герпесвирусов попал и ген, кодирующий урацил-ДНК-гликозилазу. Этот фермент вырезает урацил из ДНК, причем предпочтительно из одноцепочечной [47]. Получается, что одна из функций вирусной дезоксиуридинтрифосфатазы (снижение количества урацила в ДНК) дублируется вирусной же урацил-ДНК-гликозилазой.

Белок, связывающий одноцепочечную ДНК (UL29), тоже попал в вирусный геном из генома хозяина. Функция этого белка заключается в связывании одноцепочечной ДНК во время репликации: таким образом данный белок препятствует воссоединению комплементарных цепей и соединению инвертированных повторов на одной и той же цепи [23].

**Номера доступа (идентификаторы) в GenBank.** Исследованные последовательности можно найти в GenBank в записях о нуклеотидном составе полных геномов референтных штаммов альфа-герпесвирусов: Human herpesvirus 1 (ВПГ 1) [NC\_001806], Human herpesvirus 2 (ВПГ 2) [NC\_001798], Cercopithecine herpesvirus 16 (обезьяний герпесвирус 16) [NC\_007653], Human herpesvirus 3 (человеческий герпесвирус 3, или варицелло-зостер-вирус) [NC\_001348], Cercopithecine herpesvirus 9 (обезьяний герпесвирус 9) [NC\_002686].

Для обозначения генов альфа-герпесвирусов часто используется их локализация в уникальных областях генома у вируса герпеса 1-го типа. Например, ген, кодирующий большой капсидный белок (БКБ), расположен девятнадцатым в уникальной длинной области генома ВПГ1, а ген, кодирующий дезоксиуридинтрифосфатазу, — пятидесятым. Значит, ген, кодирующий БКБ можно сокращенно называть геном UL19, а ген, кодирующий дезоксиуридинтрифосфатазу, можно называть UL50. Ген, кодирующий гликопротеин I, расположен седьмым в уникальной короткой области генома ВПГ1. Название этого гена — US7.

#### **Описание метода.**

**Предварительные операции.** Предварительного выравнивания не требуется, поскольку «VVK in group» рассчитывает все показатели для

каждой введенной последовательности отдельно (общее количество таких последовательностей должно быть не больше пятидесяти).

**Производимые вычисления.** Для того чтобы алгоритм «VVK in group» произвел все необходимые вычисления, следует скопировать набор нуклеотидных последовательностей в соответствующие ячейки на листе «sequences». В столбцах на листе «results» появятся все вычисляемые показатели. MS Excel позволяет не только построить график, изображающий зависимость показателей друг от друга, но и получить уравнение, описывающее линейную зависимость между показателями. Уравнение это имеет общий вид  $y = kx + b$ . Постоянная величина (коэффициент)  $k$  отражает степень наклона линии тренда: чем больше  $k$ , тем круче наклон. Кроме уравнения, описывающего линейную зависимость, на график, построенный в MS Excel, можно вывести значение  $R^2$ . Величина  $R^2$  описывается в MS Excel как «показатель, характеризующий аппроксимацию зависимости к прямой линии». Чем больше  $R^2$  (он может принимать значения от 0 до 1), тем ближе зависимость к прямой. Несложно догадаться, что  $R^2$  является квадратом коэффициента корреляции. Если  $R^2 > 0,25$ , то  $R > 0,5$ , следовательно, можно говорить о сильной или средней силы зависимости. Если же  $R^2 < 0,25$ , то  $R < 0,5$ , следовательно, зависимость является слабой или вообще отсутствует. В последнем случае использовать значение коэффициента  $k$  (наклона) для характеристики зависимости не имеет смысла, поскольку можно серьезно усомниться в существовании самой зависимости.

**Определение направления мутационного давления в генах из филогенетического ряда.** О методе определения направления мутационного давления в отдельных генах подробно написано в гл. 1 второй части. О направлении, ведущих биохимических механизмах и силе мутационного давления в кодирующих геномах альфа-герпесвирусов написано в гл. 2 второй части. Еще раз хочется подчеркнуть, что ген ICP4 является более GC-насыщенным, чем остальные четыре (UL19, UL29, UL50 и US7), изученные в данной главе [84]. Это значит, что в геномах пяти симплексовирусов и в геноме герпесвируса крупного рогатого скота 5 ген ICP4 подвержен еще более сильному GC-давлению, чем гены UL19, UL29, UL50 и US7 [84].

**Определение показателя нейтральности эволюции белка по методу Н. Суеки.** Метод определения показателя нейтральности эволюции белка по Н. Суеке был описан в гл. 1 четвертой части. В данном исследовании в качестве показателя нейтральности по Н. Суеке были использованы значения наклона (коэффициента  $k$ ) зависимости « $(1GC+2GC)/2$ » от « $3GC$ ».

**Неравномерность увеличения частот использования аминокислотных остатков из группы GARP в белках под воздействием мутационного GC-давления.** Как известно, GC-богатыми (в первых и вторых положениях) кодонами в универсальном генетическом коде кодируются только четыре аминокислотных остатка (GARP: Gly, Ala, Arg и Pro) (см.

прил. 1) [116]. Частоты их использования должны расти в белках, кодируемых генами, подверженными GC-давлению, за счет замен AT на GC [121]. Рост частот использования этих четырех аминокислотных остатков обусловлен снижением частот использования остальных шестнадцати. О том, какие именно замены, приводящие к появлению аминокислот из группы GARP, являются наиболее нейтральными, написано в гл. 2 четвертой части. Но эти данные были получены в среднем по протеомам альфа-герпесвирусов [83]. Логично предположить, что в каждом конкретном белке степени нейтральности таких замен могут отличаться от «среднепротеомных».

На рис. 47 представлен график зависимости частот использования Gly, Ala, Arg и Pro в большом капсидном белке десяти альфа-герпесвирусов от уровня GC-насыщенности кодирующего его гена.

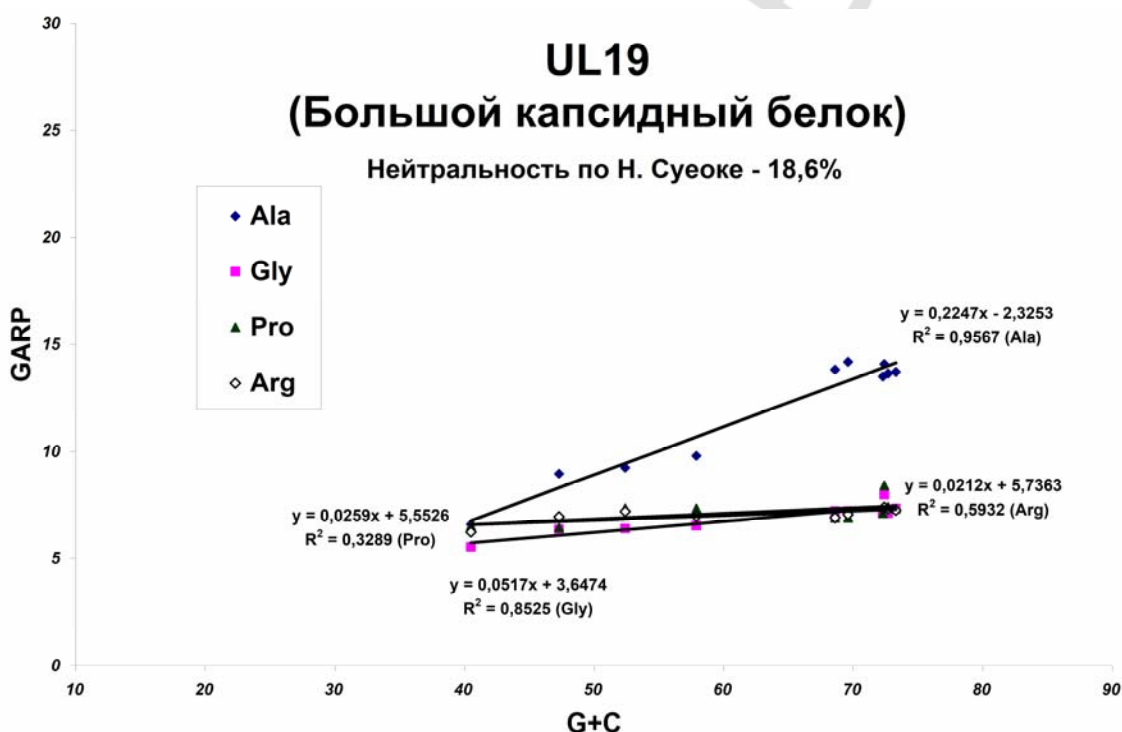


Рис. 47. Зависимость частот использования аминокислотных остатков из группы GARP в большом капсидном белке десяти альфа-герпесвирусов от GC-насыщенности кодирующих его генов

Как видно на рис. 47, ограничения, накладываемые отрицательным отбором на закрепление аминокислотных замен в большом капсидном белке, очень близки к «среднепротеомным». Частоты использования всех четырех аминокислотных остатков из группы GARP находятся в прямой линейной зависимости (сильной или средней силы) от G+C. Но это еще не означает, что нейтральность всех замен, приводящих к их появлению, одинаково высока. Несмотря на прямую зависимость от G+C, углы наклона для Pro, Arg и Gly очень низкие: коэффициент  $k$  не превышает 0,052

(рис. 47). Это означает, что замены 10AA на Pro, Arg и Gly, происходящие в результате GC-давления, фиксируются в большом капсидном белке, но очень редко.

На рис. 47 выделяется зависимость Ala от G+C. Действительно, угол наклона этой зависимости ( $k = 0,22$ ) гораздо больше, чем таковые для Pro, Arg и Gly. Это говорит о том, что наиболее часто в большом капсидном белке фиксируются замены 10AA на аланин. Действительно, аланин является «наиболее нейтральной» аминокислотой из группы GARP, по крайней мере, в белках альфа-герпесвирусов. Появление аланина на месте аминокислот из группы 10AA должно гораздо реже приводить к изменению или к потере функции белка, чем появление пролина, глицина или аргинина.

На рис. 48 представлен график зависимости частот использования аминокислотных остатков из группы GARP от G+C в белке, связывающем одноцепочечную ДНК изученных альфа-герпесвирусов. Как и на рис. 47, наибольшим углом наклона обладает зависимость частоты использования аланина от G+C ( $k = 0,27$ ). Частоты использования остальных аминокислот из группы GARP демонстрируют гораздо меньшие наклоны зависимости от G+C. Нельзя не обратить внимание на практически полное отсутствие роста частоты использования аргинина в белках, связывающих одноцепочечную ДНК, по мере увеличения GC-насыщенности соответствующих генов. Известно, что белки связываются с ДНК при помощи положительно заряженных аминокислотных остатков, таких как лизин и аргинин [23]. По всей вероятности, появление новых остатков аргинина в белке, связывающем одноцепочечную ДНК, существенно ухудшает его функцию: возможно, связывание с ДНК в таком случае начинает происходить не только в активном центре.

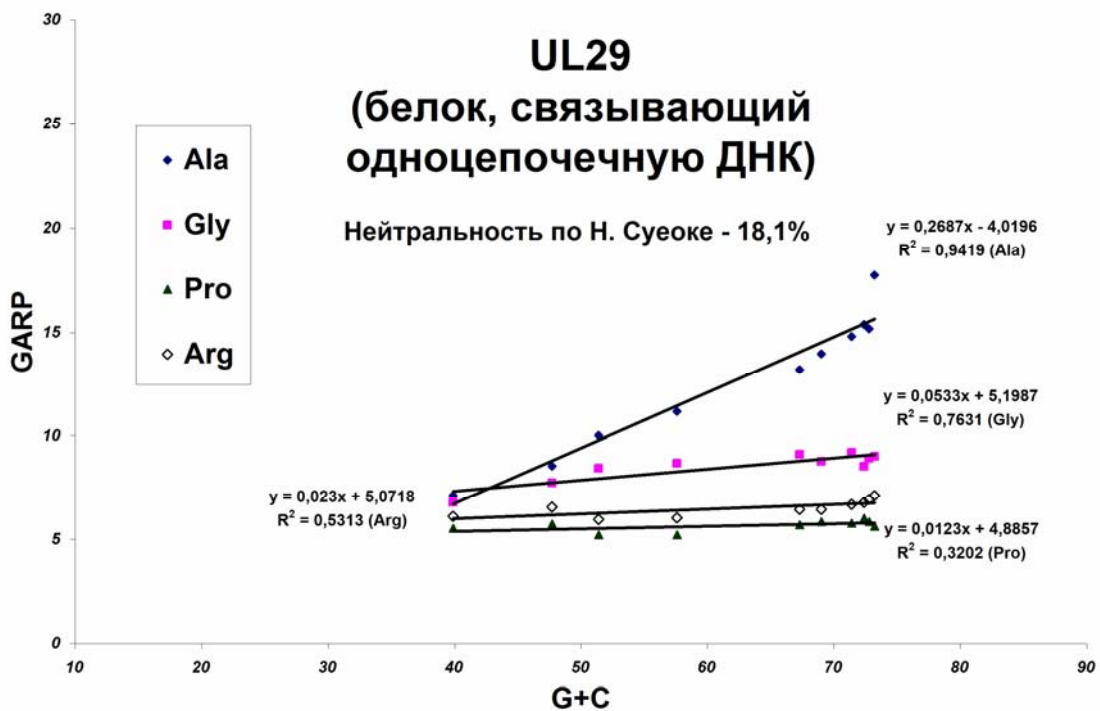


Рис. 48. Зависимость частот использования аминокислотных остатков из группы GARP в белке, связывающем одноцепочечную ДНК, у десяти альфа-герпесвирусов от GC-насыщенности кодирующих его генов

На рис. 49 представлен график зависимости частот использования аминокислотных остатков из группы GARP от G+C в гликопротеине I десяти альфа-герпесвирусов. Контраст этого графика (рис. 49) с графиками, изображенными на рис. 47 и 48 бросается в глаза. Уровень нейтральности аминокислотных замен по Н. Суеоке более чем в 2 раза выше для гликопротеина I по сравнению с большим капсидным белком.

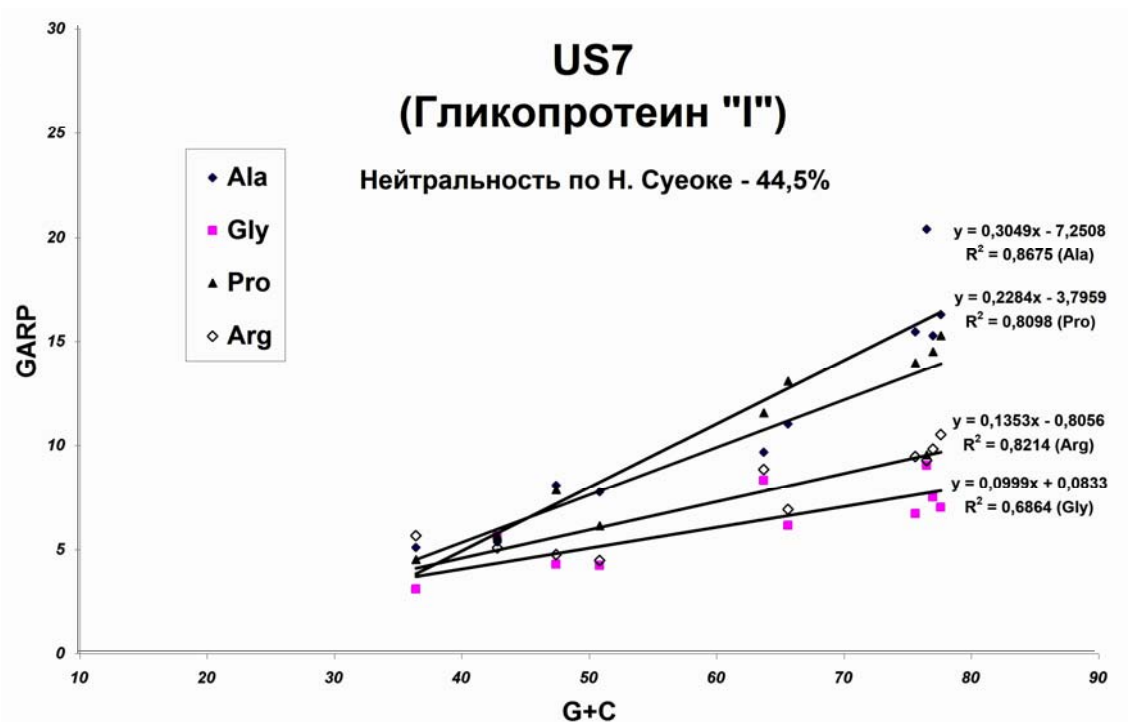


Рис. 49. Зависимость частот использования аминокислотных остатков из группы GARP в гликопротеине I десяти альфа-герпесвирусов от GC-насыщенности кодирующих его генов

Степень нейтральности аминокислотных замен 10AA на пролин приближается к таковой для замен 10AA на аланин в гликопротеине I. Углы наклона зависимостей Arg и Gly от G+C ниже, чем таковые для Pro и Ala (см. рис. 49), однако они значительно превышают аналогичные углы наклона для большого капсидного белка (см. рис. 47). В целом можно заключить, что замены 10AA на GARP в гликопротеине I фиксируются чаще, чем в большом капсидном белке. Кроме того, замены 10AA на пролин являются вторыми по вероятности фиксации (а значит, и по степени нейтральности) в гликопротеине I.

Замены, приводящие к появлению пролина, могут сопровождаться значительными перестройками третичной структуры белка и образованием новых линейных В-клеточных эпитопов [87]. Полученные нами данные свидетельствуют о том, что вероятность фиксации замены 10AA на пролин в гликопротеине I уступает лишь вероятности фиксации замены 10AA на аланин. Получить возможность быть зафиксированной может лишь та мутация, которая не снижает жизнеспособность вируса. Согласно полученным данным, гликопротеин I должен быть подвержен возникновению новых В-клеточных эпитопов (линейных и пространственных), не приводящих к значительному снижению жизнеспособности вируса, чаще, чем другие белки вирусов простого герпеса.

Рис. 50 представляет собой график зависимости частот использования аминокислотных остатков из группы GARP от G+C в дезоксиуридинтрифосфатазе десяти альфа-герпесвирусов.

Степень нейтральности аминокислотных замен по Н. Суеке практически одинаково велика как для гликопротеина I, так и для дезоксиуридинтрифосфатазы. Наш метод, однако, позволяет заключить, что существуют определенные различия между степенями нейтральности замен 10AA на GARP между гликопротеином I и дезоксиуридинтрифосфатазой.

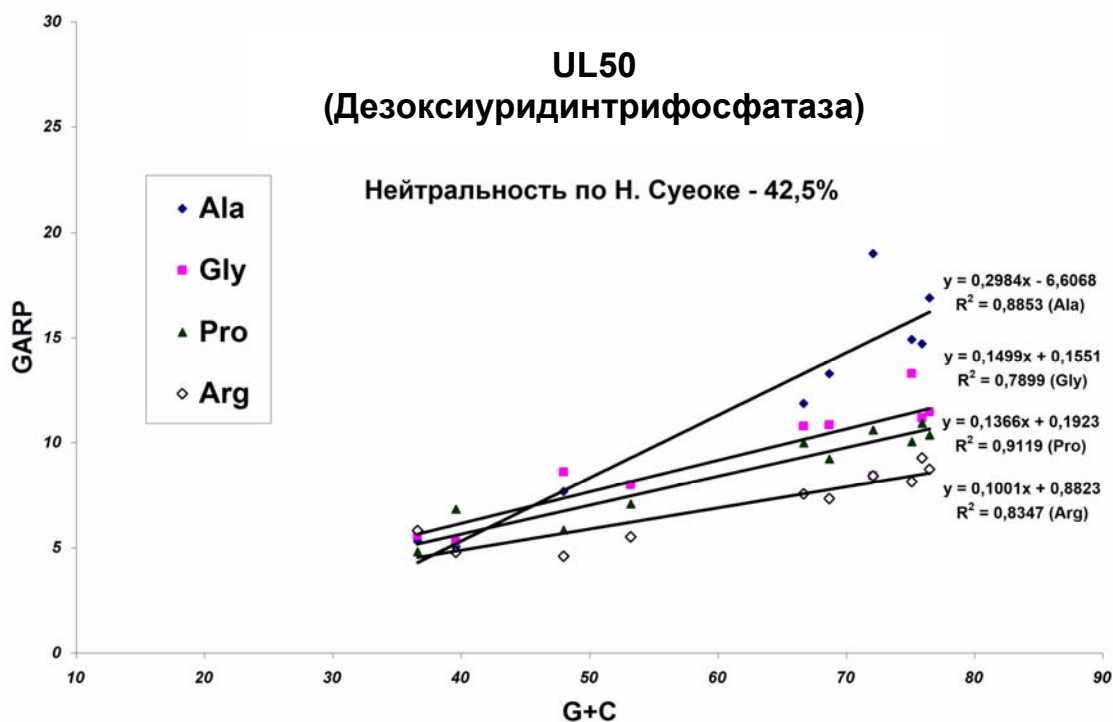


Рис. 50. Зависимость частот использования аминокислотных остатков из группы GARP в дезоксиуридинтрифосфатазе десяти альфа-герпесвирусов от GC-насыщенности кодирующих ее генов

В частности, уровни нейтральности замен 10AA на пролин, глицин и аргинин в дезоксиуридинтрифосфатазе близки друг к другу (при том, что наиболее нейтральными среди них являются замены 10AA на глицин). В гликопротеине I уровень нейтральности замен 10AA на пролин значительно превышает уровни нейтральности замен 10AA на глицин и аргинин. Нельзя не подчеркнуть, что на рис. 50, так же как и на трех предыдущих, наибольшим наклоном отличается зависимость частоты аланина от G+C.

С чем можно связать высокий уровень нейтральности аминокислотных замен, возникающих в результате GC-давления, в дезоксиуридинтрифосфатазе? По всей видимости, данный фермент не является жизненно необходимым для альфа-герпесвирусов, поскольку его функция дублируется клеточной дезоксиуридинтрифосфатазой. Как уже было сказано, снижение



содержания урацила в ДНК может быть достигнуто исключительно за счет работы урацил-ДНК-гликозилазы (как вирусной, так и клеточной) [47].

Интересно отметить, что уровень нейтральности аминокислотных замен по Н. Суеоке для жизненно необходимого белка ICP4 примерно равен таковому для дезоксиуридинтрифосфатазы (41,2 и 42,5 % соответственно). Только предложенный нами метод анализа дискретных уровней нейтральности аминокислотных замен способен помочь в разьяснении этого феномена.

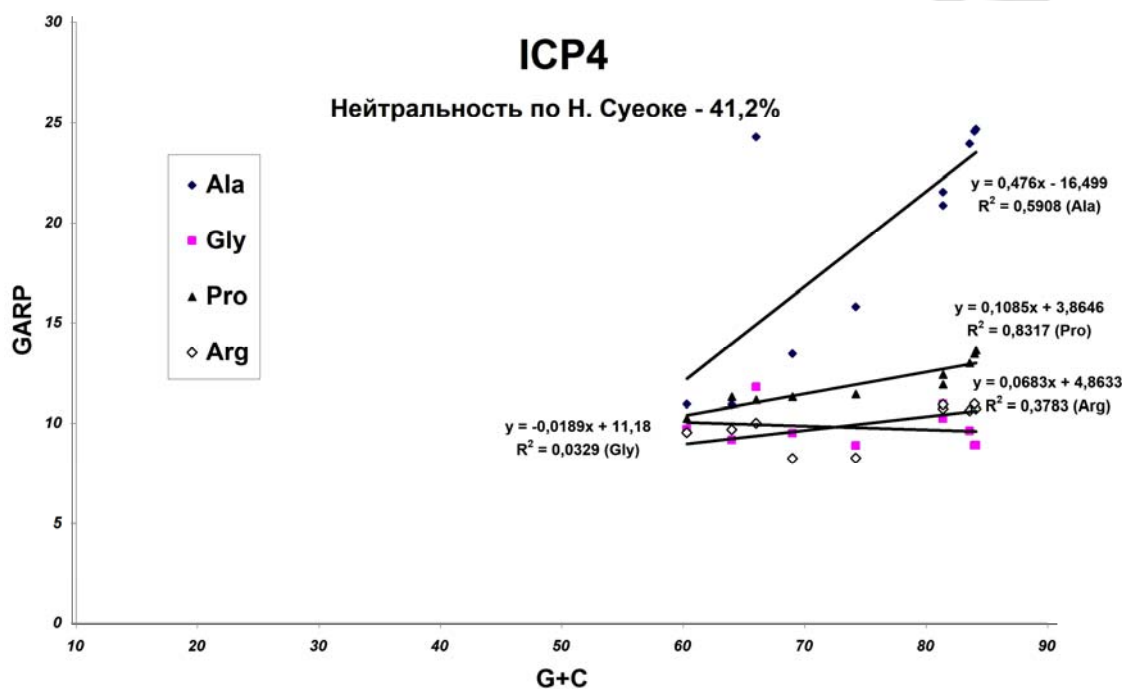


Рис. 51. Зависимость частот использования аминокислотных остатков из группы GARP в регуляторе транскрипции ICP4 десяти альфа-герпесвирусов от GC-насыщенности кодирующих его генов

На рис. 51 отчетливо видно, что частота использования аланина растет по мере увеличения G+C с наклоном, близким к 0,5, в то время как наклоны частот зависимостей остальных аминокислотных остатков от G+C значительно ниже. Зависимость частоты использования глицина от G+C вообще отсутствует. График на рис. 51 напоминает графики с рис. 47 и 48 с той лишь разницей, что угол наклона частоты использования аланина на двух последних более чем в два раза ниже.

Получается, что как в большом капсидном белке, так и в ICP4 чаще всего фиксируются замены 10AA на аланин, а вероятность фиксации замен 10AA на глицин, пролин и аргинин в них гораздо ниже. Это говорит о том, что большой капсидный белок и ICP4 подвержены гораздо более жесткому отрицательному отбору, чем гликопротеин I и дезоксиуридинтрифосфатаза. Так чем же обусловлен высокий уровень нейтральности аминокислотных замен по Н. Суеоке в ICP4?



Ген, кодирующий ICP4, расположен в инвертированных повторах, вследствие чего трансверсии АТ на GC в нем происходят с гораздо большей частотой, чем в генах, расположенных в уникальных областях геномов альфа-герпесвирусов [84]. Чем чаще происходят нейтральные замены, тем быстрее они должны фиксироваться [16, 121]. Как в большом капсидном белке, так и в ICP4 «разрешенными» являются замены 10АА на аланин. Судя по рис. 47 и 50, замены 10АА на аланин гораздо чаще фиксировались в ICP4, чем в большом капсидном белке. За счет закрепления в том числе и этих замен общая GC-насыщенность гена ICP4 у симплексовирусов достигла 80–85 %. Произошло это, опять же, за счет более высокой частоты *возникновения* нуклеотидных замен АТ на GC в LTR, по сравнению с областями UL и US [84].

В то же время, GC-насыщенность генов, кодирующих гликопротеин I и дезоксиуридинтрифосфатазу симплексовирусов, превышает GC-насыщенность гена, кодирующего большой капсидный белок тех же вирусов, за счет более частой *фиксации* замен АТ на GC при одинаковой частоте их *возникновения*. Отличить рост показателя нейтральности по Н. Суеоке за счет увеличения интенсивности биохимических процессов, вызывающих мутационное давление, от роста за счет увеличения нейтральности замен, позволяет описанный в данной главе метод.

Интересное следствие насыщения белков симплексовирусов аланином — появление «хрупкой спирали». Процент (и вероятность) альфа-спирализации этих белков должен повышаться по мере роста частоты использования аланина [83]. Еще две аминокислоты из группы GARP — глицин и пролин — являются сильными разрушителями альфа-спиралей [58]. То, что замены 10АА на пролин и глицин редко фиксируются, совсем не означает, что эти замены происходят редко. Если пролин или глицин появляется в альфа-спирализованном участке белка, этот элемент вторичной структуры, как правило, полностью или частично разрушается [58]. Такое изменение вторичной структуры белка может оказывать серьезное влияние на третичную структуру.

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью описанного метода можно определить степень нейтральности каждой отдельной аминокислотной замены, происходящей под воздействием мутационного давления в конкретном белке. Это, в частности, позволяет предсказать вероятность «безопасного» (для жизнеспособности вируса) возникновения новых В-клеточных эпитопов в данном белке. С помощью описанного метода можно предположить функциональную значимость того или иного мало исследованного белка.

**Часть 5**  
**Уровень изменчивости**  
**отдельных участков белка**

**Глава 1**

**Подтверждение результатов предсказания**  
**уровня мутабельности пространственных эпитопов поверхностного белка**  
**gp120 вируса иммунодефицита человека 1-го типа**

**Цель исследования:** определить уровень изменчивости предсказанных пространственных эпитопов поверхностного гликопротеина gp120 на основании анализа обширных выборок нуклеотидных последовательностей участков гена *env*, кодирующих их, от разных штаммов и клонов ВИЧ1.

**Область применения:** данный тип молекулярно-биологического анализа может быть применен для определения степени изменчивости (вариативности) пространственных эпитопов или любых других участков белка, в том случае, если исследователь располагает достаточным количеством нуклеотидных последовательностей, кодирующих изучаемый белок, полученных от разных штаммов и клонов вируса или бактерии.

**Используемые оригинальные алгоритмы:** «VVK Consensus» (см. гл. 5 шестой части).

**Материалы.** Для расчета частот нуклеотидных замен в участках, кодирующих 3D-эпитопы gp120 ВИЧ1, и для получения данных об изменчивости их нуклеотидного состава, мы использовали десять выборок нуклеотидных последовательностей гена *env*, полученных в результате секвенирования различных клонов ВИЧ1 у инфицированных пациентов (каждая выборка — от одного пациента, см. гл. 3 второй части). Эти последовательности были добавлены в GenBank в ходе 3 независимых исследований. Мы отобрали выборки с наибольшим количеством последовательностей.

**Номера доступа (идентификаторы) в GenBank.** Пять групп последовательностей (от пяти пациентов) были получены в ходе исследования изменчивости ВИЧ1 под воздействием гуморального иммунитета [57], их идентификаторы в GenBank: EU744175–EU743973. Две группы — в ходе работы по изучению тропизма вируса [66], их идентификаторы: EU604642–EU604549. Еще три группы генов мы выбрали среди таковых, полученных в ходе исследования мутагенеза ВИЧ1 у так называемых «элитных контроллеров» (DQ410040 to DQ410649 — номера всех генов, полученных в этом исследовании) [44].

**Описание метода.**

**Предварительные операции.** Перед тем как изучать изменчивость участков, кодирующих 3D-эпитопы, необходимо предсказать или непо-

средственно определить их границы в интересующем бактериальном или вирусном белке. Границы участков, кодирующих пространственные эпитопы белка gp120 ВИЧ1, были определены нами в гл. 2 третьей части при помощи программы «DiscoTope» [43] с использованием сведений о третичной структуре gp120 из базы данных PDB (Protein Data Bank).

Мы предлагаем определять степень изменчивости отдельных участков вирусного белка в монофилетических группах, то есть на выборках нуклеотидных последовательностей, полученных от одного пациента. Можно применять описанную ниже методику и для полифилетических групп. Однако при наличии обширных коллекций последовательностей, полученных в результате периодического секвенирования в течение заболевания, лучше прибегнуть к более чистому эксперименту.

Нуклеотидные последовательности из каждой выборки должны быть выровнены независимо друг от друга. Затем полученное выравнивание (по каждой из выборок) должно быть соотнесено с границами эпитопов в референтном штамме. После этого необходимо вырезать участки, гомологичные таковым, кодирующим 3D-эпитоп в референтном штамме (по каждому эпитопу).

В результате предварительных процедур в распоряжении должны будут появиться выборки участков, кодирующих каждый 3D-эпитоп у вирусов, полученных от каждого пациента.

Хочется напомнить, что выравнивание необходимо проводить по аминокислотным последовательностям (предварительно «транслировав» нуклеотидные последовательности в аминокислотные), которые затем следует переводить обратно в нуклеотидные.

Алгоритм «VVK Consensus» распознает символ «N» (латинскую букву «эн») как гэп. Если программа для выравнивания использует для этого другой символ, необходимо произвести замену. Автоматическими средствами, позволяющими быстро провести эту процедуру, оснащен, в частности, текстовый редактор MS Word.

**Производимые вычисления.** Алгоритм «VVK Consensus», как мы уже писали в гл. 3 второй части, производит подсчет количества всех типов нуклеотидных замен в выборке введенных в соответствующие ячейки на листе «sequences» последовательностей. Если в гл. 3 второй части мы оперировали процентным содержанием замен каждого типа, то в данном исследовании мы будем обращаться к общей частоте всех типов замен. В задачу исследования входит сравнение уровней изменчивости различных участков одних и тех же нуклеотидных последовательностей (внутри выборки). Длина этих участков не является постоянной (длины предсказанных 3D-эпитопов не равны друг другу). Поэтому, для того чтобы сравнить уровни их изменчивости, возникает необходимость отнести количество замен к длине последовательности. Величина, которая получается в ре-

зультате, называется *частотой замен*, измеряется она обычно в количестве замен на сайт за определенный промежуток времени. Поскольку промежуток времени между первым и последним секвенированием для всех участков из одной и той же выборки одинаков, в нашем случае частота измеряется в заменах на сайт.

Вычислить общую частоту замен можно путем деления общего количества замен в выборке нуклеотидных последовательностей на длину последовательностей (включая гэпы). При помощи «VVK Consensus» вычисляют также и процент замен в первых и вторых положениях кодонов. Такие замены, как правило, являются несинонимичными, в отличие от замен в третьих положениях. При умножении общей частоты замен на долю их в первых и во вторых положениях кодонов можно получить приблизительную частоту несинонимичных замен.

Количество, а значит, и частоту замен «VVK Consensus» подсчитывает по направлению от консенсусной последовательности двумя способами. Подсчет замен «*per site*» дает представление о том, каково количество вариантов нуклеотидных замен, встречающихся в данной выборке: подсчитывается количество сайтов с нуклеотидными заменами (см. формулу 3.1 из гл. 3 второй части). Подсчет «*per nucleotide*» дает представление еще и о том, насколько распространен в выборке полиморфизм по переменным нуклеотидным сайтам: подсчитывается количество мутировавших нуклеотидов во всех нуклеотидных последовательностях (см. формулу 3.2 из гл. 3 второй части).

Кроме вычисления частот замен, «VVK Consensus» рассчитывает нуклеотидный состав каждой из введенных в соответствующие ячейки на листе «sequences» последовательностей [86]. В данной главе мы оперировали средними по каждой выборке значениями частоты использования гуанина и цитозина в участках, кодирующих 3D-эпитоп. Кроме того, мы вычисляли средние частоты использования G и C в первых и вторых, а также в третьих положениях кодонов.

Результаты предсказания уровня мутабельности каждого из пяти длинных 3D-эпитопов в белке gp120 ВИЧ1 находятся в табл. 17 гл. 2 третьей части. По результатам этих предсказаний можно ожидать наименьшей изменчивости от первого пространственного эпитопа gp120.

Уровни предсказанной мутабельности **по гуанину** для участков, кодирующих 3D-эпитопы gp120 у референтного штамма ВИЧ1, можно расположить в следующем порядке: **C4-V5-C5 > V4 > V3 > C3 > C1**.

Уровни предсказанной мутабельности **по цитозину** можно расположить в следующем порядке: **V3 > V4 > C3 > C1 > C4-V5-C5** (указаны области gp120, в которых располагаются 5 длинных 3D-эпитопов).

**Определение уровня изменчивости пространственных эпитопов.** Под *изменчивостью* (*вариабельностью*) мы понимаем частоту нуклеотид-

ных замен в кодирующем участке (и аминокислотных замен в участке белка). Высокая вариабельность кодирующего участка может быть обусловлена двумя факторами: высокой вероятностью возникновения несинонимичных нуклеотидных замен (повышенной *мутабельностью*) и/или низкой функциональной значимостью данного участка белка (повышенной нейтральностью аминокислотных замен) [86]. Данный метод позволяет оценить вариабельность в целом. Метод, описанный в гл. 2 третьей части позволяет предсказать уровень мутабельности. О степени нейтральности аминокислотных замен могут дать представление результаты применения метода, описанного в гл. 1 четвертой части.

Сначала мы предлагаем оценить степень изменчивости нуклеотидного состава участков, кодирующих 3D-эпитопы, в разных выборках. Как известно, от нуклеотидного состава кодирующего участка зависит уровень мутабельности белка (или какой-либо его части) [86].

На рис. 52 изображены столбчатые диаграммы, отображающие средние значения частоты использования гуанина в каждом из участков, кодирующих 3D-эпитоп gr120, для каждой из десяти выборок. Кроме того, каждый столбец разделен на две части, отображающие процентное содержание гуанина в первых и вторых, а также в третьих положениях кодонов.

Наименьшее содержание G в первых и вторых положениях кодонов характерно для участков, кодирующих эпитоп из региона C1 (рис. 52, а). Кроме того, участки, кодирующие эпитоп из региона C1, обладают наибольшим уровнем 3G среди всех остальных участков, кодирующих 3D-эпитопы gr120.

В большинстве выборок уровень 3G в участках, кодирующих V3-регион, минимален (рис. 52, б), при том, что уровень 1G+2G в них превышает таковой в участках, кодирующих 3D-эпитопы из C1 и C3 (рис. 52, а, в).

Уровень насыщенности гуанином участков, кодирующих 3D-эпитопы из V4 и C4-V5-C5 (рис. 52, г, д), довольно сильно варьирует от выборки к выборке, причем как в третьих, так и в первых и во вторых положениях кодонов. И тем не менее наибольшей G-насыщенностью обладают участки, кодирующие 3D-эпитоп, расположенный в регионах C4-V5-C5.

Очень интересен тот факт, что участок, кодирующий V3-регион у одного из элитных контроллеров (ES8) полностью лишен защитного буфера в виде гуанина, расположенного в третьих положениях кодонов (рис. 52, б). Как же организму этого пациента удастся контролировать репликацию ВИЧ, если основная мишень для антител должна быть подвержена частым аминокислотным заменам? По нашему мнению, у данного пациента V3-регион как раз не является основной мишенью для иммунного ответа. Было доказано, что встречаются пациенты, у которых большая часть антител синтезируется против других участков gr120 [119]. Возможно, в том случае, если эти эпитопы менее изменчивы, чем V3-регион, пациент обре-

тает способность долгое время подавлять репликацию вируса и становится бессимптомным носителем ВИЧ (элитным контроллером).

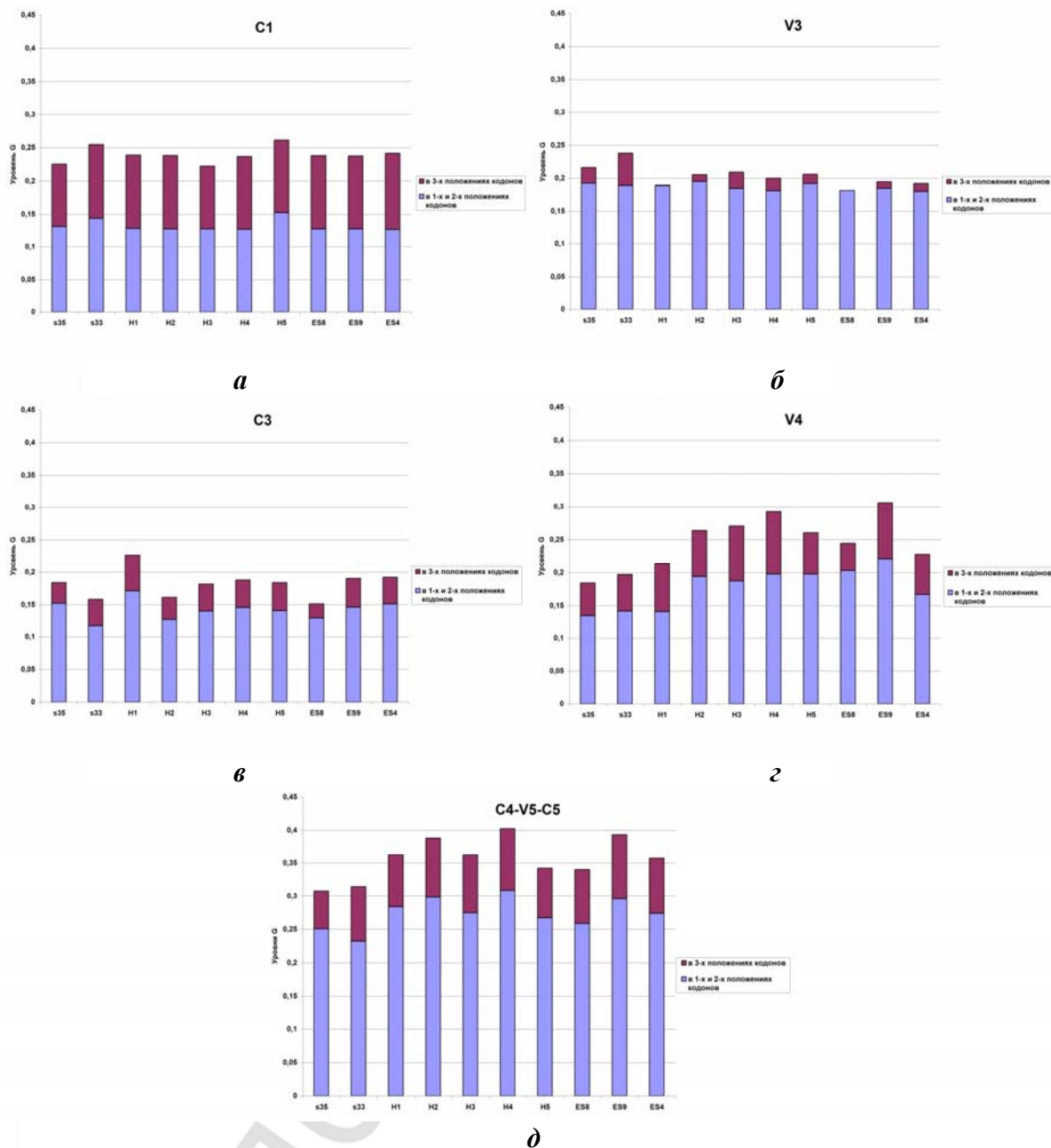


Рис. 52. Средние по каждой из десяти выборок частоты использования гуанина (G) в участках, кодирующих 3D-эпитопы, расположенные в следующих регионах gp120: а — C1; б — V3; в — C3; г — V4; д — C4-V5-C5

Повторимся (см. гл. 1 третьей части), что уровень 3G в участке, кодирующем V3-регион, можно использовать как прогностический критерий скорости развития СПИДа только в том случае, если большинство антител в организме данного пациента синтезировано именно к нему.

На рис. 53 изображены столбчатые диаграммы, отображающие средние значения частоты использования цитозина в каждом из участков,

кодирующих 3D-эпитоп gr120, для каждой из десяти выборок. Каждый столбец, как и на рис. 52, разделен на две части, отображающие процентное содержание цитозина в первых и вторых, а также в третьих положениях кодонов.

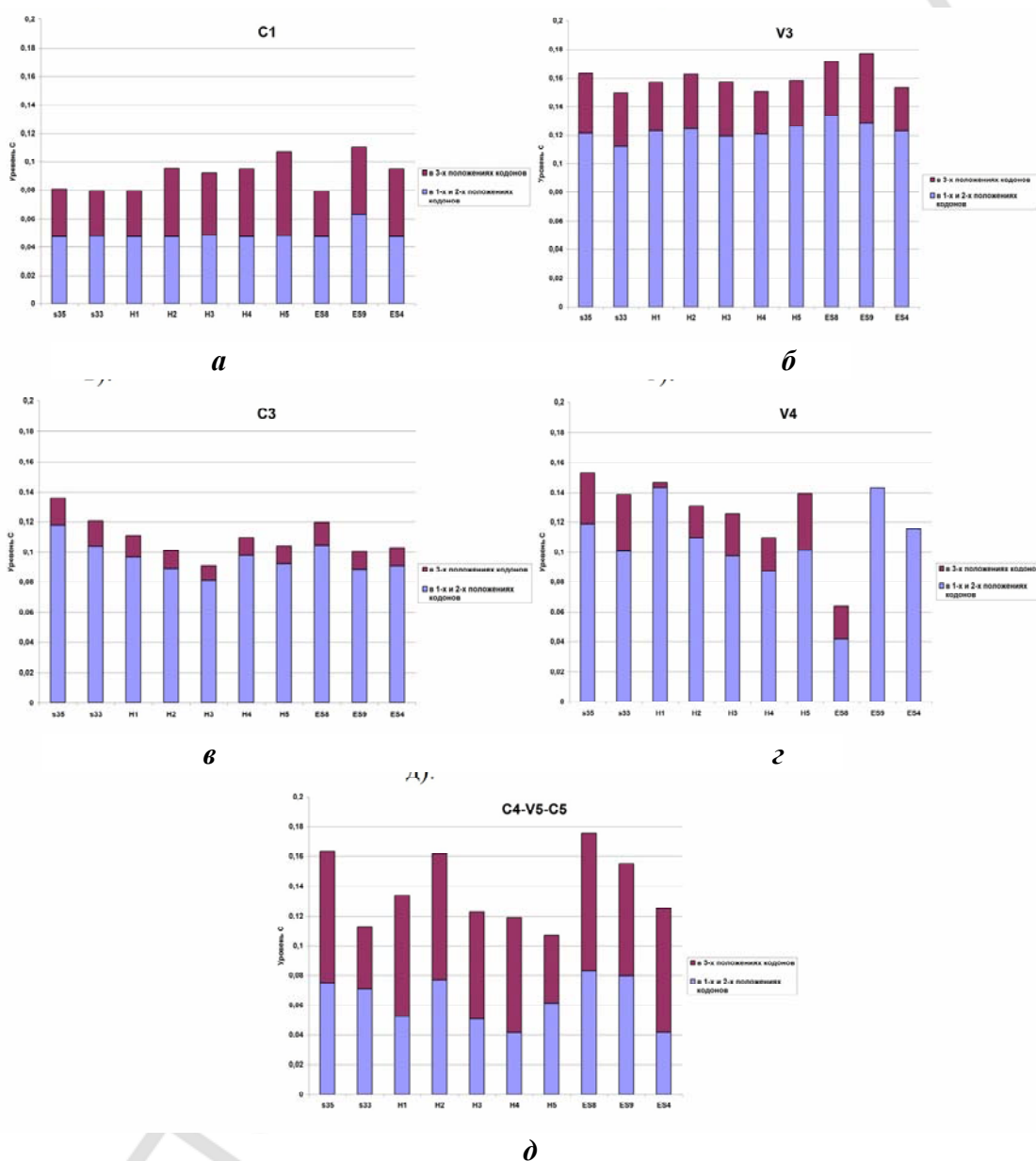


Рис. 53. Средние по каждой из десяти выборок частоты использования цитозина (С) в участках, кодирующих 3D эпитопы, расположенные в следующих регионах gr120: а — C1; б — V3; в — C3; г — V4; д — C4-V5-C5

Интересно отметить, что закономерность, найденная нами при вычислении частот использования цитозина в участке, кодирующем 3D-эпитоп из C1-региона gr120 референтного штамма, характерна и для кодирующих участков других штаммов (рис. 53, а). Это нельзя сказать об участках, кодирующих 3D-эпитоп из регионов C4-V5-C5. Как оказалось, частота

использования цитозина в первых и вторых положениях кодонов участка, кодирующего эпитоп из C4-V5-C5, чаще всего превышает таковую в участке, кодирующем эпитоп из C1 (рис. 53, д). Для участка, кодирующего эпитоп из C4-V5-C5, характерны колебания общего уровня цитозина и частот его использования в положениях кодонов. Уровень его мутабельности по цитозину может быть выше (как в случае с выборками S33 и H3) или ниже мутабельности участка, кодирующего эпитоп из C1.

Самый большой риск несинонимичной замены цитозина характерен для участков, кодирующих V3-регион (рис. 53, б). На втором месте по уровню мутабельности, вызванной высоким риском несинонимичной замены цитозина, чаще всего находятся участки, кодирующие 3D-эпитоп из V4 (рис. 53, в), на третьем — из C3-региона (рис. 53, г).

На рис. 52 и 53 четко видно, что участки, кодирующие эпитоп из региона C1 и участки, кодирующие V3-регион, подвержены гораздо меньшим колебаниям в частотах использования гуанина и цитозина между десятью выборками, чем участки, кодирующие 3D-эпитопы из регионов C3, V4 и C4-V5-C5. В чем же причина таких вариаций? Во время проведения предварительного этапа исследования мы записывали информацию о количестве встречающихся «инделов» в выровненных последовательностях. Термин «индел» — это общее название делеций и инсерций. Почему возникла необходимость объединить эти противоположные явления под одним названием? На деле практически невозможно определить, что же произошло в гене: делеция или инсерция. В выровненных последовательностях результаты инсерций и делеций будут выглядеть одинаково — они приведут к появлению гэпов. В табл. 24 мы указали количество и длину гэпов (в аминокислотных остатках — AA), обнаруженных в выровненных последовательностях участков, гомологичных предсказанным 3D-эпитопам.

Таблица 24

**Характеристика инделов в выровненных последовательностях, гомологичных предсказанным 3D-эпитопам, расположенным в C1, V3, C3, V4 и C4-V5-C5 регионах gp120 ВИЧ1**

Выборка	Количество и длина инделов				
	C1	V3	C3	V4	C4-V5-C5
s35	–	–		5AA	1AA; 1AA
s33	–	–	1AA	9AA; 2AA; 2AA	–
H1	–	–		6 – 24AA; 1 – 2AA	1AA; 1AA; 1AA
H2	–	–	2AA; 1AA	4AA	2AA; 1AA; 1AA; 1AA
H3	–	–		3 – 4AA; 3AA	2 – 4AA; 1AA
H4	–	–	1AA	1AA; 1AA	3AA; 1AA; 1AA; 1AA
H5	–	–	1AA	3 – 6AA; 1AA	2 – 3AA; 1 – 2AA
ES8	–	–	–	–	1 – 3AA; 1AA; 1AA
ES9	–	–	–	2 – 10AA	2AA
ES4	–	–	–	6AA; 1AA	–



Как видно из табл. 24, в участках, гомологичных 3D-эпиту по из региона 1C, и в V3-регионе из десяти выборок генов не было обнаружено ни одного индела (внутри выборки). В участках, гомологичных остальным трем 3D-эпиту, инделов довольно много, многие из них варьируют по своей длине (опять же, внутри выборки). Инделлы должны приводить к значительным колебаниям содержания нуклеотидов в соответствующих кодирующих участках. Изменчивость частот использования гуанина и цитозина в участках, кодирующих эпиту по из регионов C3, V4 и C4-V5-C5 во многом должна быть обусловлена большим количеством инделлов в них.

Вернемся же к частотам нуклеотидных замен в участках, кодирующих 3D-эпиту по. На рис. 54 изображены общие частоты всех нуклеотидных замен в каждом кодирующем участке из каждой выборки (подсчет велся per site). На рис. 55 изображены частоты замен в первых и вторых положениях кодонов. Напомним, что сравнивать можно частоты замен исключительно внутри одной выборки, но не между ними, поскольку время, в течение которого проводили секвенирование гена *env*, для этих выборок различно. Для доказательства достоверности разности между частотами замен подходит статистический метод парных разностей, которые вычисляются внутри каждой выборки, а потом набор разностей из всех десяти выборок обрабатывается как вариационный ряд.

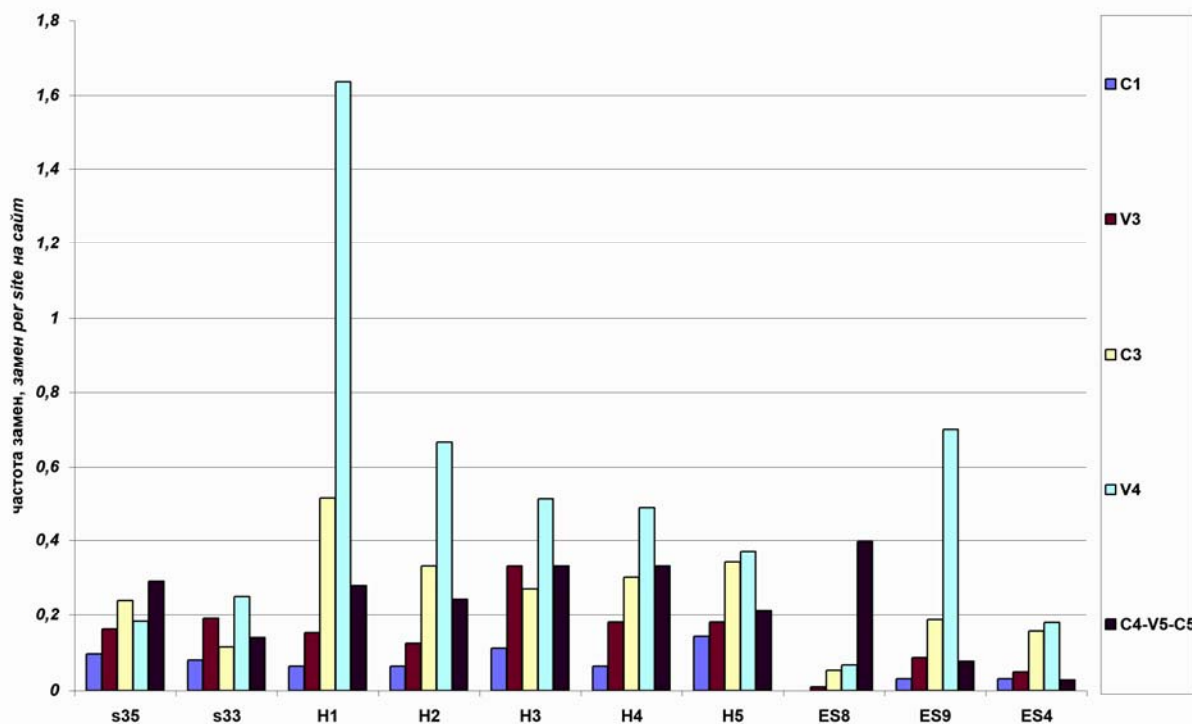


Рис. 54. Общие частоты нуклеотидных замен в участках из десяти выборок генов *env*, кодирующих 3D-эпиту по, расположенные в следующих регионах gp120: C1, V3, C3, V4, C4-V5-C5

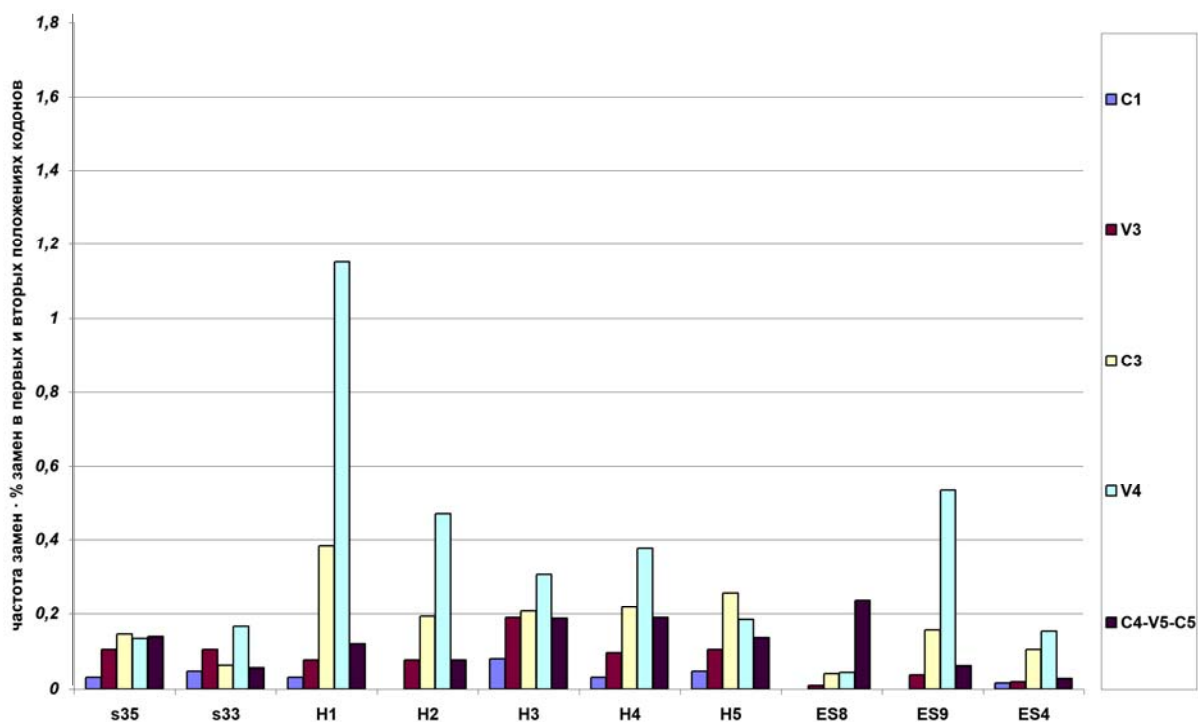


Рис. 55. Частоты нуклеотидных замен в первых и вторых положениях кодонов в участках из десяти выборок генов *env*, кодирующих 3D-эпитопы, расположенные в следующих регионах gp120: C1, V3, C3, V4, C4-V5-C5

По результатам статистической обработки можно заключить, что разница между всеми общими частотами замен, подсчитанными per site в участках, кодирующих эпитопы, достоверна ( $P < 0,05$ ), за исключением разности между участками, кодирующими эпитопы из регионов C4-V5-C5 и V3, из C4-V5-C5 и C3 и из C4-V5-C5 и V4. Действительно, частота замен в участке, кодирующем эпитоп из региона C4-V5-C5, отличается наибольшим непостоянством: в отдельных выборках она заметно выше, в других — ниже частот замен из участков, кодирующих эпитопы из регионов V3, C3 и V4 (см. рис. 54).

На рис. 54 видно, что общая частота нуклеотидных замен минимальна в участке, кодирующем эпитоп из C1-региона, для всех выборок, за исключением ES4. В последней наименьшей является общая частота замен в участке, кодирующем эпитоп из региона C4-V5-C5. Однако на рис. 55 заметно, что частота нуклеотидных замен в первых и вторых положениях кодонов минимальна в участке, кодирующем эпитоп из C1-региона, для всех выборок, включая ES4. По результатам статистической обработки можно заключить, что разница между всеми частотами замен в первых и вторых положениях кодонов, подсчитанными per site в участках, кодирующих эпитопы, достоверна ( $P < 0,05$ ), за исключением разности между участками, кодирующими эпитопы из регионов C4-V5-C5 и V3, из C4-V5-C5 и C3, из C4-V5-C5 и V4 и из C3 и V4. Характерно, что частота замен в первых

и вторых положениях кодонов в участке, кодирующем первый эпитоп (из региона C1), достоверно ниже, чем частоты таких замен во всех остальных участках, кодирующих эпитопы.

В рост частот замен в участках, кодирующих эпитопы из регионов C3, V4 и C4-V5-C5, большой вклад должны вносить делеции и инсерции, а не единичные нуклеотидные замены. Интересно сравнить частоты замен в участках, кодирующих V3-регион и эпитоп из C1-региона, не подверженных делециям и инсерциям, по крайней мере, в десяти изученных выборках генов *env*. Частота замен в менее мутабельных участках, кодирующих 3D-эпитоп из C1-региона, всегда ниже, чем в высоко мутабельных участках, кодирующих V3 регион.

На рис. 56 изображены общие частоты всех нуклеотидных замен, подсчитанных *per nucleotide* в каждом кодирующем участке из каждой выборки. По результатам статистической обработки можно заключить, что разница между всеми общими частотами замен, подсчитанными *per nucleotide* в участках, кодирующих эпитопы, достоверна ( $P < 0,05$ ), за исключением разности между участками, кодирующими эпитопы из регионов C4-V5-C5 и V3, из C4-V5-C5 и C3, из C4-V5-C5 и V4, из V3 и C3.

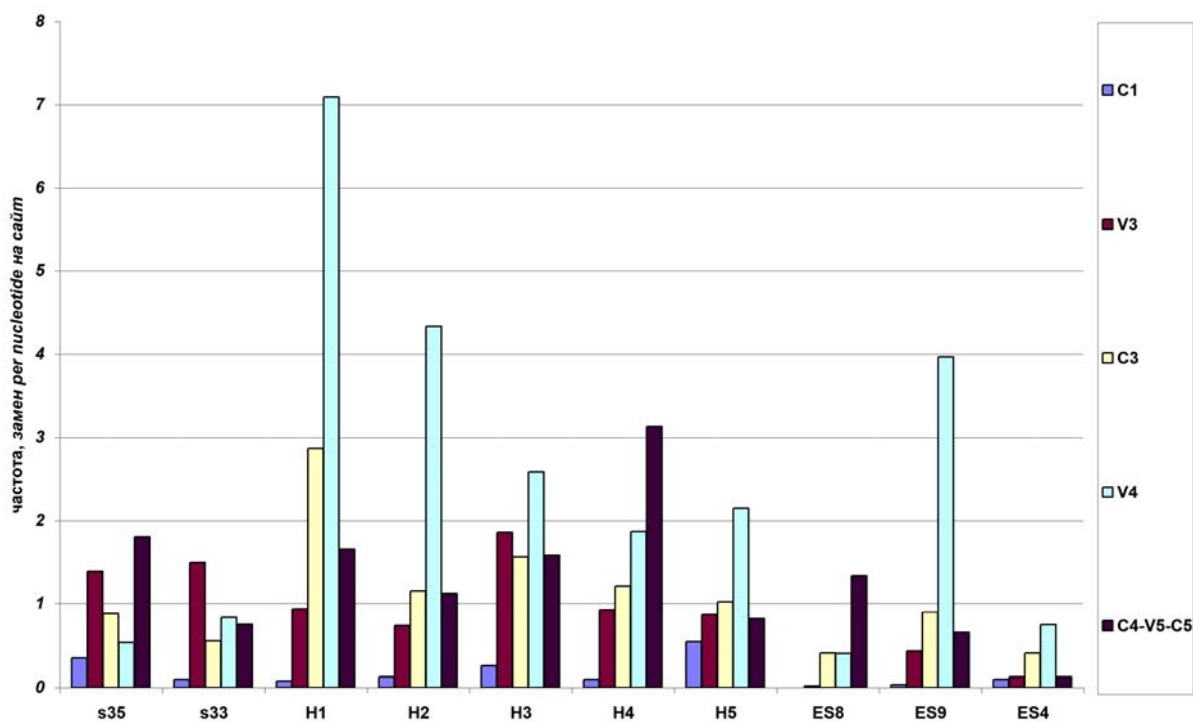


Рис. 56. Общие частоты нуклеотидных замен в участках из десяти выборок генов *env*, кодирующих 3D-эпитопы, расположенные в следующих регионах gp120: C1, V3, C3, V4, C4-V5-C5

На рис. 57 изображены частоты замен *per nucleotide* в первых и вторых положениях кодонов. По результатам статистической обработки можно заключить, что разница между всеми частотами замен в первых и вторых по-

ложениях кодонов, подсчитанными per nucleotide в участках, кодирующих эпитопы, достоверна ( $P < 0,05$ ), за исключением разности между участками, кодирующими эпитопы из регионов C4-V5-C5 и V3, из C4-V5-C5 и C3, из C4-V5-C5 и V4, из V3 и C3 и из V3 и V4.

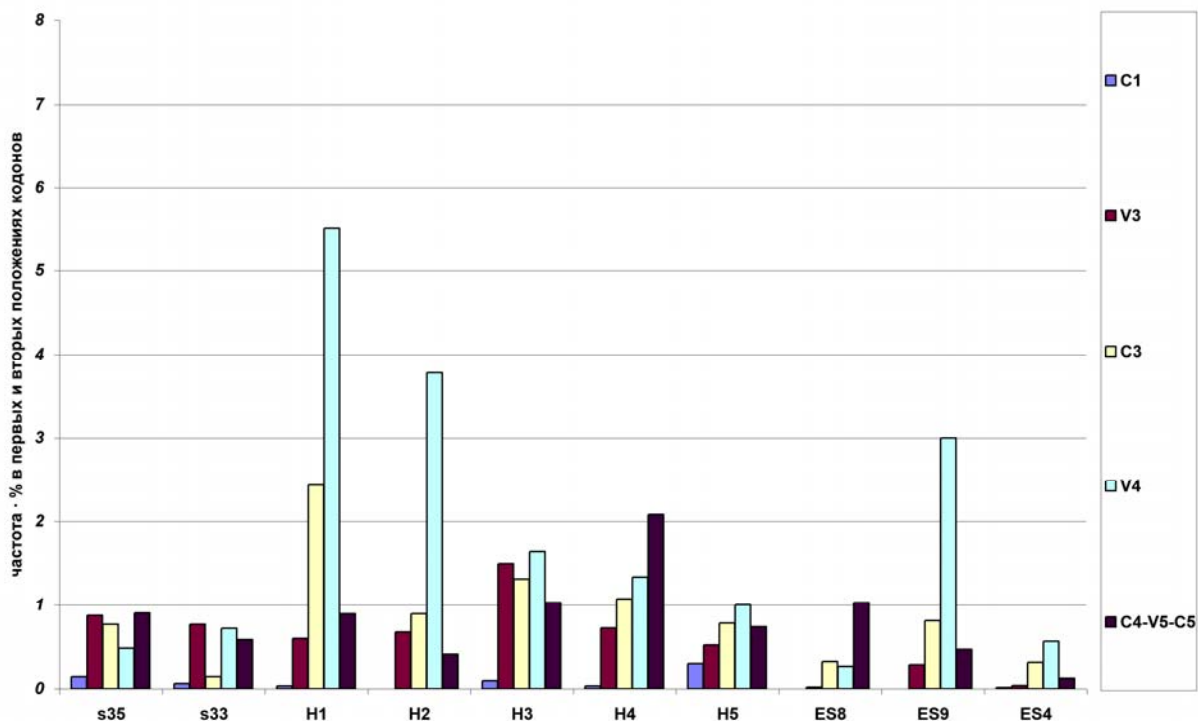


Рис. 57. Частоты нуклеотидных замен в первых и вторых положениях кодонов в участках из десяти выборок генов *env*, кодирующих 3D-эпитопы, расположенные в следующих регионах gp120: C1, V3, C3, V4, C4-V5-C5

Нельзя еще раз не подчеркнуть, что частоты общих замен и замен в первых и вторых положениях кодонов, подсчитанных как per site, так и per nucleotide, в участке, кодирующем первый 3D-эпитоп gp120, достоверно ниже соответствующих частот в участках, кодирующих все остальные эпитопы.

Разрыв между частотами замен в участках, кодирующих эпитопы из участка C1, и частотами замен во всех остальных участках, кодирующих 3D-эпитопы, на рис. 56 и 57 еще больше, чем на рис. 54 и 55. Это говорит о том, что участки, кодирующие эпитопы из регионов V3, C3, V4 и C4-V5-C5 подвержены полиморфизму в гораздо большей степени, чем участок, кодирующий эпитоп из региона C1.

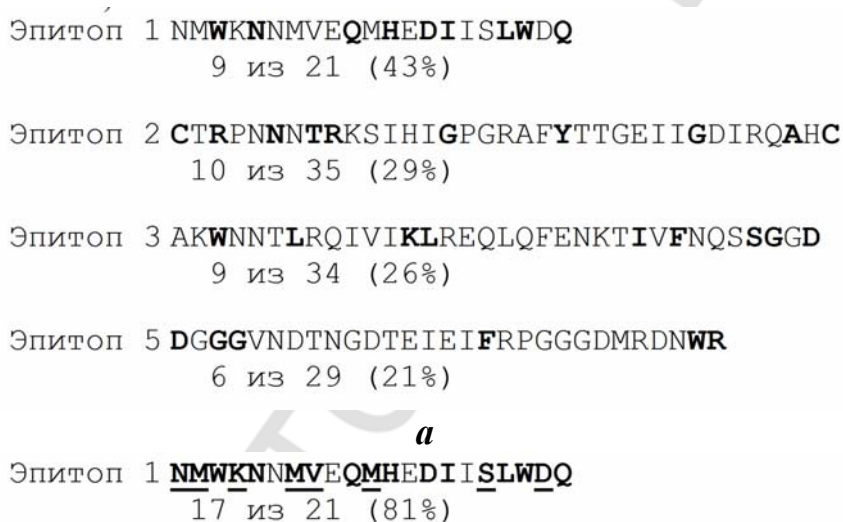
Наши данные убедительно свидетельствуют о том, что участок, кодирующий эпитоп из региона C1, не только наименее мутабельный, но и наименее вариабельный среди всех остальных участков, кодирующих пространственные эпитопы gp120 ВИЧ1.

Высокие частоты замен в участках из C3-, V4- и C4-V5-C5-регионов, вызванные, в том числе, и инделами, свидетельствуют о более низкой их

функциональной значимости по сравнению с таковой для участка из региона С1 и с таковой для V3-региона. Наличие полиморфизма, особенно связанного с инделами, говорит о том, что существует много «вариантов», не приводящих к существенным изменениям функции белка.

В финальной части нашей работы мы выровняли все аминокислотные последовательности из десяти выборок (общее их число — 385). После этого мы выяснили, насколько консервативными являются аминокислотные последовательности участков, гомологичных предсказанным 3D-эпитопам, во всех изученных выборках.

Результаты выравнивания представлены на рис. 58. Жирным шрифтом обозначены инвариантные аминокислотные остатки, процент которых указан под аминокислотной последовательностью. Рис. 58, *а* подтверждает наши результаты: наименее вариабельным является первый эпитоп gp120 (расположенный в регионе С1).



*б*

*Рис. 58.* Инвариантные (обозначены жирным шрифтом) и консервативные (обозначены подчеркнутым жирным шрифтом) аминокислотные остатки в участках, гомологичных предсказанным пространственным эпитопам, в выборке из 385 аминокислотных последовательностей gp120 ВИЧ1:

*а* — инвариантные аминокислотные остатки; *б* — консервативные

На рис. 58, *б* мы отметили подчеркнутым шрифтом аминокислотные остатки, замены которых были обнаружены только в двух последовательностях из 385 исследованных.

Аминокислотные остатки, не подверженные полиморфизму (отмеченные на рис. 58, *б* подчеркнутым шрифтом), можно считать консервативными. Общий процент инвариантных и консервативных аминокислотных остатков в 3D-эпитопе из региона С1 gp120 ВИЧ1 составляет 81 %. По нашему мнению, вакцина против ВИЧ1, созданная с использованием эпитопа

из региона С1, была бы более эффективной, чем многочисленные вакцины, созданные против V3-региона и других участков gp120.

**Биологические проблемы, которые можно решить с использованием данного метода.** С помощью метода, описанного в данной главе, можно производить отбор наименее мутабельных пространственных В-клеточных эпитопов в бактериальных и вирусных белках для дальнейшей работы по созданию вакцин. Данный метод может применяться только при наличии большого количества нуклеотидных последовательностей исследуемого гена от разных штаммов и клонов вируса или бактерии.

Результаты предсказания уровней мутабельности 3D-эпитопов gp120 ВИЧ1 (см. гл. 2 третьей части) в целом были подтверждены в исследовании, описанном в данной главе. Существенным дополнением к определению уровня мутабельности участков белка является оценка степени нейтральности аминокислотных замен и крупных делеций и инсерций в них.

Насколько эффективной будет вакцина против ВИЧ1, созданная к первому 3D-эпитопу gp120, сказать трудно. Вирус иммунодефицита человека очень изменчив. Единственное, что можно утверждать, так это то, что такая вакцина будет эффективней, чем вакцины, созданные к другим эпитопам gp120, отличающимся большим уровнем мутабельности и, следовательно, изменчивости.

По нашему мнению, применение метода, описанного в гл. 3 второй части, гл. 1 и 2 третьей части, гл. 1 пятой части, к другим вирусам (и бактериям), менее изменчивым, чем ВИЧ, имеет большие шансы на успех.

## Часть шестая

### Описание оригинальных компьютерных алгоритмов

#### Глава 1

##### Алгоритм «codonchanges»

**Объект исследования:** две нуклеотидные последовательности, кодирующие гомологичные белки или участки гомологичных белков.

**Необходимость предварительного выравнивания:** есть.

**Распознаваемые символы:** A, T, G, C, N.

**Последовательность действий:**

1. Ввести нуклеотидную последовательность, по направлению от которой планируется расчет частот замен нуклеотидов и кодонов, в ячейку на листе «sequence1».

2. Ввести нуклеотидную последовательность, по направлению к которой планируется расчет частот замен нуклеотидов и кодонов, в ячейку на листе «sequence2».

3. Ввести в ячейку на листе «codon» кодон, замены которого вы собираетесь изучить по направлению от последовательности, введенной в ячейку на листе «sequence1», к последовательности, введенной в ячейку на листе «sequence2». (Данное действие можно повторять неограниченное количество раз, вводя новые кодоны в ту же ячейку).

**Вычисляемые показатели:**

1. G+C, 1GC, 2GC, 3GC, GC4f, GC2f3p, G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T, частоты использования кодонов в гене (в том числе показатель ПТК) и частоты использования аминокислотных остатков в кодируемом белке (в том числе показатели GARP, FYMINK и 10AA), энтропия распределения аминокислотных остатков в белке (H).

**Расположение:** на листе «gene1» расположены значения этих показателей для нуклеотидной последовательности, введенной в ячейку на листе «sequence1», на листе «gene2» расположены значения этих показателей для нуклеотидной последовательности, введенной в ячейку на листе «sequence2».

2. Частоты каждого из 12 возможных типов нуклеотидных замен по направлению от последовательности, введенной в ячейку на листе «sequence1», к последовательности, введенной в ячейку на листе «sequence2», в каждом из трех положений кодонов и общая их частота.

**Расположение:** на листе «gene1».

3. Количество каждого из 64 кодонов и гэпов, оказавшихся в тех сайтах последовательности, введенной в ячейку на листе «sequence2», в которых в последовательности, введенной в ячейку на листе «sequence1», находился кодон, введенный в ячейку на листе «codon».

**Расположение:** на листе «changes».



## Глава 2

### Алгоритм «Vvk Dinucleotides»

**Объект исследования:** одна нуклеотидная последовательность.

**Распознаваемые символы:** A, T, G, C.

**Последовательность действий:** ввести нуклеотидную последовательность в ячейку на листе «sequence».

**Вычисляемые показатели:**

1. G+C, 1GC, 2GC, 3GC, G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T, частоты использования кодонов в гене.

**Расположение:** на листе «gene».

2. Общие частоты использования динуклеотидов в гене, статистическая вероятность частоты использования каждого динуклеотида в гене, частоты использования каждого динуклеотида в трех позициях относительно занимаемых положений в кодоне.

**Расположение:** на листе «gene».

## Глава 3

### Алгоритм «Vvk in Group»

**Объект исследования:** до 50 нуклеотидных последовательностей, кодирующих гомологичные белки или участки гомологичных белков.

**Необходимость предварительного выравнивания:** нет.

**Распознаваемые символы:** A, T, G, C.

**Последовательность действий:** ввести нуклеотидные последовательности в ячейки, расположенные на листе «sequences» (ячейки расположены в столбик, при копировании всего набора последовательностей, разделенных знаком конца абзаца, в первую ячейку, все остальные ячейки автоматически).

**Вычисляемые показатели:**

1. G+C, 1GC, 2GC, 3GC, GC4f, GC2f3p, G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T, частоты использования кодонов в гене (в том числе показатель ПТК) и частоты использования аминокислотных остатков в кодируемом белке, энтропия распределения аминокислотных остатков в белке (H), общие частоты использования динуклеотидов в гене, статистическая вероятность частоты использования каждого динуклеотида в гене, частоты использования каждого динуклеотида в трех позициях относительно занимаемых положений в кодоне.

**Расположение:** на каждом листе от «gene» до «gene49» расположены значения этих показателей для каждой из нуклеотидных последовательностей, введенных в ячейки на листе «sequences».

2. G+C, 1GC, 2GC, 3GC, GC4f, GC2f3p, G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T, частоты использования кодонов в гене (в том



числе показатель ПТК) и частоты использования аминокислотных остатков в кодируемом белке (в том числе показатели GARP, FYMINK и 10AA), энтропия распределения аминокислотных остатков в белке (H), общие частоты использования динуклеотидов в гене.

**Расположение:** на листе «results», в столбцах под соответствующими названиями (строки расположены в том же порядке, что и на листе «sequences»).

## Глава 4

### Алгоритм «Coding Genome Scanner»

**Объект исследования:** запись о частотах использования кодонов в каждом кодирующем участке организма данного вида (штамма) из электронной базы данных Codon Usage Database.

**Последовательность действий:** скопировать весь текст из записи о частотах использования кодонов в каждом кодирующем участке организма данного вида (штамма) из электронной базы данных Codon Usage Database на лист «all CDSs» (копировать весь текст нужно непосредственно в обозначенную ячейку, остальные заполнятся автоматически).

**Вычисляемые показатели:** G+C, 1GC, 2GC, 3GC, GC4f, GC2f3p, GC0f, G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T, частоты использования кодонов в гене (в том числе показатель ПТК) и частоты использования аминокислотных остатков в кодируемом белке (в том числе и показатели GARP, FYMINK и 10AA), энтропия распределения аминокислотных остатков в белке (H).

**Расположение:** на листе «List of indexes», в столбцах под соответствующими названиями (строки расположены в том же порядке, что и кодирующие участки в записи о частотах использования кодонов в каждом кодирующем участке организма данного вида (штамма) из электронной базы данных Codon Usage Database, скопированном на лист «all CDSs»).

## Глава 5

### Алгоритм «Vvk Consensus»

**Объект исследования:** до 100 нуклеотидных последовательностей, кодирующих гомологичные белки или участки гомологичных белков.

**Необходимость предварительного выравнивания:** есть.

**Распознаваемые символы:** A, T, G, C, N.

**Последовательность действий:** ввести нуклеотидные последовательности в ячейки, расположенные на листе «sequences» (ячейки расположены в столбик, при копировании всего набора последовательностей, разделенных знаком конца абзаца, в первую ячейку, все остальные заполняются автоматически).

**Вычисляемые показатели:**

1. G+C, 1GC, 2GC, 3GC, G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T.

**Расположение:** в столбцах на листе «content» расположены значения этих показателей для каждой из нуклеотидных последовательностей, введенных в ячейки на листе «sequences».

2. Количество нуклеотидных замен каждого из 12 возможных типов, подсчитанное per site от консенсусной последовательности в каждом из трех положений кодонов и общая их частота.

**Расположение:** на листе «results», в таблице под названием «NUMBER OF ALL SUBSTITUTIONS from consensus sequence per site».

3. Количество нуклеотидных замен каждого из 12 возможных типов, подсчитанное per nucleotide от консенсусной последовательности в каждом из трех положений кодонов и общая их частота.

**Расположение:** на листе «results», в таблице под названием «NUMBER OF ALL SUBSTITUTIONS from consensus sequence per nucleotide».

4. Количество и процентное соотношение 6 возможных типов взаимных замен нуклеотидов (G ↔ A, C ↔ T, C ↔ A, G ↔ T, A ↔ T и G ↔ C), подсчитанных per site и per nucleotide.

**Расположение:** на листе «calculations».

## Глава 6

### Алгоритм «Vvk in Length»

**Объект исследования:** одна нуклеотидная последовательность.

**Распознаваемые символы:** A, T, G, C.

**Последовательность действий:**

1. Ввести нуклеотидную последовательность в ячейку, расположенную на листе «full sequence».

2. Ввести длину (в кодонах) частей, на которые будет разрезана изучаемая последовательность, в ячейку на листе «length» (это действие может повторяться неоднократно путем введения разной длины нарезаемых частей в ту же ячейку).

**Вычисляемые показатели:** G, 1G, 2G, 3G, C, 1C, 2C, 3C, A, 1A, 2A, 3A, T, 1T, 2T, 3T.

**Расположение:** в столбцах на листе «results» расположены значения этих показателей для каждой части изучаемой нуклеотидной последовательности, нуклеотидные последовательности всех нарезанных частей расположены в строках на листе «parts».

## Глоссарий

Основные источники литературы для составления глоссария: энциклопедический словарь по генетике под редакцией академика НАН Республики Беларусь Н. А. Картеля [15]; терминологический словарь по иммунологии, составленный членом-корреспондентом НАН Республики Беларусь Л. П. Титовым [26]; толковый словарь по биологии под редакцией В. П. Андреева [1]; книга А. М. Дейчмана с соавторами под названием «Редактирование РНК. Гипотетические механизмы» [62]; руководство по молекулярной биологии «Гены и геномы» М. Сингер и П. Берга [23].

### А

**Абсолютно GC-бедный кодон (триплет)** — кодон, не содержащий *гуанин* и *цитозин* ни в одном из трех *положений*.

**Абсолютно GC-богатый кодон (триплет)** — кодон, содержащий *гуанин* или *цитозин* во всех трех *положениях*.

**AICDA (AID). Индуцируемая активацией цитозин-ДНК-дезаминаза** — фермент, дезаминирующий цитозин и 5-метилцитозин в клеточной одноцепочечной ДНК. Был открыт в активированных В-лимфоцитах; впоследствии выяснилось, что активен он не только в них [106]. Считается, что AICDA повышает уровень изменчивости генов иммуноглобулинов, внося в них замены С на U (способствует соматическому гипермутагенезу). Относится к суперсемейству *APOBEC*.

**Адаптация** — в эволюционном плане под адаптацией понимают приспособленность (за счет морфологических, биохимических, поведенческих или иных особенностей) популяции в целом к выживанию в условиях данной среды обитания.

**Аденин (6-аминопурин)** — пуриновое *азотистое основание*, входящее в состав нуклеотидов *ДНК* и *РНК*.

**Аденозин-РНК-дезаминаза (ADAR)** — семейство ферментов, дезаминирующих аденин, входящий в состав РНК. В результате дезаминирования аденозина образуется *гипоксантин (инозин)*. Ферменты из семейства ADAR предпочтительно дезаминируют аденин в двухцепочечной РНК. У отдельных представителей семейства ADAR обнаружены участки, связывающие двухцепочечную ДНК, но факт дезаминирования ими аденина в ДНК пока не доказан.

**Азотистые основания** — обширная группа органических веществ, которые проявляют основные свойства благодаря присутствию в составе их молекул атомов азота. В состав *мажорных нуклеотидов ДНК* входят четыре типа азотистых оснований: *аденин, гуанин, цитозин и тимин*. В со-

став нуклеотидов РНК также входят четыре типа азотистых оснований: аденин, гуанин, цитозин и урацил. По своей структуре аденин и гуанин являются пуриновыми азотистыми основаниями, а цитозин, тимин и урацил — пиримидиновыми.

**Акрофильность** — свойство аминокислотного остатка располагаться на поверхности белковой глобулы, отражающее высокую вероятность включения его в линейные (и пространственные) В-клеточные эпитопы [75].

**Акрофобность** — свойство аминокислотного остатка располагаться внутри белковой глобулы, отражающее низкую вероятность включения его в линейные (и пространственные) В-клеточные эпитопы [75].

**Активация тРНК** — процесс образования *аминоацил-тРНК* из *аминокислоты* и *тРНК* с помощью ферментов *аминоацил-тРНК-синтетаз*. Активация тРНК протекает в два этапа: 1) взаимодействие аминокислоты с АТФ с образованием комплекса аминокислоты и АМФ; 2) взаимодействие образовавшегося комплекса с тРНК с образованием *аминоацил-тРНК*.

**Активный центр фермента** — участок фермента, благодаря которому он проявляет специфичность при присоединении к субстрату и преобразовании его. Как правило, активный центр является наиболее консервативным участком фермента. Формируется благодаря взаимному расположению определенных аминокислотных остатков.

**Акцепторная точка сплайсинга** — участок про-мРНК на 3' (правом) конце интрона. Этот участок в ядерной про-мРНК млекопитающих содержит консервативную последовательность нуклеотидов (несколько идущих подряд пиримидиновых нуклеотидов, а затем — аденин и гуанин), которая и отмечает стык интрона и экзона.

**Аллель** — одно из нескольких возможных структурных состояний гена, возникшее в результате мутаций.

**Аллельные гены** — гены, определяющие проявление разных вариантов данного признака, расположенные в одинаковых локусах гомологичных хромосом.

**Альфа-спираль** — распространенный мотив *вторичной структуры белка*, представляющий собой спираль, в которой каждая N–H группа «остова» аминокислоты образует водородную связь с C=O группой «остова» аминокислоты, расположенной на 4 аминокислотных остатка раньше.

**Амбер-кодон** — общепринятое название *терминального* триплета TAG (UAG — в РНК).

**Амбер-мутация** — мутация, приводящая к возникновению в *кодирующей* участке ДНК (мРНК) триплета TAG (UAG — в мРНК).

**Амбер-супрессор** — ген, кодирующий тРНК, антикодон которой комплементарен кодону UAG (амбер-кодону). В результате амбер-супрессии возникает отклонение от универсального генетического кода. *Пример:* у некоторых организмов терминальный кодон UAG кодирует аминокисло-

ту глутамин. Амбер-супрессор возникает в результате мутации одного из генов, кодирующих тРНК, связывающую глутамин.

**Аминоацил-тРНК** — транспортная РНК, связавшаяся в процессе *активации* с аминокислотой. Участвует в процессе *трансляции*.

**Аминоацил-тРНК-синтетазы** — ферменты, осуществляющие процесс *активации тРНК*. Аминоацил-тРНК-синтетазы обладают специфичностью — катализируют активацию только тех тРНК, антикодоны которых могут связываться с триплетами мРНК, кодирующими данную конкретную аминокислоту. Сначала аминоацил-тРНК-синтетаза связывается с аминокислотой, а затем — с тРНК, антикодон которой комплементарен триплету, кодирующему данную аминокислоту. После этого происходит собственно активация тРНК.

**Аминоацильный центр рибосомы** — участок рибосомы, в котором во время трансляции происходит процесс комплементарного связывания антикодона аминоацил-тРНК с находящимся в нем кодоном мРНК.

**Аминокислоты** — класс органических соединений, содержащих карбоксильную группу и аминогруппу. 20 протеиногенных аминокислот кодируются соответствующими триплетами ДНК (мРНК), являются мономерами для *белков*.

**Аминотерминальный конец белка (N-конец)** — конец полипептидной цепи, имеющий свободную аминогруппу. N-конец соответствует первым аминокислотным остаткам в полипептиде (в процессе трансляции белок синтезируется от N-конца к С-концу).

**Амплификация** — увеличение числа копий участков ДНК в ходе эволюции или в процессе эксперимента (в т. ч. в ПЦР). Фактически, дупликация генов, создающая по *теории нейтральности* М. Кимуры материал для эволюции, является разновидностью амплификации. Благодаря амплификации генов возникли *генные семейства*, множественные и тканеспецифические формы ферментов, *псевдогены*.

**Антиген** — субстанция, способная вызвать развитие иммунного ответа. В антигене выделяют инертную часть (части) и эпитоп (эпитопы), при этом иммунный ответ вырабатывается непосредственно на эпитоп.

**Антикодон** — триплет тРНК, комплементарный определенному триплету мРНК. В процессе *трансляции* в аминоацильном центре рибосомы происходит временное связывание кодона мРНК с комплементарным ему антикодоном аминоацил-тРНК. Число генов тРНК в геноме всегда меньше, чем 61. Это означает, что одна молекула тРНК может связываться с несколькими синонимичными кодонами. Для того чтобы объяснить этот феномен, Крик предложил гипотезу «качающегося» взаимодействия кодона с антикодоном. По этой гипотезе, в первом положении антикодона должны находиться модифицированные (минорные) нуклеотиды, которые в большей степени, чем мажорные, способны связываться с разными видами

нуклеотидов, находящихся в третьем положении кодона мРНК. Вероятно, свой вклад в «качающееся» взаимодействие вносят процессы *редактирования РНК*. Возможность ферментативного дезаминирования цитозина в тРНК у архей была доказана экспериментально.

**Антимутагены** — факторы, снижающие частоту возникновения мутаций.

**Антитело** — молекула иммуноглобулина, образующаяся в активированных В-лимфоцитах в результате антигенной стимуляции. Антитело специфично к определенному эпитопу антигена.

**Антитерминация** — механизм регуляции транскрипции, основанный на предотвращении терминации трансляции за счет специфического воздействия антитерминаторных белков на РНК-полимеразы. При этом РНК-полимеразы продолжают трансляцию при прохождении терминаторных участков мРНК (*шпилек*). Антитерминация также возможна при связывании антитерминаторных белков с некоторыми терминаторными участками. Механизм антитерминации подробно изучен у бактериофагов.

**Антитерминаторные белки** — комплекс белков, способных связываться с РНК-полимеразами и предотвращать терминацию трансляции. Некоторые антитерминальные белки связываются непосредственно с терминаторами, тем самым подавляя их активность.

**Апоиндуктор** — активаторный белок, который может связываться с оператором и стимулировать транскрипцию в оперонах (у прокариот). Благодаря апоиндуктору осуществляется дополнительная стимуляция транскрипции, поскольку собственно *индуктор*, попадая в бактериальную клетку, конкурирует с опероном за *репрессор* и стимулирует тем самым транскрипцию. Апоиндуктор может угнетать транскрипцию при изменении его положения относительно промотора.

**Аполипротеин В редактирующая цитидиндезаминаза (АРОВЕС)** — первый открытый представитель суперсемейства цитидиндезаминаз, дезаминирующих цитозин в ДНК и РНК. Аполипротеин В редактирующая цитидиндезаминаза дезаминирует цитозин в составе мРНК, кодирующей аполипротеин В. Впоследствии термин «АРОВЕС» стал общим названием целого суперсемейства ферментов. Собственно «аполипротеин В редактирующая цитидиндезаминаза» относится к семейству АРОВЕС1, представители которого дезаминируют цитозин в разнообразных клеточных и вирусных РНК.

**АР-сайт (апуриновый/апиримидиновый сайт)** — сайт в молекуле ДНК, в котором отсутствует азотистое основание, но остается дезоксирибоза, связанная с сахарами двух соседних нуклеозидов с помощью фосфодиэфирных связей. АР-сайт образуется в результате работы ДНК-гликозилаз, вырезающих нуклеотиды из неправильно спаренных пар или из одноцепочечной ДНК. Важно подчеркнуть, что АР-сайт не приводит к раз-

рыву цепочки ДНК. Разрыв цепочки ДНК на месте AP-сайта создается ферментативно — при помощи *AP-эндонуклеаз*.

**АРОВЕС1** — семейство ферментов, дезаминирующих цитозин в разнообразных клеточных и вирусных РНК. Преимущественно дезаминируют цитозин в одноцепочечной РНК.

**АРОВЕС2** — семейство ферментов, родственных остальным представителям суперсемейства АРОВЕС (их сходство было обнаружено при помощи методов молекулярной филогенетики), свойства которых пока не охарактеризованы.

**АРОВЕС3** — семейство ферментов, дезаминирующих цитозин в одноцепочечной ДНК, в том числе в одноцепочечной ДНК вируса иммунодефицита человека.

**AP-эндонуклеаза** — фермент, функцией которого является разрыв цепочки ДНК на месте *AP-сайта*. Только после появления одноцепочечного разрыва ДНК на месте AP-сайта достраивается нуклеотид.

**Археи** — одно из трех надцарств живых организмов. Представители архей — прокариотические организмы. Ранее архей причисляли к бактериям (точнее — к археобактериям). Отделить надцарство бактерий от надцарства архей удалось лишь при помощи молекулярной филогенетики [56]. Как правило, археи используют нетрадиционные способы получения энергии: окисление железа, серы, синтез метана.

**АТ-давление** — направление мутационного давления, при котором частота возникновения и фиксации замен гуанина и цитозина на аденин и тимин превышает частоту возникновения и фиксации замен аденина и тимина на гуанин и цитозин. Термин справедлив для организмов, геном которых представлен двухцепочечной ДНК, гены в которой находятся на обеих цепочках. АТ-давление приводит к закономерному уменьшению GC-насыщенности ДНК в ряду поколений.

О силе АТ-давления в данном геноме (или в данной изохоре) можно (примерно) судить по уровню его (ее) GC-насыщенности. При GC-насыщенности в пределах от 40 до 50 % следует предполагать наличие слабого АТ-давления в данном геноме (или в данной изохоре); при GC-насыщенности менее 40 % АТ-давление считается сильным.

Более точным критерием сильного АТ-давления в геноме является низкий коэффициент корреляции ( $R < 0,5$ ) между значениями 3GC и G+C для всех кодирующих участков данного генома (при низком значении 3GC) [84].

**Аттенуация** — механизм регуляции транскрипции в оперонах (у прокариот), который работает при котранскрипционной трансляции (трансляции, протекающей одновременно с транскрипцией). Прекращение транскрипции может происходить после синтеза первого (лидерного) пептида из данного оперона, в результате образования *шпильки мРНК*. Шпилька мРНК не образуется при задержке рибосомы во время синтеза первого пептида.

В последнем случае образуется альтернативная вторичная структура мРНК (без образования шпильки) и процесс транскрипции не прерывается.

**Аттенуатор** — участок мРНК, играющий критическую роль в процессе *аттенуации*: этот участок может образовывать шпильку в процессе транскрипции оперона (у прокариот). Образование шпильки прерывает транскрипцию.

**Аутоиммунные заболевания** — заболевания, вызванные специфическими иммунологическими реакциями гуморального и клеточного типов против эпитопов собственных антигенов.

**Аффинность** — сила связывания (вероятность связывания) эпитопа определенного антигена с активным центром антитела (паратопом).

## Б

**Бактерии** — представители одного из трех надцарств живых организмов. Вместе с *археями* бактерии относятся к *прокариотическим организмам*. Четко отграничить надцарство бактерий от надцарства архей позволили лишь методы молекулярной филогенетики [56].

**Белки** — высокомолекулярные генетические биополимеры, состоящие из аминокислотных остатков, соединенных в определенной последовательности пептидными связями. В качестве мономеров для белков могут использоваться 20 протеиногенных аминокислот: фенилаланин, серин, лейцин, пролин, изолейцин, метионин, треонин, валин, аланин, тирозин, цистеин, триптофан, глутамин, гистидин, аргинин, глутаминовая кислота, аспарагин, аспарагиновая кислота, глицин. У отдельных организмов к протеиногенным аминокислотам могут относиться также селеноцистеин и пирролизин. Существуют *первичная, вторичная, третичная и четвертичная структуры* белков.

**Бета-структура** — разновидность вторичной структуры белка, которая представляет собой две цепочки аминокислот, объединенных друг с другом за счет водородных связей между «остовами» их аминокислотных остатков.

**Биохимическая мутация** — мутация, проявляющаяся тем или иным изменением метаболизма.

**Блоттинг** — 1) саузерн-блоттинг (назван по фамилии автора: «Southern») — процесс переноса фрагментов электрофоретически разделенной ДНК с геля (агарозного или полиакриламидного) на твердый носитель (бумагу, нитроцеллюлозу и др.); 2) нозерн-блоттинг — аналогичная процедура для РНК; 3) вестерн-блоттинг — аналогичная процедура для белка.

**Блуждающие гены** — см. *транспозоны*.

**В-лимфоциты** — клетки иммунной системы, в зрелом активированном состоянии синтезирующие молекулы иммуноглобулинов — антитела. Основные мембранные маркеры В-лимфоцитов — молекулы CD19 и CD20.



**Бутстрэп** — набор методов, используемых для статистического подтверждения корректности расположения ветвей на *дендрограмме* (*топологии* дендрограммы). Общий принцип бутстрэпа состоит в создании нескольких сот новых *последовательностей* из случайным образом перемешанных участков исходных и построении на основании сравнения новых последовательностей множества промежуточных дендрограмм (с использованием метода, примененного для построения исходной дендрограммы). В конечной дендрограмме ветви принимают положение, встречавшееся в большинстве промежуточных дендрограмм. Недостатком метода является нарушение структуры сравниваемых последовательностей, то есть удаление «зерен сходства» между ними.

## В

**Вариабельные сайты** — в массиве выровненных последовательностей так называются положения, в которых нуклеотиды (в выровненных последовательностях ДНК или РНК) или аминокислотные остатки (в выровненных аминокислотных последовательностях) не идентичны друг другу.

**Вид** — *современное определение*: виды живых организмов — репродуктивно изолированные единицы, эволюционирующие независимо друг от друга и расходящиеся в генетическом отношении друг от друга. Данное определение отличается от классического тем, что подчеркивает ведущее значение репродуктивной изоляции как критерия вида; *классическое определение*: вид — совокупность особей, занимающих определенный ареал, имеющих генетическое, морфологическое, физиологическое и поведенческое сходство, свободно скрещивающихся между собой и дающих плодовитое потомство.

**Видимые мутации** — мутации, вызывающие изменение каких-либо «видимых» признаков организма, то есть проявляющиеся фенотипически.

**Водородная связь** — слабое, нековалентное электростатическое взаимодействие между полярными группами, имеющими в своем составе атомы водорода и электроотрицательные атомы кислорода, азота или хлора. Водородными связями поддерживается вторичная структура белка. Комплементарность нуклеотидов обусловлена образованием водородных связей между ними.

**Время действия гена** — временной промежуток в онтогенезе, во время которого ген проявляет свою активность.

**Вторичная структура белка** — структура, образующаяся в результате образования водородных связей между группами атомов аминокислотных остатков полипептида.

**Вторичная структура РНК (одноцепочечной ДНК)** — структура, которая образуется в результате взаимодействия нуклеотидов одной цепочки РНК (ДНК) друг с другом за счет образования водородных связей.

Наиболее известными элементами вторичной структуры РНК и ДНК являются *шпильки* и *G-квардуплекс*. Водородные связи образуются между нуклеотидами, входящими в состав одной и той же цепи, как правило, при наличии идентичных и не полностью идентичных инвертированных повторов в ней. При комплементарном объединении инвертированных повторов образуется шпилька (участок двухцепочечной РНК или ДНК). Из нуклеотидной последовательности, расположенной между инвертированными повторами образуется «петля», а собственно из объединенных повторов — «стебель». G-квардуплекс образуется при образовании водородных связей между четырьмя молекулами гуанина.

**Выравнивание последовательностей** — процесс поиска участков наибольшего сходства между нуклеотидными или аминокислотными последовательностями. Осуществляется с помощью специальных компьютерных программ (ClustalW, MEGA 4 и др.). Результатом выравнивания является создание массива *выровненных последовательностей*. Благодаря выравниванию последовательностей появляется возможность корректного применения методов вычисления эволюционных дистанций к группе гомологичных генов (белков).

**Выровненные последовательности** — нуклеотидные или аминокислотные последовательности, в которых найдены участки наибольшего сходства друг с другом. Графически массив выровненных последовательностей представляет собой ряд расположенных друг под другом последовательностей, в которых участки наибольшего сходства находятся на одном (вертикальном) уровне.

**Вырожденность генетического кода** — свойство генетического кода, заключающееся в наличии в нем синонимичных кодонов. Данное свойство более корректно называется *избыточностью*: один аминокислотный остаток может быть закодирован несколькими кодонами ДНК (мРНК). В универсальном генетическом коде 61 кодон ДНК (мРНК) кодирует 20 протеиногенных аминокислот, а 3 кодона являются терминальными.

## Г

**GARP** — группа аминокислотных остатков, которые кодируются кодонами, содержащими гуанин или цитозин в первых и во вторых положениях (GC-богатыми кодонами). К этой группе относятся глицин, аланин, аргинин и пролин.

**Геликазы (хеликазы)** — ферменты, вызывающие локальное раскручивание цепей ДНК (ДНК-хеликазы) с образованием репликационной вилки (в процессе репликации, транскрипции или репарации ДНК) или шпильки РНК (РНК-хеликазы) перед началом трансляции.

**Ген** — участок ДНК (РНК у некоторых вирусов), ответственный за определенную специфическую функцию. Это самое широкое определение понятия «ген».

На современном этапе развития молекулярной генетики для эукариотических генов было сформулировано другое определение. **Эукариотический ген** — совокупность сегментов ДНК, которые вместе составляют экспрессируемую единицу, обуславливающую образование специфического функционального продукта (молекулы РНК или полипептида).

Эукариотические гены состоят из следующих сегментов ДНК: 1) единица транскрипции (кодирующий участок, интроны, 5'-лидерная и 3'-трейлерная последовательности зрелой мРНК, спейсеры); 2) последовательности, необходимые для начала транскрипции (промотор) и образования правильного 3'-конца зрелой мРНК; 3) цис-элементы — последовательности, регулирующие транскрипцию (ответственные за частоту инициации транскрипции, за тканевую, клеточную и временную специфичность транскрипции).

Транскрипция (и трансляция) эукариот регулируется не только цис-элементами, но и транс-элементами — белками или РНК (микроРНК), специфически связывающимися с определенными участками ДНК (или мРНК), влияя тем самым на процесс транскрипции (или трансляции). Последовательности ДНК, кодирующие транс-элементы, не подходят под приведенное выше определение эукариотического гена. У эукариот цис- и транс-элементы могут располагаться как в непосредственной близости от единицы транскрипции, которую они регулируют, так и в отдаленных от нее участках генома.

Под приведенное выше определение эукариотического гена не подходят также последовательности ДНК, влияющие на пространственную конфигурацию гена в хроматине и его топологию.

Для прокариотических организмов справедливо широкое определение понятия «ген». У них выделяют *структурные гены* и *гены-регуляторы*.

Прокариотические гены отличаются от эукариотических рядом особенностей: 1) единица транскрипции не содержит интронов; 2) промотор, как правило, является общим для нескольких структурных генов, расположенных последовательно в одной единице транскрипции; 3) операторы (фактически, аналоги цис-элементов эукариот) расположены, в основном, на расстоянии 50–75 нуклеотидных пар перед точкой инициации транскрипции.

**Генетико-автоматические процессы** — см. *случайный генетический дрейф*.

**Генетический груз** — насыщенность популяции или вида рецессивными генами, которые, будучи в гомозиготном состоянии, приводят (в той или иной степени) к снижению приспособленности особи к среде обитания.

**Генетический код** — система соответствия аминокислотных остатков (в белках) определенным триплетам в кодирующих участках ДНК (мРНК). Существует универсальный генетический код, который используется при расшифровке нуклеотидных последовательностей кодирующих участков ДНК у подавляющего большинства организмов, и несколько десятков неуниверсальных генетических кодов. Последние, как правило, имеют небольшое количество отклонений от универсального. Особенно часто встречаются два типа отклонений в неуниверсальных генетических кодах: 1) существование альтернативных иницирующих кодонов; 2) уменьшение числа терминальных кодонов.

**Генная мутация** — мутация, приводящая к изменению структуры отдельного гена за счет изменения его нуклеотидной последовательности.

**Геномная мутация** — мутация, приводящая к изменению числа хромосом в кариотипе.

**Генокопии** — организмы, сходные по своим фенотипам, но различные по генотипам.

**Геном** — совокупность генов, содержащихся в гаплоидном наборе хромосом. Термин используется для характеристики всего вида, а не отдельной особи. Геному противопоставляется *плазмон*. В последнее время (с увеличением количества записей о полностью расшифрованных нуклеотидных последовательностях генетического аппарата различных штаммов или подвидов организмов одного вида) термин «геном» утратил свою групповую специфичность. Стали вполне корректными выражения «в геноме данного штамма» и «в геноме данного организма».

**Генотип** — совокупность всех генов отдельного организма. В вирусологии и микробиологии под термином «генотип» подразумевают характеристику определенной внутривидовой группы. Схожие друг с другом по нуклеотидной последовательности определенных генов (по которым проводят их внутривидовую классификацию) вирусы объединяются в генотип, которому присваивается определенное название.

**Генофонд** — совокупность генов всех особей, составляющих популяцию. Набор особенностей генотипов всех особей, обеспечивающий наследственную изменчивость вида в целом.

**Геномный островок** — участок «хромосомы» прокариотического организма, GC-насыщенность в котором существенно отличается от GC-насыщенности в основной части «хромосомы».

**Гены-регуляторы** — гены, кодирующие белки-регуляторы. Белки-регуляторы связываются с операторами, регулируя тем самым процесс транскрипции у прокариот.

**Гены-мутаторы** — гены, продукты которых могут воздействовать на интенсивность мутационного процесса, повышая частоту возникновения мутаций в других генах. Чаще всего продуктами генов-мутаторов являются

ферменты, участвующие в репликации, транскрипции или репарации ДНК. Существование процесса мутационного давления было впервые экспериментально подтверждено на следующем примере.

Под воздействием свободных радикалов дезоксигуанинтрифосфат (дГТФ) окисляется с образованием 8-оксогуанинтрифосфата (8-оксо-ГТФ). При репликации ДНК 8-оксо-ГТФ часто оказывается в дочерней цепи напротив аденина. Специфический фермент, продукт гена *MutY*, вырезает аденин, оказавшийся в паре с 8-оксо-Г. Напротив 8-оксо-Г встраивается цитозин. Еще один специфический фермент, *MutM*, вырезает сам 8-оксо-Г, после чего на место последнего встраивается Г. В результате происходит трансверсия по направлению АТ на GC. Функцией фермента, кодируемого геном *MutT*, является элиминация 8-оксоГТФ из пула нуклеотидов. Штаммы *E. coli*, в которых ген *MutT* был дефективным, в ряду поколений действительно накапливали больше мутаций А на С (то есть у них повышалось общее GC-содержание генома).

**Гетерогенная картина замен нуклеотидов** — картина нуклеотидных замен в сравниваемых нуклеотидных последовательностях, при которой последние существенно отличаются друг от друга. Предполагается, что большое число различий между гомологичными последовательностями является результатом разницы в скоростях фиксации в них нуклеотидных замен. Такое отклонение от гипотезы молекулярных часов требует коррекции методов определения эволюционных дистанций. При гетерогенной картине нуклеотидных замен в сравниваемых последовательностях следует пользоваться модифицированными методами вычисления эволюционных дистанций между ними.

**Гетерогенная картина замен аминокислотных остатков** — картина аминокислотных замен в сравниваемых аминокислотных последовательностях, при которой последние существенно отличаются друг от друга. Предполагается, что большое число различий между гомологичными последовательностями является результатом разницы в скоростях фиксации в них аминокислотных замен. Такое отклонение от гипотезы молекулярных часов требует коррекции методов определения эволюционных дистанций. При гетерогенной картине аминокислотных замен в сравниваемых последовательностях следует пользоваться модифицированными методами вычисления эволюционных дистанций между ними.

**Гетерозигота** — организм, гомологичные хромосомы которого несут различные аллели того или иного гена.

**Гетерозис** — явление превосходства «ценных» качеств у гибридов первого поколения, полученных при скрещивании неродственных особей (или особей, находящихся в очень далеком родстве). При таком скрещивании у гибридов первого поколения повышается процентное содержание аллелей, находящихся в гетерозиготном состоянии. В последующих поко-

лениях эффект заметно ослабляется или даже переходит в свою противоположность.

**Гипотеза Гудмена** — гипотеза, признающая дубликацию генов ведущим механизмом возникновения функциональной дивергенции между ними. Была использована М. Кимурой для создания теории нейтральной эволюции (четвертое положение данной теории).

**Гипотеза молекулярных часов (авторы — Э. Цукеркэндл и Л. Полинг)** — гипотеза, постулирующая постоянство скорости молекулярной эволюции любого белка в разных филогенетических линиях при его неизменной третичной структуре и функции. Используется при вычислении эволюционных дистанции в случае гомогенной картины замещений в сравниваемых последовательностях; при построении дендрограмм методом попарного невзвешенного кластерирования с арифметическим усреднением (UPGMA).

**Гистоны** — низкомолекулярные высококонсервативные белки, участвующие в упаковке ДНК. Гистоны богаты положительно заряженными аминокислотами.

**G-квардуплекс** — разновидность вторичной структуры, которую могут образовывать обогащенные гуанином участки молекул ДНК и РНК. Образуется G-квардуплекс за счет водородных связей между четырьмя молекулами гуанина. Все четыре молекулы гуанина могут находиться как в одной цепи ДНК или РНК, так и в двух, трех и даже в четырех отдельных цепях. Квартет гуанинов обычно стабилизируется одновалентным катионом (особенно часто — ионом калия). G-квардуплекс впервые был обнаружен на концах хромосомной ДНК (в теломерах), но, естественно, может встречаться и в любой другой молекуле ДНК или РНК.

**Гликопротеины** — гликозилированные белки — сложные молекулы, представляющие собой белки, к которым присоединены остатки углеводов. У вирусов гликопротеины располагаются на поверхности вирионов.

**Гомогенная картина замен нуклеотидов** — картина нуклеотидных замен в сравниваемых нуклеотидных последовательностях, при которой последние несущественно отличаются друг от друга. Предполагается, что малое число различий между гомологичными последовательностями является результатом стабильной одинаковой скорости фиксации в них нуклеотидных замен. При выполнении гипотезы молекулярных часов в сравниваемых последовательностях не следует пользоваться модифицированными методами вычисления эволюционных дистанций между ними.

**Гомогенная картина замен аминокислотных остатков** — картина аминокислотных замен в сравниваемых аминокислотных последовательностях, при которой последние несущественно отличаются друг от друга. Предполагается, что малое число различий между гомологичными последовательностями является результатом стабильной одинаковой скорости

фиксации в них аминокислотных замен. При выполнении гипотезы молекулярных часов в сравниваемых последовательностях не следует пользоваться модифицированными методами вычисления эволюционных дистанций между ними.

**Гомозигота** — организм, гомологичные хромосомы которого несут одинаковые аллели того или иного гена.

**Гомологичные гены** — гены, имеющие общего эволюционного предшественника, возникшие, предположительно, вследствие его амплификации. Отдаленные гомологи могут серьезно отличаться друг от друга по своей функции и активности, сохраняя определенное сходство в строении, в связи с чем гомологичные гены и белки разделяют на *ортологов* и *парологов*.

**Группа сцепления** — совокупность генов одной хромосомы, которые наследуются совместно (в отсутствии кроссинговера).

**Гуанин (2-амино-6-оксипуридин)** — пуриновое азотистое основание, входящее в состав нуклеотидов ДНК и РНК.

**ГЦ-давление (GC-давление)** — направление мутационного давления в ДНК, при котором частота возникновения и фиксации замен аденина и тимина на гуанин и цитозин превышает частоту замен гуанина и цитозина на аденин и тимин. Термин справедлив для организмов, геном которых представлен двухцепочечной ДНК, в которой гены располагаются на обеих цепочках. GC-давление приводит к закономерному увеличению GC-насыщенности ДНК в ряду поколений.

О силе GC-давления в данном геноме (или в данной изохоре) можно судить по уровню его (ее) GC-насыщенности. При GC-насыщенности в пределах от 50 до 60 % следует предполагать наличие слабого GC-давления в данном геноме (или в данной изохоре); при GC-насыщенности более 60 % GC-давление считается сильным.

Более точным критерием сильного GC-давления в геноме является низкий коэффициент корреляции ( $R < 0,5$ ) между значениями  $3GC$  и  $G+C$  для всех кодирующих участков данного генома (при высоком значении  $3GC$ ) [84].

**ГЦ-насыщенность генома (GC-насыщенность, GC-содержание, G+C)** — доля гуанина и цитозина в геноме относительно всех нуклеотидов, входящих в состав ДНК. Гуанин и цитозин комплементарны друг другу, следовательно, в двухцепочечной ДНК частоты их использования должны быть одинаковыми и изменяться сочетанно. Уровень GC-насыщенности в геноме является следствием фиксации нуклеотидных замен, происходящих, преимущественно, под воздействием одного из направлений мутационного давления (AT- или GC-давления). GC-насыщенность влияет на физические свойства ДНК: при повышении количества GC-пар увеличивается температура плавления и устойчивость дуплекса ДНК.

GC-насыщенность — один из таксономических критериев при классификации бактерий.

**ГЦ-насыщенность гена** — доля гуанина и цитозина в кодирующем участке гена относительно всех нуклеотидов, входящих в состав ДНК данного кодирующего участка. Для большинства организмов GC-насыщенность каждого отдельного гена примерно равна GC-насыщенности генома. Традиционно считается, что данное равенство не выполняется у рептилий, птиц и млекопитающих вследствие присутствия в их геномах *изохор*. Однако даже у прокариотических организмов (бактерий и архей) существуют геномные островки, существенно отличающиеся по своей GC-насыщенности от большей части генома. К тому же, GC-насыщенность может довольно широко варьировать в пределах одного гена.

Список показателей, отражающих **GC-насыщенность**:

– **G+C** — общая насыщенность гена или генома гуанином и цитозином;

– **1GC, 2GC, 3GC** — насыщенность гуанином и цитозином первых, вторых и третьих положений кодонов генов или кодирующего генома, соответственно;

– **GC4f** — насыщенность гуанином и цитозином четырехкратно вырожденных сайтов;

– **GC2f3p** — насыщенность гуанином и цитозином двукратно вырожденных сайтов, расположенных в третьих положениях кодонов;

– **GC0f** — насыщенность гуанином и цитозином нолькратно вырожденных сайтов.

**Гэпы** — см. *инделы*.

## Д

**Дальтон** — единица измерения массы вещества, равная массе атома водорода ( $1,661 \times 10^{-24}$  грамм). В дальтонах измеряется масса молекул *генетических биополимеров*.

**Дарвинизм** — течение в эволюционном учении, созданное приверженцами теории эволюции Ч. Дарвина. По Ч. Дарвину, основной движущей силой эволюции является *естественный отбор*, возникающий в результате борьбы за существование на основе наследственности и изменчивости.

**Двукратно вырожденный сайт** — положение в нуклеотидной последовательности, в котором одна из трех возможных замен является синонимичной. Фактически, двукратно вырожденными сайтами являются третьи положения кодонов двукратно вырожденных серий и первое положение в триплетях, кодирующих *дуплеты* лейцина и аргинина.

**Двукратно вырожденная серия кодонов** — серия кодонов, состоящая из двух синонимичных триплетов.



**Двухцепочечная альфа-спирализованная суперспираль** (англ. «**coiled coil**») — надвторичная структура белка, формирующаяся из двух альфа-спиральных белков, скрученных друг с другом.

**Двухшаговая аминокислотная замена** — аминокислотная замена, вызванная заменой двух нуклеотидов в исходном кодоне.

**Делеция** — разновидность мутации, при которой утрачивается часть генетического материала. Может происходить на уровне хромосомы или внутри одного гена. В последнем случае в массиве выровненных последовательностей делетированный участок гена называется *гэпом*.

**Дендрограмма (филогенетическое дерево)** — условное графическое изображение (чертеж), отражающее предполагаемые родственные связи между генетическими макромолекулами, биологическими видами или более высокими таксонами. С помощью дендрограммы осуществляется визуализация филогенетических взаимоотношений между ними. Частями дендрограммы являются корень, ветви и последовательности (зачастую именуемые «листьями»). Взаимное расположение ветвей называется *топологией*. Группа ветвей называется *кластером*. Длина ветви — расстояние от корня или от последнего разветвления до ее конца. Под дендрограммой обычно указывается ее масштаб — отрезок, равный определенному эволюционному расстоянию. Существует традиционная, радиальная и круговая формы построения дендрограмм. Среди всех ныне существующих методов построения дендрограмм выделяются две основные группы: а) *дистанционно-матричные методы*; б) *дискретные методы*.

**Дивергенция** — в макроэволюции: дивергенция — процесс расхождения признаков в родственных таксонах. В микроэволюции: дивергенция — процесс расхождения признаков, приводящий к образованию новых группировок особей внутри вида. Молекулярно-биологическая трактовка: дивергенция — процесс увеличения числа различий в кодирующих участках гомологичных генов, приводящий к закономерному увеличению числа различий в аминокислотных последовательностях кодируемых ими белков.

**Динуклеотид** — два расположенных друг за другом нуклеотида в ДНК или в РНК. Некоторые ферменты специфически связываются с определенными динуклеотидами, вызывая модификацию нуклеотидов в них.

**Дискретные методы построения дендрограмм** — дискретные методы имеют дело непосредственно с массивом выровненных нуклеотидных или аминокислотных последовательностей, а не с коэффициентами их сходства, полученными на основе вычисления эволюционных дистанций. Сам же алгоритм вычисления эволюционных дистанций встроены в метод построения дендрограммы. Один из этих методов называется методом *максимальной экономии (MP)*.

**Дистанционно-матричные методы построения дендрограмм** — эта группа методов основана на использовании результатов вычисления эво-

люционных дистанций. Топология и длины ветвей дендрограмм рассчитываются с учетом полученных ранее коэффициентов сходства.

**Дистанции эволюционные** — среднее число нуклеотидных (аминокислотных) замен, приходящихся на одно нуклеотидное (аминокислотное) положение в сравниваемых гомологичных генах (белках). Существует целый набор *методов определения эволюционных дистанций*.

**Дезоксирибонуклеиновая кислота (ДНК)** — высокомолекулярный генетический биополимер, обеспечивающий хранение и наследование генетической информации у живых организмов. У всех живых организмов, за исключением некоторых групп вирусов, ДНК состоит из двух комплементарных друг другу антипараллельных цепочек, образующих двойную спираль. Каждая из цепочек представляет собой неразветвленную полинуклеотидную цепь, нуклеотиды в которой состоят из азотистого основания (аденина, гуанина, цитозина или тимина), дезоксирибозы и остатка фосфорной кислоты.

**ДНК-полимеразы** — ферменты, осуществляющие праймер-зависимую, детерминируемую ДНК-матрицей реакцию присоединения дезоксирибонуклеотидов в процессе репликации и репарации.

**ДНК-лигазы** — ферменты, осуществляющие соединение фрагментов вновь синтезированной цепи ДНК при репликации и репарации.

**Доминантный аллель** — аллель, проявляющийся фенотипически как в гомо-, так и в гетерозиготном состоянии (в отсутствии неполного доминирования).

**Донорная точка сплайсинга** — см. *левая граница сплайсинга*.

**Дрейф генов** — случайные колебания частот генов. Термин, применимый исключительно для характеристики эволюционного процесса в малых популяциях размножающихся половым путем организмов. Следует дифференцировать с термином «*Случайный генетический дрейф (генетико-автоматические процессы)*», который обозначает процесс фиксации нейтральных мутаций, происходящих на молекулярно-генетическом уровне. Дрейф генов является одним из процессов, вносящих свой вклад в случайный генетический дрейф.

**Дуплет** — два кодона из *шестикратно вырожденных серий* (лейцин, серин, аргинин), отличающиеся друг от друга нуклеотидами в третьем положении.

**Дупликация** — мутация, при которой происходит удвоение участка хромосомы (то есть одного, нескольких генов, какого-либо участка гена). См. также *амплификация, гипотеза Гудмена*.

## Е

**Единица эволюционного времени** — единица, равная времени (в млн. лет), необходимому для осуществления дивергенции последовательностей, приводящей к уменьшению их сходства на 1 %.

**Естественный отбор** — движущая сила эволюции (по Ч. Дарвину), возникающая в результате борьбы за существование на основе наследственности и изменчивости. Естественный отбор, по автору, выражается в избирательном выживании наиболее приспособленных к условиям среды особей, то есть особей, обладающих приспособительными (положительными) наследственными признаками. В синтетической теории эволюции выделяют три вида естественного отбора: 1) *отрицательный* (стабилизирующий, негативный) отбор; 2) *направленный* (движущий) отбор; 3) *дисруптивный* (расщепляющий) отбор. К понятию о естественном отборе Ч. Дарвин пришел через наблюдение результатов отбора искусственного (то есть отбора, осуществляемого человеком среди домашних животных и культурных растений).

### З

**Закон Харди–Вайнберга** — закон, формулирующийся следующим образом: в идеальной популяции соотношение частот генов и генотипов (гетерозигот, доминантных и рецессивных гомозигот) не изменяется в ряду поколений. Для соблюдения этого закона необходимо, чтобы каждая особь вносила свой вклад в генофонд последующих поколений.

**Закрытая рамка считывания** — рамка считывания, не позволяющая считывать информацию со всего кодирующего участка мРНК (от иницирующего кодона до терминального). При использовании закрытой рамки считывания внутри кодирующего участка образуются терминальные кодоны.

**Запаздывающая (отстающая) цепь ДНК** — цепь ДНК, репликация которой происходит отдельными фрагментами (фрагментами Оказаки) в направлении, противоположном репликации лидирующей цепи, по мере раскручивания спирали ДНК.

**Затравка** — см. *праймер*.

**Z-тест** — статистический тест, который часто используется для определения статистической достоверности результатов *селекционных тестов*.

### И

**Избыточность** — см. *вырожденность*.

**Изменчивость** — явление неполного сходства родителей и их потомков. Выделяют наследственную (генотипическую) и модификационную (фенотипическую) изменчивость.

**Изоакцепторные кодоны** — см. *синонимичные кодоны*.

**Изокодонная замена** — см. *синонимичная замена*.

**Изоляция** — важный фактор микроэволюции, заключающийся в возникновении разного рода преград (географических, экологических, репродуктивных) свободному скрещиванию особей внутри популяции. Изоляция приводит к возникновению эффекта родоначальника.

**Изохора** — участок ДНК, характеризующийся равномерным содержанием гуанина и цитозина. Между собой изохоры существенно различаются по GC-насыщенности (могут быть GC-бедными или GC-богатыми). По мнению Н. Суеоки и К. Дж. Фриксела [68], изохора — это участок генома, на который распространяется мутационное давление определенного направления. По мнению Дж. Бернарди (первооткрывателя изохор), изохоры появились в результате положительного естественного отбора: они необходимы для предотвращения вырезания дублированных генов [50]. Считается, что изохоры существуют в геномах рептилий, птиц и млекопитающих, однако недавно появились публикации, указывающие на существование изохор у земноводных и других представителей животного мира. У прокариотических организмов аналогами изохор (протоизохорами) можно назвать геномные островки.

**Иммунный ответ** — совокупность изменений в иммунной системе, направленных на сохранение постоянства внутренней среды организма и защиту его от чужеродных для него агентов. В частности, гуморальный иммунный ответ обеспечивается синтезом специфических антител В-лимфоцитами, а клеточный иммунный ответ обеспечивается синтезом специфических Т-клеточных рецепторов у Т-лимфоцитов.

**Инвариантные сайты (консервативные сайты)** — в массиве выровненных последовательностей инвариантными сайтами называют те положения нуклеотидов (аминокислоты), которые идентичны для всех сравниваемых последовательностей.

**Инверсия** — хромосомная мутация, заключающаяся в отрыве участка хромосомы, повороте его на  $180^\circ$  и прикреплении его к месту отрыва.

**Ингибитор (репрессор)** — белок-регулятор, который подавляет транскрипцию в опероне, связываясь с оператором. Может также связываться с *индуктором*.

**Инделы (пробелы, гэпы)** — образующиеся в массиве выровненных последовательностей разрывы, обусловленные инсерциями и делециями. Термин «индел» является комбинацией слов «инсерция» и «делеция». Необходимость введения такого термина вызвана тем, что практически невозможно определить, какой именно биологический процесс, инсерция или делеция, вызвал появление разрывов в массиве выровненных последовательностей.

**Индуктор** — вещество, способное конкурентно связываться с репрессором (ингибитором), освобождая оператор от связи с последним. В ре-

зультате появления в клетке индуктора запускается транскрипция соответствующего оперона.

**Инициация трансляции** — начальный этап трансляции, при котором происходит связывание рибосомы с мРНК и факторами начала трансляции, поступление первой тРНК в аминокислотный центр рибосомы. В большинстве случаев первой является тРНК с метионином, а первым триплетом кодирующего участка мРНК соответственно AUG. Тем не менее синтез белка может начинаться и с многих других триплетов (CUG, GUG, AUA и др.). Существует большое количество факторов инициации трансляции, часть из которых обладает РНК-хеликазной активностью.

**Инозин (гипоксантин)** — минорный пуриновый нуклеотид, являющийся продуктом дезаминирования аденина. Образует комплементарную связь, преимущественно, с цитозином (то есть является аналогом гуанина).

**Инсерция** — мутация, приводящая к увеличению длины ДНК или РНК за счет встраивания в данную ДНК или РНК чужеродного генетического материала.

**Интроны** — участки единицы транскрипции, которые вырезаются в процессе сплайсинга из про-мРНК. Существуют только у эукариот и у некоторых вирусов.

## К

**Канонические основания** — см. *мажорные основания*.

**Каноническая последовательность** — см. *консенсусная последовательность*.

**Капсидный белок** — вирусный белок, из которого состоит капсидная оболочка вируса, защищающая его генетический аппарат. Как правило, капсидная оболочка состоит из большого числа копий нескольких капсидных белков.

**Карбокситерминальный конец белка (С-конец)** — конец полипептидной цепи, имеющий свободную карбоксильную группу. С-конец соответствует последним аминокислотным остаткам в полипептиде (в процессе трансляции белок синтезируется от N-конца к С-концу).

**Кариотип** — совокупность постоянных морфологических признаков, по которым можно идентифицировать хромосомный набор (число, размер, форма хромосом) определенной таксономической группы (вида и более высоких таксонов) или отдельной особи.

**Квартет** — четыре кодона из *шестикратно вырожденных серий* (лейцин, серин и аргинин), отличающиеся друг от друга нуклеотидами в третьем положении.

**Кластер дендрограммы** — группа последовательностей (таксономических единиц), имеющих общего предшественника, не являющегося

предшественником для других последовательностей (таксономических единиц), не образующих данный кластер.

**Кодирующий участок ДНК** — участок ДНК, который несет информацию о первичной структуре рРНК, тРНК, мРНК (белка). Транскрипция происходит на цепи ДНК, комплементарной кодирующему участку.

**Кодон (триплет)** — три последовательно расположенных нуклеотида в кодирующем участке ДНК (мРНК). Триплет является наименьшей функциональной единицей гена, поскольку кодирует одну аминокислоту. В универсальном генетическом коде аминокислоты кодирует 61 триплет, 3 триплета являются *терминальными*.

**Комбинационная изменчивость** — разновидность наследственной изменчивости, возникающая в результате рекомбинации наследственных признаков родителей у потомков. Является результатом *кроссинговера*.

**Комплементарные нуклеотиды** — нуклеотиды, способные образовывать друг с другом водородные связи. Аденин комплементарен тимину и урацилу (образуют две водородные связи); гуанин комплементарен цитозину (образуют три водородные связи). Благодаря комплементарности нуклеотидов связываются две цепочки ДНК, происходит синтез РНК на кодирующей цепи ДНК, антикодон тРНК связывается с кодоном мРНК во время трансляции. Комплементарность — понятие относительное. Например, аденин преимущественно образует водородные связи с тимином и урацилом, но может (с меньшей вероятностью) связываться и со всеми остальными мажорными и многими минорными нуклеотидами.

**Конвергенция** — процесс приобретения сходства (в большей степени функционального, чем структурного) между органами (или отдельными генами) представителей неродственных таксонов. Результатом конвергенции считается формирование аналогичных органов (или ферментов).

**Консенсусная (каноническая, усредненная) последовательность** — искусственно создаваемая последовательность, характеризующая многочисленное семейство повторяющихся (родственных) последовательностей ДНК или белка. Создается на основе массива выровненных последовательностей. В каждом положении консенсусной последовательности находятся нуклеотиды (аминокислоты), наиболее часто встречающиеся в данном положении во всей совокупности родственных последовательностей.

**Консервативная аминокислотная замена** — замена аминокислоты в белке, которая не приводит к значительным изменениям структуры и свойств белка. Термин весьма относителен, поскольку существует большое количество противоречащих друг другу систем определения консервативности аминокислотных замен. Нами предложен способ определения степени консервативности (противопост. *радикальности*) аминокислотных

замен в отдельных белках и в протеомах на основе анализа частот их фиксации в популяциях исследуемых организмов [30, 31, 32, 83].

**Консервативный участок нуклеотидной последовательности (белка)** — последовательность нуклеотидов (аминокислот) в гомологичных генах, мало измененная в процессе эволюции. Как правило, к консервативным участкам относятся активные центры гомологичных ферментов.

**Консервативные сайты в выровненных последовательностях** — см. *инвариантные сайты*.

**Коэволюция** — длительное взаимодействие особей разных видов в процессе эволюции, приводящее к тому, что различные виды живых организмов становятся необходимыми друг другу для успешного существования (пример: покрытосеменные растения и насекомые-опылители). Коэволюция лежит в основе появления биоценозов.

**Критерий отсутствия отбора** — в селекционных тестах так называется картина нуклеотидных замен в гомологичных генах, при которой синонимичная дистанция (число фиксированных синонимичных замен деленное на число синонимичных сайтов) равна несинонимичной (число фиксированных несинонимичных замен деленное на число несинонимичных сайтов) дистанции. Авторы селекционных тестов считают, что такая картина замен предполагает одинаковую степень нейтральности синонимичных и несинонимичных замен. Тем не менее с позиции теории нейтральности, такая картина не говорит о полном отсутствии отрицательного отбора. Кроме того, при *сильном мутационном давлении* частота возникновения несинонимичных замен гораздо больше, чем частота возникновения синонимичных, поскольку большая часть замен по третьему положению кодонов в таких условиях произойти не может [82].

**Кроссинговер** — особенность конъюгации хромосом в мейозе, заключающаяся в том, что хроматиды, принадлежащие к гомологичным хромосомам, перекрещиваются. Когда конъюгация заканчивается и хромосомы расходятся, в местах перекреста могут образовываться разрывы, приводящие к обмену участками хромосом. В результате кроссинговера образуются гаметы с новыми комбинациями генов.

**КЭП** — структура на 5'-конце эукариотических мРНК, образующаяся после транскрипции. КЭП представляет собой метилированный гуанозин, присоединенный к первому нуклеотиду мРНК. Вероятнее всего, КЭП необходим для начала инициации трансляции: соединения мРНК с 40S-субъединицей рибосомы.

## Л

**Ламаркизм** — учение, основанное на одной из первых эволюционных теорий, изложенной Ж. Б. Ламарком в труде «Философия зоологии» (1809). Теория основана на предположении о том, что «натренирован-

ность» определенных органов (например, способность жирафа вытягивать шею в процессе поиска свежих листьев, которые расположены на верхних ветвях деревьев) передается по наследству последующим поколениям. В труде «Философия зоологии» высказана идея о том, что развитие происходит от низкоорганизованных видов к высокоорганизованным. В основе же усложнения (эволюции) лежит «врожденное стремление к самосовершенствованию», на «правильность» которого оказывает влияние окружающая среда.

**Латеральный перенос** — передача наследственной информации (в различном объеме) от одного вида к другому. Особенно хорошо латеральный перенос изучен и доказан на примере переноса генов хозяина в геном поражающих его вирусов.

**Левая граница сплайсинга (донорная точка сплайсинга)** — граница между концом экзона и началом интрона. 5'-конец интрона (начало интрона), как правило, консервативен: содержит гуанин и урацил.

**Летальные мутации** — мутации, несовместимые с жизнью.

**Лигазы (синтетазы)** — класс ферментов, катализирующих присоединение к какому-либо субстрату различных органических молекул с использованием энергии АТФ. Одним из представителей лигаз является *ДНК-лигаза*.

**Лидирующая цепь ДНК** — цепь ДНК, *репликация* которой происходит непрерывно на матрице 5'–3' цепи ДНК. Репликация лидирующей цепи не прерывается, поскольку направление работы ДНК-полимеразы совпадает с направлением, в котором раскручивается матрица ДНК.

**Локус** — место расположения гена в хромосоме.

## М

**Мажорные основания** — азотистые основания, наиболее часто входящие в состав нуклеотидов ДНК и РНК (аденин, гуанин, цитозин, тимин, урацил). Существуют также *минорные основания*.

**Макромолекулы генетические** — сборное понятие, обозначающее молекулы биополимеров, вовлеченные в процесс хранения и реализации генетической информации (ДНК, РНК, белки).

**Макроэволюция** — надвидовая эволюция. Эволюционные преобразования, приводящие к возникновению или исчезновению таксонов более высокого ранга, чем вид.

**Масштаб дендрограммы** — отрезок, равный определенной эволюционной дистанции. Располагается под дендрограммой.

**Матричная рибонуклеиновая кислота (мРНК)** — макромолекулярный генетический биополимер, мономерами которого являются нуклеотиды, состоящие из азотистого основания (аденин, гуанин, цитозин или урацил), рибозы и остатка фосфорной кислоты. Матричная РНК является



одноцепочечной, однако может образовывать вторичные структуры (шпильки и др.) за счет комплементарного взаимодействия друг с другом отдельных групп нуклеотидов. В эукариотических организмах на матрице цепи ДНК синтезируется про-мРНК, содержащая участки, соответствующие интронам и экзонам ДНК; далее про-мРНК подвергается процессингу: *кэпированию, полиаденилированию, сплайсингу* (вырезанию участков, соответствующих интронам, и соединению транскриптов с участками, соответствующих экзонам) с образованием собственно мРНК. У прокариот процессинг мРНК отсутствует.

**Методы определения эволюционных дистанций между нуклеотидными последовательностями (наиболее часто используемые):**

– *Джукса–Кантора* — однопараметрический метод, основанный на равенстве вероятностей фиксации всех возможных нуклеотидных замен;

– *Тадзимы–Нея* — в данном методе отдельно рассчитываются скорости замен на каждый из четырех нуклеотидов. Вероятности транзиций и трансверсий считаются равными;

– *Кимуры* — двухпараметрический метод, в котором при расчете эволюционных дистанций учитываются: 1) скорость транзиций, 2) скорость трансверсий;

– *Тамуры* — трехпараметрический метод, в котором при расчете эволюционных дистанций учитываются: 1) скорость транзиций, 2) скорость трансверсий, 3) GC-насыщенность;

– *Тамуры–Нея* — в данном методе при расчете эволюционных дистанций учитываются: 1) скорость всех трансверсий, 2) скорость транзиций аденина на гуанин, 3) скорость транзиций цитозина на урацил. В данном методе скорости замен А на Г (Ц на У) и Г на А (У на Ц) считаются равными.

Все вышеперечисленные методы были разработаны с целью корректной оценки степени дивергенции между гомологичными генами. В двух- и трехпараметрических методах разным типам нуклеотидных замещений придается различное значение. Например, по методу Кимуры, скорость трансверсий вносит в два раза больший вклад в величину вычисляемой эволюционной дистанции, чем скорость транзиций.

**Метод максимальной экономии (MP — Maximum parsimony)** — один из дискретных методов построения дендрограмм. Коэффициенты сходства, на основании которых строятся промежуточные дендрограммы, определяются только по тем положениям в выровненных последовательностях, в которых минимум два разных нуклеотида (или две разные аминокислоты) встречаются хотя бы два раза. Для каждой последовательности в каждой промежуточной дендрограмме рассчитывается минимальное число замен, которое могло бы привести к дивергенции. Из множества

промежуточных дендрограмм выбирается та, в которой общая сумма возможных замен для всех последовательностей является минимальной.

**Метод минимальной эволюции (ME — Minimum evolution)** — один из дистанционно-матричных методов построения дендрограмм, основанный на построении множества промежуточных дендрограмм *методом связывания ближайших соседей (NJ)*. Каждая из промежуточных дендрограмм отличается от остальных первой последовательностью, с которой начинается построение NJ-дендрограммы. Конечной дендрограммой является та, в которой сумма всех ветвей минимальна.

**Метод попарного невзвешенного кластерирования с арифметическим усреднением (UPGMA — Unweighted pair-group method using arithmetic averages)** — один из дистанционно-матричных методов построения дендрограмм, основанный на гипотезе молекулярных часов. В процессе построения дендрограммы пары последовательностей, между которыми эволюционные дистанции минимальны, группируются в первую очередь и образуют общий кластер; эволюционное расстояние между кластерами равно среднему от коэффициентов сходства каждой последовательности из одного кластера с каждой последовательностью из другого. Использовать UPGMA рекомендуется только для выяснения эволюционных отношений между генами с гомогенной картиной нуклеотидных замещений.

**Метод связывания ближайших соседей (NJ — neighbor-joining)** — один из дистанционно-матричных методов построения дендрограмм. В начале построения NJ-дендрограммы выбирается одна из последовательностей и сравнивается поочередно со всеми остальными. По результатам сравнений определяется наиболее сходная с ней последовательность, после чего устанавливается сходство между данной парой последовательностей и всей остальной выборкой. Далее такая же процедура производится для всех оставшихся последовательностей до момента исчерпания выборки.

**Миграция** — перемещение отдельных групп животных из одной популяции в другую (или перемещение целых популяций). Может рассматриваться как фактор микроэволюции, поскольку посредством миграции происходит изменение генофонда популяции (одни генотипы исчезают, другие появляются).

**Микроэволюция** — совокупность эволюционных процессов, протекающих в популяциях.

**Минорные основания** — азотистые основания, редко входящие в состав нуклеотидов ДНК и РНК (например инозин, 5-формил-урацил и 8-оксогуанин). К таковым относятся тем или иным образом модифицированные *мажорные основания*.

**Множественное выравнивание** — выравнивание трех и более нуклеотидных или аминокислотных последовательностей гомологичных генов (белков).

**Модификационная изменчивость** — изменчивость, возникающая в результате воздействия на организм факторов внешней среды, приводящего к вариациям признаков в пределах их генетически детерминированной нормы реакции.

**Модулирующий кодон** — определенный кодон мРНК, комплементарный антикодону аминоксил-тРНК, содержащейся в малом количестве в момент трансляции (в результате малого содержания соответствующей аминокислоты). Благодаря наличию нескольких модулирующих кодонов в определенном участке мРНК может происходить временная задержка трансляции. Модулирующие кодоны играют решающую роль в процессе *аттенуации*.

**Молекулярная эволюция** — раздел эволюции, изучающий изменения, происходящие в процессе филогенеза исключительно в молекулах генетических биополимеров. Значительная часть таких изменений либо вообще не проявляется фенотипически, либо проявляется крайне слабо.

**Молчащая мутация** — мутация, которая не проявляется фенотипически.

**Мультигенное семейство** — структурно и функционально тесно взаимосвязанные гены, произошедшие от одного и того же предкового гена путем амплификации, претерпевшие в дальнейшем определенные мутационные изменения.

**Моноцистронный транскрипт** — молекула мРНК, содержащая один кодирующий участок.

**Мутагенный фактор** — фактор внешней или внутренней среды, вызывающий возникновение мутаций.

**Мутагенез** — процесс возникновения мутаций.

**Мутационное давление** — фактор молекулярной эволюции, дающий материал для естественного отбора и случайного генетического дрейфа. Мутационное давление обусловлено повышенной частотой возникновения нуклеотидных замещений аденина и тимина на гуанин и цитозин относительно частоты возникновения замен гуанина и цитозина на аденин и тимин (*GC-давление*) или наоборот (*AT-давление*). Наиболее чувствительными к мутационному давлению являются третьи положения кодонов в кодирующих участках. При уровне  $3GC$  выше 0,5 можно предположить наличие *GC-давления* в данном гене (геноме), при уровне  $3GC$  ниже 0,5 — *AT-давления*. Причиной возникновения мутационного давления может быть не только высокая частота возникновения мутаций, но и дефект в работе ферментов системы репарации. Основными биохимическими процессами, вызывающими мутационное давление, являются окисление и дезаминиро-

вание нуклеотидов (ферментативное и спонтанное). Мутационное давление было открыто в 60-х гг. XX в. и послужило основой для создания *теории Н. Суеоки*.

**Мутации** — внезапно возникающие стойкие изменения наследственных структур. Мутации в половых клетках (гаметические мутации) передаются по наследству при половом размножении; мутации в соматических клетках (соматические мутации) передаются по наследству только в случае вегетативного размножения.

## Н

**Направленный отбор** — разновидность естественного отбора (согласно синтетической теории эволюции). Направленный отбор предполагает адаптивность менее распространенного в популяции признака, который постепенно становится наиболее распространенным. При этом наиболее распространенный ранее признак подвергается воздействию отрицательного отбора. Если адаптивными в изменившихся условиях среды оказываются сразу несколько вариантов признака (а не один, как при направленном отборе), то эту ситуацию в синтетической теории эволюции называют *дизруптивным (расщепляющим) отбором*.

**Наследственность** — свойство живых организмов передавать свои признаки следующим поколениям, осуществляемое благодаря способности ДНК (РНК у некоторых вирусов) к репликации.

**Наследственная изменчивость** — изменчивость, связанная с изменением самих генов, механизмов регуляции их экспрессии или с возникновением новых их сочетаний. Такая изменчивость наследуется последующими поколениями.

**Невырожденные серии кодонов** — кодоны, не имеющие синонимичных себе триплетов, например, кодирующие триптофан и метионин.

**Нейтрализм** — течение в молекулярной эволюции, основанное на теории нейтральности М. Кимуры. Основными факторами эволюции, действующими на молекулярном уровне, в данной теории признаются *отрицательный естественный отбор* и *случайный генетический дрейф*.

**Нейтральные аминокислотные замены** — данный термин эквивалентен термину «консервативные аминокислотные замены», противопоставляется термину «радикальные аминокислотные замены». По теории нейтральности М. Кимуры подавляющее большинство зафиксированных в популяции аминокислотных замен являются нейтральными или почти нейтральными (слабо отрицательными). Такие аминокислотные замены имеют шанс быть зафиксированными в популяции случайным генетическим дрейфом. Отрицательные аминокислотные замены (*радикальные замены*) элиминируются из популяции *отрицательным естественным отбором*.

**Неперекрываемость** — свойство генетического кода, заключающееся в наличии одной рамки считывания в кодирующем белок участке гена. У многих вирусов тем не менее рамки считывания перекрываются (чаще — частично, реже — полностью).

**Несинонимичная эволюционная дистанция** — эволюционная дистанция, рассчитанная исключительно на основании скорости фиксации несинонимичных замен на несинонимичный сайт в сравниваемых нуклеотидных последовательностях. Перед непосредственным вычислением несинонимичной эволюционной дистанции необходимо воспользоваться одним из методов, рассчитывающих число *синонимичных* и *несинонимичных сайтов* и соответствующих замен в сравниваемых последовательностях.

**Несинонимичная нуклеотидная замена** — нуклеотидная замена в кодирующем участке ДНК (мРНК), приводящая к замене кодона, в котором она произошла, на несинонимичный, то есть к замене аминокислоты в кодируемом белке.

**Несинонимичный сайт** — сайт нуклеотидной последовательности, в котором хотя бы одно из трех возможных нуклеотидных замещений является несинонимичным. Если все замены в данном сайте несинонимичны, он называется абсолютно несинонимичным. Если одна из замен в данном сайте несинонимична, он называется несинонимичным на  $\frac{1}{3}$  (соответственно, синонимичным на  $\frac{2}{3}$ ). Если две замены в данном сайте несинонимичны, он называется несинонимичным на  $\frac{2}{3}$  (соответственно, синонимичным на  $\frac{1}{3}$ ).

**Невырожденный (нолькратно вырожденный) сайт** — так называется сайт нуклеотидной последовательности, в котором любая замена является несинонимичной.

**Нонсенс-замена** — нуклеотидная замена в кодирующем участке ДНК (мРНК), приводящая к возникновению терминального кодона (из претерминального, если замена одношаговая).

**Нонсенс-кодон** — см. *терминальный кодон*.

**Норма реакции** — пределы модификационной изменчивости признака, то есть свойство данного генотипа проявлять индивидуальную изменчивость под воздействием условий среды в определенных пределах.

**Нуклеоид** — центральная часть генетического аппарата прокариот, представляющая собой одну кольцевую молекулу ДНК в комплексе с негистоновыми белками. В настоящее время нуклеоид прокариот часто называют «хромосомой», особенно, если в геноме прокариотического организма присутствуют два нуклеоида. У некоторых прокариот нуклеоид представляет собой линейную «хромосому» [73]. Допускается, что нуклеоид может менять свою форму (с кольцевой на линейную и обратно).

**Нуклеотиды** — химические соединения, состоящие из 1) *азотистого основания*, 2) моносахарида, 3) остатка фосфорной кислоты. Являются

мономерами для ДНК и РНК. Также нуклеотиды входят в состав НАД, НАДФ, АТФ и других биологически активных соединений.

**Нуклеотидная пара** — соединенные между собой водородными связями нуклеотиды в двойной спирали ДНК. Правильные нуклеотидные пары: аденин и тимин, гуанин и цитозин. Наряду с этими парами в ДНК могут образовываться и неправильные. Неправильные нуклеотидные пары являются субстратом для ферментов из системы репарации.

## О

**Обратная транскрипция** — процесс синтеза ДНК на матрице РНК. Осуществляется с помощью фермента обратной транскриптазы (ревертазы), присущей ретровирусам. Ген обратной транскриптазы входит также в состав эукариотических *ретротранспозонов*.

**Обратная трансляция** — гипотетический процесс синтеза молекулы мРНК на основе короткой последовательности аминокислот при помощи особого рода тРНК. В литературе есть единичные сведения о возможности такого процесса в митохондриях макрофагов [62], однако это является не более чем одним из гипотетических механизмов реализации генетической информации.

**Однозначность** — свойство генетического кода, заключающееся в том, что один кодон может кодировать только одну аминокислоту. Это свойство плохо сочетается с гипотезой о «качающемся» взаимодействии антикодонов тРНК с кодонами мРНК во время трансляции. Если следовать этой гипотезе и предположить, что модифицированные нуклеотиды в первых положениях антикодонов тРНК могут взаимодействовать с несколькими видами нуклеотидов, расположенных в третьих положениях кодонов мРНК, то почему не допустить вероятность появления модифицированных нуклеотидов в первых и во вторых положениях антикодонов тРНК?

**Одношаговая аминокислотная замена** — аминокислотная замена, вызванная заменой одного из трех нуклеотидов в исходном кодоне.

**Опал-кодон** — общепринятое название *терминального* триплета TGA (UGA — в РНК).

**Опал-мутация** — мутация, приводящая к возникновению в *кодирующей* участке ДНК (мРНК) триплета TGA (UGA — в мРНК).

**Опал-супрессор** — ген, кодирующий тРНК, антикодон которой комплементарен кодону UGA (опал-кодону). В результате опал-супрессии возникает отклонение от универсального генетического кода. Например у микоплазм терминальный кодон UGA кодирует аминокислоту триптофан. Опал-супрессор возникает в результате мутации одного из генов, кодирующих тРНК, связывающую триптофан. Также у некоторых организмов опал-кодон может кодировать селеноцистеин или цистеин.

**Оператор** — участок ДНК, регулирующий транскрипцию у прокариот. Представляет собой участок связывания регуляторного белка. Оператор располагается в непосредственной близости к промотору. Оператор и промотор могут перекрываться. У многих оперонов имеется не один, а несколько операторов.

**Оперон** — термин применим исключительно к прокариотическим организмам. Оперон представляет собой группу из нескольких последовательно расположенных структурных генов, продукты которых находятся в тесной функциональной связи вместе с регуляторной системой (оператор, промотор, терминатор). Встречаются опероны, кодирующие один белок. Оператор и промотор расположены перед кодирующей областью, а терминатор — после.

**Ортологи** — гомологичные гены или белки, выполняющие сходные функции.

**Открытая рамка считывания (ORF — open reading frame)** — рамка считывания, которая позволяет считывать информацию со всего кодирующего участка мРНК (от иницирующего кодона до терминального).

**Относительно GC-бедный кодон (триплет)** — кодон, обладающий меньшей GC-насыщенностью по сравнению с другими кодонами из данной серии. Относительно GC-бедный кодон не содержит гуанин и цитозин в третьем положении.

**Относительно GC-богатый кодон (триплет)** — кодон, обладающий большей GC-насыщенностью по сравнению с другими кодонами из данной серии. Относительно GC-богатый кодон содержит гуанин или цитозин в третьем положении.

**Отстающая цепь ДНК** — прерывисто реплицирующаяся цепь ДНК. Направление репликации одинаково как для отстающей, так и для лидирующей цепи (5'–3'). По этой причине, перед тем, как начинается синтез комплементарной цепи на матрице, направление нуклеотидов в которой 5'–3', ДНК-хеликаза должна раскрутить определенный участок ДНК. Синтез комплементарной цепи начинается много раз (по мере раскручивания спирали ДНК). В результате отстающая цепь синтезируется в виде коротких фрагментов (фрагментов Оказаки), которые затем сшиваются ДНК-лигазами.

**Отрицательный отбор** — разновидность естественного отбора. Отрицательный отбор приводит к элиминации нежизнеспособных или неприспособленных к условиям обитания особей из популяции. Другими словами, отрицательный отбор основан на том, что неприспособленные к условиям внешней среды особи не оставляют потомства. Согласно теории М. Кимуры, естественный отбор на молекулярном уровне за редким исключением представлен отрицательным отбором. В синтетической теории эволюции отрицательный отбор часто приравнивают к стабилизи-

рующему, то есть к отбору, элиминирующему все варианты признаков, отличные от наиболее распространенных в популяции.

**Охра-кодон** — общепринятое название *терминального* триплета ГАА (UAA — в РНК).

**Охра-мутация** — мутация, приводящая к возникновению в *кодирующей* участке ДНК (мРНК) триплета ГАА (UAA — в мРНК).

**Охра-супрессор** — ген, кодирующий тРНК, антикодон которой комплементарен кодону UAA (охра-кодону). В результате охра-супрессии возникает отклонение от универсального генетического кода. Например, у некоторых организмов терминальный кодон UAA кодирует аминокислоту (глутамин). Охра-супрессор возникает в результате мутации одного из генов, кодирующих тРНК, связывающую глутамин. Также охра-кодон у некоторых организмов может кодировать пирролизин.

**Очищающий отбор** — в селекционных тестах так называется картина нуклеотидных замен в гомологичных генах, при которой частота фиксации синонимичных замен больше несинонимичных. Авторы селекционных тестов предполагали, что такая картина — проявление отрицательного (очищающего) отбора. Данный вывод, в целом, согласуется с теорией нейтральности М. Кимуры. Однако очищающий отбор как биологический процесс может иметь место и при других результатах селекционных тестов.

## П

**Палиндром** — комплементарные друг другу, находящиеся в разных цепях ДНК последовательности нуклеотидов, которые читаются одинаково в разных направлениях считывания. Палиндромные участки часто являются местами связывания рестриктирующих эндонуклеаз. Дифференцируют с *палиндромами истинными*.

**Палиндром истинный** — последовательность нуклеотидов ДНК в одной цепи, которая читается одинаково в направлениях 5' и 3'.

**Панмиксия** — свободное скрещивание между всеми особями внутри одной популяции. Абсолютная панмиксия достижима только теоретически.

**Парологи** — гомологичные гены или белки, выполняющие разные функции.

**Пенетрантность** — частота фенотипического проявления определенного гена (находящегося в гомозиготном состоянии) в популяции (выражается в процентах).

**Пептидилтрансфераза** — фермент, катализирующий образование пептидной связи между двумя аминокислотными остатками во время трансляции. У прокариот пептидилтрансферазной активностью обладает 50S-субъединица рибосом.

**Пептидильный центр** — участок рибосомы, в котором происходит образование *пептидной* связи между двумя аминокислотными остатками.



**Пептиды** — органические молекулы, образующиеся при взаимодействии аминокислот (аминогруппа одной аминокислоты взаимодействует с карбоксильной группой другой с образованием пептидной связи).

**Первичная структура белка** — порядок расположения аминокислотных остатков в аминокислотной последовательности белка. Первичная структура белка «закодирована» в нуклеотидной последовательности гена. Она предопределяет, но не детерминирует вторичную и третичную структуру.

**Первичная структура ДНК или РНК** — порядок расположения нуклеотидов в ДНК или РНК.

**Перекрестно реагирующие антитела** — антитела, синтезированные к одному антигену, но способные к взаимодействию и с другим антигеном (антигенами).

**Пиримидиновые основания** — азотистые основания, являющиеся производными пиримидина. В ДНК к мажорным пиримидиновым азотистым основаниям относят цитозин и тимин, в РНК — цитозин и урацил.

**Первичный транскрипт** — см. *про-мРНК*.

**Плазмон** — совокупность генов, расположенных во всех внеядерных молекулах ДНК. Термин характеризует весь вид в целом, противопоставляется *геному*.

**Плазмиды** — небольшие участки ДНК, относительно автономно существующие в клетках прокариот и низших эукариот. Есть плазмиды, специфические для определенного вида прокариот (они входят в состав их полного генома), некоторые плазмиды нельзя отнести к геному какого-либо определенного вида. Некоторые плазмиды реплицируются самостоятельно; плазмиды, входящие в состав полного генома, как правило, реплицируются вместе с «хромосомами» прокариот, то есть в момент деления клетки. Плазмиды способны к встраиванию в «хромосомы» своих хозяев. При помощи плазмид происходит обмен генетической информацией между бактериями.

**Плейотропизм** — ситуация, возникающая в том случае, если продукт одного гена вызывает проявление определенных вариантов сразу нескольких фенотипических признаков [131]. При этом, если отбор ведется по одному признаку, то автоматически закрепляются (или элиминируются) и все остальные, обусловленные данным аллельным вариантом гена.

**Позитивный (положительный) отбор** — в селекционных тестах так называется картина нуклеотидных замен в двух гомологичных генах, при которой несинонимичная дистанция (число несинонимичных нуклеотидных замен деленное на число несинонимичных сайтов) достоверно выше синонимичной дистанции (число синонимичных нуклеотидных замен деленное на число синонимичных сайтов). Авторы селекционных тестов предполагали, что такая картина нуклеотидных замен свидетельствует

о существовании *направленного положительного отбора* относительно данного гена. Тем не менее с позиции теории нейтральности такая трактовка является неправомерной: высокая частота несинонимичных замен может свидетельствовать о полном отсутствии отрицательного отбора относительно структуры данного гена. Сами авторы селекционных тестов подчеркивают, что подобная картина нуклеотидных замен встречается в генах после их недавней дупликации. Подобные гены (по М. Кимуре) не являются необходимыми для жизнедеятельности организма, в связи с чем практически все мутации в них нейтральны [16]. Кроме того, несинонимичная дистанция может быть выше, чем синонимичная, после единичной крупной молекулярной перестройки (инсерции, делеции, сдвига рамки считывания). Нами впервые было обнаружено, что несинонимичная дистанция может быть больше синонимичной в генах с 99 % насыщенностью гуанином и цитозином третьих положений кодонов [82]. Этот феномен объясняется критическим снижением самой возможности возникновения синонимичных замен (в результате сильного GC-давления) [82].

**Полиаденилирование** — процесс, являющийся частью процессинга мРНК у эукариот, заключающийся в присоединении к 3'-концу мРНК 50–200 аденилатных остатков.

**Полимер** — макромолекула, состоящая из ковалентно связанных повторяющихся субъединиц (мономеров).

**Полиморфизм** — наличие в популяции нескольких генетически различных форм. Под полиморфизмом на генетическом уровне понимается наличие в популяции нескольких вариантов нуклеотидных последовательностей одного и того же гена. Если частота встречаемости единичной нуклеотидной замены в популяции выше 1 %, то по традиционным критериям такая замена считается следствием полиморфизма, а не произошедшей ранее мутации.

**Полисома** — совокупность нескольких рибосом, осуществляющих трансляцию с одной и той же мРНК одновременно.

**Полицистронный транскрипт** — молекула мРНК, содержащая более чем один кодирующий участок. Большая часть генов прокариот организована в опероны, транскрипты с которых являются полицистронными.

**Популяционные волны** — один из элементарных микроэволюционных факторов. Популяционные волны являются колебаниями численности популяции, вызываемыми периодическими изменяющимися условиями внешней среды. При благоприятных условиях внешней среды возрастает численность популяции и ее гетерогенность за счет снижения воздействия естественного отбора; при неблагоприятных условиях среды численность популяции существенно снижается, ее гетерогенность уменьшается за счет увеличения влияния естественного отбора.

**Популяция** — элементарная единица процесса микроэволюции, способная реагировать на изменение внешней среды изменением своего генофонда. Популяция представляет собой группу особей одного вида, длительно населяющих определенный ареал, свободно скрещивающихся между собой и относительно изолированных от других групп особей того же вида.

**Последовательность** — порядок расположения нуклеотидов в ДНК (РНК) или аминокислот в белке. Часто термин используется для обозначения определенного участка (секвестра — от англ. sequence) ДНК, мРНК или белка, представляющего собой ген или участок гена (белок или участок белка).

**Посттрансляционная модификация полипептидных цепей** — процесс, приводящий к изменению первичной структуры белка после трансляции. Многие вирусные белки, гормоны и нейропептиды синтезируются в виде полипротеинов, которые впоследствии разрезаются протеолитическими ферментами в определенных местах, образуя несколько меньших по длине зрелых (функционально активных) пептидов или полипептидов.

**Правила Чаргаффа.** В 1953 г. с помощью кислотного гидролиза ДНК с последующей хроматографией и количественным анализом Э. Чаргафф установил следующие закономерности:

1. Количество *пуриновых* оснований (А+G) в молекуле ДНК всегда равно количеству *пиримидиновых* оснований (Т+С).

2. Количество аденина равно количеству тимина [ $A=T$ ,  $A/T=1$ ], а количество гуанина равно количеству цитозина [ $G=C$ ,  $G/C=1$ ], причем как в двухцепочечной ДНК, так и внутри каждой отдельной цепи ДНК.

3. Соотношение количества гуанина и цитозина в ДНК к количеству аденина и тимина является постоянным для каждого вида живых организмов [ $(G+C)/(A+T)=K$ , где  $K$  — коэффициент специфичности].

Данные правила послужили основой для создания Уотсоном и Криком модели двойной спирали ДНК. Второе правило Чаргаффа внутри одной цепи ДНК, как правило, выполняется не строго: частота использования G близка к частоте использования C, но не равна ей абсолютно у большинства организмов.

**Праймер (затравка)** — короткий участок РНК, синтезирующийся на раскрученной цепи ДНК в начале процесса репликации ДНК. Новая цепь ДНК синтезируется в виде продолжения праймеров. В дальнейшем праймеры замещаются сегментами ДНК.

**Праймосома** — мультиферментный комплекс, активирующий матричную цепь ДНК, в результате чего фермент праймаза синтезирует *праймеры*.

**Претерминальные кодоны** — кодоны, которые могут стать терминальными в результате одной нуклеотидной замены. В универсальном генетическом коде существует 18 претерминальных кодонов.

**Про-мРНК** — первичный транскрипт (у эукариот), не подвергшийся *процессингу*.

**Прокариоты** — живые организмы, клетки которых лишены ядра. Прокариотические организмы разделяются на два надцарства: *бактерии* и *археи*. Разделить прокариот на два надцарства удалось исключительно благодаря использованию методов молекулярной филогенетики [56]. Долгое время постулировалось, что генетический аппарат прокариотической клетки представлен «нуклеоидом» — одной кольцевой молекулой ДНК. В настоящее время обнаружено довольно большое количество прокариот, у которых в клетке находятся две длинные кольцевые молекулы ДНК (их называют «хромосомами») и определенное количество линейных или кольцевых коротких молекул ДНК (их называют «плазмидами»). Более того, у некоторых прокариот (например, у представителей рода *Streptomyces*) были обнаружены линейные «хромосомы» [73]. В полный геном прокариотического организма входят как хромосомы, так и плазмиды. Большинство прокариот относится к строго одноклеточным организмам, меньшая же их часть способна не только к одноклеточной форме существования, но и к образованию многоклеточных образований с дифференцированными клетками (например некоторые метанпродуцирующие археи) [70].

**Промотор** — участок ДНК, детерминирующий инициацию транскрипции (связывание РНК-полимеразы).

**Протеом** — совокупность всех белков данного вида организмов.

**Процессинг** — процесс посттранскрипционной модификации РНК. У прокариот процессингу подвергаются первичные транскрипты рибосомальной и транспортной РНК (трансляция зачастую протекает параллельно с транскрипцией — мРНК не модифицируется). У эукариот процессингу подвергаются не только первичные транскрипты рРНК и тРНК, но и про-РНК. Элементами процессинга про-мРНК являются *экспонирование*, *полиаденилирование*, *сплайсинг*, приобретение мРНК вторичной структуры.

**Процессированные псевдогены** — подкласс псевдогенов, похожих скорее на ДНК-копии мРНК, чем на гены. Процессированные псевдогены не содержат участков, похожих на интроны родственных им генов, но имеют полиаденилированные 3'-концы. Они рассеяны по всему геному, встречаются в разных хромосомах. Возможно, данный подкласс псевдогенов возник в результате обратной транскрипции и транспозиции. Процессированные псевдогены являются *ретрогенами*.

**Прыгающие гены** — см. *транспозоны*.

**Псевдогены** — участки генома, близкие в структурном отношении специфическим функционально активным генам, но не являющиеся их аллельными формами и не кодирующие никаких функциональных генных продуктов. Псевдогены, вероятнее всего, являются дубликатами функционально активных генов, отличающимися от последних мутациями в регу-

ляторной части или в кодирующем участке. Некоторые псевдогены вообще не транскрибируются, некоторые могут транслироваться с образованием дефективных полипептидов.

**ПТК** — частота использования претерминальных кодонов в гене или в кодирующем геноме.

**Пуриновые основания** — азотистые основания, производные пурина. В ДНК и РНК к мажорным пуриновым азотистым основаниям относят аденин и гуанин.

## Р

**Радикальная аминокислотная замена** — замена аминокислоты в белке, которая приводит к значительным изменениям структуры и свойств белка. Термин весьма относителен, поскольку существует большое количество противоречащих друг другу систем определения радикальности аминокислотных замен [2]. Нами предложен способ определения степени радикальности (противопост. *консервативности*) аминокислотных замен в белках и протеомах на основе анализа частоты их фиксации в популяциях исследуемых организмов [30, 31, 32, 83].

**Рамка считывания** — один из трех возможных способов считывания генетической информации с кодирующего участка мРНК. Триплетность генетического кода предполагает возможность начала считывания информации как с первого, так и со второго или с третьего положения нуклеотида в кодирующем участке. Различают *открытую* и *закрытую* рамки считывания.

**Редактирование РНК** — процесс, приводящий к возникновению нуклеотидных замен в РНК под воздействием специфических ферментов. Различают три основных вида редактирования: 1) вставка поли-U, 2) дезаминирование С с образованием U, 3) дезаминирование А с образованием I (инозин), который является аналогом гуанина. Процесс дезаминирования аденина обычно происходит в местах образования двухцепочечных фрагментов РНК (в т. ч. шпилек), поскольку такие структуры являются субстратом для воздействия ферментов из семейства ADAR [62]. Процесс дезаминирования цитозина, наоборот, чаще происходит в одноцепочечных фрагментах РНК, поскольку они являются субстратом для ферментов из семейства APOBEC1. Процессы редактирования РНК, фактически, являются процессами, вызывающими *мутационное давление* в геномах РНК вирусов [7]. AICDA и ферменты из семейства APOBEC3, дезаминирующие цитозин в ДНК, являются гомологами ферментов их семейства APOBEC1, дезаминирующих цитозин в РНК [62].

**Рекогниция** — процесс образования *аминоацил-тРНК* из тРНК и соответствующей ей аминокислоты.

**Рекомбинация** — перераспределение генетического материала родителей у потомства. Происходит у всех живых организмов: у эукариот — за счет полового размножения, у прокариот — посредством конъюгации, трансформации, трансдукции.

**Рекомбинация генов** — возникновение новых наследственных комбинаций генов (изменение состава *групп сцепления*) в результате кроссинговера.

**Репарация генетического материала** — процесс исправления повреждений ДНК. Осуществляется ферментами, входящими в состав системы репарации.

**Репликационная вилка** — структура, существующая во время репликации, представляющая собой место раскручивания двойной цепи ДНК.

**Репликация ДНК** — процесс удвоения молекулы ДНК. Осуществляется с помощью набора специфических ферментов (ДНК-полимеразы, праймазы, ДНК-лигазы, ДНК-геликазы, гиразы, топоизомеразы и т. д.). Репликация ДНК полуконсервативна (в каждой новой молекуле ДНК только одна из цепочек является вновь синтезированной) и полупрерывиста (*отстающая цепь* реплицируется прерывисто, *лидирующая* — непрерывно).

**Реплисома** — гипотетический мультиферментный комплекс, состоящий из функционального праймосомо-праймазного комплекса, геликазы, ДНК-полимеразы и др. ферментов, который обеспечивает одновременное осуществление процессов инициации праймерной РНК, удлинения лидирующей цепи и достраивания отстающей.

**Репрессор** — см. *белок-репрессор*.

**Рестриктурирующие эндонуклеазы** — ферменты, разрезающие обе цепи двойной спирали ДНК в определенных местах (*палиндромах*).

**Ретрогены** — перемещающиеся по геному элементы, не имеющие в своей структуре генов, необходимых для акта транспозиции. По-видимому, ретрогены возникают в результате пассивной (осуществляемой ревертазами ретровирусов или *ретротранспозонов*) обратной транскрипции. К ретрогенам относятся *процессированные псевдогены*.

**Ретротранспозоны** — транспозоны, содержащие в себе ген, кодирующий обратную транскриптазу. Ретротранспозоны I класса напоминают ретровирусные провирусы (содержат в себе ген ревертазы и длинные концевые повторы, окружающие центральную часть), но не образуют жизнеспособных внеклеточных форм. Ретротранспозоны II класса отличаются от представителей I класса отсутствием концевых повторов.

**Рецессивный аллель** — аллель, проявляющийся исключительно в гомозиготном состоянии.

**Рибонуклеиновые кислоты (РНК)** — высокомолекулярные генетические биополимеры. Цепочка РНК представляет собой неразветвленную полинуклеотидную цепь, нуклеотиды в которой состоят из азотистого

основания (аденина, гуанина, цитозина или урацила), рибозы и остатка фосфорной кислоты. Выделяют следующие разновидности РНК: *матричная РНК (мРНК)*, *транспортная РНК (тРНК)*, *рибосомальная РНК (рРНК)*, а также *малые РНК*, в частности, *микроРНК*, функцией которой является регуляция трансляции.

**Рибосомы** — органеллы, функцией которых является синтез белка (трансляция). Состоят из двух субъединиц (большой и малой). Каждая субъединица рибосомы представляет собой комплекс рРНК и белков.

**Рибосомы прокариот** — рибосомы, состоящие из малых (30S) и больших (50S) субъединиц. 30S-субъединицы состоят из единственной молекулы рРНК размером 1542 нуклеотида и 21 белок. 50S-субъединицы содержат две молекулы рРНК (большую, размером 2904 нуклеотида; и меньшую, размером 120 нуклеотидов) и 34 разных белка.

**Рибосомы эукариот** — рибосомы, состоящие из малых (40S) и больших (60S) субъединиц. 40S-субъединицы состоят из единственной молекулы рРНК размером 1900 нуклеотидов и 30–35 белков. 60S-субъединицы содержат три цепи рРНК (длиной 120, 160 и 4800 нуклеотидов) и 40–45 белков.

## С

**Сайленсер** — один из цис-элементов в *эукариотических генах*, ослабляющий транскрипцию.

**Сайт** — положение в гене, белке или в массиве выровненных последовательностей, соответствующее определенному нуклеотиду или аминокислотному остатку. В литературе данный термин может использоваться для обозначения короткой последовательности в составе ДНК, РНК или в белке. В данном пособии, во избежание путаницы в терминах, короткая последовательность в составе ДНК, РНК или белка называется *участком или регионом*.

**Секвенирование** — процесс установления порядка расположения нуклеотидов в определенном участке ДНК или РНК.

**Селекционные тесты** — группа методов, разработанных для определения вида естественного отбора в гомологичных генах или их участках. С помощью селекционных тестов можно определить наличие *очищающего отбора*, *направленного отбора*, *критерия отсутствия отбора*. Данные тесты основаны на сравнении *синонимичных* и *несинонимичных дистанций*.

**Серия кодонов** — группа синонимичных кодонов. Серии кодонов могут быть *двукратно вырожденными*, *трехкратно вырожденными*, *четырекратно вырожденными*, *шестикратно вырожденными*, *невырожденными*.

**Синонимичная дистанция** — эволюционная дистанция, равная числу синонимичных замен на синонимичный сайт в сравниваемых нук-

леотидных последовательностях. Перед непосредственным вычислением синонимичной эволюционной дистанции необходимо воспользоваться одним из методов, рассчитывающих число *синонимичных* и *несинонимичных сайтов* и соответствующих замен в сравниваемых последовательностях.

**Синонимичная замена** — нуклеотидная замена в кодирующем участке ДНК (мРНК), приводящая к замене кодона, в котором она произошла, на синонимичный; то есть не приводящая к замене аминокислоты в кодируемой белке.

**Синонимичный сайт** — так называется нуклеотидное положение, в котором хотя бы одно из трех возможных замещений является синонимичным. Если все замены в данном сайте синонимичны, сайт называется абсолютно синонимичным. Если одна из замен в данном сайте синонимична, сайт называется синонимичным на  $\frac{1}{3}$  (соответственно, несинонимичным на  $\frac{2}{3}$ ). Если две замены в данном сайте синонимичны, сайт называется синонимичным на  $\frac{2}{3}$  (соответственно, несинонимичным на  $\frac{1}{3}$ ).

**Синонимичные кодоны** — группа триплетов, кодирующих одну и ту же аминокислоту.

**Скорость молекулярной эволюции** — показатель, равный эволюционной дистанции между двумя гомологичными генами (белками), деленной на удвоенное число лет, прошедших после их дивергенции. Эволюционная дистанция определяется исключительно молекулярно-биологическими методами. Примерное время дивергенции может быть определено с помощью палеонтологических исследований или с помощью молекулярно-биологических методов. Скорость молекулярной эволюции выражается в заменах на сайт в год.

**Случайный генетический дрейф (генетико-автоматические процессы)** — процесс, обеспечивающий фиксацию в популяции эволюционно нейтральных или почти нейтральных мутаций. Случайный генетический дрейф является сборным понятием, объединяющим множество эволюционных факторов, результатом взаимного влияния которых является фиксация (или элиминация) нейтральной мутации. Термин «случайный генетический дрейф» подразумевает под собой вероятностный характер фиксации в популяции тех мутаций, которые не снижают и не повышают приспособленность особи к условиям внешней среды. Случайный генетический дрейф является (по теории М. Кимуры) основным (наряду с *отрицательным отбором*) эволюционным фактором, действующим на молекулярном уровне.

Всеобъемлющим проявлением случайного генетического дрейфа является *эффект горлышка бутылки* (bottle neck). Этот эффект подразумевает под собой неселективное уничтожение части популяции в результате воздействия каких-либо мощных факторов (землетрясение, наводнение, засуха и т. д.). При этом погибает только та часть популяции, которая



попала под воздействие такого фактора, а остальная ее часть выживает вне зависимости от особенностей ее генофонда [123].

Далее случайный генетический дрейф начинает проявляться в виде **эффекта родоначальника**: выжившие (или мигрировавшие) организмы размножаются, существенно повышая частоты встречаемости в популяции всех своих генов [123].

Так называемый «**дрейф генов**» является частным проявлением случайного генетического дрейфа, характерным исключительно для малых популяций диплоидных организмов, размножающихся половым путем.

Только малая часть из всех генов в геноме проявляется фенотипически. Лимитирующими являются отдельные фенотипические признаки, по которым и ведется естественный отбор на популяционном уровне. Соответственно, фиксация или элиминация всех нейтральных мутаций, не затрагивающих гены, ответственные за лимитирующие признаки, являются случайностью.

**Спейсеры** — некодирующие участки ДНК в генах рРНК или тРНК. Спейсеры вырезаются во время процессинга рРНК или тРНК.

**Спаривание оснований** — процесс образования водородных связей между комплементарными друг другу азотистыми основаниями.

**Специализированные гены** — гены, активные только в определенных типах клеток многоклеточного организма.

**Сплайсинг** — процесс разрезания про-мРНК в специальных участках (*акцепторных* и *донорных точках сплайсинга*) и последующего соединения участков, соответствующих экзонам, с образованием мРНК. Сплайсинг происходит только у эукариотических организмов.

**Структурные гены** — гены, кодирующие белки, производящие непосредственный фенотипический эффект. Термин справедлив для характеристики прокариотических генов: их принято разделять на структурные гены и *гены-регуляторы*.

## Т

**Тандемные повторы** — повторяющиеся друг за другом короткие последовательности нуклеотидов. При наличии тандемных повторов в кодирующем участке, в кодируемом им белке тоже возникают повторяющиеся друг за другом группы аминокислотных остатков.

**Таксон** — группа родственных организмов, которым можно присвоить определенную таксономическую категорию (вид, род, семейство, класс, тип, царство и др.)

**Теория информации (К. Шеннона)** — математическая теория, созданная в 40-е гг. XX в. Рассматривает любую систему, в том числе и биологическую, как объем информации. Чем ближе друг к другу частоты встречаемости всех компонентов, из которых состоит система, тем больше

комбинаций они могут образовать, тем выше объем информации, который может содержать данная система. Показателем, отражающим количество информации в системе, является информационная энтропия ( $H$ ). Чем ниже энтропия, тем меньше информации в системе (тем система однородней). Чем выше энтропия, тем больше информации в системе (тем система разнообразней) [135].

**Теория мутационного давления (теория Н. Суеоки)** — теория, признающая существование *мутационного давления* и провозглашающая его основной причиной возникновения генных мутаций преимущественного направления и, как следствие, разницы в частотах использования нуклеотидов между геномами различных видов организмов и различными генами внутри генома одного вида организмов. Теория Н. Суеоки во многом базируется на *теории нейтральной эволюции*.

**Теория нейтральной эволюции (теория М. Кимуры)** — фундаментальная теория молекулярной эволюции. Главным принципом данной теории является признание *отрицательного отбора* и *случайного генетического дрейфа* основными факторами эволюции на молекулярном уровне. Пять основных положений данной теории:

1. Скорость эволюции любого белка, выраженная через число аминокислотных замен на сайт в год, приблизительно постоянна и одинакова в разных филогенетических линиях, если только функция и третичная структура этого белка остаются практически неизменными.

2. Функционально менее важные молекулы и их части эволюционируют (накапливают мутационные замены) быстрее, чем более функционально важные.

3. Мутационные замены, приводящие к меньшим нарушениям структуры и функции молекулы (консервативные замены), в ходе эволюции происходят чаще тех, которые вызывают более существенное нарушение структуры и функции данной молекулы (радикальных замен).

4. Появлению нового в функциональном отношении гена всегда должна предшествовать дупликация гена.

5. Селективная элиминация вредных мутаций (за счет отрицательного отбора) и случайная фиксация селективно нейтральных или очень слабо вредных мутаций (за счет случайного генетического дрейфа) происходят в ходе эволюции гораздо чаще, чем положительный дарвиновский отбор благоприятных мутаций.

**Терминатор транскрипции** — участок ДНК, который является сигналом к завершению процесса транскрипции. Транскрипт мРНК с терминаторного участка имеет вторичную структуру в виде шпильки. Шпилька мРНК приостанавливает работу РНК-полимеразы.

**Терминальные кодоны** — кодоны мРНК, на которых прекращается процесс трансляции. В универсальном генетическом коде существуют три

терминальных кодона мРНК: UAA, UAG и UGA. Не существует тРНК с антикодонами, комплементарными терминальным кодонам. Тем не менее при попадании в аминокатильный центр рибосомы терминальные кодоны узнаются специфическими факторами терминации трансляции, которые катализируют распад трансляционного комплекса.

**Терминация** — заключительный этап транскрипции или трансляции.

**Тимин (5-метилурацил)** — пуриновое азотистое основание, входящее в состав нуклеотидов ДНК.

**Т-лимфоциты** — клетки иммунной системы, осуществляющие реакции клеточного иммунитета. Общим мембранным маркером для всех субпопуляций Т-лимфоцитов являются молекулы комплекса CD3.

**Транзиция** — тип нуклеотидной замены. Транзицией называется замена пуринового основания на пуриновое, пиримидинового — на пиримидиновое.

**Трансверсия** — тип нуклеотидной замены. Трансверсией называется замена пуринового основания на пиримидиновое, пиримидинового — на пуриновое.

**Транскрипт** — молекула РНК, образующаяся в результате транскрипции.

**Транскрипция** — процесс синтеза на матрице цепи ДНК молекулы РНК. Процесс осуществляется ферментами ДНК-зависимыми РНК-полимеразами.

**Транслокация рибосомы** — перемещение рибосомы вдоль молекулы мРНК после включения каждой новой аминокислоты в полипептидную цепь.

**Трансляция** — процесс синтеза белка. Молекула белка синтезируется на матрице мРНК с соблюдением соответствия каждого кодона определенной аминокислоте. В процессе принимают участие *аминокатиль-тРНК* и *рибосомы*.

**Транспозоны** — группа мобильных элементов генома, которые могут менять свою локализацию в хромосомах (в нуклеоиде — у прокариот). Обладают двумя свойствами: 1) несут в себе ген (или несколько генов), необходимый для транспозиции, 2) на концах содержат специфические взаимно инвертированные повторяющиеся последовательности. Транспозиция может сопровождаться сильными мутационными эффектами. Транспозоны, в состав которых входит ген обратной транскриптазы, называются *ретротранспозонами*.

**3'-трейлерная последовательность** — нетранслируемая последовательность на 3'-конце мРНК.

**Третичная структура белка** — сложная конфигурация белковой молекулы, формирующаяся на основании вторичной структуры. Третичная структура поддерживается гидрофобными взаимодействиями аминокис-

лотных остатков, электростатическими связями между некоторыми аминокислотными радикалами, дисульфидными связями.

**Трехкратно вырожденный сайт** — так называется положение нуклеотида, в котором две замены (из всех трех возможных) являются синонимичными. Фактически, к таким сайтам можно отнести только третьи положения в кодонах изолейцина (в универсальном генетическом коде).

**Трехшаговая аминокислотная замена** — аминокислотная замена, вызванная заменой всех трех нуклеотидов в исходном кодоне.

**Триплет** — см. *кодон*.

**Триплетность** — свойство генетического кода, заключающееся в том, что одна аминокислота кодируется тремя последовательно расположенными нуклеотидами (*триплетом*).

**10АА** — группа аминокислотных остатков, которые кодируются кодонами, средними по GC-насыщенности (содержащими гуанин или цитозин либо в первом, либо во втором положении). Всего таких аминокислотных остатков десять, с чем и связано введение нами такого группового обозначения. К данной группе относятся: глутамин, серин, треонин, гистидин, лейцин, валин, цистеин, триптофан, аспарагиновая и глутаминовая кислоты.

## У

**Универсальность** — свойство *генетического кода*, заключающееся в том, что подавляющее большинство живых организмов имеют одинаковый (универсальный) генетический код. Тем не менее открыто несколько десятков неуниверсальных генетических кодов. Последние, как правило, имеют небольшое число отклонений от универсального.

**Урацил (2,4-диоксипиримидин)** — пиримидиновое азотистое основание, содержащееся в нуклеотидах РНК.

**Усредненная последовательность** — см. *консенсусная последовательность*.

**Уход от иммунитета (иммунного ответа)** — сборное понятие, включающее в себя все механизмы, благодаря которым вирус или бактерия избегает воздействия *иммунного ответа*.

**Участки сплайсинга** — см. *донорные и акцепторные точки сплайсинга*.

## Ф

**Фактор эволюции** — процесс, тем или иным образом вызывающий изменение генофонда популяции (*естественный отбор, мутационное давление, случайный генетический дрейф, популяционные волны* и др.).

**Фенокопия** — изменение фенотипа под воздействием факторов внешней или внутренней среды, внешне напоминающее проявление какой-либо определенной мутации.

**Фенотип** — комплекс всех признаков и свойств особи, формирующийся под воздействием факторов среды на основе генотипа (факторы среды могут вызывать вариации генетически детерминированных признаков в пределах нормы реакции). Особи с одинаковыми генотипами могут обладать разными фенотипами. Особи со схожими фенотипами могут иметь разные генотипы. В процессе микроэволюции отбор идет исключительно по фенотипическим признакам.

**Ферменты** — молекулы белков, являющиеся катализаторами биохимических реакций в живых организмах.

**Фиксация мутации** — о фиксации мутации в популяции можно говорить, если частота встречаемости этой мутации среди особей данной популяции приблизилась к 100 %.

**Филогенез** — процесс исторического развития отдельных таксономических единиц (или гомологичных генов). Графически отражается *дендрограммой*.

**Филогенетика** — наука, изучающая *филогенез*.

**ГУМИНК** — группа аминокислотных остатков, которые кодируются кодонами, не содержащими ни гуанина, ни цитозина в первых и во вторых положениях (GC-бедными кодонами). К этой группе относятся фенилаланин, тирозин, метионин, изолейцин, аспарагин и лизин.

**Фрагменты Оказаки** — короткие участки, образующиеся при репликации *отстающей цепи ДНК*. Фрагменты Оказаки синтезируются на участке 3'–5' цепи ДНК между двумя точками начала синтеза комплементарной цепи.

## Х

**Хеликазы** — см. *геликазы*.

**Хроматида** — одна из двух нуклеопротеидных нитей, вместе образующих метафазную хромосому. В пресинтетический период митотического цикла хромосома состоит из одной хроматиды, после репликации ДНК (репликация происходит в синтетический период) образуется вторая хроматида, копия первой. По структуре хроматида представляет собой комплекс ДНК и белков-гистонов.

**Хроматин** — нуклеопротеидные нити, из которых состоят хромосомы клеток эукариот. Представляет собой дезокси-нуклеопротеин (комплекс ДНК и белков-гистонов).

**Хромосома** — дискретная единица эукариотического генома. В промежутке между постсинтетическим периодом интерфазы и анафазой хромосома состоит из двух *хроматид*, соединенных между собой центроме-

рой. Хромосома представляет собой комплекс ДНК и белков-гистонов. Число, размер и форма хромосом строго специфичны для каждого вида.

**Хромосомные мутации (абберации)** — крупные перестройки хромосом (делеции, дупликации, инсерции, инверсии, транслокации), выявляемые при изучении кариотипа.

## Ц

**Центральная догма молекулярной биологии** — в настоящее время формулируется следующим образом: генетическая информация может передаваться от ДНК к ДНК (в процессе репликации), от ДНК к РНК (в процессе транскрипции), от РНК к ДНК (в процессе обратной транскрипции, с помощью фермента ревертазы), от мРНК к белку (в процессе трансляции).

**Цепь ДНК** — неразветвленная полинуклеотидная цепь, состоящая из нуклеотидов, связанных между собой фосфодиэфирными связями.

**Цитозин (2-окси-4-аминопиримидин)** — пиримидиновое азотистое основание, входящее в состав нуклеотидов ДНК, РНК, некоторых антибиотиков и коферментов.

**Цитоплазматическая наследственность** — явление передачи наследственных признаков через цитоплазму, а именно, через ДНК, находящуюся в органеллах, расположенных в цитоплазме (через митохондрии — у животных и грибов, через митохондрии и пластиды — у растений). У человека митохондрии содержатся исключительно в яйцеклетке, но отсутствуют в сперматозоиде.

## Ч

**Четвертичная структура белка** — структура, образующаяся в результате взаимодействия нескольких полипептидных цепей (за счет водородных, ионных связей или за счет гидрофобных взаимодействий).

**Четырехкратно вырожденный сайт** — так называется положение в нуклеотидной последовательности, в котором все три возможные замены синонимичны. Фактически, четырехкратно вырожденными сайтами являются третьи положения в кодонах четырехкратно вырожденных серий и в кодонах квартетов шестикратно вырожденных серий.

**Четырехкратно вырожденная серия кодонов** — серия кодонов, состоящая из четырех триплетов.

## Ш

**Шестикратно вырожденная серия кодонов** — серия кодонов, состоящая из шести триплетов.

**Шпилька мРНК** — вторичная структура мРНК, представляющая собой комплементарно связанные друг с другом нуклеотиды из одной

и той же цепи (инвертированные повторы). У прокариот шпилька участвует в процессе *аттенуации*: формируется из транскрипта мРНК с области терминатора транскрипции (терминатор транскрипции содержит инвертированные повторы, разделенные небольшим числом нуклеотидов) и препятствует работе РНК-полимеразы, тем самым прерывая процесс транскрипции.

**Шпилька ДНК** — комплементарно связанные друг с другом нуклеотиды из одной и той же цепи ДНК. Одноцепочечная ДНК, так же как и РНК, может образовывать вторичную структуру за счет комплементарного соединения инвертированных повторов.

Э

**Эволюционная дистанция** — см. *дистанция эволюционная*.

**Эволюционное учение** — наука о причинах, движущих силах, механизмах и общих закономерностях развития живых организмов. Объединяет в себе весь комплекс теорий эволюции.

**Эволюция** — закономерное историческое развитие органического мира. Эволюционный процесс на внутривидовом (популяционном) уровне называется микроэволюцией, на уровне выше видового — макроэволюцией. Эволюционный процесс, протекающий на уровне молекул генетических биополимеров, называется *молекулярной эволюцией*.

**Экзоны** — участки единицы транскрипции эукариотического гена, которые содержат информацию о первичной структуре соответствующего белка, то есть образуют собственно мРНК. В процессе *сплайсинга* участки *про-мРНК*, соответствующие экзонам, сшиваются и образуют мРНК.

**Экспрессивность** — степень проявления действия гена в данном организме.

**Элонгация** — этап трансляции, на котором происходит удлинение полипептидной цепи.

**Энхансер** — один из цис-элементов в *эукариотических генах*, усиливающий транскрипцию.

**Эпистаз** — взаимодействие неаллельных генов. При этом взаимодействии вероятность и степень фенотипического проявления одного гена определяется активностью других генов. Если вариант фенотипического признака, возникший в результате взаимодействия генов, повышает жизнеспособность особи, то вероятность передачи его потомству, как правило, невелика. Чем больше доминантных аллелей вызывают возникновение этого варианта признака и чем ниже количество взаимодействующих генов, тем выше вероятность передачи его по наследству.

**Эпитоп** — участок антигена, непосредственно на который вырабатывается иммунный ответ. Выделяют В-клеточные эпитопы (на которые вырабатывается гуморальный иммунный ответ) и Т-клеточные эпитопы (на которые вырабатывается клеточный иммунный ответ). В-клеточные

эпитопы разделяют на пространственные и линейные. Естественный иммунный ответ вырабатывается исключительно на пространственные эпитопы (участки, расположенные на поверхности белковой глобулы). Примерно в 10 % случаев в состав пространственного эпитопа входит линейный эпитоп (небольшой участок аминокислотной последовательности), в 90 % случаев пространственный эпитоп состоит из аминокислотных остатков, которые оказались в непосредственной близости друг к другу в результате формирования третичной структуры белка.

**Эукариоты (эукариотические организмы)** — представители одного из трех надцарств живых организмов. В эукариотических клетках имеется оформленное ядро, отделенное от цитоплазмы ядерной мембраной. В ядре находятся хромосомы, состоящие из линейной ДНК, упакованной при помощи специальных белков. Среди эукариот есть как одноклеточные, так и многоклеточные организмы. В клетках большинства эукариот присутствуют относительно автономные органеллы, содержащие собственный генетический аппарат (митохондрии и/или пластыды). Надцарство эукариот включает царства растений, грибов и животных.

**Эффект горлышка бутылки** — разновидность процессов случайного генетического дрейфа. Эффект горлышка бутылки заключается в уничтожении большей части популяции в результате какого-либо катаклизма. Выживание особей при этом обусловлено случайными причинами. Эффект горлышка бутылки приводит к возникновению эффекта родоначальника среди выжившей части популяции.

**Эффект родоначальника** — разновидность процессов случайного генетического дрейфа. Эффект родоначальника заключается в том, что генофонд популяции содержит только те гены (за исключением мутантных), которые присутствовали в геномах основавших ее особей. Набор генов в геномах родоначальников, как правило, определяется случайными факторами.



## Литература

1. *Биология*. Толковый словарь (с английскими эквивалентами) / В. П. Андреев [и др.] ; под общ. ред. В. П. Андреева. СПб. : Лань, 1999. 448 с. (Серия «Учебники для вузов, специальная литература»)
2. *Барковский, Е. В.* Мембраносвязанные аденилатциклазы : монография / Е. В. Барковский, О. В. Ачинович ; под общ. ред. Е. В. Барковского. Минск : БГМУ, 2005. 134 с.
3. *Характеристика* методов определения эволюционных расстояний между нуклеотидными последовательностями генетических макромолекул / Е. В. Барковский [и др.] // *Здравоохранение*. 2005. №5. С. 37–43.
4. *Методы* молекулярной эволюции и филогенетики : учеб.-метод. пособие / Е. В. Барковский [и др.]. Минск : БГМУ, 2005. 63 с.
5. *Барковский, Е. В.* Сравнительная характеристика мРНК аденилатциклаз актиномицетов / Е. В. Барковский, В. В. Хрусталева // *Бел. мед. журн.* 2004. № 3. С. 27–30.
6. *Барковский, Е. В.* Сравнительная характеристика матричных РНК и аминокислотных последовательностей аденилатциклаз цианобактерий / Е. В. Барковский, В. В. Хрусталева // *Бел. мед. журн.* 2005. № 1. С. 23–27.
7. *Барковский, Е. В.* Асимметричное мутационное давление у одноцепочечных РНК вирусов / Е. В. Барковский, В. В. Хрусталева // *Проблемы биохимии, радиационной и космической биологии* : тр. III Междунар. симпозиума под эгидой ЮНЕСКО. Дубна : ОИЯИ, 2007. С. 268–272.
8. *Барковский, Е. В.* Частоты использования кодонов аргинина в геномах прокариот как свидетельство ранее существовавшего сильного мутационного давления / Е. В. Барковский, В. В. Хрусталева // *Генетика и биотехнология XXI века. Фундаментальные и прикладные аспекты* : материалы Междунар. науч. конф. Минск : БГУ, 2008. С. 221–223.
9. *Берг, Р. Л.* Зависимость между степенью проявления мутаций в гетерозиготе и их концентрацией в генотипе естественных популяций / Р. Л. Берг // *Изв. АН СССР*. 1944. № 2. С. 121–128.
10. *Сравнительный* анализ современных методов изучения синонимичных и несинонимичных нуклеотидных замещений. Сообщение 1. Методы, основанные на эволюционных путях / А. В. Бутвиловский [и др.] // *Здравоохранение*. 2005. № 6. С. 37–39.
11. *Сравнительный* анализ современных методов изучения синонимичных и несинонимичных нуклеотидных замещений. Сообщение 2. Методы, основанные на модели Кимуры. Методы максимального сходства с моделями замещений кодонов / А. В. Бутвиловский [и др.] // *Здравоохранение*. 2005. № 6. С. 40–43.
12. *Основные* методы молекулярной эволюции : монография / А. В. Бутвиловский [и др.] ; под общ. ред. проф. Е. В. Барковского. Минск : Белпринт, 2009. 216 с.
13. *Справочник* терминов молекулярной эволюции и филогенетики : учеб.-метод. пособие / В.Э. Бутвиловский [и др.]. Минск : БГМУ, 2006. 40 с.
14. *Калмыков, В. Н.* Философия : учеб. пособие / В. Н. Калмыков. Минск : Выш. шк., 2006. 431 с.
15. *Картель, Н. А.* Генетика. Энциклопедический словарь / Н. А. Картель, Е. Н. Макеева, А. М. Мезенко. Минск : Технология, 1999. 488 с.
16. *Кимура, М.* Молекулярная эволюция : теория нейтральности / М. Кимура. М. : Мир, 1985. 398 с.

17. *Левонтин, Р.* Генетические основы эволюции / Р. Левонтин. М. : Мир, 1978. 352 с.
18. *Лобзин, Ю. В.* Руководство по инфекционным болезням : в 2 ч. Ч. 1 / Ю. В. Лобзин. СПб. : Лань, 2000. 190 с.
19. *Лобзин, Ю. В.* Руководство по инфекционным болезням : в 2 ч. Ч. 2 / Ю. В. Лобзин. СПб. : Лань, 2000. 180 с.
20. *Мажуль, В. М.* Развитие исследований в области протеомики в Республике Беларусь : фундаментальные и прикладные аспекты / В. М. Мажуль. Доложено на заседании Президиума НАН Беларуси 7 июля 2005 г. Минск, 2005. 49 с.
21. *Оно, С.* Генетические механизмы прогрессивной эволюции / С. Оно. М. : Мир, 1973. 228 с.
22. *Столярова, Е. А.* Молекулярно-генетическая эволюция генов иммуноглобулинов, кодирующих вариабельные области молекул антител больных ВИЧ-инфекцией и ВИЧ-ассоциированными заболеваниями / Е. А. Столярова, Т. А. Столярова, В. В. Хрусталёв // Актуальные проблемы современной медицины 2008 : материалы Междунар. науч. конф. Минск : БГМУ, 2008. С. 501–505.
23. *Сингер, М.* Гены и геномы : в 2 т. / М. Сингер, П. Берг ; пер. с англ. М. : Мир, 1998. 864 с.
24. *Тимофеев-Ресовский, Н. В.* Краткий очерк теории эволюции / Н. В. Тимофеев-Ресовский, Н. Н. Воронцов, А. В. Яблоков. М. : Наука, 1969. 408 с.
25. *Тимченко, В. В.* Детские инфекционные болезни / В. В. Тимченко. СПб. : Лань, 2006. 521 с.
26. *Титов, Л. П.* Иммунология : терминологический словарь / Л. П. Титов. Минск : БГМУ, 2002. 213 с.
27. *Фрейд, З.* Воспоминания Леонардо да Винчи о раннем детстве / З. Фрейд ; пер. с нем. Р. Додельцева, М. Попова. СПб. : Азбука-классика, 2007. 224 с.
28. *Основные* методы анализа эволюционных отношений между последовательностями генетических макромолекул / В. В. Хрусталев [и др.] // *Здравоохр.* 2005. № 8. С. 11–13.
29. *Хрусталев, В. В.* Частота использования претерминальных кодонов в матричных РНК бактериальных аденилатциклаз / В. В. Хрусталев, Е. В. Барковский // *Здравоохр.* 2006. № 2. С. 17–20.
30. *Хрусталёв, В. В.* Модель совместного влияния естественного отбора и мутационного давления на аминокислотный состав белков и использование кодонов в мРНК у бактерий. Сообщение 1 / В. В. Хрусталёв, Е. В. Барковский // *Мед. журн.* 2006. № 2. С. 101–105.
31. *Хрусталёв, В. В.* Модель совместного влияния естественного отбора и мутационного давления на аминокислотный состав белков и использование кодонов в мРНК у бактерий. Сообщение 2 / В. В. Хрусталёв, Е. В. Барковский // *Мед. журн.* 2006. № 4. С. 99–103.
32. *Хрусталёв, В. В.* Модель совместного влияния естественного отбора и мутационного давления на аминокислотный состав белков и использование кодонов в мРНК у бактерий. Сообщение 3 / В. В. Хрусталёв, Е. В. Барковский // *Мед. журн.* 2007. № 1. С. 92–96.
33. *Хрусталев, В. В.* Молекулярные механизмы мутагенеза альфа-герпесвирусов / В. В. Хрусталев, Е. В. Барковский // *Современные проблемы инфекционной патологии человека : сб. науч. тр.* Минск : Белпринт, 2008. Вып. I. С. 66–70.
34. *Хрусталев, В. В.* Молекулярный механизм гипермутабельности V3-региона поверхностного гликопротеина gp120 ВИЧ1 / В. В. Хрусталев, Е. В. Барковский // *Гене-*

тика и биотехнология XXI века. Фундаментальные и прикладные аспекты : материалы Междунар. науч. конф. Минск : БГУ, 2008. С. 32–34.

35. *Хрусталев, В. В.* Характеристика мутагенеза вируса иммунодефицита человека 1-го типа / В. В. Хрусталев, Е. В. Барковский // Достижения современной биологии, химии и медицины : сб. науч. тр. участников Республ. конф., посвящ. 100-летию со дня рождения В. А. Бандарина. Минск : Белпринт, 2009. С. 93–104.

36. *Хрусталёв, В. В.* Влияние мутационного давления на ген, кодирующий аденилатциклазу IX типа у человека / В. В. Хрусталёв // Тез. докл. 54-й итоговой студ. науч. конф. с международным участием. М. : МГМСУ, 2006. С. 133–134.

37. *Хрусталёв, В. В.* Влияние мутационного давления на частоты фиксации несинонимичных и синонимичных нуклеотидных замен в кодирующих участках бактериальных аденилатциклаз I класса / В. В. Хрусталёв // Актуальные вопросы молекулярной эволюции и биохимии : материалы Республ. конф., посвящ. 75-летию со дня основания каф. общей химии БГМУ. Минск : БГМУ, 2006. С. 150–153.

38. *Хрусталёв, В. В.* Влияние мутационного давления на гены аденилатциклаз IX типа позвоночных животных / В. В. Хрусталёв // Молодежь в науке – 2006 : сб. тр. Междунар. науч. конф. молодых ученых, Минск, 16–19 окт. 2006 г. Минск : НАН РБ, 2007. Ч. 1. С. 366–369.

39. *Хрусталёв, В. В.* Сильное мутационное GC-давление — причина частого рецидивирования герпетической инфекции / В. В. Хрусталёв // Актуальные проблемы современной медицины–2007 : материалы Междунар. науч. конф. Минск : БГМУ, 2007. С. 609–612.

40. *Хрусталёв, В. В.* Компьютерный алгоритм «CGS» — инструмент для сканирования нуклеотидного состава кодирующих геномов / В. В. Хрусталёв // Актуальные проблемы современной медицины–2008 : материалы Междунар. науч. конф. Минск : БГМУ, 2008. С. 564–568.

41. *Хрусталёв, В. В.* Происхождение гена ICP0 и его роль в реактивации вирусов простого герпеса / В. В. Хрусталёв // Первый шаг в науку – 2007 : сб. мат. Междунар. форума студ. и учащ. молодежи. Минск : НАН РБ, 2009. Ч. 2. С. 192–196.

42. *Шмальгаузен, И. И.* Факторы эволюции (теория стабилизирующего отбора) / И. И. Шмальгаузен ; ред. коллегия : Р. Л. Берг, А. А. Махотин, А. В. Яблоков. М. : Наука, 1968. 451 с.

43. *Andersen, P. H.* Prediction of residues in discontinuous B cell epitopes using protein 3D structures / P. H. Andersen, M. Nielsen, O. Lund // Protein Science. 2006. Vol. 15. P. 2558–2567.

44. *Bailey, J. R.* Neutralizing antibodies do not mediate suppression of human immunodeficiency virus type 1 in elite suppressors or selection of plasma virus variants in patients on highly active antiretroviral therapy / J. R. Bailey, K. G. Lassen, H. C. Yang // J. Virol. 2006. Vol. 80. P. 4758–4770.

45. *Barkovsky, E. V.* Inverse correlation between the GC content of bacterial genomes and their level of preterminal codon usage / E. V. Barkovsky, V. V. Khrustalev // Molecular Genetics, Microbiology and Virology. 2009. Vol. 24. P. 17–23.

46. *Bell, P.* Accumulation and intranuclear distribution of unintegrated human immunodeficiency virus type 1 DNA / P. Bell, L. J. Montaner, G. G. Maul // J. Virol. 2001. Vol. 75. P. 7683–7691.

47. *Bellamy, S. R. W.* A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1 / S. R. W. Bellamy, G. S. Baldwin // Nucleic Acids Res. 2001. Vol. 29. P. 3857–3863.

48. *Berkhout, B.* The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins / B. Berkhout, F. J. van Hemert // Nucl Acid Res. 1994. Vol. 22. P. 1705–1711.

49. *Berkhout, B.* APOBEC3G versus reverse transcriptase in the generation of HIV-1 drug-resistance mutations / B. Berkhout, A. de Ronde // AIDS. 2004. Vol. 18. P. 1861–1863.

50. *Bernardi, G.* Isochores and the evolutionary genomics of vertebrates / G. Bernardi // Gene. 2000. Vol. 241. P. 3–17.

51. *APOBEC-mediated* editing of viral RNA / K. N. Bishop [et al.] // Science. 2004. Vol. 305. P. 645.

52. *Role of emm and mrp genes in the virulence of group A streptococcal isolate 64/14 in a mouse model of skin infection* / M. D. Boyle [et al.] // J. Infect. Dis. 1998. Vol. 177. P. 991–997.

53. *Bourara, K.* Target cell APOBEC3C can induce limited G-to-A mutation in HIV-1 / K. Bourara, T. J. Liegler, R. M. Grant // PLoS Pathog. 2007. Vol. 3. N. 153.

54. *Antibody levels to the class I and II epitopes of the M protein and myosin are related to group A streptococcal exposure in endemic populations* / E. R. Brandt [et al.] // International Immunology. 2001. Vol. 13. P. 1335–1343.

55. *Genetic and biological analysis of variants derived from the cerebrospinal fluid of HIV type 1 subtype B- and D-infected patients with and without AIDS dementia complex* / K. Brengel-Pesce [et al.] // AIDS Res. Hum. Retroviruses. 1996. Vol. 12. P. 1643–1645.

56. *Brown, J. R.* Archaea and the prokaryote-to-eukaryote transition / J. R. Brown, W. F. Doolittle // Microbiology and Molecular Biology Reviews. 1997. Vol. 61. P. 456–502.

57. *Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection* / E. M. Bunnik [et al.] // J. Virol. 2008. Vol. 82. P. 7932–7941.

58. *Chou, P. Y.* Prediction of the secondary structure of proteins from their amino acid sequence / P. Y. Chou, G. D. Fasman // Adv. Enzymol. Relat. Areas Mol. Biol. 1978. Vol. 47. P. 45–148.

59. *Double-stranded RNA as a not-self alarm signal : to evade, most viruses purine-load their RNAs, but some (HTLV-1, Epstein-barr) pyrimidine-load* / A. D. Cristillo [et al.] // Journal of Theoretical Biology. 2001. Vol. 208. P. 475–489.

60. *Genomic plasticity in prokaryotes: the case of the square haloarchaeon* / S. Cuadros-Orellana [et al.] // The ISME Journal. 2007. Vol. 1. P. 235–245.

61. *Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family* / Y. Dang [et al.] // J. Virol. 2006. Vol. 80. P. 10522–10533.

62. *Deichman, A. M.* RNA-editing. Hypothetical mechanisms / A. M. Deichman, W. C. Choi, A. Yu. Baryshnikov. M. : Practical medicine, 2005. 266 p.

63. *Einolf, H. J.* Steady-state and pre-steady-state kinetic analysis of 8-oxo-7,8-dihydro-guanosine triphosphate incorporation and extension by replicative and repair DNA polymerases / H. J. Einolf, N. Schnetz-Boutaud, F. P. Guengerich // Biochemistry. 1998. Vol. 37. P. 13300–13312.

64. *Existence of two emm-like «mrp» and «emm» genes in the mga regulon of the Streptococcus pyogenes strain ST4547* / M. Eshaghi [et al.] // J. Biochem. Mol. Biol. Biophys. 2002. Vol. 6. P. 23–28.

65. *Forsdyke, D. R.* Evolutionary Bioinformatics / D. R. Forsdyke. New York : Springer, 2006. 273 p.

66. *Suppression of dualtropic human immunodeficiency virus type 1 by the CXCR4 antagonist AMD3100 is associated with efficiency of CXCR4 use and baseline virus composition* / S. Fransen [et al.] // Antimicrob Agents Chemother. 2008. Vol. 52. P. 2608–2615.

67. Fryxell, K. J. CpG mutation rates in the human genome are highly dependent on local GC content / K. J. Fryxell, W. J. Moon // *Molecular Biology and Evolution*. 2004. Vol. 22. P. 650–658.
68. Fryxell, K. J. Cytosine deamination plays a primary role in the evolution of mammalian isochores / K. J. Fryxell, E. Zuckerkandl // *Molecular Biology and Evolution*. 2000. Vol. 17. P. 1371–1383.
69. Fu, L. H. A selenoprotein in the plant kingdom / L. H. Fu, X. Wang, Y. Eyal // *Biol. Chem*. 2002. Vol. 277. P. 25983–25991.
70. Galagan, J. E. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity / J. E. Galagan // *Genome Research*. 2002. Vol. 12. P. 532–542.
71. Viral complementation allows HIV-1 replication without integration / H. C. Gelderblom [et al.] // *Retrovir*. 2008. Vol. 5. N. 60.
72. Gros, L. Enzymology of the repair of free radicals-induced DNA damage / L. Gros, M. K. Saparbaev, J. Laval // *Oncogene*. 2002. Vol. 21. P. 8905–8925.
73. Hinnebusch, J. Linear plasmids and chromosomes in bacteria / J. Hinnebusch, K. Tilly // *Mol. Microbiol*. 1993. Vol. 10. P. 917–922.
74. Hodge, T. A myosin family tree / T. Hodge, M. J. Cope // *J. Cell. Sci*. 2000. Vol. 113. P. 3353–3354.
75. Hopp, T. P. Protein antigen conformation : folding patterns and predictive algorithms ; selection of antigenic and immunogenic peptides / T. P. Hopp // *Ann. Sclavo*. 1984. Vol. 2. P. 47–60.
76. Hopp, T. P. A computer program for predicting protein antigenic determinants / T. P. Hopp, K. R. Woods // *Mol. Immunol*. 1983. Vol. 20. P. 483–489.
77. Izumi, T. Cytidine deaminases as a weapon against retroviruses and a new target for antiviral therapy / T. Izumi, K. Shirakawa, A. Takaori-Kondo // *Mini Rev. Med. Chem*. 2008. Vol. 8. P. 231–238.
78. *HSV-1* DNA in tears and saliva of normal adults / H. E. Kaufman [et al.] // *Invest. Ophthalmol. Vis. Sci*. 2005. Vol. 46. P. 241–247.
79. *Eukaryotic* genome evolution through the spontaneous duplication of large chromosomal segments / R. Koszul [et al.] // *EMBO J*. 2004. Vol. 23. P. 234–243.
80. Khrustalev, V. V. Levels of CpG and GpC dinucleotides in coding districts of archaeal genomes / V. V. Khrustalev, E. V. Barkovsky // *Computational Phylogenetics and Molecular Systematics «CPMS' 2007»*. Conference proceedings. Moscow : KMK Scientific Press Ltd, 2007. P. 354–357.
81. Khrustalev, V. V. Mutational Pressure in Genomes of Human Alphaherpesviruses / V. V. Khrustalev, E. V. Barkovsky // *Molecular Genetics, Microbiology and Virology*. 2008. Vol. 23. P. 94–100.
82. Khrustalev, V. V. An in-silico study of alphaherpesviruses ICP0 genes : Positive selection or strong mutational GC-pressure? / V. V. Khrustalev, E. V. Barkovsky // *International Union of Biochemistry and Molecular Biology Life*. 2008. Vol. 60. P. 456–460.
83. Khrustalev, V. V. Main pathways of proteome simplification in alphaherpesviruses under the influence of the strong mutational GC-pressure / V. V. Khrustalev, E. V. Barkovsky // *Journal of Proteomics and Bioinformatics*. 2009. Vol. 2. P. 88–96.
84. Khrustalev, V. V. Mutational pressure is a cause of inter- and intragenomic differences in GC-content of simplex and varicello viruses / V. V. Khrustalev, E. V. Barkovsky // *Computational Biology and Chemistry*. 2009. Vol. 33. P. 295–302.
85. Khrustalev, V. V. Common predecessor's effect in archaeal genomes and proteomes / V. V. Khrustalev, E. V. Barkovsky // *Proceedings of the 4-th International Moscow Conference on Computational Molecular Biology «MCCMB'09»*. Moscow : MSU, 2009. P. 163–164.

86. *Khrustalev, V. V.* HIV1 V3 loop hypermutability is enhanced by the guanine usage bias in the part of *env* gene coding for it / V. V. Khrustalev // *In Silico Biology*. 2009. Vol. 9. N. 4. P. 255–269.

87. *Khrustalev, V. V.* Can mutational GC-pressure create new linear B-cell epitopes in herpes simplex virus type 1 glycoprotein B? / V. V. Khrustalev // *Immunological Investigations*. 2009. Vol. 38. P. 613–623.

88. *Kozak, M.* Recognition of AUG and alternative initiator codons is augmented by G in position F4 but is not generally affected by the nucleotides in positions F5 and F6 / M. Kozak // *The EMBO Journal*. 1997. Vol. 16. P. 2482–2492.

89. *Analysis of the IgM and IgG antibody response against herpes simplex virus type 1 (HSV-1) structural and nonstructural proteins* / J. E. Kuhn [et al.] // *J. Med. Virol.* 1987. Vol. 23. P. 135–150.

90. *Larsen, J. E. P.* Improved method for predicting linear B-cell epitopes / J. E. P. Larsen, O. Lund, M. Nielsen // *Immunome Res.* 2006. Vol. 2. № 2.

91. *Lee, S. J.* Genomic conflict settled in favour of the species rather than of the gene at extreme GC% values / S. J. Lee, J. R. Mortimer, D. R. Forsdyke // *Applied Bioinformatics*. 2004. Vol. 3. P. 219–228.

92. *Lehman, I. R.* Replication of Herpes Simplex Virus DNA / I. R. Lehman, P. E. Boehmer // *J. Biol. Chem.* 1999. Vol. 274. P. 28059–28062.

93. *Liebler, D. C.* Introduction to proteomics : tools for the new biology / D. C. Liebler. New Jersey : Humana Press, 2002. 541 p.

94. *Li, Z.* RNA damage and surveillance under oxidative stress / Z. Li, J. Wu, C. J. Deleo // *IUBMB Life*. 2006. Vol. 58. P. 581–588.

95. *APOBEC3F* properties and hypermutation preferences indicate activity against HIV-1 in vivo / M. T. Liddament [et al.] // *Curr. Biol.* 2004. Vol. 14. P. 1385–1391.

96. *Lobry, J. R.* Asymmetric directional mutation pressures in bacteria / J. R. Lobry, N. Sueoka // *Genome Biol.* 2002. Vol. 3. № 0058.

97. *McGeoch, D. J.* Molecular phylogeny of the *Alphaherpesvirinae* subfamily and a proposed evolutionary timescale / D. J. McGeoch, S. Cook // *J. Mol. Biol.* 1994. Vol. 238. P. 9–22.

98. *Coiled-Coil Irregularities and Instabilities in Group A Streptococcus M1 are Required for Virulence* / C. McNamara [et al.] // *Science*. 2008. Vol. 319. P. 1405–1408.

99. *Modiano, G.* Nonrandom patterns of codon usage and of nucleotide substitutions in human  $\alpha$ - and  $\beta$ -globin genes: an evolutionary strategy reducing the rate of mutations with drastic effects? / G. Modiano, G. Baffistuzzi, A. G. Motulsky // *Proc. Natl. Acad. Sci. USA*. 1981. Vol. 78. P. 1110–1114.

100. *Moe, A.* Incision at hypoxanthine residues in DNA by a mammalian homologue of the *Escherichia coli* antimutator enzyme endonuclease V / A. Moe, J. Ringvoll, L. M. Nordstrand // *Nucleic Acids Res.* 2003. Vol. 31. P. 3893–3900.

101. *Moore, J. P.* Antibodies to discontinuous or conformationally sensitive epitopes on the gp120 glycoprotein of human immunodeficiency virus type 1 are highly prevalent in sera of infected humans / J. P. Moore, D. D. Ho // *J. Virol.* 1993. Vol. 67. P. 863–875.

102. *Activation-induced* cytidine deaminase deaminates 5-Methylcytosine in DNA and is expressed in pluripotent tissues / H. D. Morgan [et al.] // *J. Biol. Chem.* 2004. Vol. 279. P. 52353–52360.

103. *Muto, A.* The guanine and cytosine content of genomic DNA and bacterial evolution / A. Muto, S. Osawa // *Proc. Natl. Acad. Sci. USA*. 1987. Vol. 84. P. 166–169.

104. *Nakamura, Y.* Codon usage tabulated from the international DNA sequence databases : status for the year 2000 / Y. Nakamura, T. Gojobori, T. Ikemura // Nucl. Acids Res. 2000. Vol. 28. P. 292.
105. *Nei, M.* Molecular evolution and phylogenetics / M. Nei, S. Kumar. New York : Oxford University Press, 2000. 333 p.
106. *Neuberger, M. S.* Somatic hypermutation at A·T pairs : polymerase error versus dUTP incorporation / M. S. Neuberger, J. M. Di Noia, R. C. L. Beale // Nature Reviews Immunology. 2005. Vol. 5. P. 171–178.
107. *Osawa, S.* Recent evidence for evolution of the genetic code / S. Osawa, T. H. Jukes, K. Watanabe, A. Muto // Microbiol Rev. 1992. Vol. 56. P. 229–264.
108. *Murine APOBEC1* is a powerful mutator of retroviral and cellular RNA in vitro and in vivo / V. Petit [et al.] // J. Mol. Biol. 2009. Vol. 385. P. 65–78.
109. *Pillai, S. K.* Turning up the volume on mutational pressure : Is more of a good thing always better? (A case study of HIV-1 Vif and APOBEC3) / S. K. Pillai, J. K. Wong, J. D. Barbour // Retrovir. 2008. Vol. 5. № 26.
110. *Poon, B.* Human immunodeficiency virus type 1 (HIV-1) *vpr* enhances expression from unintegrated HIV-1 DNA / B. Poon, I. S. Y. Chen // J. Virol. 2003. Vol. 77. P. 3962–3972.
111. *A cytidine deaminase* edits C to U in transfer RNAs in archaea / L. Randau [et al.] // Science. 2009. Vol. 324. P. 657–659.
112. *Biological and Genetic Evolution of HIV Type 1 in Two Siblings with Different Patterns of Disease Progression* / C. Ripamonti [et al.] // AIDS Res. Hum. Retroviruses. 2007. Vol. 23. P. 1531–1540.
113. *Sagar, M.* Selection of HIV variants with signature genotypic characteristics during heterosexual transmission / M. Sagar, O. Laeyendecker, S. Lee // J. Infect. Dis. 2009. Vol. 199. P. 580–589.
114. *Inpatient* variability of the human immunodeficiency virus type 2 envelope V3 loop / J. L. Sankale [et al.] // AIDS Res. Hum. Retroviruses. 1995. Vol. 11. P. 617–623.
115. *Shikanai, T.* RNA editing in plant organelles: machinery, physiological function and evolution. / T. Shikanai // Cell. Mol. Life. Sci. 2006. Vol. 63. P. 698–708.
116. *Singer, G. A. C.* Nucleotide bias causes a genomewide bias in the amino acid composition of proteins / G. A. C. Singer, D. A. Hickey // Mol. Biol. Evol. 2000. Vol. 17. P. 1581–1588.
117. *Genome* sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks / J. C. Smoot [et al.] // Proc. Natl. Acad. Sci. USA. 2002. Vol. 99. P. 4668–4673.
118. *The residue* mass of L-pyrrolysine in three distinct methylamine methyltransferases / J. A. Soares [et al.] // J. Biol. Chem. 2005. Vol. 280. P. 36962–36969.
119. *Study* of the V3 loop as a target epitope for antibodies involved in the neutralization of primary isolates versus T-cell-line-adapted strains of human immunodeficiency virus type 1 / C. Spenlehauer [et al.] // J. Virol. 1998. Vol. 72. P. 9855–9864.
120. *Sueoka, N.* Compositional correlation between deoxyribonucleic acid and protein / N. Sueoka // Cold Spring Harb Symp Quant Biol. 1961. Vol. 26. P. 35–43.
121. *Sueoka, N.* Directional mutation pressure and neutral molecular evolution / N. Sueoka // Proc. Natl. Acad. Sci. USA. 1988. Vol. 85. P. 2653–2657.
122. *Sueoka, N.* Wide intra-genomic G+C heterogeneity in human and chicken is mainly due to strand-symmetric directional mutation pressures: dGTP-oxidation and symmetric cytosine-deamination hypotheses / N. Sueoka // Gene. 2002. Vol. 300. P. 141–154.

123. *An Introduction to Genetic Analysis* 4<sup>th</sup> ed. / D. T. Suzuki [et al.]. New York : Freeman W. H., 1989. 521 p.
124. *MEGA4* : Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 / K. Tamura [et al.] // *Mol. Biol. Evol.* 2007. Vol. 24. P. 1596–1599.
125. *The DNA* sequence of equine herpesvirus-4 / E. A. R. Telford [et al.] // *J. Gen. Virol.* 1998. Vol. 79. P. 1197–1203.
126. *Umene, K.* Conversion of a fraction of the unique sequence to part of the inverted repeats in the S component of the Herpes Simplex Virus Type 1 genome / K. Umene // *J. Gen. Virol.* 1986. Vol. 67. P. 1035–1048.
127. *Selection, recombination, and G→A* hypermutation of human immunodeficiency virus type 1 genomes / J. P. Vartanian [et al.] // *J. Virol.* 1991. Vol. 65. P. 1779–1788.
128. *Yamaguchi, Y.* Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts / Y. Yamaguchi, T. Gojobori // *Proc. Natl. Acad. Sci. USA.* 1997. Vol. 94. P. 1264–1269.
129. *Yang, W.* Widespread adaptive evolution in the human immunodeficiency virus type 1 genome / W. Yang, J. P. Bielawski, Z. Yang // *J. Mol. Evol.* 2003. Vol. 57. P. 212–221.
130. *Yoon, J. H.* Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg : G mispair / J. H. Yoon // *Nucleic Acids Research.* 2003. Vol. 31. P. 5399–5404.
131. *Williams, G. C.* Pleiotropy, natural selection, and the evolution of senescence / G. C. Williams // *Evolution.* 1957. Vol. 11. P. 398–411.
132. *Williamson, S.* Adaptation in the *env* gene of HIV-1 and evolutionary theories of disease progression / S. Williamson // *Mol. Biol. Evol.* 2003. Vol. 20. P. 1318–1325.
133. *Wu, Y.* The second chance story of HIV-1 DNA : Unintegrated? Not a problem! / Y. Wu // *Retrovir.* 2008. Vol. 5. № 61.
134. *The cytidine* deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA / H. Zhang [et al.] // *Nature.* 2003. Vol. 424. P. 94–98.
135. *Zeeberg, B.* Shannon information theoretic computation of synonymous codon usage biases in coding regions of human and mouse genomes / B. Zeeberg // *Gen. Res.* 2002. Vol. 12. P. 944–955.



Таблица универсального генетического кода

<b>Phe</b> <u>TTT</u> <u>TTC</u>	<b>Ser4</b> TCT TCC	<b>Tyr</b> <u>TAT</u> <u>TAC</u>	<b>Cys</b> TGT TGC
<b>Leu2</b> TTA TTG	TCA TCG	<b>Term</b> TAA TAG	<b>Term</b> TGA <b>Trp</b> TGG
<b>Leu4</b> CTT CTC CTA CTG	<b>Pro</b> <i>CCT</i> <i>CCC</i> <i>CCA</i> <i>CCG</i>	<b>His</b> CAT CAC <b>Gln</b> CAA CAG	<b>Arg4</b> <i>CGT</i> <i>CGC</i> <i>CGA</i> <i>CGG</i>
<b>Ile</b> <u>ATT</u> <u>ATC</u> <u>ATA</u>	<b>Thr</b> ACT ACC ACA	<b>Asn</b> <u>AAT</u> <u>AAC</u>	<b>Ser2</b> AGT AGC
<b>Met</b> <u>ATG</u>	ACG	<b>Lys</b> <u>AAA</u> <u>AAG</u>	<b>Arg2</b> <i>AGA</i> <i>AGG</i>
<b>Val</b> GTT GTC GTA GTG	<b>Ala</b> <i>GCT</i> <i>GCC</i> <i>GCA</i> <i>GCG</i>	<b>Asp</b> GAT GAC <b>Glu</b> GAA GAG	<b>Gly</b> <i>GGT</i> <i>GGC</i> <i>GGA</i> <i>GGG</i>

*Примечание.* Подчеркнутым шрифтом обозначены кодоны, кодирующие аминокислотные остатки из группы FYMINK; курсивом обозначены кодоны, кодирующие аминокислотные остатки из группы GARP; кодоны, кодирующие аминокислотные остатки из группы 10AA, не имеют шрифтового выделения.

**Таблица универсального генетического кода**

<b>Phe</b> <u>TTT</u> <u>TTC</u>	<b>Ser4</b> <i>TCT</i> <i>TCC</i>	<b>Tyr</b> <u>TAT</u> <u>TAC</u>	<b>Cys</b> <u>TGT</u> <u>TGC</u>
<b>Leu2</b> <u>TTA</u> <u>TTG</u>	<i>TCA</i> <i>TCG</i>	<b>Term</b> <u>TAA</u> <u>TAG</u>	<b>Term</b> TGA <b>Trp</b> TGG
<b>Leu4</b> <i>CTT</i> <i>CTC</i> <i>CTA</i> <i>CTG</i>	<b>Pro</b> <i>CCT</i> <i>CCC</i> <i>CCA</i> <i>CCG</i>	<b>His</b> <u>CAT</u> <u>CAC</u> <b>Gln</b> <u>CAA</u> <u>CAG</u>	<b>Arg4</b> <i>CGT</i> <i>CGC</i> <i>CGA</i> <i>CGG</i>
<b>Ile</b> ATT ATC ATA	<b>Thr</b> <i>ACT</i> <i>ACC</i> <i>ACA</i>	<b>Asn</b> <u>AAT</u> <u>AAC</u>	<b>Ser2</b> <u>AGT</u> <u>AGC</u>
<b>Met</b> ATG	<i>ACG</i>	<b>Lys</b> <u>AAA</u> <u>AAG</u>	<b>Arg2</b> <u>AGA</u> <u>AGG</u>
<b>Val</b> <i>GTT</i> <i>GTC</i> <i>GTA</i> <i>GTG</i>	<b>Ala</b> <i>GCT</i> <i>GCC</i> <i>GCA</i> <i>GCG</i>	<b>Asp</b> <u>GAT</u> <u>GAC</u> <b>Glu</b> <u>GAA</u> <u>GAG</u>	<b>Gly</b> <i>GGT</i> <i>GGC</i> <i>GGA</i> <i>GGG</i>

*Примечание.* Подчеркнутым шрифтом обозначены кодоны, содержащие двукратно вырожденные сайты в третьих положениях; курсивом обозначены кодоны, содержащие четырехкратно вырожденные сайты в третьих положениях.

Таблица универсального генетического кода

<b>Phe</b> TTT TTC	<b>Ser4</b> TCT TCC	<b>Tyr</b> <u>TAT</u> <u>TAC</u>	<b>Cys</b> <u>TGT</u> <u>TGC</u>
<b>Leu2</b> <u>TTA</u> <u>TTG</u>	<u>TCA</u> <u>TCG</u>	<b>Term</b> TAA TAG	<b>Term</b> TGA <b>Trp</b> <u>TGG</u>
<b>Leu4</b> CTT CTC CTA CTG	<b>Pro</b> CCT CCC CCA CCG	<b>His</b> CAT CAC <b>Gln</b> <u>CAA</u> <u>CAG</u>	<b>Arg4</b> CGT CGC <u>CGA</u> CGG
<b>Ile</b> ATT ATC ATA	<b>Thr</b> ACT ACC ACA	<b>Asn</b> AAT AAC	<b>Ser2</b> AGT AGC
<b>Met</b> ATG	ACG	<b>Lys</b> <u>AAA</u> <u>AAG</u>	<b>Arg2</b> <u>AGA</u> AGG
<b>Val</b> GTT GTC GTA GTG	<b>Ala</b> GCT GCC GCA GCG	<b>Asp</b> GAT GAC <b>Glu</b> <u>GAA</u> <u>GAG</u>	<b>Gly</b> GGT GGC <u>GGA</u> GGG

*Примечание.* Подчеркнутым курсивом обозначены претерминальные кодоны, подчеркнутым жирным курсивом обозначены претерминальные кодоны, способные стать терминальными в результате замены по направлению АТ-давления (GC на AT).

## Содержание

Введение .....	3
Часть первая. Теория мутационного давления.....	5
Глава 1. Предмет и задачи молекулярной биологии, геномики, протеомики, биоинформатики и вычислительной биологии .....	5
Глава 2. Теория нейтральности М. Кимуры .....	8
Глава 3. Теория мутационного давления Н. Суеоки.....	17
Глава 4. Понятие о силе мутационного давления .....	27
Глава 5. Последствия мутационного давления .....	30
Часть вторая. Мутационное давление в генах и геномах.....	37
Глава 1. Анализ частот нуклеотидных замен между генами, кодирующими отдельные белки герпесвирусов (вирусов простого герпеса человека и обезьяны, варицелло-зостер-вирусов человека и обезьяны).....	37
Глава 2. Анализ полных геномов альфа-герпесвирусов .....	47
Глава 3. Анализ выборки нуклеотидных последовательностей гена <i>env</i> , полученных от разных штаммов и клонов вируса иммунодефицита человека 1-го типа .....	57
Часть третья. Уровень мутабельности кодирующих участков.....	67
Глава 1. Поиск участков с высокой и низкой мутабельностью в генах <i>env</i> вирусов иммунодефицита человека 1-го и 2-го типа.....	67
Глава 2. Предсказание уровня мутабельности пространственных эпитопов поверхностного белка gp120 вируса иммунодефицита человека 1-го типа.....	78
Глава 3. Расчет риска нонсенс-мутации в гене, кодирующем М-протеин <i>Streptococcus pyogenes</i> .....	86
Часть четвертая. Степень селекционных ограничений, накладываемых на аминокислотный состав белка .....	94
Глава 1. Анализ механизмов упрощения аминокислотного состава препроинсулина хордовых животных под воздействием GC-давления .....	94
Глава 2. Анализ механизмов упрощения аминокислотного состава протеомов альфа-герпесвирусов под воздействием GC-давления .....	114

Глава 3. Анализ механизмов упрощения аминокислотного состава отдельных белков альфа-герпесвирусов под воздействием GC-давления .....	127
Часть пятая. Уровень изменчивости отдельных участков белка .....	137
Глава 1. Подтверждение результатов предсказания уровня мутабельности пространственных эпитопов поверхностного белка gp120 вируса иммунодефицита человека 1-го типа.....	137
Часть шестая. Описание оригинальных компьютерных алгоритмов .....	150
Глава 1. Алгоритм «CodonChanges».....	150
Глава 2. Алгоритм «VVK Dinucleotides» .....	151
Глава 3. Алгоритм «VVK in group» .....	151
Глава 4. Алгоритм «Coding Genome Scanner».....	152
Глава 5. Алгоритм «VVK Consensus» .....	152
Глава 6. Алгоритм «VVK in length» .....	153
Глоссарий .....	154
Литература .....	200
Приложение 1 .....	208
Приложение 2 .....	209
Приложение 3 .....	210